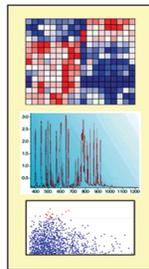


Dietary Restrictions



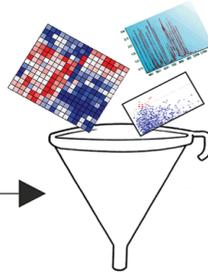
CR, VLCD,  
IF, ADF

Health-related Biomarkers



Outcomes Inclusion

Meta-Analyses



Geroprotective Footprint  
(Biomarkers for Healthy Aging)

Effect Size  
*Fixed or Random  
effect model*



PRECISION MEDICINE

## Special Collection on Biomarkers

Algorithm development for biomarkers validation of dietary restriction in human, see Lettieri-Barbato et al. - "Effects of dietary restriction on adipose mass and biomarkers of healthy aging in human."

**Biomarkers**

**AGING**

# AGING

www.aging-us.com

## EDITORIAL BOARD

### EDITORS-IN-CHIEF

Jan Vijn - Albert Einstein College of Medicine, Bronx, NY, USA

David A. Sinclair - Harvard Medical School, Boston, MA, USA

Vera Gorbunova - University of Rochester, Rochester, NY, USA

Judith Campisi - The Buck Institute for Research on Aging, Novato, CA, USA

Mikhail V. Blagosklonny - Roswell Park Cancer Institute, Buffalo, NY, USA

### EDITORIAL BOARD

Frederick Alt - Harvard Medical School, Boston, MA, USA

Vladimir Anisimov - Petrov Institute of Oncology, St.Petersburg, Russia

Johan Auwerx - Ecole Polytechnique Federale de Lausanne, Switzerland

Andrzej Bartke - Southern Illinois University, Springfield, IL, USA

Nir Barzilai - Albert Einstein College of Medicine, Bronx, NY, USA

Elizabeth H. Blackburn - University of California, San Francisco, CA, USA

Maria Blasco - Spanish National Cancer Center, Madrid, Spain

Vilhelm A. Bohr - National Institute on Aging, NIH, Baltimore, MD, USA

William M. Bonner - National Cancer Institute, NIH, Bethesda, MD, USA

Robert M. Brosh, Jr. - National Institute on Aging, NIH, Baltimore, MD, USA

Anne Brunet - Stanford University, Stanford, CA, USA

Rafael de Caba - NIA, NIH, Baltimore, MD, USA

Ronald A. DePinho - Dana-Farber Cancer Institute, Boston, MA, USA

Jan van Deursen - Mayo Clinic, Rochester, MN, USA

Lawrence A. Donehower - Baylor College of Medicine, Houston, TX, USA

Caleb E. Finch - University of Southern California, Los Angeles, CA, USA

Toren Finkel - National Institutes of Health, Bethesda, MD, USA

Luigi Fontana - Washington University, St. Louis, MO, USA

Claudio Franceschi - University of Bologna, Bologna, Italy

David Gems - Inst. of Healthy Ageing, Univ. College London, UK

Myriam Gorospe - National Institute on Aging, NIH, Baltimore, MD, USA

Leonard Guarente - MIT, Cambridge, MA, USA

Andrei Gudkov - Roswell Park Cancer Institute, Buffalo, NY, USA

Michael Hall - University of Basel, Basel, Switzerland

Philip Hanawalt - Stanford University, CA, USA

Nissim Hay - University of Illinois at Chicago, Chicago, IL, USA

**Siegfried Hekimi** - McGill University, Montreal, Canada  
**Stephen L. Helfand** - Brown University, Providence, RI, USA  
**Jan H.J. Hoeijmakers** - Erasmus MC, Rotterdam, The Netherlands  
**John O. Holloszy** - Washington University, St. Louis, MO, USA  
**Stephen P. Jackson** - University of Cambridge, Cambridge, UK  
**Heinrich Jasper** - The Buck Institute for Research on Aging, Novato, CA, USA  
**Pankaj Kapahi** - The Buck Institute for Research on Aging, Novato, CA, USA  
**Jan Karlseder** - The Salk Institute, La Jolla, CA, USA  
**Cynthia Kenyon** - University of California San Francisco, San Francisco, CA, USA  
**James L. Kirkland** - Mayo Clinic, Rochester, MN, USA  
**Guido Kroemer** - INSERM, Paris, France  
**Titia de Lange** - Rockefeller University, New York, NY, USA  
**Arnold Levine** - The Institute for Advanced Study, Princeton, NJ, USA  
**Michael P. Lisanti** - University of Salford, Salford, UK  
**Lawrence A. Loeb** - University of Washington, Seattle, WA, USA  
**Valter Longo** - University of Southern California, Los Angeles, CA, USA  
**Gerry Melino** - University of Rome, Rome, Italy  
**Simon Melov** - The Buck Institute for Research on Aging, Novato, CA, USA  
**Alexey Moskalev** - Komi Science Center of RAS, Syktyvkar, Russia  
**Masashi Narita** - University of Cambridge, Cambridge, UK  
**Andre Nussenzweig** - National Cancer Institute, NIH, Bethesda, MD, USA  
**William C. Orr** - Southern Methodist University, Dallas, TX, USA  
**Daniel S. Peepker** - The Netherlands Cancer Institute, Amsterdam, The Netherlands  
**Thomas Rando** - Stanford University School of Medicine, Stanford, CA, USA  
**Michael Ristow** - Swiss Federal Institute of Technology, Zurich, Switzerland  
**Igor B. Roninson** - Ordway Research Institute, Albany, NY, USA  
**Michael R. Rose** - University of California, Irvine, CA, USA  
**K Lenhard Rudolph** - Hannover Medical School, Hannover, Germany  
**Paolo Sassone-Corsi** - University of California, Irvine, CA, USA  
**John Sedivy** - Brown University, Providence, RI, USA  
**Manuel Serrano** - Spanish National Cancer Research Center, Madrid, Spain  
**Gerald S. Shadel** - Yale University School of Medicine, New Haven, CT, USA  
**Norman E. Sharpless** - University of North Carolina, Chapel Hill, NC, USA  
**Vladimir P. Skulachev** - Moscow State University, Moscow, Russia  
**Sally Temple** - NY Neural Stem Cell Institute, Albany, NY, USA  
**George Thomas** - University of Cincinnati, Cincinnati, OH, USA  
**Jonathan L. Tilly** - Massachusetts General Hospital, Boston, MA, USA  
**John Tower** - University of Southern California, LA, CA, USA  
**Eric Verdin** - University of California, San Francisco, CA, USA  
**Thomas von Zglinicki** - Newcastle University, Newcastle, UK  
**Alex Zhavoronkov** - Insilico Medicine, Baltimore, MD, USA

*Aging* (ISSN: 1945 - 4589) is published monthly by Impact Journals, LLC.  
6666 East Quaker St., Suite 1B, Orchard Park, NY 14127

Abstracted and/or indexed in: PubMed/Medline (abbreviated as "Aging (Albany NY)"), PubMed Central (abbreviated as "Aging (Albany NY)"), Web of Science/Science Citation Index Expanded (abbreviated as Aging-US) & listed in the Cell Biology-SCIE and Geriatrics & Gerontology category, Scopus /Rank Q1(the highest rank) (abbreviated as Aging)- Aging and Cell Biology category, Biological Abstracts, BIOSIS Previews, EMBASE, META (Chan Zuckerberg Initiative), Dimensions (Digital Science's).

This publication and all its content, unless otherwise noted, is licensed under CC-BY 3.0 Creative Commons Attribution License.  
Impact Journals, LLC meets Wellcome Trust Publisher requirements.  
IMPACT JOURNALS is a registered trademark of Impact Journals, LLC.



## **Editorial and Publishing Office Aging**

6666 E. Quaker St., Suite 1,  
Orchard Park, NY 14127  
Phone: 1-800-922-0957  
Fax: 1-716-508-8254  
e-Fax: 1-716-608-1380

### **Submission**

Please submit your manuscript on-line at <http://aging.msubmit.net>

### **Editorial**

For editorial inquiries, please call us or email [editors@impactaging.com](mailto:editors@impactaging.com)

### **Production**

For questions related to preparation of your article for publication, please call us or email [krasnova@impactaging.com](mailto:krasnova@impactaging.com)

### **Indexing**

If you have questions about the indexing status of your paper, please email [kurenova@impactaging.com](mailto:kurenova@impactaging.com)

### **Billing/Payments**

If you have questions about billing/invoicing or would like to make a payment, please call us or email [payment@impactaging.com](mailto:payment@impactaging.com)

### **Media**

If you have questions about post publication promotion, Altmetric, video interviews or social media, please email [media@impactjournals.com](mailto:media@impactjournals.com)

### **Printing**

Each issue or paper can be printed on demand. To make a printing request, please call us or email [printing@impactjournals.com](mailto:printing@impactjournals.com)

### **Publisher's Office**

Aging is published by Impact Journals, LLC  
To contact the Publisher's Office, please email: [publisher@impactjournals.com](mailto:publisher@impactjournals.com), visit [www.impactjournals.com](http://www.impactjournals.com), or call 1-800-922-0957

*Aging* (ISSN: 1945 - 4589) is published twice a month by Impact Journals, LLC.  
6666 East Quaker St., Suite 1B, Orchard Park, NY 14127

**Abstracted and/or indexed in:** PubMed/Medline (abbreviated as "Aging (Albany NY)"), PubMed Central (abbreviated as "Aging (Albany NY)"), Web of Science/Science Citation Index Expanded (abbreviated as Aging-US) & listed in the Cell Biology-SCIE and Geriatrics & Gerontology category, Scopus /Rank Q1(the highest rank) (abbreviated as Aging) - Aging and Cell Biology category, Biological Abstracts, BIOSIS Previews, EMBASE, META (Chan Zuckerberg Initiative), Dimensions (Digital Science's).

This publication and all its content, unless otherwise noted, is licensed under CC-BY 3.0 Creative Commons Attribution License.  
Impact Journals, LLC meets Wellcome Trust Publisher requirements.

IMPACT JOURNALS is a registered trademark of Impact Journals, LLC.



# Table of Contents

Effects of dietary restriction on adipose mass and biomarkers of healthy aging in human

[Originally published in Volume 8, Issue 12 pp 3341-3355](#)

Deep biomarkers of aging are population-dependent

[Originally published in Volume 8, Issue 9 pp 2253-2255](#)

Deep biomarkers of human aging: Application of deep neural networks to biomarker development

[Originally published in Volume 8, Issue 5 pp 1021-1030](#)

Telomere length is a prognostic biomarker in elderly advanced ovarian cancer patients: a multicenter GINECO study

[Originally published in Volume 7, Issue 12 pp 1066-1074](#)

Proteomic identification of prognostic tumour biomarkers, using chemotherapy-induced cancer-associated fibroblasts

[Originally published in Volume 7, Issue 10 pp 816-838](#)

What biomarkers (if any) for precise medicine?

[Originally published in Volume 7, Issue 8 pp 533-534](#)

Biomarkers and subtypes of cancer

[Originally published in Volume 7, Issue 5 pp 280-281](#)

HGF-MET as a breast cancer biomarker

[Originally published in Volume 7, Issue 3 pp 150-151](#)

Methylated TRF2 associates with the nuclear matrix and serves as a potential biomarker for cellular senescence

[Originally published in Volume 6, Issue 4 pp 248-263](#)

Plasma microRNA biomarkers for detection of mild cognitive impairment: Biomarker Validation Study

[Originally published in Volume 5, Issue 12 pp 925-938](#)

Specific lipofuscin staining as a novel biomarker to detect replicative and stress-induced senescence. A method applicable in cryo-preserved and archival tissues

[Originally published in Volume 5, Issue 1 pp 37-50](#)

The mitochondria-targeted antioxidant SkQ1 but not N-acetylcysteine reverses aging-related biomarkers in rats

[Originally published in Volume 4, Issue 10 pp 686-694](#)

Plasma microRNA biomarkers for detection of mild cognitive impairment

[Originally published in Volume 4, Issue 9 pp 590-605](#)

Circulatory miR-34a as an RNA-based, noninvasive biomarker for brain aging

[Originally published in Volume 3, Issue 10 pp 985-1002](#)

Recent developments in the use of  $\gamma$ -H2AX as a quantitative DNA double-strand break biomarker

[Originally published in Volume 3, Issue 2 pp 168-174](#)

Notch1/2/3/4 are prognostic biomarker and correlated with immune infiltrates in gastric cancer

[Originally published in Volume 12, Issue 3 pp 2595-2609](#)

Identification of biomarkers related to CD8<sup>+</sup> T cell infiltration with gene co-expression network in clear cell renal cell carcinoma

[Originally published in Volume 12, Issue 4 pp 3694-3712](#)

A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment

[Originally published in Volume 12, Issue 8 pp 6966-6980](#)

PTX3 modulates the immunoflogosis in tumor microenvironment and is a prognostic factor for patients with clear cell renal cell carcinoma

[Originally published in Volume 12, Issue 8 pp 7585-7602](#)

Aberrantly glycosylated integrin  $\alpha$ 3 $\beta$ 1; $\alpha$ 3 $\beta$ 2; $\alpha$ 1 is a unique urinary biomarker for the diagnosis of bladder cancer

[Originally published in Volume 12, Issue 11 pp 10844-10862](#)

SNAI1 is a prognostic biomarker and correlated with immune infiltrates in gastrointestinal cancers

[Originally published in Volume 12, Issue 17 pp 17167-17208](#)

TILRR (FREM1 isoform 2) is a prognostic biomarker correlated with immune infiltration in breast cancer

[Originally published in Volume 12, Issue 19 pp 19335-19351](#)

*TRIM28* is a distinct prognostic biomarker that worsens the tumor immune microenvironment in lung adenocarcinoma

[Originally published in Volume 12, Issue 20 pp 20308-20331](#)

Identification of a nomogram based on an 8-lncRNA signature as a novel diagnostic biomarker for head and neck squamous cell carcinoma

[Originally published in Volume 12, Issue 20 pp 20778-20800](#)

Immune-related biomarker risk score predicts prognosis in prostate cancer

[Originally published in Volume 12, Issue 22 pp 22776-22793](#)

Serum KIAA1199 is an advanced-stage prognostic biomarker and metastatic oncogene in cholangiocarcinoma

[Originally published in Volume 12, Issue 23 pp 23761-23777](#)

Could microtubule inhibitors be the best choice of therapy in gastric cancer with high immune activity: mutant DYNC1H1 as a biomarker

[Originally published in Volume 12, Issue 24 pp 25101-25119](#)

Identification of  $\beta$ -regulation of RhoA activity panel  $\beta$ ; as a prognostic and predictive biomarker for gastric cancer

[Originally published in Volume 13, Issue 1 pp 714-734](#)

KRAS mutations are negatively correlated with immunity in colon cancer

[Originally published in Volume 13, Issue 1 pp 750-768](#)

Heparanase is a novel biomarker for immune infiltration and prognosis in breast cancer

[Originally published in Volume 13, Issue 16 pp 20836-20852](#)

PPM1D is a potential prognostic biomarker and correlates with immune cell infiltration in hepatocellular carcinoma

[Originally published in Volume 13, Issue 17 pp 21294-21308](#)

CCR5 is a prognostic biomarker and an immune regulator for triple negative breast cancer

[Originally published in Volume 13, Issue 20 pp 23810-23830](#)

THRSP identified as a potential hepatocellular carcinoma marker by integrated bioinformatics analysis and experimental validation

[Originally published in Volume 14, Issue 4 pp 1743-1766](#)

Polo-like kinases as potential targets and PLK2 as a novel biomarker for the prognosis of human glioblastoma

[Originally published in Volume 14, Issue 5 pp 2320-2334](#)

Centrosomal protein 290 is a novel prognostic indicator that modulates liver cancer cell ferroptosis via the *Nrf2* pathway

[Originally published in Volume 14, Issue 5 pp 2367-2382](#)

Pan-cancer analysis revealed the significance of the GTPBP family in cancer

[Originally published in Volume 14, Issue 6 pp 2558-2573](#)

F12 as a reliable diagnostic and prognostic biomarker associated with immune infiltration in papillary thyroid cancer

[Originally published in Volume 14, Issue 8 pp 3687-3704](#)

Integrative analysis of expression, prognostic significance and immune infiltration of RFC family genes in human sarcoma

[Originally published in Volume 14, Issue 8 pp 3705-3719](#)

Transcriptional expressions of hsa-mir-183 predicted target genes as independent indicators for prognosis in bladder urothelial carcinoma

[Originally published in Volume 14, Issue 9 pp 3782-3800](#)

Expression status and prognostic significance of mitochondrial dynamics OPA3 in human ovarian cancer

[Originally published in Volume 14, Issue 9 pp 3874-3886](#)

# Effects of dietary restriction on adipose mass and biomarkers of healthy aging in human

Daniele Lettieri-Barbato<sup>1</sup>, Esmeralda Giovannetti<sup>1</sup>, Katia Aquilano<sup>1,2</sup>

<sup>1</sup>Department of Biology, University of Rome Tor Vergata, Rome, Italy

<sup>2</sup>IRCCS San Raffaele La Pisana, Rome, Italy

**Correspondence to:** Daniele Lettieri-Barbato; email: [d.lettieribarbato@hotmail.it](mailto:d.lettieribarbato@hotmail.it)

**Keywords:** adipose tissue, aging, calorie restriction, fasting, biomarkers, human, longevity

**Received:** October 3, 2016

**Accepted:** November 16, 2016

**Published:** November 29, 2016

## ABSTRACT

In developing countries the rise of obesity and obesity-related metabolic disorders, such as cardiovascular diseases and type 2 diabetes, reflects the changes in lifestyle habits and wrong dietary choices. Dietary restriction (DR) regimens have been shown to extend health span and lifespan in many animal models including primates. Identifying biomarkers predictive of clinical benefits of treatment is one of the primary goals of precision medicine. To monitor the clinical outcomes of DR interventions in humans, several biomarkers are commonly adopted. However, a validated link between the behaviors of such biomarkers and DR effects is lacking at present time. Through a systematic analysis of human intervention studies, we evaluated the effect size of DR (i.e. calorie restriction, very low calorie diet, intermittent fasting, alternate day fasting) on health-related biomarkers. We found that DR is effective in reducing total and visceral adipose mass and improving inflammatory cytokines profile and adiponectin/leptin ratio. By analysing the levels of canonical biomarkers of healthy aging, we also validated the changes of insulin, IGF-1 and IGFBP-1,2 to monitor DR effects. Collectively, we developed a useful platform to evaluate the human responses to dietary regimens low in calories.

## INTRODUCTION

Aging and wrong lifestyle choices, including inadequate dietary patterns, increase the risk of developing several diseases such as obesity and its-related chronic degenerative diseases. Interestingly, the aging program can be accelerated by obesity [1]. It is thus likely that obesity reduces life- and health span and plays a predominant role in the onset of age-related diseases [2]. In fact, the prevalence of obesity is globally increasing in populations and has become a burden for healthcare systems. Several studies suggest that dietary restriction (DR) regimens (e.g. intermittent fasting, calorie restriction, low calorie diet) reverse obesity and improve health in human by promoting the same molecular and metabolic adaptations that have been shown in animal models of longevity. In particular, DR in humans ameliorates several metabolic and hormonal factors that are implicated in the pathogenesis of an array of age-associated chronic metabolic diseases [3, 4].

At present it is difficult to evaluate the effectiveness of DR on lifespan in humans, so that several works proposed predictive non-invasive biomarkers to evaluate the geroprotective role of DR. However, a miscellaneous of biomarkers is investigated in human intervention studies limiting the statistical robustness of the data. Whether a “biomarker-based” approach could be suitable for evaluating the effectiveness of DR still remains a matter of debate.

Precision medicine is a medical model that proposes the customization of healthcare, with the identification of predictors that can help to find the effectiveness of health-promoting dietary interventions. Biomarkers represent potentially predictive tools for precision medicine but, although affordable 'omics'-based technology has enabled faster identification of putative biomarkers [5], their validation is still hindered by low statistical power as well as limited reproducibility of results.

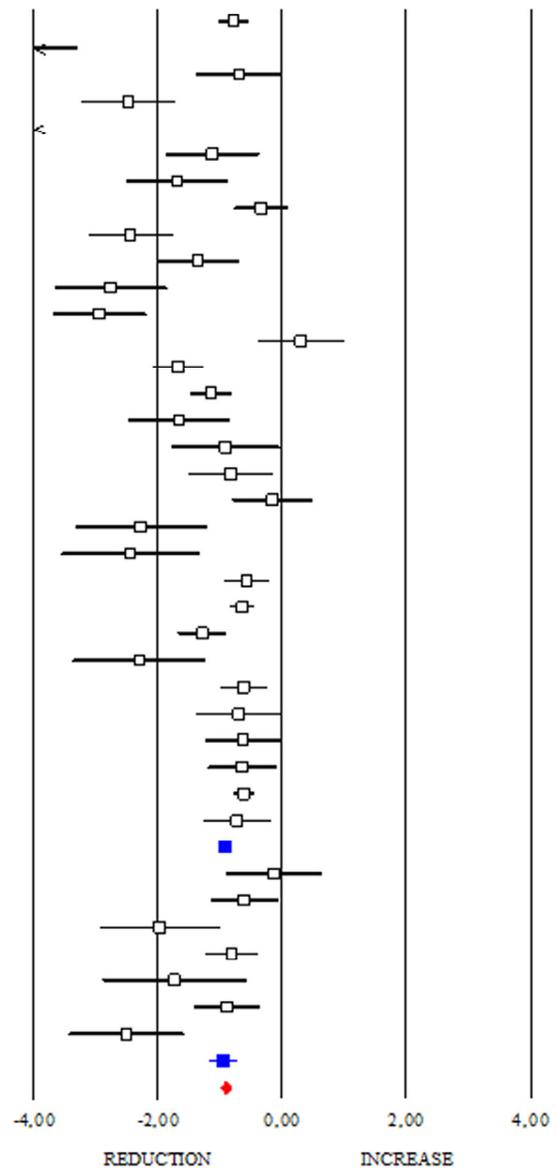
Herein, through meta-analysis we have evaluated the effect size of DR regimens on adipose mass and well-recognized biomarkers of healthy aging. Overall findings provide the geroprotective footprint of DR in humans and highlight a useful platform to validate or monitor the efficiency of dietary treatments to preserve and improve health span and longevity.

## RESULTS

### Effects of DR on total and visceral adipose mass

DR regimens are effective in slowing aging, and maintaining healthy status in animals [6, 7]. Adipose mass quickly and dynamically responds to nutrient/

Study name	Subgroup within study	Std diff in means	p-Value
Agueda, 2012	FatMass, total	-0,774	0,000
Belobrajdic, 2010	FatMass, total	-4,355	0,000
Bhutani, 2013	FatMass, total	-0,686	0,059
Cangemi, 2010	FatMass, total	-2,473	0,000
Cllessens, 2009	FatMass, total	-5,094	0,000
Cifton, 2004	FatMass, total	-1,115	0,004
Das, 2007	FatMass, total	-1,682	0,000
De Luis, 2012	FatMass, total	-0,334	0,146
Fontana, 2009	FatMass, total	-2,436	0,000
Fontana, 2006	FatMass, total	-1,353	0,000
Fontana, 2004	FatMass, total	-2,754	0,000
Fontana, 2008	FatMass, total	-2,930	0,000
Halberg, 2005	FatMass, total	0,305	0,400
Harvie, 2010	FatMass, total	-1,667	0,000
Harvie, 2010 b	FatMass, total	-1,143	0,000
Haugaard, 2009	FatMass, total	-1,658	0,000
Johnston, 2004	FatMass, total	-0,906	0,044
Kasim, 2009	FatMass, total	-0,820	0,019
Kleiner, 2006	FatMass, total	-0,150	0,654
Layman, 2003	FatMass, total	-2,274	0,000
Layman, 2005	FatMass, total	-2,437	0,000
Lee, 2009	FatMass, total	-0,567	0,002
Lejeune, 2004	FatMass, total	-0,635	0,000
Melanson, 2012	FatMass, total	-1,280	0,000
Redman, 2008	FatMass, total	-2,294	0,000
Soenen, 2012	FatMass, total	-0,613	0,001
Svendsen, 2012	FatMass, total	-0,696	0,048
Tapscott, 2009	FatMass, total	-0,624	0,048
Walker, 2008	FatMass, total	-0,633	0,029
Westerterp, 2004	FatMass, total	-0,612	0,000
Wycheley, 2010	FatMass, total	-0,723	0,010
		-0,913	0,000
Garcia, 2012	FatMass, visceral	-0,119	0,765
Klempel, 2012	FatMass, visceral	-0,600	0,034
Larson, 2006	FatMass, visceral	-1,964	0,000
Nicklas, 2009	FatMass, visceral	-0,806	0,000
Pijl, 2001	FatMass, visceral	-1,728	0,004
Weiss, 2006	FatMass, visceral	-0,891	0,001
Yoshimura, 2014	FatMass, visceral	-2,500	0,000
		-0,944	0,000
		-0,916	0,000



**Figure 1. Changes of total and visceral adipose mass after DR.** Studies were stratified according to the design of the study. A positive standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of fat mass (total or visceral). The empty black square indicates the results of each study, whereas empty blue square shows the summary results of each subgroup data. The red diamond resumes overall results of the included studies in the forest plot.

energy fluctuation and its remodelling seems to mediate the beneficial effects of DR [7]. In this section we evaluated the effects of DR on adipose mass (Fig. 1). Interestingly, all studies showed clear evidence on the efficacy of DR in reducing total adipose mass in human (SDM -0.913; 95% CI -0.994, -0.832;  $p < 0.000$ ). Interestingly, we detected higher effectiveness of DR in healthy than unhealthy subjects (SDM -1.843; 95% CI -2.144, -1.542  $p < 0.000$  and SDM -0.813; 95% CI -0.897, -0.728  $p < 0.000$ , respectively). Our data reveal that DR was also effective in reducing visceral fat mass (SDM -0.944; 95% CI -1.187, -0.700;  $p < 0.000$ ) (Fig. 1) and identify adipose mass measurement as a feasible approach to evaluate the efficacy of diets low in calories.

### Effects of DR on adipokines and DHEA

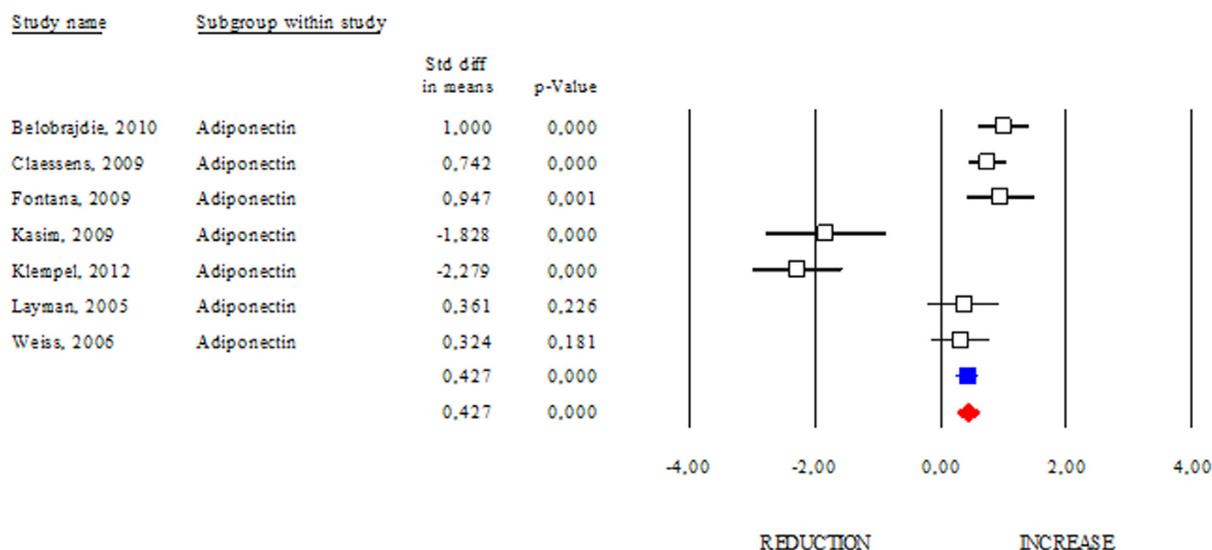
Among adipokines, adiponectin has an anti-inflammatory function and correlates with healthy metabolic profile. Reduction of adiponectin production is often revealed in obese and diabetic subjects [8]. These evidences highlight adiponectin as a good candidate to monitor healthy status in human. However, conflicting results emerge from circulating adiponectin levels in centenarians [9, 10]. Herein we determined changes of adiponectin levels occurring after DR. As shown in Fig. 2, DR increased adiponectin levels in human (SDM 0.427; 95% CI 0.243, 0.612;  $p < 0.000$ ) independently of healthy status (healthy group: SDM 0.947; 95% CI 0.395, 1.499  $p < 0.001$  and unhealthy

group: SDM 0.370; 95% CI 0.155, 0.585  $p < 0.001$ ). The “satiety hormone” leptin controls dietary behaviour and has been strongly associated with adipose mass. Indeed, reduced leptin levels are associated with diminished visceral adipose mass. However, unclear are evidences about its levels in healthy centenarians [9, 10]. Our data reveal that leptin levels were significantly reduced in DR group (SDM -1.383; 95% CI -1.511, -1.255;  $P < 0.000$ ) (Fig. 3).

The hormonal profile of aging includes a marked decrease in the adrenal hormone dehydroepiandrosterone (DHEA) [11]. DHEA is taken up by adipose tissue and seems to reduce its mass protecting against obesity [12]. Epidemiologic data in the elderly cohort of long-living Okinawans (over 65) show relatively high plasma DHEA levels at older ages than the aged-matched counterpart [13]. However, as disclosed in Suppl. Fig. 1, DHEA levels were unchanged after DR (SDM 0.149; 95% CI -0.342, 0.641  $p = 0.551$ ). Overall findings suggest a tight relationship between changes in circulating adipokines and reduction of adipose mass occurring after DR. Differently, DHEA modulation seems to be independent of calorie intake.

### Effects of DR on insulin, IGF-1, HOMA Index and IGBPs

Insulin and insulin growth factors 1 (IGF-1) signalling is an evolutionary conserved pathway linking nutrient levels to fat mass and lifespan. Generally, reduced level



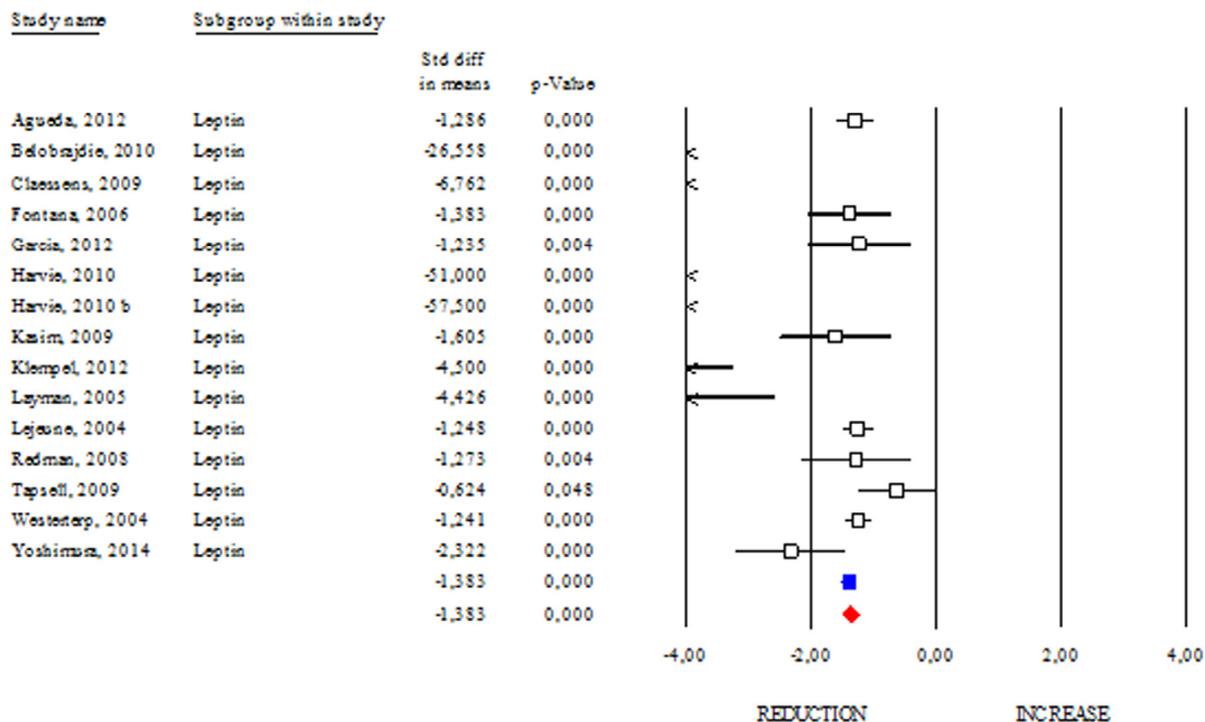
**Figure 2. DR effects on circulating adiponectin.** Studies were stratified according to the design of the study. A positive standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of circulating adiponectin. The empty black square indicates the results of each study, whereas empty blue square shows the summary results of each subgroup data. The red diamond resumes overall results of the included studies in the forest plot.

of insulin and IGF-1 is associated with increased longevity from yeasts to mammals [14]. Differently, levels of insulin and IGF-1 are commonly higher in subjects affected by age-related diseases or obesity than lean healthy subjects [15]. In our work, we reported clear evidence about DR effects on insulin and IGF-1 levels in human (Fig. 4). In particular, we observed a significant reduction in insulin both in healthy (SDM -1.019; 95% CI -1.362, -0.675  $p < 0.000$ ) and unhealthy subjects (SDM -0.811; 95% CI -0.893, -0.730  $p < 0.000$ ). The same trend was detected by analysing the IGF-1 levels (SDM -0.546; 95% CI -0.750, -0.342  $p < 0.000$ ). Overall data analyses (SDM -0.779; 95% CI -0.851, -0.706  $p < 0.000$ ) confirm decreased insulin/IGF-1 levels as downstream effect of DR in human.

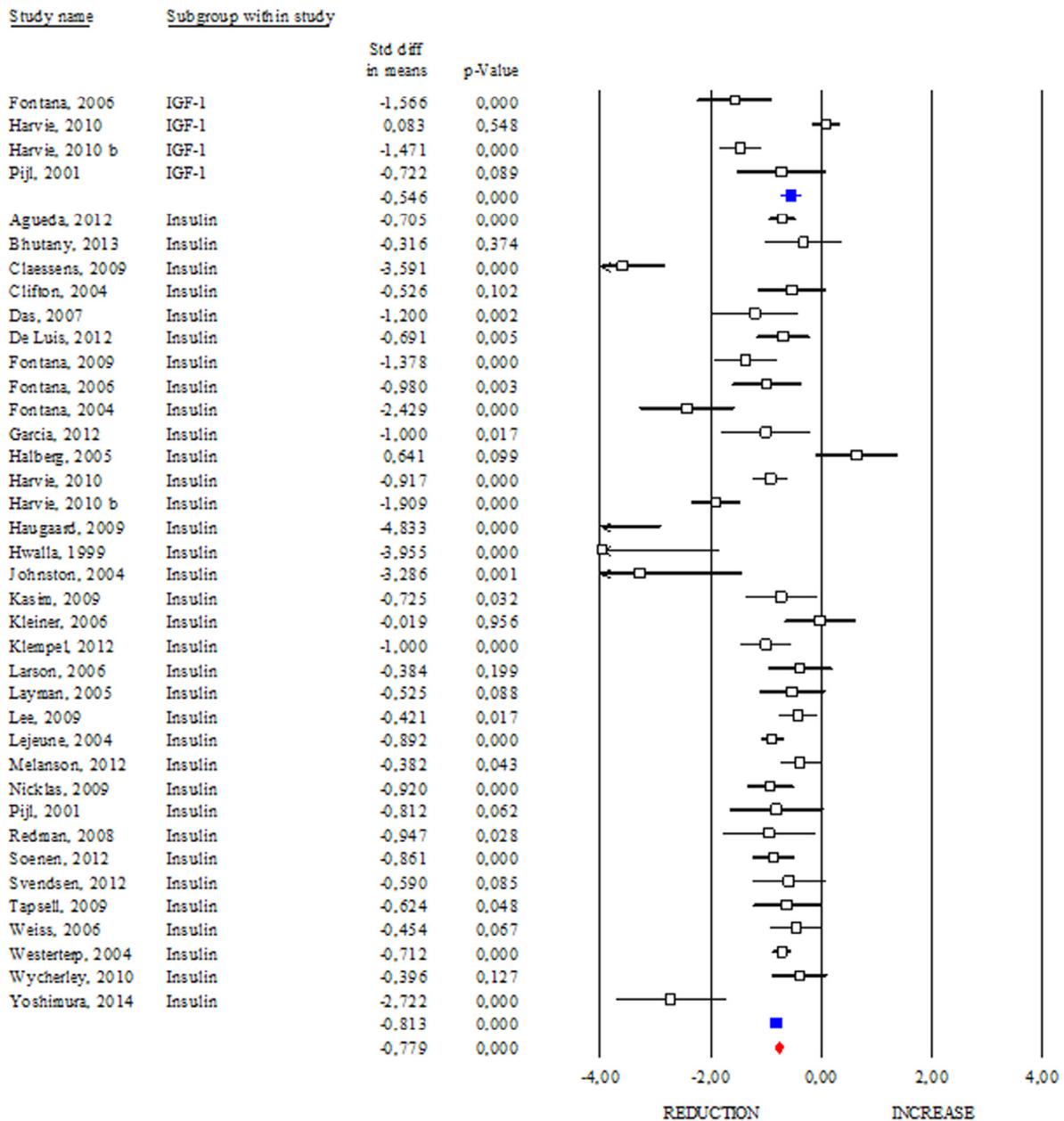
The Homeostasis Model Assessment (HOMA) Index is currently a biochemical tool to estimate insulin sensitivity by matching fasting glycaemia and insulinemia [16]. A study carried out on centenarians indicates that they seem to be protected from hyperinsulinaemia, and their insulin resistance is as low, if not lower, than that of healthy younger adults [17]. The correlation between HOMA Index with obesity or aging

suggests its prognostic capacity to evaluate the efficacy of health promoting strategies. Accordingly, we reported a significant reduction in the HOMA Index occurring after DR (SDM -0.837; 95% CI -0.990, -0.750  $p < 0.000$ ) (Fig. 5) and this effect was stronger if dietary treatment was longer than 3 months (data not shown).

The IGF-binding protein 2 (IGFBP2) is known as a carrier protein for IGF-1 limiting its biological action [18]. However, there are several characterized IGFBPs, which seem to improve metabolic status independently of IGFs binding [19]. Interestingly, some papers reported that DR regimens increase circulating levels of IGFBPs [20]. In our work, we analysed the changes in the levels of the best-known IGFBPs after DR. As shown in Fig. 6, DR similarly modulated IGFBP-1 and IGFBP-2 levels (SDM 1.527; 95% CI 1.248, 1.806  $p < 0.000$  and SDM 1.687; 95% CI 1.387, 1.986  $p < 0.000$ , respectively). Differently, DR was ineffective in increasing IGFBP-3 levels (SDM -0.045; 95% CI -0.517, 0.427  $p = 0.853$ ). These results suggest that IGFBP-1 and -2 are more sensitive to DR than IGFBP-3.



**Figure 3. DR effects on circulating leptin.** Studies were stratified according to the design of the study. A positive standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of circulating leptin. The empty black square indicates the results of each study. The red diamond resumes overall results of the included studies in the forest plot.



**Figure 4. Changes of circulating insulin and insulin growth factor-1 (IGF-1) after DR.** Studies were stratified according to the design of the study. A positive standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of circulating IGF-1 or insulin. The empty black square indicates the results of each study, whereas empty blue square shows the summary results of each subgroup data. The red diamond resumes overall results of the included studies in the forest plot.

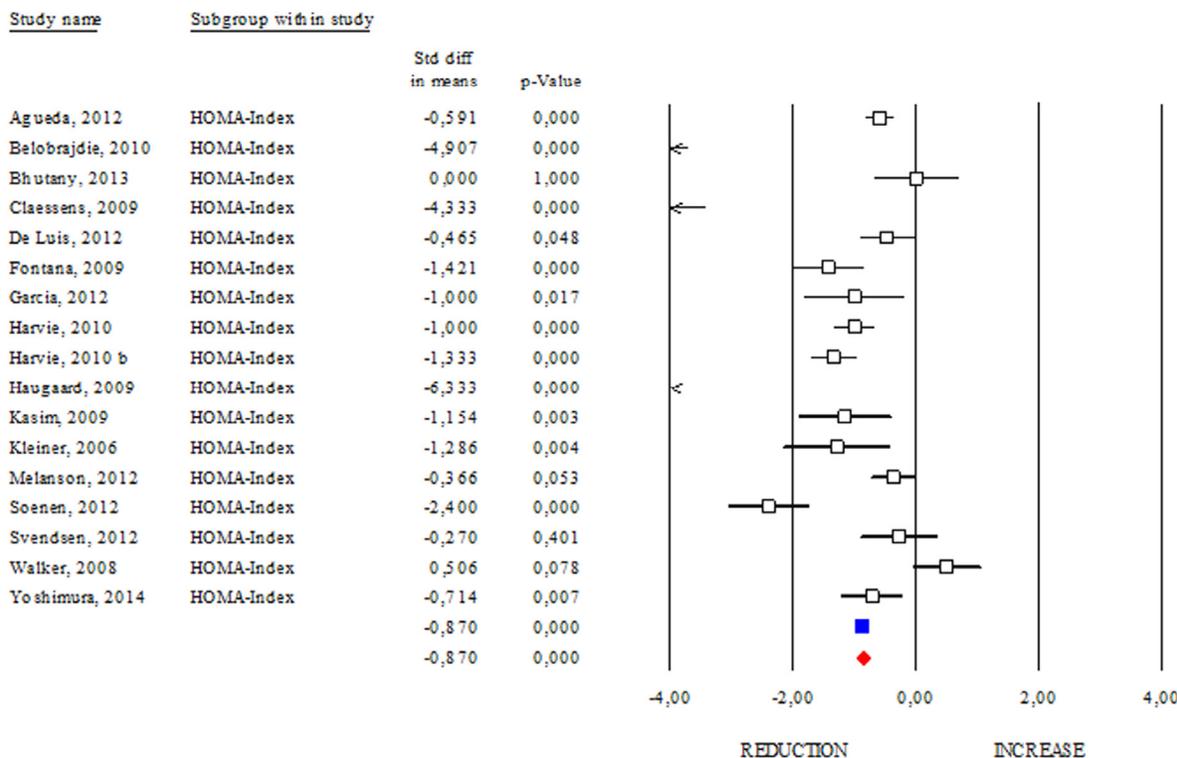
### Effects of DR on inflammatory markers

One of the common features of aging and obesity is the presence of a chronic sterile low-grade inflammatory status, which contributes to the onset of several metabolic perturbations [21]. In our work we evaluated the changes in circulating inflammatory markers observ-

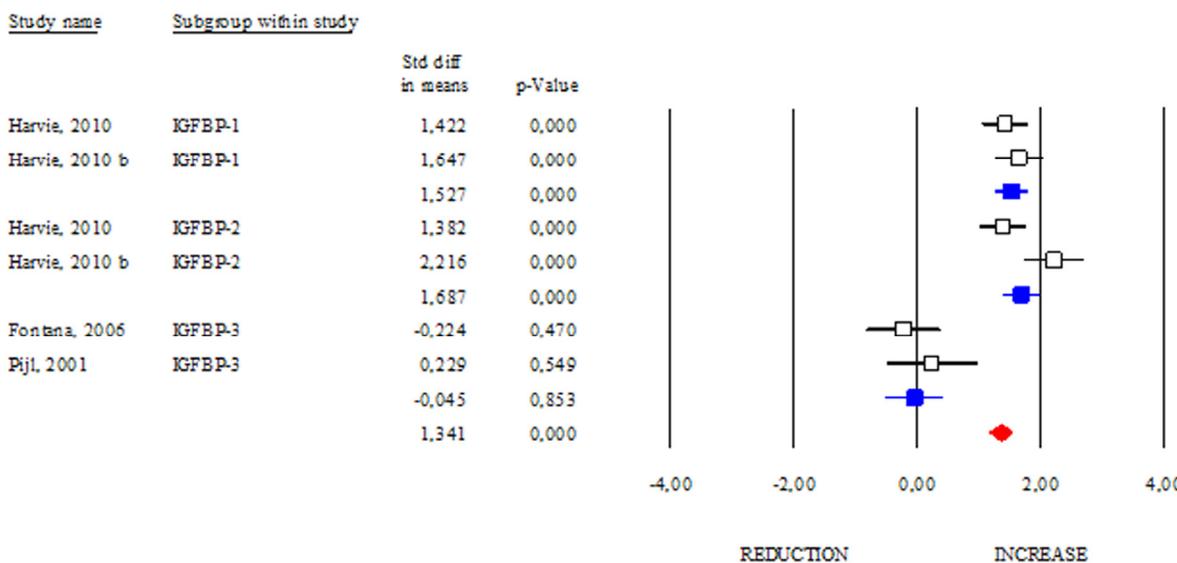
ed after DR (Fig. 7). Interestingly, among the evaluated inflammatory markers, only CRP and IL-6 displayed a significant reduction after DR (SDM -0.715; 95% CI -0.862, -0.568  $p < 0.000$  and SDM -0.316; 95% CI -0.515, -0.118  $p < 0.002$ , respectively). Although IL-1 and TNF- $\alpha$  are cytokines routinely assayed to monitor systemic inflammation, our data revealed that their level

remained unchanged after DR (SDM 0.041; 95% CI -0.181, 0.263 p=0.719 and SDM -0.079; 95% CI -0.264, 0.106 p=0.402, respectively). Overall data regarding

CRP, IL-6, IL-1 and TNF- $\alpha$  levels revealed anti-inflammatory effect of DR in human (SDM -0.351; 95% CI -0.442, -0.260 p<0.000) (Fig. 7).

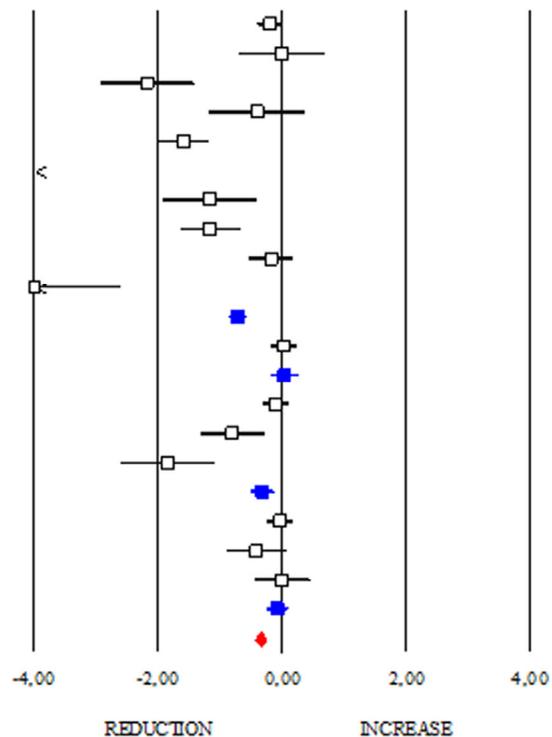


**Figure 5. Changes of HOMA Index after DR.** Studies were stratified according to the design of the study. A positive standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of HOMA Index. The empty black square indicates the results of each study. The red diamond resumes overall results of the included studies in the forest plot.



**Figure 6. Changes of circulating IGFB-1, IGFBP-2 and IGFBP-3 after DR.** Studies were stratified according to the design of the study. A standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of IGFB-1, IGFBP-2 or IGFBP-3. The empty black square indicates the results of each study, whereas empty blue square shows the summary results of each subgroup data. The red diamond resumes overall results of the included studies in the forest plot.

Study name	Subgroup within study	Std diff in means	p-Value
Agueda, 2012	CRP	-0,183	0,110
Bhutani, 2013	CRP	0,000	1,000
Fontana, 2006	CRP	-2,183	0,000
Garcia, 2012	CRP	-0,400	0,319
Harvia, 2010	CRP	-1,600	0,000
Harvia, 2010 b	CRP	-5,000	0,000
Kasim, 2009	CRP	-1,170	0,003
Klempel, 2012	CRP	-1,155	0,000
Melanson, 2012	CRP	-0,177	0,345
Yoshimura, 2014	CRP	-4,000	0,000
		-0,715	0,000
Agueda, 2012	IL-1	0,041	0,719
		0,041	0,719
Agueda, 2012	IL-6	-0,105	0,354
Fontana, 2009	IL-6	-0,795	0,004
Yoshimura, 2014	IL-6	-1,854	0,000
		-0,316	0,002
Agueda, 2012	TNF-alpha	-0,027	0,809
Weiss, 2006	TNF-alpha	-0,406	0,098
Yoshimura, 2014	TNF-alpha	0,000	1,000
		-0,079	0,402
		-0,351	0,000



**Figure 7. Changes of inflammatory markers after DR.** Studies were stratified according to the design of the study. A positive standardized difference in mean (SDM) indicates an increase, whereas a negative SDM indicates the decrease of CRP, IL-1, IL-6 or TNF-alpha. The empty black square indicates the results of each study, whereas empty blue square shows the summary results of each subgroup data. The red diamond resumes overall results of the included studies in the forest plot.

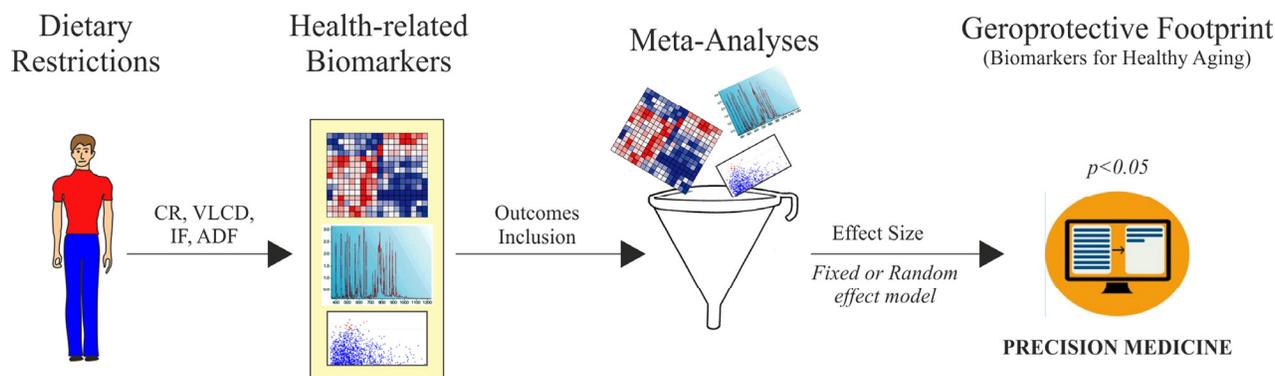
## DISCUSSION

Aging is commonly defined as a physiological decline of biological functions in the body. Aging strongly remodels adipose depots by reducing subcutaneous adipose in favour of visceral depots enlargement [22]. Aging and visceral adipose tissue expansion act in synergy in inducing a chronic low grade of inflammatory status, which triggers a systemic metabolic decline in human [21, 23]. DR is a promising and feasible strategy that ameliorates body metabolic and inflammatory profile increasing lifespan through evolutionary-conserved mechanisms [4, 22, 24, 25]. Herein we included all studies evaluating the impact of DR on several healthy-associated markers in human including adipose mass. Increased visceral adiposity leads to chronic inflammation, which is often associated with a number of comorbidities (e.g. hyperinsulinemia, hypertension, insulin resistance, glucose intolerance) and reduced life expectancy [26, 27]. Through this meta-analysis approach, we confirmed the capacity of

DR to reduce total and visceral adipose mass and, interestingly, we observed a more effective visceral adipose mass reduction after DR regimens (-20% in DR: SDM -1.081; 95% CI -1.242, -0.921 p<0.000) (-30/40% in DR: SDM -0.893; 95% CI -1.050, -0.737 p<0.000 and >-40% in DR: SDM -0.678; 95% CI -0.800, -0.555 p<0.000). These findings suggest that to obtain a more effective adipose mass loss, 20% in calorie reduction could be an elective strategy. Central or visceral adiposity perturbs systemic inflammation in animal models and human and relatively to this, the healthy effects of DR could be mediated by visceral adiposity reduction. Indeed, DR significantly diminished the markers of inflammation, highlighting the central role of DR-mediated adipose tissue remodelling in improving inflammatory profile in human. Furthermore, DR also increased adiponectin/leptin ratio, which is commonly associated with ameliorated insulin sensitivity in human. In line with this effect, we demonstrated that DR was successful in reducing insulin, IGF-1 and HOMA index.

Biomarkers		Effect Size
Fat mass	Total	Red bar
	Visceral	Red bar
Inflammation	CRP	Red bar
	IL-6	Red bar
Metabolism	Insulin	Red bar
	IGF-1	Red bar
	HOMA-Index	Red bar
Adipokines	Adiponectin	Green bar
	Leptin	Red bar
IGFBPs	IGFBP-1	Green bar
	IGFBP-2	Green bar

**Figure 8. Geroprotective footprint of dietary restriction.**



**Figure 9. Algorithm development for biomarkers validation of dietary restriction in human.** CR: calorie restriction; VLCD: very low calorie diet; IF: intermittent fasting; ADF: alternate-day-fasting.

The insulin growth factor binding proteins (IGFBPs) are a family of proteins that bind to insulin-like growth factors limiting their biological actions [28]. IGFBP-2 is the most abundant among circulating IGFBPs and its anti-diabetic role as well as direct ability to limit adipogenesis has been demonstrated [29, 30]. Actually, high serum levels of IGFBP-2 appear to protect against obesity and type 2 diabetes [30]. IGFBP-1 showed an inverse relation with insulin and BMI in human [31]. Differently, unclear are the evidences about the link between IGFBP-3 and adipose mass. In accordance with the data described above, we observed a strong responsiveness in circulating levels of IGFBP-1 and -2 occurring after DR. However some limitations emerge from this meta-analysis. In particular, statistical analyses on IL-1 and IGFBPs were carried out only evaluating the results obtained from few studies [32-35]. Moreover, it was not possible to evaluate the efficiency of DR in gender or time of treatment

subgroups because it was difficult to collect a good number of subjects.

In conclusion, by a meta-analysis approach we have provided evidences about DR efficiency on key hallmarks of aging (Fig. 8) and built a useful platform to evaluate the responses of human to dietary regimens low in calories (Fig. 9).

## MATERIALS AND METHODS

### Search strategy and included studies

In our work we analysed human intervention studies and evaluated the impact of DR regimens on adipose mass and some biomarkers of healthy aging (*Geromarkers*). The *Geromarkers* included in our meta-analysis were described in Table 1. Two investigators,

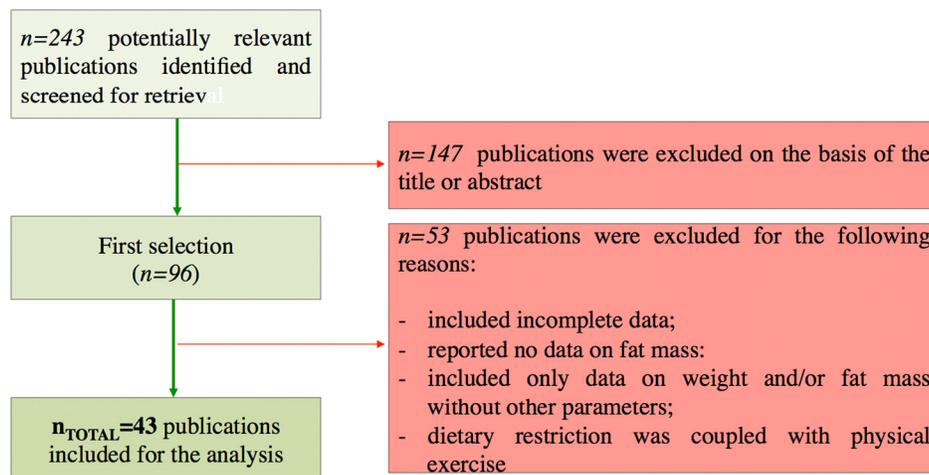
E.G. and D.L.B., independently carried out study selection and included both studies with an experimental design (EXP) and quasi-experimental design (Q-EXP). EXP studies were randomized with a control group and a parallel or crossover design; whereas Q-EXP included observational studies (pre- and post-intervention or pre- and post-data), non-randomized or uncontrolled studies [36]. Q-EXP studies were pooled together with EXP studies only after assessing whether they were in agreement with EXP studies [37]. Candidate studies were searched in PubMed (finalized February 30, 2016) using the terms ‘calorie or caloric or dietary restriction’, ‘fasting or intermittent fasting or alternate day fasting and ‘adipose tissue or fat mass or fat tissue’. Inclusion criteria were as follows: human intervention studies with long-term study design (> 3 months); healthy and unhealthy (e.g. dyslipidaemia, obesity, metabolic syndrome) subjects; numerically analysable information about results, study

duration and calories reduced in the study. Studies were excluded when: only abstracts were available; duration time of the study was lesser than 3 months; data presentation was incomplete; information about the DR was incomplete. When necessary, efforts were made to contact investigators for clarification or additional data. This research strategy produced a total of 201 studies. Furthermore, a manual research of references from clinical studies and reviews identified 42 additional studies, for a total of 243 studies to be evaluated, 9 of which are reviews [38-46]. A first screening allowed discarding 147 articles whose titles or abstracts were evidently irrelevant to our aim. Of the remaining 96 studies, 53 were rejected whenever: they presented incomplete data; DR was coupled with physical exercise; there were no reported data on adipose mass; they only presented data on weight and fat mass without other parameters (Fig. 10). Therefore, from 243 initial candidates, the 43 studies available for a formal meta-

analysis had the following characteristics: they were written in English; they had a period of intervention of at least two weeks; they were carried out exclusively on human subjects. Among the considered studies, 12 were on females [32, 34, 35, 47-55], 4 on males [56-59], and the rest mixed [60-62, 33, 63-85]. Moreover, 30 studies were intervention studies evaluating the efficacy of calorie restriction [33-35, 47-49, 51-53, 56, 58-60, 63-67, 69, 71-73, 75, 76, 79-82, 84, 85]; 4 were intervention studies evaluating the efficacy of intermittent fasting [50, 57, 61, 70]; 9 were intervention studies evaluating the efficacy of low or very low calorie diets [32, 54, 55, 62, 68, 74, 77, 78, 83]. The selected studies included human groups with different BMI. In particular, 10 were studies on obese [34, 35, 48, 50, 53, 54, 58, 61, 64, 68], 16 on overweight [51, 52, 57, 59, 62, 63, 65, 69, 71-73, 77-79, 82, 85], 12 on both obese and overweight [32, 47, 49, 55, 60, 62, 74, 75, 80, 81, 83, 84], 5 on both normal weight and over-

**Table 1. Selected biomarkers and number of the studies included in meta-analysis.**

Biomarkers	n. of the studies
Fat Mass (total and visceral)	38
Adipokines (adiponectin and leptin)	22
IGFBPs (IGFBP-1, -2, -3)	6
IGF-1	4
HOMA-Index	17
Insulin	34
Inflammation (TNFa, IL-1, IL-6, CRP)	17
DHEA	5



**Figure 10. Flow chart of the study identification and selection.**

weight [33, 66, 67, 70, 76]. Finally, the studies were on healthy subjects, with the exception of few articles in which subjects were affected by the following pathologies: chronic osteoarthritis [64]; metabolic syndrome [59]; hyperinsulinemia [58, 72], polycystic ovary syndrome [49], type 2 diabetes [84]. Hence, the meta-analysis was based on 43 studies and analysed a total of 2094 subjects. Before analyses, all studies were stratified for gender, healthy status, time of treatment and percentage of calorie reduction and the main characteristics of the included studies were reported in Table 2. Calorie restriction, intermittent or alternate-day-fasting and low calorie diet interventions were overall grouped in dietary restriction (DR) category.

random effect model was selected following evaluation of heterogeneity between studies based on the  $I^2$  test for heterogeneity. When  $I^2$  values were low, we selected a fixed effects model, whereas random effects model was selected for  $I^2$  values higher than 75%.

## CONFLICTS OF INTEREST

The authors declare no conflict of interests.

## FUNDING

This work was supported by “Uncovering Excellence” – University of Rome Tor Vergata and Ministero della Salute (GR-2011-02348047).

**Table 2. Characteristics of the included studies for the meta-analyses.**

Study Design				Gender Stratification		Healthy Status Stratification		Time of Treatment	
Unrandomized	Randomized or Controlled	Randomized and Controlled	Cross-Sectional	Yes	No	Yes	No	Brief (<3 months)	Long-Term (>3 months)
4	14	23	2	16	27	36	2	26	16

## Data analysis

Relevant data of the 43 studies available were entered for formal meta-analytic evaluation into the Comprehensive Meta-Analysis software (Biostat) [86]. Data analysis was performed as previously described [87]. In particular, for the results showed as post-data only, we selected mean, standard deviation and sample size in each group, or difference in means, sample size and p value between groups. When results were reported as pre- and post-data, we used mean, standard deviation, sample size in each group and correlation between baseline and end-point intervention period, or mean change, standard deviation difference, sample size in each group, correlation between baseline and end-point intervention period. For observational studies considering only one group (pre–post-intervention data), we used mean difference, standard deviation of difference and sample size. In all studies, we assumed the correlation between baseline and end-point study period to be 0.5 to produce the most conservative estimate [37, 88]. To enable a joint comparison, the standardized difference in mean (SDM) was calculated for each outcome. In our analysis, positive SDM indicates increased effect size of DR on outcome considered. The effect sizes of the included studies were pooled both under a ‘fixed effects model’ or ‘random effects model’. Under fixed effects model we assumed that the true effect is the same in all studies. By contrast, under the random effects model we allowed that the true effect may vary from one study to the next [37]. Fixed or

## REFERENCES

1. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, Deloukas P, Siebert R, Sipos B, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci USA*. 2014; 111:15538–43. doi: 10.1073/pnas.1412759111
2. Tchkonja T, Morbeck DE, Von Zglinicki T, Van Deursen J, Lustgarten J, Scoble H, Khosla S, Jensen MD, Kirkland JL. Fat tissue, aging, and cellular senescence. *Aging Cell*. 2010; 9:667–84. doi: 10.1111/j.1474-9726.2010.00608.x
3. Al-Regaiey KA. The effects of calorie restriction on aging: a brief review. *Eur Rev Med Pharmacol Sci*. 2016; 20:2468–73.
4. Most J, Tosti V, Redman LM, Fontana L. Calorie restriction in humans: an update. *Ageing Res Rev*. 2016; 15:1568-1637(16)30183-0.
5. Pearson ER. Personalized medicine in diabetes: the role of ‘omics’ and biomarkers. *Diabet Med*. 2016; 33:712–17. doi: 10.1111/dme.13075
6. Wiesenborn DS, Menon V, Zhi X, Do A, Gesing A, Wang Z, Bartke A, Altomare DA, Masternak MM. The effect of calorie restriction on insulin signaling in skeletal muscle and adipose tissue of Ames dwarf mice. *Aging (Albany NY)*. 2014; 6:900–12.

doi: 10.18632/aging.100700

7. Lettieri Barbato D, Tatulli G, Aquilano K, Ciriolo MR. Erratum: mitochondrial Hormesis links nutrient restriction to improved metabolism in fat cell. *Aging (Albany NY)*. 2016; 8:1571. doi: 10.18632/aging.101003
8. Ghoshal K, Bhattacharyya M. Adiponectin: probe of the molecular paradigm associating diabetes and obesity. *World J Diabetes*. 2015; 6:151–66. doi: 10.4239/wjd.v6.i1.151
9. Pareja-Galeano H, Santos-Lozano A, Sanchis-Gomar F, Fiuza-Luces C, Garatachea N, Gálvez BG, Lucia A, Emanuele E. Circulating leptin and adiponectin concentrations in healthy exceptional longevity. *Mech Ageing Dev*. 2016S0047-6374(16)30019-7.
10. Meazza C, Vitale G, Pagani S, Castaldi D, Ogliaari G, Mari D, Laarej K, Tinelli C, Bozzola M. Common adipokine features of neonates and centenarians. *J Pediatr Endocrinol Metab*. 2011; 24:953–57. doi: 10.1515/JPEM.2011.373
11. Samaras N, Samaras D, Frangos E, Forster A, Philippe J. A review of age-related dehydroepiandrosterone decline and its association with well-known geriatric syndromes: is treatment beneficial? *Rejuvenation Res*. 2013; 16:285–94. doi: 10.1089/rej.2013.1425
12. Karbowska J, Kochan Z. Effects of DHEA on metabolic and endocrine functions of adipose tissue. *Horm Mol Biol Clin Investig*. 2013; 14:65–74. doi: 10.1515/hmbci-2013-0009
13. Willcox BJ, Willcox DC, Todoriki H, Fujiyoshi A, Yano K, He Q, Curb JD, Suzuki M. Caloric restriction, the traditional Okinawan diet, and healthy aging: the diet of the world's longest-lived people and its potential impact on morbidity and life span. *Ann N Y Acad Sci*. 2007; 1114:434–55. doi: 10.1196/annals.1396.037
14. Parrella E, Longo VD. Insulin/IGF-I and related signaling pathways regulate aging in nondividing cells: from yeast to the mammalian brain. *Sci World J*. 2010; 10:161–77. doi: 10.1100/tsw.2010.8
15. Balasubramanian P, Longo VD. Growth factors, aging and age-related diseases. *Growth Horm IGF Res*. 2016; 28:66–68. doi: 10.1016/j.ghir.2016.01.001
16. Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. *Diabetes Care*. 2004; 27:1487–95. doi: 10.2337/diacare.27.6.1487
17. Curtis R, Geesaman BJ, DiStefano PS. Ageing and metabolism: drug discovery opportunities. *Nat Rev Drug Discov*. 2005; 4:569–80. doi: 10.1038/nrd1777
18. Kang HS, Kim MY, Kim SJ, Lee JH, Kim YD, Seo YK, Bae JH, Oh GT, Song DK, Ahn YH, Im SS. Regulation of IGFBP-2 expression during fasting. *Biochem J*. 2015; 467:453–60. doi: 10.1042/BJ20141248
19. Oh Y. IGF-independent regulation of breast cancer growth by IGF binding proteins. *Breast Cancer Res Treat*. 1998; 47:283–93. doi: 10.1023/A:1005911319432
20. Levine ME, Suarez JA, Brandhorst S, Balasubramanian P, Cheng CW, Madia F, Fontana L, Mirisola MG, Guevara-Aguirre J, Wan J, Passarino G, Kennedy BK, Wei M, et al. Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab*. 2014; 19:407–17. doi: 10.1016/j.cmet.2014.02.006
21. Goldberg EL, Dixit VD. Drivers of age-related inflammation and strategies for healthspan extension. *Immunol Rev*. 2015; 265:63–74. doi: 10.1111/imr.12295
22. Lettieri Barbato D, Aquilano K. Feast and famine: adipose tissue adaptations for healthy aging. *Ageing Res Rev*. 2016; 28:85–93. doi: 10.1016/j.arr.2016.05.007
23. Minamino T, Orimo M, Shimizu I, Kunieda T, Yokoyama M, Ito T, Nojima A, Nabetani A, Oike Y, Matsubara H, Ishikawa F, Komuro I. A crucial role for adipose tissue p53 in the regulation of insulin resistance. *Nat Med*. 2009; 15:1082–87. doi: 10.1038/nm.2014
24. Lettieri Barbato D, Aquilano K, Ciriolo MR. FoxO1 at the nexus between fat catabolism and longevity pathways. *Biochim Biophys Acta*. 2014; 1841:1555–60. doi: 10.1016/j.bbali.2014.08.004
25. Emran S, Yang M, He X, Zandveld J, Piper MD. Target of rapamycin signalling mediates the lifespan-extending effects of dietary restriction by essential amino acid alteration. *Aging (Albany NY)*. 2014; 6:390–98. doi: 10.18632/aging.100665
26. Fontana L, Hu FB. Optimal body weight for health and longevity: bridging basic, clinical, and population research. *Aging Cell*. 2014; 13:391–400. doi: 10.1111/ace.12207
27. Guilherme A, Virbasius JV, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat Rev Mol Cell Biol*. 2008; 9:367–77. doi: 10.1038/nrm2391
28. Clemmons DR. Insulin-like growth factor binding proteins and their role in controlling IGF actions. *Cytokine Growth Factor Rev*. 1997; 8:45–62. doi: 10.1016/S1359-6101(96)00053-6
29. Hedbacker K, Birsoy K, Wysocki RW, Asilmaz E, Ahima RS, Farooqi IS, Friedman JM. Antidiabetic effects of

- IGFBP2, a leptin-regulated gene. *Cell Metab.* 2010; 11:11–22. doi: 10.1016/j.cmet.2009.11.007
30. Wheatcroft SB, Kearney MT, Shah AM, Ezzat VA, Miell JR, Modo M, Williams SC, Cawthorn WP, Medina-Gomez G, Vidal-Puig A, Sethi JK, Crossey PA. IGF-binding protein-2 protects against the development of obesity and insulin resistance. *Diabetes.* 2007; 56:285–94. doi: 10.2337/db06-0436
31. Janssen JA, Stolk RP, Pols HA, Grobbee DE, Lamberts SW. Serum total IGF-I, free IGF-I, and IGFB-1 levels in an elderly population: relation to cardiovascular risk factors and disease. *Arterioscler Thromb Vasc Biol.* 1998; 18:277–82. doi: 10.1161/01.ATV.18.2.277
32. Harvie MN, Pegington M, Mattson MP, Frystyk J, Dillon B, Evans G, Cuzick J, Jebb SA, Martin B, Cutler RG, Son TG, Maudsley S, Carlson OD, et al. The effects of intermittent or continuous energy restriction on weight loss and metabolic disease risk markers: a randomized trial in young overweight women. *Int J Obes.* 2011; 35:714–27. doi: 10.1038/ijo.2010.171
33. Fontana L, Klein S, Holloszy JO. Long-term low-protein, low-calorie diet and endurance exercise modulate metabolic factors associated with cancer risk. *Am J Clin Nutr.* 2006; 84:1456–62.
34. Pijl H, Langendonk JG, Burggraaf J, Frölich M, Cohen AF, Veldhuis JD, Meinders AE. Altered neuroregulation of GH secretion in viscerally obese premenopausal women. *J Clin Endocrinol Metab.* 2001; 86:5509–15. doi: 10.1210/jcem.86.11.8061
35. Agueda M, Lasa A, Simon E, Ares R, Larrarte E, Labayen I. Association of circulating visfatin concentrations with insulin resistance and low-grade inflammation after dietary energy restriction in Spanish obese non-diabetic women: role of body composition changes. *Nutr Metab Cardiovasc Dis.* 2012; 22:208–14. doi: 10.1016/j.numecd.2010.06.010
36. Harris AD, Bradham DD, Baumgarten M, Zuckerman IH, Fink JC, Perencevich EN. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis.* 2004; 38:1586–91. doi: 10.1086/420936
37. Borenstein M. 2009. *Introduction to meta-analysis.* Chichester, West Sussex, UK; Hoboken: John Wiley & Sons.
38. Fontana L, Klein S. Aging, adiposity, and calorie restriction. *JAMA.* 2007; 297:986–94. doi: 10.1001/jama.297.9.986
39. Holloszy JO, Fontana L. Caloric restriction in humans. *Exp Gerontol.* 2007; 42:709–12. doi: 10.1016/j.exger.2007.03.009
40. Johansson K, Neovius M, Hemmingsson E. Effects of anti-obesity drugs, diet, and exercise on weight-loss maintenance after a very-low-calorie diet or low-calorie diet: a systematic review and meta-analysis of randomized controlled trials. *Am J Clin Nutr.* 2014; 99:14–23. doi: 10.3945/ajcn.113.070052
41. Longo VD, Fontana L. Calorie restriction and cancer prevention: metabolic and molecular mechanisms. *Trends Pharmacol Sci.* 2010; 31:89–98. doi: 10.1016/j.tips.2009.11.004
42. Omodei D, Fontana L. Calorie restriction and prevention of age-associated chronic disease. *FEBS Lett.* 2011; 585:1537–42. doi: 10.1016/j.febslet.2011.03.015
43. Sohal RS, Weindruch R. Oxidative stress, caloric restriction, and aging. *Science.* 1996; 273:59–63. doi: 10.1126/science.273.5271.59
44. Varady KA, Hellerstein MK. Alternate-day fasting and chronic disease prevention: a review of human and animal trials. *Am J Clin Nutr.* 2007; 86:7–13.
45. Varady KA. Intermittent versus daily calorie restriction: which diet regimen is more effective for weight loss? *Obes Rev.* 2011; 12:e593–601. doi: 10.1111/j.1467-789X.2011.00873.x
46. Wycherley TP, Moran LJ, Clifton PM, Noakes M, Brinkworth GD. Effects of energy-restricted high-protein, low-fat compared with standard-protein, low-fat diets: a meta-analysis of randomized controlled trials. *Am J Clin Nutr.* 2012; 96:1281–98. doi: 10.3945/ajcn.112.044321
47. Clifton PM, Noakes M, Keogh JB. Very low-fat (12%) and high monounsaturated fat (35%) diets do not differentially affect abdominal fat loss in overweight, nondiabetic women. *J Nutr.* 2004; 134:1741–45.
48. García-Unciti M, Izquierdo M, Idoate F, Gorostiaga E, Grijalba A, Ortega-Delgado F, Martínez-Labari C, Moreno-Navarrete JM, Forga L, Fernández-Real JM, Ibáñez J. Weight-loss diet alone or combined with progressive resistance training induces changes in association between the cardiometabolic risk profile and abdominal fat depots. *Ann Nutr Metab.* 2012; 61:296–304. doi: 10.1159/000342467
49. Kasim-Karakas SE, Almario RU, Cunningham W. Effects of protein versus simple sugar intake on weight loss in polycystic ovary syndrome (according to the National Institutes of Health criteria). *Fertil Steril.* 2009; 92:262–70. doi: 10.1016/j.fertnstert.2008.05.065
50. Klempel MC, Kroeger CM, Bhutani S, Trepanowski JF, Varady KA. Intermittent fasting combined with calorie restriction is effective for weight loss and cardio-

- protection in obese women. *Nutr J*. 2012; 11:98. doi: 10.1186/1475-2891-11-98
51. Layman DK, Boileau RA, Erickson DJ, Painter JE, Shiue H, Sather C, Christou DD. A reduced ratio of dietary carbohydrate to protein improves body composition and blood lipid profiles during weight loss in adult women. *J Nutr*. 2003; 133:411–17.
52. Layman DK, Evans E, Baum JI, Seyler J, Erickson DJ, Boileau RA. Dietary protein and exercise have additive effects on body composition during weight loss in adult women. *J Nutr*. 2005; 135:1903–10.
53. Nicklas BJ, Wang X, You T, Lyles MF, Demons J, Easter L, Berry MJ, Lenchik L, Carr JJ. Effect of exercise intensity on abdominal fat loss during calorie restriction in overweight and obese postmenopausal women: a randomized, controlled trial. *Am J Clin Nutr*. 2009; 89:1043–52. doi: 10.3945/ajcn.2008.26938
54. Nørrelund H, Børglum J, Jørgensen JO, Richelsen B, Møller N, Nair KS, Christiansen JS. Effects of growth hormone administration on protein dynamics and substrate metabolism during 4 weeks of dietary restriction in obese women. *Clin Endocrinol (Oxf)*. 2000; 52:305–12. doi: 10.1046/j.1365-2265.2000.00937.x
55. Svendsen PF, Jensen FK, Holst JJ, Haugaard SB, Nilas L, Madsbad S. The effect of a very low calorie diet on insulin sensitivity, beta cell function, insulin clearance, incretin hormone secretion, androgen levels and body composition in obese young women. *Scand J Clin Lab Invest*. 2012; 72:410–19. doi: 10.3109/00365513.2012.691542
56. Cangemi R, Friedmann AJ, Holloszy JO, Fontana L. Long-term effects of calorie restriction on serum sex-hormone concentrations in men. *Aging Cell*. 2010; 9:236–42. doi: 10.1111/j.1474-9726.2010.00553.x
57. Halberg N, Henriksen M, Söderhamn N, Stallknecht B, Ploug T, Schjerling P, Dela F. Effect of intermittent fasting and refeeding on insulin action in healthy men. *J Appl Physiol (1985)*. 2005; 99:2128–36. doi: 10.1152/jappphysiol.00683.2005
58. Baba NH, Sawaya S, Torbay N, Habbal Z, Azar S, Hashim SA. High protein vs high carbohydrate hypoenergetic diet for the treatment of obese hyperinsulinemic subjects. *Int J Obes Relat Metab Disord*. 1999; 23:1202–06. doi: 10.1038/sj.ijo.0801064
59. Lee K, Lee J, Bae WK, Choi JK, Kim HJ, Cho B. Efficacy of low-calorie, partial meal replacement diet plans on weight and abdominal fat in obese subjects with metabolic syndrome: a double-blind, randomised controlled trial of two diet plans - one high in protein and one nutritionally balanced. *Int J Clin Pract*. 2009; 63:195–201. doi: 10.1111/j.1742-1241.2008.01965.x
60. Belobrajdic DP, Frystyk J, Jeyaratnaganathan N, Espelund U, Flyvbjerg A, Clifton PM, Noakes M. Moderate energy restriction-induced weight loss affects circulating IGF levels independent of dietary composition. *Eur J Endocrinol*. 2010; 162:1075–82. doi: 10.1530/EJE-10-0062
61. Bhutani S, Klempel MC, Kroeger CM, Trepanowski JF, Varady KA. Alternate day fasting and endurance exercise combine to reduce body weight and favorably alter plasma lipids in obese humans. *Obesity (Silver Spring)*. 2013; 21:1370–79. doi: 10.1002/oby.20353
62. Claessens M, van Baak MA, Monsheimer S, Saris WH. The effect of a low-fat, high-protein or high-carbohydrate ad libitum diet on weight loss maintenance and metabolic risk factors. *Int J Obes*. 2009; 33:296–304. doi: 10.1038/ijo.2008.278
63. Das SK, Gilhooly CH, Golden JK, Pittas AG, Fuss PJ, Cheatham RA, Tyler S, Tsay M, McCrory MA, Lichtenstein AH, Dallal GE, Dutta C, Bhapkar MV, et al. Long-term effects of 2 energy-restricted diets differing in glycemic load on dietary adherence, body composition, and metabolism in CALERIE: a 1-y randomized controlled trial. *Am J Clin Nutr*. 2007; 85:1023–30.
64. de Luis DA, Izaola O, García Alonso M, Aller R, Cabezas G, de la Fuente B. Effect of a commercial hypocaloric diet in weight loss and post surgical morbidities in obese patients with chronic arthropathy, a randomized clinical trial. *Eur Rev Med Pharmacol Sci*. 2012; 16:1814–20.
65. Fontana L, Klein S, Holloszy JO. Effects of long-term calorie restriction and endurance exercise on glucose tolerance, insulin action, and adipokine production. *Age (Dordr)*. 2010; 32:97–108. doi: 10.1007/s11357-009-9118-z
66. Fontana L, Meyer TE, Klein S, Holloszy JO. Long-term calorie restriction is highly effective in reducing the risk for atherosclerosis in humans. *Proc Natl Acad Sci USA*. 2004; 101:6659–63. doi: 10.1073/pnas.0308291101
67. Fontana L, Weiss EP, Villareal DT, Klein S, Holloszy JO. Long-term effects of calorie or protein restriction on serum IGF-1 and IGFBP-3 concentration in humans. *Aging Cell*. 2008; 7:681–87. doi: 10.1111/j.1474-9726.2008.00417.x
68. Haugaard SB, Vaag A, Mu H, Madsbad S. Skeletal muscle structural lipids improve during weight-maintenance after a very low calorie dietary intervention. *Lipids Health Dis*. 2009; 8:34.

doi: 10.1186/1476-511X-8-34

69. Heilbronn LK, de Jonge L, Frisard MI, DeLany JP, Larson-Meyer DE, Rood J, Nguyen T, Martin CK, Volaufova J, Most MM, Greenway FL, Smith SR, Deutsch WA, et al, and Pennington CALERIE Team. Effect of 6-month calorie restriction on biomarkers of longevity, metabolic adaptation, and oxidative stress in overweight individuals: a randomized controlled trial. *JAMA*. 2006; 295:1539–48. doi: 10.1001/jama.295.13.1539
70. Heilbronn LK, Smith SR, Martin CK, Anton SD, Ravussin E. Alternate-day fasting in nonobese subjects: effects on body weight, body composition, and energy metabolism. *Am J Clin Nutr*. 2005; 81:69–73.
71. Johnston CS, Tjonn SL, Swan PD. High-protein, low-fat diets are effective for weight loss and favorably alter biomarkers in healthy adults. *J Nutr*. 2004; 134:586–91.
72. Kleiner RE, Hutchins AM, Johnston CS, Swan PD. Effects of an 8-week high-protein or high-carbohydrate diet in adults with hyperinsulinemia. *MedGenMed*. 2006; 8:39.
73. Larson-Meyer DE, Heilbronn LK, Redman LM, Newcomer BR, Frisard MI, Anton S, Smith SR, Alfonso A, Ravussin E. Effect of calorie restriction with or without exercise on insulin sensitivity, beta-cell function, fat cell size, and ectopic lipid in overweight subjects. *Diabetes Care*. 2006; 29:1337–44. doi: 10.2337/dc05-2565
74. Lejeune MP, Kovacs EM, Westerterp-Plantenga MS. Additional protein intake limits weight regain after weight loss in humans. *Br J Nutr*. 2005; 93:281–89. doi: 10.1079/BJN20041305
75. Melanson KJ, Summers A, Nguyen V, Brosnahan J, Lowndes J, Angelopoulos TJ, Rippe JM. Body composition, dietary composition, and components of metabolic syndrome in overweight and obese adults after a 12-week trial on dietary treatments focused on portion control, energy density, or glycemic index. *Nutr J*. 2012; 11:57. doi: 10.1186/1475-2891-11-57
76. Meyer TE, Kovács SJ, Ehsani AA, Klein S, Holloszy JO, Fontana L. Long-term caloric restriction ameliorates the decline in diastolic function in humans. *J Am Coll Cardiol*. 2006; 47:398–402. doi: 10.1016/j.jacc.2005.08.069
77. Redman LM, Rood J, Anton SD, Champagne C, Smith SR, Ravussin E, and Pennington Comprehensive Assessment of Long-Term Effects of Reducing Intake of Energy (CALERIE) Research Team. Calorie restriction and bone health in young, overweight individuals. *Arch Intern Med*. 2008; 168:1859–66. doi: 10.1001/archinte.168.17.1859
78. Redman LM, Veldhuis JD, Rood J, Smith SR, Williamson D, Ravussin E, Pennington CT, and Pennington CALERIE Team. The effect of caloric restriction interventions on growth hormone secretion in nonobese men and women. *Aging Cell*. 2010; 9:32–39. doi: 10.1111/j.1474-9726.2009.00530.x
79. Soenen S, Bonomi AG, Lemmens SG, Scholte J, Thijssen MA, van Berkum F, Westerterp-Plantenga MS. Relatively high-protein or ‘low-carb’ energy-restricted diets for body weight loss and body weight maintenance? *Physiol Behav*. 2012; 107:374–80. doi: 10.1016/j.physbeh.2012.08.004
80. Tapsell L, Batterham M, Huang XF, Tan SY, Teuss G, Charlton K, Oshea J, Warensjö E. Short term effects of energy restriction and dietary fat sub-type on weight loss and disease risk factors. *Nutr Metab Cardiovasc Dis*. 2010; 20:317–25. doi: 10.1016/j.numecd.2009.04.007
81. Lasker DA, Evans EM, Layman DK. Moderate carbohydrate, moderate protein weight loss diet reduces cardiovascular disease risk compared to high carbohydrate, low protein diet in obese adults: A randomized clinical trial. *Nutr Metab (Lond)*. 2008; 5:30. doi: 10.1186/1743-7075-5-30
82. Weiss EP, Racette SB, Villareal DT, Fontana L, Steger-May K, Schechtman KB, Klein S, Holloszy JO, and Washington University School of Medicine CALERIE Group. Improvements in glucose tolerance and insulin action induced by increasing energy expenditure or decreasing energy intake: a randomized controlled trial. *Am J Clin Nutr*. 2006; 84:1033–42.
83. Westerterp-Plantenga MS, Lejeune MP, Nijs I, van Ooijen M, Kovacs EM. High protein intake sustains weight maintenance after body weight loss in humans. *Int J Obes Relat Metab Disord*. 2004; 28:57–64. doi: 10.1038/sj.ijo.0802461
84. Wycherley TP, Noakes M, Clifton PM, Cleanthous X, Keogh JB, Brinkworth GD. A high-protein diet with resistance exercise training improves weight loss and body composition in overweight and obese patients with type 2 diabetes. *Diabetes Care*. 2010; 33:969–76. doi: 10.2337/dc09-1974
85. Yoshimura E, Kumahara H, Tobina T, Matsuda T, Ayabe M, Kiyonaga A, Anzai K, Higaki Y, Tanaka H. Lifestyle intervention involving calorie restriction with or without aerobic exercise training improves liver fat in adults with visceral adiposity. *J Obes*. 2014; 2014:197216. doi: 10.1155/2014/197216

86. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol.* 2000; 53:1119–29. doi: 10.1016/S0895-4356(00)00242-0
87. Lettieri-Barbato D, Tomei F, Sancini A, Morabito G, Serafini M. Effect of plant foods and beverages on plasma non-enzymatic antioxidant capacity in human subjects: a meta-analysis. *Br J Nutr.* 2013; 109:1544–56. doi: 10.1017/S0007114513000263
88. Higgins JP, Green S, and Cochrane Collaboration. 2008. *Cochrane handbook for systematic reviews of interventions.* Chichester, West Sussex; Hoboken NJ: Wiley-Blackwell.

## Deep biomarkers of aging are population-dependent

Alan A. Cohen, Vincent Morissette-Thomas, Luigi Ferrucci, Linda P. Fried

Putin et al. [1] just published an excellent article showing how machine learning methods (specifically deep neural networks, DNNs) can be used to quantify the aging process using a set of 41 standard clinical biomarkers, most of which are not specifically recognized as biomarkers of aging. DNNs provide a method to obtain a predictive algorithm from raw data (the biomarkers in this case) with minimal to no *a priori* assumptions (see Mamoshina et al. 2016 [2] for details). This is an important finding because (a) it confirms that aging is not a single specific process, but rather a suite of changes that are felt across multiple physiological systems, probably within a complex systems framework, and (b) it suggests that measurement of the aging process is feasible with simple, standard measures. Both of these agree with recent findings from our lab showing that similar sets of biomarkers perform well for measurement of physiological dysregulation [3-7]. The difference is that our models are geared toward understanding the biology, and Putin et al. [1]'s toward prediction (i.e., estimation of biological age, though they do not use the term). Their model substantially outperforms ours for age prediction, but because the underlying algorithm is sufficiently complex as to remain a black box, it can provide relatively little insight into mechanisms. The two approaches are thus complementary.

There is, however, a substantial caveat to Putin et al. [1]'s approach that was not mentioned in their article. Their algorithm was developed based on clinical data from a single source covering Eastern Europe (90% Russia), and the applicability to data from other settings or to population subsets was not verified. There are a number of reasons to suspect that their algorithm would need to be adjusted for application in other settings: (1) Aging rates may differ across countries; (2) Genetic and environmental determinants of physiology may differ across countries/cultures, independent of aging; and (3) There may be specific biases in how clinical lab samples are taken and analyzed that differ substantially across health systems. These distinctions are not trivial: a universal measure of biological age has very different practical and biological implications than one that is highly contextual. They also represent a more general challenge for machine learning in the health domain: traditional applications of such techniques (e.g. facial

recognition, sentence completion [2]) are not generally subject to bias or anything related to the epidemiological concept of confounding, whereas such problems are rife in (bio)medical fields. There is thus substantial potential for development of methodological approaches to adjust for bias in machine learning methods applied in biomedical research.

We have access to similar data to that used by Putin et al. [1] for three major aging cohort studies, the Women's Health and Aging Study I & II (WHAS)[8], the Baltimore Longitudinal Study on Aging (BLSA)[9, 10], and *Invecchiare in Chianti* (InCHIANTI)[11], as well as publicly available cross-sectional data for a representative sample of the American population from the National Health and Nutrition Examination Survey (NHANES)[12]. For each study, we randomly chose 110 participants, stratified by age when necessary to achieve a broad age range, and input their values for the 10 basic biomarkers (albumin, glucose, alkaline phosphatase, urea, erythrocytes, cholesterol, RDW, alpha-2 globulins, hematocrit, and lymphocytes) in the online tool provided by Putin et al. [1] at [www.aging.ai](http://www.aging.ai). Alpha-2-globulins were only present in InCHIANTI, so we left the field empty in the other data sets (the DNN is capable of treating missing data, though this reduces accuracy). In addition, we ran as many of the full 41 biomarkers as possible for a set of 10 individuals per study, chosen randomly by age stratum from among the 110 run with 10 biomarkers. The number of biomarkers available was: WHAS: 34 biomarkers out of 41, BLSA: 37, InCHIANTI: 38, and NHANES: 33.

We found that indeed the performance of the model was substantially diminished in all four of our data sets. In the original study, the 10-biomarker version of the DNN has a 10-year epsilon accuracy (i.e., percentage correct prediction within  $\text{age} \pm 10$  years) of 70% and  $R^2 = 0.63$ ; across our datasets the mean epsilon accuracy was 38% and mean  $R^2 = 0.37$ , with maximum epsilon accuracy = 56% (InCHIANTI) and maximum  $R^2 = 0.59$  (NHANES, Fig. 1). The 41-biomarker versions performed neither markedly better nor worse, with a mean age error (MAE) actually increasing by 0.45 (95%CI: [-2.2, 1.3]) across our 40 samples. The confidence intervals and consistency across data sets are sufficient to exclude the possibility that our core results are due to the use of the 10-biomarker rather than the 41-biomarker tool (Fig. 1).

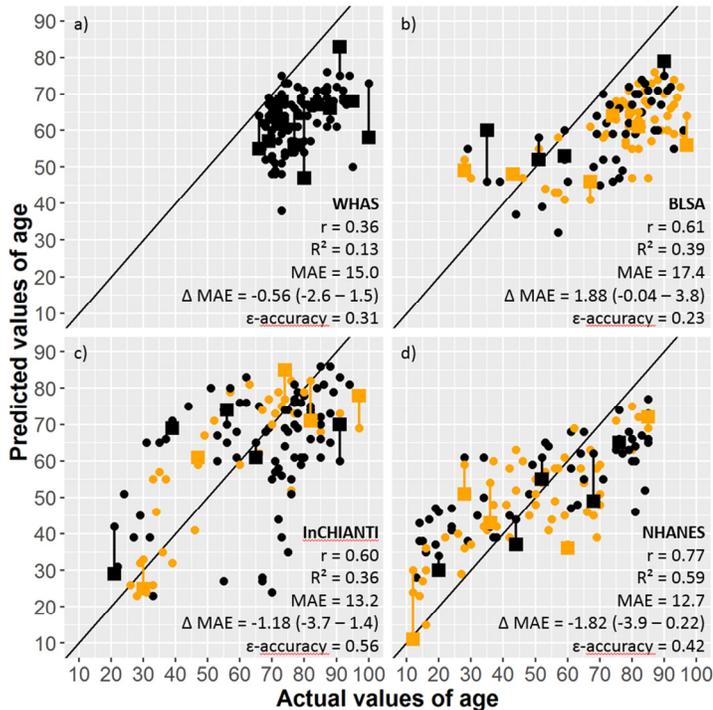
In addition to heterogeneity of performance across data sets, the DNN had a significantly better performance for men than for women globally (MAE diff= 1.8,  $p=0.04$ ) and in InCHIANTI (MAE diff= 5.5,  $p=0.002$ ) and NHANES (MAE diff= 4.2,  $p=0.007$ ), though there was no significant effect in BLSA (MAE diff= -1.5,  $p=0.39$ ). This is consistent with our findings on other measures of biological age, which for some reason consistently perform better for men, even when the methods are calibrated on women ([4] and unpublished data using methods from [13, 14]).

One potential reason for the poorer performance of the model in our datasets is the absence of children. Including children increases the age range, which by itself, all else equal, will increase  $r$  and  $R^2$  statistics [15]. Whether a measure of biological age needs to be accurate for children too is perhaps debatable or context-dependent, but clearly we would like the measure to be able to discriminate ages among adults well.

Additionally, we found a clear bias in the age estimates for BLSA and WHAS, with age substantially underestimated for almost all individuals in both data sets (Fig. 1a, b). This is actually consistent with the results of Putin et al. [1]. Their Fig. 1 a, d shows a bias toward underestimation of age for individuals aged 70+, and the BLSA and WHAS datasets largely contain individuals in this age range. For InCHIANTI and NHANES as well, ages of older individuals are underestimated and ages of younger individuals are overestimated, though less so than for BLSA and WHAS. Globally this suggests that Putin et al. [1]’s model performs well when the age range is large, but loses discriminatory power particularly at older ages. If the age bias is larger in BLSA and WHAS, as it appears to be, this might also imply that these populations age more slowly, an interesting finding.

However, such differences could also be due to something more mundane such as diet. Dietary patterns differ substantially between Eastern Europe, Italy, and the US, and diet is known to affect many clinical biomarkers (e.g. [16-18]), so it is hardly surprising that performance of algorithms based on these markers differs across these populations. Likewise, the majority of data used by Putin et al. [1] come from middle-aged individuals, and life expectancy in Russia is much lower than in Italy or the US [19], and has a substantially different cause composition [20]. We expect that many such factors contribute jointly to the patterns observed here.

In sum, these results show that there is unlikely to be a single algorithm that can predict biological age for all populations/sexes based on these clinical biomarkers. While we have not explored other population strata, such as by race, socioeconomic status, or environmental exposures, differences likely exist among these groups as well. The methods used by Putin et al. [1] are state of the art and perform well within their original dataset, suggesting that the barrier is true population differences rather than algorithm refinement. Population-specific algorithms might be an option but would require substantial work. Practically, this result is unfortunate, but biologically it is interesting. It implies that aging proceeds differently, and perhaps at different rates, in different populations. Other measures of biological age – for example, the epigenetic clock, or based on highly specific aging biomarkers such as leukocyte telomere length (LTL) – may or may not face these same hurdles [13-15, 21-23]. However, longitudinal changes in LTL depend on demographics, genes, and environment [24], implying that there will be population differences in how it works as a measure of biological age. More broadly, our results suggest that substantial caution is warranted in generalizing age-related changes in biomarkers across populations. Future work should attempt to replicate



**Figure 1.** Correlation between actual and predicted age values on 110 observations from four databases (a) WHAS, (b) BLSA, (c) InCHIANTI, and (d) NHANES] using the DNN on 10 biomarkers (small circles) or all available biomarkers (large squares). Paired observations with 10 and all available biomarkers are linked by vertical lines. Orange symbols are men and black symbols are women. MAE is mean age error and  $\Delta$  MAE is difference between MAE using 10 biomarkers and MAE using all available biomarkers, with positive values indicating better performance of the model with all biomarkers.  $\Delta$  MAE parentheses indicate 95% confidence intervals.

these findings in appropriate datasets from non-Western countries [25, 26], and to assess the performance of more diverse, integrated datasets.

## REFERENCES

1. Putin E, et al. *Aging (Albany NY)*. 2016; 8:1021-33. doi.org/10.18632/aging.100968
2. Mamoshina P, et al. *Molecular pharmaceutics*. 2016; 13:1445-54. doi.org/10.1021/acs.molpharmaceut.5b00982
3. Arbeev KG, et al. *Frontiers in Public Health*. 2016; 4. doi.org/10.3389/fpubh.2016.00003
4. Cohen AA, et al. *PLoS One*. 2015; 10:e0122541. doi.org/10.3389/fpubh.2016.00003
5. Cohen AA, et al. *Experimental Gerontology*. 2014; 57:203-10. doi.org/10.1371/journal.pone.0122541
6. Cohen AA, et al. *Mechanisms of Ageing and Development*. 2013; 134:110-17. doi.org/10.1016/j.mad.2013.01.004
7. Milot E, et al. *Mechanisms of Ageing and Development*. 2014; 141-142:56-63. doi.org/10.1016/j.mad.2014.10.001
8. Fried LP, et al. 1995. *The Women's Health and Aging Study*. Bethesda, MD: National Institute on Aging, pp. 1-8.
9. Ferrucci L. *Giornale di Gerontologia*. 2009; 57:3-8.
10. Shock NW. 1984. *Normal Human Aging*. Washington DC: National Institute of Aging, pp. 661.
11. Ferrucci L, et al. *Journal of the American Geriatrics Society*. 2000; 48:1618-25. doi.org/10.1111/j.1532-5415.2000.tb03873.x
12. National Health and Nutrition Examination Survey. <http://www.cdc.gov/nchs/nhanes.htm>. 2010. Center for Disease Control.
13. Belsky DW, et al. *Proc Natl Acad Sci USA*. 2015; 112:E4104-10. doi.org/10.1073/pnas.1506264112
14. Levine ME. *J Gerontol A Biol Sci Med Sci*. 2013; 68:667-74. doi.org/10.1093/gerona/gls233
15. Horvath S. *Genome Biol*. 2013; 14:R115. doi.org/10.1186/gb-2013-14-10-r115
16. Ajani UA, et al. *J Nutr*. 2004; 134:1181-85.
17. Don BR, Kaysen G. 2004. *Seminars in dialysis*: Wiley Online Library, pp. 432-437.
18. Mensink RP, et al. *Am J Clin Nutr*. 2003; 77:1146-55.
19. World Health Organization. 2010. *World health statistics 2010*: World Health Organization.
20. Naghavi M, et al. *Lancet*. 2015; 385:117-71. doi.org/10.1016/S0140-6736(14)61682-2
21. Hannum G, et al. *Mol Cell*. 2013; 49:359-67. doi.org/10.1016/j.molcel.2012.10.016
22. Jarman SN, et al. *Mol Ecol*. 2015; 24:4826-47. doi.org/10.1111/mec.13357
23. Mitnitski A, et al. *BMC Med*. 2015; 13:161. doi.org/10.1186/s12916-015-0400-x
24. Berglund K, et al. *Aging (Albany NY)*. 2016; 8:1398-415. doi.org/10.18632/aging.100995
25. Arai Y, et al. *EBioMedicine*. 2015; 2:1549-58. doi.org/10.1016/j.ebiom.2015.07.029
26. Qin L, et al. *Aging (Albany NY)*. 2016; 8:848-59. doi.org/10.18632/aging.100894

**Alan A. Cohen:** Department of Family Medicine, University of Sherbrooke, Sherbrooke, QC J1H 5N4, Canada

**Correspondence:** Alan A. Cohen

**Email:** [Alan.Cohen@USherbrooke.ca](mailto:Alan.Cohen@USherbrooke.ca)

**Acknowledgements:** AAC is funded by a CIHR New Investigator Salary Award and CIHR grant #s 119485, and 145585, and NSERC Discovery Grant #402079-2011.

**Keywords:** aging, biomarker, biological age, deep neural network, epidemiology

**Received:** July 05, 2016

**Published:** September 8, 2016

## Deep biomarkers of human aging: Application of deep neural networks to biomarker development

Evgeny Putin<sup>1,2</sup>, Polina Mamoshina<sup>1,3</sup>, Alexander Aliper<sup>1</sup>, Mikhail Korzinkin<sup>1</sup>, Alexey Moskalev<sup>1,4</sup>, Alexey Kolosov<sup>5</sup>, Alexander Ostrovskiy<sup>5</sup>, Charles Cantor<sup>6</sup>, Jan Vijg<sup>7</sup>, and Alex Zhavoronkov<sup>1,3</sup>

<sup>1</sup>Pharma.AI Department, Insilico Medicine, Inc, Baltimore, MD 21218, USA

<sup>2</sup>Computer Technologies Lab, ITMO University, St. Petersburg 197101, Russia

<sup>3</sup>The Biogerontology Research Foundation, Oxford, UK

<sup>4</sup>School of Systems Biology, George Mason University (GMU), Fairfax, VA 22030, USA

<sup>5</sup>Invitro Laboratory, Ltd, Moscow 125047, Russia

<sup>6</sup>Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

<sup>7</sup>Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA

**Key words:** deep learning, deep neural networks, biomarker development, aging biomarkers, human aging, machine learning

**Abbreviations:** ML, Machine Learning; SVM, Support Vector Machine; DNN, Deep Neural Network; PFI, Permutation Feature Importance; RF, Random Forests; GBM, Gradient Boosting Machine; kNN, k-Nearest Neighbors; DT, Decision Trees; LR, Linear Regression

**Received:** 09/26/15; **Accepted:** 05/09/16; **Published:** 05/18/16

**Correspondence to:** Alex Zhavoronkov, PhD; **E-mail:** [alex@biogerontology.org](mailto:alex@biogerontology.org)

**Copyright:** Putin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** One of the major impediments in human aging research is the absence of a comprehensive and actionable set of biomarkers that may be targeted and measured to track the effectiveness of therapeutic interventions. In this study, we designed a modular ensemble of 21 deep neural networks (DNNs) of varying depth, structure and optimization to predict human chronological age using a basic blood test. To train the DNNs, we used over 60,000 samples from common blood biochemistry and cell count tests from routine health exams performed by a single laboratory and linked to chronological age and sex. The best performing DNN in the ensemble demonstrated 81.5 % epsilon-accuracy  $r = 0.90$  with  $R^2 = 0.80$  and MAE = 6.07 years in predicting chronological age within a 10 year frame, while the entire ensemble achieved 83.5% epsilon-accuracy  $r = 0.91$  with  $R^2 = 0.82$  and MAE = 5.55 years. The ensemble also identified the 5 most important markers for predicting human chronological age: albumin, glucose, alkaline phosphatase, urea and erythrocytes. To allow for public testing and evaluate real-life performance of the predictor, we developed an online system available at <http://www.aging.ai>. The ensemble approach may facilitate integration of multi-modal data linked to chronological age and sex that may lead to simple, minimally invasive, and affordable methods of tracking integrated biomarkers of aging in humans and performing cross-species feature importance analysis.

### INTRODUCTION

Aging is a complex process affecting all biological systems at every level of organization [1, 2]. While many anti-aging interventions have demonstrated life-extending or other geroprotective effects in model organisms, practical limitations continue to hamper

translation to the clinic [3]. One problem is that the evaluation of aging changes and possible anti-aging remedies requires a comprehensive set of robust biomarkers [4]. Large-scale longitudinal programs like MARK-AGE [5] have been launched to analyze changes in multiple biomarkers during aging and correlation between biological and chronological age.

Several “aging clocks” able to predict human chronological age using various biomarkers have already been proposed. Methylation-based markers such as epigenetic aging clocks (Horvath [6] and Hannum [7]) are currently the most accurate, while transcriptomics [8,9] and metabolomics [10] have shown to be less so. Telomere length is commonly used to measure senescence but has lower predictive ability of human chronological age than IgG N-glycans, immunoglobulin G glycosylated at conservative N-glycation sites [11]. Recent studies show that biomarkers of age-related pathologies could be used to evaluate senescence modifications based on the connection between age-related pathologies at the signaling pathway level [12].

However, most of these biomarkers are not representative of the health state of the entire organism or individual systems and are not easily measured or targeted with known interventions. The common blood biochemistry test is one of the simplest tests used by physicians to examine the health state of patients. While being highly variable in nature, some markers from blood biochemistry are sensitive indicators of various conditions, such as inflammation and even alcoholism, and are approved for clinical use [13, 14].

Machine learning (ML) techniques, such as support vector machines (SVM), are routinely used in biomarker development [15] and rapid increases in labeled data are enabling deep neural networks (DNNs). Methods based on deep architectures have outperformed classical approaches not only in image analysis, but also in solving a wide range of genomics, transcriptomics and proteomics problems [16].

In this study, we apply a deep learning technique for predicting human chronological age that utilizes multiple DNNs stacked into an ensemble and trained on tens of thousands of blood biochemistry samples from patients undergoing routine physical examinations. We then use a custom implementation of the permutation feature importance (PFI) technique [17] to evaluate the relative importance of each blood biochemistry marker to ensemble accuracy. We also analyzed the performance and accuracy of 40 DNN architectures optimized using a variety of optimizers, identified the best DNN, and selected 21 DNNs that cumulatively provided higher accuracy and  $R^2$  as an ensemble than the best DNN in the ensemble.

## RESULTS

To perform this study, we obtained a dataset of 62,419 anonymized blood biochemistry records, where each

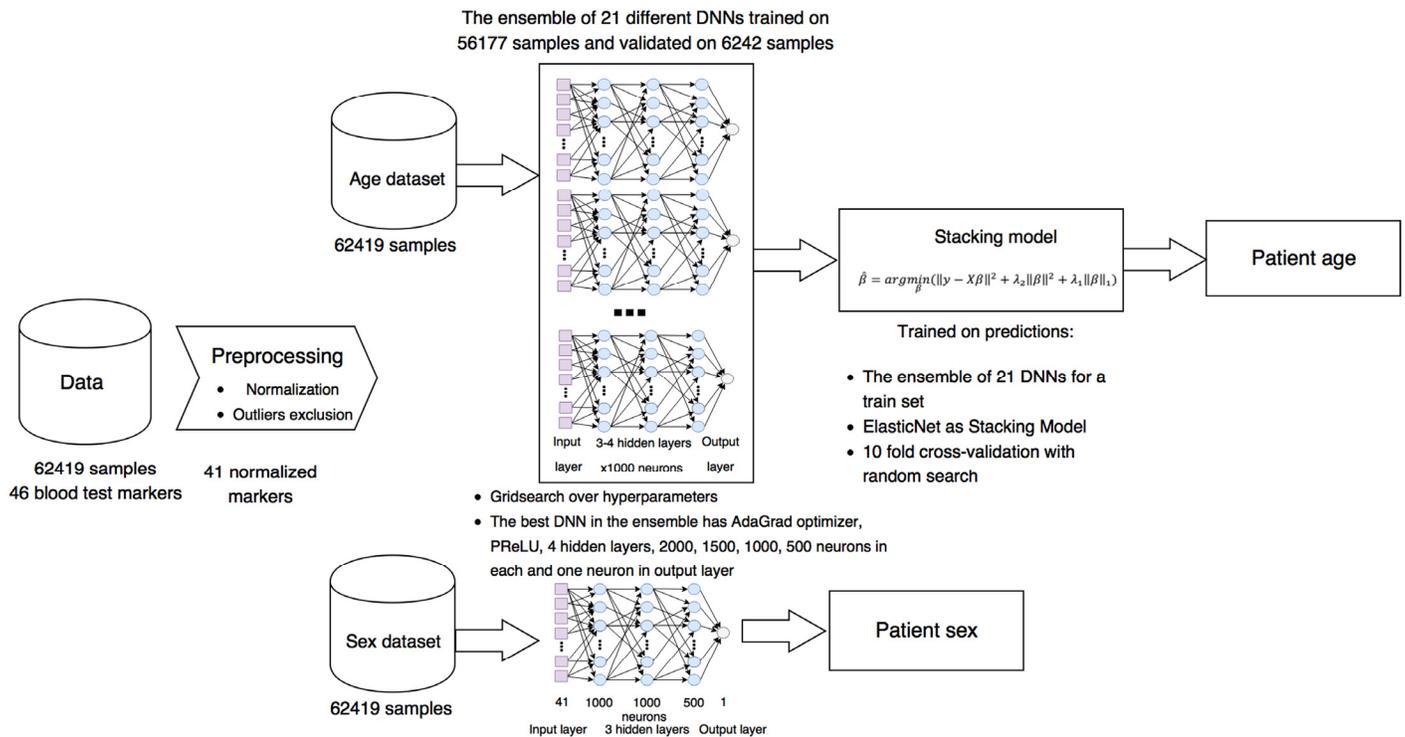
record consists of a person’s age, sex, and 46 standardized blood markers through a collaboration with one of the largest laboratory networks in Russia, Invitro Laboratory, Ltd. We aimed to draw data from a reasonably healthy population. While we did not have access to patient records, we selected only blood tests from routine health checks, avoiding obvious sources of unhealthy patients, such as hospitals, and through statistical analysis omitted blood tests with outliers.

The generalized project pipeline is depicted in Figure 1. First, we preprocessed the blood test data set, excluding highly biased markers from reference ranges, normalizing them for training the DNNs, and removing outliers (see Methods for details). The resulting data set was split into training and test sets comprised of 56,177 and 6242 samples, respectively. Then 40 different DNNs were trained on 56,177 blood test samples.

Since we treated human age prediction as a regression problem, we used two metrics to estimate the performance of the method: standard coefficient of determination ( $R^2$ ) and  $\epsilon$ -prediction (epsilon-prediction) accuracy (see Methods for details). When using epsilon-prediction accuracy, the sample is considered correctly recognized if the predicted age is in the range of  $[\text{true age} - \epsilon; \text{true age} + \epsilon]$ , where  $\epsilon$  controls the level of certainty in the prediction. So if  $\epsilon = 0$ , then it is a simple classification accuracy. In this study, we considered  $\epsilon = 10$ . The key advantage of using epsilon-prediction accuracy is that it allows cohort analysis without fixed age ranges (e.g. 10-20, 20-30).

The best single DNN performed with 0.80 of  $R^2$  and 82% within the 10 year frame of epsilon-prediction accuracy (Figure 2 A & B). Single DNN outperformed other ML models such as k-Nearest Neighbors, Support Vector Machine, Random Forests, Gradient Boosting Machine, etc (Figure 3 & B).

To further increase the coefficient of determination and accuracy of predictions, we combined these single DNNs into an ensemble based on the stacked generalization (Stacking) technique [18]. Stacking is a method that fits some ML models on the predictions of other models, in our case on the predictions of DNNs. Model selection was performed with 10 fold cross-validation and with the random search strategy for finding the best hyperparameters for considered models. The experiments with Stacking models showed (Figure 4 A & B) that the best ML model was ElasticNet.



**Figure 1. Project pipeline.** Laboratory blood biochemistry data sets were normalized and cleaned of outliers and some abnormal markers. For biological age prediction, 21 different DNNs with different parameters were combined in ensemble based on ElasticNet model. For biological sex prediction, single DNN were trained.

To successfully combine the predictions of DNNs into the Stacking ensemble model, the predictions of DNNs should closely approximate the target variable and differ from one another, or be less correlated. To achieve this, DNNs should be trained with different hyperparameters, varying in the number of layers, counts of neurons in each layer, activation functions, regularization techniques, etc. We investigated 40 DNNs, each unique in terms of hyperparameters. Pearson correlations of these DNNs are presented in a heat map on Figure 2 F, showing a high degree of similarity among many of the networks regarding predictions ( $r$  approaching 1) but also some major distinctions.

To determine how many of these trained DNNs were necessary for constructing the Stacking ensemble model, we performed an iterative process of adding each DNN's predictions vector into the ensemble. Two iterative strategies were employed: adding predictions by decreasing  $R^2$  of each network, i.e. adding better networks considering  $R^2$  earliest in the ensemble, and

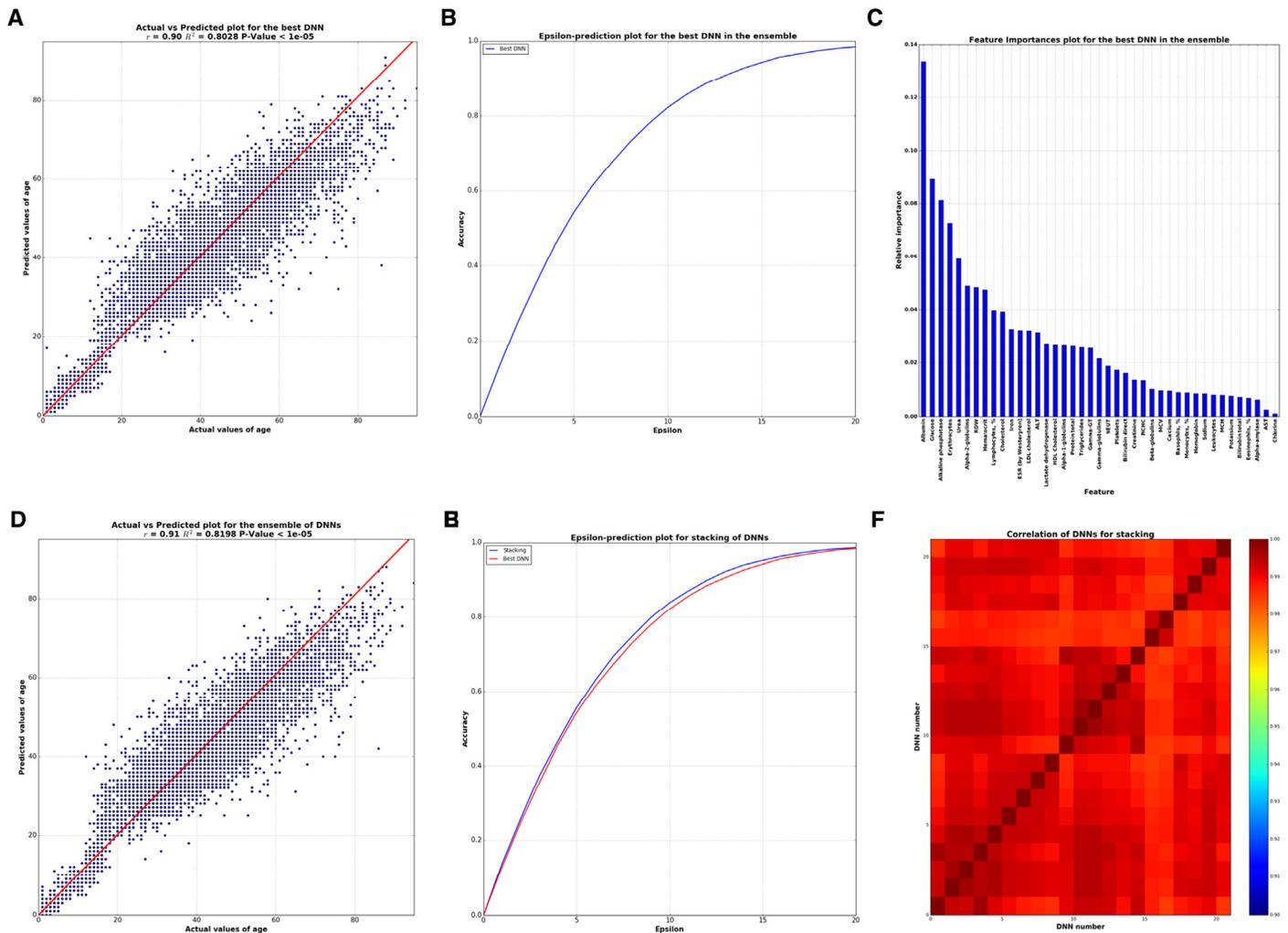
increasing the correlation between DNNs, i.e. adding less correlated networks first. The results of this assay are presented in Figure S2. Both strategies showed that no more than 21 DNNs were needed in the ensemble. The ensemble resulting from distinguishing the correlations of DNNs and ordering the addition of DNNs into the ensemble demonstrated  $R^2=0.82$  and 83,5% within a 10 year frame of epsilon-prediction accuracy (Figure 2 D & E).

We compared our deep-learned predictor with several published epigenetics and transcriptomics markers of human age. Surprisingly, despite the fact that we used only blood biochemistry data with 41 values for each patient, our biomarker outperformed blood transcriptomics biomarkers presented by Peters et al with  $R^2=0,6$  for the best model [8]. Due to the nature of the data, epigenetics markers show a stronger correlation with chronological age, with  $R^2=0,93$  for Horvath's methylation clock and  $R^2=0,89$  for the Hannum et methylation clock [6, 7].

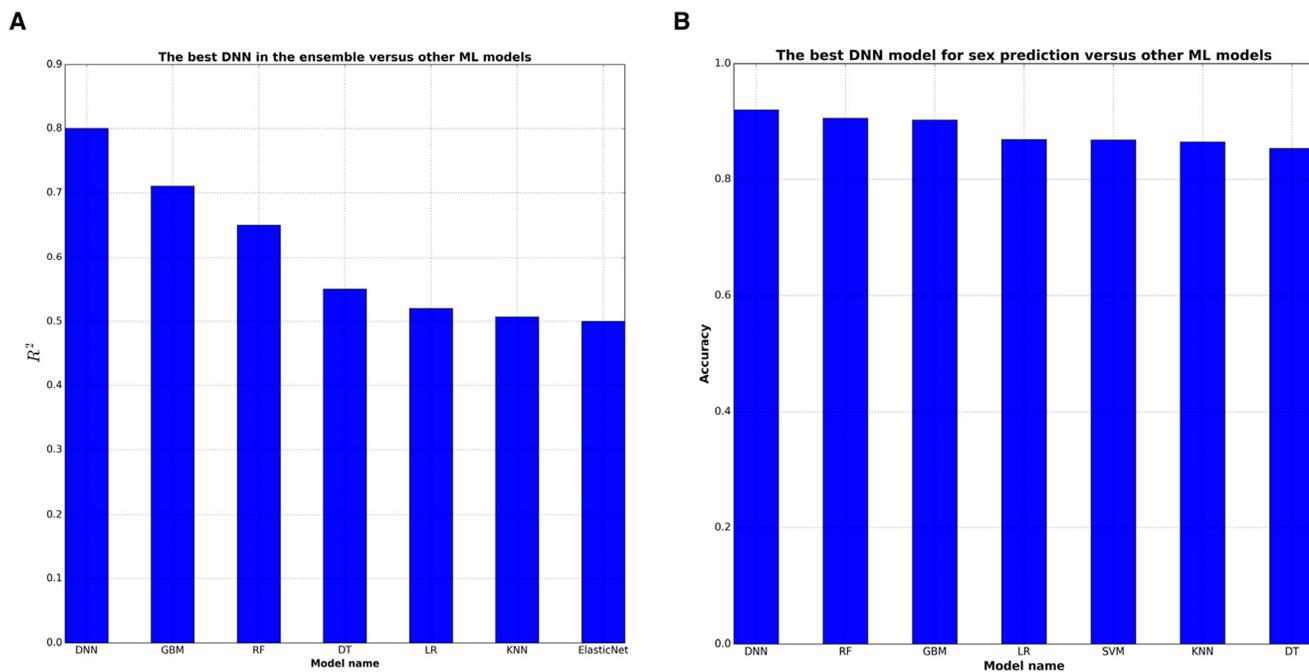
## Marker importance

In order to analyze the importance of blood test markers via neural networks, some wrapper feature (selection) importances approaches are required. We used a modification of the Permutation Feature Importance (PFI) method (see Methods for details). By applying this method, one receives a list sorted by the importance of

markers via DNN. This technique has two benefits: 1) it is native and simple to interpret and 2) as other wrapper methods it relies on DNN performance, which in this case is better than other ML models, thus produces more robust and meaningful features. Marker importance analysis by PFI method, the results of which are presented in Figure 2 C, reveals the five important markers: albumin, glucose, alkaline phosphatase, urea, and erythrocytes.



**Figure 2. Analysis of best DNN model in the ensemble and the whole ensemble.** (A) Correlation between actual and predicted age values by the best DNN in the ensemble. (B) Biological age epsilon-prediction accuracy plot for the best DNN. (C) Biological age marker Importance, performed using PFI method. (D) Correlation between actual and predicted age values by whole ensemble based on ElasticNet model. (E) Biological age epsilon-prediction accuracy plot for the ensemble. (F) Heat map for Pearson's correlation coefficients between 40 DNNs. Scale bar colors indicate the sign and magnitude of Pearson's correlation coefficient between predictions of DNNs.



**Figure 3. DNNs outperform baseline ML approaches in terms of  $R^2$  statistics.** DNN were compared with 7 ML techniques: GBM (Gradient Boosting Machine), RF (Random Forests), DT (Decision Trees), LR (Linear Regression), kNN (k-Nearest Neighbors), ElasticNet, SVM (Support Vector Machines). (A) GBM shows the higher 0,72  $R^2$  among ML models for biological age prediction. (B) All ML models have comparable high  $R^2$  for biological sex prediction.

### Top features

We also performed so-called top features analysis, which answers how the performance of a single DNN will decrease as the number of markers used in the model decreases. To select the smaller number of markers for training the DNN, the sorted list of all PFI scores is used. The results of this analysis for both  $R^2$  and epsilon-prediction accuracy are presented on Figure 5 A & B. For the top 10 features by PFI, the DNN got  $R^2=0.63$  and 70% of 10 year frame epsilon-accuracy prediction. In practical terms, the fact that this drop in performance was so small supports the top 10 markers received by PFI as robust and reliable features for predicting age.

### Use case

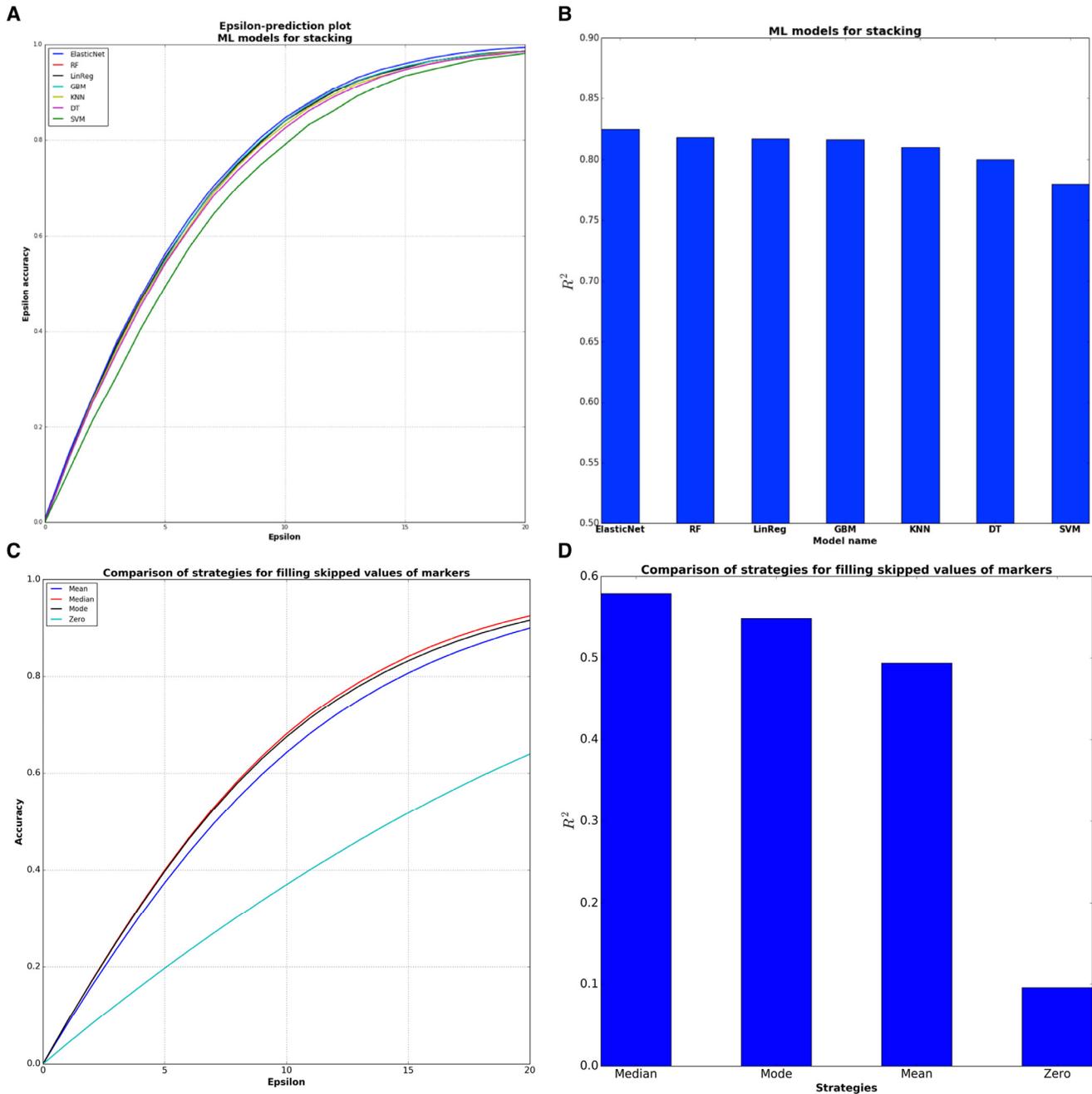
To make this deep network ensemble available to the public, we placed our system online ([www.Aging.AI](http://www.Aging.AI)), allowing any patient with blood test data to predict their age and sex. In order to validate our approach, we collected the blood biochemistry reports that were uploaded on the site from 25 January to 15 March 2016.

The total number of collected reports with indicated real age was 1,563 samples. Many users expressed no desire to specify all 41 parameters of the blood test, so we added an option to enter only the 10 most important markers. The average number of missing values provided by the volunteer testers was 18.5 markers per person. There are several strategies for filling skipped values, including zero, mean, mode and median over all values of each marker. Evaluation of these 4 strategies on the aging.ai data showed that median filling strategy has the best performance in terms of both  $R^2$  and epsilon-prediction accuracy (Figure 4 C & D).

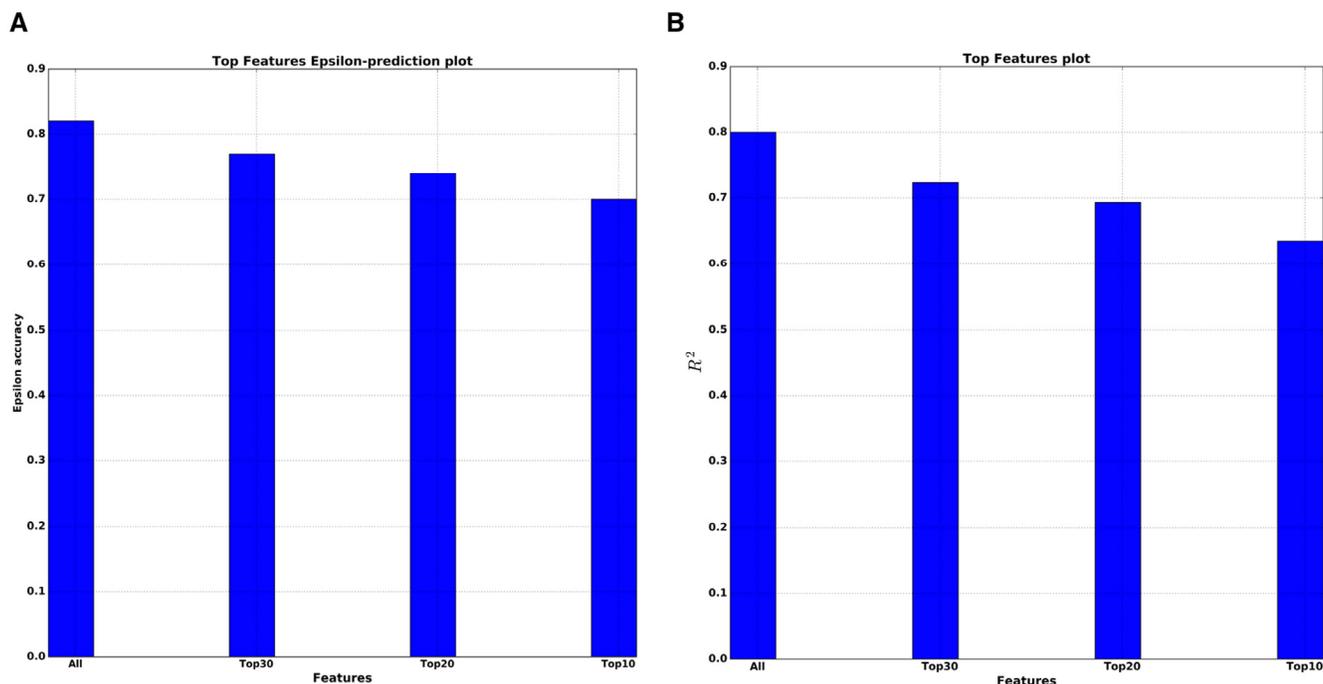
Aging.AI provides a proof of concept for a simple and inexpensive blood-based predictor of chronological age, which may be used for speculate on the biological age of the patient. However, it has many limitations. When it comes to developing predictors using deep neural networks, one of the major difficulties is building large data sets. In this study we were constrained by the limited number of features available to us in large numbers of blood test results. Some of the features, for example globulin fractures, are no longer frequently

used in diagnostic medicine and are excluded from the newer standard tests. However, these features were

present in historical tests available in large numbers and were used for training.



**Figure 4. Comparison of sub-models for stacking ensemble and evaluation of filling strategies.** (A) ElasticNet model has the higher epsilon-prediction accuracy among the stacking models. (B) ElasticNet is the best model for stacking from the point of  $R^2$  statistics. (C) Median filling strategy has higher epsilon-prediction accuracy than other strategies. Median filling strategy shows 64,5 % epsilon accuracy within 10 years frame. (D) Median filling strategy is better from the point of  $R^2$  statistics.



**Figure 5. Top features analysis.** (A) Dependence of the epsilon-prediction accuracy from the number of features. (B) Dependence of  $R^2$  statistics from the number of features.

## DISCUSSION

Aging is a complex process and occurs at different rates and to different extents in the various organ systems, including respiratory, renal, hepatic, and metabolic [19, 20]. The analysis of relative feature importance within the DNNs helped deduce the most important features that may shed light on the contribution of these systems to the aging process, ranked in the following order: metabolic, liver, renal system and respiratory function. The five markers related to these functions were previously associated with aging and used to predict human biological age [21, 22]. Another interesting finding was the extraordinarily high importance of albumin, which primarily controls the oncotic pressure of blood. Albumin declines during aging and is associated with sarcopenia [23]. The second marker by relative importance is glucose, which is directly linked to metabolic health. Cardiovascular diseases associated with diabetes mellitus are major causes of death within the general population [24].

Our approach of using an ensemble of DNNs outperformed other ML models in terms of  $R^2$  and epsilon-prediction accuracy (Figure 3 A & B).

Application of DNNs uncovered complex nonlinear interactions between markers resulting in robust ensemble performance. This ensemble may also be expanded with DNNs trained on different sources and types of biological data allowing for complex multi-modal markers to be created and relative contributions of each input analyzed.

Current and future directions of this work include adding other sources of features including transcriptomic and metabolomics markers from blood, urine, individual organ biopsies and even imaging data as well as testing the system using data from patients with accelerated aging syndromes, multiple diseases and performing gender-specific analysis. Similar tests may be performed by research teams working on rare diseases or working with athletic groups by using <http://www.Aging.AI> system or contacting the authors to perform a high-throughput analysis. Developing similar systems for model organisms and performing PFI analysis may help perform cross-species analysis and of the relative importance of individual markers and organ systems in predicting chronological and biological age.

## MATERIALS AND METHODS

**Data.** Anonymized statistical data of human blood tests was kindly provided by an independent laboratory, Invitro (www.Invitro.ru). No patient records were used in the study. In total, the data contains 62419 records where each record consists of person's age and 46 standardized blood markers, such as Glucose, Cholesterol, Alpha-1-globulins, etc. (Table S1) Histograms of human age for training sets and descriptive statistics of top 10 blood markers used in the research are depicted in the Figure S1 A.

One can see from the Figure S1 B that minimum and maximum values of each marker are far distributed from their normal range values. This distribution reflects patients' tendencies to self-report symptoms and test their health with professional health-care services only in complex cases, which affects their health condition and thus test results [25]. Moreover, we found that there were no patients that could be considered as healthy and who have blood test values within a reference range. The most frequently abnormal markers in the distribution were white blood cell count markers: basophils, abs., eosinophils, abs., lymphocytes abs. monocytes, abs, neutrophils, abs. These types of test provide the total number (absolute number, abs.) of white blood cells in blood microliter. Here, this routine analysis was conducted using a hematology automated analyzer, which counts cells precisely with low error rate [26]. In this case, these aberrant values of markers are more likely linked to the major function of white blood cells; immune function, infections, allergies, smoking [27] or even sleep duration [28] could affect the rate of white blood cells. Additionally, recent studies show a connection between metabolic diseases such as diabetes and range of white blood cells [29, 30]. For this reason, levels of basophils, eosinophils, lymphocytes, monocytes and neutrophils are extremely variable in the general population. To prevent DNN predictions from being highly biased with respect to abnormal ranges of blood markers, we excluded these 5 markers. Processed data was presented in a tabular format of 62419 rows and 42 columns (age and sex + 41 markers).

Then, specifically for training deep neural network, we normalized all blood markers to 0-1 range by using the formula:

$$X_{0-1} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where  $X$  is the origin values of each blood marker,  $X_{min}$  and  $X_{max}$  are its minimum and maximum, respectively and  $X_{0-1}$  is the marker within 0-1 range.

We split the data to the training and test sets with 90/10 ratio. Thus, the size of training and test sets were 56177 and 6242 samples, respectively. The DNN was built by adjusting its hyperparameters (such as a number of layers, activation function, etc.) on the training set and measuring the performance of the trained neural network on the test set. The comparison of performances of 6 best DNNs with different values of hyperparameters is depicted on Table S1. All experiments were conducted on Nvidia Tesla K80 graphics processing unit.

There are two reasons why in the study we treated the prediction of human age as a regression problem: 1) age has natural order, so it is an order variable and 2) one may be interested in the difference in values of the markers with difference in ages, which is the natural way to perform the analysis of marker influence. In this case, it was better to use regression instead of classification methods.

So, in all evaluations 4 metrics were measured:

1.  $r$ , which is a Pearson's correlation coefficient defined as:  $r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$ ; where  $x_i$  is real value and  $\bar{x}$  is the mean of  $x$ ,  $y_i$  is predicted value and  $\bar{y}$  is the mean of  $y$ , and  $N$  is number of samples.
2.  $R^2$ , which is a standard coefficient of determination defined as:  $R^2 = \frac{\sum_{i=1}^N (y_i - f_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$ ; where  $y_i$  is the real value,  $f_i$  is the predicted value and  $\bar{y}$  is the mean of  $y$ .
3. Mean absolute error (MAE), which is defined as  $MAE = \frac{1}{N} \sum_{i=1}^N |f_i - y_i|$ ; where  $f_i$  is a prediction of the model,  $y_i$  is a true value and  $N$  is a number of samples.
4.  $\epsilon$ -prediction accuracy defined as:  $\epsilon - prediction = \frac{\sum_{i=1}^N 1_A(f_i)}{N}$ ; where  $A$  is  $[y_i - \epsilon; y_i + \epsilon]$ ;  $y_i$  is the real value,  $f_i$  is the predicted value and  $\epsilon$  is a parameter that controls the range of correctness of predictions. So for example if  $\epsilon$  is 10 and the true value of age is 45 the deep neural network correctly recognized sample if it is in the [35, 55] range.

**Feature importance method.** The idea behind the algorithm stemmed from the feature randomization technique used in Random Forest (RF) [31]. PFI computes significance scores for all features by determining the accuracy of a model to random permutations of the values of those feature variables. The main underlying assumption is that permuting the values of important features results in a more significant reduction in a model's performance compared to the

effect of less important ones. But when cross-validation is not performed, one should improve the robustness of the method.

To do this, we shuffled each feature  $k$  times and then computed the average PFI score for the feature, concretely the PFI score for one feature is defined as follows:

$PFI_{feature} = R_{total}^2 - \frac{1}{N} \sum_{i=1}^N R_{shuffle}^2$ ; where  $R_{total}^2$  is a total  $R^2$  for the model without any permutations and  $R_{shuffle}^2$  is a  $R^2$  for the model with permuted feature and  $K$  is a parameter that controls how many times the feature is permuted.

Note that PFI is a wrapper method, so it would significantly depend on applying ML model, but because DNNs show better performance than other ML models, it was suitable for the problem.

Architecture of DNN. We used simple feed-forward neural networks trained with the standard backpropagation algorithm as our deep (more than 3 layers) learning models. For each DNN in the resulting ensemble, multiple hyperparameters were adjusted, including the number of hidden layers, the number of neurons in each layer, choice of activation function, choice of optimization method, and regularization techniques. The table with experiments of different hyperparameters for the DNNs are presented in Table S1.

The best DNN in the ensemble had 5 hidden layers with 2000, 1500, 1000, 500, and 1 neurons in each, respectively. The last layer, with one neuron, corresponds to regression output. The optimization loss function was simple mean squared error (MSE) with regularization terms. The DNN used PReLU activation function [32] in each layer, AdaGrad [33] as optimizer of the loss function, Dropout [34] with probability of 0.2 after each layer, and l2 weight decay [35]. To further cope with over fitting and make more stable convergence of models, we used Batch normalization technique [36] after the first 2 layers.

## ACKNOWLEDGEMENTS

We thank Vladimir Parfenov for assistance and helpful comments, Mark Berger from NVIDIA for assistance with the GPU equipment and Dr. Elena Kondrasheva, who made major contribution to data selection and processing and interpretation, but decided that this major contribution did not justify authorship. We would like to thank Dr. Leslie C. Jellen for editing this manuscript. The Aging.AI system was inspired by

How-Old.net by Microsoft, Inc and we would like to thank Dr. Joseph Sirosh for presentations of this system. Finally, we are grateful to all users of the Aging.AI system for helping test and improve the system.

## Funding

This work was financially supported by the Government of Russian Federation, Grant 074-U01 and by a research grant from the Life Extension Foundation 2016-LEF-AA-INSIL. Insilico Medicine is grateful to Nvidia Corporation for providing Tesla K80 GPUs and early access to the NVIDIA DevBox used in this study.

## Conflict of interest statement

The authors are affiliated with Insilico Medicine, Inc, a commercial company developing differential pathway activation scoring-based and deep learned biomarkers of multiple diseases and aging and engaging in drug discovery and drug repurposing. The company has developed a range of drug candidates addressing specific diseases and geroprotector interventions addressing human aging processes that need to be validated in human patients. The company intends to use blood biochemistry and multi-parametric markers, including the one published in this paper to test the efficacy of these compounds. Despite company's commitment to best academic practices and in silico veritas, the authors may have a conflict of interest.

## REFERENCES

1. Zhavoronkov A, Cantor CR. Methods for structuring scientific knowledge from many areas related to aging research. *PLoS One*. 2011; 6:e22597.
2. Moskalev A, Zhikrivetskaya S, Shaposhnikov M, Dobrovolskaya E, Gurinovich R, Kuryan O, et al. Aging Chart: a community resource for rapid exploratory pathway analysis of age-related processes. *Nucleic Acids Res*. 2016; 44:D894–899.
3. Moskalev A, Chernyagina E, de Magalhães JP, Barardo D, Thoppil H, Shaposhnikov M, et al. Geroprotectors.org: a new, structured and curated database of current therapeutic interventions in aging and age-related disease. *Aging (Albany NY)*. 2015; 7:616–728. doi: 10.18632/aging.10.100799.
4. Zhavoronkov A, Alex Z, Bhupinder B. Classifying Aging as a Disease in the context of ICD-1 . *Frontiers in Genetics*. 2015; 6:1–16.
5. Bürkle A, Moreno-Villanueva M, Bernhard J, Blasco M, Zondag G, Hoeijmakers JHJ, et al. MARK-AGE biomarkers of ageing. *Mech. Ageing Dev*. 2015; 151:2–12.
6. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013; 14:R115.
7. Hannum G, Gregory H, Justin G, Ling Z, Li Z, Guy H, et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell*. 2013; 49:359–367.
8. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN,

- Powell J, et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* 2015; 6:8570.
9. Nakamura S, Kawai K, Takeshita Y, Honda M, Takamura T, Kaneko S, et al. Identification of blood biomarkers of aging by transcript profiling of whole blood. *Biochem. Biophys. Res. Commun.* 2012; 418:313–318.
  10. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai P-C, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int. J. Epidemiol.* 2013; 42:1111–1119.
  11. Krištić J, Kri ti J, Vu kovi F, Menni C, Klari L, Keser T, et al. Glycans Are a Novel Biomarker of Chronological and Biological Ages. *J. Gerontol. A Biol. Sci. Med. Sci.* 2013; 69:779–789.
  12. Aliper AM, Csoka AB, Buzdin A, Jetka T, Roumiantsev S, Moskalev A, et al. Signaling pathway activation drift during aging: Hutchinson-Gilford Progeria Syndrome fibroblasts are comparable to normal middle-age and old-age cells. *Aging (Albany NY)*. 2015; 7:26–37. doi: 10.18632/aging.100717.
  13. Kaysen GA. Biochemistry and biomarkers of inflamed patients: why look, what to assess. *Clin. J. Am. Soc. Nephrol.* 2009; 4 Suppl 1:S56–63.
  14. Peterson K. Biomarkers for alcohol use and abuse—a summary. *Alcohol Res. Health.* 2004; 28:30–37.
  15. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 2015; 16:321–332.
  16. Mamoshina P, Polina M, Armando V, Evgeny P, Alex Z. Applications of Deep Learning in Biomedicine. *Mol. Pharm. Internet.* 2016; Available from: <http://dx.doi.org/10.1021/acs.molpharmaceut.5b00982>.
  17. Altmann A, Tološi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics.* 2010;26:1340–1347.
  18. Wolpert DH. Stacked generalization. *Neural Netw.* 1992;5:241–259.
  19. Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J. Gerontol. A Biol. Sci. Med. Sci.* 2013; 68:667–674.
  20. Cheng S, Larson MG, McCabe EL, Murabito JM, Rhee EP, Ho JE, et al. Distinct metabolomic signatures are associated with longevity in humans. *Nat. Commun.* 2015; 6:6791.
  21. Cohen AA, Milot E, Li Q, Bergeron P, Poirier R, Dusseault-Bélanger F, et al. Detection of a novel, integrative aging process suggests complex physiological integration. *PLoS One.* 2015; 10:e0116489.
  22. Park J, Cho B, Kwon H, Lee C. Developing a biological age assessment equation using principal component analysis and clinical biomarkers of aging in Korean men. *Arch. Gerontol. Geriatr.* 2009; 49:7–12.
  23. Visser M, Kritchevsky SB, Newman AB, Goodpaster BH, Tyllavsky FA, Nevitt MC, et al. Lower serum albumin concentration and change in muscle mass: the Health, Aging and Body Composition Study. *Am. J. Clin. Nutr.* 2005; 82:531–537.
  24. World Health Organization. Health in 2015: from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals. WHO (World Health Organization); 2015.
  25. Elnegaard S, Andersen RS, Pedersen AF, Larsen PV, Søndergaard J, Rasmussen S, et al. Self-reported symptoms and healthcare seeking in the general population—exploring “The Symptom Iceberg.” *BMC Public Health.* 2015; 15:685.
  26. Devreese K, De Logi E, Francart C, Heyndrickx B, Philippé J, Leroux-Roels G. Evaluation of the automated haematology analyser Sysmex NE-8000. *Eur. J. Clin. Chem. Clin. Biochem.* 1991; 29:339–345.
  27. Hansen LK, Grimm RH, Neaton JD. The Relationship of White Blood Cell Count to Other Cardiovascular Risk Factors. *Int. J. Epidemiol.* 1990; 19:881–888.
  28. Boudjeltia KZ, Faraut B, Stenuit P, Esposito MJ, Dyzma M, Brohée D, et al. Sleep restriction increases white blood cells, mainly neutrophil count, in young healthy men: a pilot study. *Vasc. Health Risk Manag.* 2008; 4:1467–1470.
  29. Babio N, Ibarrola-Jurado N, Bulló M, Martínez-González MÁ, Wärnberg J, Salaverría I, et al. White blood cell counts as risk markers of developing metabolic syndrome and its components in the PREDIMED study. *PLoS One.* 2013; 8:e58354.
  30. Twig G, Afek A, Shamiss A, Derazne E, Tzur D, Gordon B, et al. White blood cells count and incidence of type 2 diabetes in young men. *Diabetes Care.* 2013; 36:276–282.
  31. Breiman L, Leo B, Michael L, John R. Random Forests: Finding Quasars. *Statistical Challenges in Astronomy.* 2001; 243–254.
  32. He K, Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV) Internet. 2015. Available from: <http://dx.doi.org/10.1109/iccv.2015.123>
  33. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* 2011;12:2121–2159.
  34. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 2014; 15:1929–1958.
  35. Moody J, Hanson S, Krogh A, Hertz JA. A simple weight decay can improve generalization. *Adv. Neural Inf. Process. Syst.* 1995; 4:950–957.
  36. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint.* 2015; arXiv:1502.03167.

## Telomere length is a prognostic biomarker in elderly advanced ovarian cancer patients: a multicenter GINECO study

Claire Falandry<sup>1</sup>, Béatrice Horard<sup>2</sup>, Amandine Bruyas<sup>3</sup>, Eric Legouffe<sup>4</sup>, Jacques Cretin<sup>5</sup>, Jérôme Meunier<sup>6</sup>, Jérôme Alexandre<sup>7</sup>, Valérie Delecroix<sup>8</sup>, Michel Fabbro<sup>9</sup>, Marie-Noëlle Certain<sup>10</sup>, Raymonde Maraval-Gaget<sup>11</sup>, Eric Pujade-Lauraine<sup>7</sup>, Eric Gilson<sup>12</sup>, and Gilles Freyer<sup>13</sup>

<sup>1</sup> Geriatrics and Oncology Unit, HCL Cancer Institute, LBMC, CarMEN Laboratory, Lyon 1 University, Lyon, France;

<sup>2</sup> LBMC, ENS/Lyon, Lyon 1 University, CGphiMC Lyon 1 University, Lyon, France;

<sup>3</sup> Oncology Unit, Lyon Sud University Hospital, Lyon University, Pierre-Bénite, France;

<sup>4</sup> Clinique Valdegour, Department of Medical Oncology, Nîmes, France;

<sup>5</sup> Clinique Bonnefont, Oncology and Radiotherapy Department, Alès, France;

<sup>6</sup> Centre Hospitalier Régional d'Orléans, Department of Medical Oncology, Orléans, France;

<sup>7</sup> Paris Descartes University, AP-HP, Hôpitaux Universitaires Paris Centre, Site Hôtel Dieu, Paris, France;

<sup>8</sup> Clinique Mutualiste de l'Estuaire, Cité Sanitaire, Department of Medical Oncology, Saint-Nazaire, France;

<sup>9</sup> Institut du Cancer Montpellier, Medical Oncology, Montpellier, France;

<sup>10</sup> Centre Hospitalier d'Auxerre, Department of Medical Oncology, Auxerre, France;

<sup>11</sup> Lyon Sud University Hospital, Pierre-Bénite, France;

<sup>12</sup> LBMC, Lyon 1 University, IRCAN, CNRS UMR 7284, INSERM U1081, Nice Sophia-Antipolis University; CHU of Nice, Nice, France;

<sup>13</sup> HCL Cancer Institute, Department of Medical Oncology, Lyon 1 University, Lyon, France.

**Key words:** telomere, ovarian cancer, elderly, geriatric, prognostic factor

**Received:** 09/06/15; **Accepted:** 10/30/15; **Published:** 12/03/15

**Written:** on behalf of the GINECO (French National Group of Investigators for the Study of Ovarian and Breast Cancer).

**Presented:** at the 48th Annual Meeting of the American Society of Clinical Oncology, Chicago, IL, June 1-5, 2012.

**Correspondence to:** Claire Falandry, PhD; **E-mail:** [claire.falandry@chu-lyon.fr](mailto:claire.falandry@chu-lyon.fr)

**Copyright:** Falandry et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** Purpose: Age induces a progressive decline in functional reserve and impacts cancer treatments. Telomere attrition leads to tissue senescence. We tested the hypothesis that telomere length (TL) could predict patient vulnerability and outcome with cancer treatment.

**Patients and methods:** An ancillary study in the Elderly Women GINECO Trial 3 was performed to evaluate the impact of geriatric covariates on survival in elderly advanced ovarian cancer patients receiving six cycles of carboplatin. TL was estimated from peripheral blood at inclusion using standard procedures.

**Results:** TL (in base pairs) was estimated for 109/111 patients (median 6.1 kb; range [4.5-8.3 kb]). With a cut-off of 5.77 kb, TL discriminated two patient groups, long telomere (LT) and short telomeres (ST), with significantly different treatment completion rates of 0.80 (95%CI [0.71-0.89]) and 0.59 (95%CI [0.41-0.76]), respectively (odds ratio [OR]=2.8, p=0.02). ST patients were at higher risk of serious adverse events (SAE, OR=2.7; p=0.02) and had more unplanned hospital admissions (OR=2.1; p=0.08). After adjustment on FIGO stage, TL shorter than 6 kb was a risk factor of premature death (HR=1.57; p=0.06).

**Conclusion:** This exploratory study identifies TL as predictive factor of decreased treatment completion, SAE risk, unplanned hospital admissions and OS after adjustment on FIGO stage.

## INTRODUCTION

Aging is associated with a progressive decline in the functional reserve of multiple organ systems [1]. Given that the process of aging is heterogeneous, this decline should ideally be assessed individually and care of an elderly person adapted accordingly rather than solely on the basis of chronological age. Such assessments are currently entirely clinical, based on a geriatric evaluation.

During normal ageing, the gradual loss of telomeric DNA in dividing somatic cells contributes to replicative senescence [2]. Importantly, this telomere length dynamics plays an important signaling role in determining cell fate during aging and cancer [3]. There is growing evidence linking pathologic aging to telomere shortening in prospective studies recruiting elderly patients, although there is some controversy associated with this research. Patients with shorter telomeres tend to develop more functional disabilities [4], have increased cognitive loss [5], higher cardiovascular morbidity [6], more degenerative diseases [7] and higher mortality [8].

In an oncologic context, the impact of aging on a patient's survival is challenged by the nature of the tumor itself, which in turn means a differential impact of geriatric covariates on overall survival (OS). In 1997, the French National Group of Investigators for the Study of Ovarian and Breast Cancer (GINECO) established a research program focused on the treatment of ovarian cancer in elderly women. The feasibility of carboplatin-cyclophosphamide and standard carboplatin AUC5-paclitaxel protocols in patients over 70 years of age was demonstrated in two studies [9,10], with treatment completion rates of 76% and 68% respectively [10]. A multivariate analysis performed in a non-randomized retrospective review of these trials reported a significant negative impact of various geriatric covariates on survival [10]. A prospective trial, the Elderly Woman GINECO Trial 3, was thus initiated to evaluate the impact of geriatric covariates on survival in elderly patients with advanced ovarian cancer treated with six cycles of carboplatin AUC5. A geriatric vulnerability score (GVS) was developed which segregates patients into two groups with significantly different outcomes in terms of treatment completion rates and risk of treatment toxicities (serious and severe adverse events, unplanned hospital admissions, [11]).

An ancillary study was envisaged in the original design of the GINECO Trial 3 with a working hypothesis that telomere biology influences patients' future outcomes and may correlate with a clinical geriatric assessment.

The results reported here pinpoint an association between short TL, treatment tolerance and completion in ovarian cancer patients.

## RESULTS

### **An overall decrease in telomere length with age in the patient cohort used in this study**

Duplicate telomere length (TL) distribution measurements were performed on blood samples from 109 of the 111 patients included between August 2007 and January 2010. Patient characteristics and outcome of the geriatric assessment are shown in Table 1. Median follow-up was 16.4 months (range 0.2-49.6).

TL ranged from 4.52 to 8.33 kilobases (kb) with a mean of 6.05 kb (SD 0.71 kb). A weak inverted linear correlation with age was demonstrated, with every 1-year increase in age associated with a 26-base pair decrease in mean TL, with a  $R^2$  ratio of 0.0341 (Figure 1).

### **Longer TL in patients that completed their treatment and exhibited a better tolerance**

Since treatment completion is considered a meaningful short-term marker of patient outcome, we analyzed for a possible correlation between TL distribution and the fact that patients completed their treatment. In order to analyze variations in the non-gaussian distributions of human TL, we segregated the patients according to TL quartiles (Table 2). Patients with telomeres longer than 6.54 kb had a 52% higher chance of treatment completion than those with telomeres shorter than 5.77 kb (95%CI: 1.00-2.32,  $P=0.04$ ). Thus, we can identify two groups of patients with different treatment completion rates according to their TL distribution: 80% (95%CI: 71% to 89%) in the group with telomeres longer than 5.77 kb (LT group) versus 59% (95%CI: 41% to 76%) in the group with shorter telomeres (ST group,  $P=0.02$ ).

Using the same cut-off of 5.77 kb, TL segregates the same two groups (i.e., ST vs LT) as having different outcomes in terms of tolerance. Serious adverse events were significantly more frequent in the ST group, with an odds ratio of 2.7 ( $P=0.02$ ). Unplanned hospital admissions and grade 3-4 non-hematological toxicity also tended to be more frequent (Table 3). No significant difference between TL groups could be identified in terms of hematological toxicity, however blood cell counts were only routinely evaluated 1 day prior to chemotherapy.

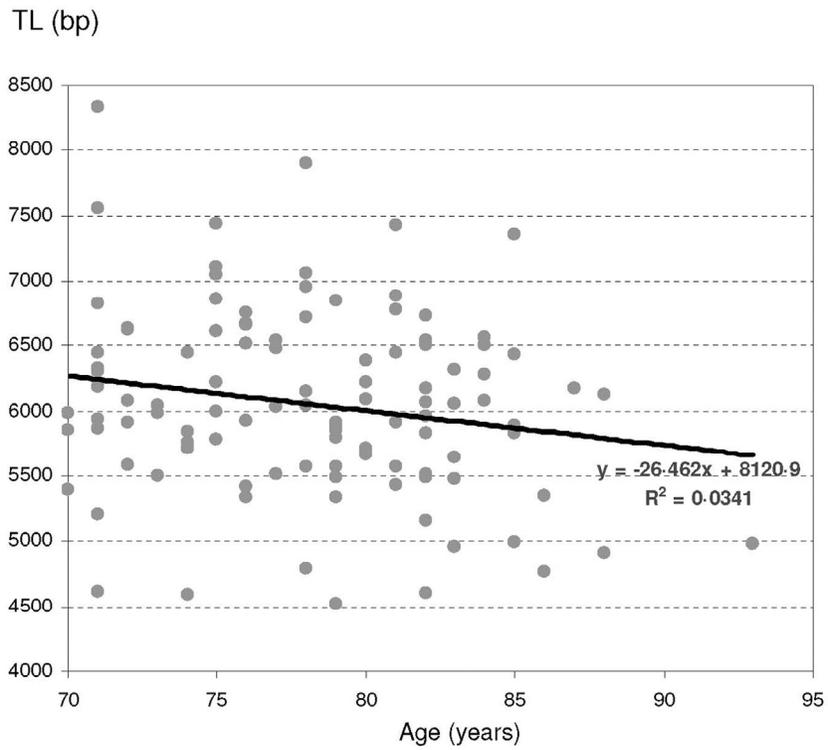


Figure 1. Telomere length repartition according to age.

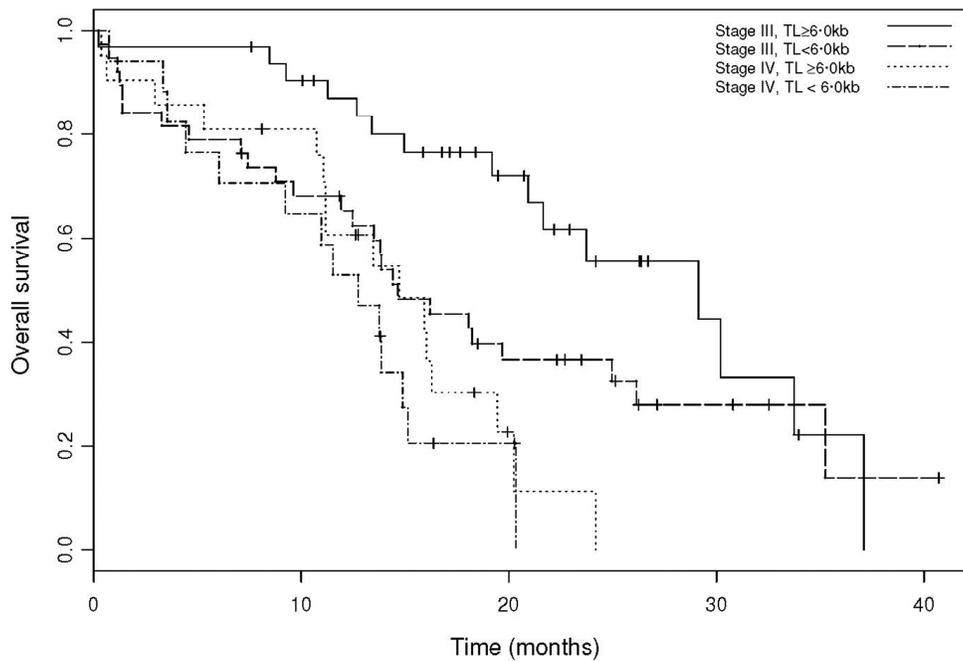


Figure 2. Overall survival by TL groups, adjusted for FIGO stage.

**Table 1. Patient and disease characteristics and geriatric assessment**

	N of patients (%)
Median age in years (range)	78 (70-93)
≥80 years	44 (40.3)
Performance status (ECOG) ≥2	51 (46.8)
Tumor assessment	
FIGO stage IV	38 (34.9)
Complete primary cytoreduction	18 (16.5)
Geriatric assessment	
≥3 comorbidities	26 (23.9)
N comedications	
1-3	32 (29.4)
4-6	44 (40.4)
≥7	30 (27.5)
Functional assessment	
ADL score <6	60 (55.0)
IADL score <25	76 (69.7)
Nutritional assessment	
Albuminemia <35 g/L	64 (58.8)
BMI <21 kg/m <sup>2</sup>	24 (22.0)
Lymphocyte count <1 G/L	27 (24.8)
Psychocognitive assessment	
MMS score <25	32 (29.4)
HADS score >14	40 (36.7)
GDS score >10	39 (35.6)

ADL: Activities of Daily Living; IADL: Instrumental Activities of Daily Living; BMI: Body Mass Index; ECOG: Eastern Cooperative Oncology Group; GDS: Geriatric Depression Scale; HADS: Hospital Anxiety and Depression Scale; MMS: Mini-Mental Scale

### Correlation between TL distribution and overall survival

A survival analysis using Cox proportional hazards was conducted in order to evaluate the impact of telomere length on survival (Figure 2). Reasons for death were the most frequently related to cancer progression (105 patients), 2 patients died from treatment toxicity (septic shock), 4 from other reasons (colic perforation: 1; pulmonary embolism: 1, suspicion of pulmonary embolism: 1, major depression and denutrition: 1). After adjustment on FIGO stage (IV versus III), TL less than 6.00 kb was identified as a risk factor for premature death, with an HR of 1.57 (95%CI: 0.98 to 2.51,  $P=0.06$ ). This pejorative trait staid robust in different models including FIGO stage and age [HR = 1.58 (95%CI: 0.99 to 2.53,  $P=0.06$ )], FIGO stage and GVS ≥ 3 [HR = 1.56 (95%CI: 0.97 to 2.49,  $P=0.07$ )]

and FIGO stage, GVS ≥ 3 and age [HR = 1.57 (95%CI: 0.98 to 2.53,  $P=0.06$ )] (Supplementary table 1).

### A tendency towards correlation between TL and geriatric vulnerability parameters

We then tested the correlation between TL groups and geriatric vulnerability groups using the Geriatric Vulnerability Score [11]. Patients displaying at least three geriatric vulnerability parameters had a 2.94-fold higher risk of mortality in a univariate analysis (95%CI 1.79-4.84,  $P < 0.0001$ ) and this was 2.89-fold in a multivariate analysis after adjustment for FIGO stage (95%CI 1.74-4.78,  $P < 0.0001$ ). Despite the absence of a significant correlation with patients' characteristics except age (Supplementary table 2) and any of the individual geriatric vulnerability parameters, TL groups and geriatric vulnerability groups showed a tendency towards correlation ( $P=0.08$ , Table 3).

**Table 2. Association between telomere length parameters and patient outcomes**

	Telomere length	Odds ratio for treatment completion (95%CI)	P-value	Hazard ratio for death (95%CI) (adjusted for FIGO stage)	P-value
TL mean	6.05 kb	0.56 (0.22-1.27)	0.15	1.42 (0.88-2.29)	0.15
TL median	6.00 kb	0.50 (0.21-1.2)	0.11	1.57 (0.98-2.51)	0.06
TL quartiles		<b>1.50 (1.01-2.23)</b>	<b>0.04</b>	<b>0.82 (1.67-1.01)</b>	<b>0.07</b>
	<5.77 kb	<b>0.36 (0.15-0.87)</b>	<b>0.02</b>	1.49 (0.91-2.44)	0.12
	5.77-6.06 kb	1.60 (0.54-4.74)	0.38	1.02 (0.59-1.77)	0.94
	6.06-6.54 kb	1.28 (0.46-3.58)	0.64	0.97 (0.55-1.72)	0.91
	>6.54 kb	2.08 (0.65-6.67)	0.19	0.62 (0.34-1.14)	0.12

**Table 3. TL repartition of vulnerability criteria and clinical end points between**

	Observed risk: short/long telomere group	95% CI	P-value
Treatment completion	0.36	0.15-0.87	0.020
Serious Adverse Events	2.69	1.17-6.19	0.019
Unplanned hospital admissions	2.14	0.92-4.95	0.076
Grade $\geq$ 3 non-hematological toxicity	2.04	0.88-4.71	0.095
Grade $\geq$ 3 hematological toxicity	1.32	0.58-3.00	0.51
<b>Geriatric vulnerability parameters :</b>			
ADL score < 6	1.78	0.77-4.12	0.17
IADL score < 25	1.31	0.53-3.22	0.56
HADS score >14	1.89	0.82-4.33	0.13
Albuminemia <35 g/L	1.27	0.55-2.90	0.57
Lymphocytes <1 x 10 <sup>9</sup> /L	1.76	0.71-4.36	0.22
<b>Geriatric vulnerability score</b>			
GVS $\geq$ 3	2.06	0.90-4.70	0.08

## DISCUSSION

This analysis was planned as an ancillary study in the prospective multicentric Elderly Woman GINECO Trial 3. The working biological hypothesis - identifying TL as a putative prognostic biomarker in elderly cancer patients - was based on the pooled results of two prior studies, Elderly Woman GINECO Trials 1 and 2. In a retrospective multivariate analysis of overall survival (OS), several factors - FIGO stage IV, use of paclitaxel, age, emotional disorders and lymphopenia - were significantly associated with an increased risk of premature death ([10] and unpublished data). Moreover,

a significant correlation was also shown between emotional disorders and lymphopenia. TL shortening, previously shown to correlate with age [13,14], lifestyle stress [15] and survival [10], is considered to be both an actor and a witness of the pathologic aging process. The myeloid skewing of hematological progenitors that accompanies aging [16] and telomere dysfunction [17,18] could explain why there is an increasing amount of clinical data proposing lymphopenia as a marker of pathologic aging [19].

In previous epidemiological cohorts, a correlation was observed between TL and lifespan [8,20,21], aging

associated diseases [7], and OS in healthy subjects [22]. These findings are of limited clinical use since associations only appear when large cohorts are investigated and the predictive value on an individual basis is poor. TL in blood leucocytes has also gained considerable interest as a potential biomarker of cancer risk, and direct measurement of TL and telomerase activity in tumors are considered to be cancer prognosis markers [23, 24].

To our knowledge, this study is the first to investigate the impact of TL on treatment feasibility as an individual risk factor. In spite of a relatively small patient sample size (111 patients), we were able to control several putative biased errors typically associated with TL measurement in clinical trials [25-27]. All patients were female, post-menopausal, almost exclusively Caucasian and fell into a narrow age bracket (70 to 93 years). According our biologic working model, the impact of TL shortening was expected to be challenged by the competition between tumor-related and host-related covariates on patients' outcomes. Due to the context of the trial, the study did not include any external reference, as for example TL samples from non-cancer elderly patients, that would have evaluated the impact of the tumor itself on TL.

Despite these constraints, in this particular oncologic context of ovarian cancer, our results reveal a clear correlation between TL and patient immediate outcomes. Indeed, TL distribution identified a subgroup of elderly patients with short telomeres who have a lower probability of completing treatment and a higher risk of severe adverse events and unplanned hospitalization [28]. However, the translation to clinical practice of these results might be difficult for the following reasons. Firstly and as usually for TL analyses, the cut-off between short and long telomeres were made *a posteriori*, being highly dependent on the technical conditions and the population studied. Moreover, different cut-offs separated the immediate outcomes (treatment completion, severe adverse events, unplanned hospital admissions) and the risk of premature death. Secondly, TL remained less discriminating than the GVS, based on simple clinical tests and routine bioassays, for immediate outcomes and survival. Finally, TL was estimated using the gold standard technique, namely the mean length terminal restriction fragments. Even if this technique is highly feasible, it is time-consuming and difficult to implement in routine analysis. In this respect, many of large epidemiological cohort analyses have preferred an alternative method of TL estimation, namely quantitative PCR [29,30]. However, this alternative technique suffers from a number of technical

disadvantages, notably a high coefficient of variation and a lack of good reference standards, making it difficult to evaluate absolute TL [31]. Studies in mouse models have revealed that the number of dysfunctional telomeres is a more accurate factor than mean telomere length in the evolution of tissue pathology during aging [32]. Thus, evaluation of blood markers of dysfunctional telomeres [33] appear to be a promising method for telomere biology analysis in clinical trials [34, 35].

In conclusion, our study demonstrates a correlation between telomere length and patient outcomes in an oncogeriatric context [34]. This finding opens the way for future work aimed at identifying telomere biomarkers which can be implemented in routine practice for outcome prediction as well as on the evaluation of the impact of cancer treatment - mainly chemotherapy - on biomarkers of aging.

## MATERIALS AND METHODS

**Study design.** The Elderly Woman GINECO Trial 3 was an open-label phase II multicentric trial approved by the Independent Ethics Committee of Lyon University Hospital (EUDRACT No. 2006-005504-13). The study design, population and assessments have been described elsewhere [1]. Written informed consent was obtained from each patient and included authorization for collection of a blood sample for TL measurement. Patients were treated with up to six cycles of carboplatin AUC5 (5 mg/mL/min for 30 min every 3 weeks). In this ancillary study, TL at inclusion was evaluated, along with the impact of TL on patient outcome and the correlation between TL and geriatric covariates.

**Patient population.** Eligible patients were  $\geq 70$  years old, with a life expectancy  $\geq 3$  months, and histologically or cytologically proven epithelial FIGO stage III-IV ovarian cancer. Cytology consistent with ovarian cancer was considered sufficient if associated with both a CA125 rise and a radiological pelvic mass. Patients were considered ineligible if they had any prior malignancy except basal cell carcinoma or carcinoma *in situ* of the cervix or urinary bladder, prior chemo- or radiotherapy, serious medical or psychiatric illness that might affect treatment, major disturbance of hepatic parameters (alanine aminotransferase or aspartate aminotransferase  $> 3$  times the upper limit of normal, total bilirubin  $> 2$  times the upper limit of normal), severe renal insufficiency (creatinine clearance  $< 30$  mL/min), or abnormal hematological parameters (neutrophils  $< 1.5 \times 10^9/L$ , platelets  $< 100 \times 10^9/L$ ). Patients with planned interval debulking surgery were also excluded.

**Assessments.** A multidimensional pre-inclusion geriatric assessment was performed at baseline. Data concerning the patient's medical charts, nutrition, functionality and an extensive psychocognitive assessment were collected, including comorbidities, comedication, body mass index (BMI), serum albumin levels, and functional scores for Activities of Daily Living (ADL), Instrumental ADL (IADL), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and the Mini-Mental Scale (MMS).

Patient outcomes of treatment completion rate (defined as receiving six courses of chemotherapy without premature discontinuation for death, treatment toxicity or tumor progression), survival, serious adverse events, unplanned hospital admissions, and grade  $\geq 3$  toxicities were recorded.

The GVS was calculated for each patient as described previously [11]. This score is the addition of the following geriatric vulnerability parameters: ADL score  $< 6$ , IADL score  $< 25$ , HADS score  $> 14$ , albuminemia  $< 35$  g/L, and lymphopenia  $< 1 \times 10^9$ /L.

**Measurement of telomere length.** A blood sample was collected at inclusion. DNA extraction was performed within 14 days using the PAXgene Blood DNA System (PreAnalytix GmbH, Hombrechtikon, Switzerland) according to the manufacturer's instructions. DNA integrity was assessed by electrophoresis on 1.0% agarose gels. DNA samples (4  $\mu$ g) were digested overnight with the restriction digest set HinfI (33 U) /RsaI (33 U) (New England Biolabs, France) and resolved using field inversion electrophoresis (FIGE) on FIGE Mapper System (Bio-Rad Life Science, France). Briefly, samples were precipitated, resuspended in 12  $\mu$ l of H<sub>2</sub>O and run on 1% pulse field grade agarose gel (20cm x 13cm) containing 0.5X TBE at room temperature for 13h. The switch time ramp was between 0.1 and 0.5s (linear shape) with forward and reverse voltages of 160 and 100 V, respectively. A combination of two DNA molecular weight size standards was run on each gel: 1 Mix Marker 19 that spans 48.5 – 1.5kb and MassRuler™ DNA Ladder mix that spans 10 – 0.08kb (Thermo Scientific Molecular Biology Inc., France) and used to establish a standard curve (molecular size as a function of migration distance). Digested DNA were blotted to N+ Hybond membrane (GE Healthcare, France) by capillary transfer using SCC 20 $\times$  transfer buffer and then UV cross-linked. Hybridization was carried out overnight at 65°C in hybridization buffer (0.5M NaPO<sub>4</sub> pH7.2, 7% SDS, 0.1% BSA, 1mM EDTA) containing a digoxigenin (DIG)-labeled probe specific for telomeric repeats (400

bp of repeated 5'-T2AG3-3' motif). Membranes were then washed twice at room temperature in 2X SSC, 0.1% SDS (5min) and twice at 50°C in 0.2X SSC, 0.1% SDS (25min). Chemiluminescence detection was carried out according to TeloTAGGG Telomere Length Assay (Roche Applied Science, France) instructions. Telomeric Restriction Fragments (TRF) chemiluminescence signals were captured using a LAS-3000 Imager (FujiFilm Life Science, France) and images were processed using ImageJ software (<http://rsb.info.nih.gov/ij/>). The optical densities (OD) versus mean TRF length were calculated according to the formula ( $\sum ODi / \sum (ODi/MWi)$ ), where ODi is the chemiluminescent signal and MWi is the length of the TRF at position i ([13] and Supplementary figure 1). Measurements were performed on each sample at least twice in different gels and the mean was used for statistical analyses. Pearson's correlation coefficient for duplicates was 0.74, with an average coefficient of variation for pair sets of 7.3%. The laboratory conducting the TL measurement was blinded to all patient characteristics.

**Statistical analyses.** The sample size of 110 patients was calculated on the basis of the primary objective of the main part of the study (to confirm the impact of psychogeriatric covariates on OS) as reported elsewhere [12]. TL was analyzed both as a continuous ordinal variable and a categorical variable. For the former analysis, non-parametric two-sample Wilcoxon rank-sum tests were performed to evaluate the impact of TL on patient outcome. For the latter analysis, TL was transformed into quartiles, categorized into TL groups (shorter vs longer) and introduced as a dichotomous trait into linear regression models and Cox's proportional-hazards regression models. Different cut-offs were used based on the outcome under consideration. Survival curves were estimated using the Kaplan-Meier method and OS models were adjusted for FIGO stage (IV versus III). Odd ratios (ORs) and hazard ratios (HRs), 95% confidence intervals (CIs), and p values (P) were calculated. Analyses were performed using R statistical package (R Foundation for Statistical Computing, Austria) and Splus, version 6.2 (Insightful Corp., WA, USA).

## ACKNOWLEDGEMENTS

The authors would like to acknowledge Douglas Micheau-Bonnier, Nicolas Gane, and Benedicte Votan from the GINECO study office and the Delegation a la Recherche Clinique et a l'Innovation of the Hospices Civils de Lyon. We also thank the following investigators who participated in the trial: Pr. H. Cure (Institut Jean Godinot, Reims), Dr J. Salvat (Hopitaux

du Leman, Thonon-les-Bains), Dr M. Combe (Centre Hospitalier du Mans, Le Mans), Dr. M.C. Kaminsky (Centre Alexis Vautrin-Brabois, Vandoeuvre-les-Nancy), Dr. I. Ray-Coquard (Centre Leon Berard, Lyon), Dr. G. Yazbek, (Institut Jean Godinot, Reims), Dr. J. Meunier (Centre Hospitalier Regional d'Orleans, Orleans), Dr. J. Cretin (Clinique Bonnefon, Ales), Dr. L. Chauvenet, (Hopital Hotel-Dieu, Paris), Dr J. Provencal (Centre Hospitalier de la region d'Annecy, Pringy), Dr. M. Fabbro (Hopital Arnaud de Villeneuve, Montpellier), Dr. M.N. Certain (Centre Hospitalier d'Auxerre, Auxerre), Dr. J.P. Guastalla (Centre Leon Berard, Lyon), Dr. S. Kalla (Groupe Hospitalier Saint-Joseph, Paris), Pr. J. Alexandre (Hopital Hotel-Dieu, Paris), Dr. E. Legouffe (Clinique de Valdegour, Nimes), Dr. F. Savinelli (Groupe Hospitalier Saint-Joseph, Paris), Dr. J.M. Tigaud (Hopital Hotel-Dieu, Paris), Dr. R. Largillier (Centre Azureen de Cancerologie, Mougins), Pr. O. Gisserot (Hopital d'Instruction des Armees Sainte-Anne, Toulon), Dr. C. Ligeza-Poisson (Clinique Mutualiste de l'Estuaire, Saint-Nazaire), Dr. G. Deplanque (Groupe Hospitalier Saint-Joseph, Paris), Dr. P. Deguiral (Clinique Mutualiste de l'Estuaire, Saint-Nazaire), Dr. E. Luporsi (Centre Alexis Vautrin-Brabois, Vandoeuvre-les-Nancy), Dr. F. Priou (Centre Hospitalier Departemental Les Oudairies, La Roche-sur-Yon), Dr. F. Rousseau (Institut Paoli Calmettes, Marseille), Dr. A. Le Rol (Hopital Perpetuel Secours, Levallois-Perret). We thank Sarah MacKenzie, PhD for language editing (funded by ARCAGY-GINECO).

## Funding

This work was supported by research grants from the French Ministry of Health (Projet Hospitalier de Recherche Clinique) and La Fondation de France and promoted by the Hospices Civils de Lyon. EG Lab was supported by "La Ligue Nationale Contre le Cancer" (Équipe labellisée), the "Institut National du Cancer" (EG Lab: INCa, TELOFUN program) and the "Agence Nationale de la Recherche" (ANR, TELOREP program). CF held fellowships from "la Région Rhône-Alpes", the "Association de Recherche contre le Cancer" and Astra-Zeneca.

## Conflict of interest statement

The authors declare no conflicts of interest.

## REFERENCES

1. Balducci L, Extermann M. Management of cancer in the older person: A practical approach. *Oncologist* 2000; 5:224-237.
2. Gilson E, Geli V. How telomeres are replicated. *Nat Rev Mol Cell Biol* 2007; 8:825-838.

3. Ye J, Renault VM, Jamet K, Gilson E. Transcriptional outcome of telomere signalling. *Nat Rev Genet* 2014; 15:491-503.
4. Risques RA, Arbeev KG, Yashin AI, Ukrainitseva SV, Martin GM, Rabinovitch PS, Oshima J. Leukocyte telomere length is associated with disability in older u.S. Population. *J Am Geriatr Soc* 2010; 58:1289-1298.
5. Yaffe K, Lindquist K, Kluse M, Cawthon R, Harris T, Hsueh WC, Simonsick EM, Kuller L, Li R, Ayonayon HN, Rubin SM, Cummings SR: Telomere length and cognitive function in community-dwelling elders: Findings from the health abc study. *Neurobiol Aging* 2011; 32:2055-2060.
6. Epel ES, Merkin SS, Cawthon R, Blackburn EH, Adler NE, Pletcher MJ, Seeman TE. The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. *Aging (Albany NY)* 2008; 1:81-88.
7. Atzmon G, Cho M, Cawthon RM, Budagov T, Katz M, Yang X, Siegel G, Bergman A, Huffman DM, Schechter CB, Wright WE, Shay JW, Barzilai N, Govindaraju DR, Suh Y. Evolution in health and medicine sackler colloquium: Genetic variation in human telomerase is associated with telomere length in ashkenazi centenarians. *Proc Natl Acad Sci U S A* 2010; 1:1710-1717.
8. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* 2003; 361:393-395.
9. Freyer G, Geay JF, Touzet S, Provencal J, Weber B, Jacquin JP, Ganem G, Tubiana-Mathieu N, Gisserot O, Pujade-Lauraine E. Comprehensive geriatric assessment predicts tolerance to chemotherapy and survival in elderly patients with advanced ovarian carcinoma: A gineco study. *Ann Oncol* 2005; 16:1795-1800.
10. Trédan O, Geay JF, Touzet S, Delva R, Weber B, Cretin J, Provencal J, Martin J, Pujade-Lauraine E, Freyer G. Carboplatin cyclophosphamide or carboplatin paclitaxel in elderly with advanced ovarian cancer? Analysis of two consecutive trials from the gineco. *Annals of Oncology* 2007; 18:256-262.
11. Falandry C, Weber B, Savoye A-M, Tinquaut F, Tredan O, Sevin E, Stefani L, Savinelli F, Atlassi M, Salvat J, Pujade-Lauraine E, Freyer G. Development of a geriatric vulnerability score in elderly patients with advanced ovarian cancer treated with first-line carboplatin: A gineco prospective trial. *Ann Oncol* 2013; 24:2808-2813.
12. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature* 1990; 345:458-460.
13. Counter CM: The roles of telomeres and telomerase in cell life span. *Mutat Res* 1996; 366:45-63.
14. Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD, Cawthon RM. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci U S A* 2004; 101:17312-17315.
15. Janzen V, Forkert R, Fleming HE, Saito Y, Waring MT, Dombkowski DM, Cheng T, DePinho RA, Sharpless NE, Scadden DT. Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16ink4a. *Nature* 2006; 443:421-426.
16. Rudolph KL, Chang S, Lee HW, Blasco M, Gottlieb GJ, Greider C, DePinho RA. Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* 1999; 96:701-712.
17. Ju Z, Jiang H, Jaworski M, Rathinam C, Gompf A, Klein C, Trumpp A, Rudolph KL. Telomere dysfunction induces environmental alterations limiting hematopoietic stem cell function and engraftment. *Nat Med* 2007; 13:742-747.
18. Cretel E, Veen I, Pierres A, Binan Y, Robert P, Loundou AD, Baumstarck-Barrau K, Hubert AM, Bongrand P, Heim M.

[immune profile of elderly patients admitted in a geriatric short care unit]. *Rev Med Interne* 2011; 32:275-282.

**19.** Martin-Ruiz C, Jagger C, Kingston A, Collerton J, Catt M, Davies K, Dunn M, Hilkens C, Keavney B, Pearce SH, Elzen WP, Talbot D, Wiley L, Bond J, Mathers JC, Eccles MP, Robinson L, James O, Kirkwood TB, von Zglinicki T. Assessment of a large panel of candidate biomarkers of ageing in the newcastle 85+ study. *Mech Ageing Dev* 2011; 132:496-502.

**20.** Fitzpatrick AL, Kronmal RA, Kimura M, Gardner JP, Psaty BM, Jenny NS, Tracy RP, Hardikar S, Aviv A. Leukocyte telomere length and mortality in the cardiovascular health study. *J Gerontol A Biol Sci Med Sci* 2011; 66:421-429.

**21.** Njajou OT, Hsueh WC, Blackburn EH, Newman AB, Wu SH, Li R, Simonsick EM, Harris TM, Cummings SR, Cawthon RM. Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. *J Gerontol A Biol Sci Med Sci* 2009; 64:860-864.

**22.** Gertler R, Rosenberg R, Stricker D, Friederichs J, Hoos A, Werner M, Ulm K, Holzmann B, Nekarda H, Siewert JR. Telomere length and human telomerase reverse transcriptase expression as markers for progression and prognosis of colorectal carcinoma. *J Clin Oncol* 2004; 22:1807-1814.

**23.** Umehara N, Ozaki T, Sugihara S, Kunisada T, Morimoto Y, Kawai A, Nishida K, Yoshida A, Murakami T, Inoue H. Influence of telomerase activity on bone and soft tissue tumors. *J Cancer Res Clin Oncol* 2004; 130:411-416.

**24.** Bischoff C, Graakjaer J, Petersen HC, Jeune B, Bohr VA, Koelvræ S, Christensen K. Telomere length among the elderly and oldest-old. *Twin Res Hum Genet* 2005; 8:425-432.

**25.** Lee DC, Im JA, Kim JH, Lee HR, Shim JY. Effect of long-term hormone therapy on telomere length in postmenopausal women. *Yonsei Med J* 2005; 46:471-479.

**26.** Eisenberg DT, Salpea KD, Kuzawa CW, Hayes MG, Humphries SE. Substantial variation in qPCR measured mean blood telomere lengths in young men from eleven European countries. *Am J Hum Biol* 2011; 23:228-231.

**27.** Freyer G, Weber BE, Lebrun-Jezekova D, Tinquaut F, Tredan O, Sevin E, Stefani L, Pujade-Lauraine E, Falandry C. Development of a geriatric vulnerability score (gvs) in elderly advanced ovarian cancer (aoc) patients (pts) treated in first line: A prospective gineco trial. *J Clin Oncol* 2012; 30:Abstr 9079.

**28.** Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res* 2002; 30:e47.

**29.** Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 2009; 37:e21.

**30.** Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB, Aviv A. Measurement of telomere length by the southern blot analysis of terminal restriction fragment lengths. *Nat Protoc* 2010; 5:1596-1607.

**31.** Hemann MT, Strong MA, Hao LY, Greider CW. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* 2001; 107:67-77.

**32.** Jiang H, Schiffer E, Song Z, Wang J, Zurbig P, Thedieck K, Moes S, Bantel H, Saal N, Jantos J, Brecht M, Jenö P, Hall MN, Hager K, Manns MP, Hecker H, Ganser A, Dohner K, Bartke A, Meissner C, Mischak H, Ju Z, Rudolph KL. Proteins induced by telomere dysfunction and DNA damage represent biomarkers of human aging and disease. *Proc Natl Acad Sci U S A* 2008; 105:11299-11304.

**33.** Falandry C, Gilson E, Rudolph KL. Are aging biomarkers clinically relevant in oncogeriatrics? *Crit Rev Oncol Hematol* 2013; 85:257-265.

**34.** Augereau A, T'Kint de Roodenbeke C, Simonet T, Bauwens S, Horard B, Callanan M, Leroux D, Jallades L, Salles G, Gilson E, Poncet D. Telomeric damage in early stage of chronic lymphocytic leukemia correlates with shelterin dysregulation. *Blood* 2011; 118:1316-1322.

**35.** Falandry C, Weber B, Savoye AM, Tinquaut F, Tredan O, Sevin E, Stefani L, Savinelli F, Atlassi M, Salvat J, Pujade-Lauraine E, Freyer G. Development of a geriatric vulnerability score in elderly patients with advanced ovarian cancer treated with first-line carboplatin: A gineco prospective trial. *Ann Oncol* 2013; 2013:2022.

## Proteomic identification of prognostic tumour biomarkers, using chemotherapy-induced cancer-associated fibroblasts

Maria Peiris-Pagès<sup>1,2</sup>, Duncan L. Smith<sup>3</sup>, Balázs Győrffy<sup>4,5</sup>, Federica Sotgia<sup>1,2</sup>, and Michael P. Lisanti<sup>1,2</sup>

<sup>1</sup> The Breast Cancer Now Research Unit, Institute of Cancer Sciences, University of Manchester, UK;

<sup>2</sup> The Manchester Centre for Cellular Metabolism (MCCM), Institute of Cancer Sciences, University of Manchester, UK;

<sup>3</sup> The Cancer Research UK Manchester Institute, University of Manchester, UK;

<sup>4</sup> MTA TTK Lendület Cancer Biomarker Research Group, Budapest, Hungary;

<sup>5</sup> Semmelweis University 2nd Dept. of Pediatrics, Budapest, Hungary.

**Key words:** chemotherapy, metabolism, catabolism, cancer-associated fibroblasts, second primary tumours, tumour microenvironment, quantitative proteomics, markers, cancer survival

**Abbreviations:** AZA, azathioprine; TAX, taxol; CAF, cancer-associated fibroblast; MCT4, monocarboxylate transporter 4; IL6, interleukin 6; ROS, reactive oxygen species;  $\alpha$ SMA,  $\alpha$ -smooth muscle actin

**Received:** 08/13/15; **Accepted:** 09/12/15; **Published:** 10/23/15

**Correspondence to:** Michael Lisanti, PhD; **E-mail:** [Michael.Lisanti@ics.manchester.ac.uk](mailto:Michael.Lisanti@ics.manchester.ac.uk)

**Copyright:** Peiris-Pagès et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** Cancer cells grow in highly complex stromal microenvironments, which through metabolic remodelling, catabolism, autophagy and inflammation nurture them and are able to facilitate metastasis and resistance to therapy. However, these changes in the metabolic profile of stromal cancer-associated fibroblasts and their impact on cancer initiation, progression and metastasis are not well-known. This is the first study to provide a comprehensive proteomic portrait of the azathioprine and taxol-induced catabolic state on human stromal fibroblasts, which comprises changes in the expression of metabolic enzymes, myofibroblastic differentiation markers, antioxidants, proteins involved in autophagy, senescence, vesicle trafficking and protein degradation, and inducers of inflammation. Interestingly, many of these features are major contributors to the aging process. A catabolic stroma signature, generated with proteins found differentially up-regulated in taxol-treated fibroblasts, strikingly correlates with recurrence, metastasis and poor patient survival in several solid malignancies. We therefore suggest the inhibition of the catabolic state in healthy cells as a novel approach to improve current chemotherapy efficacies and possibly avoid future carcinogenic processes.

### INTRODUCTION

Unlike normal healthy fibroblasts, aged or senescent fibroblasts are pro-tumorigenic [1]. Cellular damage, which is widely considered to be the general cause of aging, occasionally may provide cells with abnormal advantages that can eventually give rise to cancer. Thus, cancer and aging are two different faces of the same underlying process: the accumulation of cellular damage in cells and tissues over the years, which eventually become senescent. Indeed, accumulation of

senescent cells has been detected after examination of aged tissues, and it contributes to tissue degeneration during aging [1]. Whether senescence of the stroma is sufficient to initiate tumorigenesis still remains unclear. However, senescent cells can have profound effects on the surrounding microenvironment, through the expression and secretion of a range of pro-inflammatory factors, which is known as the senescence-associated secretory phenotype (SASP). The onset of SASP may help explain the increased tumor incidence observed in aged individuals [2].

The tumour stroma comprises the majority of the neoplastic mass and is mainly composed of fibroblasts [3]. Nevertheless, our comprehension of the tumour microenvironment is rather limited in comparison with that of cancer cells. The emergence of a reactive microenvironment via metabolic stress and inflammation fuels cancer cells, enables tumour growth and invasion, and leads to treatment failure [3-10]. However, the mechanisms by which the metabolic remodelling of cancer-associated stromal fibroblasts (CAFs) regulates the evolution of malignancy or even may control the susceptibility of incompletely transformed cells to become fully malignant are not fully understood.

We have previously established that exposure to anti-cancer agents independently drives metabolic stress and catabolism, autophagy, senescence, myofibroblastic differentiation and production of the pro-inflammatory cytokine Interleukin 6 (IL6) in human stromal fibroblasts *in vitro* (Figure 1) [11]. Thus, according to our model, chemotherapy promotes the same effects in stromal fibroblasts as their interaction with cancer cells, the so-called catabolic tumour stroma phenotype, which creates an energy-rich, pro-inflammatory niche ideal for cancer development and possibly initiation.

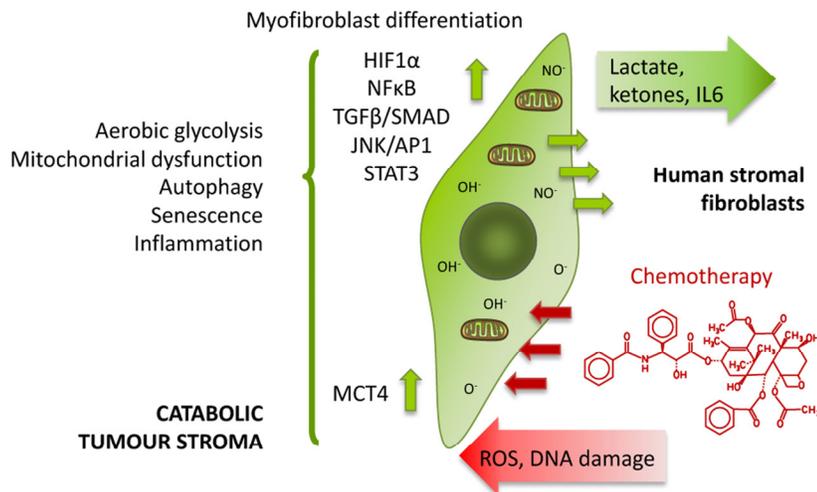
Despite the significant number of markers and secreted proteins already related to CAFs, there is little evidence of the contribution of chemotherapy-induced CAF transformation to metastasis or the growth of a second primary tumour after therapy. Indeed, only one report links secretion of factors associated with inflammation and cancer progression in therapy-damaged senescent fibroblasts with *de novo* tumorigenesis [12]. Thus, novel biomarkers are required to improve the prediction of recurrence, metastasis and, in particular, the prediction of therapy-related carcinogenesis. So far, there is one study investigating transcriptomic changes in stromal fibroblasts upon chemotherapeutic treatment, but none investigating phenotypic changes by proteomics [11]. Azathioprine and taxol (paclitaxel) are drugs widely used in chemotherapy for a variety of cancers and in particular taxol is used as the first-line chemotherapeutic agent for ovarian cancer [13-15]. In this study, we describe a strategy based on a label-free quantitative proteomic profiling of fibroblasts obtained after treatment with azathioprine or taxol, which allows us to measure numerous markers of the CAF phenotype. Likewise, the data presented here attempt to identify novel biomarkers of the catabolic remodelling in human stromal fibroblasts that are associated with chemoresistance, metastasis and second primary tumours by reporting their impact on cancer survival. The expression of several over-expressed proteins found

in taxol-treated fibroblasts that are involved in metabolism, antioxidant response, autophagy, vesicle trafficking, protein degradation and myofibroblastic transformation correlate with poor prognosis in chemotherapy-treated breast, lung, gastric and ovarian cancer patients. We conclude that a strategy that targets constituents of the tumour microenvironment in combination with conventional chemotherapy may help improving treatment efficacy and avoiding the growth of future malignancies.

## RESULTS

To identify differentially regulated proteins upon chemotherapeutic treatment, hTERT-BJ1 fibroblasts were exposed for 48 h to either vehicle or sub-lethal concentrations of azathioprine (100  $\mu$ M) or taxol (100 nM) (Figure S1), and cell lysates were subject to label-free quantitative proteomics. Following protein digestion with trypsin, peptide fractions were processed on an LTQ-Orbitrap XL mass spectrometer. The experimental workflow used for the present study is depicted in Figure 2. Those peptides identified were further analyzed to find proteomic changes between chemotherapy-treated and vehicle-treated fibroblasts. To define differential regulation, those identified proteins that showed a fold change difference of 1.15 or higher, and p values of <0.05 (ANOVA) compared to vehicle were considered. In the azathioprine-treated fibroblasts, 1640 proteins were identified as differentially expressed, from which 779 were up-regulated and 861, down-regulated. In the taxol treatments, 2967 proteins were found as differentially expressed compared to vehicle, from which 1624 were up-regulated and 1343, down-regulated (Figure 3A).

To obtain additional functional insights into pathways that are differentially regulated in stromal fibroblasts upon treatment, bioinformatics analyses of our proteomics datasets were conducted. All differentially expressed proteins were analysed using Ingenuity Pathway Analysis (IPA) to seek altered canonical pathways and toxicity functions. IPA was able to analyse 633 proteins out of 1640 in the azathioprine-treated fibroblasts, and 841 out of 2967 proteins in the taxol-treated fibroblasts. We further examined the subcellular distribution and the nature of these differentially regulated proteins in azathioprine and taxol-treated fibroblasts. Over 80% of all proteins analyzed were intracellular, in particular cytoplasmic proteins (Figure 3B). Likewise, the largest portion of classified proteins, accounting for one fourth of all analysed proteins, were enzymes, over 8% of all proteins were transporter proteins, and over 7% were transcription regulators (Figure 3C). Finally, a comparison



**Figure 1. Chemotherapy induces the catabolic tumour stroma phenotype.** ROS production and DNA damage induced by chemotherapy generate oxidative stress to stromal cells, which in turn brings about several changes in them such as differentiation into CAFs, activation of HIF1 $\alpha$ , NF $\kappa$ B, TGF $\beta$ , STAT3 or JNK/AP1 signalling pathways, switch to aerobic glycolysis and loss of functional mitochondria, acquisition of an autophagic and senescent phenotype, and release of pro-inflammatory cytokines. Thus, these stromal fibroblasts acquire the catabolic tumour stroma phenotype.

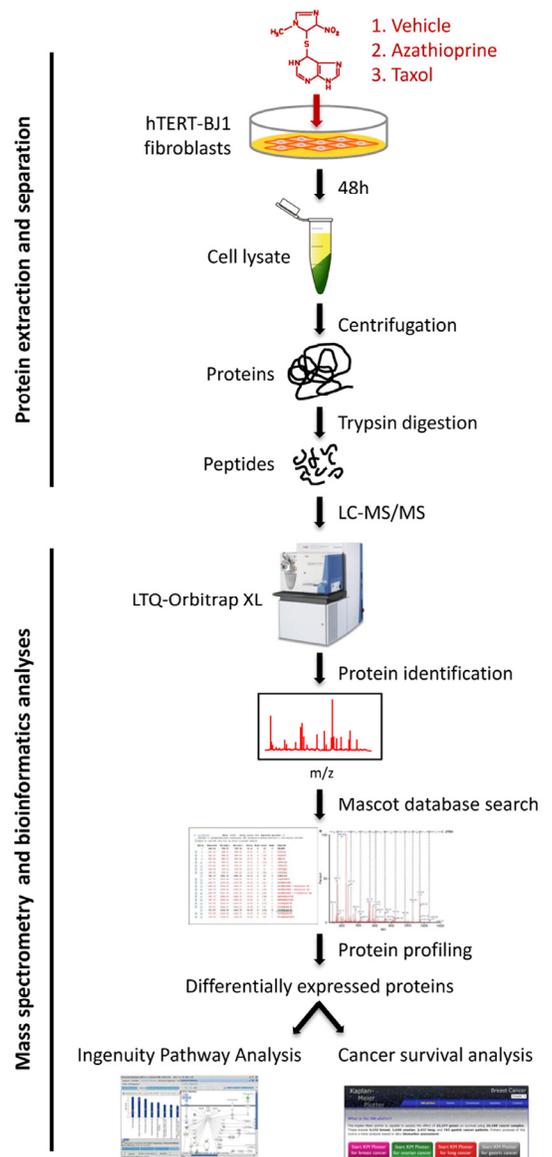
analysis revealed that 352 proteins were differentially regulated in both treatment conditions compared to vehicle (Figure 3D).

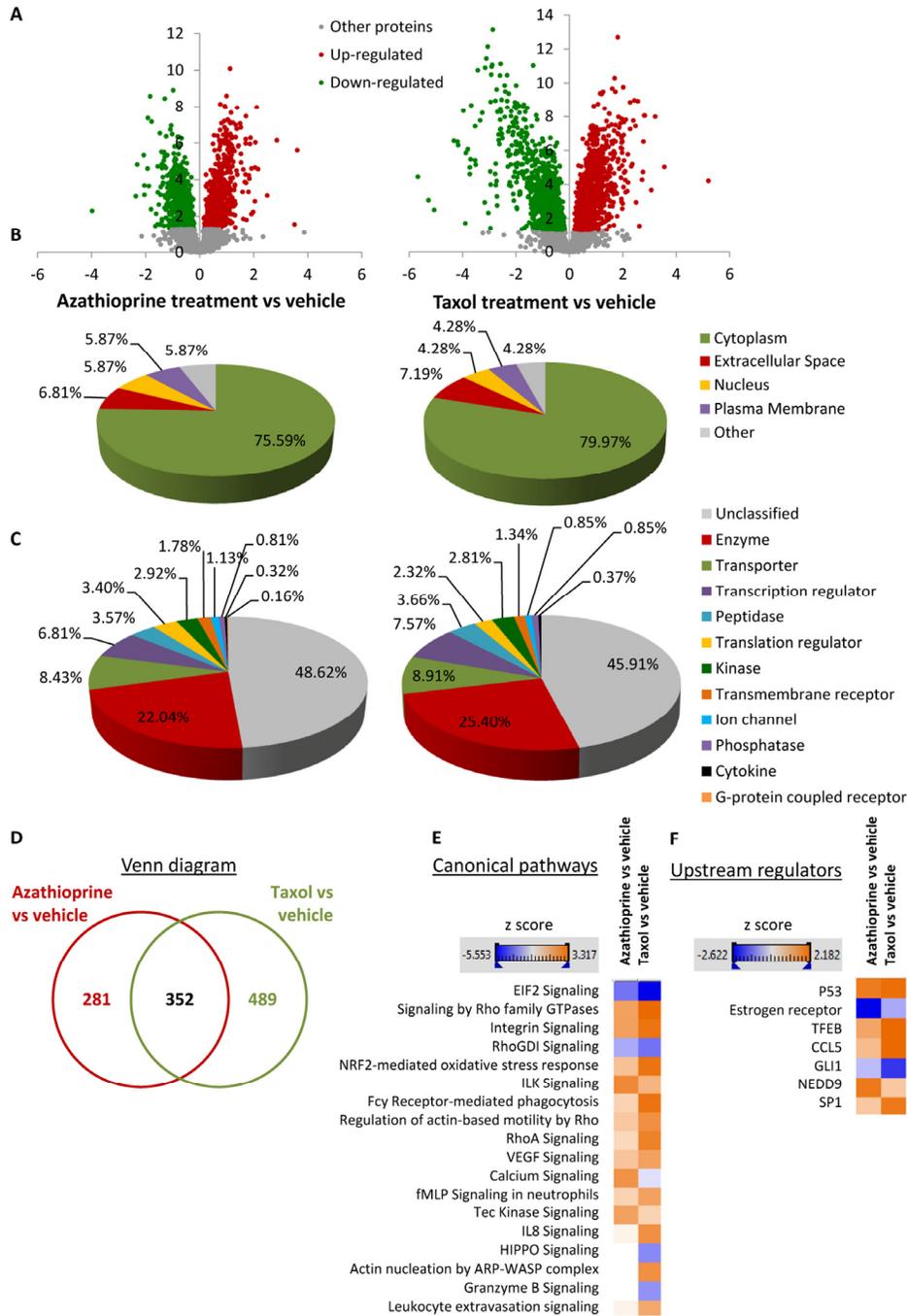
### Cellular pathways affected by chemotherapy in stromal fibroblasts

#### 1. Metabolism

One of the major contributors to the aging process is mitochondrial dysfunction, which involves a decrease in the oxidative phosphorylation efficacy and an increase in the electron leakage resulting in reduced ATP generation [17]. Similarly, one of the hallmarks of the catabolic tumour stroma is the induction of metabolic stress that favours glycolysis to the detriment of mitochondrial metabolism. We have showed in a previous study that several chemotherapeutic agents, including azathioprine and taxol were able to stimulate stromal fibroblasts to consume more glucose and produce more lactate, which was released via enhanced MCT4 expression, hence increasing extracellular acidification (Figure 1). The cellular ATP content was also minor upon treatment suggesting a decrease in mitochondrial respiration [11].

**Figure 2. Workflow for the comparative proteome analysis of hTERT-BJ1 fibroblasts treated with azathioprine, taxol or vehicle.** Protein lysates were obtained from hTERT-BJ1 fibroblasts after 48 h treatment with azathioprine, taxol or vehicle. Peptides obtained after trypsin digestion were analysed via LC-MS/MS on an LTQ-Orbitrap XL mass spectrometer. Label-free quantitative proteomics was used to detect changes in protein abundances across vehicle-treated and drug-treated fibroblast extracts. The proteomics data sets were further analysed using Ingenuity Pathway Analysis and a cancer survival analysis tool (kmplot.com).





**Figure 3. Overview of proteins and pathways identified as differentially regulated in the lysates of azathioprine-treated and taxol-treated fibroblasts relative to vehicle by Ingenuity Pathway Analysis.** (A) Volcano plot representation of protein abundance changes in hTERT-BJ1 fibroblasts upon treatment with azathioprine and taxol compared to vehicle treatment. A total of 1640 differentially regulated proteins with fold changes  $\geq 1.15$  and p values  $< 0.05$  were identified in azathioprine-treated fibroblasts, and 2967 differentially regulated proteins in taxol-treated fibroblasts. X axis represents log<sub>2</sub>(fold change). Y axis represents -log(p value). Non-significantly regulated proteins are shown in grey, in green, significantly down-regulated proteins and in red, significantly up-regulated proteins. (B) Subcellular localization of differentially regulated proteins identified in azathioprine and taxol treatments compared to vehicle treatment. (C) Classification of differentially regulated proteins identified in azathioprine and taxol treatments by type. Cellular enzymes account for 22.04% and 25.40% of total differentially regulated proteins identified in azathioprine treatment and taxol treatment, respectively. (D) Overlap of differentially regulated proteins identified in azathioprine and taxol treatments compared to vehicle treatment. Of all proteins identified by quantitative proteomics, 352 were proteins the expression of which was found altered in both treatments compared to vehicle. (E) Canonical pathways and (F) upstream regulators identified or predicted as altered in both treatment conditions compared to vehicle. A positive z score is indicated in orange and points towards an activation of the pathway, and a negative z score, in blue, indicates an inhibition of the pathway.

To validate that chemotherapy-induced glycolytic phenotype and to identify other cellular metabolic pathways potentially altered by chemotherapy, we searched the proteomic data for changes in the expres-

sion of metabolic enzymes. Of all differentially regulated proteins, 22.04% and 25.40% were actually enzymes in azathioprine-treated and taxol-treated fibroblasts, respectively (Figure 3C).

**Table 1. Changes in the expression of enzymes involved in glucose metabolism after treatment with azathioprine and taxol for 48 h as measured by quantitative proteomics.**

GLYCOLYSIS		Azathioprine	Taxol
Solute carrier family 2 (facilitated glucose transporter), member 1 (GLUT1)	SLC2A1	↑ 1.18	
Hexokinase	HK1		↑ 1.43
	HK2	↑ 2.46	
Glucose-6-phosphate isomerase	GPI		↓ 1.27
6-phosphofructokinase type C	PFKP		↑ 1.56
Fructose-bisphosphate aldolase A	ALDOA	↑ 2.10	↓ 11.58
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	↑ 1.90	↑ 1.89
Triosephosphate isomerase 1	TPI1		↑ 1.33
Phosphoglycerate kinase 1	PGK1		↓ 1.58
Phosphoglycerate mutase	PGAM1		↓ 1.38
	PGAM4	↓ 1.45	
Enolase 1	ENO1	↓ 1.37	
Pyruvate kinase	PKM		↑ 1.65
POST-GLYCOLYSIS PROCESSES		Azathioprine	Taxol
L-lactate dehydrogenase A chain	LDHA	↓ 1.47	↓ 1.46
Monocarboxylate transporter 4 (MCT4)	SLC16A3		↑ 2.21
Pyruvate dehydrogenase	PDHB	↓ 1.20	↓ 1.40
GLUCONEOGENESIS		Azathioprine	Taxol
Glucose-6-phosphate isomerase	GPI		↓ 1.27
Aldolase A, fructose-bisphosphate	ALDOA	↑ 2.10	↓ 11.57
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	↑ 1.90	↑ 1.89
Phosphoglycerate kinase 1	PGK1		↓ 1.58
Phosphoglycerate mutase	PGAM1		↓ 1.38
	PGAM4	↓ 1.45	
Enolase 1	ENO1	↓ 1.37	
Malate dehydrogenase	MDH1		↓ 1.80
	MDH2		↓ 1.23
Malic enzyme 2, NAD(+)-dependent, mitochondrial	ME2		↓ 1.62
PENTOSE PHOSPHATE PATHWAY		Azathioprine	Taxol
<b>Glucose-6-phosphate dehydrogenase</b>	<b>G6PD</b>	↓ 2.03	↑ 2.69
	H6PD	↑ 1.46	
6-phosphogluconolactonase	PGLS		↑ 1.54
Phosphogluconate dehydrogenase	PGD		↓ 1.57
Transketolase	TKT	↓ 1.88	
Transaldolase	TALDO	↑ 2.13	
HEXOSAMINE BIOSYNTHESIS PATHWAY		Azathioprine	Taxol
Glutamine-fructose-6-phosphate transaminase 1	GFPT1		↑ 1.36
<b>Glucosamine-6-phosphate deaminase 1</b>	<b>GNPDA1</b>		↑ 2.46

Chemotherapy increased the expression of several enzymes involved in glycolysis, pentose phosphate pathway, and hexosamine biosynthesis. Enzymes that show amplified expression after treatment are shown in red, and enzymes that are decreased are shown in green. Number indicates the fold increase or fold decrease in protein expression in chemotherapy-treated versus vehicle-treated hTERT-BJ1 fibroblasts.

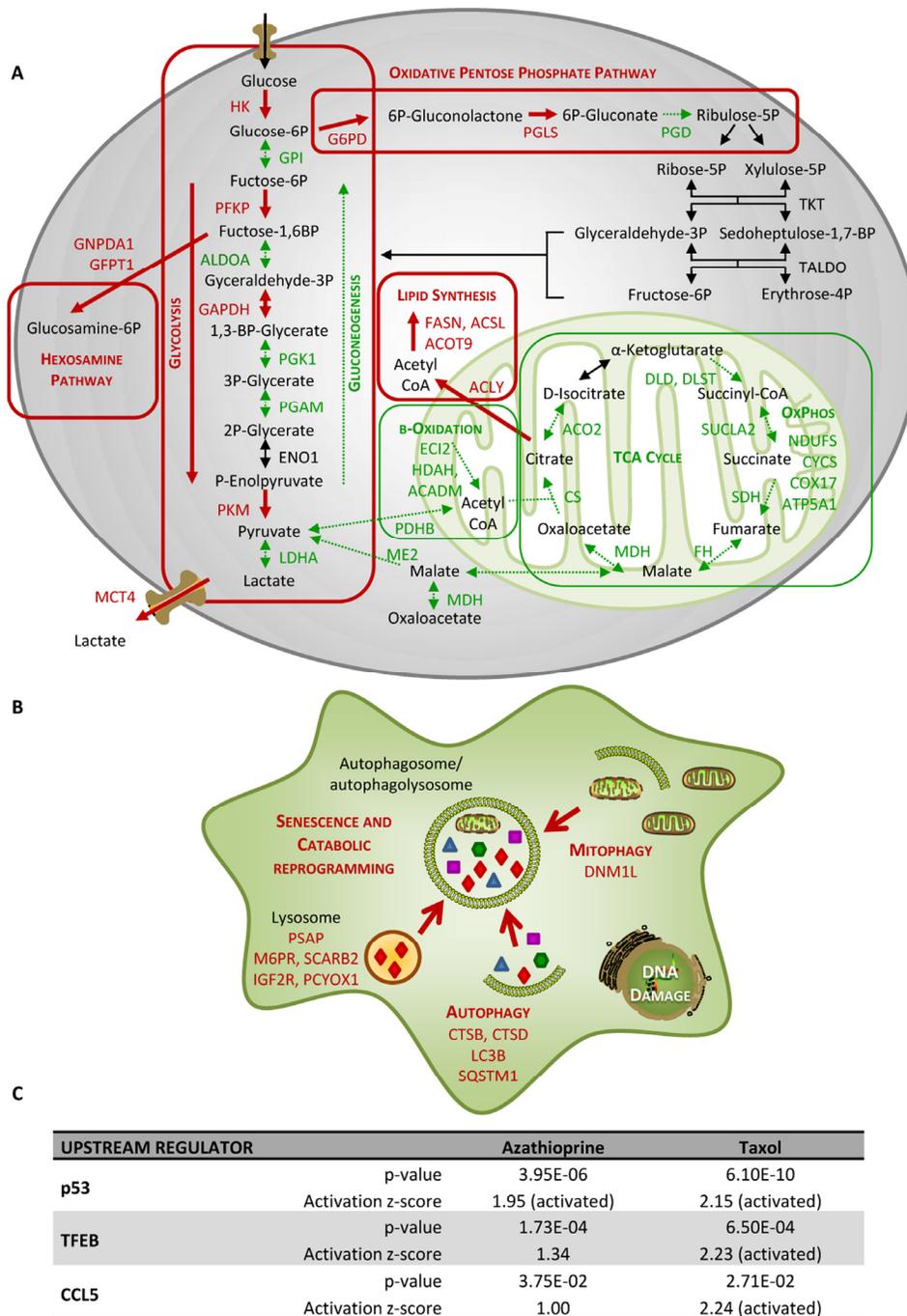
**Table 2. Changes in the expression of enzymes involved in mitochondrial glucose metabolism after treatment with azathioprine and taxol for 48 h as measured by quantitative proteomics.**

TCA CYCLE		Azathioprine	Taxol
Citrate synthase, mitochondrial	CS		↓ 1.52
Aconitate hydratase	ACO2	↑ 1.41	↓ 1.23
Dihydrolipoyl dehydrogenase	DLD		↓ 1.48
Alpha-ketoglutarate dehydrogenase complex dihydrolipoyl succinyltransferase	DLST	↓ 1.23	↓ 1.70
2-oxoglutarate dehydrogenase	OGDH	↓ 1.37	
Beta-succinyl CoA synthetase	SUCLA2		↓ 1.25
Succinate dehydrogenase complex	SDHA		↓ 2.43
	SDHB	↓ 1.51	
Fumarate hydratase	FH		↓ 1.27
Malate dehydrogenase	MDH1		↓ 1.80
	MDH2		↓ 1.22
OXIDATIVE PHOSPHORYLATION		Azathioprine	Taxol
NADH dehydrogenase (complex I)	NDUFV1		↑ 1.54
	NDUFV2		↓ 1.45
	NDUFS1		↓ 1.73
	NDUFS7		↓ 1.31
	NDUFB10		↓ 1.29
Succinate dehydrogenase complex (complex II)	SDHA		↓ 2.43
	SDHB	↓ 1.51	
Coenzyme Q – cytochrome c reductase (complex III)	CYCS		↓ 1.61
	UQCRC1		↓ 1.42
	UQCRB		↑ 1.62
Cytochrome c oxidase (complex IV)	COX17		↓ 2.00
	COX6C		↓ 1.62
	COX6A1	↑ 1.47	
	COX5A		↓ 1.39
	COX5B	↑ 1.37	
ATP synthase (complex V)	ATP5A1	↓ 1.51	↓ 1.16
	ATP5F1	↓ 1.70	↓ 1.51
	ATP5H	↑ 1.56	
	ATP5J	↑ 1.36	
	ATP5B		↑ 2.36

Taxol treatment remarkably decreased the expression of enzymes of the TCA cycle and oxidative phosphorylation. Enzymes that show increased expression after treatment are shown in red, and enzymes that are decreased are shown in green. Number indicates the fold increase or fold decrease in protein expression in chemotherapy-treated versus vehicle-treated hTERT-BJ1 fibroblasts.

The expression of most glycolytic enzymes was significantly altered (Table 1). Interestingly, most of the glycolytic enzymes that were found to be up-regulated are enzymes that perform an irreversible reaction in the glycolytic pathway and all down-regulated enzymes are reversible enzymes, able to perform the opposite re-

action, hence also involved in gluconeogenesis. Curiously, LDHA, the enzyme involved in transforming pyruvate into lactate was found to be down-regulated in both treatments suggesting that pyruvate was further processed into the TCA cycle instead of being transformed into lactate.



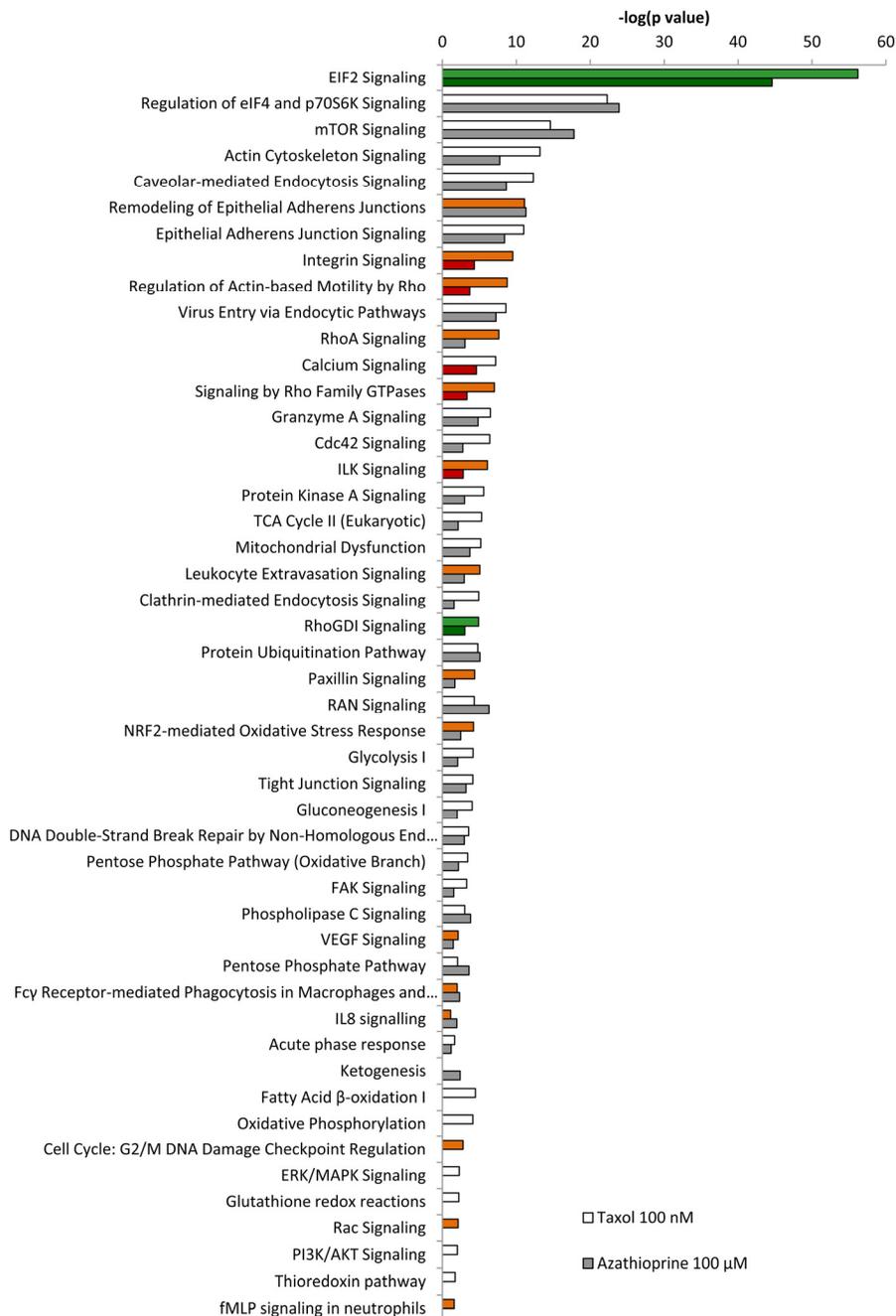
**Figure 4. Taxol treatment alters several cellular metabolic pathways and induces autophagy and senescence in hTERT-BJ1 fibroblasts.** (A) Summary of changes observed in the expression of numerous metabolic enzymes after treatment with taxol for 48 h, as measured by quantitative proteomics. Taxol treatment increased the expression of enzymes involved in glycolysis, pentose phosphate, and hexosamine biosynthesis and lipid synthesis pathways in detriment of those involved in gluconeogenesis and mitochondrial metabolism (TCA cycle, oxidative phosphorylation, and mitochondrial  $\beta$ -oxidation). Enzymes that show amplified expression after taxol treatment are shown in red, and enzymes that are decreased are shown in green. Cellular metabolic pathways that may be up-regulated are boxed in red, and those that may be down-regulated are boxed in green. (B) Summary of changes observed in the expression of numerous autophagy and senescence-related proteins after treatment with taxol for 48 h, as measured by quantitative proteomics. Taxol amplified the expression of proteins involved in senescence, autophagy, mitophagy and vesicle formation and trafficking. Proteins that show increased expression after treatment are shown in red. (C) Ingenuity Pathway Analysis of azathioprine and taxol-treated hTERT-BJ1 fibroblasts predicted p53, TFEB and CCL5 to be activated in these cells.

**Table 3.** Changes in the expression of Nrf2-target proteins and other proteins involved in the antioxidant response after treatment with azathioprine and taxol for 48 h as measured by quantitative proteomics and indicated by IPA analysis.

NRF2-MEDIATED ANTIOXIDANT RESPONSE		Azathioprine	Taxol
ATP-binding cassette, sub-family C (CFTR/MRP), member 1	ABCC1	↓ 2.34	
Actin, alpha 2, smooth muscle, aorta	ACTA2	↑ 1.34	↑ 1.78
Actin, beta	ACTB		↑ 1.30
Actin, alpha, cardiac muscle 1	ACTC1	↑ 1.75	↑ 1.20
Carbonyl reductase 1	CBR1		↑ 1.29
Chaperonin containing TCPI1, subunit 7 (eta)	CCT7	↑ 2.89	
DnaJ (Hsp40) homolog	DNAJA1	↑ 1.61	↓ 9.61
	DNAJB11	↓ 1.34	
	DNAJC8	↓ 1.79	↓ 1.58
	DNAJC13		↓ 4.30
Epoxide hydrolase 1, microsomal (xenobiotic)	EPHX1		↑ 2.06
Ferritin, light polypeptide	FTL	↑ 1.61	↑ 6.97
Glutathione S-transferase	GSTK1		↑ 1.34
	GSTO1		↑ 1.57
	GSTP1		↓ 1.53
Heme oxygenase (decycling) 1	HMOX1	↑ 4.48	↑ 2.83
Mitogen-activated protein kinase 3	MAPK3	↑ 1.54	
Peptidylprolyl isomerase B (cyclophilin B)	PPIB	↑ 1.60	↑ 1.36
Peroxiredoxin 1	PRDX1		↑ 1.68
Protein kinase C, alpha	PRKCA	↑ 1.96	
Related RAS viral (r-ras) oncogene homolog	RRAS		↑ 2.44
Superoxide dismutase 1, soluble	SOD1		↑ 1.48
Superoxide dismutase 2, mitochondrial	SOD2	↑ 1.69	↓ 1.87
Sequestosome 1	SQSTM1		↑ 3.50
Thioredoxin	TXN		↑ 2.75
Thioredoxin reductase 1	TXNRD1		↑ 2.66
Valosin-containing protein	VCP	↓ 1.29	
ANTIOXIDANTS (OTHER)		Azathioprine	Taxol
Aminopeptidase N	ANPEP	↑ 1.78	↑ 1.56
Glutaredoxin-1	GLRX		↑ 1.25
Glutathione peroxidase	GPX8		↓ 1.85

Peroxidasin	PXDN		↑ 3.57
	PRDX4	↑ 1.50	
Peroxiredoxin	PRDX5		↑ 1.33
	PRDX6		↓ 1.23

Chemotherapy mostly increased the expression of proteins involved in antioxidant response, mainly in taxol-treated fibroblasts. Proteins that show amplified expression after treatment are shown in red, and proteins that are decreased are shown in green. Number indicates the fold increase or fold decrease in protein expression in chemotherapy-treated versus vehicle-treated hTERT-BJ1 fibroblasts.

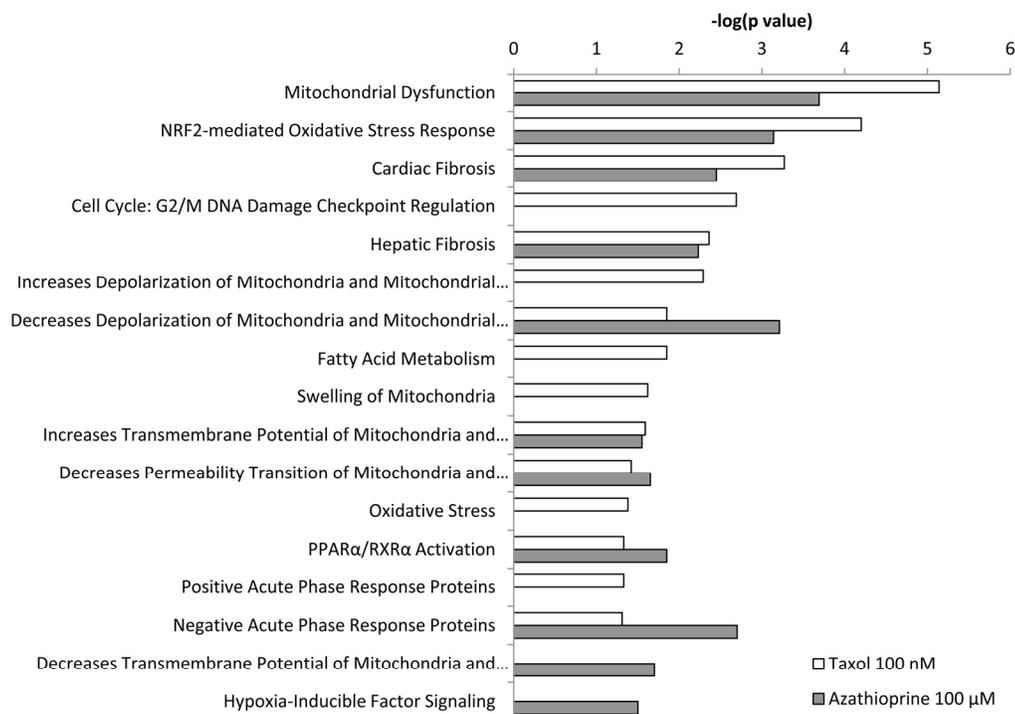


**Figure 5. Pathway analysis of differentially expressed proteins in hTERT-BJ1 fibroblasts treated with azathioprine or taxol compared to vehicle-treated cells.** Ingenuity Pathway Analysis showed canonical pathways significantly altered by the proteins differentially expressed in hTERT-BJ1 fibroblasts treated with azathioprine or taxol ( $P < 0.05$ ). The p value for each pathway is indicated by the bar and is expressed as -1 times the log of the p value. Green coloured bars show a predicted inhibition of the pathway (z score < -1.9) and red coloured bars indicate a predicted activation of the pathway (z score > 1.9).

Nevertheless, LDH is an enzyme that exhibits feedback inhibition, by which high lactate concentrations can suppress it. In fact, MCT4, the monocarboxylate transporter responsible for the secretion of lactate, turned out to be up-regulated in taxol-treated hTERT-BJ1 cells, and regarding the further processing of pyruvate into acetyl-CoA and the citric acid (TCA) cycle, the expression of practically all enzymes were found to be down-regulated in taxol-treated fibroblasts compared to vehicle indicating a dramatic down-regulation of mitochondrial metabolism. A similar down-regulation trend was observed in numerous proteins involved in all oxidative phosphorylation complexes (Table 2) and mitochondrial function proteins (Table S1). Finally, several enzymes involved in mitochondrial fatty acid  $\beta$ -oxidation were also identified as down-regulated, and a few enzymes involved in fatty acid biosynthesis, up-regulated (Table S2). Other metabolic changes included enzymes of the oxidative pentose phosphate pathway, responsible for the generation of antioxidant power (NADPH), the hexosamine synthesis pathway, account-

able for the production of amino sugars used for the synthesis of glycoproteins, glycolipids and proteoglycans, and a few enzymes involved in the generation of ketone bodies (Table 1 and 5). Figure 4A summarises all changes observed in the expression of metabolic enzymes in taxol-treated fibroblasts and their contribution to different cellular metabolic pathways.

IPA revealed glycolysis, gluconeogenesis and pentose phosphate pathway as altered canonical pathways in both azathioprine and taxol treatments. Similarly, IPA revealed mitochondrial dysfunction, TCA cycle, and PPAR $\alpha$ /RXR $\alpha$  activation, responsible for ketone body production and fatty acid metabolism, as two of the top canonical pathways affected by both treatments, and oxidative phosphorylation and fatty acid  $\beta$ -oxidation, as altered pathways also in taxol-treated cells (Figure 5 and Table S5). The toxicity impact of both drugs extensively involved mitochondrial dysfunction and damage, and also fatty acid metabolism and PPAR $\alpha$ /RXR $\alpha$  activation (Figure 6 and Table S6).



**Figure 6. Toxicity effects of differentially expressed proteins in hTERT-BJ1 fibroblasts treated with azathioprine or taxol compared to vehicle-treated cells.** Ingenuity Pathway Analysis showed toxicity functions significantly enriched by the proteins differentially expressed in hTERT-BJ1 fibroblasts treated with azathioprine or taxol ( $P < 0.05$ ). The p value for each pathway is indicated by the bar and is expressed as -1 times the log of the p value.

**Table 4.** Changes in the expression of proteins involved in myofibroblastic transformation and muscle-related proteins after treatment with azathioprine and taxol for 48 h as measured by quantitative proteomics.

CANCER-ASSOCIATED FIBROBLAST AND MUSCLE-RELATED PROTEINS		Azathioprine	Taxol
Actin, alpha, cardiac muscle	ACTC1	↑ 1.75	↑ 1.20
Actin, alpha 2, aortic smooth muscle ( $\alpha$ SMA)	ACTA2	↑ 1.34	↑ 1.78
Caldesmon	CALD1		↑ 2.49
Calponin	CNN1		↑ 2.76
	CNN2	↑ 1.81	↑ 2.76
	CNN3	↓ 1.28	↑ 2.46
Desmin	DES	↑ 1.21	↓ 1.79
Dysferlin	DYSF		↑ 1.24
Fibroblast activation protein	FAP	↑ 1.74	↑ 1.56
Fibronectin 1	FN1	↑ 2.02	↑ 2.03
Filamin	FLNA	↓ 1.92	↑ Infinity
	FLNB	↓ 1.23	
	FLNC		↑ 1.56
Moesin	MSN	↑ 1.37	↑ 1.60
Myoferlin	MYOF	↑ 4.15	↑ 1.95
	MYH4		↑ 1.62
Myosin	MYH9	↑ 1.64	
	MYH10	↑ 1.37	
	MYH11	↑ 1.47	↑ 1.90
	MYH14	↓ 3.06	
	MYO1B		↓ 1.72
	MYO1C	↑ 2.21	↑ 1.59
	MYO10	↑ 1.37	
	MYO18B	↓ 1.50	↑ 2.62
Myosin, light polypeptide kinase	MYLK		↑ 1.88
Myosin phosphatase Rho interacting protein	MPRIP		↑ 1.45
Myosin regulatory light polypeptides	MYL6	↑ 1.33	↑ 1.74
	MYL9		↑ 1.77
	MYL12A		↑ 2.21
Palladin	PALLD		↑ 3.88
Platelet-derived growth factor receptor, beta	PDGFRB		↑ 1.59
Prolyl 4-hydroxylase	P4HA1		↑ 3.14
	P4HA2	↓ 1.32	↓ 2.50
	P4HB	↓ 1.76	
Talin	TLN1	↑ 1.46	↑ 2.95
	TLN2	↓ 1.80	

Transgelin 2	TAGLN2		↑ 1.93
	TPM1	↓ 2.63	↓ 4.23
	TPM2	↑ 1.24	
Tropomyosin	TPM3	↑ 1.43	
	TPM4		↑ 2.12
Vimentin	VIM	↑ 1.72	↑ 1.76

Chemotherapy enormously increased the expression of proteins involved in CAF transformation. Proteins that show amplified expression after treatment are shown in red, and proteins that are decreased are shown in green. Number indicates the fold increase or fold decrease in protein expression in chemotherapy-treated versus vehicle-treated hTERT-BJ1 fibroblasts.

To recapitulate, the previously identified chemotherapy-induced metabolic stress in stromal fibroblasts is clearly detected by quantitative proteomics and it represents not only an increase in glycolysis and a reduction in mitochondrial function, as observed in our previous study, but also affects other metabolic pathways, including hexosamine synthesis and pentose phosphate pathways, fatty acid metabolism and ketogenesis.

## 2. Antioxidant response and stress-related pathways

Another feature of the catabolic tumour stroma is the induction of oxidative stress, which can be induced by chemotherapy in cancer cells and also healthy tissues (Figure 1) [18, 19]. Particularly, we previously showed increased ROS production and antioxidant response in hTERT-BJ1 fibroblasts after treatment with taxol [11]. We have seen in this study that several enzymes of the oxidative branch of the pentose phosphate pathway, which is responsible for the generation of antioxidant power (NADPH), were up-regulated in chemotherapy-treated stromal fibroblasts. Thus, we sought for other antioxidant response proteins in the proteomics datasets. The expression of numerous proteins involved in Nrf2-mediated antioxidant response and other antioxidant proteins was significantly altered, in most cases up-regulated (Table 3), suggesting an activation of the pathway in both azathioprine and taxol-treated hTERT-BJ1 fibroblasts, although it was found to be notably higher in taxol-treated cells.

A higher presence of antioxidant response proteins was further confirmed using IPA, which revealed Nrf2-mediated antioxidant response to be one of the top canonical pathways affected by azathioprine and taxol treatments, although z score values demonstrated a significant activation of this pathway exclusively in

taxol-treated fibroblasts. Similarly, taxol-treated cells showed altered glutathione and thioredoxin antioxidant pathways (Figure 3E and 5, and Table S5). The toxicity of these chemotherapeutic drugs also involved oxidative stress. HIF signalling, which is activated in response to stress and a central player in the regulation of cellular metabolism, was also identified as a toxic effect of azathioprine treatment (Figure 6 and Table S6), and had been previously identified as activated in hTERT-BJ1 fibroblasts treated with azathioprine and taxol [11]. Thus, proteomics analysis clearly detects an activation of the antioxidant response after treatment with chemotherapy.

## 3. Myofibroblastic differentiation

Cancer-associated fibroblasts are commonly identified by their expression of alpha smooth muscle actin ( $\alpha$ SMA) [20] (Figure 1). Indeed, we previously showed up-regulation of  $\alpha$ SMA by immunoblotting in taxol-treated hTERT-BJ1 [11]. Quantitative proteomics profiling of azathioprine and taxol-treated hTERT-BJ1 fibroblasts also revealed a significantly higher presence of  $\alpha$ SMA and many other myofibroblastic markers, such as fibroblast activation protein (FAP) or vimentin, as well as muscle-related proteins compared to the vehicle-treated control (Table 4).

Myofibroblasts are mostly responsible for the presence of fibrosis [21], and the toxicity functions of these chemotherapeutic drugs involved tissue fibrosis, as analysed by IPA (Figure 6 and Table S6). Therefore, chemotherapy can independently induce the differentiation of hTERT-BJ1 fibroblasts into cancer-associated fibroblasts, which can be detected by a numerous increase of myofibroblastic markers in quantitative proteomics analysis.

**Table 5.** Changes in the expression of autophagy, mitophagy and senescence markers and ketogenesis enzymes after treatment with azathioprine and taxol for 48 h as measured by quantitative proteomics.

AUTOPHAGY MARKERS		Azathioprine	Taxol
Cathepsin	CATB	↑ 1.56	↑ 1.55
	CATD		↑ 1.62
Lysosomal-associated membrane protein 1	LAMP1	↑ 1.55	
Microtubule-associated protein 1 light chain 3 beta	MAP1LC3B		↑ 1.58
Sequestrosome 1 (p62)	SQSTM1		↑ 3.50
MITOPHAGY MARKERS		Azathioprine	Taxol
Dynamin-1-like protein (fission)	DNM1L		↑ 1.71
Dynamin-like 120 kDa protein, mitochondrial (fusion)	OPA1	↓ 1.44	
LYSOSOMAL PROTEINS (OTHER)		Azathioprine	Taxol
Cation-dependent mannose-6-phosphate receptor	M6PR		↑ 2.38
<b>Cation-independent mannose-6-phosphate receptor</b>	<b>IGF2R</b>	↑ 3.58	↑ 1.88
Galactosidase, beta 1	GLB1	↑ 1.53	
Late endosomal/lysosomal adaptor, MAPK and mTOR activator 1	LAMTOR1	↑ 1.85	
Lysosome membrane protein 2	SCARB2	↑ 1.63	↑ 1.49
N-acetylglucosamine-6-sulfatase	GNS		↑ 1.43
Preylcysteine oxidase 1	PCYOX1		↑ 1.55
Prosaposin	PSAP	↑ 2.21	↑ 2.13
KETOGENESIS		Azathioprine	Taxol
Hydroxyacyl-CoA dehydrogenase	HADH		↓ 1.43
	HADHA	↓ 1.57	
	HADHB	↑ 1.36	
3-Hydroxymethyl-3-methylglutaryl-CoA lyase	HMGCL	↑ 1.67	

Chemotherapy incremented the expression of proteins involved in autophagy and senescence. Proteins that show increased expression after treatment are shown in red, and proteins that are decreased are shown in green. Number indicates the fold increase or fold decrease in protein expression in chemotherapy-treated versus vehicle-treated hTERT-BJ1 fibroblasts.

#### 4. Autophagy and senescence

Autophagy and senescence represent a common response to stresses such as exposure to DNA-damaging exogenous cytotoxic agents, including chemotherapy or radiation [22, 23]. Indeed our previous study indicated that azathioprine and taxol induce autophagic vesicle formation and increase  $\beta$ -galactosidase activity in stromal fibroblasts [11]. Thus, to further examine changes in autophagy and senescence upon chemotherapeutic exposure, the differential expression

of several autophagy, mitophagy and lysosomal markers was analysed using the proteomics datasets. Numerous autophagy and senescence markers such as sequestrosome 1, also known as p62, cathepsin B or the lysosomal enzyme  $\beta$ -galactosidase were found to be up-regulated in either azathioprine or taxol treatments (Table 5). In addition, DNM1L, a protein involved in mitochondrial fission, was up-regulated in taxol-treated hTERT-BJ1 fibroblasts, and OPA1, a protein involved in mitochondrial fusion, was down-regulated in azathioprine-treated cells. Likewise, VAT1, a vesicle

membrane protein that inhibits mitochondrial fusion, was up-regulated in both treatments, suggesting that mitophagy might be also activated in response to chemotherapy (Table 5 and Table S3).

During autophagy, protein and lipid degradation occur, the latter leading to the generation of ketone bodies. A long list of up-regulated proteins involved in vesicle formation and trafficking, in protein ubiquitination pathway and proteasomal degradation, and a few enzymes involved in ketogenesis were also identified (Table 5, S3 and S4). The boost in vesicle formation and trafficking proteins could explain the greater presence of enzymes involved in fatty acid synthesis and hexosamine biosynthesis (Figure 4A and Table 1 and S2). The observed effects of taxol treatment on autophagy, mitophagy and senescence are summarised in Figure 4B.

IPA analysis confirmed an alteration of ketogenesis in fibroblasts exposed to azathioprine (Figure 5 and Table S5). IPA also showed a robust inhibitory effect of azathioprine and taxol on EIF2 signalling, responsible for protein synthesis, as measured by z score, and also revealed an alteration in the regulation of eIF4 and p70S6K signaling and in the protein ubiquitination pathway. Finally, PPAR $\alpha$ /RXR $\alpha$  activation, Rho, caveolar and clathrin-mediated signalling pathways, all involved in vesicle trafficking and motility, and the mTOR signalling, known for its role in autophagy, mitochondrial metabolism and lipid metabolism, as well as cytoskeleton dynamics, were some of the top canonical pathways altered by chemotherapy, (Figure 3E, 5, and Table S5). Interestingly, p53, known mediator of senescence,

was predicted to be activated in both azathioprine-treated and taxol-treated hTERT-BJ1 cells according to IPA, and TFEB, a transcription factor that coordinates the expression of lysosomal hydrolases, membrane proteins and genes involved in autophagy, was also predicted to be activated by taxol treatment (Figure 3F and 4C). Our previous study also reported an increased expression of p53 in hTERT-BJ1 fibroblasts treated with azathioprine and taxol by immunoblotting [11].

Thus, quantitative proteomics analysis reveals a higher presence of markers of autophagy and senescence and proteins involved in protein degradation and vesicle trafficking in hTERT-BJ1 fibroblasts, upon treatment.

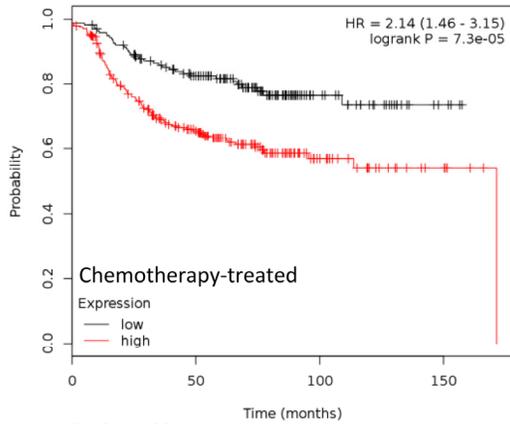
## 5. Inflammation

Senescent cells dramatically alter their secretome, enriching it with pro-inflammatory cytokines and matrix metalloproteinases. This senescence-associated secretory phenotype (SASP) can lead to chronic inflammation, which is a hallmark of aging [17]. Stromal fibroblasts secrete inflammatory cytokines when in contact with cancer cells [24]. Chemotherapy is also able to induce cytokine production in healthy tissues [12], and in particular taxol treatment induces IL6 secretion in stromal fibroblasts (Figure 1) [11]. STAT3, a known inducer of inflammation in response to stress [25, 26] was found significantly up-regulated in both azathioprine (1.69 fold increase) and taxol treatments (2.35 fold increase) relative to vehicle. STAT3 signalling was found to be activated in response to taxol also in our previous study [11].

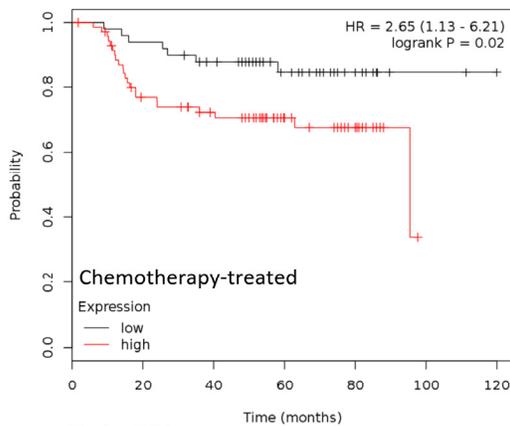
**Table 6. Proteins of the catabolic stroma signature and their contribution to CAF transformation, metabolism, antioxidant response, autophagy and vesicle trafficking.**

CATABOLIC STROMA SIGNATURE		Taxol	Process
Ubiquitin-like modifier activating enzyme 1	UBA1	↑ 2.36	Protein degradation
26S protease regulatory subunit 8	PSMC5	↑ 3.17	Protein degradation
Cation-independent mannose-6-phosphate receptor	IGF2R	↑ 1.88	Autophagy
Synaptic vesicle membrane protein VAT-1	VAT1	↑ 2.44	Vesicle trafficking / oxidoreductase /inhibits mitochondrial fusion
Heme oxygenase (decycling) 1	HMOX1	↑ 2.83	Antioxidant / oxidoreductase
Calponin 2	CNN2	↑ 2.76	Myofibroblastic differentiation
Talin 1	TLN1	↑ 2.95	Myofibroblastic differentiation
Glucosamine-6-phosphate deaminase 1	GNPDA1	↑ 2.46	Carbohydrate metabolism
Glucose-6-phosphate dehydrogenase	G6PD	↑ 2.69	Carbohydrate metabolism /generates antioxidant power
Fatty acid synthase	FASN	↑ 1.77	Fatty acid metabolism

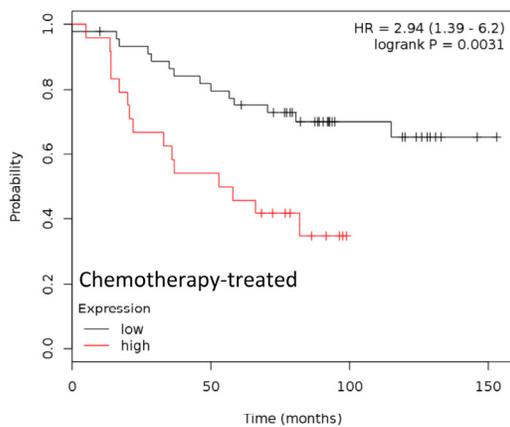
**A. UBA1**



Number at risk			
low	173	117	30
high	252	125	28

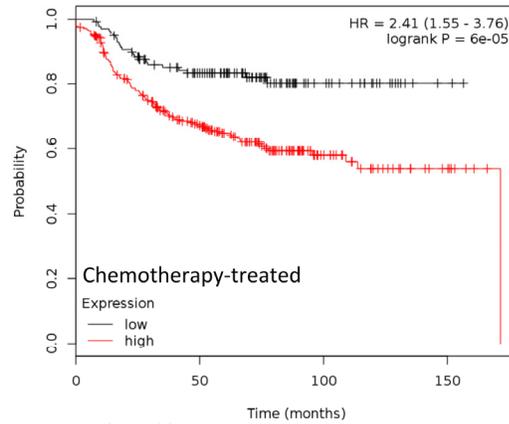


Number at risk						
low	50	47	41	24	10	2
high	72	51	43	26	15	0

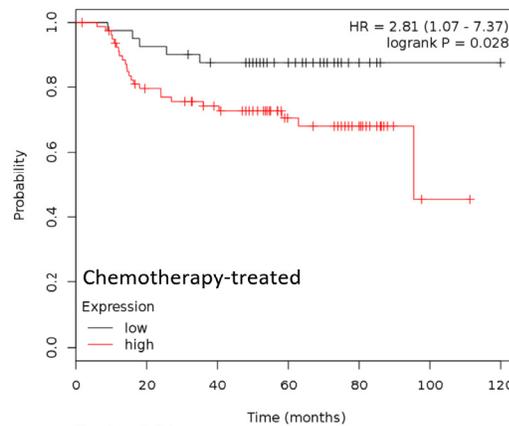


Number at risk			
low	45	36	15
high	24	13	0

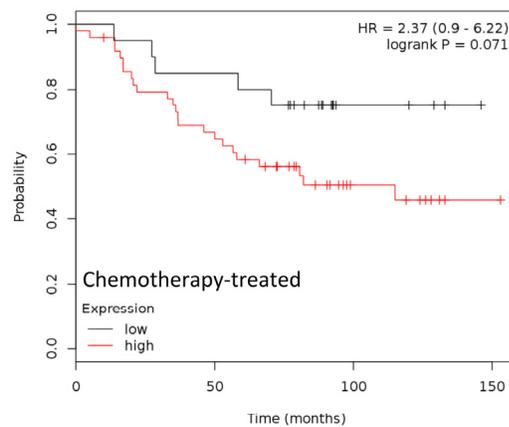
**B. Catabolic stroma signature**



Number at risk			
low	138	92	22
high	287	150	36



Number at risk						
low	41	38	34	21	7	1
high	81	60	50	29	18	1



Number at risk			
low	20	17	4
high	49	32	11

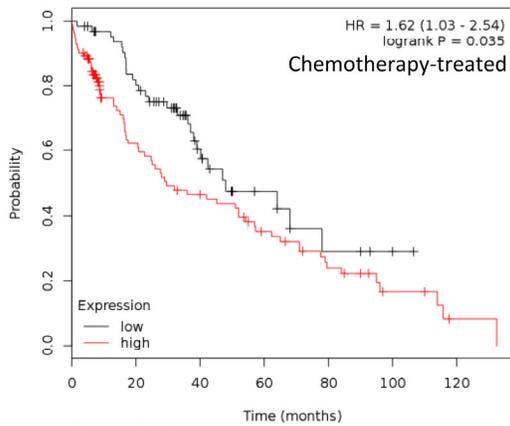
Relapse-free survival

Distant metastasis-free survival

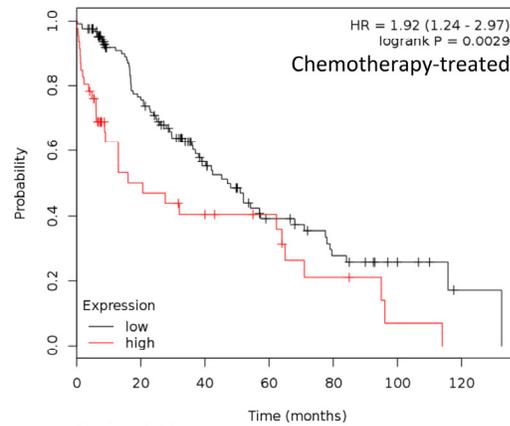
Overall survival

**Figure 7. Clinical correlations of UBA1 expression and the UBA1, PSMC5, IGF2R, VAT1, HMOX1, CNN2, TLN1, GNPDA1, G6PD and FASN signature in chemotherapy-treated breast cancer patients. (A) Correlations of UBA1 expression with relapse-free survival, distant metastasis-free survival and overall survival in breast cancer. (B) Correlations of the catabolic stroma signature with relapse-free survival, distant metastasis-free survival and overall survival in breast cancer. All graphs are calculated using microarray data from 425, 122 and 69 chemotherapy-treated breast cancers, respectively, determined using an online survival analysis tool. Kaplan-Meier correlations are plotted for high (above median, in red) and low (below median, in black) gene expression.**

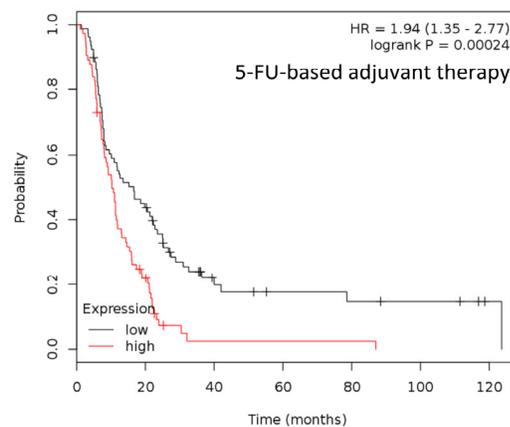
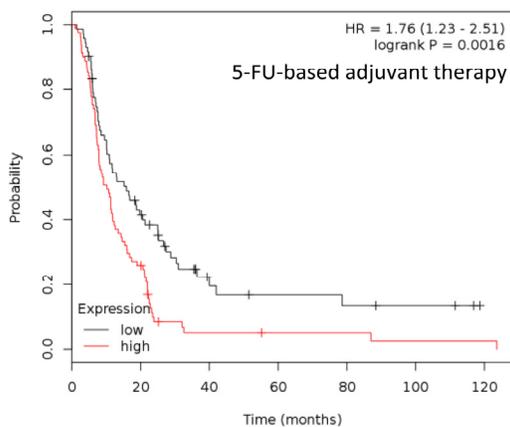
**A. UBA1**



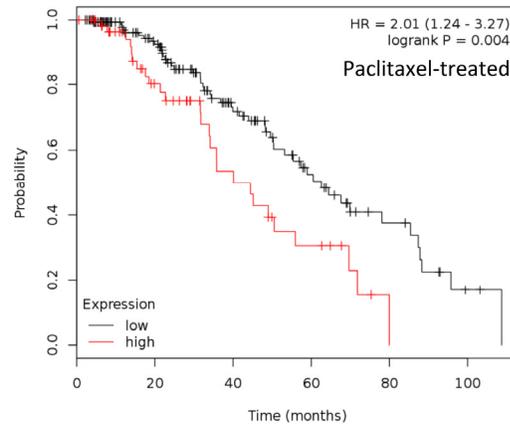
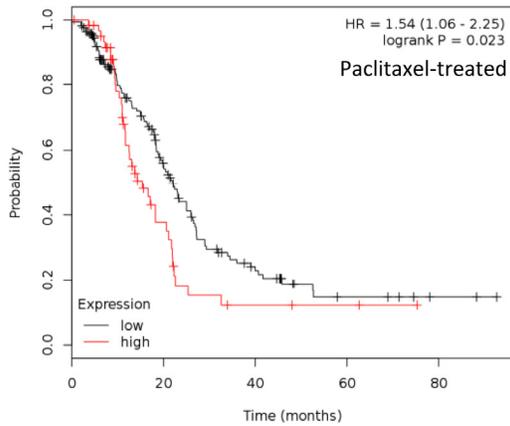
**B. Catabolic stroma signature**



Lung cancer; Overall survival



Gastric cancer; Overall survival



Ovarian cancer; Progression-free survival

**Figure 8. Clinical correlations of UBA1 expression and the UBA1, PSMC5, IGF2R, VAT1, HMOX1, CNN2, TLN1, GNPDA1, G6PD and FASN signature in chemotherapy-treated lung, gastric and ovarian cancer patients. (A) Correlations of UBA1 expression with overall survival in lung and gastric cancer, and with progression-free survival in ovarian cancer. (B) Correlations of the catabolic stroma signature with overall survival in lung and gastric cancer, and with progression-free survival in ovarian cancer. All graphs are calculated using microarray data from 176 chemotherapy-treated lung cancers, 153 5-FU-based adjuvant therapy-treated gastric cancers, and 229 paclitaxel-treated ovarian cancers, determined using an online survival analysis tool. Kaplan-Meier correlations are plotted for high (above median, in red) and low (below median, in black) gene expression.**

Interestingly, several pathways involved in the inflammatory process such as IL8, acute phase response, leukocyte extravasation or Fcγ receptor-mediated phagocytosis signalling were amongst altered canonical pathways in both azathioprine-treated and taxol-treated hTERT-BJ1 cells relative to vehicle, according to IPA. N-formyl-Met-Leu-Phe (fMLP) signalling was also found to be altered in taxol-treated fibroblasts (Figure 3E and 5 and Table S5). Most of these pathways were clearly activated in taxol-treated fibroblasts as indicated by z score values, suggesting an induction of the inflammatory response in chemotherapy-treated stromal fibroblasts. Likewise, the chemokine (C-C motif) ligand 5 (CCL5 or RANTES), which plays a role in recruiting leukocytes into inflammatory sites, was identified as an upstream regulator in both treatments, particularly activated in taxol-treated fibroblasts (Figure 3F and 4C). Finally, acute phase response, which occurs soon after the onset of an inflammatory process, was identified as one of the toxic effects of both drugs (Figure 6 and Table S6). Therefore, pathway analysis of the proteomics results indicates an induction of the inflammatory response in stromal fibroblasts after exposure to chemotherapy.

#### **Differentially up-regulated proteins in taxol-treated fibroblasts correlate with recurrence, metastasis and poor cancer survival**

To further investigate the clinical implications of our proteomics datasets, we decided to test the impact of over-expressed proteins in taxol-treated hTERT-BJ1 fibroblasts in cancer prognosis. To do so, we used an on-line survival analysis tool that uses microarray gene expression data from multiple studies on breast, ovarian, lung and gastric cancer [16], which was suitable to our purpose since taxol is a chemotherapeutic drug currently used as therapy for most of these malignancies. Only those proteins with a fold change difference of 1.75 or higher and p values of <0.05 (ANOVA) compared to vehicle were used for survival analyses. The expression of several proteins was found to correlate with survival in breast, ovarian, gastric and lung cancer patients. In particular, high expression of ubiquitin-like modifier activating enzyme 1 (UBA1), implicated in protein catabolism and degradation, showed a striking correlation with poor relapse-free survival, distant metastasis-free survival and overall survival in breast cancer patients previously treated with chemotherapy (425, 122 and 69 patients, respectively) (Figure 7A). The same correlation was not observed when patients who did not receive systemic treatment were considered (1000, 533 and 375 patients, respectively) (Figure S2A). Similarly, high expression of UBA1 correlated with poor overall survival in lung

and gastric cancer patients previously treated with chemotherapeutic drugs (176 and 153 patients, respectively), and with progression-free survival in ovarian cancer patients particularly treated with paclitaxel (229 patients) (Figure 8A). Once more, the correlation was lost when patients who did not receive systemic treatment were considered in lung cancer (227 patients) or when patients who underwent only surgery were examined in gastric cancer (174 patients) (Figure S3A). No data from untreated ovarian cancer patients was available.

A taxol-induced catabolic stroma signature was created comprising UBA1, and other proteins representative of the catabolic CAF-like phenotype, including PSMC5, and VAT1, involved in catabolism and vesicle trafficking, several metabolic enzymes including FASN, G6PD and GNPDA, an autophagy marker, IGF2R, the oxidoreductase HMOX1, and myofibroblastic markers including CNN2 and TLN1. See Table 6 for details and abundances of these proteins in taxol-treated fibroblasts. That catabolic stroma signature showed a strong correlation with survival in treated patients (Figure 7B and 8B), whereas that correlation was not seen in non-treated patients (Figure S2B and S3B). Hence, the over-expression of proteins upon treatment with taxol is strongly linked to poor survival, treatment failure and metastasis in breast, lung, gastric and ovarian cancers. Interestingly, most markers used for the signature are also proteins found to be significantly up-regulated in azathioprine-treated hTERT-BJ1 fibroblasts relative to vehicle, including IGF2R, PSMC5, VAT1, HMOX1, CNN2, TLN1 and FASN (Tables 3, 4 5, S2, S3 and S4). Thus, we conclude that chemotherapy-mediated changes in the abundance of stromal proteins related to the CAF-like catabolic phenotype measured by quantitative proteomics associate with reduced survival, enhanced recurrence and metastasis incidence in several solid malignancies.

## **DISCUSSION**

In this study we have analysed the impact of chemotherapy in the acquisition of the CAF-like, catabolic tumour stroma phenotype, which emerges in stromal fibroblasts in contact with cancer cells, and is characterised by increased glucose uptake, lactate production and extracellular acidification, increased expression of αSMA, augmented production of ROS, an activation of the JNK/AP1, HIF1, TGFβ/SMAD, STAT3 and NFκB stress-induced pathways, senescence and autophagy, and a greater secretion of inflammatory cytokines. To do so, we aimed at comprehensively characterising the cellular and metabolic changes that take place in stromal fibroblasts exposed to two

common chemotherapeutic drugs: azathioprine and taxol. Label-free quantitative proteomics and extensive bioinformatics analyses revealed a protein profile characteristic for chemotherapy-treated fibroblasts that included alterations in energy metabolism, antioxidant response, autophagy and senescence, vesicle formation and trafficking, protein degradation, myofibroblastic differentiation and inflammation.

### **Proteomic map of the stromal catabolic state induced by chemotherapy**

The main protein set whose expression was altered in chemotherapy-treated fibroblasts was that of enzymes. Azathioprine and taxol effects on the glycolytic function of stromal fibroblasts were characterised in a previous study, which showed the induction of a glycolytic switch in hTERT-BJ1 fibroblasts via increasing their lactate production and extracellular acidification, and decreasing their ATP content [11]. We know now that this behaviour is likely to be a result of the observed up-regulation of enzymes involved in glycolysis and down-regulation of mitochondrial respiration enzymes. Indeed, our proteomic analysis shows alterations not only in glycolysis but also in pyruvate conversion to acetyl-coenzyme A and the TCA cycle. The reduction in ATP production through decreased TCA cycle activity and mitochondrial respiration may be forcing a metabolic switch that allows the cell to obtain energy from alternative metabolic processes such as glycolysis or protein catabolism. This metabolic remodelling also includes an alteration of other pathways of the carbohydrate metabolism, such as the phosphate pentose and hexosamine biosynthesis pathways, and several pathways of the fatty acid metabolism. In line with our findings, a recent study demonstrates that as an adaptive response to mitochondrial respiratory chain dysfunction and ATP deficiency, human fibroblasts up-regulate the expression of glycolytic enzymes, suggesting the induction of anaerobic glycolysis and a cellular catabolic state, in particular protein catabolism, together with autophagy [27].

According to our data, the hexosamine biosynthesis pathway is also altered in chemotherapy-treated fibroblasts, which goes in line with the increased presence of proteins involved in vesicle formation and trafficking. GNPDA1 is an enzyme of the hexosamine biosynthesis pathway, the end-products of which are used for the synthesis of membrane components such as glycolipids and proteoglycans. However, GNPDA1 catalyzes a reversible conversion between D-fructose-6-phosphate and D-glucosamine-6-phosphate, and therefore could also be used in reverse to produce

substrates for glycolysis, by sacrificing structural components of the cell.

The phosphate pentose pathway is a metabolic pathway also altered by chemotherapy. During the oxidative phase of the pentose phosphate pathway, most of the reducing power of the cell is generated as NADPH. Taxol-treated fibroblasts show increased levels of the initial rate-limiting enzyme of the pentose phosphate pathway, glucose-6-phosphate dehydrogenase (G6PDH), which converts glucose-6-phosphate and NADP<sup>+</sup> into 6-phosphoglucono- $\delta$ -lactone and NADPH. NADPH is also necessary for lipid and nucleic acid biosynthesis, and its re-oxidation to NADP<sup>+</sup> constitutes an essential step to prevent damage by oxidative stress. Increased amounts of NADPH go in line with the up-regulation of several proteins involved in antioxidant response, such as Nrf2-mediated antioxidant response proteins, as well as the greater ROS levels that we have previously seen in hTERT-BJ1 fibroblasts after taxol treatment [11]. The accumulation of ROS as a consequence of mitochondrial dysregulation during aging is also associated with DNA damage [28]. Therefore, oxidative stress and aging can be coupled in a positive feedback mechanism that accelerates cellular damage and generates a permissive metabolic microenvironment for cancer development and progression [5, 28-30].

Oxidative stress in healthy cells and tissues by most current anti-cancer therapies is thought to occur via p53 activation, which causes mitochondrial dysfunction, ROS production and downstream STAT3 signalling, promoting inflammation-related cancer [19, 26]. Our proteomics results revealed a predicted activation state of p53, mitochondrial dysfunction, increased expression of antioxidant proteins, autophagy and senescence markers, and specifically STAT3 up-regulation in chemotherapy-treated fibroblasts compared to vehicle. Our observations in vesicle trafficking and protein transport and degradation also suggest a general intensification in the transport of proteins and other cellular components to autophagosomes and lysosomes. Indeed, taxol is able to induce the formation of autophagic vesicles in stromal fibroblasts [11]. By increasing that transport, the stressed cell could be attempting to remove dysfunctional organelles such as mitochondria and other molecules that accumulate as a result of oxidative damage, or alternatively it could also be trying to retrieve molecules for re-utilisation in detriment of *de novo* protein synthesis, which is more energy consuming.

Myofibroblasts are abundant components of the reactive tumour microenvironment and are mostly accountable for the development of fibrosis [21], one of the side-

effects of cancer therapy [31]. Fibroblast-to-myofibroblast differentiation increases in the stroma with age as well, resulting in an increased incidence of fibrosis-associated diseases, such as cancer [32]. Here, we report that exposure to chemotherapy can independently induce the transformation of fibroblasts into CAFs, by detecting an increase in the presence of numerous markers of myofibroblastic differentiation as measured by quantitative proteomics.

Finally, we identify several inflammation-related pathways as altered in taxol-treated fibroblasts, suggesting an induction of the inflammatory response. We previously described IL6 as one of the cytokines secreted by chemotherapy-treated fibroblasts and STAT3 target genes as being activated after treatment with taxol [11]. However, to comprehensively characterize the pro-inflammatory cytokines released by hTERT-BJ1 cells, secretome analysis should be performed for better insights.

In the present study, proteomic analyses revealed the impact of chemotherapy in the acquisition of a catabolic CAF-like phenotype in stromal fibroblasts, gradually conforming an energy-rich, pro-inflammatory micro-environment able to succour cancer cells in their battle for survival, in their purpose to metastasise, or even through which chemotherapy-damaged epithelial cells may become more susceptible to develop a fully malignant phenotype, giving rise to a new cancer. Taxol induces the catabolic stroma phenotype to a bigger extent than azathioprine. The different nature of these two drugs and the concentrations selected for the study could account for their different impact on catabolic remodelling.

### **Clinical implications of the catabolic stroma signature**

Quantitative proteomics allowed us to obtain a protein profile characteristic for fibroblasts exposed to azathioprine and taxol, which suggests a general alteration in their energy metabolism and a shift towards glycolysis and catabolic processes: a whole metabolic reprogramming to adapt to chemotherapy-driven mitochondrial dysfunction and oxidative stress. Likewise, the quantitative proteomics analysis allowed us to identify biomarkers of this stromal catabolic state, which were further analysed for their potential clinical implications. We generated a signature to figure out whether the chemotherapy-induced catabolic state in the stroma would have prognostic value. Indeed, the catabolic phenotype in stromal fibroblasts strikingly correlated with poor survival, treatment failure and metastatic growth in a set of breast, ovarian, lung and gastric cancer patients who were subject to chemo-

therapy, correlation that was lost when untreated patients were considered.

Our knowledge of the role of healthy stromal cells in metastasis and in particular in the emergence of therapy-related malignancies is still very scarce. A piece of evidence for chemotherapy-induced tumour-promoting paracrine activities of non-malignant cells has only been recently published, showing that pre-treatment of tumour-free mice with a single dose of doxorubicin is sufficient to stimulate the engraftment of lung carcinoma cells and to elevate the mitogenic activity of the serum from treated animals [12]. By label-free quantitative proteomics we detect the acquisition of a catabolic state in stromal cells due to chemotherapy-induced DNA damage, which potentially leads to the generation of an autophagic, nutrient-rich, senescent, pro-inflammatory microenvironment, the ideal niche to encourage the development of a secondary tumour or even a new carcinogenic process. However, the tumour-promoting function and the stability and reversibility of the chemotherapy-induced CAF phenotype needs to be further investigated, as well as the differential expression of chemotherapy-induced catabolic stress markers on stromal fibroblasts and cancer cells *in vivo*.

To conclude, proteomic analyses revealed a significant metabolic reprogramming in response to chemotherapy. Our data provide information about novel protein targets that might enable and support different stages of the tumorigenic process, thereby opening new doors for future research. Given the essential contribution of the catabolic tumour stroma in cancer progression, it emerges as a new interesting therapeutic target. A promising approach would be the preventive inhibition of the catabolic state transformation. Indeed, TGF $\beta$ -induced myofibroblastic transformation in fibroblasts can be reversed by using antioxidants [33]. Most importantly, an antibody against fibroblast activation protein (FAP) is already being tested in clinical trials [34]. Therefore, inhibition of the catabolic stress in the tumour stroma and healthy tissues in parallel to conventional chemotherapy could help avoiding recurrence, metastasis and the growth of second primary tumours.

## **MATERIALS AND METHODS**

Cell culture. All cell culture experiments were carried out using human foreskin fibroblasts immortalised with the human telomerase reverse transcriptase (hTERT-BJ1 cells). hTERT-BJ1 fibroblasts were originally purchased from ATCC (CRL-4001) and maintained in DMEM media (D6546, Sigma) supplemented with 10

% fetal bovine serum (FBS) (F7524, Sigma), 100 units/ml of penicillin, 100 µg/ml streptomycin (P0781, Sigma) and 1% Glutamax (#35050087, Life Technologies) at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub>.

Chemotherapeutic agents. Azathioprine (A4638, Sigma) and taxol or paclitaxel (Y0000698, Sigma) were used for this study at 100 µM and 100 nM, respectively.

Sulforhodamine B (SRB) assay. SRB (S9012, Sigma) measures total biomass by staining cellular proteins. After 48 h treatment, cells were fixed in 10% trichloroacetic acid (T9159, Sigma) for 1h at 4°C, stained with SRB (S9012, Sigma) for 15 minutes, and washed 3 times with 1% acetic acid (27225, Sigma). The incorporated dye was solubilized with 10 mM Tris Base, pH 8.8 (T1503, Sigma). Absorbance was spectrophotometrically measured at 562 nm in a FluoStar Omega plate reader (BMG Labtech). Background measurements were subtracted from all values.

#### Label-free quantitative proteomics

*Chemicals and sample preparation.* Formic acid, trifluoroacetic acid, ammonium formate (10 M), ammonium bicarbonate TCEP (Tris (2-carboxyethyl)phosphine hydrochloride), MMTS (Methyl methanethiosulfonate) and trypsin were all obtained from Sigma. HPLC gradient grade acetonitrile was obtained from Fisher Scientific. Briefly, 2 x 10<sup>6</sup> hTERT-BJ1 fibroblasts were seeded in 150 cm plates until cells were attached. Cells were then treated with azathioprine or taxol at the concentrations indicated. As control, vehicle-treated cells were processed in parallel. After 48 hours of treatment, cells were lysed in RIPA buffer (R0278, Sigma) and kept at 4°C for 20 minutes with rotation. Lysates were cleared by centrifugation for 10 minutes at 10,000 x g and supernatants were collected and kept frozen at -80°C.

*Protein digestion.* Lysate samples were thawed to room temperature and their concentrations equalised to 1 µg/µL (50 µL volume) with RIPA buffer, and further processed for trypsin digestion by sequential reduction of disulphide bonds with TCEP and alkylation with MMTS. Briefly, 1 µL benzonase (Novagen) was added to the 50 µL aliquot and placed on ice for 15 minutes. The sample was then taken to dryness using a SpeedVac, and resuspended in 22.5 µL trypsin reaction buffer (40 mM ammonium bicarbonate and 9% acetonitrile). One µL of 50 mM TCEP solution was added to each sample, mixed briefly and placed on a heater block at 60°C for 60 minutes. After cooling to

room temperature, 0.5 µL of 200 mM MMTS solution was added to each sample and allowed to react for 15 minutes. Trypsin was added in two waves to ensure efficient digestion of the sample. Firstly, 20 µg of sequencing grade trypsin was resuspended in 1800 µL of trypsin reaction buffer; 225 µL of this solution were added to each sample for digestion, and the reactions were left at 37°C overnight with shaking (600 rpm). The following morning, a further aliquot of trypsin was added. Two ml of trypsin reaction buffer was added to 20 µL of sequencing grade trypsin; 250 µL of this solution were added to each of the digest samples from overnight, and the reactions were left at 37°C for 4 hours with shaking (600 rpm). Thirty-five µL 10% formic acid were added to the 500 µL digest sample (0.7% final concentration of formic acid) to stop the digestion. The digested solution was diluted in 7.5 mL of acetonitrile containing 0.3% formic acid.

*HILIC solid phase extraction (SPE) of peptides.* PolyhydroxyethylA SPE 12 µm, 300A, 300mg cartridges (obtained from PolyLC) were used for the HILIC procedure. Prior to use, cartridges required an overnight soak in 50 mM formic acid followed by rinsing with water the following day. Cartridges were preconditioned with 2 mL of Buffer A (90% acetonitrile, 5 mM ammonium formate, pH 2.7) followed by 2 mL of Buffer B (5mM ammonium formate, pH 2.7) and finally re-equilibrated with 10 mL Buffer A. The diluted samples were loaded onto the cartridges and washed with a further 10 mL Buffer A. Finally, peptides were eluted in 1 mL Buffer C (9 parts Buffer B plus 1 part Buffer A) and the samples dried on a Speedvac to remove organic solvent prior to LC-MS/MS analysis.

*LC-MS/MS analysis.* Lyophilised digests were resuspended in 50 µL of 0.1% TFA to give an approximate concentration of 1 µg/µL. One µL injection volumes were used throughout resulting in an on-column peptide loading of approximately 1 µg per injection. Analysis was performed in quintuplicate for each sample. All LC-MS/MS analyses were performed on an LTQ Orbitrap XL mass spectrometer coupled to an Ultimate 3000 RSLCnano system (Thermo Scientific). One µL injection volumes were used throughout and samples loaded directly onto the analytical column, PepMap RSLC C18, 2 µm x 75 µm id x 50 cm (Thermo Scientific). The composition (v/v) of LC buffers were as follows; Buffer A - 99.9% water plus 0.1% formic acid and Buffer B - 80% acetonitrile, 19.9% water and 0.1% formic acid. Peptides were loaded directly onto the column at a flow rate of 400 nl/min with an initial mobile phase composition of 1% B. The organic strength was increased linearly from 1%

to 22.5% B over 22.5 minutes again at 400 nl/min, followed by an increase to 24.8% B over the next 2.6 minutes with a concomitant reduction in flow rate to 300 nl/min, and to 39% B over a further 14 minutes. A further increase to 60% B over the next 5 minutes was followed by a ramp to 95% B over 2.5 minutes where it was held for a further 2 minutes. The column was then allowed to re-equilibrate to 1% B for a total analysis time of 74 minutes. The mass spectrometer was instructed to perform data dependent acquisition on the top six precursor ions, which were measured in the Orbitrap FTMS detector over the mass range 370–1200 m/z, at a nominal resolution of 60,000. MS/MS spectra were acquired in the ion trap under CID conditions with normalized collision energy of 35, isolation width of 3 Th, Q value of 0.25 and 30 ms activation time. Gas-phase fractionation was performed on the five replicate injections such that MS/MS data was collected for precursor ion range 370-494 m/z Injection 1, 494-595 m/z Injection 2, 595-685 m/z Injection 3, 685-817 m/z Injection 4 and 817-1200 m/z Injection 5.

*Statistical analysis.* Xcalibur raw data files acquired on the LTQ-Orbitrap XL were directly imported into Progenesis LCMS software (Waters Corp) for peak detection and alignment. Data were analysed using the Mascot search engine. Five replicates were analysed for each sample type (N = 5). Statistical analyses were performed using ANOVA and only fold-changes in proteins with a p-value less than 0.05 were considered significant.

Ingenuity pathway analyses. Pathway and function analyses were generated using Ingenuity Pathway Analysis (IPA) (Ingenuity systems, <http://www.ingenuity.com>), which assists with proteomics data interpretation via grouping differentially expressed genes or proteins into known functions and pathways. Pathways with a z score >1.9 were considered as significantly activated, and pathways with a z score <-1.9 were considered as significantly inhibited.

Graphs and correlation analyses. All graphs were done in Microsoft Excel except for correlation graphs. Correlations between protein expression and patient survival were calculated using a survival analysis tool available online (kmplot.com) [16].

## ACKNOWLEDGEMENTS

We would like to thank the mass spectrometry services from the Cancer Research UK Manchester Institute (Paterson building), for performing the unbiased proteomics and the initial statistical analysis. Ingenuity

Pathway Analysis was supported by the University of Manchester Wellcome Trust grant 097820/Z/11/B.

## Funding

We thank the University of Manchester for providing start-up funds that contributed to the success of this study. The Sotgia and Lisanti Laboratories were supported, in part, by funding from the European Union (ERC Advanced Grant), Breast Cancer Now, and the Manchester Cancer Research Centre (MCRC).

## Author contributions

MPL first proposed this UK-based project. MPP performed all the experiments, analyzed the data and generated the figures with experimental data. MPP wrote the first draft of the manuscript, which was edited by MPL and FS. Comments and corrections from DLS and BG were also incorporated. Unbiased proteomics and the statistical analysis of the proteomic results were performed by the Biological Mass Spectrometry Core Facility, at the Cancer Research UK Manchester Institute, under the supervision of DLS.

## Conflict of interest statement

The authors declare there are no potential conflicts of interest.

## REFERENCES

1. Elkhattouti A, Hassan M and Gomez CR. Stromal Fibroblast in Age-Related Cancer: Role in Tumorigenesis and Potential as Novel Therapeutic Target. *Frontiers in oncology*. 2015; 5:158.
2. Lasry A and Ben-Neriah Y. Senescence-associated inflammatory responses: aging and cancer perspectives. *Trends in immunology*. 2015; 36:217-228.
3. Hanahan D and Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell*. 2012; 21:309-322.
4. Lisanti MP, Sotgia F, Pestell RG, Howell A and Martinez-Outschoorn UE. Stromal glycolysis and MCT4 are hallmarks of DCIS progression to invasive breast cancer. *Cell Cycle*. 2013; 12:2935-2936.
5. Balliet RM, Capparelli C, Guido C, Pestell TG, Martinez-Outschoorn UE, Lin Z, Whitaker-Menezes D, Chiavarina B, Pestell RG, Howell A, Sotgia F and Lisanti MP. Mitochondrial oxidative stress in cancer-associated fibroblasts drives lactate production, promoting breast cancer tumor growth: understanding the aging and cancer connection. *Cell Cycle*. 2011; 10:4065-4073.
6. Bartling B, Hofmann HS, Silber RE and Simm A. Differential impact of fibroblasts on the efficient cell death of lung cancer cells induced by paclitaxel and cisplatin. *Cancer Biol Ther*. 2008; 7:1250-1261.
7. Ertel A, Tsigirgos A, Whitaker-Menezes D, Birbe RC, Pavlides S, Martinez-Outschoorn UE, Pestell RG, Howell A, Sotgia F and

- Lisanti MP. Is cancer a metabolic rebellion against host aging? In the quest for immortality, tumor cells try to save themselves by boosting mitochondrial metabolism. *Cell Cycle*. 2012; 11:253-263.
8. Nieman KM, Kenny HA, Penicka CV, Ladanyi A, Buell-Gutbrod R, Zillhardt MR, Romero IL, Carey MS, Mills GB, Hotamisligil GS, Yamada SD, Peter ME, Gwin K and Lengyel E. Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat Med*. 2011; 17:1498-1503.
  9. Zhang W, Trachootham D, Liu J, Chen G, Pelicano H, Garcia-Prieto C, Lu W, Burger JA, Croce CM, Plunkett W, Keating MJ and Huang P. Stromal control of cystine metabolism promotes cancer cell survival in chronic lymphocytic leukaemia. *Nat Cell Biol*. 2012; 14:276-286.
  10. Sun Y, Campisi J, Higano C, Beer TM, Porter P, Coleman I, True L and Nelson PS. Treatment-induced damage to the tumor microenvironment promotes prostate cancer therapy resistance through WNT16B. *Nature medicine*. 2012; 18:1359-1368.
  11. Peiris-Pagès MS, F.; Lisanti, MP. Chemotherapy induces the cancer-associated fibroblast phenotype, activating paracrine Hedgehog-Gli signaling in breast cancer cells. *Oncotarget*. 2015; 6:10728-10745.
  12. Porter DC, Farmaki E, Altilia S, Schools GP, West DK, Chen M, Chang BD, Puzyrev AT, Lim CU, Rokow-Kittell R, Friedhoff LT, Papavassiliou AG, Kalurupalle S, Hurteau G, Shi J, Baran PS, et al. Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. *Proc Natl Acad Sci U S A*. 2012; 109:13799-13804.
  13. Karran, P. & Attard, N. Thiopurines in current medical practice: molecular mechanisms and contributions to therapy-related cancer. *Nat Rev Cancer*. 2008; 8:24-36.
  14. Raja FA, Chopra N and Ledermann JA. Optimal first-line treatment in ovarian cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2012; 23 Suppl 10:x118-127.
  15. Vyas D, Laput G and Vyas AK. Chemotherapy-enhanced inflammation may lead to the failure of therapy and metastasis. *OncoTargets and therapy*. 2014; 7:1015-1023.
  16. Györfy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q and Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*. 2010; 123:725-731.
  17. Lopez-Otin C, Blasco MA, Partridge L, Serrano M and Kroemer G. The hallmarks of aging. *Cell*. 2013; 153:1194-1217.
  18. Kumari KK and Setty OH. Protective effect of *Phyllanthus fraternus* against mitochondrial dysfunction induced by co-administration of cisplatin and cyclophosphamide. *J Bioenerg Biomembr*. 2012; 44:179-188.
  19. Velez JM, Miriyala S, Nithipongvanitch R, Noel T, Plabplueng CD, Oberley T, Jungsuwadee P, Van Remmen H, Vore M and St Clair DK. p53 Regulates oxidative stress-mediated retrograde signaling: a novel mechanism for chemotherapy-induced cardiac injury. *PLoS One*. 2011; 6:e18005.
  20. Paunescu V, Bojin FM, Tatu CA, Gavriliuc OI, Rosca A, Gruia AT, Tanasie G, Bunu C, Crisnic D, Gherghiceanu M, Tatu FR, Tatu CS and Vermesan S. Tumour-associated fibroblasts and mesenchymal stem cells: more similarities than differences. *J Cell Mol Med*. 2011; 15:635-646.
  21. Desmouliere A, Darby IA and Gabbiani G. Normal and pathologic soft tissue remodeling: role of the myofibroblast, with special emphasis on liver and kidney fibrosis. *Lab Invest*. 2003; 83:1689-1707.
  22. Campisi J and d'Adda di Fagagna F. Cellular senescence: when bad things happen to good cells. *Nature reviews Molecular cell biology*. 2007; 8:729-740.
  23. Goehe RW, Di X, Sharma K, Bristol ML, Henderson SC, Valerie K, Rodier F, Davalos AR and Gewirtz DA. The autophagy-senescence connection in chemotherapy: must tumor cells (self) eat before they sleep? *J Pharmacol Exp Ther*. 2012; 343:763-778.
  24. Martinez-Outschoorn UE, Whitaker-Menezes D, Lin Z, Flomenberg N, Howell A, Pestell RG, Lisanti MP and Sotgia F. Cytokine production and inflammation drive autophagy in the tumor microenvironment: role of stromal caveolin-1 as a key regulator. *Cell Cycle*. 2011; 10:1784-1793.
  25. Grivennikov SI, Greten FR and Karin M. Immunity, inflammation, and cancer. *Cell*. 2010; 140:883-899.
  26. Kamp DW, Shacter E and Weitzman SA. Chronic inflammation and cancer: the role of the mitochondria. *Oncology*. 2011; 25:400-410, 413.
  27. Marin-Buera L, Garcia-Bartolome A, Moran M, Lopez-Bernardo E, Cadenas S, Hidalgo B, Sanchez R, Seneca S, Arenas J, Martin MA and Ugalde C. Differential proteomic profiling unveils new molecular mechanisms associated with mitochondrial complex III deficiency. *Journal of proteomics*. 2015; 113:38-56.
  28. Lisanti MP, Martinez-Outschoorn UE, Pavlides S, Whitaker-Menezes D, Pestell RG, Howell A and Sotgia F. Accelerated aging in the tumor microenvironment: connecting aging, inflammation and cancer metabolism with personalized medicine. *Cell cycle*. 2011; 10:2059-2063.
  29. Martinez-Outschoorn UE, Curry JM, Ko YH, Lin Z, Tuluc M, Cognetti D, Birbe RC, Pribitkin E, Bombonati A, Pestell RG, Howell A, Sotgia F and Lisanti MP. Oncogenes and inflammation rewire host energy metabolism in the tumor microenvironment: RAS and NFkappaB target stromal MCT4. *Cell Cycle*. 2013; 12:2580-2597.
  30. Lisanti MP, Martinez-Outschoorn UE, Lin Z, Pavlides S, Whitaker-Menezes D, Pestell RG, Howell A and Sotgia F. Hydrogen peroxide fuels aging, inflammation, cancer metabolism and metastasis: the seed and soil also needs "fertilizer". *Cell cycle*. 2011; 10:2440-2449.
  31. Salata C, Ferreira-Machado SC, De Andrade CB, Mencalha AL, Mandarim-De-Lacerda CA and de Almeida CE. Apoptosis induction of cardiomyocytes and subsequent fibrosis after irradiation and neoadjuvant chemotherapy. *International journal of radiation biology*. 2014; 90:284-290.
  32. Sampson N, Berger P and Zenzmaier C. Therapeutic targeting of redox signaling in myofibroblast differentiation and age-related fibrotic disease. *Oxidative medicine and cellular longevity*. 2012; 2012:458276.
  33. Cat B, Stuhlmann D, Steinbrenner H, Alili L, Holtkotter O, Sies H and Brenneisen P. Enhancement of tumor invasion depends on transdifferentiation of skin fibroblasts mediated by reactive oxygen species. *Journal of cell science*. 2006; 119:2727-2738.
  34. Scott AM, Wiseman G, Welt S, Adjei A, Lee FT, Hopkins W, Divgi CR, Hanson LH, Mitchell P, Gansen DN, Larson SM, Ingle JN, Hoffman EW, Tanswell P, Ritter G, Cohen LS, et al. A Phase I dose-escalation study of sibtuzumab in patients with advanced or metastatic fibroblast activation protein-positive cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2003; 9:1639-1647.

## **SUPPLEMENTARY DATA**

Please follow the links in Full text version of this manuscript to see Supplementary Figures and Tables.

## What biomarkers (if any) for precise medicine?

Sabrina Strano, Paola Muti, and Giovanni Blandino

The advent of the *OMIC* technologies has strongly evolved the knowledge about the origin, the type and the response to therapy of a given tumor. To date we are aware that the epigenetic and genomic landscapes of tumors which origin, histopathological diagnoses and clinical stages are almost identical can be highly heterogeneous. Initially, the Human Genome Project represented the reference map for the human genome and provided the ideal background for the development of technology and analytic tools to decipher and rationalize enormous quantities of genomic data [1]. Subsequently, the National Research Council reported on the requirement of a precise taxonomy of human disease based on the continuous flow of molecular data originating from the *OMIC* approaches. This led The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) toward the molecular taxonomy of different human cancers. A large spectrum of gene mutations has been identified [1]. They can be categorized in: (a) *passenger* mutations that are the majority and may be biologically inactive and clinically irrelevant; (b) *driver* mutations whose activity is required for the aberrant growth, survival and chemoresistance of human cancers. Driver mutations have been the main molecular targets to be tackled with “smart” drugs, thus providing the rationale for precise medicine. Next Generation Sequence (NGS) technology has enabled to identify actionable targets such as EGFR in lung cancer and BRAF in melanoma [1,2]. Since these drugs benefit only those patients carrying specific driver mutations the identification of biomarkers that can predict treatment responses is vital for the success of the precise cancer therapy and for the development of anticancer drugs. EGFR mutations are considered biomarkers for selecting lung cancer patients for the treatment with EGFR inhibitors [3]. Gefitinib and erlotinib represent the first choice for the treatment of lung cancer patients carrying EGFR mutations and prolong significantly the progression-free survival of the selected patients. Despite it, both gefitinib and erlotinib cannot be used to treat all lung cancer patients harbouring EGFR mutations due to mutation site heterogeneity which negatively impacts on the affinity of EGFR inhibitors to the mutated EGFR and consequently of the efficacy of the treatment. Lung cancer patients develop resistance to EGFR inhibitors

due mostly common (50% of EGFR mutated lung cancer patients) to additional EGFR T90M mutation [3]. Unlike EGFR, other *driver* mutations as those affecting the p53 gene, the most frequent target of genetic alterations in human cancers, have not yet led to the development of targeted drugs to be used in the treatment of human cancers carrying mutant p53 proteins [4]. This clearly says, that while thousands of cancer genome profiles have enormously improved the molecular taxonomy of human cancers, they have only paved a background for precise cancer therapy which urges to be continuously fed towards the identification of precise cancer biomarkers. The improvement of methodologies for the isolation of circulating tumoral DNA from patients enrolled in cancer genome-driven trials coupled with NGS might contribute to tailor more precisely cancer therapy [1]. At the same time, we have learned from the *OMIC* technologies that what so called non-coding portion of the human genome plays a fundamental role in regulating the expression and the activity of the genomic coding regions [5]. The last two decades have witnessed the identification of non-coding transcripts which accordingly to their respective lengths have been distinguished in long non-coding RNAs (lncRNAs), microRNAs, small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). MicroRNAs, which regulate gene expression at the posttranscriptional level either inhibiting translation or promoting degradation of target mRNAs, emerge to be powerful to distinguish tumor tissues from their matched surrounding non-tumoral samples, to classify tumor histotypes, to predict tumor recurrence, to identify responders vs non-responders and to monitor response to cancer therapy [5,6,7]. MicroRNAs might represent early indicators of future breast cancer incidence. Previous evidence has shown that metabolic and environmental risk factors may alter the expression of microRNAs. MicroRNA profiling of the leucocytes of healthy pre-menopausal women recruited in the ORDET prospective cohort study over a follow-up period of 20 years revealed that microRNA downregulation represents a very early alteration in the development of breast cancer [8]. Selected microRNA alterations identified in ORDET were also found in different breast cancer databases, thus strengthening their value as early long-term predictors of breast cancer

occurrence [8]. MicroRNAs can also be found in blood and other biological fluids as circulating factors lined into exosomal vesicles. Despite the molecular mechanisms underlying the production and the release from tumoral cells and the intrinsic processing occurring in the exosomes are yet underexplored their potential to unveil powerful and precise cancer biomarkers is certainly promising and might provide with an additional option to treat cancer successfully.

## REFERENCES

1. Roychowdhury S. et al. *Annu. Rev. Genomics Hum. Genet.* 2014; 15:395-415.
2. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014; 511:543-550.
3. Yasuda H. et al. *Lancet Oncol.* 2011; 12:1182-1194.
4. Lehamann S. et al. *J. Clin. Oncology.* 2012; 29:3633-3639.
5. Biagioni F. et al. *Cell Cycle.* 2013; 15:2371-2375.
6. Gangi F. et al. *Annals of Oncology.* 2013; 24:3082-3088.
7. Mori F. et al. *Oncotarget.* 2015; 6:13947-13961.
8. Muti P. et al. *Cancer Epidemiol. Biomarkers Pre.* 2014; 11:2471-2481.

Giovanni Blandino: Translational Oncogenomic Unit,  
Regina Elena National Cancer Institute, Rome, Italy

**Correspondence:** *Giovanni Blandino*

**Email:** [blandino@ifo.it](mailto:blandino@ifo.it)

**Received:** August 18, 2015

**Published:** August 29, 2015

## Biomarkers and subtypes of cancer

Maud H.W. Starmans and Paul C. Boutros

The identification of prognostic and predictive biomarkers is a key research area in medicine. These biomarkers aim to contribute to personalize medicine. Ultimately, in personalized medicine treatment will be tailored towards each patient's specific disease and genetics to optimize treatment outcome and minimize side effects. In cancer research large efforts are made to screen for biological entities like gene mutations and transcription-based biomarkers for this purpose, however the identified markers are most of the time not accurate enough for clinical use. Recently we have shown that confounding factors play an important role in the limited performance of such (bio)markers [1]. Mutations in the RAS gene, a gene frequently mutated in lung cancer, were not prognostic [2], however they largely influenced accuracy of transcription-based biomarkers for non-small cell lung cancer. Taking RAS mutations to define patient subgroups and define transcription-based biomarkers for these specific patient subgroups resulted in an increase in prognostic power. While screening for prognostic or predictive markers it will thus be key to be aware of and correct for potential confounders. Therefore to create clinically useful biomarkers it will be detrimental to define clinically relevant patient subgroups rather than generalize across patients.

This general principle might apply to a broad range of other variables and studies. For example, one can imagine different biomarkers being optimal in older vs. younger patients, in men vs. women and especially based on a broad range of other tumour genetic information. To this last point, large studies such as those initiated by The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) will provide a wealth of data to exploit these findings. These studies can be used to define clinically-relevant patient subgroups based on genetic heterogeneity, rather than investigating single entities. For example, one can imagine systematic studies to identify genes that, while not themselves prognostic, confound the accuracy of other prognostic markers. Or, indeed, confound the accuracy of other biomarkers entirely: diagnostic or predictive markers, or markers for monitoring disease progress could all follow this general template.

To perform such analyses, it will be critical to rigorously assess the information content of different classes of biomarkers in different clinical situations. For example, we established interplay between RAS mutation and expression of a set of 14 genes; a gene expression-based classifier could be used to predict RAS mutation status. A large number of random gene sets were used to show this RAS predictor had optimal performance. Further large permutation studies, testing millions of random gene sets for their prognostic power, established that predicting prognosis for patients with RAS mutations should be done with different gene sets than for patients without RAS mutations. Testing large sets of random gene sets also provides valuable information for performance of transcriptome-based biomarkers. Comparing performance of the biomarker against the performance distribution of the random gene sets will immediately show whether these perform better than random and are worthwhile proceeding with [3, 4].

Taken together, these data point at a sea-change in the development of biomarkers. Rather than simply focusing on finding the best "signature" to predict a specific clinical event [5, 6], we will look to further sub-stratify patient populations into subtypes that can be accurately prognosed. Indeed, while these subtypes themselves may not be inherently informative, they may provide the structure or framework upon which more accurate biomarkers can be developed. We can foresee the adoption of information content methods like those described above to try to identify proactively specific genomic events that mark groups of patients with coherently predictable clinical outcome.

## REFERENCES

1. Starmans MH et al. *Clin Cancer Res.* 2015; 21: 1477-1486.
2. Shepherd FA et al. *J Clin Oncol.* 2013; 31: 2173-2181.
3. Starmans MH et al. *PLoS One.* 2011; 6: e28320.
4. Boutros PC et al. *Proc Natl Acad Sci U S A.* 2009; 106: 2824-2848.
5. Lalonde E et al. *Lancet Oncol.* 2014; 15: 1521-1532.
6. Starmans MH et al. *Br J Cancer.* 2008; 99: 1884-1890.

Paul C. Boutros: Informatics & Biocomputing Program,  
Ontario Institute for Cancer Research, Toronto, Canada

**Correspondence:** *Paul C. Boutros*

**Email:** [Paul.Boutros@oicr.on.ca](mailto:Paul.Boutros@oicr.on.ca)

**Received:** April 28, 2015

**Published:** May 3, 2015

## HGF-MET as a breast cancer biomarker

Shuying Liu

Although targeting MET has yielded promising results in preclinical studies, few clinical trials of MET inhibitors have demonstrated the expected therapeutic benefits. This inconsistency raises the possibilities that there are different subsets of MET/HGF-aberrant breast cancer with different responses to MET/HGF-targeted therapies and that MET inhibitors benefit only a particular subgroup of patients. To fully realize the benefits of MET inhibitors, we must clarify the patient population that will benefit from them. Therefore, it is crucial to identify biomarkers that will optimize the use of MET inhibitors in individual breast cancer patients.

**MET aberrations.** MET overexpression with or without amplification has been reported in many cancers [1]. MET protein expression and its phosphorylation were aberrantly upregulated in around 70% and 48% breast cancers, respectively, which independently predict poor outcome [2]. The MET signaling also can be aberrantly activated as a consequence of *MET* mutation or functional single nucleotide polymorphisms (SNPs). Several *MET* gene mutations have been reported in various cancers, but oncogenic *MET* mutations occur spontaneously in only 2-3% [1]. Mutation in the tyrosine kinase domain renders the enzyme constitutively active, while mutation in the juxtamembrane domain reduces MET degradation. *MET* sequence changes occur in 9% of patients with breast cancer. However, these appear to represent SNPs rather than somatic mutations. The sequence changes were associated with higher metastatic burden and high-grade histology [3]. A recent study demonstrated that *MET* is functionally altered by an uncommon germline SNP, *MET-T1010I*, which is present twice as frequently in patients with metastatic breast cancer as in the general population. *MET-T1010I* transforms mammary epithelial cells and drives tumor formation and invasion in human *HGF* transgenic mice [4], suggesting that it potentially alters tumor pathophysiology and response to MET-targeted therapies. Therefore, *MET-T1010I* should be considered a potential biomarker when implementing clinical trials of MET-targeted agents.

**Hepatocyte growth factor (HGF)/scatter factor upregulation.** Not only the MET receptor but also its ligand HGF drives tumor formation, metastasis, and drug resistance [1]. We recently established a mouse model system in which the host mice express human

HGF at varying levels and the xenografts express human MET receptor of varying status. The models faithfully mimic patients with different HGF levels and different MET receptor status in their breast cancer. Using this model, we found that MCF-10A cells transformed with aberrant MET formed tumors in the mice with transgenic human *HGF* but not in the negative litters. Comparing mutant MET, wild type MET is more stringently dependent on its natural ligand HGF. These data suggest that not only MET status affects cell behavior but also level of its ligand HGF in the tumor microenvironment plays a key role in determining the functional outcomes of MET aberrations. Indeed, overexpression of HGF has been demonstrated in breast cancer, and HGF levels are increased in the serum of patients with breast cancer. Both primary and metastatic tumor cells (autocrine mechanism) and stromal cells (paracrine mechanism) secrete high levels of HGF and aberrantly induce ligand-dependent MET signaling. High HGF level correlates with poor prognosis in breast cancer [1]. Hypoxic conditions stimulate production of both MET receptor and HGF, rendering tumor cells more sensitive to HGF stimulation in the invasion process [5]. Therefore, HGF level also should be considered a potential biomarker when testing MET-targeted agents.

**Crosstalk between MET and other pathways.** Accumulating evidence suggests that MET plays a key role in resistance to targeted therapies for cancer through crosstalk between MET and other pathways, such as the EGFR family [6]. Inhibition of either MET or EGFR was insufficient to fully block signaling in gefitinib-resistant cell lines, whereas the combination completely inhibited signaling. Indeed, combined targeting of MET with onartuzumab and EGFR with erlotinib in a clinical trial prolonged progression-free survival and overall survival in patients with lung cancer expressing high levels of MET [7]. However, the mechanisms of acquired resistance to MET inhibition remain little known. Crosstalk with other oncogenic pathways might induce acquired resistance to MET inhibition. We demonstrated that concurrent aberration of *MET* and *PIK3CA* greatly increased in breast cancer. Our unpublished data show that concurrent aberration of MET and PI3K significantly increased cell proliferation and invasion *in vitro* and in mice with

similar human HGF levels. Targeting both MET and PI3K yielded greater inhibitory efficacy than targeting either agent alone, suggesting that response to MET-targeted therapy in breast cancer is dependent on an aberrant MET-HGF/PI3K axis.

Taken together, these findings indicate that, to optimize the use of therapies targeting MET signaling and improve treatment efficacy in individual breast cancer patients, MET status, HGF level, and activation of the MET-HGF/PI3K and EGFR-MET axes should be considered as potential biomarkers when implementing clinical trials of MET-targeted agents.

## REFERENCES

1. Sierra JR, Tsao MS. *Ther Adv Med Oncol* 2011; 3:S21-S35.
2. Raghav KP et al. *Clin Cancer Res*. 2012; 18:2269-2277.
3. de Melo Gagliato D et al. *Clin Breast Cancer*. 2014; 14:468-474.
4. Liu S et al. *Oncotarget*. 2015; 6:2604-2614.
5. Pennacchietti S et al. *Cancer Cell*. 2003; 3:347-361.
6. Guo A et al. *Proc Natl Acad Sci U S A*. 2008; 105:692-697.
7. Spigel DR et al. *J Clin Oncol*. 2013; 31:4105-4014.

Shuying Liu

*Departments of Breast Medical Oncology and Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA*  
Email: [slju@mdanderson.org](mailto:slju@mdanderson.org)

Received: 3/29/15; Published: 3/30/15

# Methylated TRF2 associates with the nuclear matrix and serves as a potential biomarker for cellular senescence

Taylor R. H. Mitchell and Xu-Dong Zhu

Department of Biology, McMaster University, Hamilton, Ontario, Canada L8S 4K1

**Key words:** TRF2, arginine methylation, nuclear matrix, cellular senescence

**Received:** 11/29/13; **Accepted:** 4/03/14; **Published:** 4/05/14

**Correspondence to:** Xu-Dong Zhu, PhD; **E-mail:** [zhuxu@mcmaster.ca](mailto:zhuxu@mcmaster.ca)

**Copyright:** © Mitchell and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** Methylation of N-terminal arginines of the shelterin component TRF2 is important for cellular proliferation. While TRF2 is found at telomeres, where it plays an essential role in maintaining telomere integrity, little is known about the cellular localization of methylated TRF2. Here we report that the majority of methylated TRF2 is resistant to extraction by high salt buffer and DNase I treatment, indicating that methylated TRF2 is tightly associated with the nuclear matrix. We show that methylated TRF2 drastically alters its nuclear staining as normal human primary fibroblast cells approach and enter replicative senescence. This altered nuclear staining, which is found to be overwhelmingly associated with misshapen nuclei and abnormal nuclear matrix folds, can be suppressed by hTERT and it is barely detectable in transformed and cancer cell lines. We find that dysfunctional telomeres and DNA damage, both of which are potent inducers of cellular senescence, promote the altered nuclear staining of methylated TRF2, which is dependent upon the ATM-mediated DNA damage response. Collectively, these results suggest that the altered nuclear staining of methylated TRF2 may represent ATM-mediated nuclear structural alteration associated with cellular senescence. Our data further imply that methylated TRF2 can serve as a potential biomarker for cellular senescence.

## INTRODUCTION

Cellular senescence refers to a state of permanent arrest of cell proliferation and it is generally thought to be a response to potentially oncogenic stimuli [1, 2], which include telomere shortening, DNA damage either at telomeres or elsewhere in the genome, strong mitogenic stimuli and epigenomic perturbations. Since it was first described approximately fifty years ago [3], cellular senescence has emerged as an important mechanism linked to both tumor suppression and aging. In most human somatic cells, telomeres shorten with each round of DNA replication, in part because DNA polymerases are unable to fill in the gap generated from removal of the last RNA primer [4]. When telomeres become critically short, they activate a DNA damage response and trigger the induction of cellular senescence [5]. Telomerase, a reverse transcriptase that can replenish the repetitive telomeric DNA *de novo*, is able to circumvent telomere shortening and allows cells to gain

unlimited growth potential, a feature associated with cancer. Therefore cellular senescence is widely considered as a tumor suppressive mechanism. In addition, cellular senescence is also implicated in tissue repair and inflammation associated with aging and cancer progression [1, 2].

The nuclear matrix is a filamentous network of protein, DNA and RNA that is refractory to high salt extraction [6-8]. This structure serves as an architectural skeleton to the nucleus and provides support for chromatin organization and various nuclear functions including DNA replication, transcription and DNA repair [8]. Changes in the composition of the nuclear matrix have been observed in senescent cells [9]. Alterations in the nuclear matrix have been implicated in restricting cellular proliferation in mortal human fibroblasts [10]. Furthermore, nuclei with abnormal nuclear structure are found to accumulate in aged and prematurely senescent cells [11, 12].

Telomeres, heterochromatic structures found at the ends of linear eukaryotic chromosomes, function to protect natural chromosome ends from being recognized as damaged DNA. Mammalian telomeric DNA consists of tandem repeats of TTAGGG and is bound by a six-subunit protein complex [13, 14], referred to as shelterin or telosome, which is composed of TRF1, TRF2, TIN2, POT1, TPP1 and hRap1. Telomeres are attached to the nuclear matrix [15, 16] and components of the shelterin complex have been reported to be associated with the nuclear envelope as well as the nuclear matrix [10, 17-19]. TIN2L, an isoform of TIN2, has been suggested to mediate the interaction between telomeres and the nuclear matrix [17] whereas hRap1 has been implicated in tethering telomeres to the nuclear envelope [18]. TRF2, a shelterin protein that interacts directly with both TIN2 and hRap1 [20-23], has also been implicated in mediating the interaction between telomeres and the nuclear matrix [19]. However, little is known about the role of post-translational modification in mediating the interaction between shelterin proteins and the nuclear matrix.

TRF2, a subunit of the shelterin complex, binds to duplex telomeric DNA [24, 25] and plays a crucial role in telomere protection [26]. TRF2 contains an N-terminal basic domain rich in glycines and arginines (the GAR domain), a central TRFH dimerization domain, a flexible linker region and a C-terminal Myb-like DNA binding domain [24, 25]. It has been shown that loss of TRF2 from telomeres through either TRF2 knockout or overexpression of a dominant-negative allele of TRF2 promotes the formation of telomere end-to-end fusions [26, 27]. On the other hand, overexpression of TRF2 lacking the basic/GAR domain induces telomere rapid deletion [28] whereas overexpression of TRF2 carrying amino acid substitutions of arginines to lysines in the basic/GAR domain promotes the formation of fragile telomeres [29]. In all aforementioned cases, the formation of dysfunctional telomeres resulting from disruption of TRF2 function results in the induction of cellular senescence [26-29].

TRF2 undergoes extensive post-translational modification [30], which in turn regulates its stability, DNA binding and cellular localization. Ubiquitylation of TRF2 by Siah1, an E3 ligase, promotes TRF2 degradation and replicative senescence of human primary fibroblasts [31]. Acetylation, SUMOylation and poly(ADP-ribosyl)ation have been implicated in modulating TRF2 binding to telomeric DNA [32-35]. TRF2 is phosphorylated in response to DNA damage and this phosphorylation has been implicated in DNA double strand break repair [36, 37]. Arginines in the N-

terminal basic/GAR domain of TRF2 are methylated by protein arginine methyltransferase 1 (PRMT1) [29]. Loss of arginine methylation in TRF2 induces DNA-damage response foci at telomeres and triggers cellular senescence [29]. Arginine methylation is also implicated in negatively regulating the amount of telomere-bound TRF2 [29], raising a question as to whether methylated TRF2 is associated with telomeres *in vivo*.

Here we report that the majority of methylated TRF2 is not released by the treatment with high salt buffer and DNase I digestion and that it co-fractionates with lamin A, a component of the nuclear matrix [38, 39], suggesting that methylated TRF2 interacts with the nuclear matrix. We find that methylated TRF2 is largely not localized at telomeres, indicating that association of TRF2 with the nuclear matrix is likely independent of telomeres. We demonstrate that methylated TRF2 dramatically changes its nuclear staining as normal human primary fibroblast cells approach and enter replicative senescence. We find that the altered nuclear staining of methylated TRF2 is predominantly associated with misshapen nuclei and abnormal nuclear matrix folds. Introduction of hTERT into human primary fibroblast cells suppresses the altered nuclear staining of methylated TRF2, suggesting that progressive telomere shortening may contribute to the altered staining of methylated TRF2 in normal primary fibroblasts. In addition to telomere shortening, dysfunctional telomeres and ionizing radiation-induced DNA damage, both of which are potent inducers of cellular senescence, also promote the altered nuclear staining of methylated TRF2. Furthermore, we show that the lack or inhibition of ATM (ataxia telangiectasia mutated), a master regulator of the DNA damage response [40, 41], blocks the formation of ionizing irradiation (IR)-induced altered nuclear staining of methylated TRF2, indicating that the formation of the altered nuclear staining of methylated TRF2 is mediated by the ATM-dependent DNA damage response. Taken together, our results reveal that methylated TRF2-associated nuclear matrix undergoes an ATM-mediated structural alteration during the process of cellular senescence. Our data further imply that methylated TRF2 may serve as a potential biomarker for cellular senescence.

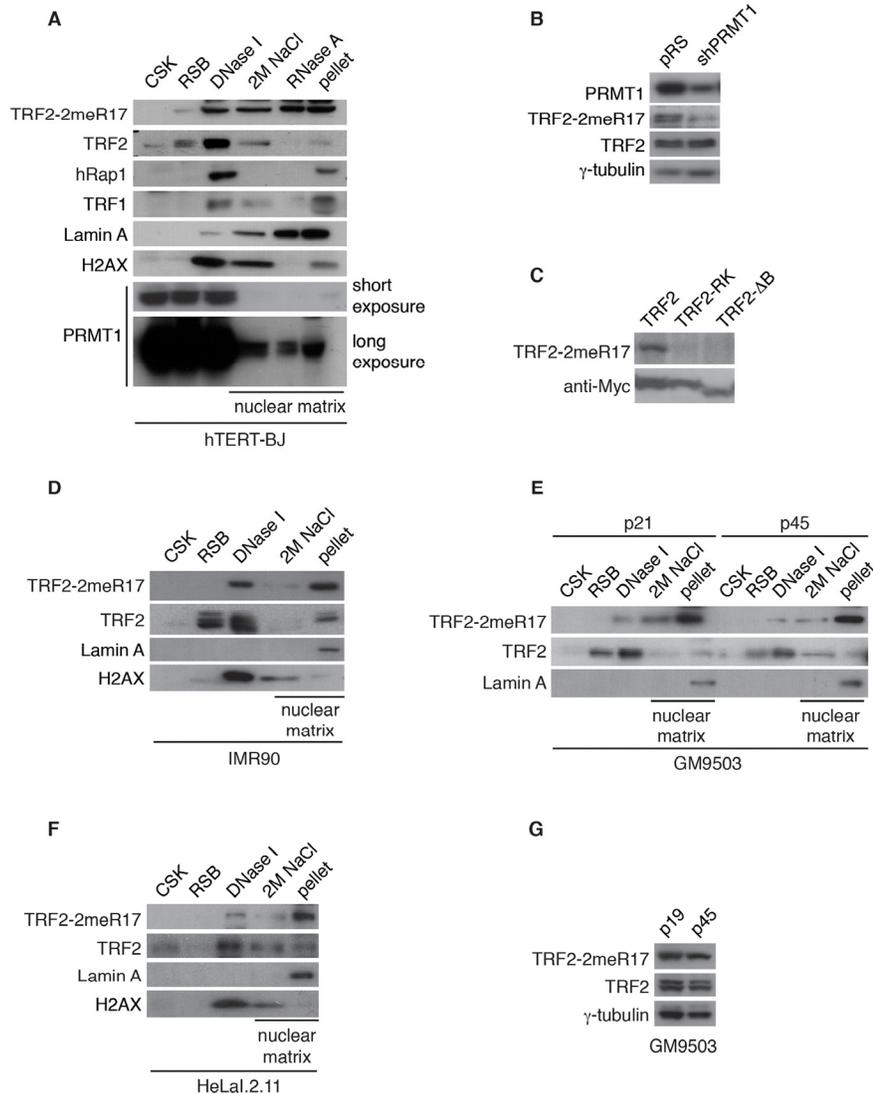
## RESULTS

### Methylated TRF2 is associated with nuclear matrix

We have previously reported that PRMT1 methylates arginines in the N-terminal GAR domain of TRF2 and that this arginine methylation negatively regulates TRF2

association with telomere chromatin [29], suggesting that methylated TRF2 is not associated with telomeres *in vivo*. To investigate the nuclear compartmentalization of methylated TRF2, we first subjected hTERT-immortalized BJ (hTERT-BJ) cells to analysis of sequential extraction of the nuclear matrix, which began with the treatment of cells with the CSK buffer to remove the majority of soluble proteins. The treatment of the RSB-magik buffer further removed the cytoskeleton,

leaving behind the nuclei and their attached filaments. Digestion of the nuclei with DNase I released chromatin bound proteins and the DNase I-resistant pellet was then further fractionated to release the outer nuclear matrix components by 2M NaCl. The treatment with RNase A disassembled ribonucleo-proteins and the final insoluble pellet contained the core nucleofilament proteins. The DNaseI-resistant fractions including the pellet are referred to as the nuclear matrix-associated fractions.



**Figure 1.** Methylated TRF2 is associated with the nuclear matrix. (A) Sequential extraction of the nuclear matrix from hTERT-BJ cells. Immunoblotting was performed with anti-TRF2-2meR17, anti-TRF2, anti-hRap1, anti-TRF1, anti-Lamin A, anti-H2AX, or anti-PRMT1 antibody. (B) Western analysis of 293T cells expressing shPRMT1 or the vector alone. Immunoblotting was performed with anti-PRMT1, anti-TRF2-2meR17 or anti-TRF2 antibody. The  $\gamma$ -tubulin blot was used a loading control. (C) Western analysis of 293T cells overexpressing Myc-tagged wild type TRF2, TRF2 carrying amino acid substitutions of arginines to lysines (TRF2-RK) or TRF2 lacking the N-terminal GAR/basic domain (TRF2- $\Delta$ B). Immunoblotting was carried out with anti-TRF2-2meR17 or anti-Myc antibody. (D) Sequential extraction of the nuclear matrix from IMR90 cells. Immunoblotting was performed with anti-TRF2-2meR17, anti-TRF2, anti-Lamin A or anti-H2AX antibody. (E) Sequential extraction of the nuclear matrix from GM9503 cells. Immunoblotting was performed with anti-TRF2-2meR17, anti-TRF2 or anti-Lamin A antibody. (F) Sequential extraction of the nuclear matrix from HeLa.2.11 cells. Immunoblotting was performed with anti-TRF2-2meR17, anti-TRF2, anti-Lamin A or anti-H2AX antibody. (G) Western analysis of early and late passage GM9503 cells. Immunoblotting was performed with anti-TRF2-2meR17 and anti-TRF2 antibody. The  $\gamma$ -tubulin blot was used as a loading control.

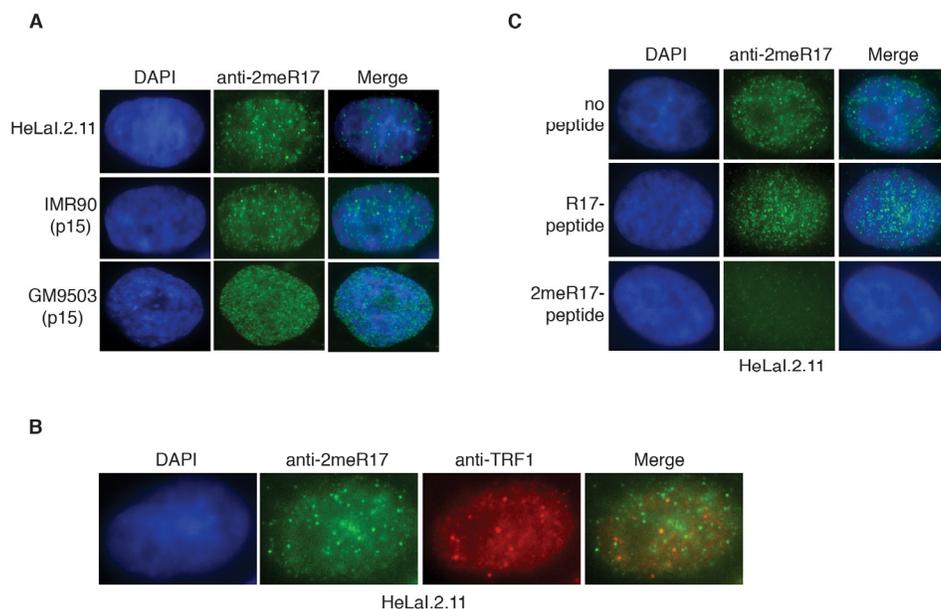
Examination of cell fractionations of hTERT-BJ with anti-Lamin A antibody revealed that Lamin A, a nuclear matrix-associated protein [38, 39], was predominantly found in the final pellet as well as fractions treated with 2M NaCl and RNase A (Fig. 1A), in agreement with these DNase I-resistant fractions being nuclear matrix. On the other hand, we found that the majority of chromatin-bound histone H2AX protein was released by the DNase I digestion (Fig. 1A), suggesting that the sequential cell fractionation protocol was working as expected. A very small amount of PRMT1 was detected in fractions associated with the nuclear matrix, consistent with previous findings [42]. In addition, a small amount of shelterin proteins including TRF1, TRF2 and hRap1 was also found in fractions associated with nuclear matrix (Fig. 1A), indicative of the nuclear matrix association of the shelterin proteins.

We have previously raised an antibody specifically against TRF2 methylated at R17 (anti-2meR17) [29], which specifically recognized methylated TRF2 (Fig. 1B) but not TRF2 carrying amino acid substitutions of arginines to lysines in its N-terminal domain (TRF2-RK) or lacking its N-terminal domain (TRF2-ΔB) (Fig. 1C). Using anti-2meR17 antibody, we found that although some methylated TRF2 was released by DNase I digestion, the majority of methylated TRF2 was recovered in nuclear matrix-associated fractions

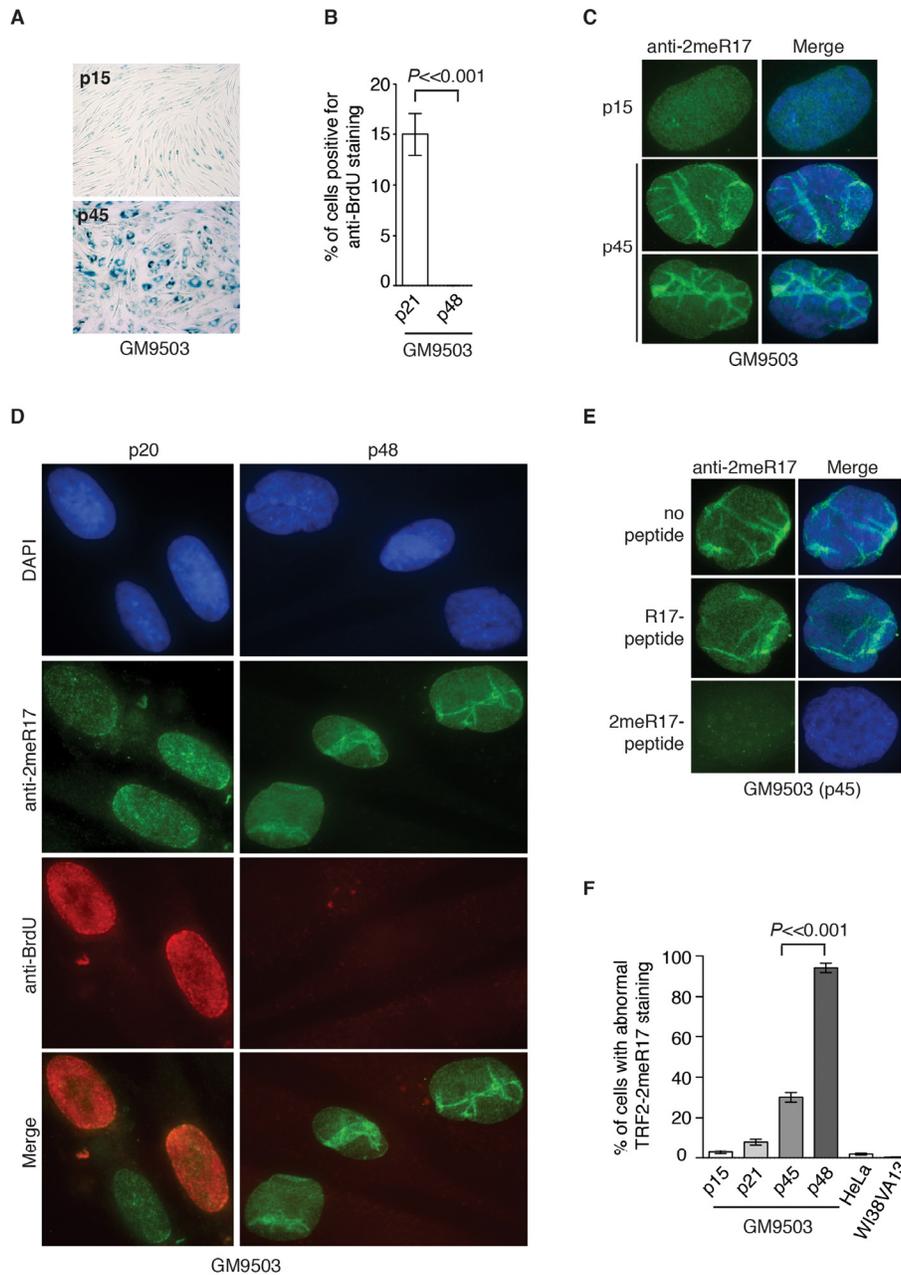
(Fig. 1A). The association of methylated TRF2 with the nuclear matrix was also observed in two other primary fibroblasts IMR90 and GM9503 cells as well as cancer cell line HeLaI.211 (Fig. 1D-1F). Taken together, these results suggest that methylated TRF2 preferentially associates with the nuclear matrix.

### Methylated TRF2 exhibits nuclear staining that is predominantly not associated with human telomeres

To further investigate the nuclear localization of methylated TRF2, we performed indirect immunofluorescence with anti-2meR17. We found that methylated TRF2 exhibited nuclear staining in both human primary (IMR90, GM9503) and cancer (HeLaI.2.11) cells (Fig. 2A). Analysis of dual indirect immunofluorescence with anti-2meR17 in conjunction with antibody against TRF1 [43], a marker for interphase telomeres, revealed that although there appeared to be some overlap between anti-2meR17 staining and anti-TRF1 staining (Fig. 2B), the majority of methylated TRF2 was not found to localize at telomeres (Fig. 2B). To investigate whether the observed anti-2meR17 staining might be due to any non-specific binding, we performed indirect immunofluorescence with anti-2meR17 antibody in the presence of TRF2 peptide containing either unmodified R17 or dimethylated R17.



**Figure 2.** Methylated TRF2 exhibits nuclear staining largely free of human telomeres. (A) Analysis of indirect immunofluorescence of three different cell lines with anti-TRF2-2meR17 antibody. Cell nuclei were stained with DAPI in blue. (B) Analysis of dual indirect immunofluorescence with anti-TRF2-2meR17 (green) in conjunction with anti-TRF1 antibody (red). HeLaI.2.11 cell nuclei were stained in blue. (C) Analysis of indirect immunofluorescence with anti-TRF2-2meR17 in conjunction with 100 ng of TRF2 peptide containing either modified or unmodified arginine 17. HeLaI.2.11 cell nuclei were stained with DAPI in blue.



**Figure 3.** Methyated TRF2 exhibits an altered nuclear staining associated with induction of replicative senescence in normal human primary fibroblast GM9503 cells. **(A)** Senescence-associated  $\beta$ -galactosidase assays for GM9503 cells at either p15 or p45. **(B)** Quantification of percentage of young and senescent GM9503 cells with BrdU incorporation. A total of 300 cells in triplicate were scored for either early passage (p21) or senescent GM9503 (p48) cells. Standard deviations from three independent experiments are indicated. **(C)** Analysis of indirect immunofluorescence with anti-TRF2-2meR17 antibody in GM9503 cells at either p15 or p45. Cell nuclei were stained with DAPI in blue. **(D)** Analysis of dual indirect immunofluorescence with anti-TRF2-2meR17 antibody (green) in conjunction with anti-BrdU antibody (red). The early passage (p20) and senescent GM9503 (p48) cells were incubated for six hours in growth media containing 10  $\mu$ M BrdU prior to being processed for immunofluorescence. Cell nuclei were stained with DAPI in blue. **(E)** Analysis of indirect immunofluorescence with anti-TRF2-2meR17 in conjunction with 100 ng of TRF2 peptide containing either modified or unmodified arginine 17. Cell nuclei were stained with DAPI in blue. **(F)** Quantification of percentage of cells with altered nuclear staining of methylated TRF2. At least 900 cells in triplicate were scored in blind for each transformed cell line or each normal primary fibroblast cell line at a given passage as indicated. Both HeLa and WI38VA13 are transformed cell lines. Standard deviations from three independent experiments are indicated.

We found that while TRF2 peptide containing unmodified R17 had little effect on the nuclear staining of methylated TRF2 in these cell lines (Fig. 2C and data not shown), TRF2 peptide containing dimethylated R17 abrogated the nuclear staining of methylated TRF2 (Fig. 2C and data not shown). These results suggest that the observed anti-2meR17 staining is unlikely due to non-specific binding. Taken together, these results suggest that methylated TRF2 localizes in nuclear domains largely free of human telomeres. These results are in agreement with our previous report that arginine methylation negatively regulates TRF2 association with telomere chromatin [29].

### **Replicative senescence induces altered nuclear staining of methylated TRF2**

We have shown that methylated TRF2 is associated with the nuclear matrix, which is known to undergo alterations in the process of replicative senescence [44]. To investigate whether the association of methylated TRF2 with the nuclear matrix might be affected by cellular senescence, we performed the sequential extraction of nuclear matrix in both early and late passages of primary skin fibroblast GM9503 cells. GM9503 cells exhibited an accumulation of senescent cells at passage 45 (Fig. 3A). Analysis of cell fractionations with anti-2meR17 antibody revealed that the association of methylated TRF2 with the nuclear matrix in late passage GM9503 (p45) cells was indistinguishable from that in young GM9503 (p21) cells (Fig. 1E). We did not observe any significant change in the level of methylated TRF2 and the total TRF2 between early and late passages of GM9503 cells (Fig. 1G). Taken together, these results suggest that methylated TRF2 does not dissociate from the nuclear matrix during the process of cellular senescence.

To investigate whether the nuclear staining of methylated TRF2 might undergo any changes during the process of replicative senescence, we performed indirect immunofluorescence with anti-2meR17 antibody in GM9503 cells of various passages from young to senescent. At passage 45, the majority of GM9503 cells were found to be positive for senescence-associated  $\beta$ -galactosidase staining (Fig. 3A). After they reached passage 48, GM9503 cells appeared to have fully entered replicative senescence since they failed to gain one population doubling over a period of at least two weeks and they showed little incorporation of BrdU (Fig. 3B), a marker used to measure DNA synthesis in cycling cells.

We found that methylated TRF2 showed a rather homogenous nuclear staining in young GM9503 (p15

and p20) cells (Fig. 3C and 3D), however, its staining was drastically altered in late passage and senescent GM9503 (p45 and p48) cells, (Fig. 3C and 3D and Supplementary Fig. 1). To address a concern that this altered nuclear staining might be due to any non-specific binding of anti-2meR17 antibody, we performed indirect immunofluorescence with anti-2meR17 antibody in the presence of TRF2 peptide containing either unmodified R17 or dimethylated R17. We found that TRF2 peptide containing unmodified R17 had little effect on the altered nuclear staining of methylated TRF2 whereas TRF2 peptide containing dimethylated R17 abolished TRF2 staining in senescent cells (Fig. 3E), arguing against the possibility that the observed altered nuclear staining of methylated TRF2 is due to non-specific binding of the anti-2meR17 antibody.

While less than 3% of young GM9503 (p15) cells displayed the altered methylated TRF2 staining (Fig. 3F), about 30% of late passage GM9503 (p45) cells showed the altered methylated TRF2 staining (Fig. 3F). To investigate whether the altered nuclear staining of methylated TRF2 might occur in non-dividing cells, we cultured GM9503 cells in media containing BrdU. Analysis of dual indirect immunofluorescence with anti-2meR17 antibody in conjunction with anti-BrdU antibody revealed that the altered methylated TRF2 staining was always associated with nuclei lacking incorporation of BrdU (Fig. 3D). We found that in the senescent GM9503 (p48) cell culture with little BrdU incorporation (Fig. 3B), over 90% of nuclei exhibited the altered methylated TRF2 staining (Fig. 3F). Taken together, these results suggest that the altered methylated TRF2 staining is associated with replicative senescence in GM9503 cells.

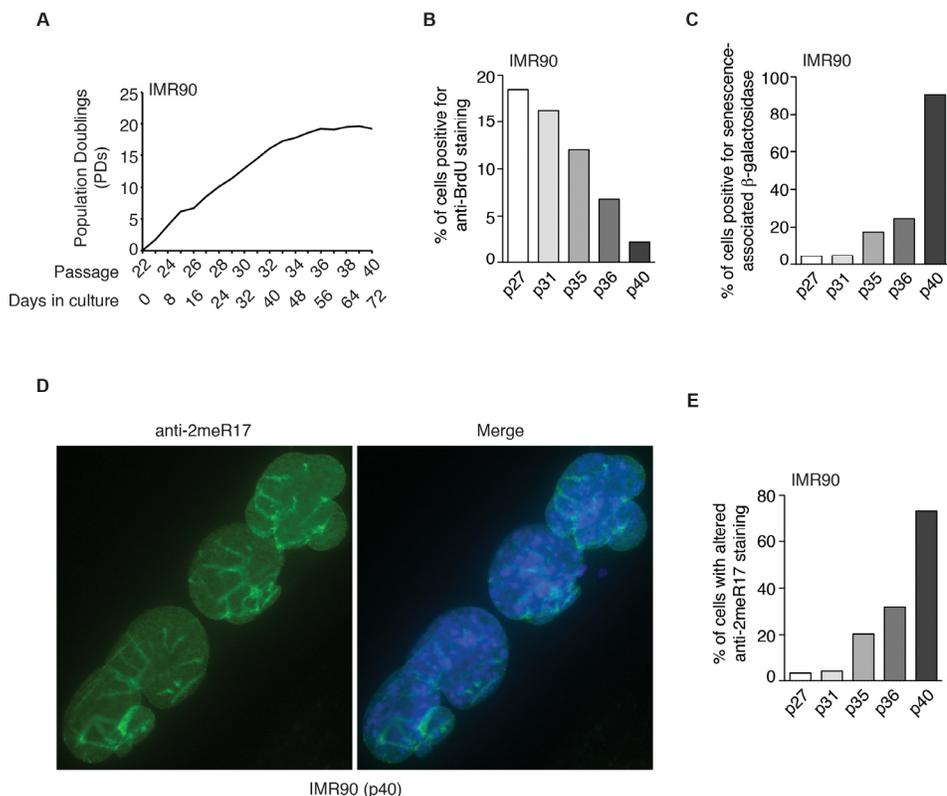
We also observed an induction of the altered nuclear staining of methylated TRF2 in several other normal human primary fibroblast cell lines (AG02661 and GM1706) as they aged in cultures (Supplementary Fig. 2). To further investigate whether the altered methylated TRF2 staining might be a general feature of replicative senescence, we passaged young IMR90 (p22) cells continuously every four days for 72 days until they reached passage 40 (Fig. 4A). In the last two weeks of culturing, IMR90 failed to gain one population doubling (Fig. 4A), suggesting that the IMR90 cell culture had entered replicative senescence. In agreement with this notion, we found that only 2% of IMR90 (p40) cells were stained positive for anti-BrdU antibody (Fig. 4B) and that about 90% of IMR90 cells were stained positive for senescence-associated  $\beta$ -galactosidase (Fig. 4C). Analysis of indirect immunofluorescence with anti-2meR17 antibody revealed that the number of

IMR90 cells with the altered methylated TRF2 staining increased substantially as IMR90 cells aged in culture (Fig. 4D and 4E). While less than 4% of IMR90 (p27) cells showed the altered nuclear staining of methylated TRF2, the altered methylated staining was observed in over 73% of senescent IMR90 (p40) cells (Fig. 4E). On the other hand, we did not detect any significant accumulation of the altered staining of methylated TRF2 in cancer and transformed cell lines including HeLa and WI38VA13/2RA (Fig. 3F). Collectively, these results suggest that the altered nuclear staining of methylated TRF2 is a general characteristic associated with replicative senescence.

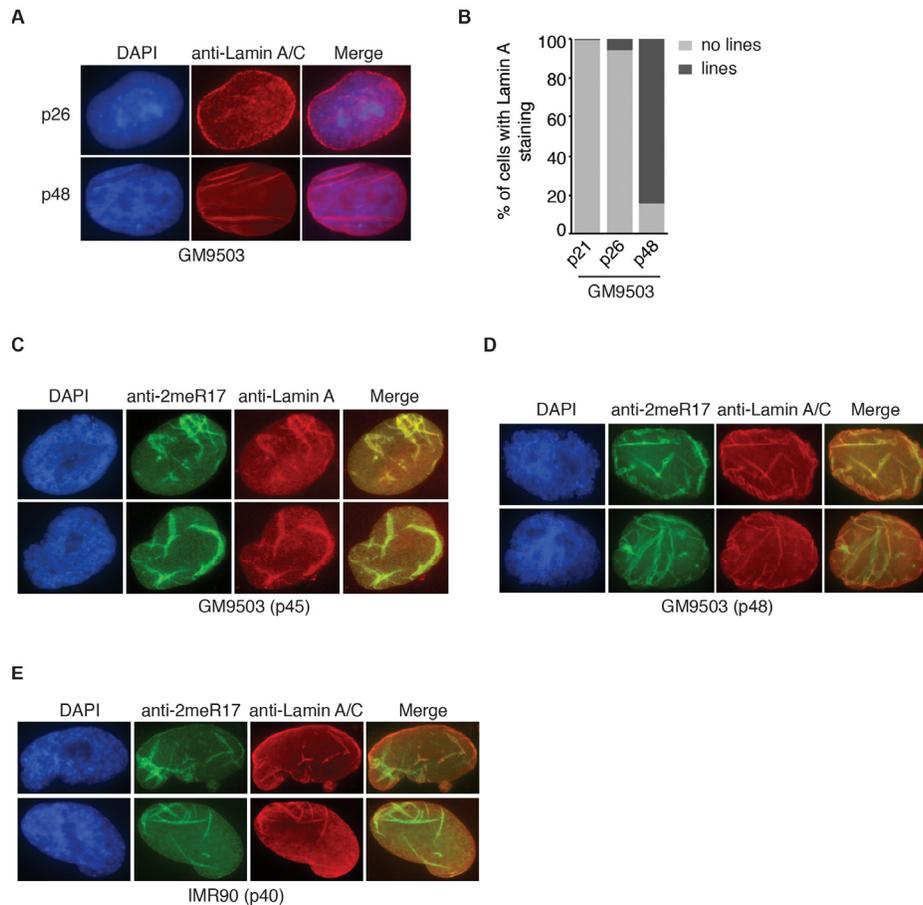
### Altered nuclear staining of methylated TRF2 is associated with the altered nuclear structure in senescent cells

It has been reported that senescent and aged cells are associated with distorted nuclear defects [11, 12, 44]. We examined 1182 senescent GM9503 (p48) cells and found that over 96% of the nuclei with altered methylated TRF2 staining contained blebbings and/or herniations, suggesting that the altered methylated TRF2 staining is overwhelmingly associated with misshapen nuclei. Lamin A has been suggested to serve as a marker for

distorted nuclear defects in aged cells [11, 12] and thus we also examined the relationship between the altered methylated TRF2 staining and Lamin A staining in senescent GM9503 (p48) cells. In young GM9503 (p21) cells, we observed strong Lamin A staining associated with the nuclear periphery as well as less intense Lamin A staining in the nucleoplasm (Fig. 5A). However, as cells aged in culture, we observed the appearance of line-shaped Lamin A staining in the nucleus. We found that in the senescent GM9503 (p48) culture, over 85% of cells exhibited line-shaped Lamin A/C staining (Fig. 5B). Analysis of dual indirect immunofluorescence with anti-2meR17 antibody in conjunction with either anti-Lamin A or anti-Lamin A/C antibody revealed that the line-shaped Lamin A staining overlapped well with the altered methylated TRF2 staining in GM9503 cells (Fig. 5C and 5D). The overlap between the altered nuclear staining of methylated TRF2 and the line-shaped Lamin A staining was also observed in senescent IMR90 (p40) cells (Fig. 5E). We examined 1073 nuclei of senescent GM9503 (p48) cells and found that 88% of nuclei had both the altered methylated TRF2 staining and the line-shaped Lamin A staining. Taken together, these results suggest that the altered nuclear staining of methylated TRF2 is predominantly associated with senescence-induced distorted nuclear structure.



**Figure 4.** The altered nuclear staining of methylated TRF2 is a general characteristics of replicative senescence. (A) Growth curve of IMR90 cells. IMR90 cells (p22) were passaged every four days continuously for 72 days. (B) Quantification of percentage of IMR90 cells with BrdU incorporation. A total number of 952, 1027, 1146, 1720 and 1418 cells were scored for passages 27, 31, 35, 36 and 40, respectively. (C) Quantification of percentage of IMR90 cells with senescence-associated  $\beta$ -galactosidase staining. A total number of 1887, 1606, 1286, 1712 and 1047 cells were scored for passages 27, 31, 35, 36 and 40, respectively. (D) Analysis of indirect immunofluorescence with anti-2meR17 antibody in senescent IMR90 (p40) cells. Cell nuclei were stained with DAPI in blue. (E) Quantification of percentage of IMR90 cells with altered nuclear staining of methylated TRF2. A total number of 952, 1027, 1146, 1720 and 1418 cells were scored for passages 27, 31, 35, 36 and 40, respectively.



**Figure 5.** The altered nuclear staining of methylated TRF2 is associated with altered nuclear structure in senescent cells. **(A)** Analysis of indirect immunofluorescence with anti-Lamin A/C antibody on young (p26) and senescent GM9503 (p48) cells. Cell nuclei were stained with DAPI in blue. **(B)** Quantification of percentage of GM9503 cells showing nuclear Lamin A staining that are either line-shaped or not associated with lines. A total number of 1014, 1050 and 1073 cells were scored for passages 21, 26 and 48, respectively. **(C)** Analysis of dual indirect immunofluorescence in late passage GM9503 (p45) with anti-TRF2-2meR17 antibody in conjunction with anti-Lamin A antibody. Cell nuclei were stained with DAPI in blue. **(D)** Analysis of dual indirect immunofluorescence in senescent GM9503 (p48) with anti-TRF2-2meR17 antibody in conjunction with anti-Lamin A/C antibody. Cell nuclei were stained with DAPI in blue. **(E)** Analysis of dual indirect immunofluorescence in senescent IMR90 (p40) with anti-TRF2-2meR17 antibody in conjunction with anti-Lamin A/C antibody. Cell nuclei were stained with DAPI in blue.

### Overexpression of hTERT prevents the formation of senescence-induced altered nuclear staining of methylated TRF2

Overexpression of hTERT can prevent replicative senescence, resulting in immortalization of normal primary fibroblasts [45, 46]. To investigate whether hTERT may repress the formation of the altered nuclear staining of methylated TRF2, we introduced hTERT into GM9503 cells of passage 38 (p38) and cultured hTERT-GM9503 cells in parallel with parental GM9503 cells (p38) until GM9503 cells approached

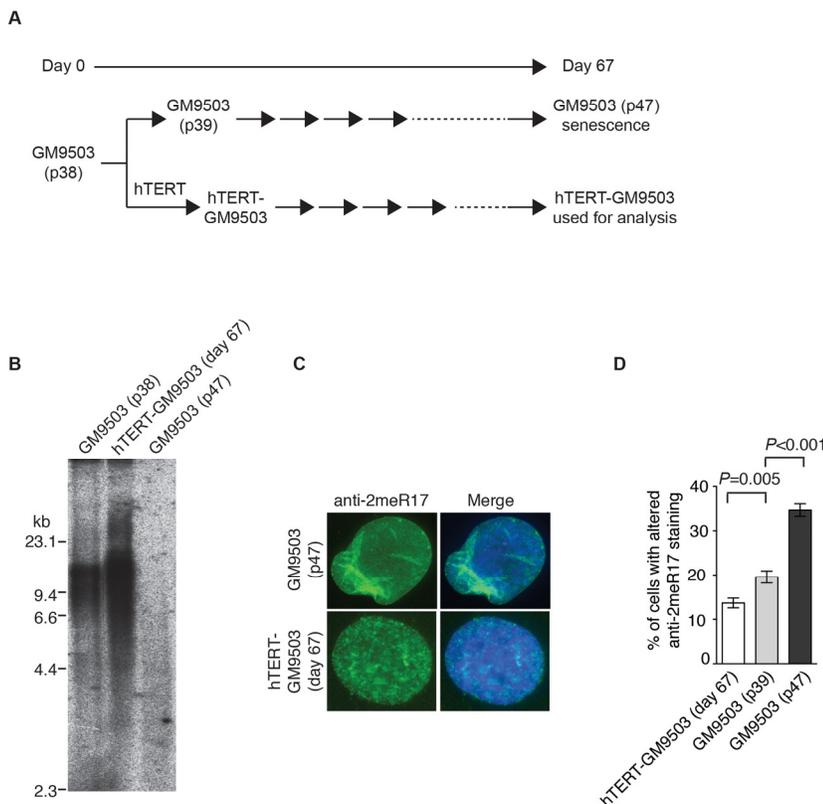
replicative senescence at p47 (Fig. 6A). Exogenous expression of hTERT was sufficient to immortalize GM9503 cells and prevented telomere shortening associated with replicative senescence (Fig. 6B). Analysis of indirect immunofluorescence with anti-2meR17 antibody revealed an induction of an altered nuclear staining of methylated TRF2 in GM9503 cells at passage 47 compared to those at passage 39 (Fig. 6C and 6D), consistent with our earlier finding. On the other hand, no increase in the altered nuclear staining of methylated TRF2 was observed in hTERT-immortalized GM9503 cells. Instead, we detected a 30% decrease

( $P=0.005$ ) in the number of hTERT-GM9503 cells with an altered nuclear staining of methylated TRF2 when compared to GM9503 cells of passage 39 (Fig. 6D). These results suggest that overexpression of hTERT suppresses the formation of replicative senescence-induced altered nuclear staining of methylated TRF2.

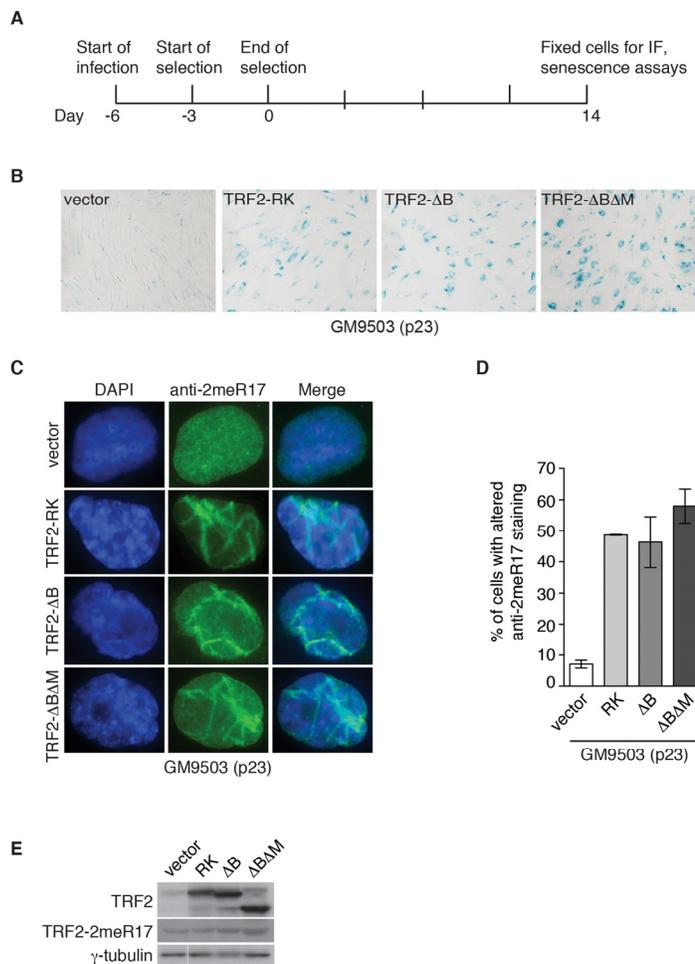
### Dysfunctional telomeres induce the formation of the altered nuclear staining of methylated TRF2

In addition to programmed telomere shortening, dysfunctional telomeres resulting from disruption of TRF2 function can also induce cellular senescence [26, 28, 29]. It has been well documented that overexpression of TRF2 lacking the N-terminal basic/GAR domain (TRF2- $\Delta$ B) promotes telomere rapid deletion whereas overexpression of TRF2 lacking both the N-terminal basic/GAR and C-terminal Myb-like DNA binding domains (TRF2- $\Delta$ B $\Delta$ M) induces the formation of telomere fusions, both of which result in the induction of cellular senescence [26, 28]. Overexpression of TRF2 carrying amino acid substitutions of arginines to lysines in its N-terminal

domain (TRF2-RK) has been shown to induce fragile telomeres, triggering cellular senescence [29]. To investigate whether the altered nuclear staining of methylated TRF2 might be associated with cellular senescence induced by dysfunctional telomeres, we infected GM9503 (p23) cells with retrovirus expressing TRF2-RK, TRF2- $\Delta$ B, TRF2- $\Delta$ B $\Delta$ M or the vector alone. Fourteen days post infection (Fig. 7A), we observed an induction of cellular senescence in GM9503 cells overexpressing either TRF2-RK, TRF2- $\Delta$ B or TRF2- $\Delta$ B $\Delta$ M but not in GM9503 cells expressing the vector alone (Fig. 7B), consistent with previous reports [26, 28, 29]. Analysis of indirect immunofluorescence revealed a drastic accumulation of the altered nuclear staining of methylated TRF2 in GM9503 cells overexpressing either TRF2-RK, TRF2- $\Delta$ B or TRF2- $\Delta$ B $\Delta$ M when compared to GM9503 cells expressing the vector alone (Fig. 7C and 7D). Overexpression of TRF2-RK, TRF2- $\Delta$ B or TRF2- $\Delta$ B $\Delta$ M had little effect on the level of endogenous methylated TRF2 (Fig. 7E). These results suggest that the altered nuclear staining of methylated TRF2 is associated with cellular senescence induced by dysfunctional telomeres.



**Figure 6.** Introduction of hTERT into normal primary fibroblast cells suppresses the formation of senescence-associated altered nuclear staining of methylated TRF2. **(A)** Schematic diagram of the experimental setup. At day 0, GM9503 cells were infected with retrovirus expressing hTERT, generating hTERT-GM9503 cells. Both GM9503 and hTERT-GM9503 cells were cultured continuously for 67 days. **(B)** Genomic blots of telomeric restriction fragments from GM9503 (p38), GM9503 (p47) and hTERT-GM9503 at day 67. About 3  $\mu$ g of RsaI/HinfI-digested genomic DNA from each sample was used for gel electrophoresis. The DNA molecular size markers are shown to the left of the blots. **(C)** Analysis of indirect immunofluorescence with anti-TRF2-2meR17 antibody. Cell nuclei of GM9503 and hTERT-GM9503 were stained with DAPI in blue. **(D)** Quantification of percentage of cells with altered nuclear staining of methylated TRF2. At least 900 cells in triplicate were scored in blind for each cell line as indicated. Standard deviations from three independent experiments are indicated.

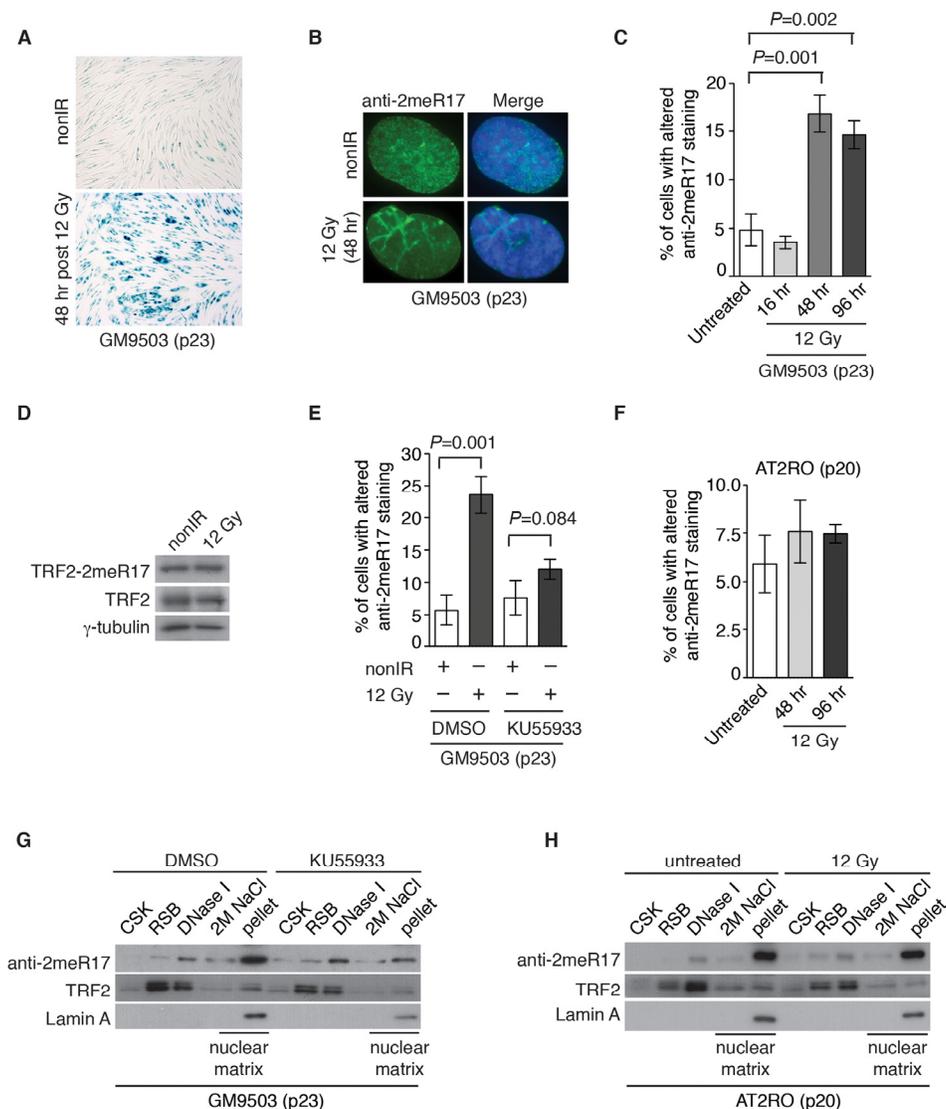


**Figure 7.** Dysfunctional telomeres induce altered nuclear staining of methylated TRF2. **(A)** Schematic diagram of experimental setup. GM9503 cells (p23) were infected with retrovirus expressing various TRF2 mutant alleles at day -6. After the three-day selection ended on day 0, the cells were cultured for 14 days and then subjected to analysis of immunofluorescence (IF) and cellular senescence. **(B)** Senescence-associated  $\beta$ -galactosidase assays of GM9503 cells overexpressing the vector alone or various TRF2 mutant alleles as indicated. **(C)** Indirect immunofluorescence with anti-TRF2-2meR17 antibody in fixed GM9503 cells overexpressing the vector alone or various TRF2 mutant alleles as indicated. Cell nuclei were stained with DAPI in blue. **(D)** Quantification of percentage of cells with altered nuclear staining of methylated TRF2. A total of 1000 cells in triplicate were scored in blind for each cell line. Standard deviations from three independent experiments are indicated. **(E)** Western analysis of GM9503 cells overexpressing various TRF2 mutant alleles as indicated. Immunoblotting was carried out with anti-TRF2, anti-TRF2-2meR17 or anti- $\gamma$ -tubulin antibody.

### DNA damage induces the formation of the altered nuclear staining of methylated TRF2 in an ATM-dependent manner

Stress induced premature senescence is known to be a DNA damage response [47] and therefore we asked whether the altered nuclear staining of methylated TRF2 might be associated with DNA damage-induced senescence. To address this question, GM9503 cells (p23) were either mock-treated or treated with 12 Gy ionizing irradiation (IR). We found that 48 h post IR, all GM9503 cells entered cellular senescence as evidenced by analysis of senescence-associated  $\beta$ -galactosidase assays (Fig. 8A). Analysis of indirect immunofluorescence with anti-2meR17 revealed an over 3-fold induction in the number of GM9503 cells with the altered staining of methylated TRF2 48 h post 12 Gy IR (Fig. 8B and 8C). We did not detect any significant change in the level of methylated TRF2 (Fig. 8D). Collectively, these results suggest that the altered nuclear staining of methylated TRF2 is associated with DNA damage-induced cellular senescence.

ATM is a master regulator of DNA damage response following ionizing radiation [40, 41] and therefore we also examined whether IR-induced altered nuclear staining of methylated TRF2 might be dependent upon the ATM-mediated DNA damage response. GM9503 cells (p23) were treated with DMSO or KU55933, a potent and specific ATM inhibitor [48], prior to 12 Gy IR. Analysis of indirect immunofluorescence with anti-2meR17 antibody revealed that treatment with KU55933 impaired the induction of IR-induced altered nuclear staining of methylated TRF2 (Fig. 8E). The impairment of IR-induced altered nuclear staining of methylated TRF2 was also observed in primary ataxia telangiectasia (AT) fibroblast cells (AT2RO) lacking functional ATM (Fig. 8F). Treatment with KU55933 or loss of ATM did not affect the association of methylated TRF2 with the nuclear matrix (Fig. 8G and H). Taken together, these results suggest that ATM is important for the formation of IR-induced altered nuclear staining of methylated TRF2.



**Figure 8.** Ionizing radiation induces altered nuclear staining of methylated TRF2 in a ATM-dependent manner. **(A)** Ionizing radiation induces cellular senescence in GM9503 cells. GM9503 (p23) cells were treated with 12 Gy IR. Senescence-associated  $\beta$ -galactosidase assays were performed 48 h post IR. **(B)** Indirect immunofluorescence with anti-TRF2-2meR17 antibody in mock- or IR-treated GM9503 cells (p23). Cell nuclei were stained with DAPI in blue. **(C)** Quantification of percentage of cells with altered nuclear staining of methylated TRF2. A total of 1000 cells in triplicate were scored in blind for untreated or IR-treated cells fixed at various time points post IR as indicated. Standard deviations from three independent experiments are indicated. **(D)** Western analysis of GM9503 cells (p23) that were either mock- or IR-treated. Immunoblotting was performed with anti-TRF2-2meR17 or anti-TRF2 antibody. The  $\gamma$ -tubulin blot was used as a loading control. **(E)** ATM inhibition abrogates IR-induced altered nuclear staining of methylated TRF2. GM9503 cells (p23) were treated with DMSO or KU55933 prior to 12 Gy IR treatment. Forty-eight hours post IR, cells were processed for indirect immunofluorescence with anti-TRF2-2meR17 antibody. Quantification of percentage of cells with altered nuclear staining of methylated TRF2. A total of 1000 cells in triplicate were scored in blind. Standard deviations from three independent experiments are indicated. **(F)** Little IR-induced altered nuclear staining of methylated TRF2 is observed in AT2RO cells lacking functional ATM. Quantification of percentage of cells with altered nuclear staining of methylated TRF2. A total of 1000 cells in triplicate were scored in blind. Standard deviations from three independent experiments are indicated. **(G)** Sequential extraction of the nuclear matrix from GM9503 cells treated with either DMSO or KU55933. Immunoblotting was performed with anti-TRF2-2meR17, anti-TRF2 or anti-Lamin A antibody. **(H)** Sequential extraction of the nuclear matrix from AT2RO cells that were either untreated or treated with 12 Gy IR. Immunoblotting was performed with anti-TRF2-2meR17, anti-TRF2 or anti-Lamin A antibody.

## DISCUSSION

In this report, we have shown that methylated TRF2 is associated with the nuclear matrix and that this localization is largely free of human telomeres. In addition we have uncovered that the nuclear staining of methylated TRF2 is drastically altered upon induction of cellular senescence and that this altered staining is predominantly associated with misshapen nuclei and abnormal nuclear matrix folds. Our findings suggest that methylated TRF2 can serve as a potential biomarker for cellular senescence.

Through sequential extraction of the nuclear matrix, we have shown that the nuclear matrix contains a small amount of endogenous TRF1, TRF2 and hRap1, suggesting that components of the shelterin complex are associated with the nuclear matrix, in agreement with previous reports [10, 17]. We have previously estimated that about 1-5% of endogenous TRF2 is methylated by PRMT1 [29]. Our finding that the majority of methylated TRF2 is associated with the nuclear matrix suggests that arginine methylation plays an important role in regulating TRF2 interaction with the nuclear matrix.

We have previously reported that overexpression of TRF2-RK carrying amino acid substitutions of arginines to lysines (TRF2-RK), which cannot undergo PRMT1-dependent arginine methylation [29], promotes the formation of telomeres with multiple telomere signals [29], also known as fragile telomeres [49, 50]. Our finding that methylated TRF2 is associated with the nuclear matrix raises a possibility that TRF2-RK-induced fragile telomeres might arise in part from its inability to interact with the nuclear matrix. As fragile telomeres are thought to result from a defect in telomere replication [49, 50], it would be of interest to know if the arginine methylation-dependent TRF2 interaction with the nuclear matrix might play a role in supporting efficient telomere replication.

Although we have shown that methylated TRF2 co-fractionates with Lamin A, we have not been able to detect a direct interaction between TRF2 and Lamin A (T. Mitchell and X.-D. Zhu, unpublished data), suggesting that the association of methylated TRF2 with the nuclear matrix may not be mediated by Lamin A. While human telomeres are attached to the nuclear matrix [16], we have shown through indirect immunofluorescence that the punctate nuclear staining of methylated TRF2 is largely free of telomere signals, suggesting that methylated TRF2 interaction with the nuclear matrix is unlikely mediated through telomeric DNA. How methylated TRF2 interacts with the

nuclear matrix remains unknown, but our finding nevertheless suggests that post-translational modification plays a role in regulating TRF2 interaction with the nuclear matrix. Future studies would be needed to investigate the nature of the punctate nuclear staining of methylated TRF2 observed in young primary cells and cancer cells.

We have shown that methylated TRF2 exhibits an altered nuclear staining upon induction of replicative senescence in normal human primary fibroblasts and that this altered staining can be suppressed by overexpression of hTERT, suggesting that telomere erosion may be a trigger for the altered nuclear staining of methylated TRF2. It has been suggested that ionizing radiation (IR)- and oncogene-induced cellular senescence is caused by irreparable DNA damage at telomeres [51, 52]. Our observation that the altered nuclear staining of methylated TRF2 is associated with IR- and dysfunctional telomeres-induced cellular senescence further supports the notion that telomere damage resulting from either genotoxic insults or programmed telomere erosion is a major inducer of the altered nuclear staining of methylated TRF2.

Using two independent antibodies against Lamin A, we have observed that senescent fibroblast cells (GM9503 and IMR90) exhibit abnormal line-shaped Lamin A staining, similar to the Lamin A staining previously reported for senescent human mesenchymal stem cells [11, 44]. Our finding that the altered nuclear staining of methylated TRF2 overlaps well with the abnormal line-shaped Lamin A staining suggests that the altered methylated TRF2 staining is associated with distorted nuclear structures in senescent cells. Our finding further implies that methylated TRF2 may serve as a potential biomarker for senescence-associated altered nuclear matrix.

It has been well established that the ATM-dependent DNA damage response is needed for initiating and maintaining cellular senescence [53-56]. Consistent with previous reports [53-56], we have observed that primary AT2RO fibroblasts lacking functional ATM fail to undergo cell cycle arrest and start to die within days after ionizing radiation (T.R.H Mitchell and X.-D. Zhu, unpublished data). We have demonstrated that loss or inhibition of ATM abrogates IR-induced altered nuclear staining of methylated TRF2, indicating that the ATM-dependent DNA damage response is needed for the formation of the altered nuclear staining of methylated TRF2. Our finding suggests that the methylated TRF2 may serve as a potential biomarker for aging- and stress-induced cellular senescence.

## METHODS

**DNA constructs and antibodies.** The retroviral construct expressing shRNA against PRMT1 (pRS-shPRMT1) has been previously described [29]. The retroviral construct expressing hTERT (pBabe-hTERT) was generously provided by Robert Weinberg, Whitehead Institute for Biomedical Research.

Antibody specifically raised against TRF2 methylated at R17 (2meR17) has been previously described [29]. Antibodies against TRF1, TRF2 and hRap1 were kind gifts from Titia de Lange, Rockefeller University. Other antibodies used include Lamin A (Millipore), Lamin A/C (Cell Signaling), PRMT1 (Millipore) and H2A.X (Upstate).

**Cell culture and retroviral infection.** Cells were grown in DMEM medium with 10% fetal bovine serum (FBS) for HeLaI.2.11, 293T, WI38VA13/2RA, hTERT-BJ and Phoenix cells supplemented with non-essential amino acids, L-glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin. Supplementary DMEM medium plus 15% FBS was used to culture normal primary human fibroblasts (IMR90, GM9503, AG02261 and GM1706) (Coriell) and ATM-deficient primary fibroblast AT2RO (a kind gift from Jan Hoeijmakers). For inhibition of ATM, cells were treated with KU55933 (20  $\mu$ M, Sigma) for 3 h before 12 Gy ionizing irradiation treatment. Ionizing irradiation was delivered from a Cs-137 source at McMaster University (Gammacell 1000).

Retroviral gene delivery was carried out essentially as described [29, 57]. Phoenix amphotropic retroviral packaging cells were transfected with the desired DNA constructs using Lipofectamine 2000 (Invitrogen) according to the manufacturer. At 36, 48, 60, 72, and 84 h post-transfection, the virus-containing medium was collected and used to infect cells in the presence of polybrene (4  $\mu$ g/ml). Twelve hours after the last infection, puromycin (2  $\mu$ g/ml) was added to the medium, and the cells were maintained in the selection media for the entirety of the experiments.

**Sequential extraction of nuclear matrix.** Extraction of nuclear matrix components was conducted essentially as described [58, 59]. Briefly, PBS-washed cells were resuspended in 5X pellet volume cytoskeleton (CSK) buffer (10 mM Pipes pH 6.8, 100 mM NaCl, 300 mM sucrose, 3 mM MgCl<sub>2</sub>, 1 mM EGTA, 2 mM vanadyl ribonucleoside complex, 1.2 mM phenylmethylsulfonyl fluoride, 1 mM DTT, 1  $\mu$ g/ml aprotinin, 1  $\mu$ g/ml leupeptin, 10  $\mu$ g/ml pepstatin and 0.5% Triton X-100). Following centrifugation at 1000g for 5 min, the cytoskeleton framework was further extracted by

incubating the pellet in RSB-magik buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 2 mM vanadyl ribonucleoside complex, 1.2 mM phenylmethylsulfonyl fluoride, 1 mM DTT, 1  $\mu$ g/ml aprotinin, 1  $\mu$ g/ml leupeptin, 10  $\mu$ g/ml pepstatin, 1% Tween 20 and 0.5% sodium deoxycholate) for 5 min. Upon centrifugation, the pellet was treated with 30-50 U of RNase-free DNaseI (Fermentas) per 10<sup>6</sup> cells in digestion buffer (10 mM Pipes pH6.8, 50 mM NaCl, 300 mM sucrose, 3 mM MgCl<sub>2</sub>, 1 mM EGTA, 2 mM vanadyl ribonucleoside complex, 1.2 mM phenylmethylsulfonyl fluoride, 1 mM DTT, 1  $\mu$ g/ml aprotinin, 1  $\mu$ g/ml leupeptin, 10  $\mu$ g/ml pepstatin and 0.5% Triton X-100) for 1 h at room temperature. Chromatin was then removed by elution with 0.25 M ammonium sulfate, leaving a complete nuclear matrix-intermediate filament scaffold containing nuclear ribonuclear-protein complexes [58]. The complete nuclear matrix was further extracted with 2M NaCl to release the outer nuclear matrix proteins, and in some cases followed by digestion with DNase-free RNase A to remove the core filaments of the matrix. All incubation and centrifugations were performed at 4°C except where indicated.

**Immunofluorescence.** Immunofluorescence was performed as described [29, 60, 61]. All cell images were recorded on a Zeiss Axioplan 2 microscope with a Hammamatsu C4742-95 camera and processed using the Openlab software package.

For BrdU labeling, cells were seeded on and processed in 8-well chamber slides (Lab Tek). Two days later, cells were incubated for 6 hours in media containing 10  $\mu$ M BrdU (Sigma) and then fixed at room temperature (RT) for 7 min in PBS-buffered 3% paraformaldehyde and 2% sucrose. Following permeabilization at RT for 7 min in Triton X-100 buffer (0.5% Triton X-100, 20 mM Hepes-KOH, pH 7.9, 50 mM NaCl, 3 mM MgCl<sub>2</sub>, 300 mM sucrose), cells were incubated in 4M HCl for 10 min at RT. Cells were then processed for immunofluorescence as described [29, 60, 61] using anti-BrdU antibody (Novus Biologicals) and anti-2meR17 antibody.

**Telomere length analysis and senescence-associated  $\beta$ -galactosidase assays.** For telomere length analysis, genomic DNA isolated from cells was digested with *RsaI* and *HinfI* and loaded onto a 0.7% agarose gel in 0.5xTBE. Blotting for telomeric fragments was carried out as described [62, 63].

Senescence-associated (SA)  $\beta$ -galactosidase assays were carried out using the SA- $\beta$ -gal senescence kit (Cell Signaling) according to the manufacturer's instructions.

The cells were seeded two to four days prior to processing.

Statistical analysis. A student's two-tailed t test was used to derive all *P* values.

## ACKNOWLEDGEMENTS

We thank John R. Walker and Nicole Batenburg for their critical comments. This work was supported by Ontario Early Researcher Award program and grants from Canadian Institutes of Health Research to X.D.Z. T.R.H.M. was a holder of Ontario Graduate Scholarship.

## Conflict of interest statement

The authors of this manuscript have no conflict of interest to declare.

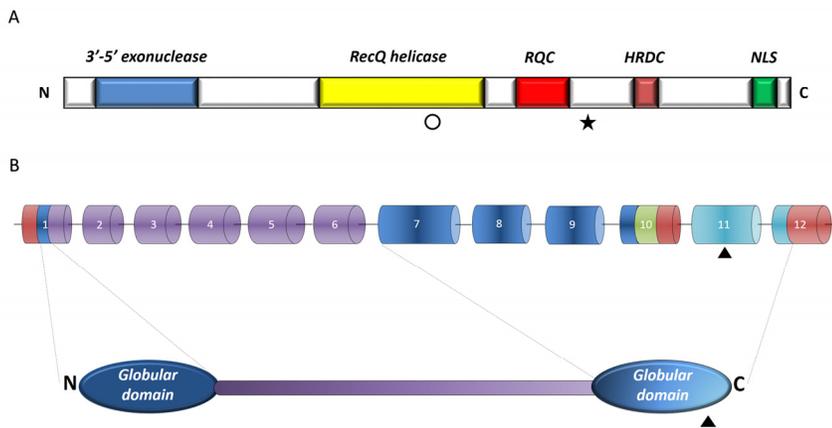
## REFERENCES

1. Campisi J. Aging, cellular senescence, and cancer. *Annu Rev Physiol.* 2013; 75:685-705.
2. Rodier F and Campisi J. Four faces of cellular senescence. *The Journal of cell biology.* 2011; 192:547-556.
3. Hayflick L. The limited in vitro lifetime of human diploid cell strains. *Exp Cell Res.* 1965; 37:614-636.
4. Levy MZ, Allsopp RC, Futcher AB, Greider CW and Harley CB. Telomere end-replication problem and cell aging. *J Mol Biol.* 1992; 225:951-960.
5. Reaper PM, di Fagagna F and Jackson SP. Activation of the DNA damage response by telomere attrition: a passage to cellular senescence. *Cell Cycle.* 2004; 3:543-546.
6. Berezney R, Mortillaro MJ, Ma H, Wei X and Samarabandu J. The nuclear matrix: a structural milieu for genomic function. *International review of cytology.* 1995; 162A:1-65.
7. Berezney R. The nuclear matrix: a heuristic model for investigating genomic organization and function in the cell nucleus. *Journal of cellular biochemistry.* 1991; 47:109-123.
8. Mehta IS, Figgitt M, Clements CS, Kill IR and Bridger JM. Alterations to nuclear architecture and genome behavior in senescent cells. *Annals of the New York Academy of Sciences.* 2007; 1100:250-263.
9. Dell'Orco RT and Whittle WL. Nuclear matrix composition and in vitro cellular senescence. *Experimental gerontology.* 1994; 29:139-149.
10. Okabe J, Eguchi A, Wadhwa R, Rakwal R, Tsukinoki R, Hayakawa T and Nakanishi M. Limited capacity of the nuclear matrix to bind telomere repeat binding factor TRF1 may restrict the proliferation of mortal human fibroblasts. *Human molecular genetics.* 2004; 13:285-293.
11. Raz V, Vermolen BJ, Garini Y, Onderwater JJ, Mommaas-Kienhuis MA, Koster AJ, Young IT, Tanke H and Dirks RW. The nuclear lamina promotes telomere aggregation and centromere peripheral localization during senescence of human mesenchymal stem cells. *Journal of cell science.* 2008; 121:4018-4028.
12. Scaffidi P and Misteli T. Lamin A-dependent nuclear defects in human aging. *Science.* 2006; 312:1059-1063.
13. Palm W and de Lange T. How shelterin protects Mammalian telomeres. *Annu Rev Genet.* 2008; 42:301-334.
14. Liu D, O'Connor MS, Qin J and Songyang Z. Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J Biol Chem.* 2004; 279:51338-51342.
15. Luderus ME, van Steensel B, Chong L, Sibon OC, Cremers FF and de Lange T. Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *J Cell Biol.* 1996; 135:867-881.
16. de Lange T. Human telomeres are attached to the nuclear matrix. *Embo J.* 1992; 11:717-724.
17. Kaminker PG, Kim SH, Desprez PY and Campisi J. A novel form of the telomere-associated protein TIN2 localizes to the nuclear matrix. *Cell Cycle.* 2009; 8:931-939.
18. Crabbe L, Cesare AJ, Kasuboski JM, Fitzpatrick JA and Karlseder J. Human telomeres are tethered to the nuclear envelope during postmitotic nuclear assembly. *Cell Rep.* 2012; 2:1521-1529.
19. Voronin AP, Lobov IB, Gilson E and Podgornaya OI. A telomere-binding protein (TRF2/MTBP) from mouse nuclear matrix with motives of an intermediate filament-type rod domain. *J Anti Aging Med.* 2003; 6:205-218.
20. Li B, Oestreich S and de Lange T. Identification of human Rap1: implications for telomere evolution. *Cell.* 2000; 101:471-483.
21. Ye JZ, Donigian JR, van Overbeek M, Loayza D, Luo Y, Krutchinsky AN, Chait BT and de Lange T. TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *The Journal of biological chemistry.* 2004; 279:47264-47271.
22. Kim SH, Beausejour C, Davalos AR, Kaminker P, Heo SJ and Campisi J. TIN2 mediates functions of TRF2 at human telomeres. *The Journal of biological chemistry.* 2004; 279:43799-43804.
23. Houghtaling BR, Cuttonaro L, Chang W and Smith S. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr Biol.* 2004; 14:1621-1631.
24. Broccoli D, Smogorzewska A, Chong L and de Lange T. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature genetics.* 1997; 17:231-235.
25. Bilaud T, Brun C, Ancelin K, Koering CE, Laroche T and Gilson E. Telomeric localization of TRF2, a novel human telobox protein. *Nat Genet.* 1997; 17:236-239.
26. van Steensel B, Smogorzewska A and de Lange T. TRF2 protects human telomeres from end-to-end fusions. *Cell.* 1998; 92:401-413.
27. Celli GB and de Lange T. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nat Cell Biol.* 2005; 7:712-718.
28. Wang RC, Smogorzewska A and de Lange T. Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell.* 2004; 119:355-368.
29. Mitchell TR, Glenfield K, Jeyanthan K and Zhu XD. Arginine methylation regulates telomere length and stability. *Molecular and cellular biology.* 2009; 29:4918-4934.
30. Walker JR and Zhu XD. Post-translational modification of TRF1 and TRF2 and their roles in telomere maintenance. *Mechanisms of ageing and development.* 2012; 133:421-434.

31. Fujita K, Horikawa I, Mondal AM, Jenkins LM, Appella E, Vojtesek B, Bourdon JC, Lane DP and Harris CC. Positive feedback between p53 and TRF2 during telomere-damage signalling and cellular senescence. *Nature cell biology*. 2010; 12:1205-1212.
32. Her YR and Chung IK. p300-mediated acetylation of TRF2 is required for maintaining functional telomeres. *Nucleic acids research*. 2013; 41:2267-2283.
33. Potts PR and Yu H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nature structural & molecular biology*. 2007; 14:581-590.
34. Dantzer F, Giraud-Panis MJ, Jaco I, Ame JC, Schultz I, Blasco M, Koering CE, Gilson E, Menissier-de Murcia J, de Murcia G and Schreiber V. Functional interaction between poly(ADP-Ribose) polymerase 2 (PARP-2) and TRF2: PARP activity negatively regulates TRF2. *Molecular and cellular biology*. 2004; 24:1595-1607.
35. Gomez M, Wu J, Schreiber V, Dunlap J, Dantzer F, Wang Y and Liu Y. PARP1 is a TRF2-associated poly(ADP-ribose)polymerase and protects eroded telomeres. *Molecular biology of the cell*. 2006; 17:1686-1696.
36. Tanaka H, Mendonca MS, Bradshaw PS, Hoelz DJ, Malkas LH, Meyn MS and Gilley D. DNA damage-induced phosphorylation of the human telomere-associated protein TRF2. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:15539-15544.
37. Huda N, Tanaka H, Mendonca MS and Gilley D. DNA damage-induced phosphorylation of TRF2 is required for the fast pathway of DNA double-strand break repair. *Molecular and cellular biology*. 2009; 29:3597-3604.
38. Hozak P, Sasseville AM, Raymond Y and Cook PR. Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. *Journal of cell science*. 1995; 108:635-644.
39. Neri LM, Raymond Y, Giordano A, Capitani S and Martelli AM. Lamin A is part of the internal nucleoskeleton of human erythroleukemia cells. *Journal of cellular physiology*. 1999; 178:284-295.
40. Shiloh Y. ATM and related protein kinases: safeguarding genome integrity. *Nat Rev Cancer*. 2003; 3:155-168.
41. Lukas J, Lukas C and Bartek J. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nature cell biology*. 2011; 13:1161-1169.
42. Boisvert FM, Hendzel MJ, Masson JY and Richard S. Methylation of MRE11 regulates its nuclear compartmentalization. *Cell Cycle*. 2005; 4:981-989.
43. Chong L, van Steensel B, Broccoli D, Erdjument-Bromage H, Hanish J, Tempst P and de Lange T. A human telomeric protein. *Science*. 1995; 270:1663-1667.
44. Righolt CH, van 't Hoff ML, Vermolen BJ, Young IT and Raz V. Robust nuclear lamina-based cell classification of aging and senescent cells. *Aging (Albany NY)*. 2011; 3:1192-1201.
45. Morales CP, Holt SE, Ouellette M, Kaur KJ, Yan Y, Wilson KS, White MA, Wright WE and Shay JW. Absence of cancer-associated changes in human fibroblasts immortalized with telomerase. *Nat Genet*. 1999; 21:115-118.
46. Jiang XR, Jimenez G, Chang E, Frolkis M, Kusler B, Sage M, Beeche M, Bodnar AG, Wahl GM, Tlsty TD and Chiu CP. Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype. *Nature genetics*. 1999; 21:111-114.
47. Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, Luise C, Schurra C, Garre M, Nuciforo PG, Bensimon A, Maestro R, Pelicci PG and d'Adda di Fagagna F. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature*. 2006; 444:638-642.
48. Hickson I, Zhao Y, Richardson CJ, Green SJ, Martin NM, Orr AI, Reaper PM, Jackson SP, Curtin NJ and Smith GC. Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res*. 2004; 64:9152-9159.
49. Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL and de Lange T. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*. 2009; 138:90-103.
50. Martinez P, Thanasoula M, Munoz P, Liao C, Tejera A, McNees C, Flores JM, Fernandez-Capetillo O, Tarsounas M and Blasco MA. Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes & development*. 2009; 23:2060-2075.
51. Suram A, Kaplunov J, Patel PL, Ruan H, Cerutti A, Boccardi V, Fumagalli M, Di Micco R, Mirani N, Gurung RL, Hande MP, d'Adda di Fagagna F and Herbig U. Oncogene-induced telomere dysfunction enforces cellular senescence in human cancer precursor lesions. *The EMBO journal*. 2012; 31:2839-2851.
52. Fumagalli M, Rossiello F, Clerici M, Barozzi S, Cittaro D, Kaplunov JM, Bucci G, Dobrev M, Matti V, Beausejour CM, Herbig U, Longhese MP and d'Adda di Fagagna F. Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. *Nature cell biology*. 2012; 14:355-365.
53. d'Adda di Fagagna F, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, Von Zglinicki T, Saretzki G, Carter NP and Jackson SP. A DNA damage checkpoint response in telomere-initiated senescence. *Nature*. 2003; 426:194-198.
54. Takai H, Smogorzewska A and de Lange T. DNA damage foci at dysfunctional telomeres. *Curr Biol*. 2003; 13(17):1549-1556.
55. Herbig U, Jobling WA, Chen BP, Chen DJ and Sedivy JM. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Molecular Cell*. 2004; 14:501-513.
56. Rodier F, Coppe JP, Patil CK, Hoeijmakers WA, Munoz DP, Raza SR, Freund A, Campeau E, Davalos AR and Campisi J. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nature cell biology*. 2009; 11:973-979.
57. Batenburg NL, Mitchell TR, Leach DM, Rainbow AJ and Zhu XD. Cockayne Syndrome group B protein interacts with TRF2 and regulates telomere length and stability. *Nucleic acids research*. 2012.
58. Nickerson JA, Krockmalnic G, Wan KM, Turner CD and Penman S. A normally masked nuclear matrix antigen that appears at mitosis on cytoskeleton filaments adjoining chromosomes, centrioles, and midbodies. *The Journal of cell biology*. 1992; 116:977-987.
59. He DC, Nickerson JA and Penman S. Core filaments of the nuclear matrix. *The Journal of cell biology*. 1990; 110:569-580.
60. McKerlie M, Lin S and Zhu XD. ATM regulates proteasome-dependent subnuclear localization of TRF1, which is important for telomere maintenance. *Nucleic acids research*. 2012; 40:3975-3989.

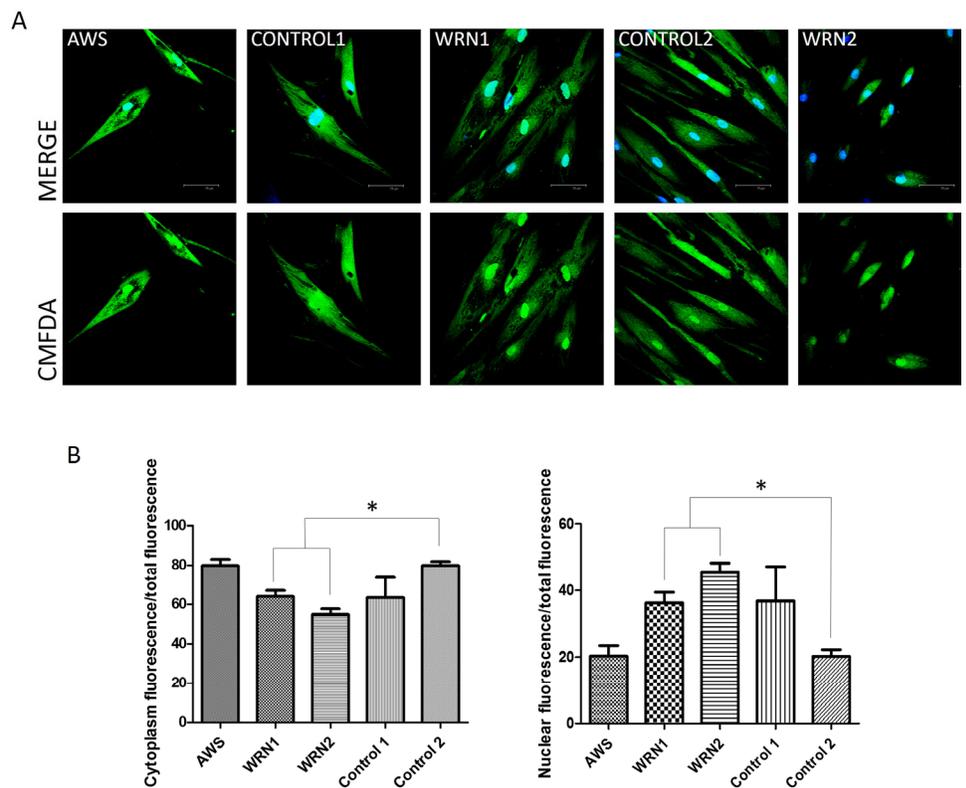
61. Mckerlie M and Zhu XD. Cyclin B-dependent kinase 1 regulates human TRF1 to modulate the resolution of sister telomeres. *Nat Commun.* 2011; 2:371 doi: 10.1038/ncomms1372.
62. Wu Y, Zagal NJ, Rainbow AJ and Zhu XD. XPF with mutations in its conserved nuclease domain is defective in DNA repair but functions in TRF2-mediated telomere shortening. *DNA Repair (Amst).* 2007; 6:157-166.
63. Wu Y, Xiao S and Zhu XD. MRE11-RAD50-NBS1 and ATM function as co-mediators of TRF1 in telomere length control. *Nat Struct Mol Biol.* 2007; 14:832-840.

## SUPPLEMENTARY FIGURES



**Supplementary Figure 1.** Methylated TRF2 exhibits an altered nuclear staining associated with induction of replicative senescence in normal human primary fibroblast cells. Analysis of indirect immunofluorescent with anti-TRF2-2meR17 antibody. Cell nuclei were stained with DAPI in blue.

**Supplementary Figure 2.** Induction of the altered methylated TRF2 staining in late passage normal human primary fibroblast cells. (A) Analysis of indirect immunofluorescence with anti-TRF2-2meR17 antibody in normal human primary fibroblast AG02261 cells at either p15 or p45. Cell nuclei were stained with DAPI in blue. (B) Senescence-associated  $\beta$ -galactosidase assays for AG02261 cells at either p15 or p45. (C) Quantification of percentage of AG02261 cells with altered nuclear staining of methylated TRF2. A total of at least 900 cells in triplicate were scored in blind for each passage as indicated. Standard deviations from three independent experiments are indicated. (D) Quantification of percentage of normal human primary fibroblast GM1706 cells with altered nuclear staining of methylated TRF2. A total of at least 900 cells in triplicate were scored in blind for each passage as indicated. Standard deviations from three independent experiments are indicated.



## Plasma microRNA biomarkers for detection of mild cognitive impairment: biomarker validation study

Kira S. Sheinerman<sup>1</sup>, Vladimir G. Tsivinsky<sup>1</sup>, Laila Abdullah<sup>2</sup>, Fiona Crawford<sup>2</sup>, and Samuil R. Umansky<sup>1</sup>

<sup>1</sup>*DiamiR, LLC, Princeton, NJ 08852, USA;*

<sup>2</sup>*Roskamp Institute, Sarasota, FL 34243, USA*

**Key words:** *microRNA, plasma, circulating, Mild Cognitive Impairment (MCI), Alzheimer's disease, neurodegeneration, brain, synapse*

**Received:** 11/19/13; **Accepted:** 12/20/13; **Published:** 12/22/13

**Correspondence to:** Samuil R. Umansky, PhD; **E-mail:** [sumansky@diamirbio.com](mailto:sumansky@diamirbio.com)

**Copyright:** © Sheinerman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** A minimally invasive test for early detection and monitoring of Alzheimer's and other neurodegenerative diseases is a highly unmet need for drug development and planning of patient care. Mild Cognitive Impairment (MCI) is a syndrome characteristic of early stages of many neurodegenerative diseases. Recently, we have identified two sets of circulating brain-enriched miRNAs, the miR-132 family (miR-128, miR-132, miR-874) normalized per miR-491-5p and the miR-134 family (miR-134, miR-323-3p, miR-382) normalized per miR-370, capable of differentiating MCI from age-matched control (AMC) with high accuracy. Here we report a biomarker validation study of the identified miRNA pairs using larger independent sets of age- and gender- matched plasma samples. The biomarker pairs detected MCI with sensitivity, specificity and overall accuracy similar to those obtained in the first study. The miR-132 family biomarkers differentiated MCI from AMC with 84%-94% sensitivity and 96%-98% specificity, and the miR-134 family biomarkers demonstrated 74%-88% sensitivity and 80-92% specificity. When miRNAs of the same family were combined, miR-132 and miR-134 family biomarkers demonstrated 96% and 87% overall accuracy, respectively. No statistically significant differences in the biomarker concentrations in samples obtained from male and female subjects were observed for either MCI or AMC. The present study also demonstrated that the highest sensitivity and specificity are achieved with pairs of miRNAs whose concentrations in plasma are highly correlated.

### INTRODUCTION

The importance of early diagnosis, treatment and prevention of Alzheimer's disease attracts the attention of scientific and medical communities, regulatory agencies, such as the US Food and Drug Administration (FDA), and industry and government leaders in many countries [1-3]. The number of AD patients and those in high risk populations grows quickly, especially in developed countries, due to increased lifespan. A number of investigational anti-AD drugs, targeting various processes characteristic of AD pathogenesis, have failed in recent clinical trials [1,4-6], likely due to massive neuronal loss and advanced stages of the

disease in the enrolled patients [3-5]. It has been demonstrated that AD dementia is preceded by 10-20 years of the disease development, initially without clinical symptoms (pre-symptomatic AD), and then manifested as MCI [7-9]. It is important to note that the detailed analysis of failed clinical trials has demonstrated a therapeutic benefit in the sub-groups of patients with mild and moderate symptoms of AD [6, see also <https://investor.lilly.com/releaseDetail.cfm?ReleaseID=702211> and <http://www.alzforum.org/new/detail.asp?id=3288>]. The high need for development of new methods for early AD detection is also emphasized in recent publications from the FDA [10, see also <http://www.fda.gov/downloads/Drugs/GuidanceCompli>

anceRegulatoryInformation/Guidances/UCM338287.pdf] and the U.S. Department of Health and Human Services (“National Alzheimer’s Project Act”, available at: <http://aspe.hhs.gov/daltcp/napa/>). Since cognitive testing cannot identify patients in pre-symptomatic stages of AD, effective biomarkers are necessary for successful patient stratification and treatment monitoring [3-5].

Due to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) in the US (<http://www.adni-info.org/>) and similar projects in other countries, a significant progress in early detection of AD with high sensitivity and specificity by imaging techniques and analysis of protein biomarkers in cerebrospinal fluid has been achieved [7]. However, the high cost and invasiveness of these methods make their application for primary screening of large populations impractical [11]. Various approaches to the development of non-invasive or minimally invasive assays for early detection of AD have been tested [12-19]. Currently there is no reliable molecular test for diagnosing AD at the pre-symptomatic or MCI stage. Recently we proposed an approach for early detection of MCI based on analysis of cell-free circulating miRNAs in plasma by RT-qPCR [20]. Several innovations were demonstrated to be effective for selection of potential miRNA biomarkers. First, we hypothesized that changes in concentrations of circulating miRNAs enriched in the brain, and more specifically in hippocampus and frontal cortex, were more likely to reflect AD-associated pathologic processes in the brain than ubiquitous or other organ-enriched miRNAs. Second, we analyzed miRNAs present in neurites and synapses, dysfunction and destruction of which is characteristic of early stages of neurodegeneration, and therefore, could affect expression and secretion of these miRNAs. Third, to compensate for processes unrelated directly to MCI, e.g. changes in blood-brain barrier permeability, we used the “biomarker pair” approach [20-23] normalizing neurite/synapse miRNAs by other brain-enriched miRNAs, which could be expressed in brain areas or cell types not involved in early stages of AD and MCI, as well as miRNAs with levels in plasma changing differently when compared with neurite/synapse miRNAs. Two sets of biomarker pairs, miR-132 (miR-128/miR-491-5p, miR-132/miR-491-5p and miR-874/miR-491-5p) and miR-134 (miR-134/miR-370, miR-323-3p/miR-370 and miR-382/miR-370), capable of differentiating MCI from AMC with sensitivity and specificity of 79%-100% were identified. In a separate small longitudinal study, the identified biomarker miRNA pairs successfully detected MCI in a majority of patients at the asymptomatic stage 1-5 years prior to clinical diagnosis. These miRNA pairs also differentiated AD from AMC ( $P < 0.001$ ) and appeared effective in detecting age-related brain changes in

younger and older controls. Thus, while biomarkers of miR-132 and miR-134 sets do not seem to be specific to AD, they detect some common processes (possibly neurite/synapse dysfunction and destruction), characteristic of AD and other neurodegenerative diseases, and are capable of detecting MCI early. The initial report described analysis of 30 plasma samples in each group (AMC, MCI and AD; 10 in the pilot study for miRNA selection, and 20 in the feasibility study); all plasma samples were collected at the Roskamp Institute (Sarasota, FL).

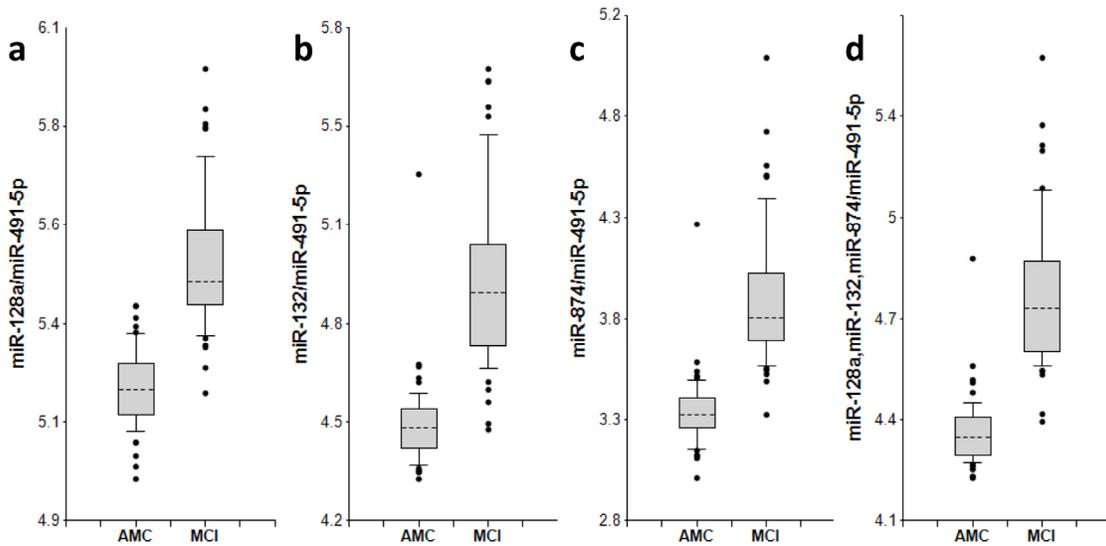
In the present biomarker validation study we analyzed new larger sets of gender- and age-matched plasma samples (50 MCI and 50 AMC) collected at different sites.

## RESULTS

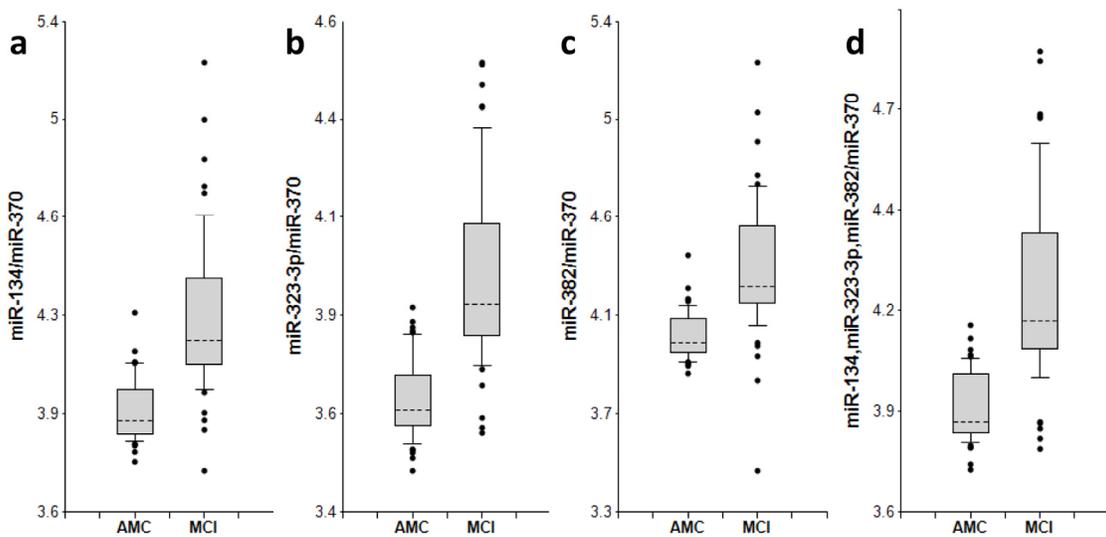
### Biomarker validation

The concentrations of 8 miRNAs were measured by RT-qPCR analysis in plasma samples from 50 MCI patients and 50 AMC subjects (Table 1). The ratios for miRNAs from the miR-132 family to miR-491-5p and for miRNAs from the miR-134 family to miR-370 ( $2^{-\Delta Ct}$ ) are presented as box-plots in Fig. 1 and 2, respectively. Fig. 3 presents Receiver-Operating Characteristic (ROC) curves for miR-132 and miR-134 families. The area under the ROC curve (AUC) for miR-128/miR-491-5p, miR-132/miR-491-5p and miR-874/miR-491-5p is 0.97, 0.97 and 0.98, respectively. These biomarker pairs (Set 1) differentiated MCI from AMC with 84%-94% sensitivity and 96%-98% specificity (Table 2). Further, biomarker pairs miR-134/miR-370, miR-323-3p/miR-370 and miR-382/miR-370 (Set 2) demonstrated 74%-88% sensitivity and 80-92% specificity (Table 2). AUC for miR-134/miR-370, miR-323-3p/miR-370 and miR-382/miR-370 are 0.92, 0.92 and 0.89, respectively. Combining biomarker miRNA pairs within the same set further improves sensitivity and specificity (Fig. 3 and Table 2). Combining biomarker miRNA pairs from miR-132 and miR-134 sets results in sensitivity and specificity that range between values obtained for the two sets of biomarker pairs.

It is important to analyze factors that could affect the test accuracy. The data presented in Figs. 4 and S1 and Table 3 show no statistically significant difference between female and male cohorts of AMC and MCI samples, although a trend toward slightly higher accuracy for MCI differentiation from AMC in the male cohort by miRNA pairs from the miR-132 family is observed, and the opposite trend is observed for miRNA pairs of the miR-134 family.



**Figure 1. miR-132 family biomarker pairs in plasma of AMC and MCI subjects.** The concentrations of miRNA in plasma samples of MCI and age-matched donors with normal cognitive function, 50 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta Ct} \times 100$ . Here and in other figures with box and whisker plots the results are presented in the Log10 scale. The upper and lower limits of the boxes and the lines inside the boxes indicate the 75th and 25th percentiles and the median, respectively. The upper and lower horizontal bars denote the 90th and 10th percentiles, respectively. The points indicate assay values located outside of 80% data. AMC: age-matches controls; MCI: MCI patients.



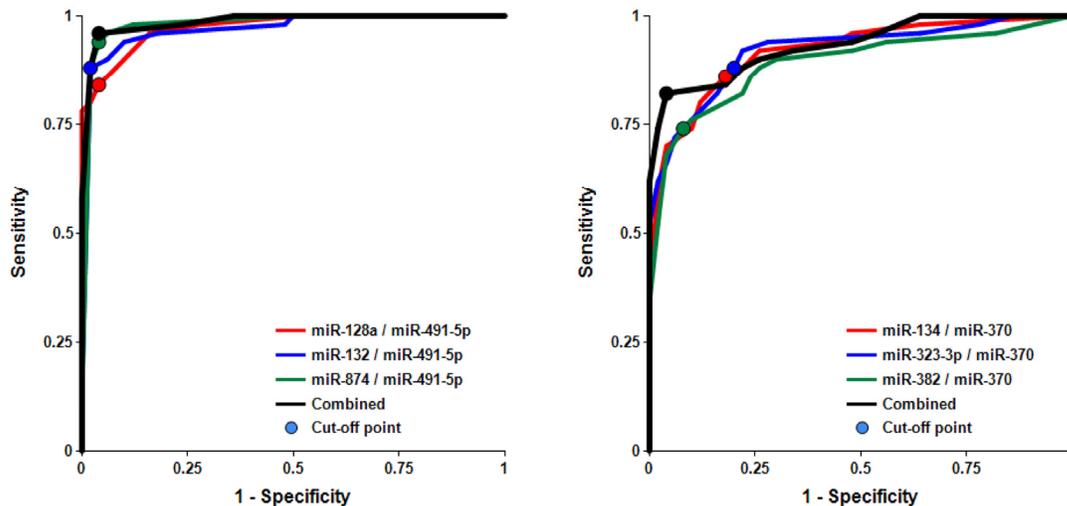
**Figure 2. miR-134 family biomarker pairs in plasma of AMC and MCI subjects.** The concentrations of miRNA in plasma samples of MCI and age-matched donors with normal cognitive function, 50 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta Ct} \times 100$ . See the legend to Fig. 1 for the description of the box and whisker plots. AMC: age-matches controls; MCI: MCI patients.

**Table 1. Demographics of plasma donors**

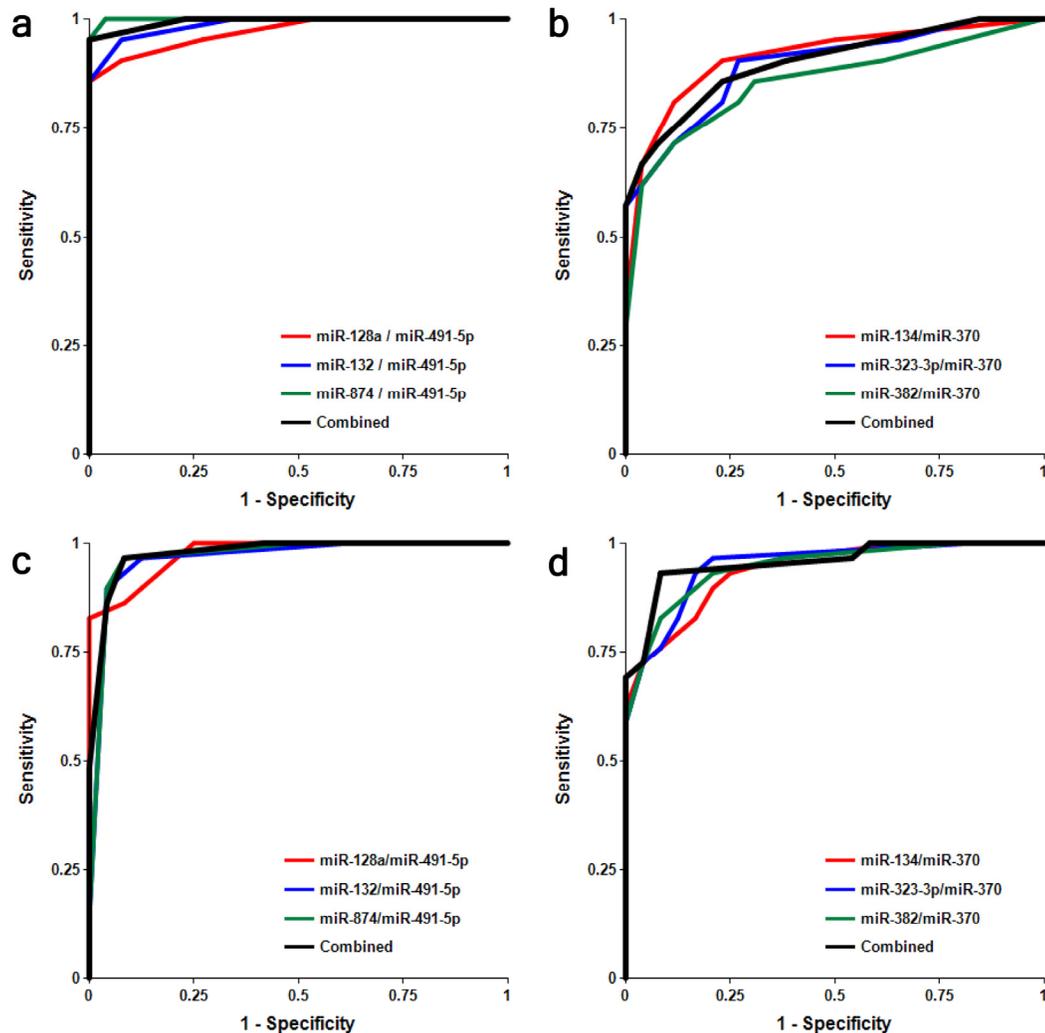
Clinical Diagnosis	Number of Subjects	Age		Sex		MMSE
		Mean	Range	Male	Female	(mean±SD)
AMC	50	65.1	50-82	26	24	29.6 ± 0.6
MCI	50	68.2	51-82	21	29	26.0 ± 1.4

**Table 2. Differentiation of MCI from AMC by miRNA biomarker pairs**

Family/normalizer	miRNA	AUC	Sensitivity	Specificity	Accuracy	P-value (MCI vs. AMC)
miR-132/miR-491-5p	miR-128	0.97	84%	96%	90%	3.53E-16
	miR-132	0.97	88%	98%	93%	1.60E-15
	miR-874	0.98	94%	96%	95%	3.16E-16
	3 pairs	0.98	96%	96%	96%	1.51E-16
miR-134/miR-370	miR-134	0.92	86%	82%	84%	1.55E-12
	miR-323-3p	0.92	88%	80%	84%	9.46E-13
	miR-382	0.89	76%	90%	83%	5.37E-11
	3 pairs	0.93	80%	94%	87%	2.29E-12
miR-132/370	miR-128	0.80	62%	82%	72%	7.10E-7
	miR-132	0.82	74%	76%	75%	3.57E-8
	miR-874	0.85	88%	64%	76%	1.92E-9
	3 pairs	0.83	86%	66%	76%	1.30E-8
miR-134/miR-491-5p	miR-134	0.65	36%	88%	62%	1.00E-2
	miR-323-3p	0.63	38%	88%	63%	2.08E-2
	miR-382	0.63	38%	80%	59%	2.97E-2
	3 pairs	0.65	36%	88%	62%	1.82E-2



**Figure 3. Receiver-Operating Characteristic (ROC) curve analysis of differentiation between MCI patients and age-matched controls obtained with different biomarker pairs.** The areas under the ROC curve (AUC), sensitivity, specificity and accuracy for each biomarker/normalizer pair presented in Table 2 are calculated for the “cutoff” point (indicated as a dot on each plot) - the value of the ratio of paired miRNA where the accuracy of predictions is the highest (see Materials and Methods).



**Figure 4. Receiver-Operating Characteristic (ROC) curve analysis of differentiation between MCI patients and age-matched controls obtained with different biomarker pairs in male (a, b) and female (c, d) cohorts.** The areas under the ROC curve (AUC), sensitivity, specificity and accuracy for each biomarker/normalizer pair presented in Table 3 are calculated for the “cutoff” point - the value of the ratio of paired miRNA where the accuracy of predictions is the highest.

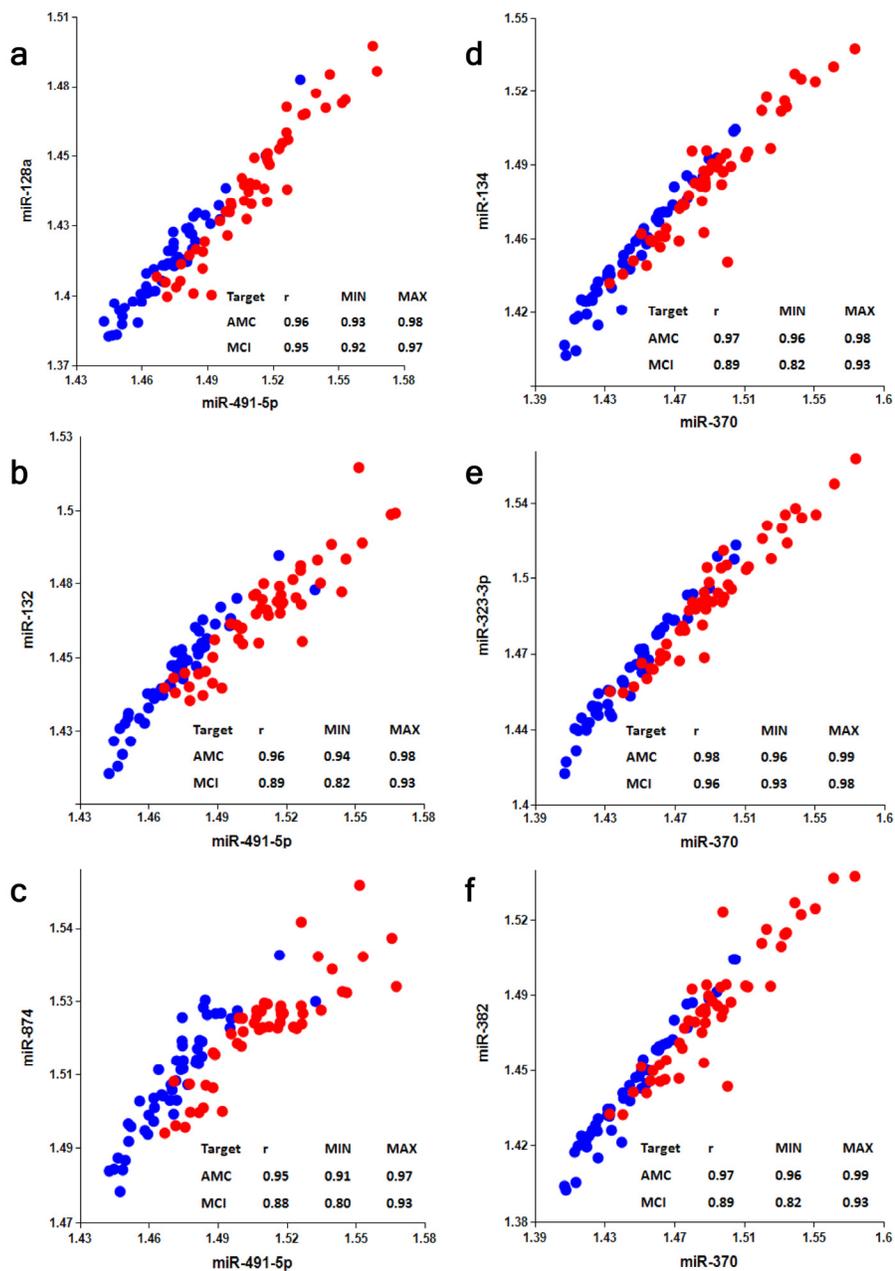
### A role of miRNA normalizer

Selection of an optimal denominator (normalizer) for each miRNA family was shown to be essential [20]. miR-491-5p and miR-370 were found to be effective when paired with miRNAs of the miR-132 and miR-134 families, respectively. This finding has been further tested in the present study. Figs. S2-S4 and Table 2 show that if normalizers are switched between the two families, pairs miR-128/miR-370, miR-132/miR-370, miR-874/miR-370, miR-134/miR-491-5p, miR-323-3p/miR-491-5p and miR-382/miR-491-5p differentiate MCI from AMC with much lower sensitivity and specificity. Concentrations of miRNAs in plasma depend

on numerous factors, including (i) levels of miRNA expression in various organs and tissues; (ii) levels of miRNA secretion from different cell types; (iii) stability of miRNAs in extracellular space and their appearance in plasma in different forms, such as exosomes and other micro-vesicles, complexes with proteins, lipids and, possibly, other molecules; and (iv) blood-brain barrier permeability for brain-enriched miRNAs. A pathological process may affect some or all of these factors. It is, therefore, logical to expect that a numerator and a denominator of an effective biomarker miRNA pair should share some of these basic common factors (e.g. both are brain-enriched and secreted in exosomes) and would change differently in

response to a pathology). In such cases, one can expect a high correlation between miRNAs of miR-132 and miR-134 families and their optimal respective normalizers, miR-491-5p and miR-370. Data presented in Fig. 5 demonstrate that in the AMC cohort Spearman test  $r$  values for the correlation between miRNAs of the miR-132 family with miR-491-5p are in the 0.95-0.96 range and for the correlation between miRNAs of the miR-134 family with miR-370 are in

the 0.97-0.98 range. In the MCI cohort, the correlation between the same miRNAs is slightly lower, indicating that the pathology differently affects plasma levels of miRNAs of the miR-132 family and of miR-491-5p, as well as levels of miRNAs of the miR-134 family and of miR-370. Correlations between neurite/synapse-enriched miRNAs from one family with the optimal normalizer of another family are significantly weaker (Fig. S5).



**Figure 5. Analysis of correlation between members of miR-132 and miR-134 families and their optimal normalizers miR-491-5p and miR-370, respectively.** Spearman's rank correlation coefficient  $r$  along with 95% confidence intervals (MIN & MAX) is shown for AMC (blue dots) and MCI (red dots) subjects.

**Table 3. Comparison of biomarker miRNA pairs of miR-132 and miR-134 families in male and female subjects**

Male								
miRNA pair	128a/491-5p	132/491-5p	874/491-5p	miR-132 Fam. combined	134/370	323-3p/370	382/370	miR-134 Fam. combined
AUC	0.97	0.99	1.00	0.99	0.91	0.89	0.86	0.90
Female								
AUC	0.98	0.97	0.97	0.98	0.94	0.95	0.95	0.96
Male – Female Comparison								
P-value	0.763	0.479	0.201	0.686	0.601	0.315	0.167	0.285

## DISCUSSION

The main objective of the present work was to validate previously identified sets of plasma biomarker miRNA pairs [20] in a larger study with clinical samples collected at sites different from the one used in the original study. The data have validated miRNAs of the miR-132 and miR-134 families, paired with miR-491-5p and miR-370 respectively, as highly sensitive biomarkers for detection of MCI. The overall accuracy for differentiating MCI from AMC is 90%-96% and 83%-87% for the biomarker miRNA pairs of miR-132 and miR-134 sets, respectively. The corresponding values obtained in the first feasibility study were 86% - 92% and 82% - 89%. Since a large number of MCI patients will progress to AD dementia [24-26], it is reasonable to suggest that these biomarker pairs detect early stages of AD as well, although they do not differentiate AD from MCI caused by other conditions. As was the case in the feasibility study, the miR-132 family biomarkers detected MCI with higher accuracy than the miR-134 family biomarkers. Although the roles of most miRNAs tested in this study in neuronal differentiation, function and pathology have not been elucidated yet, it has been demonstrated that miR-132 and miR-134 have opposite effect on neurons: miR-132 stimulates [27,28] and miR-134 [29] inhibits neurite growth. Also, the level of miR-132 has been shown to be lower in the hippocampus and temporal neocortex of AD patients [30,31]. Lau et al. [32] have demonstrated that downregulation of miR-132 occurs at Braak stages III and IV, prior to loss of neuron specific miRNAs. They have also found that deregulation of miR-132-3p in the AD brain appears to occur mainly in neurons displaying Tau hyper-phosphorylation and that the transcription factor FOX01a is a key target of miR-132 in the Tau network. Interestingly, the concentration of miR-128, which promotes neuronal maturation [33], has been shown to increase in the hippocampus in an intermediate stage and to decrease in a late stage of AD

[34,35]. Aging-associated increase in the concentrations of miR-134 and miR-874 in serum has been demonstrated [36]. We plan to further analyze the utility of both sets of biomarker miRNA pairs for MCI detection in larger longitudinal studies.

The present study has not shown statistically significant differences between male and female cohorts in differentiating MCI from AMC, suggesting that a combined control group could be used in further studies. These results need to be confirmed in larger follow-on studies.

The present study further validated the use of effective “miRNA pairs”, i.e. pairing of an optimal miRNA normalizer (denominator in biomarker pair) with a particular miRNA as the numerator. In the previous study [20] we analyzed levels of neurite and/or synapse miRNAs and other brain-enriched miRNAs in plasma of MCI and AMC subjects, and then the ability of all possible miRNA pairs to differentiate MCI from AMC was tested. Neurite/synapse miRNAs (miR-132 and miR-134 families) were found to be the best nominators in the identified and selected biomarker pairs. These data supported our initial hypothesis: neurite/synapse miRNAs can be effective biomarkers of neurodegeneration, because synapse dysfunction and subsequent neurite and synapse destruction are early events in the progression of neurodegenerative diseases. We also demonstrated that miR-491-5p was a preferred normalizer for the miR-132 family, and miR-370 was a preferred normalizer for the miR-134 family, although the nature of these preferences was not clear at the time. Here we have further analyzed this phenomenon and found that a high correlation between numerator and denominator of biomarker miRNA pair in plasma samples from different subjects is an important parameter for their compatibility. It is currently unclear on what factors such a correlation depends, since many factors likely affect concentrations of cell-free miRNAs

in plasma. Intuitively, it seems reasonable to suggest that an efficient miRNA pair should include two plasma miRNAs, which share common properties (for example, miRNAs secreted/excreted by the same mechanism, miRNAs bound to the same protein in plasma or present in similar exosomes, etc.), but differ in their response to investigated pathology. Hence, correlation analysis could be a useful approach for selecting the effective biomarker pairs among bodily fluid miRNAs for various diagnostic applications.

Thus, the present study has validated two sets of plasma biomarker miRNA pairs for the early detection of MCI, providing a basis for a large longitudinal study for determining the biomarkers' ability to detect MCI and AD at pre-symptomatic stages. The described approach is complementary to other diagnostic technologies, such as neuroimaging and CSF analysis.

## MATERIALS AND METHODS

Plasma samples. K2EDTA Plasma samples from 50 MCI patients and 50 AMC were obtained from a commercial vendor PrecisionMed (Solana Beach, California). The samples were collected in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and a written consent was obtained from each subject. All samples were frozen at -20°C within 2 hours from collection, then transferred to -80°C, and stored and shipped at -80°C.

MCI diagnosis was based on several tests evaluating cognition: (i) ADAS-Cog (Alzheimer's Disease Assessment Scale-Cognitive subscale); (ii) CDRS (Clinical Dementia Rating Scale); (iii) Wechsler Memory Scale; and (iv) MMSE (Mini Mental State Examination). MCI classification requirements included the following parameters: (i)  $28 \geq \text{MMSE} \geq 22$ ; (ii) not demented; (iii) memory complaint; (iv) preserved general cognitive function; (v) intact activities of daily living (allowed problems with 2 or less of the following: phone calls, meal preparation, handling money, completing chores); (vi) abnormal memory function documented by scoring below the education adjusted cutoff on the Logical Memory II subscale (delayed paragraph recall) from the Wechsler Memory Scale-Revised (maximum score = 25): (a)  $< 8$  for 16 years or more of education; (b)  $< 4$  for 8-15 years of education; (c)  $< 2$  for 0-7 years of education. Patients with other neurological disorders were excluded from the study.

Cognitive status of AMC subjects was also evaluated using metrics listed above. AMC subjects had MMSE scores of 29 or 30, maintained independent activities of

daily living, and did not have a known history of neurological illnesses, psychiatric disorders, or other medical conditions that could potentially interfere with their cognitive performance.

Demographic characteristics of the study groups are summarized in Table 1.

Plasma RNA extraction and qRT-PCR miRNA analysis. miRNA isolation and qRT-PCR analysis were performed by Asuragen Inc. (Austin, TX, USA) as previously described [20]. Briefly, RNA was extracted from 200  $\mu\text{l}$  aliquots of plasma using Trizol treatment and silica binding. Single target qRT-PCR was performed using the TaqMan® Reverse Transcription Kit and miRNA specific stem-loop primers (Applied Biosystems, Foster City, CA, USA). The RT step was performed in triplicate and 2  $\mu\text{l}$  plasma equivalents were present in final PCR

Bioinformatics analysis and statistical methods. All statistical calculations were performed with the use of custom software developed at DiamiR LLC (Princeton, NJ), as previously described [16]. Briefly, Mann-Whitney U-tests were used to evaluate significance of differentiation of any two patient groups by various miRNA pairs, and Spearman's rank correlation coefficient was calculated to estimate associations between various miRNAs. Receiver-Operating Characteristic (ROC) curves were constructed and the area under ROC curves (AUC) was calculated to evaluate sensitivity and specificity of various biomarker sets. The cutoff points on the ROC curves, at which accuracy of MCI detection is maximal, were selected.

## Conflicts of Interest Statement

Kira S. Sheinerman, Vladimir G. Tsvinsky, and Samuil R. Umansky are shareholders of DiamiR, LLC.

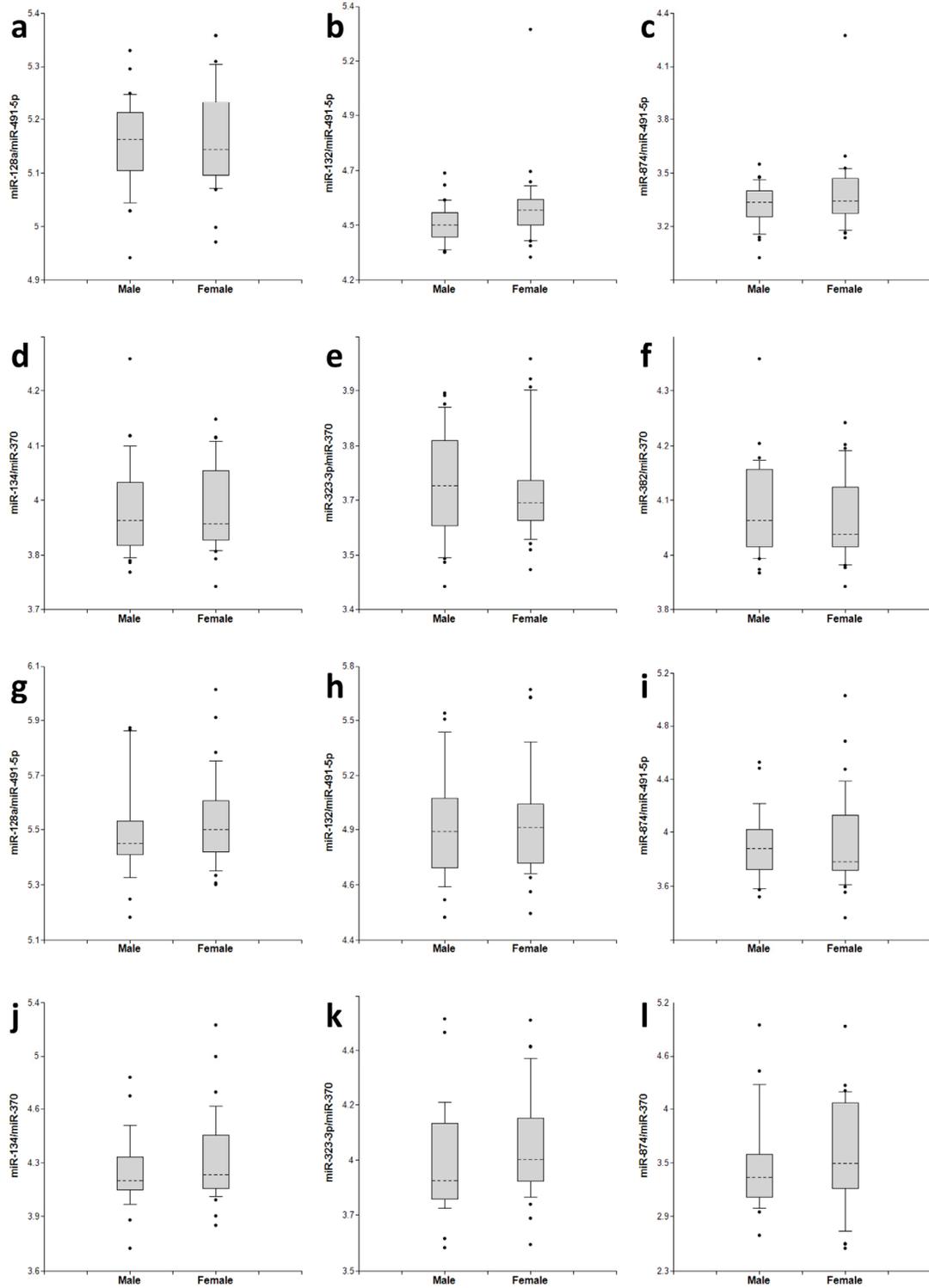
## REFERENCES

1. Gerald Z, Ockert W. Alzheimer's disease market: hope deferred. *Nat Rev Drug Discov.* 2013; 12:19-20.
2. Dorsey ER, George BP, Leff B, Willis AW. The coming crisis: obtaining care for the growing burden of neurodegenerative conditions. *Neurology.* 2013; 80:1989-1996.
3. Snyder EM, Olin J, David FS. Maximizing the value of diagnostics in Alzheimer's disease drug development. *Nat Rev Drug Discov.* 2012; 11:183-184.
4. Karakaya T, Fußer F, Schröder J, Pantel J. Pharmacological Treatment of Mild Cognitive Impairment as a Prodromal Syndrome of Alzheimer's Disease. *Curr Neuropharmacol.* 2013; 11:102-108.
5. Henriksen K, O'Bryant SE, Hampel H, Trojanowski JQ, Montine TJ, Jeromin A, Blennow K, Lönneborg A, Wyss-Coray T, Soares H,

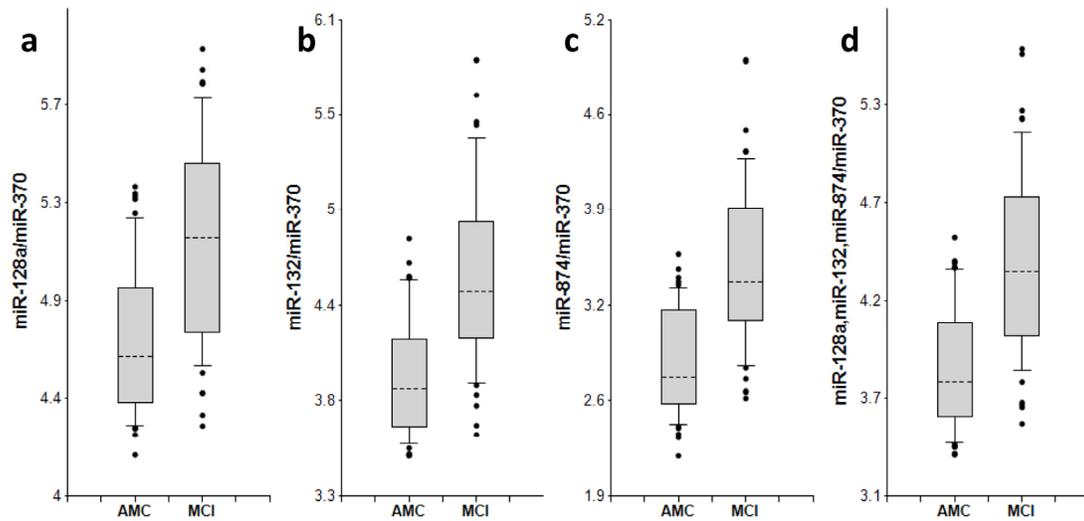
- Bazenet C, Sjögren M, Hu W, Lovestone S, Karsdal MA, Weiner MW; Blood-Based Biomarker Interest Group. The future of blood-based biomarkers for Alzheimer's disease. *Alzheimers Dement*. 2013 Jul 11 [Epub ahead of print].
6. Cunningham EL, Passmore AP. Drug development in dementia. *Maturitas*. 2013 May 23. [Epub ahead of print].
  7. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, Morris JC, Petersen RC, Saykin AJ, Schmidt ME, Shaw L, Shen L, Siuciak JA, Soares H, Toga AW, Trojanowski JQ; Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimers Dement*. 2013 Aug 6 [Epub ahead of print].
  8. Pillai JA, Cummings JL. Clinical trials in predementia stages of Alzheimer disease. *Med Clin North Am*. 2013; 97:439-457.
  9. Sperling RA, Jack CR Jr, Aisen PS. Testing the right target and right drug at the right stage. *Sci Transl Med*. 2011; 3:111cm33.
  10. Delbecke D, Alessio A, Iagaru A. FDA AD Drug Development Guidance. *J Nucl Med*. 2013; 54:16N.
  11. Bredesen D. miRNA-synaptic crystal ball? *Aging (Albany NY)*. 2012; 4:732-733.
  12. Ray S, Britschgi M, Herbert C, Takeda-Uchimura Y, Boxer A, Blennow K, Friedman LF, Galasko DR, Jutel M, Karydas A, Kaye JA, Leszek J, Miller BL, Minthon L, Quinn JF, Rabinovici GD, et al. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat Med*. 2007; 13:1359-1362.
  13. Reddy MM, Wilson R, Wilson J, Connell S, Gocke A, Hynan L, German D, Kodadek T. Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. *Cell*. 2011; 144:132-142.
  14. Nagele E, Han M, Demarshall C, Belinka B, & Nagele R. Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. *PLoS One*. 2011; 6:e23112.
  15. Leidinger P, Backes C, Deutscher S, Schmitt K, Muller SC, Frese K, Haas J, Ruprecht K, Paul F, Stahler C, Lang CJ, Meder B, Bartfai T, Meese E, Keller A. A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biol*. 2013; 14:R78.
  16. Geekiyanage H, Jicha GA, Nelson PT, Chan C. Blood serum miRNA: non-invasive biomarkers for Alzheimer's disease. *Exp Neurol*. 2012; 235:491-496.
  17. Provost P. Interpretation and applicability of microRNA data to the context of Alzheimer's and age-related diseases. *Aging (Albany NY)*. 2010; 2:166-169.
  18. Kumar P, Dezsó Z, Mackenzie C, Oestreich J, Agoulnik S, Byrne M, Bernier F, Yanagimachi M, Aoshima K, Oda Y. Circulating miRNA Biomarkers for Alzheimer's Disease. *PLoS One*. 2013; 8:e69807.
  19. Bekris LM, Lutz F, Montine TJ, Yu CE, Tsuang D, Peskind ER, Leverenz JB. MicroRNA in Alzheimer's disease: an exploratory study in brain, cerebrospinal fluid and plasma. *Biomarkers*. 2013; 18:455-466.
  20. Sheinerman KS, Tsvinsky VG, Crawford F, Mullan MJ, Abdullah L, Umansky SR. Plasma microRNA biomarkers for detection of mild cognitive impairment. *Aging (Albany NY)*. 2012; 4:590-605.
  21. Hennessey PT, Sanford T, Choudhary A, Mydlarz WW, Brown D, Adai AT, Ochs MF, Ahrendt SA, Mambo E, Califano JA. Serum microRNA biomarkers for detection of non-small cell lung cancer. *PLoS One*. 2012; 7:e32307.
  22. Matthaei H, Wylie D, Lloyd MB, Dal Molin M, Kempainen J, Mayo SC, Wolfgang CL, Schulick RD, Langfield L, Andruss BF, Adai AT, Hruban RH, Szafranska-Schwarzbach AE, Maitra A. miRNA biomarkers in cyst fluid augment the diagnosis and management of pancreatic cysts. *Clin Cancer Res*. 2012; 18:4713-4724.
  23. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabrò E, Croce CM, Pastorino U, Sozzi G. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci U S A*. 2011; 108:3713-3718.
  24. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999; 56:303-308.
  25. Mariani E, Monastero R, Mecocci P. Mild cognitive impairment: a systematic review. *J Alzheimers Dis*. 2007; 12:23-35.
  26. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*. 2009; 119:252-265.
  27. Vo N, Klein ME, Varlamova O, Keller DM, Yamamoto T, Goodman RH, Impey S. A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. *Proc Natl Acad Sci U S A*. 2005; 102:16426-16431.
  28. Wayman GA, Davare M, Ando H, Fortin D, Varlamova O, Cheng HY, Marks D, Obrietan K, Soderling TR, Goodman RH, Impey S. An activity-regulated microRNA controls dendritic plasticity by down-regulating p250GAP. *Proc Natl Acad Sci U S A*. 2008; 105:9093-9098.
  29. Schrott GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME. A brain-specific microRNA regulates dendritic spine development. *Nature*. 2006; 439:283-289.
  30. Cogswell JP, Ward J, Taylor IA, Waters M, Shi Y, Cannon B, Kelnar K, Kempainen J, Brown D, Chen C, Prinjha RK, Richardson JC, Saunders AM, Roses AD, Richards CA. Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. *J Alzheimers Dis*. 2008; 14:27-41.
  31. Hébert SS, Wang WX, Zhu Q, Nelson PT. A study of small RNAs from cerebral neocortex of pathology-verified Alzheimer's disease, dementia with lewy bodies, hippocampal sclerosis, frontotemporal lobar dementia, and non-demented human controls. *J Alzheimers Dis*. 2013; 35:335-348.
  32. Lau P, Bossers K, Janky R, Salta E, Frigerio CS, Barbash S, Rothman R, Sierksma AS, Thathiah A, Greenberg D, Papadopoulou AS, Achsel T, Ayoubi T, Soreq H, Verhaagen J, Swaab DF, Aerts S, De Strooper B. Alteration of the microRNA network during the progression of Alzheimer's disease. *EMBO Mol Med*. 2013; 5:1613-1634.
  33. Bruno IG, Karam R, Huang L, Bhardwaj A, Lou CH, Shum EY, Song HW, Corbett MA, Gifford WD, Gecz J, Pfaff SL, Wilkinson MF. Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol Cell*. 2011; 42:500-510.
  34. Lukiw WJ. Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus. *Neuroreport*. 2007; 18:297-300.
  35. Müller M, Kuiperij HB, Claassen JA, Küsters B, Verbeek MM. MicroRNAs in Alzheimer's disease: differential expression in hippocampus and cell-free cerebrospinal fluid. *Neurobiol Aging*. 2013 Aug 17 [Epub ahead of print].

**36.** Dhahbi JM, Spindler SR, Atamna H, Yamakawa A, Guerrero N, Boffelli D, Mote P, Martin DI. Deep sequencing identifies circulating mouse miRNAs that are functionally implicated in manifestations of aging and responsive to calorie restriction. *Aging (Albany NY)*. 2013; 5:130-141.

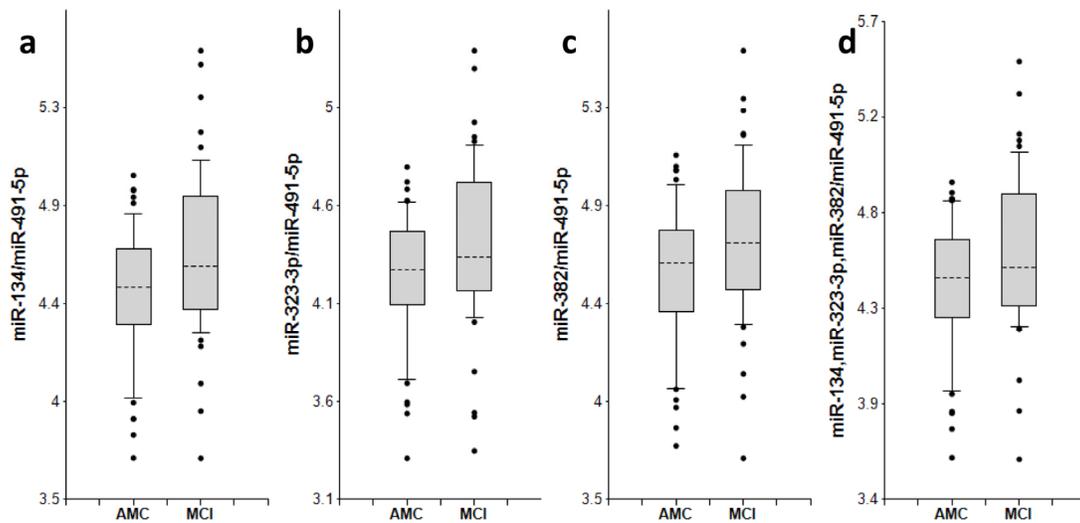
SUPPLEMENTAL MATERIALS



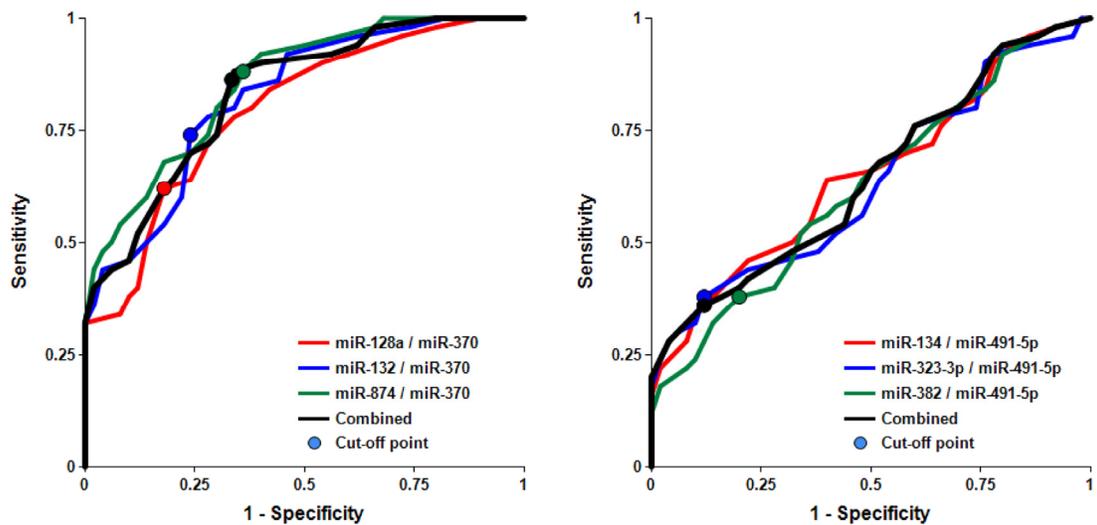
**Figure S1. Comparison of two biomarker families in male and female subjects. a-f: AMC; g-l : MCI.** The concentrations of miRNA in plasma samples of MCI and age-matched donors with normal cognitive function were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta Ct} \times 100$ . Here and in other figures with box and whisker plots the results are presented in the Log10 scale. The upper and lower limits of the boxes and the lines inside the boxes indicate the 75th and 25th percentiles and the median, respectively. The upper and lower horizontal bars denote the 90th and 10th percentiles, respectively. The points indicate assay values located outside of 80% data.



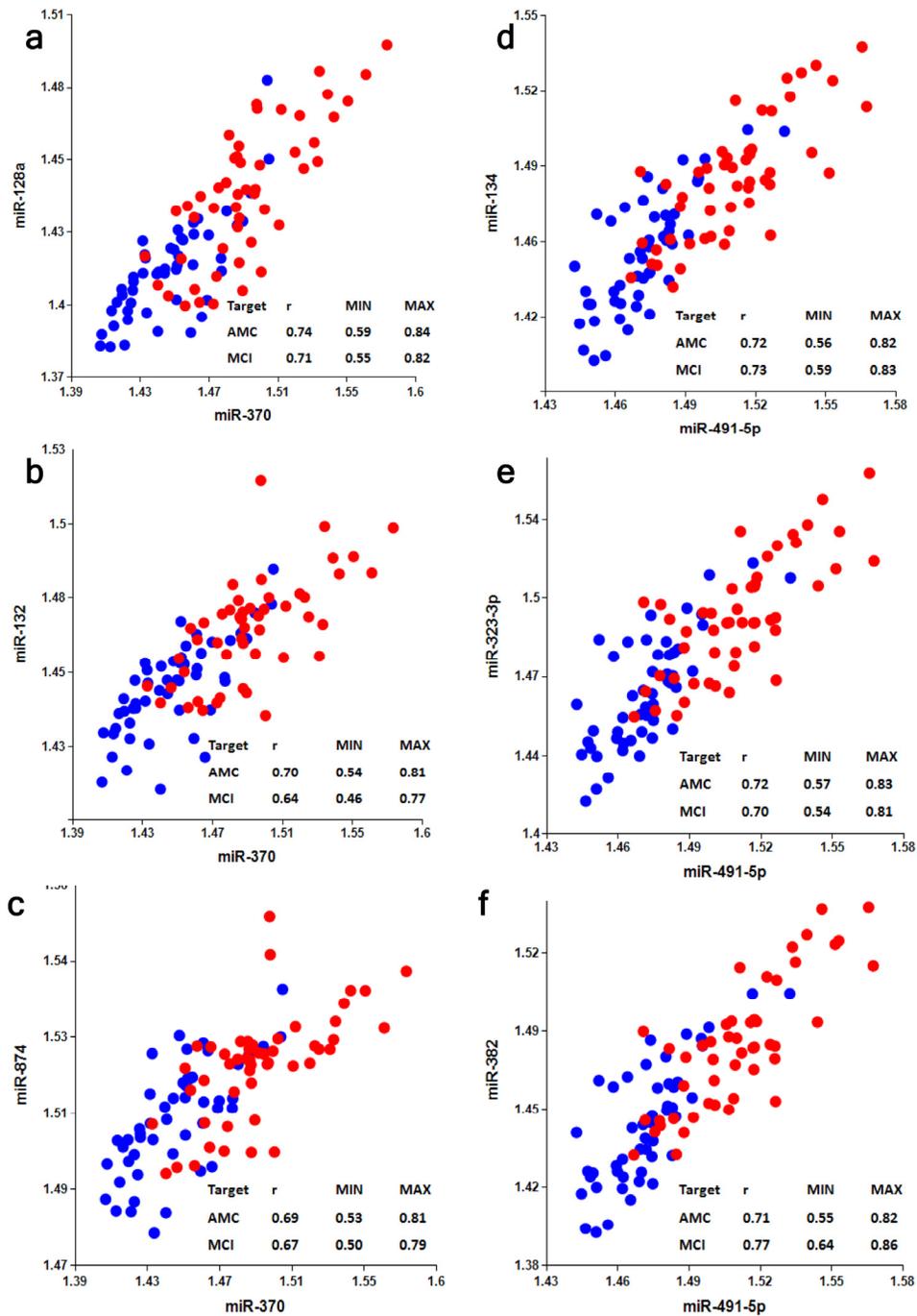
**Figure S2. miRNAs of miR-132 family paired with miR-370 in plasma of AMC and MCI subjects.** The concentrations of miRNA in plasma samples of MCI and age-matched donors with normal cognitive function, 50 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta Ct} \times 100$ . See the legend to Fig. S1 for the description of the box and whisker plots.



**Figure S3. miRNAs of miR-134 family paired with miR-491-3p in plasma of AMC and MCI subjects.** The concentrations of miRNA in plasma samples of MCI and age-matched donors with normal cognitive function, 50 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta Ct} \times 100$ . See the legend to Fig. S1 for the description of the box and whisker plots.



**Figure S4. Receiver-Operating Characteristic (ROC) curve analysis of differentiation between MCI patients and age-matched controls obtained with miR-132 paired with miR-370 and miR-134 family members paired with miR-491-5p.** The areas under the ROC curve (AUC), sensitivity, specificity and accuracy for each biomarker/normalizer pair presented in Table 2 are calculated for the “cutoff” point (indicated as a dot on each plot) - the value of the ratio of paired miRNA where the accuracy of predictions is the highest (see Materials and Methods).



**Figure S5. Analysis of correlation between members of miR-132 and miR-134 families and normalizers optimal for another family, miR-370 and miR-491-5p, respectively.** Spearman's rank correlation coefficient  $r$  along with 95% confidence intervals (MIN & MAX) is shown for AMC (blue dots) and MCI (red dots) subjects.

## Specific lipofuscin staining as a novel biomarker to detect replicative and stress-induced senescence. A method applicable in cryo-preserved and archival tissues

Georgakopoulou EA<sup>1\*</sup>, Tsimaratou K<sup>1\*</sup>, Evangelou K<sup>1\*</sup>, Fernandez Marcos-PJ<sup>2</sup>, Zoumpourlis V<sup>3</sup>, Trougakos IP<sup>4</sup>, Kletsas D<sup>5</sup>, Bartek J<sup>6,7</sup>, Serrano M<sup>2</sup>, and Gorgoulis VG<sup>1,8</sup>

<sup>1</sup> Molecular Carcinogenesis Group, Department of Histology and Embryology, School of Medicine, University of Athens, Greece

<sup>2</sup> Tumor Suppression Group, Molecular Oncology Program, Spanish National Cancer Centre (CNIO), Melchor Fernández Almagro 3, Madrid E-28029, Spain

<sup>3</sup> National Hellenic Research Foundation, Institute of Biological Research and Biotechnology, Athens, Greece

<sup>4</sup> Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 15784, Greece

<sup>5</sup> Laboratory of Cell Proliferation and Ageing, Institute of Biology, National Centre for Scientific Research 'Demokritos', Athens, Greece

<sup>6</sup> Danish Cancer Society Research Center, Copenhagen, Denmark

<sup>7</sup> Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic

<sup>8</sup> Biomedical Research Foundation, Academy of Athens, 11527 Athens, Greece

\*The authors contributed equally to this work

**Key words:** lipofuscin, cellular senescence, Sudan Black B, formalin-fixed paraffin-embedded, biomarker

**Received:** 12/16/12; **Accepted:** 12/27/12; **Published:** 12/29/12

**Correspondence to:** Vassilis G Gorgoulis, PhD; **E-mail:** [vgorg@med.uoa.gr](mailto:vgorg@med.uoa.gr); [vgorgoulis@gmail.com](mailto:vgorgoulis@gmail.com)

**Copyright:** © Georgakopoulou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** There is shortage of extensive clinicopathologic studies of cellular senescence because the most reliable senescence biomarker, the detection of Senescence-Associated-beta-galactosidase activity (SA-β-gal), is inapplicable in archival material and requires snap-frozen tissues. We validated the histochemical Sudan-Black-B (SBB) specific stain of lipofuscin, an aggregate of oxidized proteins, lipids and metals, known to accumulate in aged tissues, as an additional reliable approach to detect senescent cells independently of sample preparation. We analyzed cellular systems in which senescence was triggered by replicative exhaustion or stressful stimuli, conditional knock-in mice producing precancerous lesions exhibiting senescence, and human preneoplastic lesions known to contain senescent cells. In the above settings we demonstrated co-localization of lipofuscin and SA-β-gal in senescent cells *in vitro* and *in vivo* (cryo-preserved tissue), strongly supporting the candidacy of lipofuscin for a biomarker of cellular senescence. Furthermore, cryo-preserved tissues positive for SA-β-gal were formalin-fixed, paraffin-embedded, and stained with SBB. The corresponding SA-β-gal positive tissue areas stained specifically for lipofuscin by SBB, whereas tissues negative for SA-β-gal were lipofuscin negative, validating the sensitivity and specificity of the SBB staining to visualize senescent cells in archival material. The latter unique property of SBB could be exploited in research on widely available retrospective tissue material.

## INTRODUCTION

Cellular senescence is the state of irreversible cellular growth arrest in which the cell remains metabolically active [1, 2]. Two types of cellular senescence have been described in mammalian cells [3]. Replicative Senescence (RS) that is triggered by the arrest of cellular proliferation after a certain number of divisions due to telomere attrition [1, 3] and Stress Induced Premature Senescence (SIPS) that is a more acute phenomenon in which the cells stop to proliferate under various stress conditions, regardless of telomere length [3]. The phenomenon of cellular senescence was originally described *in vitro* [4]. More recently, senescent cells were also identified in aged skin [2], benign tumors and premalignant lesions [5-9] as well as in age-related pathologies [10]. Also, the number of senescent fibroblasts reportedly increases exponentially in the skin of aging primates, reaching >15% of all cells in very old individuals [11]. The evidence so far from *in vitro* and *in vivo* studies suggests that cellular senescence acts as a tumor barrier, whereas it contributes to the processes of tissue aging and age-related diseases [12]. The significance of cellular senescence in carcinogenesis and age-related disorders, renders the detection of these phenomena essential. This urgent need of reliable biomarkers of senescence is even more apparent given the evidence for cellular senescence induced in response to anticancer therapy [13].

The most widely used biomarker of cellular senescence reported so far is the detection of Senescence-Associated  $\beta$ -Galactosidase activity (SA- $\beta$ -gal) in sub-optimal pH [2, 14]. Nevertheless, a major disadvantage in designing large-scale studies of cellular senescence in human lesions is that SA- $\beta$ -gal staining requires fresh tissue as it is based on an enzymatic reaction [14]. This fact seriously limits the exploitation of the widely available formalin-fixed paraffin-embedded (FFPE) archival tissues, including tissue microarrays [1]. In an effort to establish a biomarker of cellular senescence that could be applicable for FFPE archival tissue material, we focused on lipofuscin, also known as an "age-pigment" [15]. Lipofuscin is an aggregate of oxidized proteins that accumulates progressively mostly in aged post mitotic cells [16]. It is considered a hallmark of aging and is also involved in the pathogenesis of certain age related pathologies such as macular degeneration [16]. Sudan Black B (SBB) is a lipophilic histochemical stain that identifies lipofuscin and is applicable for *in vitro* and *in situ* studies [17-19]. Here we employed SBB in a series of experiments designed to demonstrate that lipofuscin accumulates *in vitro* in normal human cells during RS or SIPS, as well as in stressed human cancer cells. Furthermore, we

sought to identify lipofuscin deposits in benign lesions already known to contain senescent cells. As a control marker of the cellular senescence state we used the SA- $\beta$ -gal assay. Our results show that the SBB-stained lipofuscin is present in all the cells that express SA- $\beta$ -gal activity and it is absent in SA- $\beta$ -gal-negative cells. Hence, SBB positivity could be used as an additional cellular senescence biomarker. Moreover, SBB staining was applicable in FFPE tissue sections, providing evidence that this assay can provide a reliable biomarker for detection of senescent cells in archival clinical material that is stored in paraffin.

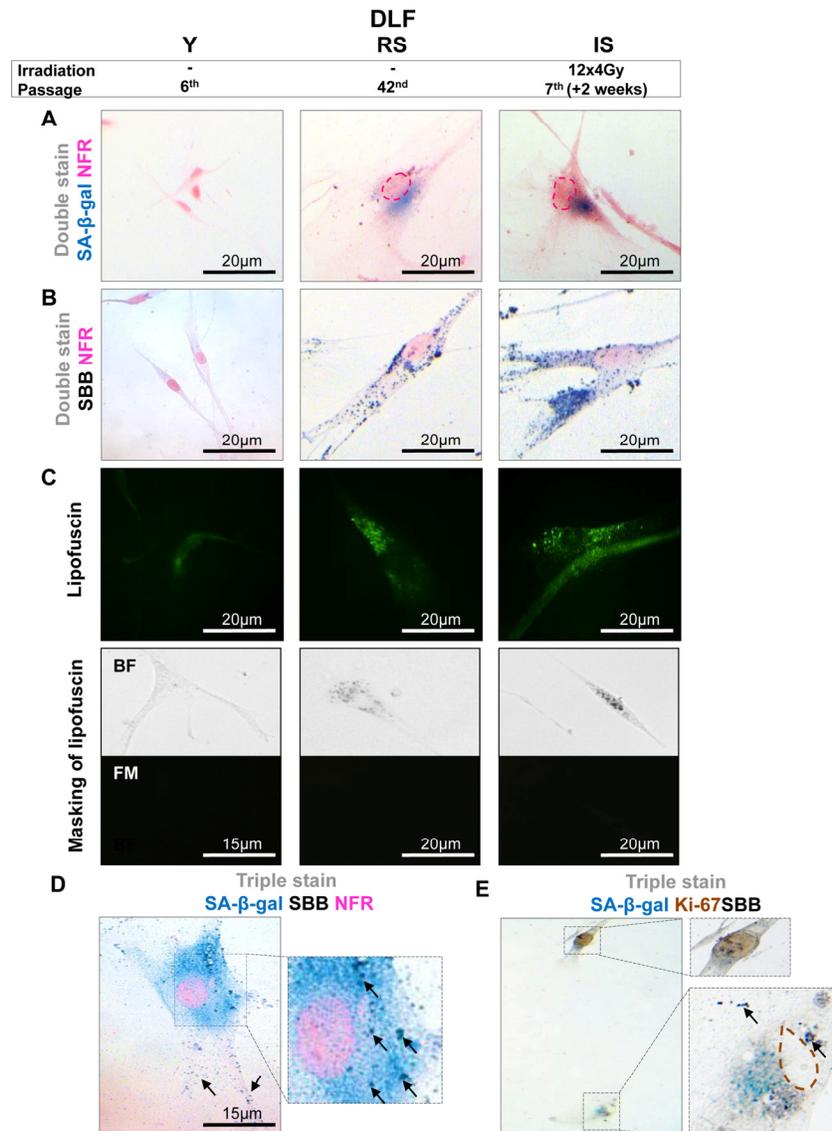
## RESULTS

To assess the value of lipofuscin as a potential biomarker of cellular senescence *in vitro*, five cellular systems of normal diploid cells and cancer cells were applied. In these experimental settings cellular senescence was triggered by means of proliferative exhaustion (Replicative Senescence, RS) or Stress Induced Premature Senescence (SIPS). Specifically, we used young proliferating primary human diploid lung fibroblasts (DLF) (at passage 6), along with replicatively senescent cells (at passage 42) and cells (at passage 7) in which SIPS was triggered by  $\gamma$ -irradiation [20] (Fig. 1). The effect of p53 and p21<sup>WAF-1</sup>, two well established effectors of the senescence program, was studied in two inducible osteosarcoma cell lines, namely Saos-2 p21-Tet-ON and Saos-2 p53-Tet-ON [21-24]. After 8 days of p21<sup>WAF-1</sup> or p53 induction senescent cells were evident (Fig. 2A, 3A). Finally, the inducible osteosarcoma U2OS E2F1-ER cell line was also tested. Over-expression of E2F1 has been shown to trigger SIPS via activation of the DNA damage response (DDR) pathway [25]. The induction of E2F1 yielded senescent cells in 10 days (Fig. 4A) as concluded by positive SA- $\beta$ -gal staining.

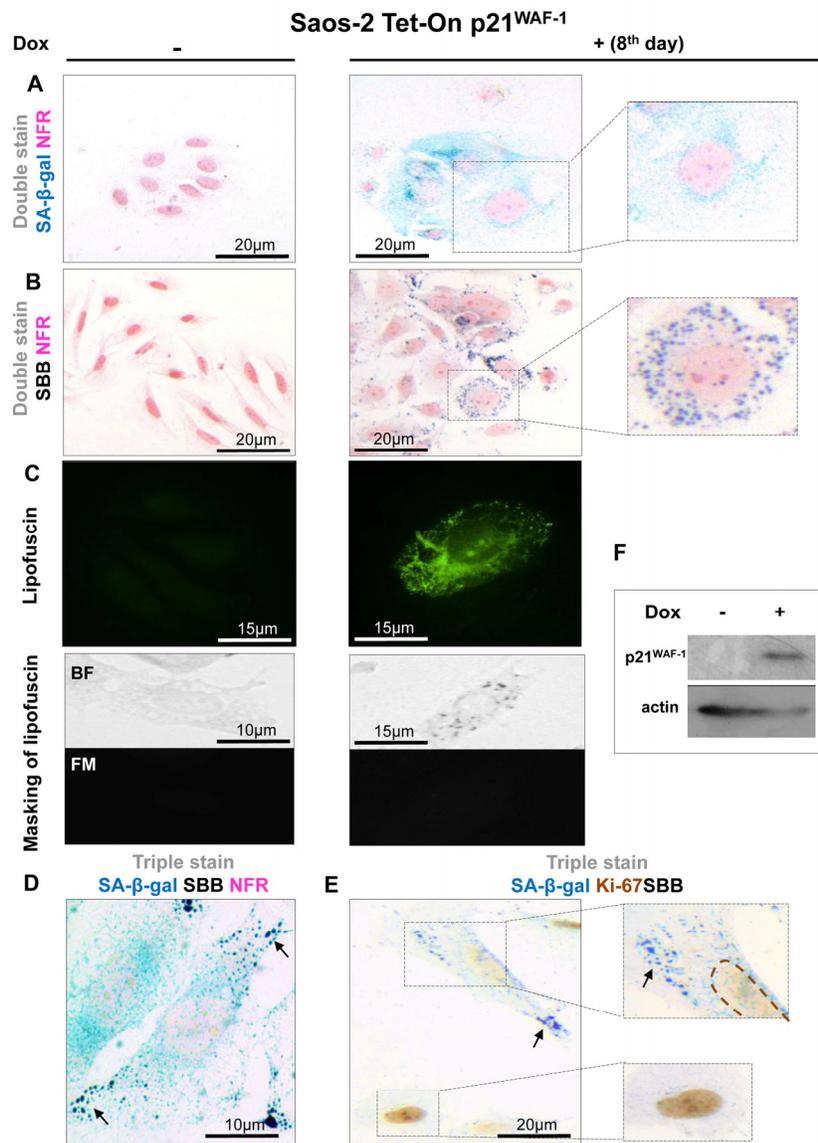
In all the above systems cells that acquired the morphological features of senescence (i.e. became enlarged and flattened) were positive for SA- $\beta$ -gal (Fig. 1A, 2A, 3A, 4A). These preparations were also stained with SBB and as illustrated in Fig. 1B, 2B, 3B and 4B all contained clearly visible perinuclear and cytoplasmic aggregates of lipofuscin, stained by SBB as dark blue-black granules. In addition, as lipofuscin produces auto-fluorescence, we could verify that the blue-black granules, stained by SBB, represented lipofuscin aggregates, by masking the lipofuscin's auto-fluorescence with SBB staining (Fig. 1C, 2C, 3C, 4C) [26]. This unique property of the SBB was the reason it was selected among other methods that detect lipofuscin [19, 27]. To assess the extent of agreement between the results from the two assays, we performed SA- $\beta$ -gal and

SBB co-staining in the same cells. Indeed, co-staining results showed a complete overlap (Fig. 1D, 2D, 3D, 4D). Likewise, concomitant staining of SA- $\beta$ -gal, SBB

and the proliferative marker Ki67 verified that cells positive for lipofuscin and SA- $\beta$ -gal activity were not proliferating (Fig. 1E, 2E, 3E, 4E).



**Figure 1. Lipofuscin accumulates and co-localizes with Senescence-Associated beta-galactosidase (SA- $\beta$ -gal) in sub-confluent senescent primary human diploid lung fibroblasts (DLF).** Y (Young): Early- passage cells, RS: Replicative-senescent cells and IS (irradiated): Early passage cells that became prematurely senescent after irradiation (12x4Gy). Collected cells were fixed on slides with 4% paraformaldehyde (A) All three cultures were stained with SA- $\beta$ -gal and nuclear fast red as counterstain (NFR). Cells from RS and IS cultures acquired the characteristic senescent morphological phenotype (enlarged and flattened) and were positive for SA- $\beta$ -gal staining (turquoise color). (B) All cultures were stained with Sudan Black B (SBB) and NFR. Cells from RS and IS cultures, which had the morphological phenotype of senescence, were also positive for SBB (dark blue-black granules). (C) **Top panels:** green pseudocolor represents visualization of lipofuscin's autofluorescence at 450-490 nm. **Bottom panels:** RS and IS cells that stained with SBB (BF, bright field microscopy) show no auto-fluorescence of lipofuscin (FM, fluorescence microscopy without pseudocolor), indicating that SBB stains lipofuscin. Cells with the morphological phenotype of senescence were positive for both SA- $\beta$ -gal and SBB (D), while cells that were positive for Ki67 were negative for both SA- $\beta$ -gal and SBB (E). Insets: Cells at higher magnification, pink dashed lines: indicate NFR-stained nuclei, brown dashed lines: indicate Ki67- negative nuclei, black arrows: show SBB granules.



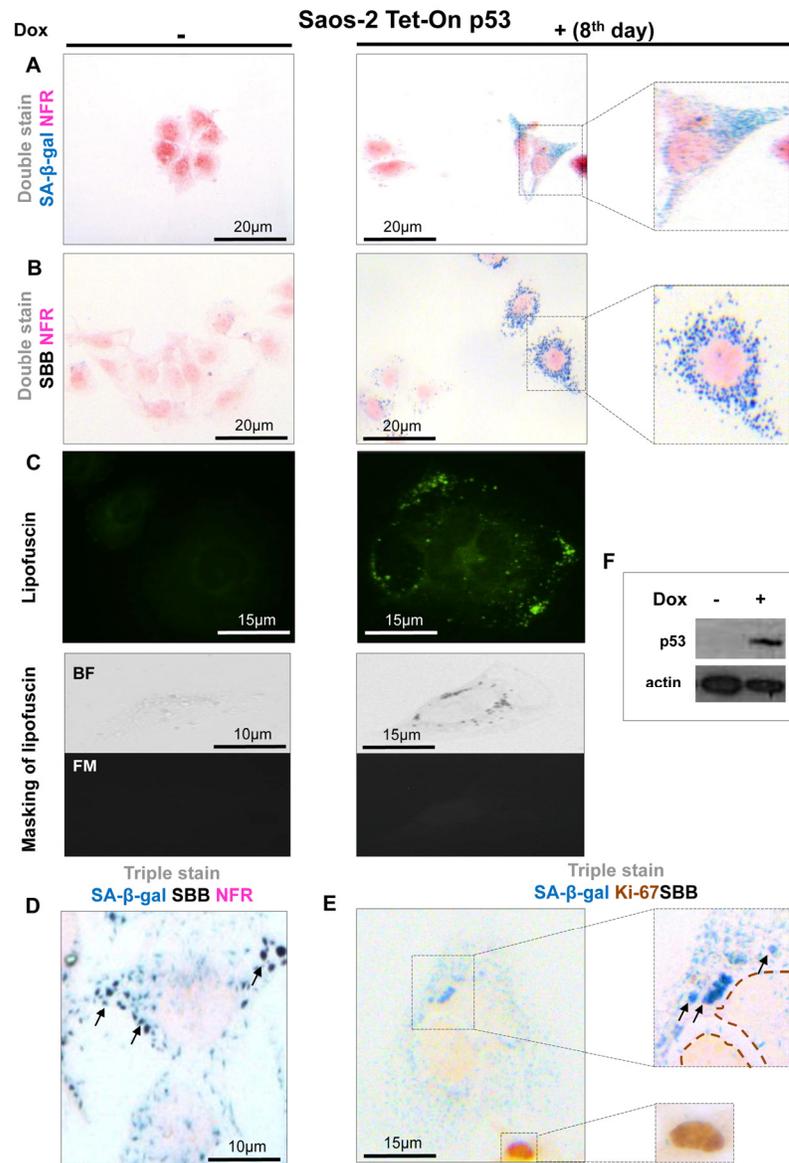
**Figure 2. Lipofuscin accumulates and co-localizes with Senescence-Associated beta-galactosidase (SA-β-gal) in senescent Saos-2 cells triggered by p21<sup>WAF-1</sup>.** (A) SA-β-gal staining (turquoise color) in the Saos-2 p21<sup>WAF-1</sup> Tet-On cell system on the 8<sup>th</sup> day of doxycycline (5 μg/ml) addition. **Inset:** Senescent cells acquired the characteristic senescent morphological phenotype (enlarged and flattened). (B) Sudan Black B (SBB) positivity (dark blue-black granules) in cells with senescent morphological phenotype (**inset**). (C) **Top panels:** Lipofuscin's auto-fluorescence in induced Saos-2 p21<sup>WAF-1</sup> Tet-On cells, by fluorescence microscopy at 450-490 nm (green pseudocolor). **Bottom panels:** Cytochemical SBB staining (BF, bright field microscopy) quenches the auto-fluorescence of lipofuscin (FM, fluorescence microscopy), indicating that SBB stains lipofuscin. SA-β-gal and SBB staining coincided in cells that had the morphological phenotype of senescence (D) and were absent in cells that were positive for the proliferative marker Ki67 (E). (F) Addition of doxycyclin (Dox) triggers p21<sup>WAF-1</sup> expression. Brown dashed lines: Ki67- negative nuclei. Black arrows: SBB granules. NFR: nuclear fast red counterstain.

Next we examined whether SBB staining of lipofuscin could serve as a reliable senescence bio-marker *in vivo*. To this end the same experimental procedure was followed in frozen tissue sections from mice lung adenomas. The lung adenomas were developed in a

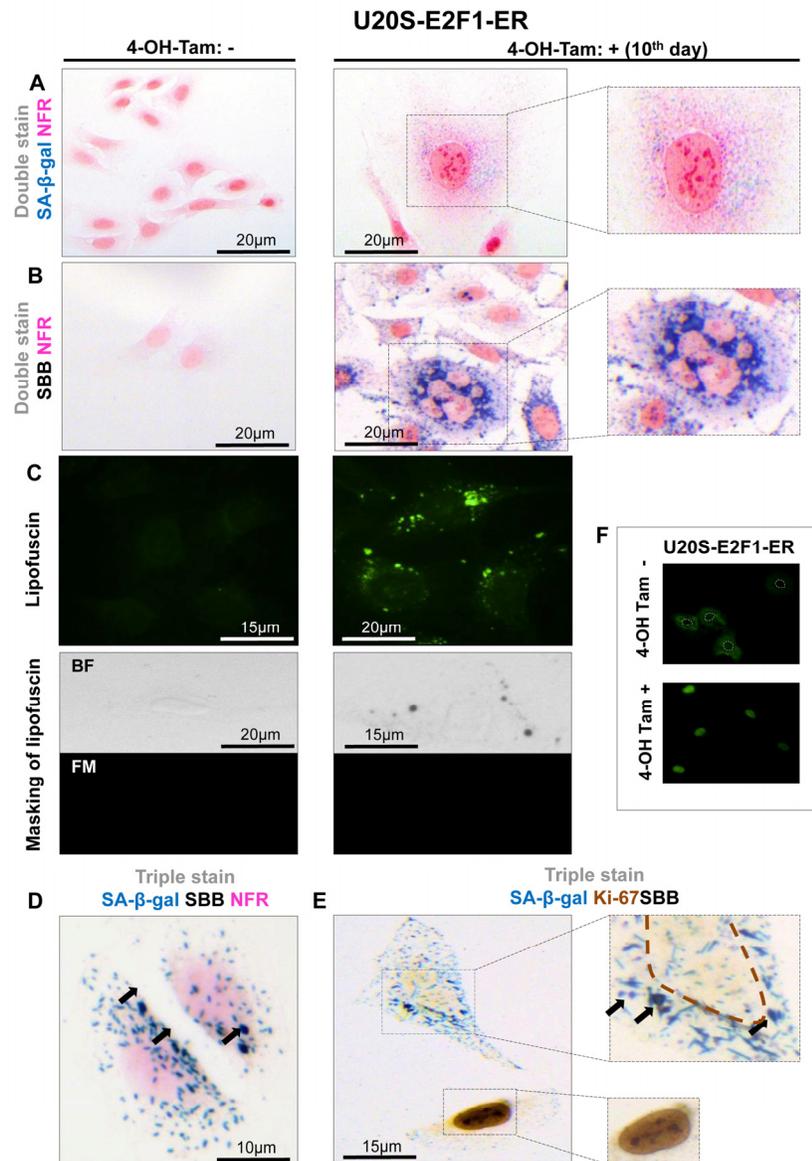
mouse model expressing conditionally *K-rasV12* in the lung [5], one of the first *in vivo* settings used to demonstrate the role of senescence as an anti-tumor barrier in premalignant lesions [5]. In line with the *in vitro* findings, the lung adenomas that demonstrated

strong SA- $\beta$ -gal activity stained positive for lipofuscin while normal lung tissues were negative (Fig. 5). Next, we examined frozen human samples from patients with benign prostatic hyperplasia (BPH) from enlarged

prostates (>55gr). These lesions had been previously shown to feature senescence [7, 28]. As shown in Fig.6, SA- $\beta$ -gal activity and SBB staining co-localized, whereas adjacent normal prostatic glands were negative.



**Figure 3. Lipofuscin accumulates and co-localizes with Senescence-Associated beta-galactosidase (SA- $\beta$ -gal) in p53-mediated Saos-2 senescent cells.** (A) Senescent cells with the characteristic morphology (enlarged and flattened) and positivity for SA- $\beta$ -gal staining (turquoise color), on the 8<sup>th</sup> day of induced with doxycycline of the Saos-2 p53 Tet-On system. (B) Sudan Black B (SBB) perinuclear accumulation as dark blue-black granules, in cells with senescent morphology. (C) **Top panels:** Perinuclear appearance of lipofuscin in apparently senescent cells: pseudocolor visualization of lipofuscin's auto-fluorescence (450-490 nm) is represented in green. **Bottom panels:** Lipofuscin's auto-fluorescence (FM, fluorescence microscopy) is masked by SBB staining (BF, bright field microscopy). (D) Co-localization of SA- $\beta$ -gal and SBB staining in senescent cells and, (E) Ki67 positive cells are negative for SBB and SA- $\beta$ -gal. (F) Addition of 5 $\mu$ g/ml doxycycline (Dox) leads to p53 expression. Brown dashed lines: Ki67- negative nuclei, black arrows: SBB granules, NFR: nuclear fast red counterstain.



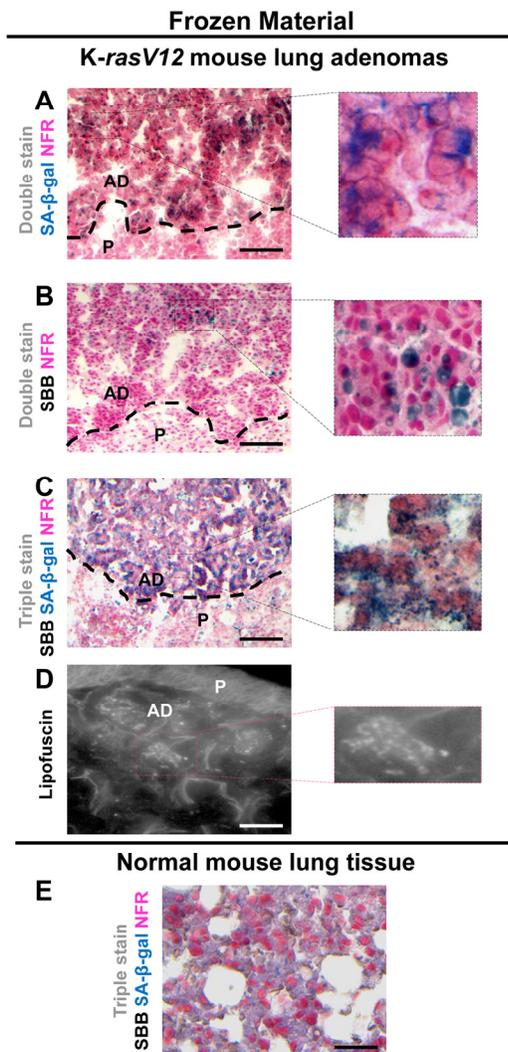
**Figure 4. Lipofuscin accumulates and co-localizes with Senescence-Associated beta-galactosidase (SA-β-gal) in E2F-1 induced U2OS senescent cells.** (A) On the 10<sup>th</sup> day of induction with 4-OH-Tamoxifen, cells were positive for SA-β-gal activity (turquoise color); cells also demonstrated the morphological phenotype of senescence (enlarged and flattened) (B) Cells demonstrating the characteristic senescent phenotype show Sudan Black B (SBB) dark blue-black granules (C) **Top panels:** Lipofuscin's auto-fluorescence at 450-490 nm is represented in green pseudocolor. **Bottom panels:** blocking of lipofuscin auto-fluorescence (FM, fluorescence microscopy) with SBB staining (BF, bright field microscopy) indicates that SBB stains lipofuscin (D) Concurrent positivity for SA-β-gal activity and SBB staining in the same cell, which is also negative for the proliferative marker Ki67 (E). (F) Addition of 300 nmol/L 4-OH-Tamoxifen (4-OH-Tam) leads to nuclear translocation of E2F1 (indirect immunofluorescence). E2F1-negative nuclei are indicated with white dashed lines. Brown dashed lines: Ki67- negative nuclei, black arrows: SBB granules, NFR: nuclear fast red counterstain.

Having verified specific staining of senescent cells by SBB we then asked whether this approach could be also applicable in archival tissues. Thus, SBB stain was performed in FFPE tissue samples prepared from the

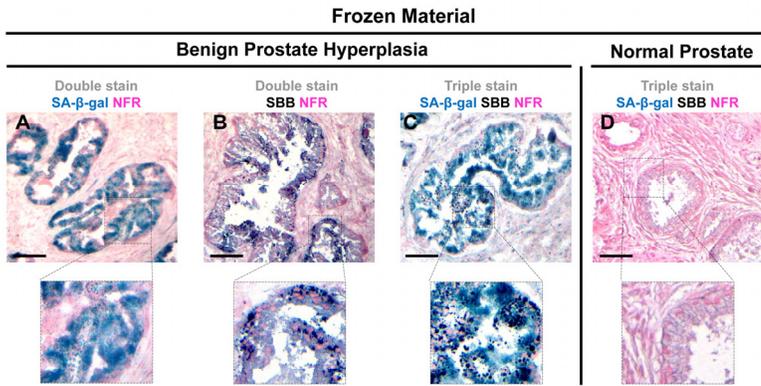
above *in vivo* settings. As demonstrated in Fig.7, SBB staining clearly demonstrated lipofuscin in the lung adenomas (Fig.7). Strikingly, adjacent adenocarcinomas that spontaneously developed in these mice [5] were

negative (Fig. 7). This finding supports the reliability of SBB as a method for staining senescent cells in FFPE (Fig. 7). Also, it is in line with the finding of Collado *et al.*, that spontaneously developing adenocarcinomas in this model bypass the senescence anti-tumor barrier [5]. In addition, as representatively shown in Fig.8, SBB specifically stained blue-black lipofuscin granules in the FFPE material from BPH (see Fig.6) that were characterized as SA- $\beta$ -gal positive. To further validate SA- $\beta$ -gal and SBB staining tissue co-localization, we applied triple staining for SA- $\beta$ -gal, SBB and NFR on BPH samples.

Specifically, the frozen samples were fixed in formaldehyde for 2h, stained with SA- $\beta$ -gal; routinely processed for FFPE as previously described [6], sectioned and then stained with SBB (Fig.9). Although, SA- $\beta$ -gal activity was detected mostly in the periphery of the sections, leaving the core of the tissue unstained (likely due to slow penetration of the SA- $\beta$ -gal stain into the tissue) a clear co-localization of SA- $\beta$ -gal activity and SBB staining was noted close to the tissue periphery marked by pathological features of BPH. Interestingly, SBB foci were also observed in the core of the tissue (Fig. 9).

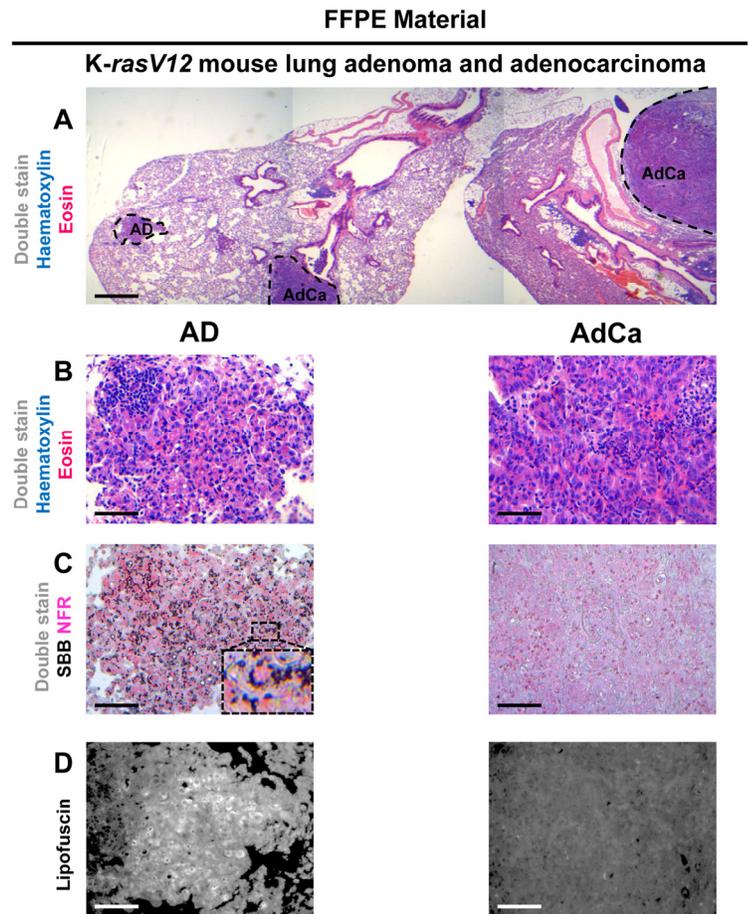


**Figure 5. Lipofuscin and Senescence-Associated beta-galactosidase (SA- $\beta$ -gal) activity co-localize in lung adenomas demonstrating senescence in a mouse model conditionally expressing K-rasV12 in the lung.** Frozen sections derived from mouse lung K-rasV12 adenomas. (A) Cells from the adenomas show SA- $\beta$ -gal activity. (B) Characteristic perinuclear deposition of blue black granules in cells stained with Sudan Black B (SBB), representing positivity for lipofuscin. (C) Cells from lung adenomas positive for both, SA- $\beta$ -gal activity and lipofuscin. (D) Fluorescence microscopy (at 450-490 nm) verifying lipofuscin presence. (E) Normal mouse lung tissue negative for SA- $\beta$ -gal activity and lipofuscin. P: Parenchyma, AD: Adenoma. Scale bars: A-C, 200  $\mu$ m; D, 25  $\mu$ m; E, 50  $\mu$ m. Insets: Cells at higher magnification.



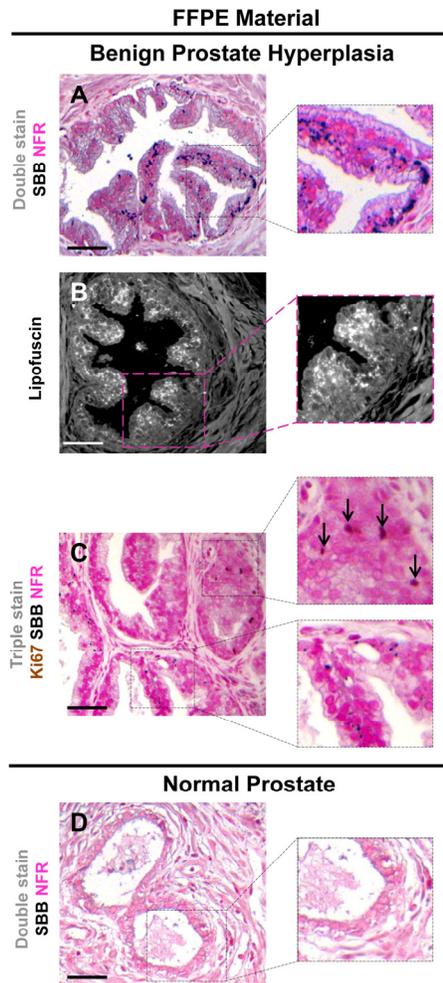
**Figure 6. Lipofuscin accumulates and co-localizes with Senescence-Associated beta-galactosidase (SA-β-gal) in senescent cells detected in cryo-preserved material from benign prostatic hyperplasia (BPH).** Frozen material from patients with BPH in enlarged prostates (prostate weight greater than 55gr) was thin-sectioned (5μm). The sections were immediately double stained for SA-β-gal activity (turquoise color) and Nuclear Fast Red (NFR) as counterstain (A), and double stained with SBB and NFR (B). Areas with characteristic BPH pathology showed SA-β-gal activity and lipofuscin positivity (C). Normal prostate regions adjacent to BPH, were found negative for SA-β-gal activity and lipofuscin (D). Scale bars: A-C, 100 μm; D, 50 μm. Insets: Cells at higher magnification.

**Figure 7. Sudan Black B (SBB) staining demonstrates lipofuscin accumulation in lung adenomas (AD) and absence in adenocarcinomas (AdCa), in formalin-fixed paraffin-embedded (FFPE) lung sections from mice conditionally expressing K-rasV12.** (A) Haematoxylin and Eosin staining demonstrates a lung adenoma (AD) and an adenocarcinoma (AdCa) on the same FFPE section (B) Histological features of the adenoma and the adenocarcinoma shown in A section. (C) Characteristic perinuclear deposition of blue black granules in adenoma cells stained with Sudan Black B (SBB), representing positivity for lipofuscin, while adenocarcinoma cells are negative for SBB. (D) Fluorescence microscopy (at 450-490 nm) verifying lipofuscin presence in the adenoma, and absence in the adenocarcinoma. Scale bars: A, 600 μm; B-D, 100 μm. Inset: Cells at higher magnification.



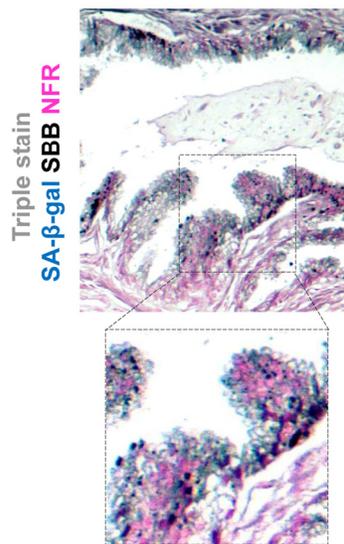
It has been reported that false positive SA- $\beta$ -gal staining may be detected in confluent cultures of quiescent cells [29, 30]. To investigate whether lipofuscin may be present in confluent cultures, we performed the SA- $\beta$ -gal assay and SBB stain in confluent DLF cultures. We observed that as soon as the cells reached confluence they showed SA- $\beta$ -gal activity while, in contrast, such cultures showed only negligible lipofuscin staining (Fig.10A). Extending the two assays for 72 hours we observed that all the cells clearly demonstrated SA-

gal activity, whereas cells containing lipofuscin granules were fewer (Fig. 10B). Furthermore, the enzymatic activity of galactosidase [31] may produce positive results of SA- $\beta$ -gal assay in non-senescent cells, when incubated for prolonged time. We incubated sub-confluent cultures of young DLFs in SA- $\beta$ -gal for 72 hours (prolonged incubation) and then stained immediately with SBB. In these cells we observed SA- $\beta$ -gal activity while there was no detectable SBB staining for lipofuscin (Fig. 10C).



**Figure 8. Accumulation of lipofuscin in formalin-fixed paraffin-embedded (FFPE) tissues from benign prostatic hyperplasia (BPH) that corresponds to senescent areas as depicted by Senescence-Associated beta-galactosidase (SA- $\beta$ -gal) in cryo-preserved material.** FFPE sections from patients with BPH in enlarged prostates (prostate weight greater than 55gr) demonstrate accumulation of lipofuscin. Sections were deparaffinized and double stained with SBB (dark blue-black granules) and NFR as counterstain (A). Lipofuscin's presence was verified with fluorescence microscopy (B). Immunostaining for Ki-67 shows no matching with lipofuscin accumulation (C). Normal prostate regions adjacent to BPH, were negative for lipofuscin (D). Scale bars: BPH, 100  $\mu$ m; Normal Prostate, 50  $\mu$ m. Insets: Cells at higher magnification.

## Benign Prostate Hyperplasia



**Figure 9. Co-localization of Senescence-Associated beta-galactosidase (SA-β-gal) activity and lipofuscin depiction in fresh-frozen tissue sample of benign prostatic hyperplasia (BPH) pretreated with SA-β-gal and subsequently embedded in paraffin.** Fresh samples with BPH were snap frozen, fixed in 4% formaldehyde, washed with buffer, incubated in SA-β-gal solution (turquoise color), subsequently fixed with formaldehyde, and then embedded in paraffin, as previously shown (Michaloglou *et al* 2005). Sections were then double stained with Sudan Black B (SBB) (dark blue-black granules) and Nuclear Fast Red (NFR) as counterstain. Areas with the characteristic pathology of BPH showed SA-β-gal activity and lipofuscin positivity. Note the weak intensity of the Sa-β-gal staining.

## DISCUSSION

In this study, we examined whether lipofuscin, “a hallmark of aging”, [15] is also a cellular senescence biomarker and especially whether it is a marker of SIPS. Our working hypothesis was based on the notion that senescent cells accumulate in aged tissues [2]. We argued that the presence of lipofuscin in senescent cells could complement SA-β-gal by providing and additional marker of cellular senescence. To this end a series of experiments were designed in which various cellular systems, based on normal and cancer cells, were driven to either RS or SIPS. In line with our hypothesis all cells that displayed the senescence phenotype under either RS, or SIPS scenarios demonstrated co-localization of SA-β-gal activity and SBB-detected

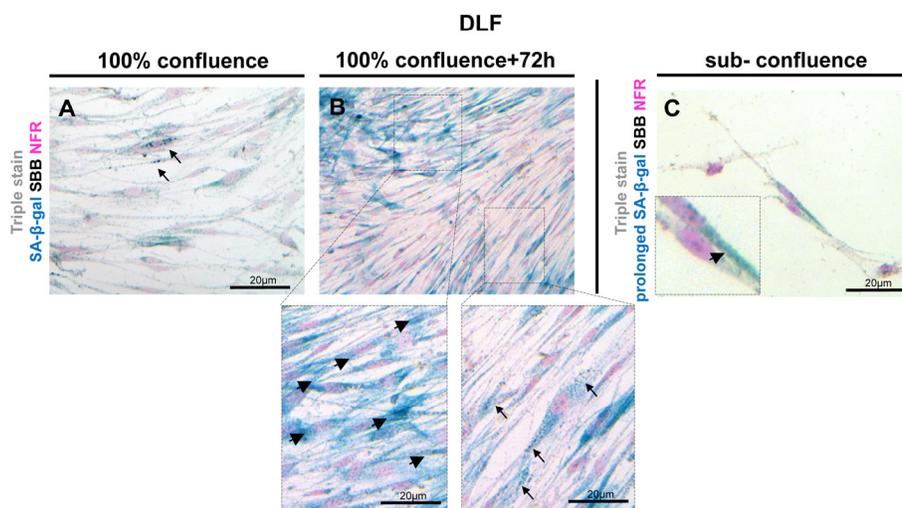
lipofuscin granules, both *in vitro* and *in vivo*, (Fig. 1D, 2D, 3D, 4D, Fig. 1E, 2E, 3E, 4E and Fig. 5C, 6C) whereas tissues negative for SA-β-gal activity were negative for lipofuscin, as well (Fig. 5E, 6D).

Lipofuscin is a non degradable aggregate of oxidized proteins, lipids and metals which accumulates inside the lysosomes of cells that do not replicate [16]. Such accumulation possibly reflects the inability of nonproliferating cells to dispose of lipofuscin by cellular division, a process that naturally results in dilution of lipofuscin [32]. Lipofuscin accumulation in aged tissues and age-related pathologies is considered a progressive phenomenon, probably as a consequence of the decline of the cellular clearance systems of misfolded (lipo)proteins and possibly other ‘aberrant’ metabolites [16, 33, 34]. We observed the lipofuscin granules in cells shortly after they became senescent. This fact implies that lipofuscin formation is possibly related to the senescent state, rather than just a random time-coincident event. Our findings reinforce the observation that senescent cells are hyper-metabolic and full of lipids in the cytoplasm. [35]. As no specific antibody for lipofuscin exists, we have used the SBB histochemical stain to demonstrate the presence of lipofuscin [16]. The dye is diluted in ethanol and due to its lipophilic nature, when in contact with lipids; it assembles on the lipid surface, as it is more soluble in lipids than in ethanol [26]. Accordingly, SBB stains the lipid component of lipofuscin [26]. Several methods were reported to detect lipofuscin, including fluorescence microscopy due to the natural auto-fluorescence of lipofuscin [36], as well as histochemical dyes like SBB, Berlin Blue, Nile Blue, Ziehl-Neelsen and Periodic acid Schiff [37]. As SBB is shown to specifically mask lipofuscin’s fluorescence, it was considered the most reliable histochemical stain to apply [19, 26, 27, 38]. SBB is suitable for use both in frozen and FFPE material [27]. We also verified that SBB staining for lipofuscin may depict senescent cells in FFPE sections from precancerous lesions already shown to contain senescent cells [5, 7, 28] (Fig. 7, 8A-C and Fig.9), while it is negative in the adjacent normal tissue (Fig. 8D) and the related carcinomas (Fig. 7) that have over-come the senescence anti-tumor barrier [5]. These characteristics support the candidacy of SBB for a highly desirable tool to study senescence in archival material.

There is a broad range of potential senescence markers whose reliability, however, varies [39,40]. None of the candidate markers proposed to date, however, is considered entirely specific for cellular senescence, especially for *in vivo* applications [3]. The most commonly used senescence biomarker is SA-β-gal

activity [14]. On the other hand, even the SA- $\beta$ -gal assay apparently produces false positive results, under certain culture conditions such as confluence and serum starvation [29, 30]. Also, cells that do not express the galactosidase gene, show no SA- $\beta$ -gal activity, but fully execute the senescence program [31]. SA- $\beta$ -gal, can only be used in fresh-frozen tissue as it is based on the enzymatic activity of galactosidase [14]. This limitation prevents to use SA- $\beta$ -gal for studies of senescence in FFPE- archival material [1]. Furthermore, if cells or tissues are left in SA- $\beta$ -gal solution for a prolonged time, all cells eventually will acquire the characteristic turquoise stain, due to their normal enzymatic galactosidase activity (Fig. 10C) [31]. To perform the technique a control sample is always required and the process is usually stopped when the sample under investigation starts to stain. As different tissues and cells require different times to stain and the desired stain is selected subjectively by the observer, the density and intensity of the turquoise color considered

as positive varies between studies. On the other hand, SBB is a fast dye and it takes only few minutes to stain lipofuscin in tissues and cells, while the results are homogeneous and reproducible. This technique for identifying senescent cells is therefore easier, more rapid, likely more reproducible, and especially more suited for a wider spectrum of applications in diagnostic pathology laboratories. Moreover, SBB is a technique employable in frozen tissue [27], so it is also ideal for use in parallel with SA- $\beta$ -gal as an additional biomarker of the senescent state. Of note, like all the other senescence biomarkers, lipofuscin is not 100% specific for senescence as it aggregates in degenerative conditions such as macular degeneration [16]. As shown (Fig. 10A, 10B) in confluent cultures, cells containing lipofuscin may be detected. However, in contrast to SA- $\beta$ -gal activity which is immediately demonstrated in confluence and is present in all the cells (Fig. 10A), lipofuscin granules appear later and to a significantly lesser extent (Fig. 10B).



**Figure 10. Lipofuscin staining and Senescence-Associated beta-galactosidase (SA- $\beta$ -gal) activity in primary human diploid lung fibroblasts (DLF).** Triple staining of early passage (6th) DLF cells with Sudan Black B (SBB) (dark blue-black granules), SA- $\beta$ -gal (turquoise color) and Nuclear Fast Red (NFR), as counterstain. (A) Cells that had just reached 100% confluence showed SA- $\beta$ -gal activity with negligible lipofuscin (arrows). (B) In the same assay 72 hours later, some cells demonstrated SA- $\beta$ -gal staining (arrowheads) without lipofuscin, and there were also cells with concurrent SA- $\beta$ -gal and SBB staining (arrows). (C) In sub-confluent DLF cells, prolonged SA- $\beta$ -gal incubation (72 hours) followed by immediate SBB staining, demonstrated SA- $\beta$ -gal activity (inset) without lipofuscin appearance.

The majority of pathologies associated with lipofuscin are age-related diseases [16]. The role of cellular senescence in age-related pathologies and cancer, conditions that are both considered integral components of “aging”, is a widely growing field of biomedical research [39,40]. Future studies will focus on the role of cellular senescence in cancer, especially considering the increasing evidence that senescence is one of major outcomes and a determinant of treatment response in oncology, broadly analogous to significance of apoptosis [13, 41]. Furthermore, there is evidence that presence of senescent cells in certain malignant tumors may be a sign of better prognosis [42]. There is little doubt that a convenient bio-marker applicable in archival tissue material would greatly facilitate research on cellular senescence in cancer. The detection of lipofuscin with SBB could be applied in studies that evaluate the effects of chemotherapy and other antineoplastic treatments.

Taken together, our present study showed that lipofuscin can provide a senescence biomarker comparable to the SA- $\beta$ -gal activity. Detection of lipofuscin with SBB stain can be applied as a stand-alone or auxiliary to SA- $\beta$ -gal technique. Furthermore, SBB may visualize senescent cells in FFPE tissues, the most common form of archival clinical material, thereby extending the applicability of currently available candidate senescence biomarkers to a much wider selection of research topics related to diverse diseases and aging.

## METHODS

Cells and inducible cellular systems. Primary human diploid lung fibroblast cultures (DLFs) were used as previously described [20]. Briefly, cells were grown in DMEM supplemented with 10% FCS (Gibco, AntiSel, Greece), 2 mmol/L L-Glutamine (Gibco, AntiSel, Greece), and 100  $\mu$ g/mL penicillin and streptomycin (Gibco, AntiSel, Greece), respectively, at 37°C and 5% CO<sub>2</sub>. The cells were serially sub-cultured at 1:2 split ratio until replicative senescence (42 passages). Alternatively, early passage fibroblasts (7 passages) were repeatedly exposed, twice a day, to sub-lethal doses (4 Gy) of  $\gamma$ -radiation up to a cumulative dose of approx. 50 Gy (12x4Gy), in a <sup>60</sup>Co gamma source (Gamma Chamber 4000A, Isotope Group, Bhadha Atomic Research Company, Trombay, Bombay, India) at a rate of 8 Gy/min. The cells were subcultured and after additional two weeks, fixed and analyzed. The regulatable osteosarcoma cell models Saos-2 p21 Tet-On, Saos-2 p53 Tet-On and U2OS-E2F1-ER were cultured as previously described [25, 43]. Briefly, these cell lines were maintained under the same culture

conditions as DLFs, except for the use of FBS tetracycline-free medium (Clontech, Lab Supplies, Greece) on Saos2-p21 Tet-On and Saos2-p53 Tet-On systems. The U2OS-E2F1-ER system induction was accomplished with 300nM 4-OH-Tamoxifen, for 10 days. Saos-2 p21 Tet-On and Saos-2 p53 Tet-On systems were induced with 5 $\mu$ g/ml doxycyclin for 8 days before fixation. All cell lines were fixed in 4% paraformaldehyde (5 min, room temperature) then washed with sterile PBS and kept at 4°C until staining.

Human samples and animal models. Tissue samples from lung adenomas and adenocarcinomas from a mouse model expressing conditionally *K-rasV12* in the lung were analyzed [5]. Normal mouse lung tissue was obtained from the Laboratory of Cell Proliferation and Ageing, Institute of Biology, National Centre for Scientific Research ‘Demokritos’. Subsequently, surgically removed material from patients with benign prostate hyperplasia was obtained (with consent of patients according to the National Kapodistrian University of Athens ethical committee guidelines). From each sample, material was partitioned and either stored at -80°C [14] or routinely formalin-fixed and paraffin-embedded (FFPE).

SA- $\beta$ -galactosidase Assay. The activity of SA- $\beta$ -gal in cell cultures and frozen tissues was detected according to Debacq-Chainiaux *et al.* [14]. Cells with cytoplasmic staining were scored as positive. Fresh-frozen, formaldehyde-fixed SA- $\beta$ -gal pre-incubated tissues were stained as described elsewhere [6].

Lipofuscin staining protocol: Sudan Black B (SBB) staining. Combining the protocols of Gatenby *et al.* [17], and Rasmussen, [18] for SBB staining, we achieved optimal lipofuscin visualization in cell cultures and tissues with the following methodology:

Preparation of SBB solution: 0.7gr of SBB (BDH, Vizas, Athens, Greece) was dissolved in 70% ethanol, covered with parafilm and thoroughly stirred overnight at room temperature. Filtered through filter paper and then filtered again through frittered glass filter of medium porosity with suction. Throughout the process, it was important to avoid ethanol evaporation, which results in precipitation of the stain, so the solution was stored in an airtight container.

Staining Procedure: OCT-Frozen-sections mounted onto superfrost slides were fixed in 1% (wt/vol) formaldehyde/PBS for 1 min at room temperature and then washed three times (approx.1 min) at room temperature, with PBS. Sections were then incubated for 5 min in 50% ethanol and then for another 5 min into 70% ethanol. Coverslips with fixed cells were incubated for 2 min in 70% ethanol. Tissue samples

were dewaxed with xylene and dehydrated until 70% ethanol. In order to avoid precipitation of SBB on cells or tissues the following two steps are crucial: 1) a drop from freshly prepared SBB was dropped on a clean slide. The coverslip with the cells or the dehydrated tissue on a slide was placed facing down on the drop of SBB on the slide. The staining was observed under the microscope. The desirable outcome with no precipitation was accomplished by 2-8 minutes. 2) The coverslip or the slide, were carefully lifted and the SBB on the edges of the coverslip or the tissue-slide was wiped out manually from the back and along the edges of the coverslip or the slide with the help of a soft paper. The cells or the tissues were then embedded into 50% ethanol, transferred and washed in distilled water, counterstained with 0.1% Nuclear Fast Red (NFR) (Sigma, BioLine, Athens Greece) for 10 min., and mounted into 40% Glycerol/TBS mounting medium. Lipofuscin staining was considered positive when perinuclear and cytoplasmic aggregates of blue-black granules were evident inside the cells.

**Immunofluorescence.** For indirect immunofluorescence analysis, cells were fixed with 4% paraformaldehyde in PBS and subsequently incubated with the primary antibody anti-E2F1 (1:100) (KH-95, Santa Cruz, Bioanalytica, Athens, Greece), as previously described [25].

**Immunoblotting analysis.** Total protein extraction from cells and SDS-polyacrylamide gel electrophoresis was performed as previously described [25]. The antibodies used were: anti-p21 (1:500) (F-5, Santa Cruz, Bioanalytica, Athens, Greece) and anti-p53 (1:500) (DO-1, Santa Cruz, Bioanalytica, Athens, Greece).

**Auto-fluorescence Detection of Lipofuscin.** The FFPE tissue sections were deparaffinized, hydrated and mounted into 40% glycerol/TBS mounting medium. The cells after fixation were also mounted into the same mounting medium. Lipofuscin auto-fluorescence was then evidenced by excitation at 450-490nm, using a dichromatic mirror at 510nm and a long-pass filter at 515nm [26]. We used the Leica DMRAZ microscope equipped with the Leica DFC350FX camera. This analysis was performed in cultured young and senescent cells as well as in all tissue sections. Lipofuscin auto-fluorescence was quenched with 0.7% SBB in 70% ethanol [38].

## ACKNOWLEDGEMENTS

Financial support was obtained from the Inflammation and Cancer Research in Europe (contract number 223151), INsPiRE (Integrating the Emerging Research

Potential of the University of Athens Cancer Research Group in the European Research Area [REGPOT]; contract number 284460), the Lundbeck Foundation, and the European Commission (projects DDRresponse and Biomedreg).

## Author contribution

EAG, KT, EK equally contributed in the experimental procedures and manuscript preparation. PJFM and MS prepared the K-rasV12 mouse model and participated in data analysis, VZ participated in data analysis, IPT and DK prepared the DLF cellular system and participated in data analysis, JB prepared the U2OS-E2F1 cellular system and participated in data analysis, VGG designed the study supervised all the procedures and writing of the manuscript.

## Conflict of Interest Statement

The authors of this manuscript have no conflict of interests to declare.

## REFERENCES

1. Gorgoulis VG, and Halazonetis T. Oncogene-induced senescence: the bright and dark side of the response. *Curr Opin Cell Biol* 2010; 22:816-827.
2. Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I, Pereira-Smith O, Peacocke M, and Campisi J. A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc Natl Acad Sci U S A*. 1995; 92:9363-9367.
3. Sikora E, Arendt T, Bennett M, and Narita M. Impact of cellular senescence signature on ageing research. *Ageing Res Rev*. 2011; 10:146-152.
4. Hayflick L. The limited in vitro lifespan of human diploid cell strains. *Exp Cell Res*. 1965; 614-636.
5. Collado M, Gil J, Efeyan A, Guerra C, Schuhmacher AJ, Barradas M, Benguría A, Zaballos A, Flores JM, Barbacid M, Beach D, and Serrano M. Tumour biology: senescence in premalignant tumours. *Nature*. 2005; 436:642.
6. Michaloglou C, Vredeveld LC, Soengas MS, Denoyelle C, Kuilman T, van der Horst CM, Majoor DM, Shay JW, Mooi WJ, and Peeper DS. BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature*. 2005; 436:720-724.
7. Chen Z, Trotman LC, Shaffer D, Lin HK, Dotan ZA, Niki M, Koutcher JA, Scher HI, Ludwig T, Gerald W, Cordon-Cardo C, and Pandolfi PP. Crucial role of the p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature*. 2005; 436:725-730.
8. Braig M, Lee S, Loddenkemper C, Rudolph C, Peters AH, Schlegelberger B, Stein H, Dörken B, Jenuwein T, and Schmitt CA. Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature*. 2005; 436:660-665.
9. Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, Issaeva N, Vassiliou LV, Kolettas E, Niforou K, Zoumpourlis VC, Takaoka M, Nakagawa H, Tort F, Fugger K, Johansson F, Sehested

- M, Andersen CL, Dyrskjot L, Ørntoft T, Lukas J, Kittas C, Helleday T, Halazonetis TD, Bartek J, and Gorgoulis VG. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*. 2006; 444:633-637.
10. Gorgoulis VG, Pratsinis H, Zacharatos P, Demoliou C, Sigala F, Asimacopoulos PJ, Papavassiliou AG, and Kletsas D. p53-dependent ICAM-1 overexpression in senescent human cells identified in atherosclerotic lesions. *Lab Invest*. 2005; 85:502-511.
  11. Herbig U FM, Condel L, Carey D, and Sedivy JM. Cellular senescence in aging primates. *Science*. 2006; 311:1257.
  12. Rodier F and Campisi J. Four faces of cellular senescence. *J Cell Biol*. 2011; 192:547-556.
  13. Shay JW, and Roninson IB. Hallmarks of senescence in carcinogenesis and cancer therapy. *Oncogene*. 2004; 23:2919-2933.
  14. Debacq-Chainiaux F, Erusalimsky JD, Campisi J, and Toussaint O. Protocols to detect senescence-associated beta-galactosidase (SA-beta-gal) activity, a biomarker of senescent cells in culture and in vivo. *Nat Protoc*. 2009; 4:1798-1806.
  15. Terman A, and Brunk UT. Lipofuscin. *Int J Biochem Cell Biol*. 2004; 36:1400-1404.
  16. Jung T, Bader N, and Grune T. Lipofuscin: formation, distribution, and metabolic consequences. *Ann N Y Acad Sci*. 2007; 1119:97-111.
  17. Gatenby JB, and Moussa TA. The sudan black B technique in cytology. *J R Microsc Soc*. 1949; 69:72-75.
  18. Rasmussen GL. A method of staining the statoacoustic nerve in bulk with Sudan black B. *Anat Rec*. 1961; 139:465-469.
  19. Jung T, Hohn A, and Grune T. Lipofuscin: detection and quantification by microscopic techniques. *Methods Mol Biol*. 2010; 594:173-193.
  20. Papadopoulou A, and Kletsas D. Human lung fibroblasts prematurely senescent after exposure to ionizing radiation enhance the growth of malignant lung epithelial cells in vitro and in vivo. *Int J Oncol*. 2011; 39:989-999.
  21. Serrano M, Lin AW, McCurrach ME, Beach D, and Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*. 1997; 88:593-602.
  22. Campisi J. Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. *Cell*. 2005; 120:513-522.
  23. Chang BD, Watanabe K, Broude EV, Fang J, Poole JC, Kalinichenko TV, and Roninson IB. Effects of p21Waf1/Cip1/Sdi1 on cellular gene expression: implications for carcinogenesis, senescence, and age-related diseases. *Proc Natl Acad Sci U S A*. 2000; 97:4291-4296.
  24. Kramer DL, Chang BD, Chen Y, Diegelman P, Alm K, Black AR, Roninson IB, Porter CW: Polyamine depletion in human melanoma cells leads to G1 arrest associated with induction of p21WAF1/CIP1/SDI1, changes in the expression of p21-regulated genes, and a senescence-like phenotype. *Cancer Res*. 2001; 61:7754-7762.
  25. Lontos M, Niforou K, Velimezi G, Vougas K, Evangelou K, Apostolopoulou K, Vrtel R, Damalas A, Kontovazenitis P, Kotsinas A, Zoumpourlis V, Tsangaris GT, Kittas C, Ginsberg D, Halazonetis TD, Bartek J, and Gorgoulis VG. Modulation of the E2F1-driven cancer cell fate by the DNA damage response machinery and potential novel E2F1 targets in osteosarcomas. *Am J Pathol*. 2009; 175:376-391.
  26. Romijn HJ, van Uum JF, Breedijk I, Emmering J, Radu I, and Pool CW. Double immunolabeling of neuropeptides in the human hypothalamus as analyzed by confocal laser scanning fluorescence microscopy. *J Histochem Cytochem*. 1999; 47:229-236.
  27. Dako: Special Stains and H & E Education Guide. 2010, Edited by George L. Kumar JAK.
  28. Castro P, Giri D, Lamb D, and Ittmann M. Cellular senescence in the pathogenesis of benign prostatic hyperplasia. *Prostate*. 2003; 55:30-38.
  29. Severino J, Allen RG, Balin S, Balin A, and Cristofalo VJ: Is beta-galactosidase staining a marker of senescence in vitro and in vivo?. *Exp Cell Res*. 2000; 257:162-171.
  30. Yang NC, and Hu ML. The limitations and validities of senescence associated-beta-galactosidase activity as an aging marker for human foreskin fibroblast Hs68 cells. *Exp Gerontol*. 2005; 40:813-819.
  31. Lee BY, Han JA, Im JS, Morrone A, Johung K, Goodwin EC, Kleijer WJ, DiMaio D, and Hwang ES. Senescence-associated  $\beta$ -galactosidase is lysosomal  $\beta$ -galactosidase. *Aging cell*. 2006; 5:187-195.
  32. Brunk UT, and Terman A. The mitochondrial-lysosomal axis theory of aging: accumulation of damaged mitochondria as a result of imperfect autophagocytosis. *Eur J Biochem*. 2002; 269:1996-2002.
  33. Strehler BL, Mark D, and Mildvan D. Rate and magnitude of age pigment accumulation in the human myocardium. *J Gerontol*. 1959; 14:430-439.
  34. Ottis P, Koppe K, Onisko B, Dynin I, Arzberger T, Kretzschmar H, Requena JR, Silva CJ, Huston JP, and Korth C. Human and rat brain lipofuscin proteome. *Proteomics*. 2012;12:2445-2454.
  35. Blagosklonny MV. Tumor suppression by p53 without apoptosis and senescence: conundrum or rapalog-like gerosuppression? *Aging*. 2012;4:450-455.
  36. Dowson JH, and Harris SJ. Quantitative studies of the autofluorescence derived from neuronal lipofuscin. *J Microsc*. 1981; 123:249-258.
  37. Charles C: "Pigments and Minerals", Theory and Practice of Histological Techniques Edited by Bancroft JD GM. Churchill Livingstone, 2002, p. 208.
  38. Viegas MS, Martins TC, Seco F, and do Carmo A. An improved and cost-effective methodology for the reduction of autofluorescence in direct immunofluorescence studies on formalin-fixed paraffin-embedded tissues. *Eur J Histochem*. 2007; 51:59-66.
  39. Jeyapalan J and Sedivy J. Cellular senescence and organismal aging. *Mech Ageing Dev*. 2008; 129 :467-474.
  40. Collado M, and Serrano M. The power and the promise of oncogene-induced senescence markers. *Nat Rev Cancer*. 2006; 6:472-476.
  41. Cairney CJ, Bilsland AE, Evans TR, Roffey J, Bennett DC, Narita M, Torrance CJ, and Keith WN. Cancer cell senescence: a new frontier in drug development. *Drug Discov Today*. 2012; 17:269-276.
  42. Haugstetter AM, Loddenkemper C, Lenze D, Gröne J, Standfuss C, Petersen I, Dörken B, and Schmitt CA.: Cellular senescence predicts treatment outcome in metastasised colorectal cancer. *Br J Cancer* .2010, 103:505-509.
  43. Gorgoulis VG, Vassiliou LV, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, Ditullio RA, Jr., Kastrinakis NG, Levy B, Kletsas D, Yoneta A, Herlyn M, Kittas C, and Halazonetis TD. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature*.2005; 434: 907-913.

## The mitochondria-targeted antioxidant SkQ1 but not N-acetylcysteine reverses aging-related biomarkers in rats

Nataliya G. Kolosova<sup>1,2</sup>, Natalia A. Stefanova<sup>1</sup>, Natalia A. Muraleva<sup>1</sup>, Vladimir P. Skulachev<sup>2,3</sup>

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Mitoengineering, Moscow, Russia

<sup>3</sup> Lomonosov Moscow State University, Belozersky Institute of Physico-Chemical Biology, Moscow, Russia

**Key words:** aging, SkQ1, NAC, growth hormone, IGF-1, testosterone

**Received:** 8/21/12; **Accepted:** 10/27/12; **Published:** 10/28/12

**Correspondence to:** Natalia G. Kolosova, PhD; **E-mail:** [kolosova@bionet.nsc.ru](mailto:kolosova@bionet.nsc.ru)

**Copyright:** © Kolosova et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** Although antioxidants have been repeatedly tested in animal models and clinical studies, there is no evidence that antioxidants reduce already developed age-related decline. Recently we demonstrated that mitochondria-targeted antioxidant 10-(6'-plastoquinonyl) decyltriphenylphosphonium (SkQ1) delayed some manifestations of aging. Here we compared effects of SkQ1 and N-acetyl-L-cysteine (NAC) on age-dependent decline in blood levels of leukocytes, growth hormone (GH), insulin-like growth factor-1 (IGF-1), testosterone, dehydroepiandrosterone (DHEA) in Wistar and senescence-accelerated OXYS rats. When started late in life, supplementation with SkQ1 not only prevented age-related decline but also significantly reversed it. With NAC, all the observed effects were of the lower magnitude compared with SkQ1 (in spite of that dose of NAC was 16000 times higher). We suggest that supplementation with low doses of SkQ1 is a promising intervention to achieve a healthy ageing.

### INTRODUCTION

Aging is commonly defined as progressive deleterious alterations that lead to increased risk of disease and death with advancing age [1-3]. Several potential therapeutic approaches are now available to slow down the age-related functional decline of the organism [4-17] including treatment with antioxidants [1, 18, 19]. Indeed, generation of reactive oxygen species (ROS) by mitochondria is considered as one of mechanisms of aging [20-28]. However, current antioxidants are not selective to mitochondria and that might hamper their effectiveness [29]. Over the course of seven years we investigated retardation of aging with a mitochondria-targeted antioxidant SkQ1 [10-(6'-plastoquinonyl) decyltriphenylphosphonium]. SkQ1 is an antioxidant selectively targeted to mitochondria that protects mitochondria from oxidative damage and which has been shown to decrease mitochondrial damage in animal models of oxidative stress [29, 30]. We have shown that SkQ1 increased the median lifespan of

organisms and also delayed, arrested, and in some cases even reversed development of many age-related pathological traits [29-34].

A complex change of the immune system occurs with aging. Immuno-senescence is defined as decreased cellular reactivity, and imbalance between inflammatory and anti-inflammatory networks, which results in low-grade, chronic, pro-inflammatory condition also known as "inflammaging" [35-39]. Aging affects all immune cells, and it leads to high susceptibility to infections and increased mortality observed in the elderly. Therefore age-related changes in immune cells could serve as a marker of health, biological age and longevity [40].

Here we evaluated the effect of the mitochondria-targeted antioxidant SkQ1 on markers of aging in the old OXYS rats, a unique animal model of accelerated senescence and age-related diseases, as well as normal Wistar rats. OXYS rats spontaneously develop several pathological phenotypes similar to human geriatric

disorders including cataract, retinopathy, osteoporosis, high blood pressure and behavioral alterations [41-50]. Most of these manifestations of accelerated aging develop in OXYS rats between 1 and 3 months of life. Recently we showed that SkQ1 at nanomolar concentrations is capable not only to prevent decline of the immune system and development of cataract and retinopathy but also can reverse already developed pathological changes in the lens and retina of OXYS rats as well as some age-related alterations in behavior [29, 51-53].

For comparison we used N-acetyl-L-cysteine (NAC) as a non-targeted antioxidant. NAC has been shown to prevent age-related cognitive defects and oxidative decline of mitochondrial functions in the brain [54, 55].

## RESULTS

### Body weight

The body weight of 19-month-old rats was measured before the treatment with antioxidants. Rats were randomly assigned to control and experimental groups. OXYS rats are characterized by lower body weight in comparison with Wistar rats. Accordingly, before treatment at the age of 19 months the body weight was dependent only on the genotype ( $F_{1,82} = 472.7$ ,  $p < 0.000$ ) and was lower in OXYS rats (Table 1). At the start of treatment with antioxidants experimental groups did not differ in weight ( $p = 0.34$  for Wistar,  $p = 0.52$  for OXYS). At the end of the 4-month treatment with NAC and SkQ1 the body weight remained lower in OXYS rats ( $F_{1,82} = 245.6$ ,  $p < 0.000$ ) and it was not

affected by the antioxidants ( $F_{2,82} = 1.6$ ,  $p = 0.2$ ). A paired dependent comparison showed that body weight of OXYS rats treated with NAC at the age of 23 months was even lower than their weight at the age of 19 months ( $p < 0.015$ ). OXYS rats treated with SkQ1 showed tendency to lose weight, whereas the body weight of control OXYS rats was not changed by the age of 23 months.

### The white blood cell analyses

In agreement with earlier findings [56], we observed an age-dependent decrease in the number of lymphocytes and an increase in neutrophils (Tabl. 2). The counts of neutrophils in the peripheral blood of Wistar rats at the age of 19 months were higher than those in 3 months old rats ( $p < 0.05$ , comparison of group mean values); and number of neutrophils in 23 months old animals was higher than that in 19 months old rats ( $p < 0.009$ , paired dependent comparisons of the same animals). At the same time the counts of lymphocytes in the blood of Wistar rats decreased ( $p < 0.009$  for 19 months old and  $p < 0.04$  for 23 months old). In OXYS rats, these parameters also varied with age, but the increase in the counts of neutrophils and decrease in lymphocytes were statistically significant only at the age of 19 months compared with the 3-month-old animals ( $p < 0.05$ ). Treatment of Wistar rats with NAC almost completely prevented the drop in lymphocytes/neutrophils ratio, while 4 months treatment with SkQ1 not only prevented its decrease, but actually increased this ratio almost to the level of young rats. In OXYS rats, neither aging from 19 to 23 months nor antioxidants significantly affected the white blood cell counts.

**Table 1.** Body weight (g) of control and SkQ1-or NAC-treated Wistar and OXYS rats at the age of 19 and 23 months

Strain	Wistar			OXYS		
	0	SkQ1	NAC	0	SkQ1	NAC
19 months	612±18	630±18	637±17	404±9*	424±14	411±6
23 months	560±27	642±20	604±27	393±8*	390±8	385±9^

\* – statistically significant differences between the strains of the same age; ^ - statistically significant differences between 19- and 23-month-old rats (paired-dependent comparisons of the same animals).

**Table 2.** White blood cell counts in Wistar and OXYS rats at the age of 3, 19 and 23 months. Effects of 4 months treatment with SkQ1 or NAC. The results are given as % of the total number of white blood cells

Strain	Age, month	Treatment	Lymphocytes	Neutrophils	Eosinophils	Monocytes	Lymphocytes/Neutrophils ratio
Wistar	3	0	74.1±5.8	20.4±0.21	2.13±0.25	4.10±0.25	3.63±0.41
	19	0	63.0±1.32 <sup>^</sup>	29.4±1.27 <sup>^</sup>	3.36±0.29 <sup>^</sup>	3.43±0.20	2.46±0.14 <sup>^</sup>
	23	0	54.5±4.42 <sup>^</sup>	38.6±4.47 <sup>^</sup>	4.50±1.26	2.38±0.34 <sup>^</sup>	1.77±0.29 <sup>^</sup>
		SkQ1	66.9±3.19 <sup>#</sup>	25.9±2.92 <sup>#</sup>	4.00±0.83	3.22±0.28	2.92±0.39 <sup>#</sup>
		NAC	62.4±3.72	31.7±3.79	2.93±0.66	3.00±0.33	2.26±0.34
OXYS	3	0	73.1±6.1	19.36±1.6	2.22±0.21	5.27±0.52	3.80±0.35
	19	0	66.0±0.91 <sup>^</sup>	28.0±0.82 <sup>^</sup>	2.77±0.20	3.0±0.21 <sup>^</sup>	2.54±0.10 <sup>^</sup>
	23	0	64.3±1.78 <sup>*</sup>	30.3±1.91	2.53±0.45 <sup>*</sup>	2.87±0.31	2.28±0.19
		SkQ1	57.1±3.17	38.4±3.05 <sup>#</sup>	1.83±0.53 <sup>#</sup>	2.65±0.31	1.76±0.22
		NAC	58.3±2.14	36.6±2.23	2.23±0.28	2.86±0.18	1.77±0.15

\* – statistically significant difference between the strains of the same age; # – significant effect of the drug within the strain; ^ – significant age-related differences from the previous age within the strain.

### Serum GH levels

The level of GH naturally was decreased with age in rats of both strains albeit more profoundly in OXYS rats than in Wistar rats ( $p < 0.0001$  for age of 3 months,  $p < 0.009$  for 19 months and  $p < 0.008$  for 23 months). Before the treatment with antioxidants GH levels depended only on the genotype ( $F_{1,42} = 60.96$ ,  $p < 0.000$ ) and at the age of 19 months OXYS rats had lower levels of GH (Table 3). At the start of supplementation with antioxidants the experimental groups did not differ in the levels of GH ( $p = 0.49$  for Wistar,  $p = 0.52$  for OXYS).

After 4-month of treatment with NAC or SkQ1 GH levels remained lower in OXYS rats ( $F_{1,42} = 49.6$ ,  $p < 0.000$ ) and were affected by antioxidants ( $F_{2,42} = 8.8$ ,  $p < 0.0008$ ). In both Wistar and OXYS rats treated with SkQ1, GH levels were significantly higher than in control groups of the corresponding strain ( $p < 0.004$  and  $p < 0.02$ , respectively). In Wistar rats treated with NAC - no difference was found. Yet a paired dependent comparison showed that GH levels in Wistar rats treated with SkQ1 and NAC were higher at the age of 23 months than at the age of 19 months ( $p < 0.009$  and  $p < 0.046$ , respectively). A paired dependent comparison showed a small but significant increase in GH levels from the age of 19 months to 23 months in OXYS rats treated with SkQ1 ( $p < 0.000$ ) or NAC ( $p < 0.046$ ). In control OXYS rats GH levels decreased significantly ( $p < 0.02$ ) by the age of 23 months. In

addition, SkQ1 treated OXYS rats and control Wistar rats of the same age (23 months) did not differ in GH levels.

### Serum IGF-I levels

At the age of 3 months, serum IGF-1 levels were maximal in Wistar and OXYS rats and there was no difference in IGF-1 levels between the strains. IGF-1 levels decreased in rats with age, more profoundly in OXYS rats than in Wistar rats ( $p < 0.002$  for the age of 19 months and  $p < 0.048$  for 23 months). Before treatment with antioxidants IGF-1 levels were dependent on the genotype ( $F_{1,42} = 12.3$ ,  $p < 0.002$ ) and at the age of 19 months IGF-1 was lower in OXYS rats than in Wistar rats (Table 3). At the start of treatment with antioxidants experimental groups did not differ in the levels of IGF-1 ( $p = 0.87$  for Wistar and  $p = 0.81$  for OXYS rats).

After 4 months treatment with antioxidants IGF-1 levels were dependent on the genotype, ( $F_{1,42} = 8.1$ ,  $p < 0.008$ ) and were affected by antioxidants ( $F_{2,42} = 19.4$ ,  $p < 0.00$ ). In Wistar rats treated with either SkQ1 or NAC, IGF-1 levels were significantly higher ( $p < 0.0002$  and  $p < 0.014$ , respectively) than in the control group. The paired dependent comparison showed that IGF-1 levels in 23-month-old Wistar rats treated with either NAC or SkQ1 were even higher than they have been at the age of 19 months ( $p = 0.032$ ), while in the control rats IGF-1 was significantly lower ( $p < 0.001$ ).

**Table 3.** Levels of GH, IGF-1, DHEA-S and testosterone (ng/ml) in serum of intact Wistar and OXYS rats at the age of 3, 19 and 23 months and of rats treated with SkQ1 or NAC from the age of 19 months to 23 months

Strain	Age, month	Treatment	GH	IGF-1	Testosterone	DHEA
Wistar	3	0	8.14±0.2	535±62	1.42±0.3	-
	19	0	5.28±0.1+	483±7	0.61±0.04+	0.66±0.01
	23	0	4.98±0.2+	448±12+	0.60±0.07	0.63±0.02+
		SkQ1	5.50±0.1 <sup>#^</sup>	500 ± 9 <sup>#</sup>	0.61±0.05	0.66±0.01
		NAC	5.37±0.2 <sup>^</sup>	486±10 <sup>#</sup>	0.61±0.03	0.66±0.03
OXYS	3	0	6.24±0.3*	498±64	1.99±0.31	-
	19	0	4.57±0.1 <sup>*+</sup>	444±7 <sup>*+</sup>	0.64±0.06+	0.58±0.01*
	23	0	4.27±0.1 <sup>*+</sup>	399±9 <sup>*+</sup>	0.56±0.02	0.55±0.03 <sup>*+</sup>
		SkQ1	4.80±0.1 <sup>#^</sup>	499±9 <sup>#^</sup>	0.58±0.02	0.60±0.02 <sup>^</sup>
		NAC	4.58±0.08 <sup>^</sup>	449±13 <sup>#</sup>	0.57±0.02	0.59±0.02

\* – statistically-significant difference between the strains of the same age; # – significant effect of the drug within the strain; + – significant age-related differences from the previous age within the strain, ^ – statistically-significant difference between the levels before and after treatment with SkQ1 or NAC (paired-dependent comparisons of the same animals).

From the age of 19 to 23 months IGF-1 levels in control group of OXYS rats also decreased significantly ( $p < 0.001$ ). In OXYS rats treated with either SkQ1 or NAC, IGF-1 levels were significantly higher than in the control group ( $p < 0.013$  and  $p < 0.031$ , respectively). The paired dependent comparison showed that in rats treated with NAC, at the age of 23 months IGF-1 levels remained similar to the 19-month-old animals, whereas, in the rats treated with SkQ1 its level increased significantly ( $p < 0.000$ ). In addition, in 23-month-old OXYS rats treated with NAC, IGF-1 levels were similar to those in Wistar control rats of the same age, but in OXYS rats treated with SkQ1, IGF-1 levels were similar to IGF-1 levels of young Wistar rats.

#### Serum testosterone and DHEA levels

ANOVA analyses showed that the level of testosterone in Wistar and OXYS rats was maximal at the age of 3 months and decreased by the age of 19 months in both strains ( $F_{2,62} = 74.7$ ,  $p < 0.00$ ) and remained unchanged by the age of 23 months. Treatment with NAC and SkQ1 had no effect on the testosterone level (Table 3). We did not measure DHEA in three-month old animals. At the age of 19 months, DHEA level was slightly lower in OXYS rats ( $F_{1,42} = 19.2$ ,  $p < 0.0001$ ). In 23 months old Wistar and OXYS rats, level of DHEA differs only slightly from that in 19 months old animals (Table 3).

## DISCUSSION

Our results indicate that when started late in life, treatment with SkQ1 not only prevented age-related decline, but also partially reversed it. Effects of NAC were of the lower magnitude compared to SkQ1, despite the higher dose of NAC used.

One reason for body weight loss in old age is sarcopenia - a gradual decline in muscle mass. After 4 months treatment with antioxidants the body weight of OXYS and Wistar rats decreased only slightly. In SkQ1 treated group weight was very similar to weight of control rats in both rat strains. These findings are consistent with unpublished data of L.E. Bakeeva and V.B. Saprunova, who found that SkQ1 reduced the age-related decline in muscle mass in OXYS and Wistar rats. Another reason for body weight reduction in old age is osteoporosis, which results in increased risk of fractures. Feeding SkQ1 prevented loss of mineral content associated with senile osteoporosis in OXYS rats [34]. In our study the paired dependent comparison showed that body weight decreased significantly ( $p < 0.015$ ) only in OXYS rats treated with NAC.

White blood cells (WBC) counts could serve as a marker of health, biological age and longevity. In humans, lymphocytes increase early in life until age of 16-21 years [56]. Some studies indicate that number of

lymphocyte decreases with age [57-59]. In line with these findings, we observed that the lymphocyte/neutrophil ratio was high in the 3-month-old Wistar and OXYS rats and decreased with age: the counts of lymphocytes were decreased and neutrophils were increased. Lymphocytes are important effector cells and therefore their activation is essential for immune responses [57]. Diminished lymphocyte production and function are major contributors to disease in elderly [56]. 4 months treatment with NAC almost completely prevented the decrease in ratio between lymphocytes and neutrophils in 23-month-old Wistar rats compared with 19-month-old ones. SkQ1 increased the lymphocyte/neutrophil ratio, and thereby partially reversed decline of this parameter, which was observed at the age of 19 months. Our present results in Wistar rats are in line with the previous reports that NAC [60, 61] and SkQ1 [32,58] prevent the age-linked decrease in lymphocyte level in mice. However, both antioxidants did not affect the blood lymphocyte/neutrophil ratio in OXYS rats. The age-related decrease in lymphocytes is a consequence of involution of thymus, the major organ of lymphocyte maturation. The OXYS rats exhibit accelerated involution of the thymus and SkQ1 reduces age-related thymic involution in both OXYS and normal Wistar rats [52]. It is possible that lack of SkQ1 and NAC effects on the WBC count in old OXYS rats is associated with impairment in bone marrow hematopoiesis. We have previously reported the age-associated changes in the functional status of hematopoietic stem cells in OXYS rats [62]. It can be assumed that in OXYS rats the lymphocyte/neutrophil ratio was already stabilized at low level in the 19 month-old rats so that antioxidants could not have a favorable effect. Antioxidants had no effect on the levels of eosinophils and monocytes in both rat strains.

Circulating GH levels are at the highest during the neonatal period; they decrease during childhood, peak again during puberty and fall dramatically in the elderly [63]. A reduction in GH level in older humans and rodents correlates with a decline in serum levels of an anabolic mediator IGF-1 [64]. The present study confirmed an age-dependent GH and IGF-1 decrease in rats of both strains. In addition, we showed that the serum GH level in all studied groups of OXYS rats was lower than in age-matched Wistar rats. Interstrain differences were highest in the three-month old animals (23%), while at the age of 19 and 23 months difference was 13% and 14%, respectively. There were no interstrain differences in the blood levels of IGF-1 in the 3-month-old animals but at the ages of 19 and 23 months they were slightly (but statistically significant) reduced in OXYS rats (by 9 and 11%, respectively).

Our study also showed for the first time that NAC supplementation from the age 19 to 23 months fully prevented the GH and IGF-I decline in both Wistar and OXYS rats. SkQ1 not only stopped the decline in hormone levels between the ages of 19 and 23 months, but it also increased the levels of GH and IGF-I above the levels of those found in 19 month-old animals (Table 3). It is well known that GH/IGF-I plays an important role in brain aging [65, 66]. Age-related reduction in the activities of somatotrophic axis may influence brain function in the elderly [67]. Recently we have shown that SkQ1 treatment of the middle-aged (12 month) Wistar and OXYS rats had beneficial effects on the locomotor and exploratory activity. SkQ1 also decreased anxiety compared to age-matched controls as well as significantly improved visual ability of the OXYS rats, which suffered from retinopathy and cataract [49]. In the present study, we observed a positive effect of SkQ1 and NAC on the behavior of rats of both strains (data are not shown). In the last series of experiments, we studied effect of SkQ1 on the levels of growth hormone and IGF-1 in 3-month-old rats (data are not shown). We found that SkQ1 caused small (about 25%) but statistically valid increase in the blood hormone level. In addition, SkQ1 prevented the development of retinopathy and cataract and had beneficial effects on behavior, learning ability and memory of OXYS rats. We suggest that the recovery of GH - IGF-I in old age to the levels of those in young age can have a positive impact on the function of the aging brain and the immune system.

At the age between 3 and 19 months, the testosterone levels fell and then remained unchanged between 19 and 23 months in rats of both strains. It was not surprising that there was no interstrain difference in the testosterone level even in old animals. As was previously shown in our group, OXYS males demonstrate an early decrease in sexual motivation; however a decrease in hormonal component of sexual behavior was not detected in aged OXYS males [68]. Recently we showed that SkQ1 is effective not only in preventing but also in reducing already developed age-related decline in male sexual behavior [69]. In the present study, we did not evaluate the sexual behavior, and neither SkQ1 nor NAC treatments affected testosterone levels in period between 19 and 23 months (Table 3). Noteworthy, NAC partially inhibits the mTOR (Target of Rapamycin) pathway [70]. Given the involvement of mTOR in cellular and organismal aging as well as age-related diseases [71-78], slight inhibition of mTOR may contribute to the therapeutic effects of this non-selective antioxidant.

DHEA and its sulfate-bound form (DHEAS) are important precursors of sex steroid hormones. Structure of DHEA is similar to testosterone and levels of both hormones reach their maximal levels in puberty and decrease dramatically with age [79]. Given its multiple metabolic effects, this decline in DHEA levels has been thought to play a role in the aging process [80]. In this study, we can assume that the DHEA levels are maximal in 3-month-old Wistar and OXYS rats. At the age between 19 and 23 months, DHEA levels decreased and SkQ1 increased DHEA only slightly (by 5%).

Deficiencies in multiple hormones are a biomarker of health status in older persons [79].

Here we conclude that SkQ1 not only prevented age-associated hormonal alterations but partially reversed them. These results suggest that supplementation with low doses of SkQ1, even in chronologically and biologically aged subjects seem to be a promising strategy to maintain health and retard the aging process.

## MATERIALS AND METHODS

**Animals and diet.** Male senescence-accelerated OXYS and age-matched male Wistar rats were obtained from the Breeding Experimental Animal Laboratory of the Institute of Cytology and Genetics (ICG), Siberian Division of the Russian Academy of Sciences (Novosibirsk, Russia). All the experiments on rats were carried out according to Animal Care Regulations of ICG Institute of Cytology and Genetics, Novosibirsk. The OXYS rat strain was established based on Wistar rat strain at the Institute of Cytology and Genetics as described earlier [51, 52] and registered in the Rat Genome Database (<http://rgd.mcw.edu/>). At the age of 4 weeks, the pups were taken away from their mothers and housed in groups of five animals per cage (57×36×20 cm) and kept under standard laboratory conditions (at 22±2°C, 60% relative humidity, and natural light), provided with a standard rodent feed, PK-120-1, Ltd. (Laboratorsnab, Russia), and given water ad libitum.

Starting from the age of 19 months OXYS and Wistar rats were randomly assigned to three groups (n = 17-24): control diet, diet supplemented with 250 nmol SkQ1 (synthesized as described earlier [33]) or 650 mg NAC (MP Biomedicals, LLC, France) per kg of body weight per day. The weight was measured before the start of treatment and at the end of experiment.

**Hormone levels and the white blood cell counts.** The levels of GH, IGF-1, testosterone and DHEA-S in the blood serum were analyzed before and after treatment

with SkQ1 or NAC in OXYS and Wistar rats (two times total for each rat) and compared with intact 3-month-old rats (n=10) of the same strains. GH, DHEA were measured by ELISA (Creative Diagnostics, USA). IGF-1 was measured by ELISA (ALPCO Diagnostics, Salem, NH). Testosterone was measured by ELISA (JSC Vector-Best, Russia). The assays were run using the manufacturer's instructions. For WBC counts peripheral blood was drawn from the tail vein. Blood smears were fixed in methanol and subsequently stained with Wright-Giemsa.

**Statistical analysis.** The data were analyzed using repeated measures ANOVA and nonparametric tests with the statistical package Statistica 6.0. Two-way ANOVA was used to evaluate the differences between rat strains (genotypes) and effects of treatment (antioxidants). To validate the effect of the diets on parameters, the genotype and antioxidants were chosen as independent variables. A Newman-Keuls *post hoc* test was applied to significant main effects and interactions in order to estimate the differences between particular sets of means. One-way ANOVA was used for individual group comparisons. Data are represented as mean ± S.E.M. Comparisons between means were analyzed with one-way or repeated measures analysis of variance (ANOVA). Results were considered statistically significant if p value was less than 0.05.

## ACKNOWLEDGEMENTS

This work was supported by grants from Russian Foundation for Basic Research (projects N 11-04-00666 and N 12-04-00091).

## Conflict of Interest Statement

The authors of this manuscript have no conflict of interests to declare.

## REFERENCES

1. Harman D. Aging: a theory based on free radical and radiation chemistry. *J Gerontol.* 1956; 11: 298-300.
2. Harman D. The free radical theory of aging. *Antioxid Redox Signal.* 2003; 5: 557-561.
3. Skulachev VP. Aging as a particular case of phenoptosis, the programmed death of an organism (a response to Kirkwood and Melov "On the programmed/non-programmed nature of ageing within the life history"). *Aging (Albany NY).* 2011; 3: 1120-1123.
4. Anisimov VN, Piskunova TS, Popovich IG, Zabezhinski MA, Tyndyk ML, Egormin PA, Yurova MV, Rosenfeld SV, Semenchenko AV, Kovalenko IG, Poroshina TE, Berstein LM. Gender differences in metformin effect on aging, life span and spontaneous tumorigenesis in 129/Sv mice. *Aging (Albany NY).* 2010; 2: 945-958

5. Anisimov VN, Berstein LM, Popovich IG, Zabezhinski MA, Egormin PA, Piskunova TS, Semenchenko AV, Tyndyk ML, Yurova MN, Kovalenko IG, Poroshina TE. If started early in life, metformin treatment increases life span and postpones tumors in female SHR mice. *Aging (Albany NY)*. 2011; 3: 148-157.
6. Anisimov VN, Zabezhinski MA, Popovich IG, Piskunova TS, Semenchenko AV, Tyndyk ML, Yurova MN, Rosenfeld SV, Blagosklonny MV. Rapamycin increases lifespan and inhibits spontaneous tumorigenesis in inbred female mice. *Cell Cycle*. 2011; 10: 4230-4236.
7. Baur JA, Sinclair DA. Therapeutic potential of resveratrol: the in vivo evidence. *Nat Rev Drug Discov*. 2006; 5: 493-506.
8. Burks TN, Cohn RD. One size may not fit all: anti-aging therapies and sarcopenia. *Aging (Albany NY)*. 2011; 3: 1142-1153.
9. Chateau MT, Araiz C, Descamps S, Galas S. Klotho interferes with a novel FGF-signalling pathway and insulin/Igf-like signalling to improve longevity and stress resistance in *Caenorhabditis elegans*. *Aging (Albany NY)*. 2010; 2: 567-581.
10. Goldberg AA, Richard VR, Kyryakov P, Bourque SD, Beach A, Burstein MT, Glebov A, Koupaki O, Boukh-Viner T, Gregg C, Juneau M, English AM, Thomas DY, Titorenko VI. Chemical genetic screen identifies lithocholic acid as an anti-aging compound that extends yeast chronological life span in a TOR-independent manner, by modulating housekeeping longevity assurance processes. *Aging (Albany NY)*. 2010; 2: 393-414.
11. Guarente L. Sirtuins in aging and disease. *Cold Spring Harb Symp Quant Biol*. 2007; 72: 483-488.
12. Hofseth LJ, Singh UP, Singh NP, Nagarkatti M, Nagarkatti PS. Taming the beast within: resveratrol suppresses colitis and prevents colon cancer. *Aging (Albany NY)*. 2010; 2: 183-184.
13. Khanna A, Kapahi P. Rapamycin: killing two birds with one stone. *Aging (Albany NY)*. 2011; 3: 1043-1044.
14. Menendez JA, Cufi S, Oliveras-Ferraro C, Vellon L, Joven J, Vazquez-Martin A. Gerosuppressant metformin: less is more. *Aging (Albany NY)*. 2011; 3: 348-362.
15. Morselli E, Galluzzi L, Kepp O, Criollo A, Maiuri MC, Tavernarakis N, Madeo F, Kroemer G. Autophagy mediates pharmacological lifespan extension by spermidine and resveratrol. *Aging (Albany NY)*. 2009; 1: 961-970.
16. Moskalev A, Shaposhnikov M. Pharmacological inhibition of NF-kappaB prolongs lifespan of *Drosophila melanogaster*. *Aging (Albany NY)*. 2011; 3: 391-394.
17. Wood JG, Rogina B, Lavu S, Howitz K, Helfand SL, Tatar M, Sinclair D. Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature*. 2004; 430: 686-689.
18. Rattan SIS. Anti-ageing strategies: prevention or therapy. *EMBO Rep*. 2005; 6: S25-29.
19. Tong JJ, Schriener SE, McCleary D, Day BJ, Wallace DC. Life extension through neurofibromin mitochondrial regulation and antioxidant therapy for neurofibromatosis-1 in *Drosophila melanogaster*. *Nat Genet*. 2007; 39: 476-485.
20. Dröge W, Schipper HM. Oxidative stress and aberrant signaling in aging and cognitive decline. *Aging Cell*. 2007; 6: 361-370.
21. Guachalla LM, Rudolph KL. ROS induced DNA damage and checkpoint responses: influences on aging? *Cell Cycle*. 2010; 9: 4058-4060.
22. Hwang AB, Lee SJ. Regulation of life span by mitochondrial respiration: the HIF-1 and ROS connection. *Aging (Albany NY)*. 2011; 3: 304-310.
23. Lindner AB, Demarez A. Protein aggregation as a paradigm of aging. *Biochim Biophys Acta*. 2009; 1790: 980-996.
24. Pani G. P66SHC and ageing: ROS and TOR? *Aging (Albany NY)*. 2010; 2: 514-518.
25. Papaconstantinou J, Hsieh CC. Activation of senescence and aging characteristics by mitochondrially generated ROS: how are they linked? *Cell Cycle*. 2010; 9: 3831-3833.
26. Sanz A, Fernandez-Ayala DJ, Stefanatos RK, Jacobs HT. Mitochondrial ROS production correlates with, but does not directly regulate lifespan in *Drosophila*. *Aging (Albany NY)*. 2010; 2: 200-223.
27. Vigneron A, Vousden KH. p53, ROS and senescence in the control of aging. *Aging (Albany NY)*. 2010; 2: 471-474.
28. Witte ME, Geurts JJ, de Vries HE, van der Valk P, van Horssen J. Mitochondrial dysfunction: a potential link between neuroinflammation and neurodegeneration? *Mitochondrion*. 2010; 10: 411-418.
29. Skulachev VP, Anisimov VN, Antonenko YN, Bakeeva LE, Chernyak BV, Elichev VP, Filenko OF, Kalinina NI, Kapelko VI, Kolosova NG, Kopnin BP, Korshunova GA, Lichinitser MR, Obukhova LA, Pasyukova EG, Pisarenko OI, Roginsky VA, Ruuge EK, Senin II, Severina II, Skulachev MV, Spivak IM, Tashlitsky VN, Tkachuk VA, Vyssokikh MY, Yaguzhinsky LS, Zorov DB. An attempt to prevent senescence: a mitochondrial approach. *Biochim Biophys Acta*. 2009; 1787: 437-461.
30. Skulachev MV, Antonenko YN, Anisimov VN, Chernyak BV, Cherepanov DA, Chistyakov VA, Egorov MV, Kolosova NG, Korshunova GA, Lyamzaev KG, Plotnikov EY, Roginsky VA, Savchenko AY, Severina II, Severin FF, Shkurat TP, Tashlitsky VN, Shidlovsky KM, Vyssokikh MY, Zamyatnin AA, Zorov DB, Skulachev VP. Mitochondrial-targeted plastoquinone derivatives. Effect on senescence and acute age-related pathologies. *Curr Drug Targets*. 2011; 12: 800-826.
31. Anisimov VN, Bakeeva LE, Egormin PA, Filenko OF, Isakova EF, Manskikh VN, Mikhelson VM, Panteleeva AA, Pasyukova EG, Piliipenko DI, Piskunova TS, Popovich IG, Roshchina NV, Rybina OY, Saprunova VB, Samoylova TA, Semenchenko AV, Skulachev MV, Spivak IM, Tsybul'ko EA, Tyndyk ML, Vyssokikh MY, Yurova MN, Zabezhinsky MA, Skulachev VP. Mitochondria-targeted plastoquinone derivatives as tools to interrupt execution of the aging program. 5. SkQ1 prolongs lifespan and prevents development of traits of senescence. *Biochemistry (Mosc)*. 2008; 73: 1329-1342.
32. Anisimov VN, Egorov MV, Krasilshchikova MS, Lyamzaev KG, Manskikh VN, Moshkin MP, Novikov EA, Popovich IG, Rogovin KA, Shabalina IG, Shekarova ON, Skulachev MV, Titova TV, Vygodin VA, Vyssokikh MY, Yurova MN, Zabezhinsky MA, Skulachev VP. Effects of the mitochondria-targeted antioxidant SkQ1 on lifespan of rodents. *Aging (Albany NY)*. 2011; 3: 1110-1119.
33. Antonenko YuN, Avetisyan AV, Bakeeva LE, Chernyak BV, Chertkov VA, Domnina LV, Ivanova OYu, Izyumov DS, Khailova LS, Klishin SS, Korshunova GA, Lyamzaev KG. Mitochondria-targeted plastoquinone derivatives as tools to interrupt execution of an aging program. 1. Cationic plastoquinone derivatives: synthesis and in vitro studies. *Biochemistry (Moscow)*. 2008; 73: 1273-1287
34. Neroev VV, Archipova MM, Bakeeva LE, Fursova AZh, Grigorian EN, Grishanova AY, Iomdina EN, Ivashchenko ZhN, Katargina LA, Khoroshilova-Maslova IP, Kilina OV, Kolosova NG, Kopenkin EP, Korshunov SS, Kovaleva NA, Novikova YP, Philippov

- PP, Pilipenko DI, Robustova OV, Saprunova VB, Senin II, Skulachev MV, Sotnikova LF, Stefanova NA, Tikhomirova NK, Tsapenko IV, Shchipanova AI, Zinovkin RA, Skulachev VP. Mitochondriatargeted plastoquinone derivatives as tools to interrupt execution of the aging program. 4. Agerelated eye disease. SkQ1 returns vision to blind animals. *Biochemistry (Moscow)*. 2008; 73: 1317-1328.
35. Franceschi C, Capri M, Monti D, Giunta S, Olivieri F, Sevini F, Panourgia MP, Laura Invidia L, Celani L, Scurti S, Cevenini E, Castellani GC, Salvioli S. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech. Ageing Dev.* 2007; 128: 92-105.
36. Lisanti MP, Martinez-Outschoorn UE, Pavlides S, Whitaker-Menezes D, Pestell RG, Howell A, Sotgia F. Accelerated aging in the tumor microenvironment: connecting aging, inflammation and cancer metabolism with personalized medicine. *Cell Cycle*. 2011; 10: 2059-2063.
37. Pavlides S, Tsirigos A, Vera I, Flomenberg N, Frank PG, Casimiro MC, Wang C, Pestell RG, Martinez-Outschoorn UE, Howell A, Sotgia F, Lisanti MP. Transcriptional evidence for the "Reverse Warburg Effect" in human breast cancer tumor stroma and metastasis: similarities with oxidative stress, inflammation, Alzheimer's disease, and "Neuron-Glia Metabolic Coupling". *Aging (Albany NY)*. 2010; 2: 185-199.
38. Weiss EP, Villareal DT, Fontana L, Han DH, Holloszy JO. Dehydroepiandrosterone (DHEA) replacement decreases insulin resistance and lowers inflammatory cytokines in aging humans. *Aging (Albany NY)*. 2011; 3: 533-542.
39. Ye J, Keller JN. Regulation of energy metabolism by inflammation: a feedback response in obesity and calorie restriction. *Aging (Albany NY)*. 2010; 2: 361-368.
40. Wessels I, Jansen J, Rink L, Uciechowski P. Immunosenescence of polymorphonuclear neutrophils. *ScientificWorld Journal*. 2010; 10: 145-160.
41. Agafonova IG, Rjtelnikov VN, Mishchenko NP, Kolosova NG. Comparative study of the influence histochrome and mexidol on structural and functional characteristics of premature aging of the brain of rats OXYS by MRI. *Bull Exp Biol Med*. 2010; 151: 380-384.
42. Kolosova NG, Stefanova NA, Sergeeva SV. OXYS rats: a prospective model for evaluation of antioxidant availability in prevention and therapy of accelerated aging and age-related cognitive decline: *Handbook of Cognitive Aging: Causes, Proceses., Gariépy Q and Ménard R, eds. Nova Science Publishers, NY. 2009; 47-82.*
43. Kolosova NG, Akulov, AE, Stefanova NA, Moshkin MP, Savelov AA, Koptyug IV, Panov AV, Vavilin VA. Effect of malate on the development of rotenone-induced brain changes in Wistar and OXYS rats: An MRI study. *Dokl Biol Sci*. 2011; 437: 72-75.
44. Loskutova LV, Kolosova NG. Emotional state and one trial learning in OXYS rats with hereditarily elevated production of oxygen radicals. *Bull Exp Biol Med*. 2000; 130: 746-748.
45. Loskutova LV, Zelenkina LM. Impairment of latent inhibition in OXYS rats with hereditary syndrome of premature aging. *Zh Vyssh Nerv Deiat Im I P Pavlova*. 2002; 52: 366-370.
46. Markova EV, Obukhova LA, Kolosova NG. Activity of cell immune response and open field behavior in Wistar and OXYS rats. *Bull Exp Biol Med*. 2003; 136: 377-379.
47. Markova EV, Kozlov VA, Trofimova NA, Kolosova NG. Stimulation of cell component of the immune response activates exploratory behavior in senescence accelerated OXYS rats. *Bull Exp Biol Med*. 2005; 140: 345-347.
48. Sergeeva S, Bagryanskaya E, Korbolina E, Kolosova N. Development of behavioral dysfunctions in accelerated-senescence OXYS rats is associated with early postnatal alterations in brain phosphate metabolism. *Exp Gerontol*. 2006; 41: 141-150.
49. Stefanova NA, Fursova AZh, Kolosova NG. Behavioral Effects Induced by Mitochondria-Targeted Antioxidant SkQ1 in Wistar and Senescence-Accelerated OXYS Rats. *J. Alzheimers Dis*. 2010; 21: 479-491.
50. Stefanova NA, Fursova AZ, Sarsenbaev KN, Kolosova NG. Effects of *Cistanche deserticola* on behavior and signs of cataract and retinopathy in senescence-accelerated OXYS rats. *J. Ethnopharmacol*. 2011; 138: 624-632.
51. Markovets AM, Saprunova VB, Zhdankina AA, Fursova AZh, Bakeeva LE, Kolosova NG. Alterations of retinal pigment epithelium cause AMD-like retinopathy in senescence-accelerated, XYS rats. *Aging (Albany NY)*. 2011; 3: 44-54.
52. Obukhova LA, Skulachev VP, Kolosova NG. Mitochondria-targeted antioxidant SkQ1 inhibits age-dependent involution of the thymus in normal and senescence-prone rats. *Aging (Albany NY)*. 2009; 1: 389-401.
53. Skulachev VP. SkQ1 treatment and food restriction--two ways to retard an aging program of organisms. *Aging (Albany NY)*. 2011; 3: 1045-1050.
54. Farr SA, Poon HF, Dogrukol-Ak D, Drake J, Banks WA, Eyerman E, Butterfield DA, Morley JE. The antioxidants alpha-lipoic acid and N-acetylcysteine reverse memory impairment and brain oxidative stress in aged SAMP8 mice. *J Neurochem*. 2003; 84: 1173-1183.
55. Kondratov RV, Vykhoanets O, Kondratova AA, Antoch MP. Antioxidant N-acetyl-L-cysteine ameliorates symptoms of premature aging associated with the deficiency of the circadian protein BMAL1. *Aging (Albany NY)*. 2009; 1: 979-987.
56. Cho RH, Sieburg HB, Muller-Sieburg CE. A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. *Blood*. 2008; 111: 5553-5561.
57. De la Fuente M, Baeza I, Guayerbas N, Puerto M, Castillo C, Salazar V, Ariznavarreta C, F-Tresguerras JA. Changes with ageing in several leukocyte functions of male and female rats. *Biogerontology*. 2004; 5: 389-400.
58. Shipounova IN, Svinareva DA, Petrova TV, Lyamzaev KG, Chernyak BV, Drize NI, Skulachev VP. Reactive oxygen species produced in mitochondria are involved in age-dependent changes of hematopoietic and mesenchymal progenitor cells in mice. A study with the novel mitochondria-targeted antioxidant SkQ1. *Mech Ageing*. 2010; 131: 415-421.
59. Uzenbaeva LB, Vinogradova IA, Golubeva AG, Churov AV, Iliukha VA. Age related changes in rat blood leucocyte count and morpho-metrical parameters of large granular lymphocytes under different light regimens. *Adv Gerontol*. 2006; 19: 79-84.
60. De la Fuente M. Murine models of premature ageing for the study of diet-induced immune changes: improvement of leucocyte functions in two strains of old prematurely ageing mice by dietary supplementation with sulphur-containing antioxidants. *Proc Nutr Soc*. 2010; 69: 651-659.
61. Puerto M, Guayerbas N, Víctor V, De la Fuente M. Effects of N-acetylcysteine on macrophage and lymphocyte functions in

a mouse model of premature ageing. *Pharmacol Biochem Behav.* 2002; 73: 797-804.

62. Orlovskaya IA, Toporkova LB, Feofanova NA, Kolosova NG. Peculiarities of bone marrow hemopoiesis in early aging OXYS rats. *Bull Exp Biol Med.* 2007; 144: 86-88.

63. Taub DD, Murphy WJ, Longo DL. Rejuvenation of the aging thymus: growth hormone-mediated and ghrelin-mediated signaling pathways. *Curr Opin Pharmacol.* 2010; 10: 408-424.

64. Dorshkind K, Montecino-Rodriguez E, Signer RA. The ageing immune system: is it ever too old to become young again? *Nat Rev Immunol.* 2009; 9: 57-62.

65. Aleman A, Torres-Alemán I. Circulating insulin-like growth factor I and cognitive function: neuromodulation throughout the lifespan. *Prog Neurobiol.* 2009; 89: 256-265.

66. Conrad CD, Bimonte-Nelson HA. Impact of the hypothalamic-pituitary-adrenal/gonadal axes on trajectory of age-related cognitive decline. *Prog Brain Res.* 2010; 182: 31-76.

67. Li E, Kim DH, Cai M, Lee S, Kim Y, Lim E, Ryu JH, Unterman TG, Park S. Hippocampus-dependent spatial learning and memory are impaired in growth hormone-deficient spontaneous dwarf rats. *Endocrine Journal.* 2011; 58: 257-267.

68. Belousova II, Gladkikh DV, Zhelezova AI, Stefanova NA, Kolosova NG, Amstislavskaya TG. Age-related aspects of male reproductive function in rats with normal and accelerated senescence. *Russ Fiziol Zh Im I M Sechenova.* 2009; 95: 1258-1267.

69. Amstislavskaya TG, Maslova LN, Gladkikh DV, Belousova II, Stefanova NA, Kolosova NG. Effects of the mitochondria-targeted antioxidant SkQ1 on sexually motivated behavior in male rats. *Pharmacol Biochem Behav.* 2010; 96: 211-216.

70. Leontieva OV, Blagosklonny MV. Yeast-like chronological senescence in mammalian cells: phenomenon, mechanism and pharmacological suppression. *Aging (Albany NY).* 2011; 3: 1078-1091.

71. Blagosklonny MV. Aging and immortality: quasi-programmed senescence and its pharmacologic inhibition. *Cell Cycle.* 2006; 5: 2087-2102.

72. Blagosklonny MV. Progeria, rapamycin and normal aging: recent breakthrough. *Aging (Albany NY).* 2011; 3: 685-691.

73. Dazert E, Hall MN. mTOR signaling in disease. *Curr Opin Cell Biol.* 2011.

74. Kapahi P, Chen D, Rogers AN, Katewa SD, Li PW, Thomas EL, Kockel L. With TOR, less is more: a key role for the conserved nutrient-sensing TOR pathway in aging. *Cell Metab.* 2010; 11: 453-465.

75. Korotchkina LG, Leontieva OV, Bukreeva EI, Demidenko ZN, Gudkov AV, Blagosklonny MV. The choice between p53-induced senescence and quiescence is determined in part by the mTOR pathway. *Aging (Albany NY).* 2010; 2: 344-352.

76. Leontieva OV, Blagosklonny MV. DNA damaging agents and p53 do not cause senescence in quiescent cells, while consecutive re-activation of mTOR is associated with conversion to senescence. *Aging (Albany NY).* 2010; 2: 924-935.

77. Stanfel MN, Shamieh LS, Kaeberlein M, Kennedy BK. The TOR pathway comes of age. *Biochim Biophys Acta.* 2009; 1790: 1067-1074.

78. Zhao C, Vollrath D. mTOR pathway activation in age-related retinal disease. *Aging (Albany NY).* 2011; 3: 346-347.

79. Maggio M, Lauretani F, Ceda GP, Bandinelli S, Ling SM, Metter EJ, Artoni A, Carassale L, Cazzato A, Ceresini G, Guralnik JM, Basaria S, Valenti G, Ferrucci L. Relationship between low

levels of anabolic hormones and 6-year mortality in older men: the aging in the Chianti Area (InCHIANTI) study. *Arch Intern Med.* 2007; 167: 2249-2254.

80. Sorwell KG, Urbanski HF. Dehydroepiandrosterone and age-related cognitive decline. *Age (Dordr).* 2010; 32: 61-67.

## Plasma microRNA biomarkers for detection of mild cognitive impairment

Kira S. Sheinerman<sup>1</sup>, Vladimir G. Tzivinsky<sup>1</sup>, Fiona Crawford<sup>2</sup>, Michael J. Mullan<sup>2</sup>, Laila Abdullah<sup>2</sup>, and Samuil R. Umansky<sup>1</sup>

<sup>1</sup> *DiamiR, LLC, Princeton, NJ 08540, USA*

<sup>2</sup> *Roskamp Institute, Sarasota, FL 34243, USA*

**Key words:** *microRNA, Mild Cognitive Impairment (MCI), Alzheimer's disease, neurodegeneration, aging, brain, synapse, neurite*

**Received:** 9/10/12; **Accepted:** 9/18/12; **Published:** 9/19/12

**Correspondence to:** Samuil Umansky, PhD; **E-mail:** [sumansky@diamirbio.com](mailto:sumansky@diamirbio.com)

**Copyright:** © Sheinerman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** Early stages of many neurodegenerative diseases, such as Alzheimer's disease, vascular and frontotemporal dementia, and Parkinson's disease, are frequently associated with Mild Cognitive Impairment (MCI). A minimally invasive screening test for early detection of MCI may be used to select optimal patient groups in clinical trials, to monitor disease progression and response to treatment, and to better plan patient clinical care. Here, we examined the feasibility of using pairs of brain-enriched plasma microRNA (miRNA), at least one of which is enriched in synapses and neurites, as biomarkers that could differentiate patients with MCI from age-matched controls. The identified biomarker pairs fall into two sets: the "miR-132 family" (miR-128/miR-491-5p, miR-132/miR-491-5p and mir-874/miR-491-5p) and the "miR-134 family" (miR-134/miR-370, miR-323-3p/miR-370 and miR-382/miR-370). The area under the Receiver-Operating Characteristic curve for the differentiation of MCI from controls using these biomarker pairs is 0.91-0.95, with sensitivity and specificity at 79%-100% (miR-132 family) and 79%-95% (miR-134 family), and  $p < 0.001$ . In a separate longitudinal study, the identified miRNA biomarker pairs successfully detected MCI in majority of patients at asymptomatic stage 1-5 years prior to clinical diagnosis. The reported biomarker pairs also appear useful for detecting age-related brain changes. Further testing in a larger study is necessary for validation of these results.

### INTRODUCTION

Neurodegenerative diseases comprise a large group of pathologies caused by metabolic changes in brain cells, loss of synapses and other compartments of neurons, and, ultimately, neuronal death [1]. Due to increased lifespan, neurodegenerative diseases, such as Alzheimer's Disease (AD), Parkinson's Disease (PD), Huntington Disease, vascular dementia and others, have become very common in developed countries. 13.9% of people age 71 and older in the United States have dementia [2]. Currently, an estimated 5.4 million people have AD in the US alone [2]. A brain's ability to compensate for the dysfunction and loss of neurons, occurring over a long period of time, results in late clinical manifestation of symptoms of AD and other dementias. At late stages of neurodegeneration, serious morphologic changes in the brain, including a massive

loss of neurons, have already occurred, and, as a consequence, successful pharmacological intervention is not feasible. Thus, diagnostic methods based on detection of early events in the development of AD and of other dementias are highly desirable.

Mild cognitive impairment (MCI) is usually defined as an intermediate state between normal aging, and AD and other dementias, representing the first stage when clinical symptoms become evident [3-5]. On average, MCI patients convert to dementia at a rate of 10-15% annually [5,6]. Currently, the disease progression of MCI patients cannot be reliably predicted. First, up to 40% of MCI patients revert to normal status [7,8], and autopsy studies demonstrate that a substantial percentage of MCI patients do not develop AD pathology [9,10]. Second, approximately 20% of MCI patients, who progress to dementia, are diagnosed with

neurodegenerative diseases other than AD, such as vascular, Lewy body, Huntington, Parkinson, and other dementias [9,11]. Third, disease progression varies from slow to intermediate to rapid [12]. Moreover, MCI is not a homogeneous pathology and is currently described as two clinical conditions - with amnesic symptoms (aMCI) and without amnesic symptoms [8,13]. Some publications have reported that aMCI converts to dementia more frequently [14,15]. However, other authors have not found significant difference in the conversion rate for the two MCI forms [16,17].

Currently, diagnosis of AD and other forms of dementia is based on analysis of the patient's cognitive function. Amyloid plaques between neurons, neurofibrillary tangles, and an overall shrinkage of the brain tissue are the hallmarks of AD, and there have been many attempts to develop diagnostic tests based on these phenomena. Recently published data have demonstrated high sensitivity of AD detection by measuring concentrations of the three protein biomarkers in the cerebrospinal fluid (CSF): beta-amyloid protein 1-42, total tau protein, and phosphorylated tau181P protein [18,19]. However, the invasiveness of the CSF collection procedure makes such assays challenging for everyday clinical use. New imaging techniques, including PET scan for *in vivo* detection of beta-amyloid deposition, are becoming more sensitive and specific but are not suitable for first line screening [20-22]. Several groups have reported encouraging early data on the development of blood assays for AD diagnosis based on analysis of a large number of proteins or antibodies in human blood [23-25].

Neurodegenerative diseases are characterized by neuronal death in specific areas of the brain, for example, hippocampus and cortex for AD, midbrain for PD, frontal and temporal lobes for frontotemporal dementia. However, loss of neurons is a relatively late event in the progression of neurodegenerative diseases that is typically preceded by metabolic changes, such as formation of beta-amyloid plaques and tau protein tangles in AD [1], followed by synaptic dysfunction, synaptic loss, neurite retraction, and the appearance of other abnormalities, such as axonal transport defects [26-29]. Numerous studies are devoted to description of axon destruction with shedding of membrane-enclosed "axosomes", axon, dendrite and spine pruning, and disassembly of synapses [30-33]. Thus, different processes are characteristic of early and late stages of neurodegeneration and different molecular tests may be needed for early detection of the pathology and monitoring of the pathology progression versus diagnosis and monitoring of a late stage disease.

The present study evaluates the hypothesis that neurite and synapse destruction, which are pathologic processes characteristic of early stages of AD, other neurodegenerative diseases, and MCI syndrome in general, can be detected *in vitro* by quantitative analysis of brain-enriched cell-free miRNA in the blood. MicroRNA (miRNA) is a class of non-coding RNA, whose final product is an approximately 22 nt functional RNA molecule. They play important roles in the regulation of target genes by binding to complementary regions of messenger transcripts and repressing their translation or regulating degradation [34,35]. Thus, miRNA are important epigenetic regulators of numerous cellular processes [35-37]. Many of miRNA are specific to or are over-expressed in certain organs/tissues/cells [38-41]. Some miRNA, including those that are cell-specific, are also enriched in certain cellular compartments, particularly in axons, dendrites and synapses [42-46]. Changes in expression of some miRNA were found in neurons of patients with AD and other neurodegenerative diseases [47-49], as well as in animal models of AD [50,51]. Importantly, cell-free miRNA have been shown to be stable in blood samples [52].

Our approach for developing a non-invasive assay for detection of MCI is based on analysis of levels of brain-enriched miRNA, including neurite- and synapse-enriched miRNA, in plasma and identification of miRNA biomarker pairs capable of successfully differentiating MCI patients from aged-matched controls.

## RESULTS

### Selection of miRNA for pilot study

Two approaches are frequently used for the selection of promising miRNA biomarkers for detection of various cancers and other diseases. The first approach is based on analysis of hundreds of miRNA using miRNA arrays with subsequent validation of potential biomarkers by RT-PCR. In spite of an obvious advantage of this approach (i.e., the analysis of huge miRNA numbers), its disadvantages, namely lower sensitivity and higher variability, make it less suitable for the analysis of cell-free circulating miRNA in plasma or serum: (i) concentrations of many miRNA in plasma are low, and (ii) dramatic changes in miRNA levels should not be expected for a chronic pathology. The second approach is based on analysis of miRNA, whose expression level changes due to a pathology development. This approach also has certain limitations due to potential involvement of the same miRNA in diseases of various organs and because higher expression of miRNA in an affected

**Table 1. List of miRNA tested in the pilot study.**  
(Highlighted are miRNA selected as potential biomarkers for further analysis)

Number	MicroRNA
1	has-miR-7
2	has-miR-9
3	has-miR-9*
4	has-miR-98
5	has-miR-124
6	has-miR-125b
7	has-miR-127-3p
8	has-miR-128
9	has-miR-132
10	has-miR-134
11	has-miR-137
12	has-miR-138
13	has-miR-149
14	has-miR-181a
15	has-miR-181b
16	has-miR-181a*
17	has-miR-218
18	has-miR-323-3p
19	has-miR-330-3p
20	has-miR-370
21	has-miR-382
22	has-miR-383
23	has-miR-409-3p
24	has-miR-433
25	has-miR-485-3p
26	has-miR-487b
27	has-miR-491-5p
28	has-miR-539
29	has-miR-770-5p
30	has-miR-874
31	has-miR-935
32	has-miR-939

organ is not necessarily accompanied by an increase in its plasma level [53,54]. In this study we selected the initial pool of miRNA among brain- and neuron-enriched miRNA, suggesting that variations of their concentrations in plasma, if any, are most likely caused by changes in neurons and not in other cell types or organs. Since MCI and early stages of AD are associated with neurite and synapse destruction, we included in the study miRNA, which are not only

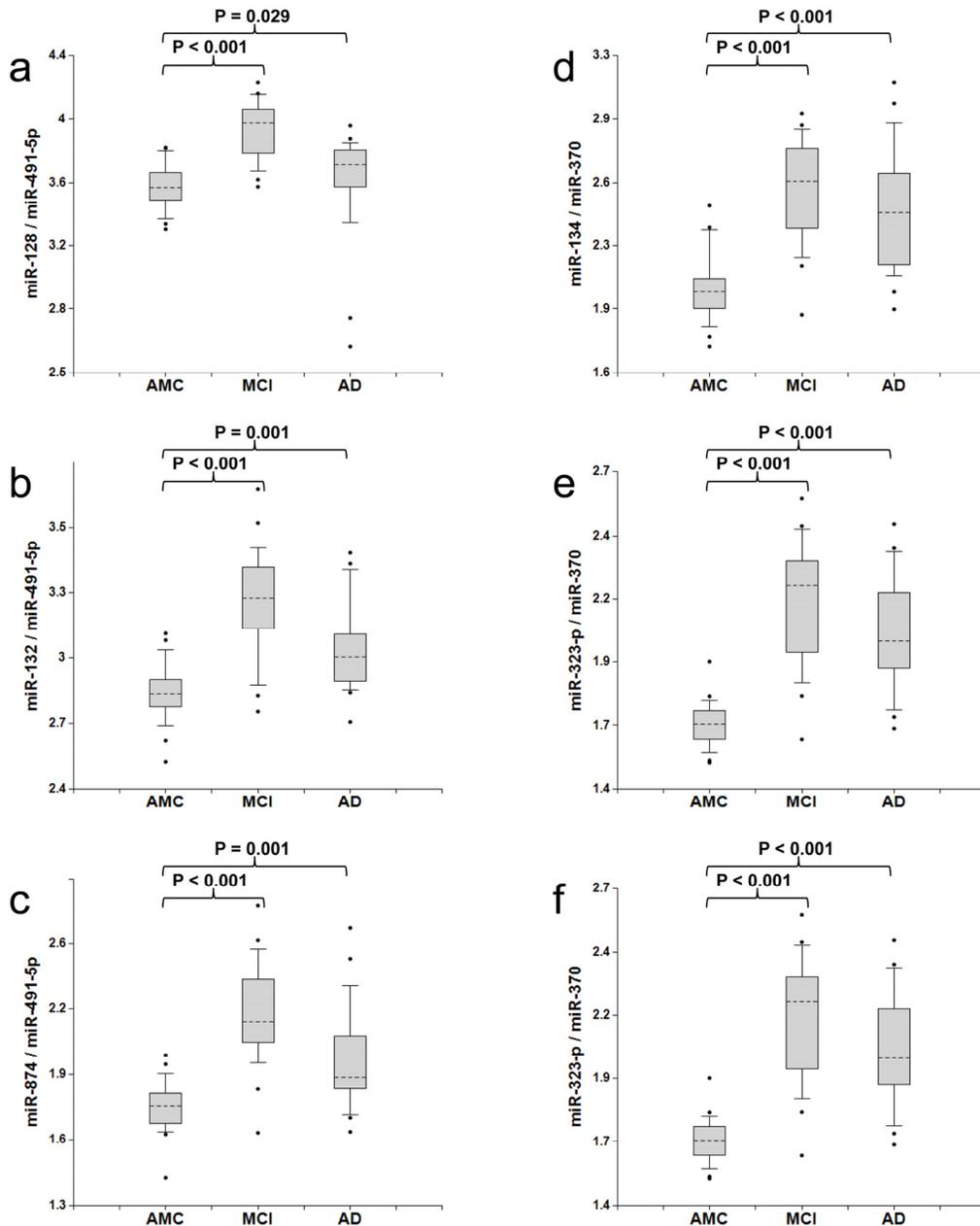
enriched in neurons but are also known to be present in neurite and synapses [38-46] and involved in neurite- and synapse-associated processes (The miR-Ontology Data Base: <http://ferrolab.dmi.unict.it/miro/>), suggesting that axon, dendrite and spine pruning and synaptic loss can lead to appearance of these miRNA in the extracellular space and ultimately in the bloodstream. 32 miRNA (Table 1) were selected for the pilot study based on the criteria described above and analyzed by individual RT-PCR, currently the most sensitive and the least variable technique.

Pilot study for selecting promising miRNA biomarkers. The concentrations of miRNA were measured in plasma samples of MCI and age-matched donors with normal cognitive function (Table 2), 10 samples in each group, by RT-PCR. miRNA with low (mean Ct>36) or undetectable plasma concentrations were excluded from the analysis. The ratios of levels of all possible miRNA pairs ( $2^{-\Delta Ct}$ ) were calculated using a software algorithm developed at DiamiR (see Supplemental Materials). Thirteen miRNA, miR-7, miR-125b, miR-128, miR-132, miR-134, miR-323-3p, miR-382, miR-874, miR-9, miR-127-3p, miR-181a, miR-370, and miR-491-5p, formed pairs differentiating MCI from age-matched controls with  $p<0.05$ ; these miRNA were selected for further analysis.

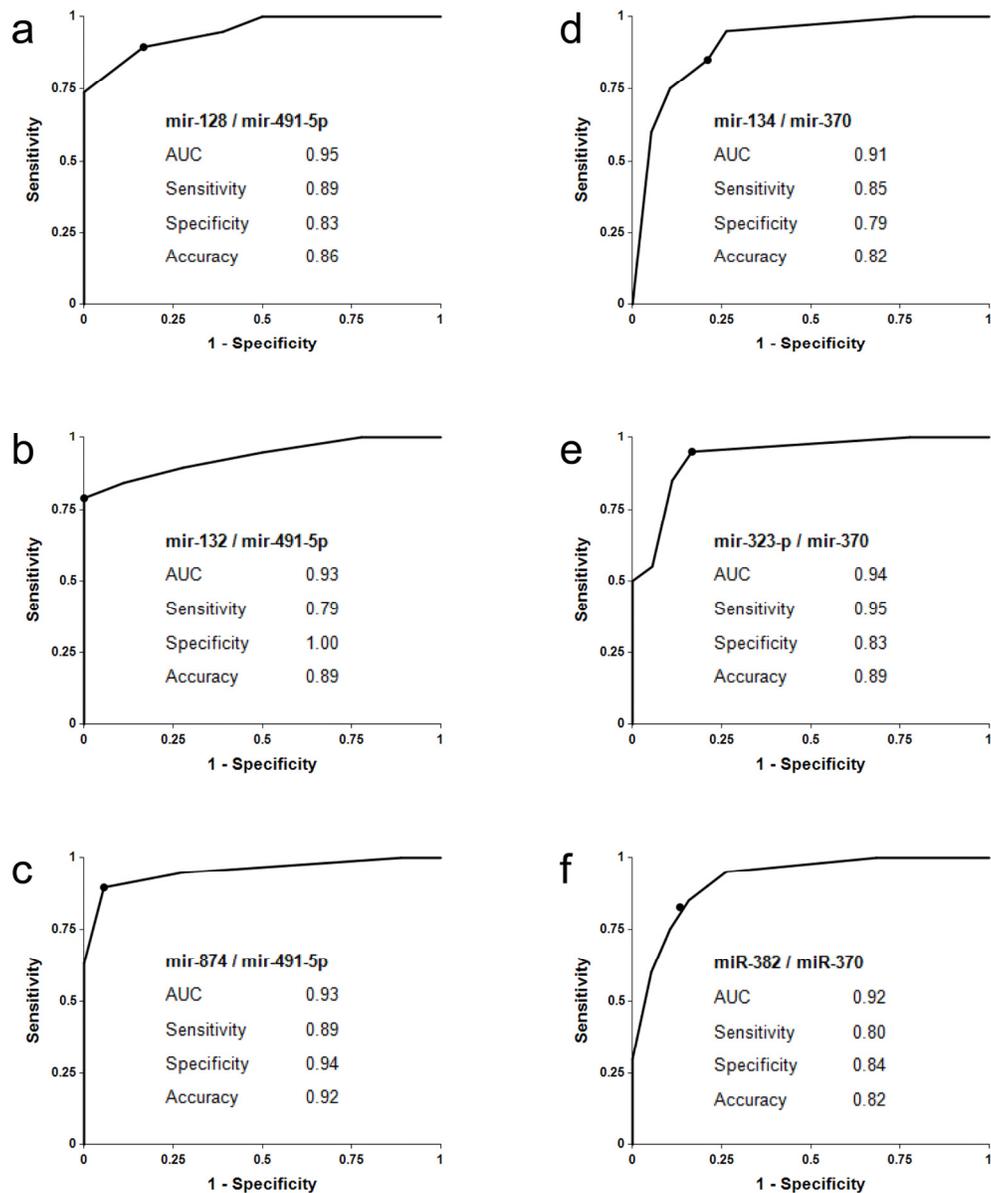
#### Feasibility study for differentiation of MCI and AD from Age-Matched Controls

The concentrations of 13 miRNA selected in the pilot study were determined by the single target TaqMan® miRNA qRT-PCR assay (Applied Biosystems) in the plasma samples of amnesic MCI patients, AD patients and age-matched donors, 20 samples in each group (Table 2). The ratios of levels of all possible miRNA pairs were calculated. The data obtained in this set of experiments are reported in Fig. 1 and S1. Receiver-Operating Characteristic (ROC) curves for miRNA pairs with the highest sensitivity and specificity are presented in Fig. 2.

Biomarker pairs miR-128/miR-491-5p, miR-132/miR-491-5p and miR-874/miR-491-5p (Set 1) differentiated MCI from age-matched control with 79%-89% sensitivity and 83%-100% specificity (Fig. 1a-c and 2a-c). The area under the ROC curve (AUC) for miR-128/miR-491-5p, miR-132/miR-491-5p and miR-874/miR-491-5p is 0.95, 0.93 and 0.95, respectively. In addition, biomarker pairs miR-134/miR-370, miR-323-3p/miR-370 and miR-382/miR-370 (Set 2) demonstrated 80%-95% sensitivity and 79-84% specificity (Fig. 1d-f and 2d-f). AUC for miR-134/miR-370, miR-323-3p/miR-370 and miR-382/miR-370 are 0.91, 0.94 and 0.92, respectively.



**Figure 1. Ratios of miRNA levels (biomarker pairs) in plasma of age-matched controls, MCI, and AD patients.** The concentrations of miRNA in plasma samples of MCI and AD patients, and age-matched donors with normal cognitive function, 20 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta Ct} \times 100$ . Here and in other figures with box and whisker plots the results are presented in the Log10 scale. The upper and lower limits of the boxes and the lines inside the boxes indicate the 75th and 25th percentiles and the median, respectively. The upper and lower horizontal bars denote the 90th and 10th percentiles, respectively. The points indicate assay values located outside of 80% data. AMC: age-matches controls; MCI: MCI patients; AD: AD patients.



**Figure 2. Receiver-Operating Characteristic (ROC) curve analysis of differentiation between MCI patients and age-matched controls obtained with different biomarker pairs.** The areas under the ROC curve (AUC) are reported. Sensitivity, specificity and accuracy for each biomarker/normalizer pair are calculated for the “cutoff” point (indicated as a dot on each plot); the cutoff point is the ratio of paired miRNA, at which a sample is equally likely to belong to the AMC and the MCI groups (see Supplementary materials for more details).

**Table 2. Demographics of plasma donors**

Clinical Diagnosis	Number	Age		Sex	MMSE
		Mean	Range	Male/Female	(mean ± SD)
<b><u>Pilot Study</u></b>					
<i>Age matched controls (AMC)</i>	10	77.4	71-85	5/5	28.9 ± 1.1
<i>Mild cognitive impairment (MCI)</i>	10	81.7	75-87	5/5	28.1 ± 1.4
<b><u>Main Study</u></b>					
<i>Young control (CY)</i>	20	36.5	21-50	11/9	29.7 ± 2.6
<i>AMC</i>	20	80.2	76-86	12/8	29.2 ± 1.3
<i>MCI</i>	20	79.9	72-89	15/5	25.8 ± 3.5
<i>AD</i>	20	76.9	63-89	13/7	20.8 ± 8.7
<b><u>Longitudinal study</u></b>	19	77.0	73-84	10/9	28.8 ± 1.3

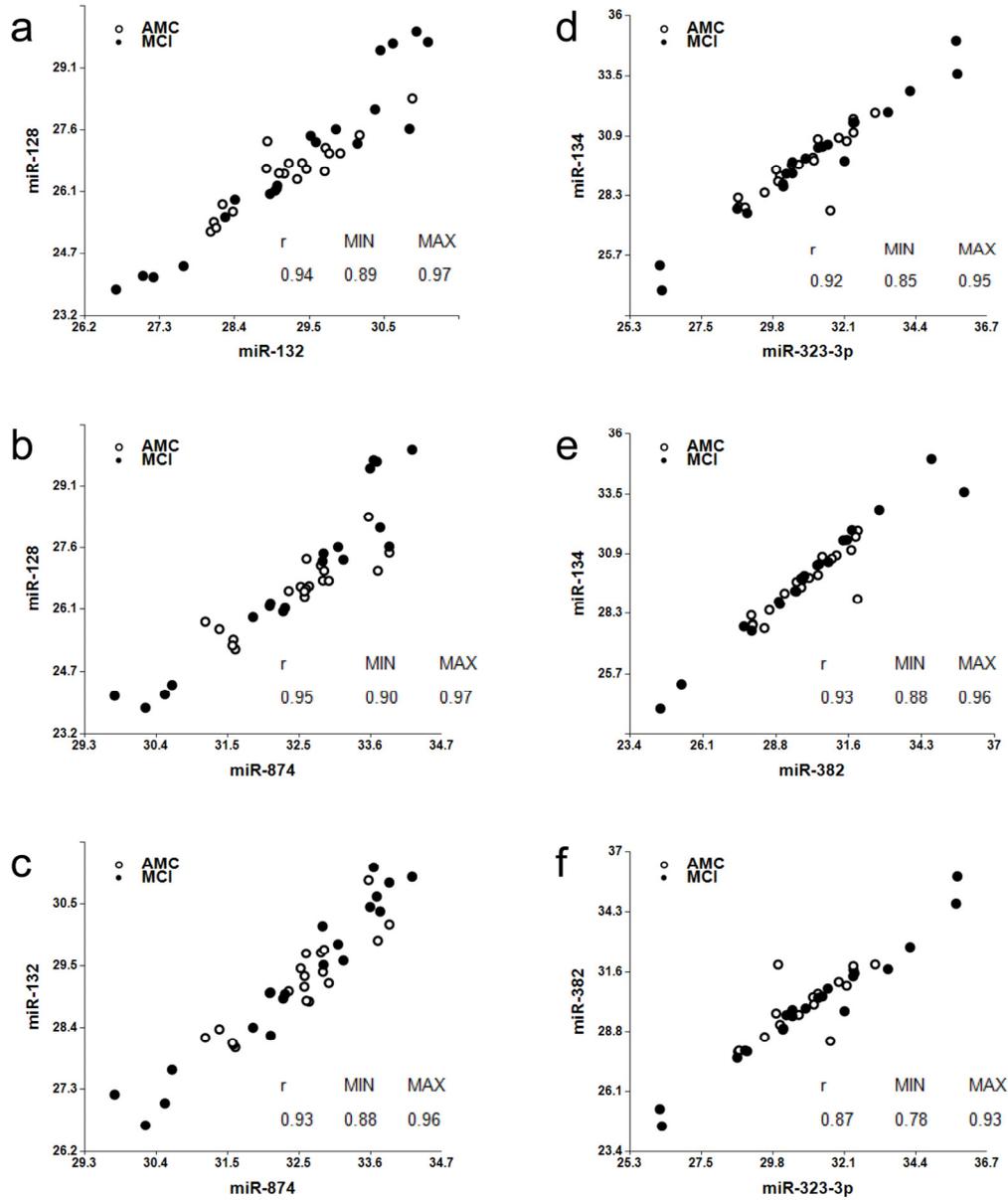
Each biomarker Set 1 and 2 includes three different miRNA (numerators) paired with the same miRNA (denominator): miR-128, miR-132 and miR-874 are paired with miR-491-5p and miR-134, miR-323-3p and miR-382 are paired with miR-370. miR-128, miR-132 and miR-134 are located in neurites and synapses [42-46]. miR-323-3p and miR382 are enriched in synaptoneuroosomes of rat cortex and hippocampus [55]. The predicted targets of miR-874 indicate its involvement in axonogenesis, neurotransmitter secretion, dendrite morphogenesis, synaptogenesis, synaptic transmission and synaptic vesicle exocytosis [The miR-Ontology Data Base: <http://ferrolab.dmi.unict.it/miro/>]. Thus, each biomarker pair includes a neurite/synapse-enriched miRNA. A correlation analysis shown in Fig. 3 demonstrates that miR-128, miR-132 and miR-874 form one family of biomarkers (miR-132 family) (Spearman test r values in the pair comparison are in the 0.93-0.95 range) and miR-134, miR-323-3p and miR-382 form another family of biomarkers (miR-134 family) (Spearman test r values in the pair comparison are in the 0.87-0.93 range).

Biomarker Set 1 and Set 2 also differentiated AD dementia from the age-matched control with  $p < 0.05$ , which is not surprising, since about 50% MCI patients progress to AD; however, these biomarkers did not distinguish AD from MCI, and moreover, the overlap between the distributions of biomarkers for AD and age-matched control was greater than the overlap for MCI and age-matched control (Fig. 1). Two factors may help explain this outcome: (i) as numerous synapses and neuritis are destroyed during earlier stages of the disease, the total amount of excreted synapse/neurite miRNA decreases in later stages of AD, and (ii) in later stages of AD, concentrations of other brain-enriched miRNA (denominator in a biomarker pair) in blood may increase due to their presence in neuronal compartments, glial cells or brain areas, which are involved in the pathology progression.

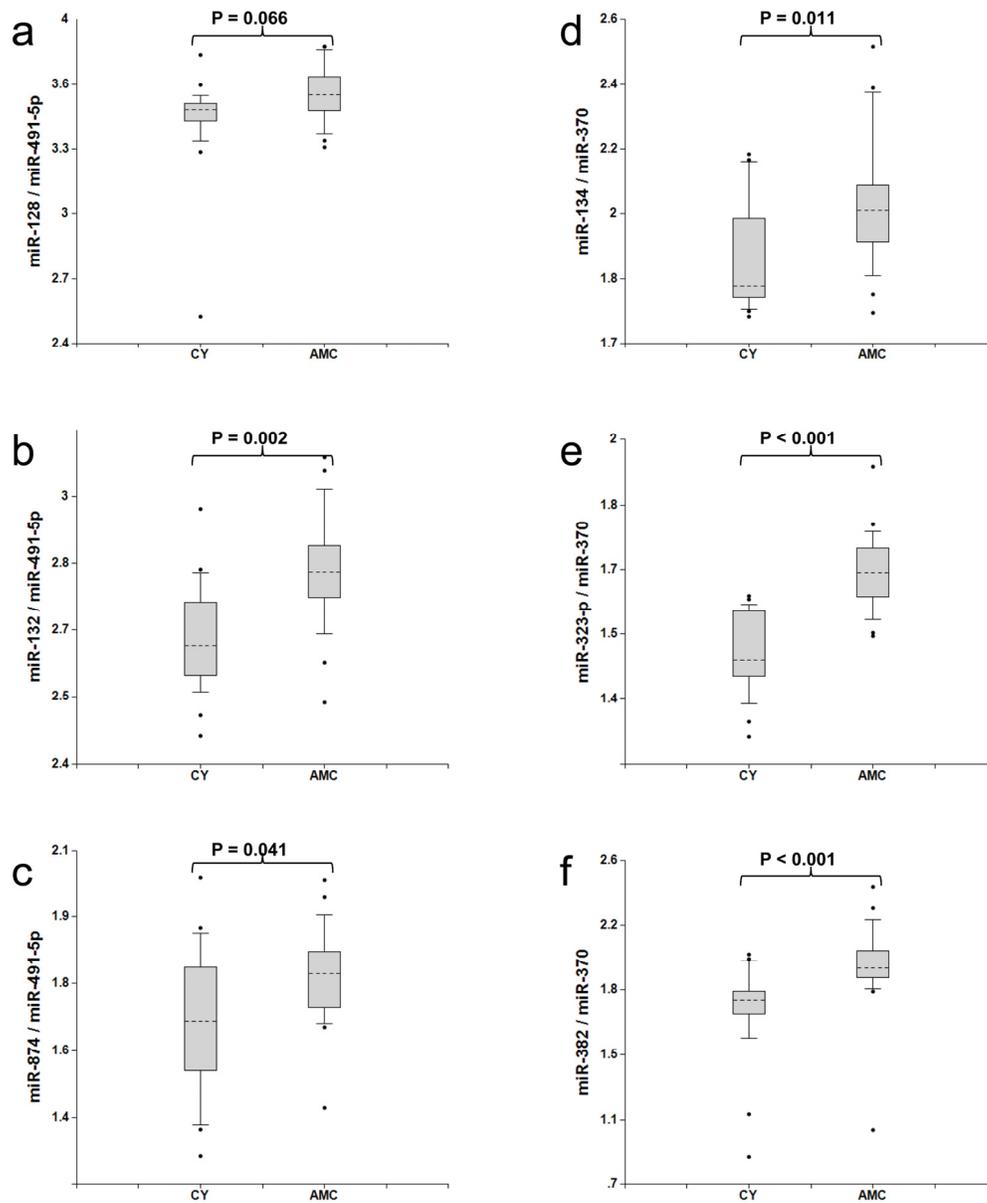
Supplemental Figure S1 summarizes the results obtained for other miRNA pairs tested for MCI differentiation from age-matched control. These miRNA pairs detect smaller sub-groups of MCI and

further studies are necessary to address the question of whether they can be used for detection of particular MCI subsets and for prediction of the disease outcome

Retrospective longitudinal study of MCI development in elderly patients with normal cognitive function at enrollment.



**Figure 3. Analysis of associations among miR-128, miR-132, and miR-874 ("miR-132 family"); and miR-134, miR-323-3p and miR-382 ("miR-134 family"). Spearman's rank correlation coefficients  $r$  along with 95% confidence intervals (MIN & MAX) are shown.**



**Figure 4. Comparison of miRNA biomarker pairs in plasma of Group 1 (30-50 years old, “CY”) and Group 2 (70-80 years old, “AMC”) individuals with normal cognitive functions.** The concentrations of miRNA in plasma samples of Group1 (30-50 years old, CY) and Group2 (70-80 years old, AMC) donors with normal cognitive function, 20 samples in each group, were measured by RT-PCR and the ratio of various miRNA was calculated as  $2^{-\Delta\Delta Ct} \times 100$ . See the legend to Fig. 1 for the description of the statistical analysis.

The three biomarker pairs of the miR-132 family (Set 1) have shown overall the highest sensitivity and specificity in differentiating MCI from the age-matched control (Fig. 1, 2). These biomarker pairs were, therefore, used to analyze the development of MCI in elderly patients with initially normal cognitive function, recruited in a small longitudinal study at the Roskamp Institute in Florida. Subjects with normal cognitive functions who were at least 70 years old were enrolled and followed for 2-5 years with cognitive assessment and regular collection of plasma. In the course of the study, some subjects remained cognitively normal, while others progressed to MCI. The plasma samples from the 19 subjects, 10 of whom progressed to MCI, were used for miRNA extraction and analysis. In an effort to minimize the effect that a prolonged storage could have had on quality of the samples, patients were

classified disease-positive only if in two samples collected at consecutive time points, the concentrations of at least two of the three biomarker pairs, miR-128/miR-491-5p, miR-132/miR-491-5p, and miR-874/miR-491-5p, were higher than the cutoffs determined in the previous experiment (Fig. 2); i.e. if the positive diagnosis made based on the first sample was confirmed using the blood sample collected from the same patient during the next visit. The data, reported in Table 3, demonstrate that in 7 of the 10 subjects who progressed to MCI (patients 10, 12-16, and 19) the increase in plasma levels of miRNA biomarkers is detectable at asymptomatic disease stage, preceding MCI diagnosis by 6 to 61 months. Among the nine patients who remained MCI free, none were classified disease-positive by our assay according to criteria described above.

**Table 3. Clinical and miRNA-based diagnosis of MCI in elderly subjects with normal cognitive function at the time of enrollment over the course of 2-5 years**

Patient	Clinical diagnosis	Time of clinical diagnosis (number of months past enrollment)	miRNA-based diagnosis	Time of miRNA-based diagnosis (number of months past enrollment)	Number of months the miRNA-based diagnosis preceded the clinical diagnosis
C-1	Normal	NA	Normal	NA	NA
C-2	Normal	NA	Normal	NA	NA
C-3	Normal	NA	Normal	NA	NA
C-4	Normal	NA	Normal	NA	NA
C-5	Normal	NA	Normal	NA	NA
C-6	Normal	NA	Normal	NA	NA
C-7	Normal	NA	Normal	NA	NA
C-8	Normal	NA	Normal	NA	NA
C-9	Normal	NA	Normal	NA	NA
MCI-1	MCI	18	MCI	0	18
MCI-2	MCI	33	Normal	NA	NA
MCI-3	MCI	12	MCI	0	12
MCI-4	MCI	23	MCI	0	23
MCI-5	MCI	6	MCI	0	6
MCI-6	MCI	19	MCI	0	19
MCI-7	MCI	61	MCI	0	61
MCI-8	MCI	19	Normal	NA	NA
MCI-9	MCI	16	Normal	NA	NA
MCI-10	MCI	28	MCI	0	28

NA- Not Applicable

## Analysis of normal brain aging with selected miRNA biomarker pairs

The development of MCI, AD and other neurodegenerative diseases on one hand, and normal aging on the other hand share certain common processes, e.g. neurite and synapse destruction and ultimately neuronal death. In this experiment we analyzed whether normal aging could be detected by the same miRNA biomarker pairs. miRNA concentrations in plasma samples from two groups, each comprised of 20 cognitively normal subjects, Group 1 (21-50 years old, "CY") and Group 2 (76-86 years old, "AMC"), were measured and compared as described above. The data presented in Fig. 4 (Sets 1 and 2) and Fig. S2 (other miRNA pairs) demonstrate that biomarker levels are higher in the plasma of Group 2, "AMC" subjects compared to Group 1, "CY" subjects ( $p < 0.05$  to  $p < 0.001$ ). Thus, a larger prospective longitudinal analysis of these biomarkers in plasma could potentially provide important information on brain processes associated with normal aging.

## DISCUSSION

The objective of the present study was to search for plasma miRNA biomarkers that can be used to detect MCI. The results obtained in our experiments have demonstrated for the first time that a minimally invasive test based on analysis of cell-free miRNA circulating in plasma could be feasible for detection of MCI, AD and even asymptomatic stages of neurodegeneration. The use of brain-enriched neurites/synapses miRNA enables detection of early pathologic events occurring in neurons. Further, combination of neurite/synapse miRNA with other experimentally selected brain-enriched miRNA significantly increases assay sensitivity and specificity at early stages of the pathology, most likely due to compensation for a number of variables, such as blood supply, changes in blood-brain barrier permeability and others.

Two sets of biomarkers have demonstrated high sensitivity and specificity in differentiating MCI from age-matched controls - the miR-132 and miR-134 families paired with miR-491-5p and miR-370, respectively. Although a relatively small number of patients was used in the feasibility study to identify the efficient miRNA biomarker pairs, the data obtained in the longitudinal study (Table 3) and the study of normal brain aging (Fig. 4 and S2) support the findings. Total of 171 plasma samples were analyzed in the experiments reported here. High correlation among members of miR-134 set, namely miR-134, miR-323-3p and miR-382, can be explained by the fact that these

miRNA belong to the same cluster (<http://www.diana.pcbi.upenn.edu/cgi-bin/miRGen/v3/Cluster.cgi>) and are expressed in the same cell types. Close functional relatedness among members of miR-132 set, namely miR-128, miR-132 and miR-874, has not been described before. It is also interesting to note that miR-132 and miR-134 biomarker sets demonstrate higher sensitivity and specificity when paired with different brain-enriched miRNA. The miR-132 set is a strong match with miR-491-5p, miR-181a, and miR-9, while the miR-134 set demonstrates the strongest differentiation between MCI and age-matched controls when paired with miR-370 and miR-127. Correlation between the two miRNA sets (data not shown) is relatively low ( $r$  values in the pair comparison Spearman test are in the 0.56-0.79 range) indicating that they possibly reflect distinct pathological processes, or are enriched in different brain areas. A mechanistic explanation for this observation is currently missing, and could be provided by a detailed analysis of expression of all these miRNA in various brain areas and cell types.

It is important to mention that most of elderly patients and age-matched controls, as well as some of young controls had various non-neurological conditions unrelated to MCI. However, since this is expected to be the case in a real-life test application, such a test should be capable of detecting MCI in subjects with accompanying diseases. Thus, only patients with a history of a stroke or other neurologic pathologies were excluded from the present study. We believe that the ability of selected miRNA pairs to differentiate MCI (and AD) from age-matched control in spite of the presence of other pathologies supports our approach to biomarker selection from brain-enriched miRNA. The same consideration applies to the comparison between younger and older groups. Additional larger studies are necessary for further data validation, including a prospective longitudinal study. miR-132 and miR-134 families paired with other brain-enriched miRNA effectively distinguish MCI and AD from age-matched control but do not differentiate MCI and AD from each other. Thus, other biomarkers are necessary for prediction of MCI progression to AD and other dementia. The experiments aimed at detection of the MCI sub-types that will progress to AD dementia are currently in progress at DiamiR. The differentiation of AD from other dementias (vascular, frontotemporal, Lewy bodies, etc.) is another important goal and we hope that analysis of miRNA enriched in different brain areas could be useful for differential diagnosis. Further, there are other brain-enriched miRNA, which were not included in the present study but could be found useful as potential biomarkers in the future. Additional

promising miRNA along with those described in the present study could be used for detecting other neurodegenerative diseases and for differential diagnosis.

Early detection of MCI patients by a minimally invasive, screening test may make more invasive and expensive tests for detection of AD and other neurodegenerative diseases more practical, since the latter can be applied to the pre-selected cohorts of patients.

Numerous data demonstrate changes in miRNA expression associated with cellular senescence and *in vivo* aging [56-59]. Li et al. described increase in levels of miR-34a in the brain, peripheral blood mononuclear cells, and plasma during aging in mice [60]. It is intriguing that in our study the miRNA biomarker pairs found to differentiate MCI from age-matched controls can be used to register changes during normal brain aging, suggesting that the approach reported in the present study enables detection of processes common for normal aging and MCI development, e.g. destruction of synapses, and could be helpful in basic neurophysiology research of aging. A larger study with subjects representing various age groups (20-30, 30-40...80-90 y. o.) is necessary for validation of these initial findings.

Recently, the National Institute of Aging and Alzheimer's Association has developed new diagnostic guidelines for AD [61-63]. The guidelines contain updated classification of the AD phases, namely the dementia phase, the symptomatic pre-dementia phase (MCI), and the asymptomatic, preclinical phase of AD (pre-MCI). The new guidelines also provide recommendations for the diagnosis of pre-MCI, MCI and AD dementia and stress the current lack of and a great need for reliable biomarkers, which can be used for detection of MCI and preclinical phases of AD. We believe the current study makes a significant contribution towards this objective.

## **MATERIALS AND METHODS**

Plasma samples. The plasma samples used in the present study were collected at the Roskamp Institute Memory Center between 2005 and 2009 under the protocol approved by the Western Institutional Review Board (WIRB). An IRB approved written consent was obtained from each subject recruited in the study and the informed consent process was conducted in accordance with the International Conference on Harmonization (ICH) guidelines. If a subject was not medically capable or legally competent to provide

consent for participation in the study, a written consent was obtained from a family member, a legally authorized representative (LAR) or health care surrogate (under 21 CFR 50: exceptions from general requirements for informed consent). An assent was obtained from the participant. Venous blood was collected in EDTA vacutainers (BD Diagnostics), which were immediately centrifuged at 1380 x g for 5 minutes. Samples were maintained at 4°C during the plasma preparation process and aliquoted immediately in 1.5ml Eppendorf tubes for storage at -80°C until further use. The use of the samples in the present study was additionally approved by the WIRB in 2010. The quantity and the type of the samples used in the present study are as follows (Table 2): Pilot Study: amnesic MCI and age-matched donors (> 70 years old) with normal cognitive function, 10 samples in each group; MCI and AD detection: amnesic MCI patients, AD patients and age-matched donors (> 70 years old), 20 samples in each group; Retrospective longitudinal study: samples from 19 subjects, each subject at least 70 years old and having normal cognitive function at the time of the first plasma collection, multiple samples collected from each subject over the course of 2-5 years; Detection of normal brain aging: samples from 20 subjects, 30-50 years old and normal cognitive function, as well as the samples from 20 subjects, each at least 70 years old and having normal cognitive function, which were used as control in the MCI and AD detection study.

MCI and AD diagnosis. The age-matched controls (AMC) were either recruited from the Roskamp Institute Memory Clinic screening programs conducted in Tampa and Sarasota, FL or through the Alzheimer's disease Anti-inflammatory Prevention Trial (ADAPT) Tampa, FL site. The subjects from ADAPT underwent a brief neuropsychological assessment at enrollment as described elsewhere to determine cognitively normal status [64]. For all AMC, mini mental status examination (MMSE) was also administered to determine cognitive status. In addition, AMC subjects maintained independent activities of daily living and were free of any active neurological illness, psychiatric disorders, or other medical conditions that would potentially interfere with their cognitive performance. Individuals suspected of having MCI or AD underwent a comprehensive dementia work-up which included physical and neurological examinations, laboratory studies (i.e., CBC, chemistry count, sedimentation rate, vitamin B12 and folic acid levels, thyroid test and syphilis serological test) and neuroimaging (i.e., MRI or CT), as applicable. A more comprehensive neuropsychological assessment was also administered as part of the dementia work-up and consisted of

expanded Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery [65]. Learning and memory functions were evaluated using the CERAD 10-word, 3-trial list learning task and CERAD delayed recall measure and Logical Memory I and II of the Wechsler Memory Scale – Revised [66]. The CERAD Constructional Praxis test and Judgment of Line Orientation Test measured visuospatial ability [67]. Language and/or executive measures included 15-item Boston Naming Test, Animal Fluency, the Control Oral Word Association Test (COWAT; CFL); and the similarities subtest from the Wechsler Adult Intelligence Test – 3rd Revision (WAIS-III) [68]. The Trails A of the Trail Making Test and Digit Symbol from the WAIS-III were utilized to measure visual scanning and processing speed. Set-shifting (an executive ability) was measured using Trails B and the Letter Number Sequencing subtest from the WAIS-III [69]. Following dementia work-up, a consensus team determined cognitive status using published diagnostic criteria. The diagnosis of AD was made using NINCDS-ADRDA [70] and amnesic MCI according to the Petersen criteria [71]. MCI and AD patients as well as control subjects with a known history of a stroke or other neurologic pathologies were excluded from the study.

Plasma RNA extraction and qRT-PCR miRNA analysis miRNA isolation and qRT-PCR analysis of the initial set of 32 miRNA, performed in the course of the Pilot study, were performed by Gene Logic (an Ocimum Biosolutions Company, Gaithersburg, MD, USA) according to the following protocol. RNA was extracted from 250  $\mu$ l aliquots using mirVana™ Paris Extraction Kit and protocol (Ambion).  $2.5 \times 10^7$  copies of *Arabidopsis thaliana* miR-159a (*ath-mir-159a*) were spiked per 0.25  $\mu$ l plasma after addition of guanidine-containing denaturing solution for evaluating miRNA yield. Single target qRT-PCR was performed using the TaqMan® Reverse Transcription Kit and miRNA-specific stem-loop primers (Applied Biosystems). Final PCR was performed in triplicate using 3.3  $\mu$ l plasma equivalents. Based on the quantitative measurement of spiked *ath-miR-159a*, average yield of miRNA from plasma was about 70%.

miRNA isolation and qRT-PCR analysis in all other experiments were performed by Asuragen Inc. (Austin, TX, USA) according to the following protocol. RNA was extracted from 200  $\mu$ l aliquots using Asuragen's proprietary protocol, which is based on Trizol treatment and silica (Ambiom Glass Fiber Microcolumn) binding. Single target qRT-PCR was performed using the TaqMan® Reverse Transcription Kit and miRNA-specific stem-loop primers (Applied Biosystems). RT

step was performed in triplicate and 2  $\mu$ l plasma equivalents were present in final PCR. The concentrations of the 13 miRNA (8 neurite/synapse miRNA and 5 other brain-enriched miRNA), were determined in the plasma samples of amnesic MCI patients, AD patients and age-matched donors, 20 samples in each group (Table 2). The sample size for this study was determined by a standard formula for a case-control study [72] using power = 0.8, significance level = 0.05 and the ratio of standard deviation to difference between comparison groups set to 1. The ratios of levels of all possible miRNA pairs were calculated.

Bioinformatics analysis and statistical methods. In addition to biological factors, such as levels of expression, secretion, blood-brain barrier permeability, etc., miRNA yield from plasma may depend on a purification technique. Further, presence of RT-PCR inhibitors in the blood may vary from subject to subject and distort an experimental outcome. Therefore, data normalization becomes an issue of critical importance. Two normalization approaches that are commonly used in miRNA studies include: (i) normalization per the least variable miRNA, such as spiked non-human miRNA or ubiquitous miRNA, whose concentration is expected to be minimally changed by a pathology being analyzed [73], and (ii) normalization based on an experimental search for miRNA pairs, which most effectively differentiate two populations, e.g. pathology versus control [74,75]; ratios of levels of all possible miRNA pairs from the same sample are calculated and the most promising pairs (self-normalizing biomarkers) are selected for further testing and validation. The advantage of the second approach is that in certain cases miRNA, whose concentrations are changed due to a pathology in opposite directions, can be effective in differentiating investigated populations. We use the latter approach and, in addition to brain-enriched miRNA present in neurites and synapses, measure other brain-enriched miRNA to compensate for variations in blood supply, blood/brain barrier permeability, and other brain-specific factors.

All statistical calculations were performed with the use of custom software developed at DiamiR LLC (Princeton, NJ), as described in the Supporting Information (Software, Calculations, Graphical Interface). Mann-Whitney U-tests were used to evaluate significance of differentiation of any two patient groups by various miRNA pairs. Spearman's rank correlation coefficient was calculated to estimate associations between various biomarkers. P-value < 0.05 was considered significant; actual p-values are reported for each experiment. Receiver-Operating Characteristic

(ROC) curves were constructed and the area under ROC curves (AUC) was calculated to evaluate sensitivity and specificity of various biomarker sets. The cutoff points on the ROC curves, at which accuracy of MCI detection is maximal, were selected.

## ACKNOWLEDGEMENTS

This work was supported by the IRS Qualifying Therapeutic Discovery Project Grant.

## Conflict of Interest Statement

Kira S. Sheinerman, Vladimir G. Tsivinsky, and Samuil R. Umansky are shareholders of DiamiR, LLC.

## REFERENCES

1. Neurodegenerative diseases: From Molecular Concepts to Therapeutic Targets. Editors: R. von Bernhardi, N.C. Inestrosa, Nova Publishers, 2008.
2. 2011 Alzheimer's Disease Facts and Figures. Report of Alzheimer's Association. [http://www.alz.org/downloads/Facts\\_Figures\\_2011.pdf](http://www.alz.org/downloads/Facts_Figures_2011.pdf).
3. DeCarli C. Mild cognitive impairment: prevalence, prognosis, etiology, and treatment. *Lancet Neurol.* 2003; 2:15-21.
4. Markesbery WR. Neuropathologic Alterations in Mild Cognitive Impairment: A Review. *J. Alzheimers Dis.* 2010; 19:221-228.
5. Apostolova LG, Thompson PM, Green AE, Hwang KS, Zoumalan C, Jack CR Jr, Harvey DJ, Petersen RC, Thal LJ, Aisen PS, Toga AW, Cummings JL, Decarli CS. 3D comparison of low, intermediate, and advanced hippocampal atrophy in MCI. *Hum Brain Mapp.* 2010; 31:786-799.
6. Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Ritchie K, Rosser M, Thal L, Winblad B. Current concepts in mild cognitive impairment. *Arch Neurol.* 2001; 58:1985-1992.
7. Larrieu S, Letenneur L, Orgogozo JM, Fabrigoule C, Amieva H, Le Carret N, Barberger-Gateau P, Dartigues JF. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology.* 2002; 59:1594-1599.
8. Brooks LG & Loewenstein DA. Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions. *Alzheimers Res Therapy.* 2010; 2:28-36.
9. Jicha GA, Parisi JE, Dickson DW, Johnson K, Cha R, Ivnik RJ, Tangalos EG, Boeve BF, Knopman DS, Braak H, Petersen RC. (2006) Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Arch Neurol.* 2006; 63:674-681.
10. Khan TK & Alkon DL. Early diagnostic accuracy and pathophysiologic relevance of an autopsy-confirmed Alzheimer's disease peripheral biomarker. *Neurobiol Aging.* 2010; 31:889-900.
11. Stephan BCM, Matthews FE, Khaw K-T, Dufouil C & Brayne C. Beyond mild cognitive impairment: vascular cognitive impairment, no dementia (VCIND). *Alzheimers Res Therapy.* 2009; 1:4-12.
12. Doody RS, Pavlik V, Massman P, Rountree S, Darby E, Chan W. Predicting progression to Alzheimer's disease. *Alzheimers Res Ther.* 2010; 2:2-10.
13. Dlugaj M, Weimar C, Wege N, Verde PE, Gerwig M, Dragano N, Moebs S, Jöckel KH, Erbel R, Siegrist J; Heinz Nixdorf Recall Study Investigative Group. Prevalence of Mild Cognitive Impairment and Its Subtypes in the Heinz Nixdorf Recall Study Cohort. *Dement Geriatr Cogn Disord.* 2010; 30:362-373.
14. Mariani E, Monastero R & Mecocci P. Mild cognitive impairment: a systematic review. *J Alzheimers Dis.* 2007; 12:23-35.
15. Luck T, Busse A, Hensel A, Angermeyer MC & Riedel-Heller SG. Mild cognitive impairment and development of dementia. *Psychiatr Prax.* 2008; 35:331-336.
16. Koivunen J, Scheinin N, Virta JR, Aalto S, Vahlberg T, Någren K, Helin S, Parkkola R, Viitanen M, Rinne JO. Amyloid PET imaging in patients with mild cognitive impairment: a 2-year follow-up study. *Neurology.* 2011; 76:1085-1090.
17. Rountree SD, Waring SC, Chan WC, Lupo PJ, Darby EJ, Doody RS Importance of subtle amnesic and nonamnesic deficits in mild cognitive impairment: prognosis and conversion to dementia. *Dement Geriatr Cogn Disord.* 2007; 24:476-482.
18. Fagan AM, Shaw LM, Xiong C, Vanderstichele H, Mintun MA, Trojanowski JQ, Coart E, Morris JC, Holtzman DM. Comparison of analytical platforms for cerebrospinal fluid measures of  $\beta$ -amyloid 1-42, total tau, and P-tau(181) for identifying Alzheimer disease amyloid plaque pathology. *Arch Neurol.* 2011; 68:1137-1144.
19. De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, Coart E, Hansson O, Minthon L, Zetterberg H, Blennow K, Shaw L, Trojanowski JQ; Alzheimer's Disease Neuroimaging Initiative. Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Arch Neurol.* 2010; 67:949-956.
20. Mistur R, Mosconi L, Santi SD, Guzman M, Li Y, Tsui W, de Leon MJ. Current Challenges for the Early Detection of Alzheimer's Disease: Brain Imaging and CSF Studies. *J Clin Neurol.* 2009; 5:153-166.
21. Miller G. Alzheimer's biomarker initiative hits its stride. *Science.* 2009; 326:386-389.
22. Schmand B, Eikelenboom P. & van Gool WA. Value of Neurophysiological Tests, Neuroimaging, and Biomarkers for Diagnosing Alzheimer's Disease in Younger and Older Age Cohorts. *J Am Geriatr Soc.* 2001; 59:1705-1710.
23. Ray S, Britschgi M, Herbert C, Takeda-Uchimura Y, Boxer A, Blennow K, Friedman LF, Galasko DR, Jutel M, Karydas A, Kaye JA, Leszek J, Miller BL, Minthon L, Quinn JF, Rabinovici GD, et al. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat Med.* 2007; 13:1359-1362.
24. Reddy MM, Wilson R, Wilson J, Connell S, Gocke A, Hynan L, German D, Kodadek T. Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. *Cell.* 2011; 144:132-142.
25. Nagele E, Han M, Demarshall C, Belinka B, & Nagele R. Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. *PLoS One.* 2011; 6:e23112.
26. Crews L & Masliah E. Molecular mechanisms of neurodegeneration in Alzheimer's disease. *Human Mol Gen.* 2010; 19:R12-R20.

27. Bredesen DE. Neurodegeneration in Alzheimer's disease: caspases and synaptic element independence. *Mol Neurodegener.* 2009; 4:27-36.
28. Nimmrich V & Ebert U. Is Alzheimer's disease a result of presynaptic failure? Synaptic disfunctions induced by oligomeric beta-amyloid. *Rev Neurosci.* 2009; 20:1-12.
29. Yoshiyama Y, Higuchi M, Zhang B, Huang SM, Iwata N, Saido TC, Maeda J, Suhara T, Trojanowski JQ, Lee VM. Synapse Loss and Microglial Activation Precede Tangles in P301S Tauopathy Mouse Model. *Neuron.* 2007; 53:337-351.
30. Eaton BA & Davis GW. Synapse disassembly. *Genes Dev.* 2003; 17:2075-2082.
31. Koirala S & Ko C-P. Pruning an Axon Piece by Piece. *Neuron.* 2004; 44:578-580.
32. Bishop DL, Misgeld T, Walsh MK, Gan W-B & Lichtman JW. Axon branch removal at developing synapses by axosome shedding. *Neuron.* 2004; 44:651-661.
33. Low LK & Cheng H-J. Axon pruning: an essential step underlying the developmental plasticity of neuronal connections. *Phil Trans R Soc B.* 2006; 361:1531-1544.
34. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A & Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 2006; 34, Database issue: D140-D144.
35. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009; 136:215-233.
36. Brett JO, Renault VM, Rafalski VA, Webb AE, Brunet A. The microRNA cluster miR-106b~25 regulates adult neural stem/progenitor cell proliferation and neuronal differentiation. *Aging (Albany NY).* 2011; 3:108-124.
37. McNeill E, Van Vactor D. MicroRNAs Shape the Neuronal Landscape. *Neuron.* 2012; 75:363-379.
38. Hua Y-J, Tang Z-Y, Tu K, Zhu L, Li Y-X, Xie L, Xiao HS. Identification and target prediction of miRNAs specifically expressed in rat neural tissue. *BMC Genomics.* 2009; 10:214-225.
39. Liang Y, Ridzon D, Wong L & Chen C. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics.* 2007; 8:166-185.
40. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foà R, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell.* 2007; 129:1401-1414.
41. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, Schmittgen TD. Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA.* 2008; 14:35-42.
42. Lugli G, Torvik VI, Larson J & Smalheiser NR. Expression of microRNAs and their precursors in synaptic fractions of adult mouse forebrain. *J Neurochem.* 2008; 106:650-661.
43. Kye MJ, Liu T, Levy SF, Xu NL, Groves BB, Bonneau R, Lao K, Kosik KS. Somatodendritic microRNAs identified by laser capture and multiplex RT-PCR. *RNA.* 2007; 13:1224-1234.
44. O'Carroll D, Schaefer A. General Principles of miRNA Biogenesis and Regulation in the Brain. *Neuropsychopharmacology.* 2012 Jun 6. [Epub ahead of print].
45. Schrott G. microRNAs at the synapse. *Nat Rev Neurosci.* 2009; 10:842-849.
46. Pichardo-Casas I, Goff LA, Swerdel MR, Athie A, Davila J, Ramos-Brossier M, Lapid-Volosin M, Friedman WJ, Hart RP, Vaca L. Expression profiling of synaptic microRNAs from the adult rat brain identifies regional differences and seizure-induced dynamic modulation. *Brain Res.* 2012; 1436:20-33.
47. Satoh J-i. MicroRNAs and Their Therapeutic Potential for Human Diseases: Aberrant MicroRNA Expression in Alzheimer's Disease Brain. *J Pharmacol Sci.* 2010; 114:269-275.
48. Cogswell JP, Ward J, Taylor IA, Waters M, Shi Y, Cannon B, Kelnar K, Kempainen J, Brown D, Chen C, Prinjha RK, Richardson JC, Saunders AM, Roses AD, Richards CA. Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. *J Alzheimers Dis.* 2008; 14:27-41.
49. Tacutu R, Budovsky A, Yanai H, Fraifeld VE. Molecular links between cellular senescence, longevity and age-related diseases - a systems biology perspective. *Aging (Albany NY).* 2011; 3:1178-1191.
50. Provost P. Interpretation and applicability of microRNA data to the context of Alzheimer's and age-related diseases. *Aging (Albany NY).* 2010; 2:166-169.
51. Yao J, Hennessey T, Flynt A, Lai E, Beal MF, Lin MT. MicroRNA-related cofilin abnormality in Alzheimer's disease. *PLoS One.* 2010; 5:e15546.
52. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Brian KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A.* 2008; 105:10513-10518.
53. Mizuno H, Nakamura A, Aoki Y, Ito N, Kishi S, Yamamoto K, Sekiguchi M, Takeda S, Hashido K. Identification of muscle-specific microRNAs in serum of muscular dystrophy animal models: promising novel blood-based markers for muscular dystrophy. *PLoS One.* 2011; 6:e18388.
54. Cuk K, Zucknick M, Heil J, Madhavan D, Schott S, Turchinovich A, Arlt D, Rath M, Sohn C, Benner A, Junkermann H, Schneeeweiss A, Burwinkel B. Circulating microRNAs in plasma as early detection markers for breast cancer. *Int J Cancer.* 2012 Aug 28. [Epub ahead of print].
55. Pichardo-Casas I, Goff LA, Swerdel MR, Athie A, Davila J, Ramos-Brossier M, Lapid-Volosin M, Friedman WJ, Hart RP, Vaca L. Expression profiling of synaptic microRNAs from the adult rat brain identifies regional differences and seizure-induced dynamic modulation. *Brain Res.* 2012; 1436:20-33.
56. Martinez I, Almstead LL, DiMaio D. MicroRNAs and senescence. *Aging (Albany NY).* 2011; 3:77-78.
57. Inukai S, de Lencastre A, Turner M, Slack F. Novel MicroRNAs Differentially Expressed during Aging in the Mouse Brain. *PLoS One.* 2012; 7:e40028.
58. Benhamed M, Herbig U, Ye T, Dejean A, Bischof O. Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. *Nat Cell Biol.* 2012; 14:266-275.
59. Smith-Vikos T, Slack FJ. MicroRNAs and their roles in aging. *J Cell Sci.* 2012; 125(Pt 1):7-17.
60. Li X, Khanna A, Li N, Wang E. Circulatory miR34a as an RNA-based, noninvasive biomarker for brain aging. *Aging (Albany NY).* 2011; 3:985-1002.
61. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from National Institute on Aging-Alzheimer's

Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:263-269.

62. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:270-279.

63. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:280-292.

64. Abdullah L, Luis C, Paris D, Mouzon B, Ait-Ghezala G, Keegan AP, Wang D, Crawford F, Mullan M. Serum Abeta levels as predictors of conversion to mild cognitive impairment/Alzheimer disease in an ADAPT subcohort. *Mol Med.* 2009; 15:432-437.

65. Morris J, Heyman A, Mohs R, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C. The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology.* 1989; 39:1159-1165.

66. Wechsler D. (1987) Wechsler Memory Scale – Revised. San Antonio, TX: The Psychological Corporation.

67. Benton A, Sivan A, Hamsher K, Varney N, & Spreen O. (1983) Judgment of Line Orientation. Psychological Assessment Resources, Inc., Lutz, FL.

68. Wechsler D. (1997) Wechsler Adult Intelligence Scale – 3rd Edition. San Antonio, TX: The Psychological Corporation.

69. Luis CA, Abdullah L, Ait-Ghezala G, Mouzon B, Keegan AP, Crawford F, Mullan M. Feasibility of Predicting MCI/AD Using Neuropsychological Tests and Serum  $\beta$ -Amyloid. *Int. J. Alzheimers Dis.* 2011; 2011 Article ID 786264.

70. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984; 34:939-944.

71. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol.* 1999; 56:303-308.

72. Eng J. Sample Size Estimation: How Many Individuals Should Be Studied? *Radiology.* 2003; 227:309-313.

73. Peltier HJ, Latham GJ. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA.* 2008; 14:844-852.

74. Hennessey PT, Sanford T, Choudhary A, Mydlarz WW, Brown D, Adai AT, Ochs MF, Ahrendt SA, Mambo E, Califano JA. Serum microRNA biomarkers for detection of non-small cell lung cancer. *PLoS One.* 2012; 7:e32307.

75. Matthaehi H, Wylie D, Lloyd MB, Dal Molin M, Kempainen J, Mayo SC, Wolfgang CL, Schulick RD, Langfield L, Andruss BF, Adai AT, Hruban RH, Szafranska-Schwarzbach AE, Maitra A. miRNA Biomarkers in Cyst Fluid Augment the Diagnosis and Management of Pancreatic Cysts. *Clin Cancer Res.* 2012 Aug 21. [Epub ahead of print].

## SUPPLEMENTARY MATERIALS

### Software, Calculations, Graphical Interface Program overview

- The data is imported from Excel spreadsheet using special wizard; no template for submitted data needed. Data can be raw Ct values, or processed, e.g. normalized data;
- Several groups of patients (Target groups) can be specified. Any group can be designated as "Control"; differentiation of all other groups from the Control group can be evaluated;
- ROC parameters are calculated for any miRNA combination, including complex biomarkers, comprising more than one miRNA as numerator, denominator or both;
- Several kinds of customized graphs are displayed, including Points and Box Plots for a set of data, 2D graphs for two markers, ROC graphs for several markers and their combinations, histograms.

### Data Flow and Technology

- The software is created in .Net V4.0 as a WinForm application, using C# and VS 2010 / VSS as Development environment. For Math and Statistic calculations, MS Excel COM interface is utilized. For histogram support, MathIridium library is used.
- After importing, data converted to internal format and saved for future use in this format, preventing any inadvertent data altering. Along with source data, normalization settings & results, the last status of data analysis (Control and Target groups), are saved. For Markers and Targets, aliases are allowed.

### Normalization

Calculations of normalized values is performed according to the following equation:

$C = 2^{-(Ct [\text{miRNA1}] - Ct [\text{miRNA2}])} * M$ , where

**Ct [miRNA2]** – experimentally determined Ct value for a potential denominator;

**Ct [miRNA1]** – experimentally determined Ct value for a potential numerator;

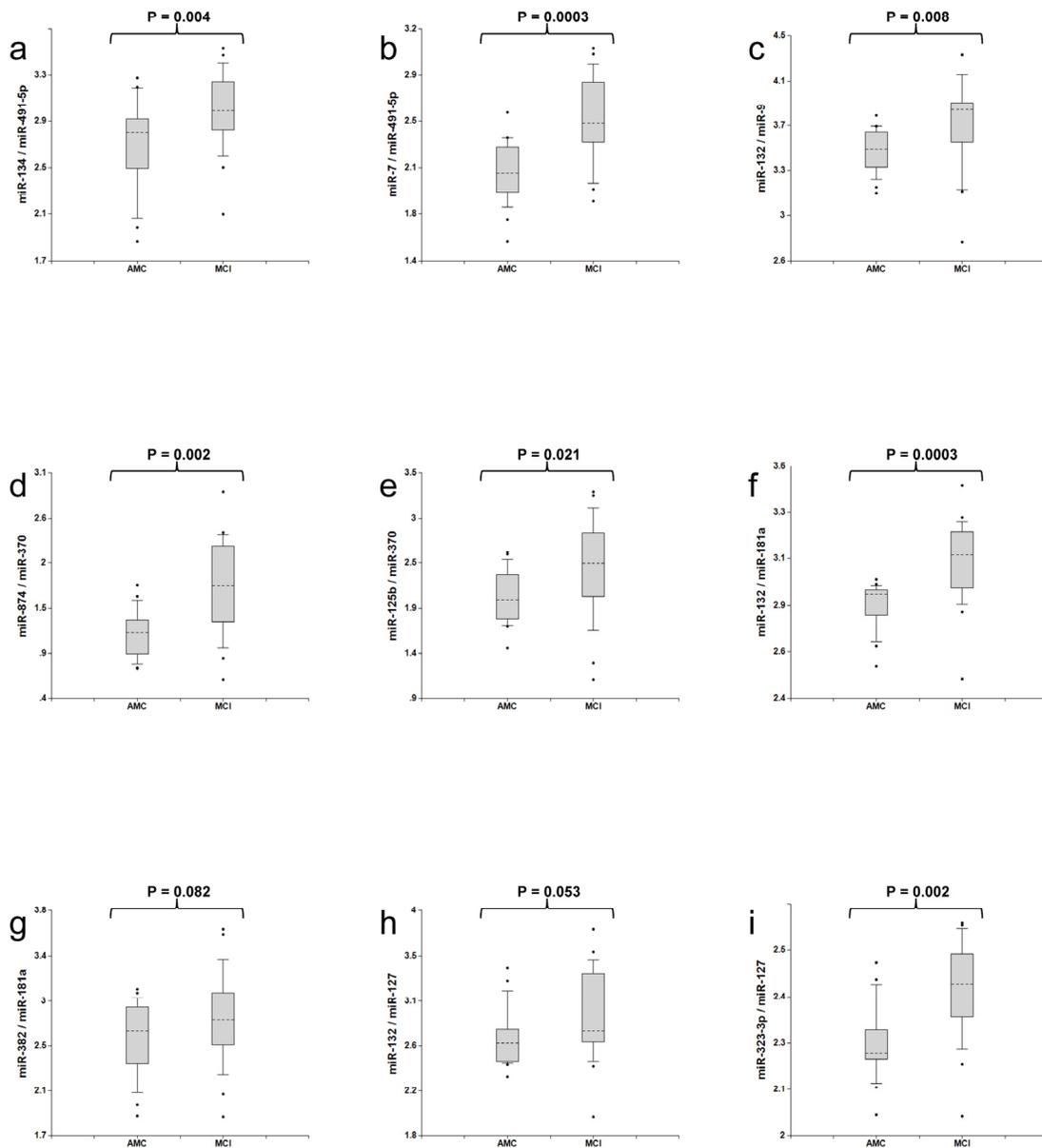
M is a constant multiplier, chosen, so that to keep the normalized values C in a computationally convenient range, 10 – 1000.

### Statistics

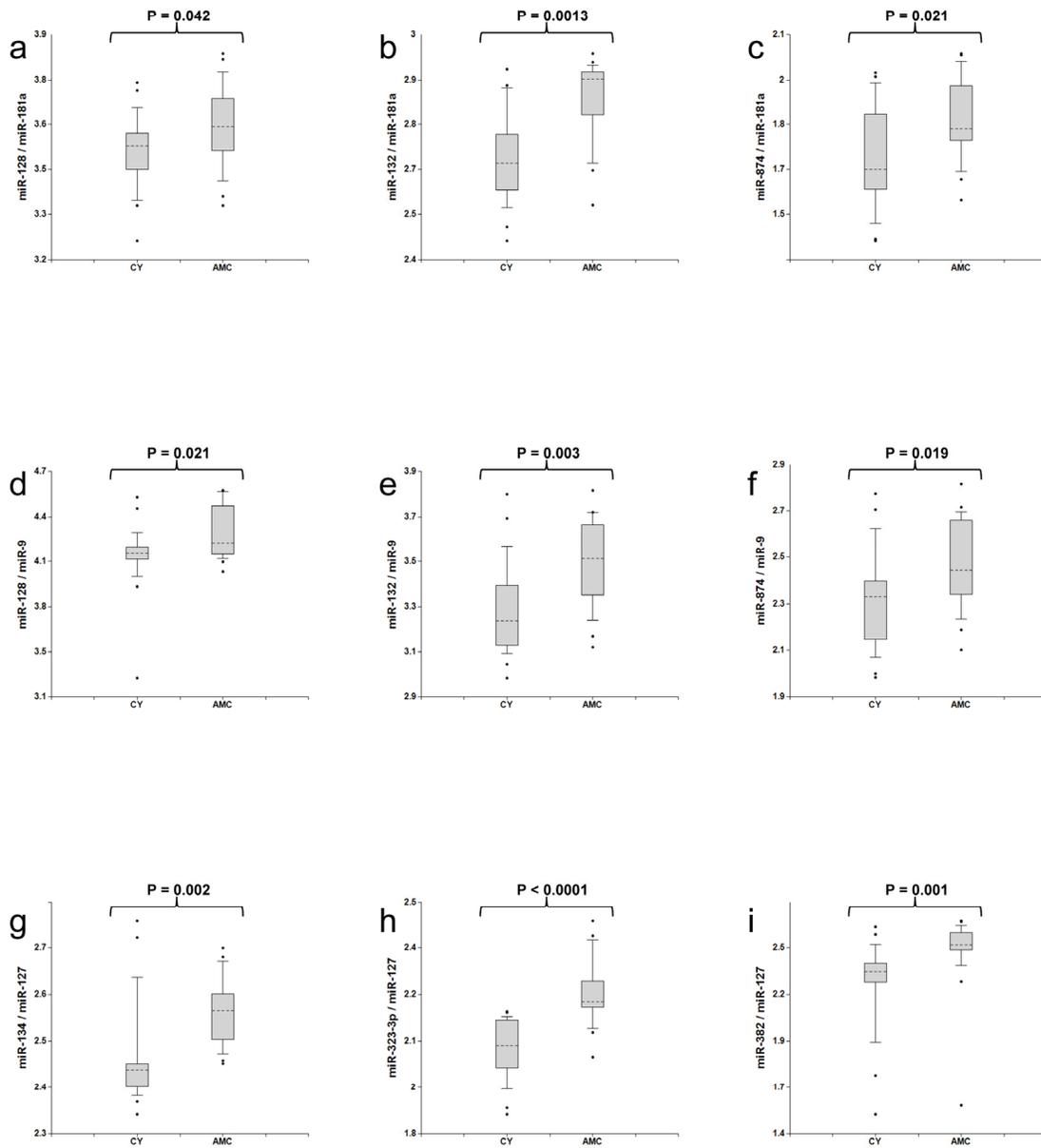
Along with regular data set parameters (average, median, standard deviation, finding outliers), some parametric and not-parametric statistics are used:

- Chi-square test with histograms – to define normality of data set;
- T – test (Student) of significance for normally distributed data;





**Figure S1: Ratios of miRNA levels (biomarker pairs) in plasma of MCI patients and age-matched controls.** The ratios of concentrations of miRNA pairs in plasma samples of MCI and age-matched donors with normal cognitive function, 20 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as  $2^{-\Delta\Delta Ct} \times 100$ . See the Fig. 1 legend for the description of the statistical analysis.



**Figure S2. Comparison of miRNA biomarker pairs in plasma of Group 1 (30-50 years old, "CY") and Group 2 (70-80 years old, "AMC") individuals with normal cognitive functions.** The ratios of concentrations of miRNA in plasma samples of Group1 (30-50 years old, CY) and Group2 (70-80 years old, AMC) donors with normal cognitive function, 20 samples in each group, were measured by RT-PCR and the ratio of various miRNA was calculated as  $2^{-\Delta\Delta C_t} \times 100$ . See the legend to Fig. 1 for the description of the statistical analysis.

## Circulatory miR-34a as an RNA-based, noninvasive biomarker for brain aging

Xiaoli Li<sup>2</sup>, Amit Khanna<sup>2</sup>, Na Li<sup>2</sup>, and Eugenia Wang<sup>1,2</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, University of Louisville School of Medicine, Louisville, Kentucky, USA

<sup>2</sup>Gheens Center on Aging, University of Louisville School of Medicine, Louisville, Kentucky, USA

**Running title:** Circulatory miR-34a as a biomarker for aging

**Keywords:** MicroRNAs, aging, biomarker, SIRT1, Bcl2, miR-34a, and miR-196a

Received: 9/22/11; Accepted: 10/11/11; Published: 10/16/11

Correspondence to Eugenia Wang, Eugenia.Wang@Louisville.edu

**Copyright:** © Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** MicroRNAs in blood samples have been identified as an important class of biomarkers, which can reflect physiological changes from cancer to brain dysfunction. In this report we identify concordant increases in levels of expression of miR-34a in brain and two components of mouse blood samples, peripheral blood mononuclear cells (PBMCs) and plasma, from 2 day old neonates through young adulthood and mid-life to old age at 25 months. Levels of this microRNA's prime target, silent information regulator 1 (SIRT1), in brain and the two blood-derived specimens decrease with age inversely to miR-34a, starting as early as 4 months old, when appreciable tissue aging has not yet begun. Our results suggest that: 1. Increased miR-34a and the reciprocal decrease of its target, SIRT1, in blood specimens are the accessible biomarkers for age-dependent changes in brain; and 2. these changes are predictors of impending decline in brain function, as early as in young adult mice.

### INTRODUCTION

The list of genetic factors pertinent to life-span determination has been growing by leaps and bounds, to define age-related physiological changes [1]. A steady loss of function in multiple vital organs has been shown to correlate with aging, accompanied by increased incidence of a wide range of diseases such as neurological disorders, metabolic disorders *e.g.* diabetes, and cancer [2]. Modification of signaling cascades, such as highly conserved signaling pathways including insulin/Insulin-like growth factor 1 (IGF1), target of rapamycin (TOR) and sirtuin, for enhanced cellular responses to stress, is recognized as key to life span extension and accompanying reduced age-related pathology [3]. The search for biomarkers for aging has led to the identification of unique leads in blood samples, a relatively noninvasive method to acquire experimental specimens. However, the main findings in

circulating blood reflecting systemic phenomena are protein-based changes, mostly connected to age-related perils such as cancer, atherosclerosis, and cardiovascular dysfunction [4]. A few studies linking blood-based changes to normal aging involve a serum protein pattern that proves to be a reliable index for aging in rat, independent of pathologies [5]; the techniques commonly used, 2-D gel electrophoresis or proteomic profiling, have led to identifying various proteins as potent blood-based biomarkers [6]. Another such example is the observation that circulating IGF-1 in centenarians with robust cognition is low [7], while those suffering Alzheimer's disease (AD) show high levels of this protein in their blood [8].

Non-coding RNAs are prominent epigenetic factors, along with nucleic acid modifications such as DNA methylation and histone/chromatin modification [9]. Among all the small non-coding RNAs, microRNAs

(miRNAs) have been studied in detail; they are generally transcribed by RNA polymerase II, rarely by RNA polymerase III [10]. MiRNAs modulate protein regulation at the post-transcriptional level, because their seed sequences have perfect or partial complementarity to the coding region or the 3'- untranslated region (UTR) of one or more target mRNAs, leading to mRNA degradation in the former case, and inhibition of protein translation in the latter; either way, they serve as negative regulators of gene expression [10, 11]. Early studies illustrating the role of miRNAs in aging found that expression of *C. elegans* lineage 4 (*lin-4*) miRNA is needed for extended lifespan, while abridged expression leads to shortened lifespan [11, 12]. We also reported a set of miRNA expressions unique to peripheral blood mononuclear cells (PBMCs) of AD victims [13], as well as changes in miRNA expression during aging in brain and liver of mice and rats, including long-lived calorie-restricted and mutant mice with extended life span [14-17].

In general, blood samples can be separated into two major components: PBMCs, composed of lymphocytes, monocytes, megakaryocytes, platelets, etc.; and plasma or serum, depending on blood- collecting procedures (adding an anticoagulant produces the former instead of the latter). Our group has shown that miRNA profiles in PBMCs of AD victims differ from those of normal elderly controls (NEC); this led us to suggest that lead microRNAs in PBMCs of AD victims may be biomarkers as a blood-based diagnostic for this disease [18]. In cell-free serum/plasma, microRNAs are repeatedly reported to be not only present [19], but also a means for cell-cell communication, secretion, and many other cellular functions associated with cell death, etc. [5, 20, 21]. Recently, major differences in miRNA expression between plasma microvesicles and PBMCs were reported, owing to the origin of these cell-free miRNA expressions in the former case [20]. Plasma miRNA expression is a predictive and diagnostic tool for lung cancer; changes in plasma miRNA expression imply staging and prognosis [22]. Blood associated microvesicles contain miRNAs suggested for inter-cellular and inter-organ communication [23]. In brain, circulating microRNAs have been suggested to be vital for neuronal communication [24], and lead microRNAs in blood serve as circulating biomarkers for bipolar disorder and early Huntington's disease [25].

In this study we report that miR-34a levels in mouse PBMCs and plasma increase with age, as do those in brain, starting at 4 months of age, as documented in

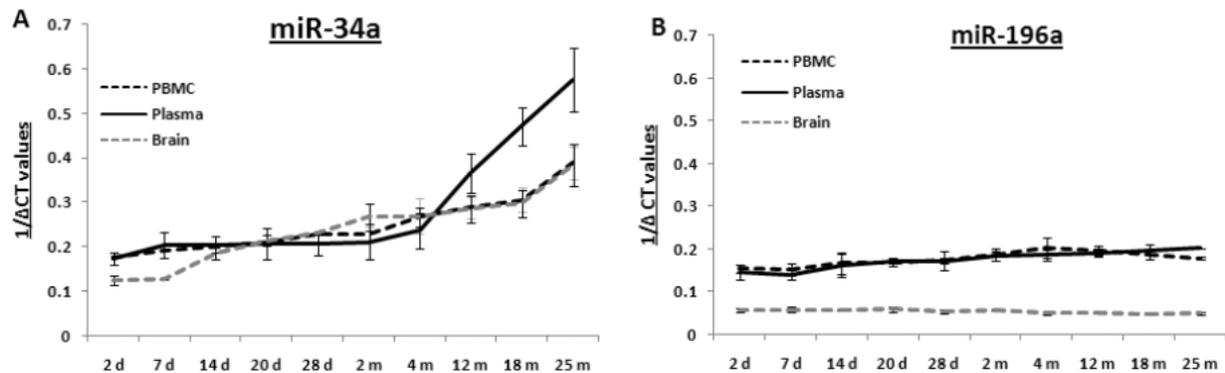
samples from 2 days old till 25 months of age. Parallel study of miR-196a shows that the level of this microRNA remains stable with age in all three specimens examined, and therefore serves as a control for age-independent regulation of its expression. Corresponding to this increase of miR-34a, plasma expression of SIRT1, the major target of this microRNA, shows a precipitous reciprocal decrease starting at 4 months of age, while PBMCs exhibit a gradual decrease in SIRT1 from this age onward. In contrast, the SIRT1 level in brain rises from 20 days old till 4 months, and remains at this level without decrease until 12 months of age. The reciprocal expression between miR-34a and SIRT1 in plasma samples is not observed with another target of this microRNA, Bcl2, whose expression in plasma actually declines as early as 7 days after birth and continues to 25 months, while in brain and PBMCs it shows a slight decrease during the same time frame. In brief, our results show here that levels of miR-34a in plasma and PBMCs may serve as a non- invasive biomarker, a circulating 'footprint' of brain aging; in particular the rise in the plasma is detected as early as 4 months of age, before any impending decline in this organ begins.

## RESULTS

### Expression levels of miRNAs in blood and brain samples of C57/B6 mice

Mouse microRNA 34a (*mmu-miR-34a*) expression increases in brain of older rodents, while a decline in its expression over age is linked to longevity in calorie-restricted mouse brain [13, 16]. SIRT1 and Bcl2 are two major target genes suppressed by miR-34a [17, 26]. In contrast, although miR-196a is elevated in Crohn's disease [27], changes in its level of expression during mouse aging is not yet reported. Thus, our study of microRNA expression over a life span from neonatal to old age was performed with miR-34a as our focus of interest, and miR-196a as a control.

Total RNA samples extracted from PBMCs, plasma and brain specimens from age groups from 2 days to 25 months were reverse transcribed for quantitative PCR (qPCR) to determine their levels of *mmu-miR-34a* and *mmu-miR-196a*. qPCR for these two miRNAs was performed using total RNA samples from all three specimens (Supplemental Figure 1-2); composite graphs of expression of these two miRNAs are shown in Figure 1. Levels of miR-34a are stable from 2 days till 4 months, with subsequent rises in all three tissues; the increases in brain and PBMCs exhibit a similar



**Figure 1: Age-dependent expression levels of two miRNAs in blood and tissue samples of C57/B6 mice** A graphical representation of expression levels of miRNAs using qualitative PCR, represented as  $1/\Delta Ct$  values as box plots for brain, PBMC and plasma samples. Age groups included are from early to old age [from 2 days (d) to 25 months (m)]. Panel (A) shows composite graphs for expression of miR-34a, while panel (B) shows the levels of expression of miR-196a in brain, PBMCs, and plasma. (n = 3; three different animals were used from each age group selected for the study.)

gradual trend, while the increase in plasma samples rises steeply starting at 4 months. All assays for each animal were repeated three times, and three animals were used per age group; thus, the qPCR results presented here were calculated from nine data points, with three repeats of three different animals. In contrast to the increased miR-34a expression with age, levels of miR-196a remain stable throughout all age groups studied here, from 2 days till 25 months of age. Statistical analysis for each age group is presented in Supplemental Table 1.

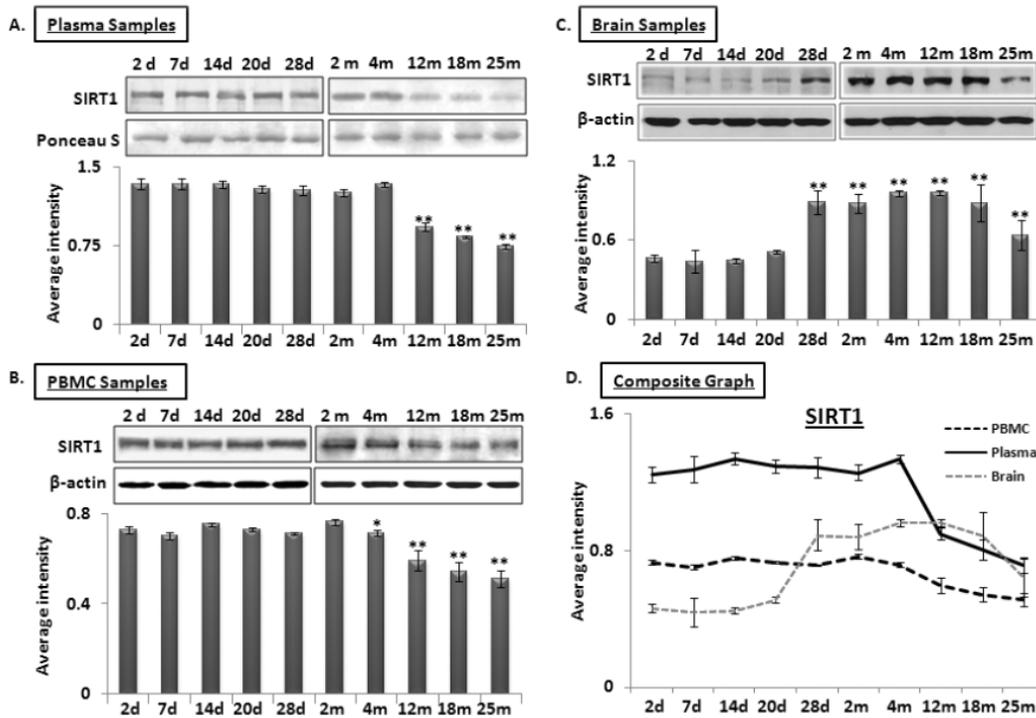
### Inverse levels of expression between miR-34a and its target, SIRT1, in blood and brain during aging

MicroRNA-34a regulates SIRT1 expression both at the pre-transcriptional level, by regulating expression of transcription factor SP1 [16], and also at the post-transcriptional level, by binding to the 3' untranslated region (UTR) of SIRT1 transcript [28] to repress its expression. Total protein samples extracted from plasma, PBMCs, and brain were subjected to Western blot analysis to determine a possible inverse relationship between SIRT1 and miR-34a levels from 2 days to 25 months old. Decreased SIRT1 expression was observed in both plasma and PBMC samples at 4 months of age, with the former showing a precipitous drop in SIRT1 abundance (Figure 2A & B). However, levels of SIRT1 in brain increase from a steady state during the postnatal period to a stable high level from 28 days to 12 months, before decreasing to 25 months (Figure 2C). A composite graph depicting the expression trends in the three specimens across all age groups shows the precipitous decline in plasma, but gradual decline in PBMCs from 4 months on, and the

brain-specific rise at 28 days followed by the decline at 12 months old (Figure 2D). In all, the only sharp inverse relationship detected at early adult life (4-months) between miR-34a and its target, SIRT1, is observed in plasma. All immunoblots were further verified by densitometric measurements of three repeats with three different animals of the same age group, after normalization with  $\beta$ -actin in brain and PBMC samples. Ponceau S stained bands were used to validate equal loading, and selected bands showing consistent levels across the blots were used to normalize SIRT1 expression in plasma samples. The individual graphs may be found as supplemental data (Supplemental Figure 3).

### Expression of miRNA-34a and its target SIRT1 in brain tissue sections from middle and old age C57/B6 mice

A lack of age-dependent increase in miR-34a expression in the hippocampal region has been previously reported in calorie-restricted mice, as compared to littermate wild type (WT) controls [17]. Here, we performed similar in situ histochemical (ISH) determination of miR-34a, to validate the qPCR results with total brain RNA studies by locked nucleic acid (LNA) probes for this microRNA. Two selected age groups, 10 and 31 months, were chosen to represent specimens before SIRT1 decreases, and beyond the maximal decline. Brain sections of these two age groups were processed for levels of miR-34a expression by in situ hybridization, binding with a locked nucleic acid probe for miR-34a, as well as a control probe for background binding reaction (Figure 3 A & B).



**Figure 2: Age-dependent expression levels of SIRT1 in blood and tissue samples of C57/B6 mice** Western blot analysis of SIRT1 expression in age groups from 2 days (d) to 25 months (m); panels (A) SIRT1 expression in plasma samples normalized with selected Ponceau S stained band, showing constant levels for all samples used, along with histograms presenting average densitometric values, (B) SIRT1 expression in PBMC samples normalized with  $\beta$ -actin, along with histograms presenting average densitometric values, (C) SIRT1 expression in brain samples normalized with  $\beta$ -actin, along with histograms presenting average densitometric values, (D) Composite graph presenting SIRT1 expression in plasma, PBMC and brain samples. (\* $p < 0.01$ , \*\* $p < 0.0001$ ; all histograms represent Mean  $\pm$  SD;  $n = 3$ ; three different mice used from each selected age group.)

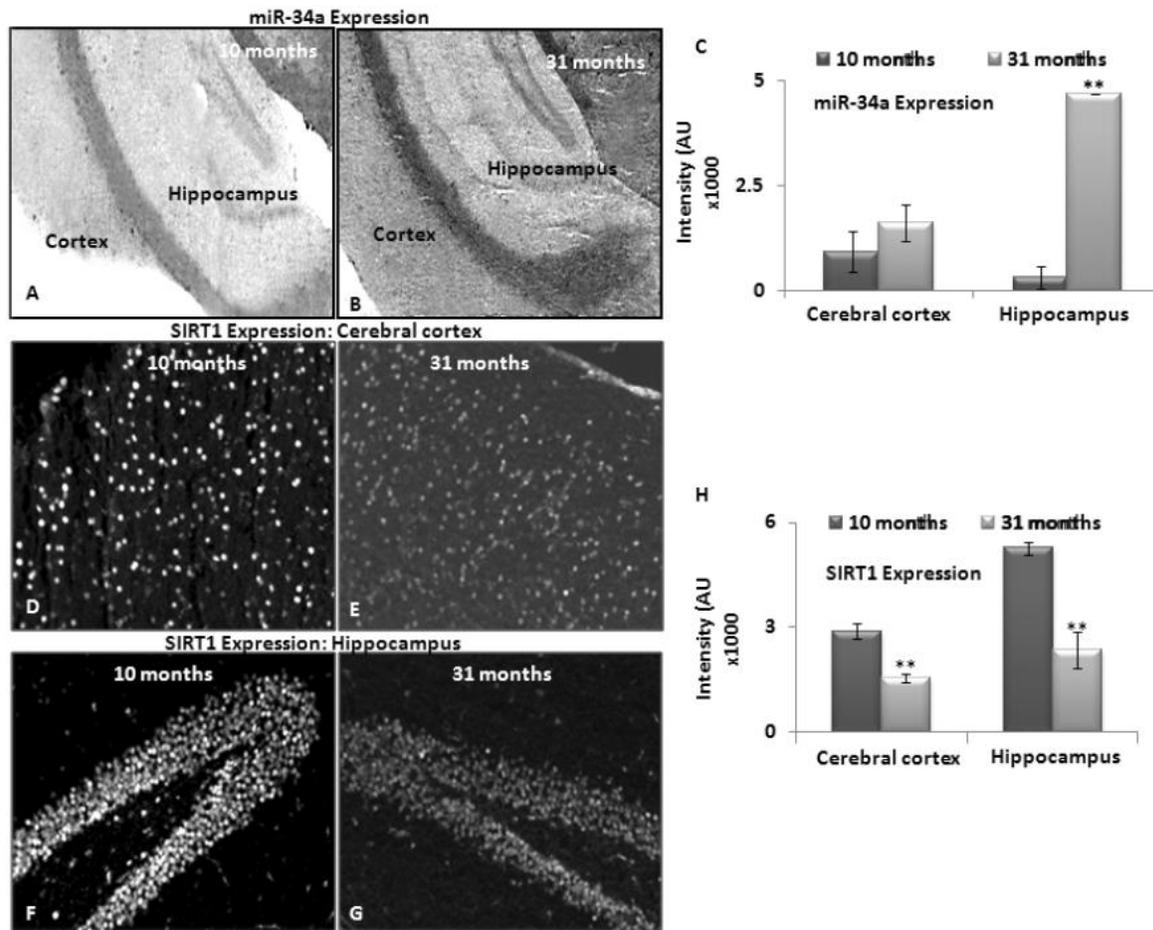
Intensity of LNA probe reaction was most noted in the hippocampus, more than the cortex, in both age groups. This intensity differential was further verified by densitometric measurements of three serial sections each from three different animals of the same age groups (Figure 3C). Expression data obtained using in situ hybridization confirmed the qPCR results, and validated the trend of increased miR-34a expression from 10 to 31 months of age in brain.

Immunostaining to detect SIRT1 in middle and old age, i.e. 10 and 31 months respectively, was performed on sister specimens of those used for in situ hybridization with the LNA probe for miR-34a. Decreased SIRT1 from 10 to 31 months was observed most significantly in hippocampus, validating the inverse relationship with miR34a expression in brain, described by the grind-and-find qPCR assays and Western blotting for SIRT1 levels. Figure 3D-G shows in situ hybridization results of heightened miR-34a expression in hippocampus, corresponding to decreased SIRT1 presence. All immunostaining results were further verified by

densitometric measurements of three repeats with three different animals ( $n = 3$ ) (Figure 3C, H).

#### Levels of Bcl-2, another target of miR-34a, in blood and tissue samples of C57/B6 mice.

Another target of interest of miR-34a is Bcl2. Western blots for 26 kilodalton Bcl2 expression were performed using protein samples of brain, PBMCs and plasma from all age groups used for the qPCR assays of the two microRNAs. A decline in Bcl2 was observed starting as early as 7 days after birth in plasma (Figure 4A-C), continuing until old age at 25 months. In contrast, Bcl2 in brain and PBMCs show stable abundance, with a gradual decrease from 2 months to old age. A composite graph showing the decline in Bcl2 is shown in Figure 4D; the individual graphs may be found as supplemental data (Supplemental Figure 4). Interestingly, the decrease of Bcl2 in plasma starts at 7 neonatal days, and continues its downward trend in plasma samples. All immunoblot data were verified by densitometric measurements of three repeats with three



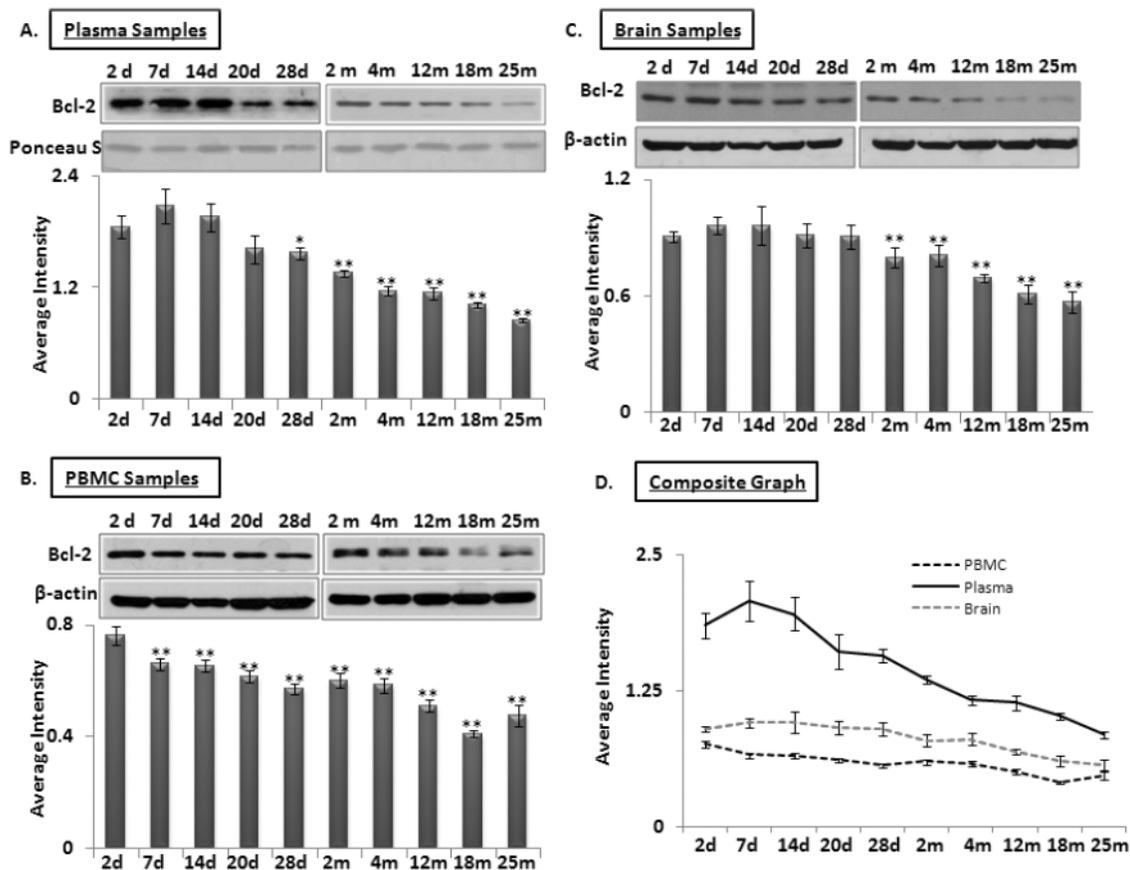
**Figure 3: Expression of miRNA-34a and SIRT1 in brain tissue sections from middle and old age C57/B6 mice** *In situ* hybridization detected miR-34a expression using LNA probes. Panels (A) and (B) show expression of miR-34a in cortex and hippocampus of 10 and 31 month old mice. Scrambled LNA probes were used as control for this experiment (Image not shown); (C) graphical presentation of mean intensity values of miR-34a expression in cortex and hippocampus of 10 and 31 month old mice, by immunostaining assay to detect SIRT1 expression. Panels (D-E) and (F-G) show expression of SIRT1 in cortex and hippocampus of 10 and 31 month old mice, by DAPI staining to detect nucleus and section integrity (Image not shown); (H) graphical presentation of mean intensity values of SIRT1 expression in cortex and hippocampus of 10 and 31 month old mice. (\* $p < 0.01$ , \*\* $p < 0.0001$ ; all histograms represent Mean  $\pm$  SD;  $n = 3$ ; three different biological samples from each age group.)

different animals, after normalization with  $\beta$ -actin for tissues and PBMC samples. For normalization of Bcl2 expression in plasma samples, Ponceau S stained bands were used to confirm equal loading ( $n = 3$ ).

### Age-dependent changes of levels of p53 and its acetylated form in blood and brain samples of C57/B6 mice.

SIRT1 is an evolutionarily conserved molecule with deacetylation properties; one of its targets is p53, whose acetylation leads to increased transcriptional activation of miR-34a expression [29]. Western blot analysis was performed using total protein samples from PBMCs, plasma and brain from mice aged 2 days to 25 months. Expression of p53 and its acetylated form was detected;

Ponceau S stained bands for plasma samples and  $\beta$ -actin for PBMCs and brain were used as normalization controls ( $n = 3$ ). Figure 5 shows levels of total and acetyl-p53 in plasma (panel 5A), PBMCs (panel 5B) and brain (panel 5C), with the percentage of the latter in the total protein pool illustrated in Panel 5D. The acetyl-p53 proportion remains stable from 2 days until 2 months in both plasma and PBMC specimens; this is followed by a steady increase between 4 and 12 months, after which it stays the same in the remaining old age groups, as observed in brain. The acetyl-p53 fraction in brain exhibits a steady increase from 2 days old till old age. Interestingly, proportions of acetyl-p53 in the total p53 protein pool in all three specimens, plasma, PBMCs, and brain, are similar from 12 to 25 months of age, again suggesting that circulating blood



**Figure 4: Age-dependent expression levels of Bcl-2 in blood and brain tissue samples of C57/B6 mice** Western blot analysis for Bcl-2 levels in age groups from 2 days (d) to 25 months (m); panels (A-B-C) Bcl-2 expression normalized with Ponceau S stained bands in plasma, and with β-actin in brain and PBMC samples, along with histograms presenting average Bcl-2 expression densitometric values; (D) Composite graph presenting Bcl-2 expression in plasma, PBMC and brain samples as average densitometric intensity. All graphs represent Mean ± SD. n = 3; three different mice from each age group.

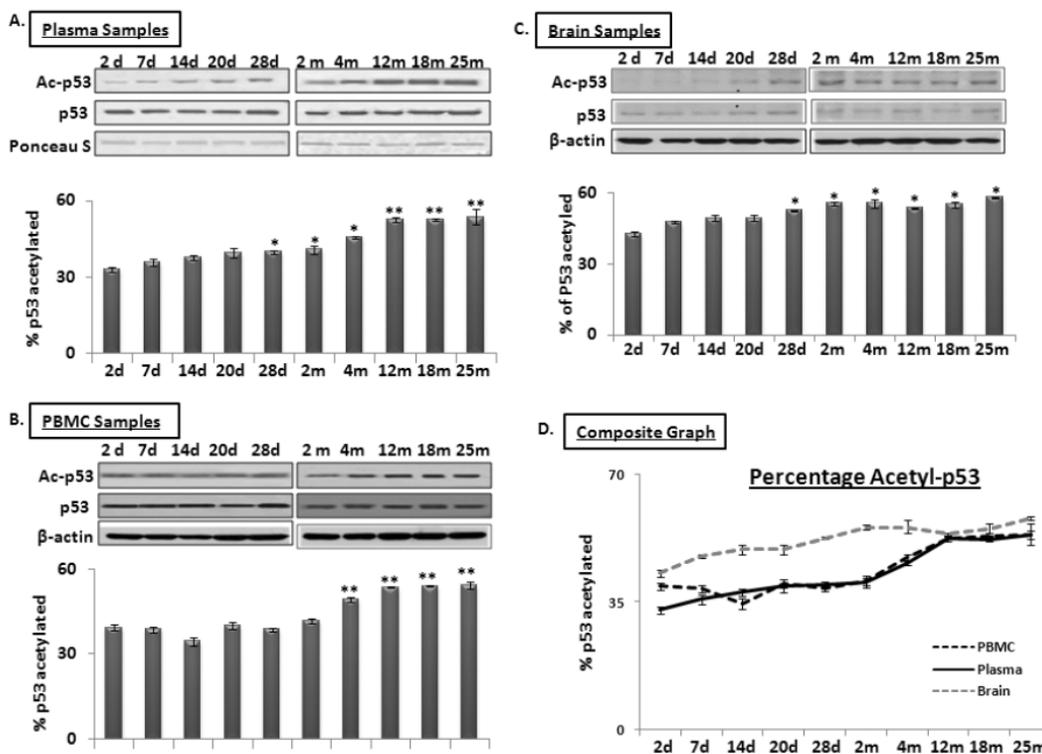
acetyl/total p53 may be an additional systemic marker for age-related changes in brain.

## DISCUSSION

Central to our findings is the observation that miRNA-34a expression in PBMCs and brain shows concordant steady increase from neonatal to old age, with levels of miR-34a in plasma showing a sharp rise from 4 to 25 months of age. This rapid gain of miR-34a in plasma is inversely related to SIRT1 abundance in this blood specimen, which declines precipitously from 4 to 25 months. However, in brain and PBMCs, this reciprocal expression between increased miR-34a and decreased SIRT1 is not observed until after 12 months of age. This pattern, starting at 4 months in plasma and 12 months in brain and PBMCs, is not observed with another major target of this microRNA, Bcl2, whose downward trend starts at neonatal time in plasma, while in brain and PBMCs it remains stable. In situ hybridization data and

histochemical studies indicate the hippocampus as the venue of the most prominent local increase in expression level of miR-34a with age, and reciprocal decrease of SIRT1 in the same brain region in old mice [26]. Taken together, our results suggest that: 1. Increased levels of expression of miR-34a in plasma and PBMCs correspond to those observed in brain, with the former most dramatically preceding the latter two; and 2. the sharp increase in plasma miR-34a, and the decrease of SIRT1, starting at 4 months of age, may serve as noninvasive biomarkers for impending age-dependent brain decline at 12 months or later in mice.

The major difference between the two blood samples, PBMCs and plasma, is the source of their RNA and protein specimens. PBMCs largely consist of cells making up this component, including lymphocytes, monocytes, macrophages, etc., while plasma components are released from various tissues into circulating blood through secretion, exocytosis, cell-to-



**Figure 5: Age-dependent changes in ratio of acetylated to total p53 in blood and brain samples of C57/B6 mice.**

Western blot analysis of acetylated p53 proportion in age groups from 2 days to 25 months; **Panels A, B, and C** show acetylated P53 proportion normalized with Ponceau S in plasma (Panel A), and with  $\beta$ -actin in PBMCs and brain samples, along with histograms presenting the average densitometric values. **Panel D** presents a composite graph of the percentage of acetylated P53 expression in the total p53 protein pool in plasma, PBMC and brain samples. (\* $p < 0.01$ , \*\* $p < 0.0001$ ; all histograms represent Mean  $\pm$  SD;  $n = 3$ ; three different biological samples from each age group)

cell communication, and even cell death, including both apoptosis and necrosis [30]. Thus, in mammals, cell-free plasma RNA and protein may be true ‘foot-prints’ of the entire organism’s health status. Obviously, changes in PBMC RNA and protein profiles follow the same age-dependent regulation as other organs, i.e. cell-type specific. Nevertheless, our observations suggest that increased levels of miR-34a in PBMCs and plasma may be a noninvasive biomarker reflecting changes in brain during aging, with increased levels of this microRNA in plasma potentially serving as an early biomarker for impending changes in the brain.

Functionally, miR-34a is well recognized as a tumor suppressor in brain and many other tissues; its absence is associated with neoplastic growth, including glioma and brain tumor [31], explaining the needed increase during early adult life for protection against neoplasms of many cell types, including those in the central nervous systems (CNS). However, its continuing increase presents a classical case of antagonistic pleiotropy, i.e. a genetic trait beneficial in early life but

posing adverse consequences in later life. This is largely due to the fact that, as do many other miRs, miR-34a silences multiple targets, suppressing cell-cycle traverse genes such as cdks and cyclins; but its silencing action on SIRT1 may be the most detrimental [31-33]. Although the notion of a direct role for SIRT1 and its sisters in extending life span has come into question recently [34], they clearly suppress age-dependent pathologies, reducing diabetes, obesity, neurodegeneration, etc. [35]. Thus, miR-34a’s continuing rise in late life may cause a reduction of SIRT1, a loss ill afforded by older organisms [35-37].

Complicating further the see-saw relationship between miR-34a and SIRT1 expression is the role of p53 as transcriptional activator for this microRNA’s expression [33]. Activation of miR-34a is regulated via binding of acetylated p53, which in turn is controlled by deacetylation by mammalian SIRT1, thus forming a feedback loop [33]. At first glance, our results may suggest that this loop may be disrupted in older mouse brain and PBMCs, as well as in many other tissues with

their 'foot-print' in the plasma, due to increased miR-34a, decreased Sirt1 levels, and increased levels of acetylated p53. However, a recent report by Lee, et al. [36] of activation of nuclear bile acid receptor, Farnesoid X Receptor (FXR), recruiting the Small Heterodimer Partner (SHP) to the p53 binding site in miR-34a's promoter region, and thus preventing this microRNA's activation and its downstream suppressing action on SIRT1, presents yet another layer of control. Abnormality of this positive feedback between FXR/SHP to decrease miR-34a and increase SIRT1 is observed in many age-related metabolic diseases [36]. Our finding of continuing rise of miR-34a with age suggests that either continuing acetyl-p53 increase overwhelming the SHP-binding or other putative factors involved, thus disable the FXR/SHP role in inhibiting this microRNA activation, and cause the absence of SIRT1 in many age-related metabolic diseases.

As noted above, among the various functional impacts of p53, from tumor suppression to induction of apoptosis, there emerges yet another vital role in regulating signaling networks through controlling miR-34a transcription activation. This adds to the complexity of how p53 governs life span determination; contrasting scenarios report high tumor incidence in p53 knockout mice, thus shortening life span, while p53 overexpression induces accelerated aging [38, 39]. Therefore, p53 regulation is another case of antagonistic pleiotropy, protective in the young but deleterious in older organisms. For example, in the context of regulating intertwining networks, p53 seems to participate in life extension as a downstream activator for apoptosis in the well-known IGF1/mTOR pathway, suppressing tumor development; but its continuous increase in older life may overwhelm SHP's binding to miR-34a and become the dominant activator for miR-34a, resulting in the SIRT1 loss described above. To further thicken the soup, in a large human population study of p53 polymorphism, the p53-P72 genotype is associated with reduced fertility and increased longevity [40]. In brain, the role of p53 is associated with cognitive robustness, by regulating glucose metabolism, as well as neuronal apoptosis, eliminating deleterious, damaged neurons to make room for neurogenesis [40]. Here, SIRT1 is involved in neuroprotective signaling, reducing the formation of  $\beta$ -amyloid, the pathogenic form of amyloid precursor protein (APP), one of the two main diagnostic histopathologies of AD, along with neurofibrillary tangles. Paradoxically,  $\beta$ -amyloid activates p53, which in turn activates miR-34a and suppresses SIRT1 [41], as one of the putative

mechanisms underlying accelerated aging when this tumor-suppressing transcriptional factor is overexpressed. The complexities presented here in p53-dependent regulation, and how miR-34a participates in this puzzle for brain aging, will demand future system biology investigation to unravel their mysterious but fascinating roles in regulating aging and life span determination. Nevertheless, our finding of the continuous increase of both acetyl-p53 and miR-34a in blood and brain during aging suggests them as two noninvasive biomarkers for aging in the central nervous system.

Circulating microRNAs, specifically in plasma, may reflect particular disease states such as cancer, cardiovascular disorders, and even neuronal dysfunction [18, 29]. Increases of miR-34a in various tissues during aging are well recognized in our own work and that of others, from liver to brain, and even in autopsy brains and PBMCs of AD victims [13, 16]. Our previous report [13] documented increased expression of miR-34a in AD patients' PBMCs, compared to NEC, associated with allelic inheritance of APOE4 [19, 42]. Although this systemic change in miRNA expression in PBMCs of AD patients is of prime significance, expression profiles in PBMCs may also indicate changes in the immune system in Alzheimer's disease [42]. This notion led to the recent suggestion that aberrant inflammation is an vital underlying AD pathogenesis, challenging the popular emphasis on amyloid plaques as the essential manifestation of the disease [43]. The kernel of this suggestion is the overexpression of cytokine interleukin-1 (IL-1) by microglia in the brain, thought to lead to neuronal deterioration in Alzheimer's disease [44]. Interestingly, upstream regulators for IL-1, i.e. IL-6 and IL-8, are secreted by senescent fibroblasts, associated with miR-146a/b, key microRNAs controlling inflammatory response [45]. Moreover, increased miR-146a is noted in AD brains and the brains of older mice, perhaps due to increased inflammatory response [46]. Linking the secretion from senescent fibroblasts to brain aging and neurodegeneration may suggest that heightened inflammatory response in aged animals and Alzheimer brain is a systemic manifestation involving all tissue types; plasma samples may be the best source to identify the factors involved. Future work using our same approach, i.e. baseline studies of the composition of this blood component, with miR-146a and others secreted by cells such as senescent fibroblasts, may yield more blood-based, age-dependent biomarkers in the category of increased inflammatory response.

Clearly, the present study is not comprehensive, covering all the miRNAs found in either PMBCs or plasma as biomarkers of brain aging, as noted above for miR-146a. However, our finding of miR-34a increasing monotonically with age, as a first example, demonstrates that circulating blood is a powerful and accessible window to visualize changes in vital organs which are not otherwise available, in the case of human studies during aging. Including our own work with Alzheimer's disease (AD), most human studies are limited by available resources and time required, and thus designed with a cross-sectional approach, comparing disease victims with age-matched controls, as reviewed by Provost, 2010 [47]. Longitudinal follow-up studies are rare and costly, and thus most microRNA studies for AD pathogenesis are limited to autopsy brain samples and animal models. Neither approach is ideal; the former is noted for its graveyard nature, imperfect neither for disease initiation nor progression, and animal models mostly use transgenic mice carrying human APP and/or Tau mutations mostly composed of familial AD polymorphisms as surrogates. The present work shows baseline concordant changes between PMBCs, plasma specimens, and brain, thus validating our human peripheral blood sample study with AD patients, pointing to circulating RNAs as an accessible and noninvasive biological source to detect changes in brain. In particular, changes in plasma microRNA profile may occur much earlier, seen here at 4 months, than the impending decline in brain. Our study presents the first example that circulating blood microRNAs, exemplified by miR-34a, may serve as biomarkers for brain aging. The RNA samples collected for this study will be an invaluable resource, when the entire repertoire of microRNA profiles from birth to old age in mice is obtained by future studies using deep sequencing; data obtained will serve as a baseline database for animal disease model studies in general, and as a possible surrogate for human studies as well, because of the cross-species conserved nature of this noncoding RNA species. The present study paves the way for the ultimate discovery of tissue-specific biomarkers in blood samples for inaccessible organs such as brain, and possibly even blood-based universal biomarkers for entire organismic aging.

## **MATERIALS AND METHODS**

### **Animals and Tissue Collection**

Mice of the C57/black 6 strain, from 2 days till 31 months old, were used in this study. For

immunohistochemistry and in situ hybridization assays, brains of males 10 and 31 months old were used, while the other assays used 2 days old till 25 months (2, 7, 14, 20 and 28 days, 2, 4, 12, 18, and 25 months). All animal work was approved by institutional (University of Louisville) biosafety board protocol #05-001. Brain and blood specimens, including both peripheral blood mononuclear cells (PBMC) and plasma, were obtained from at least three mice of each age group.

### **Processing blood for PBMC and plasma fractions**

After collection, individual mouse blood specimens were layered onto Ficoll-Paque Plus solution (GE Healthcare, Piscataway, NJ), containing EDTA to prevent coagulation, and centrifuged for 30 minutes at 1,500 x g to separate the blood samples into four layers, the plasma in the upper layer the next PBMC-containing fraction, as the white layer, then the Ficoll-Paque plus solution with the red blood cells at the bottom. The plasma and PBMC fractions were then collected and stored at -80°C until further processing for RNA and protein isolation.

Snap-frozen coronal brain sections were homogenized with Trizol reagent (Invitrogen, Carlsbad, CA), followed by the total RNA isolation, in parallel with the isolated PBMC specimens, with the RNease Mini Kit (Qiagen, Valencia, CA). The isolated RNA fractions were then dissolved in RNase-free water and stored at -80°C until use. The quality of the isolates was determined by the Agilent 2100 Bioanalyzer with the Agilent RNA 6,000 Nano kit (Agilent Technologies, Foster City, CA) by the RNA integrity number (RIN); samples with values > 7 are of acceptable integrity.

The RNA fraction was isolated by adding 0.75 ml of Trizol LS reagent (Invitrogen, Carlsbad, CA) to 0.25 ml plasma, followed by incubation at room temperature for 5 min. To this solution, 0.2 ml of chloroform was added; after vigorous mixing, it was further incubated for 5 minutes at room temperature, before centrifugation at 15,000 x g for 15 min at 4°C, to obtain the RNA aqueous phase. This was collected, and after addition of 0.5 ml isopropanol (Sigma, St. Louis, MO), RNA was precipitated and collected by further centrifugation to a pellet, before washing with 75% alcohol and storage in RNase-free water. Since plasma contains little or no 28S or 18S RNA, the integrity of the isolated RNA was determined by the small RNA peak between 0 and 150 nt (Supplemental Figure 5C). Concentrations of RNA specimens isolated from

brain, PBMC and plasma were determined by Nanodrop 2000 (Thermo Scientific, Wilmington, DE).

### **Determination of microRNA expression levels**

Primers specific for miR-34a and miR-196a were obtained from Applied Biosystems (Foster City, CA) to perform quantitative RT-PCR (qPCR), according to this vendor's protocol for TaqMan microRNA assays. Isolated total RNA fractions from brain, PBMC and plasma were initially processed for reverse transcription (RT) using two miRNA-specific primers (miR-34a: AB Assay IDs 000426; miR-196a: 241070\_mat) to obtain their RT products, used subsequently for qPCR analysis on a 7500 Fast System Real-Time PCR cyclor (Applied Biosystems). Small RNA 202 was used as control to calculate the expression levels of two microRNAs, miR-34a and -196a in isolated biological specimens. Numerical indices of these expression levels are expressed by the  $1/\Delta CT$  method, and obtain the values for individual microRNA after subtraction of the CT value for snoRNA202 (AB assay ID 001232, Applied Biosystems).

### **Isolation of Protein fraction from brain, PBMCs and plasma**

Sister blocks of brain coronal sections used for RNA isolation were used for protein extraction. In brief, brain specimens of ~100  $\mu$ g were solubilized in 300  $\mu$ l RIPA buffer before isolating the protein fraction, as described in our earlier report [16]. For the PBMC specimens, the organic phase after RNA isolation, as described above, was dialyzed on Spectra Pro 6 dialysis membranes (Spectrum Laboratories Inc., Rancho Dominguez, CA) in 0.3% SDS buffer at 4°C until the precipitate was completely dissolved, with further incubation in 0.1% SDS solution for an additional 24 hours. Plasma contains a huge amount of serum albumin and immunoglobulin (IgG); these were removed by the Vivapure anti-human albumin serum (HAS)/IgG kit (Sartorius Stedim Biotech GmbH, Göttingen, Germany). BCA protein assays from Pierce Biotechnology Inc (Rockford, IL) were used to determine the concentrations of the isolated protein samples from brain, PBMC and plasma.

### **Immunoblotting for protein levels of SIRT1, Bcl2 and p53**

SIRT1 and Bcl2, two main targets of miR-34a, and the activator of this microRNA, acetyl-p53, as well as the

total p53 protein pool were selected for determination of their protein abundance in brain, PBMC and plasma. Twenty-five micrograms of isolated protein samples from these three specimens were used for SDS-PAGE; afterwards, the electrophoretically separated bands were transferred from the gels to nitrocellulose membranes (Schleicher & Schuell BioScience, Keene, NH). Identification and quantification of the three proteins of interest were performed by incubating the membranes with antibodies including rabbit anti-SIRT1 (1:500, 75435, Abcam Inc., Cambridge, MA), rabbit anti-Bcl-2 (1:100, 2870, Cell Signaling, Danvers, MA), rabbit anti-acetyl-p53 (1:500, 06-758, Upstate (Millipore), Billerica, MA), rabbit anti-p53 (1:500, SC-6243, Santa Cruz Inc., CA) and rabbit anti- $\beta$ -actin (1:1000, 8226, Abcam Inc., Cambridge, MA) overnight at 4°C.  $\beta$ -actin was used as a loading control for PBMC and brain specimens, but not for plasma protein samples, since this protein's levels in plasma are not always stable. Instead, we selected a stable band with equal intensities across all lanes by Ponceau S staining [48]. Blots were developed following the method published previously [17], and levels of protein abundance were detected by the Enhanced Chemiluminescence (ECL) method, according to the manufacturer's instructions (Pierce Biotechnology, Rockford, IL). The intensities of bands on the ECL- developed films were quantified by densitometry using ImageQuant software (Molecular Dynamics Inc. Sunnyvale, CA).

### **Localization of miR-34a distribution by in situ hybridization in brain**

A microRNA 34a-specific locked nucleic acid (LNA) conjugated with digoxigenin (DIG), and a scrambled control, were purchased from Exiqon (Woburn, MA; miR-34a LNA probe: 38487-05; Scramble-miR LNA probe: 99004-01). These probes were incubated with the snap-frozen coronal brain sections, following our reported protocol [3]. Bound LNA probes were revealed by further incubating the sections with sheep anti-digoxigenin (DIG) antibody conjugated with alkaline phosphatase (AP) (1:2000; Roche, Indianapolis, IN) at 4°C overnight, followed by further detection of the DIG-AP substrate by nitrobluetetrazoliumchloride (NBT)/5-bromo-4-chloro-3-indolyl phosphate, toluidine salt (BCIP). Positive signal of the distribution of the LNA probes was detected by microscopic visualization, followed by image quantification with densitometry software (ImageQuant version 5.2, Molecular Dynamics).

Selected segments of images in the hippocampal and cortical regions were evaluated for miR-34a labeling intensity, compared to the scrambled control probe. Image quantification was performed using three animals per age group; the intensities of the LNA probes were obtained numerically from three repeats with three serial sections each. Thus, the final intensities for the two probes in each age group were obtained from three different animals, to control for inter-animal difference, and three repeats with three serial sections to control for inter-experiment variation. The scrambled control was used to normalize for background labeling noise.

### **Immunohistochemistry localization of SIRT1 distribution in brain**

Sister serial sections to those used above for in situ hybridization were processed to determine SIRT1 distribution by initially fixing them with 4% paraformaldehyde (PFA), followed by blocking with 10% goat serum (Invitrogen, Carlsbad, CA) and incubation with rabbit anti-SIRT1 at 1:200 dilution (75435, Abcam Inc., Cambridge, MA) at 4°C overnight. The bound antibody was revealed by further incubation with goat anti-rabbit (1:400; Invitrogen) conjugated with Alexa fluor 594 for 40 minutes at 37°C. Antibody labeling was evaluated on a Zeiss fluorescence microscope (Carl Zeiss, Brighton, MI) and AxioVision Rel.4.6 imaging system; image analysis of the distribution and intensities of SIRT1 protein in brain specimens of the two age groups follow the same procedure described above.

### **Data Analysis for statistical significance**

The  $1/\Delta CT$  values of qPCR results for the expression levels of miR-34a, miR-196, and controls among different age groups were analyzed by Student's t tests, with p values < 0.05 as statistically significant difference between any two groups presented in all the graphs for this assay. ANOVA was used to analyze data from all age groups included in all the assays. Mean  $\pm$  standard deviation (SD) was calculated for data obtained from quantitative evaluation for immunoblotting, in situ hybridization and immunofluorescence assays. Data analysis for these results proceeded as described for qPCR values. In all experiments, three different animals were used, to control for inter-animal variation; some of them performed three repeats, to control for inter-assay variance.

## **ACKNOWLEDGEMENTS**

We thank Mr. Alan N. Bloch for proofreading, and Mr. Vikranth Shetty for statistical analysis services, as well as Dr. Sang Lee for much of the technical work, such as isolation of the RNA and protein specimens. EW is supported by Kentucky's Research Challenge Trust Fund "Bucks-for-Brains" program.

## **REFERENCES**

1. Miller RA. The aging immune system: primer and prospectus. *Science*. 1996; 273: 70-74.
2. Esiri MM. Ageing and the brain. *J Pathol*. 2007; 211: 181-187.
3. Haigis MC, Yankner BA. The aging stress response. *Mol Cell*. 2010; 40: 333-344.
4. Carallo C, Irace C, De Franceschi MS, Coppoletta F, Tiriolo R, Scicchitano C, Scavelli F, Gnasso A. The effect of aging on blood and plasma viscosity. An 11.6 years follow-up study. *Clin Hemorheol Microcirc*. 2011; 47: 67-74.
5. Sprott RL. Biomarkers of aging. *Exp Gerontol*. 1988; 23: 1-3.
6. Miura Y, Kano M, Yamada M, Nishine T, Urano S, Suzuki S, Endo T, Toda T. Proteomic study on X-irradiation-responsive proteins and ageing: search for responsible proteins for radiation adaptive response. *J Biochem*. 2007; 142: 145-155.
7. Sonntag WE, Ramsey M, Carter CS. Growth hormone and insulin-like growth factor-1 (IGF-1) and their influence on cognitive aging. *Ageing Res Rev*. 2005; 4: 195-212.
8. Yang H, Lyutvinskiy Y, Soininen H, Zubarev RA. Alzheimer's Disease and Mild Cognitive Impairment are associated with elevated levels of isoaspartyl residues in blood plasma proteins. *J Alzheimers Dis*. 2011. doi: 10.3233/JAD-2011-110626.
9. Grolleau-Julius A, Ray D, Yung RL. The role of epigenetics in aging and autoimmunity. *Clin Rev Allergy Immunol*. 2010; 39: 42-50.
10. Liang R, Bates DJ, Wang E. Epigenetic control of MicroRNA expression and aging. *Curr Genomics*. 2009; 10: 184-193.
11. Kato M, Slack FJ. MicroRNAs: small molecules with big roles - C. elegans to human cancer. *Biol Cell*. 2008; 100: 71-81.
12. Boehm M, Slack F. A developmental timing microRNA and its target regulate life span in C. elegans. *Science*. 2005; 310: 1954-1957.
13. Maes OC, Chertkow HM, Wang E, Schipper HM. MicroRNA: Implications for Alzheimer Disease and other human CNS disorders. *Curr Genomics*. 2009; 10: 154-168.
14. Bates DJ, Li N, Liang R, Sarojini H, An J, Masternak MM, Bartke A, Wang E. MicroRNA regulation in Ames dwarf mouse liver may contribute to delayed aging. *Aging Cell*. 2010; 9: 1-18.
15. Li N, Bates DJ, An J, Terry DA, Wang E. Up-regulation of key microRNAs, and inverse down-regulation of their predicted oxidative phosphorylation target genes, during aging in mouse brain. *Neurobiol Aging*. 2011; 32: 944-955.
16. Li N, Muthusamy S, Liang R, Sarojini H, Wang E. Increased expression of miR-34a and miR-93 in rat liver during aging, and their impact on the expression of Mgst1 and Sirt1. *Mech Ageing Dev*. 2011; 132: 75-85.
17. Khanna A, Muthusamy S, Liang R, Sarojini H, Wang E. Gain of survival signaling by down-regulation of three key miRNAs in

brain of calorie-restricted mice. *Aging* (Albany NY). 2011; 3: 223-236.

18. Maes OC, Schipper HM, Chertkow HM, Wang E. Methodology for discovery of Alzheimer's disease blood-based biomarkers. *J Gerontol A Biol Sci Med Sci*. 2009; 64: 636-645.

19. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanian EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008; 105: 10513-10518.

20. Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, Xiao T, Schafer J, Lee ML, Schmittgen TD, Nana-Sinkam SP, Jarjoura D, Marsh CB. Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One*. 2008; 3: e3694.

21. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol*. 2008; 110: 13-21.

22. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabro E, Croce CM, Pastorino U, Sozzi C. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci U S A*. 2011; 108: 3713-3718.

23. Benner SA. Extracellular 'communicator RNA'. *FEBS Lett*. 1988; 233: 225-228.

24. Tsang J, Zhu J, van Oudenaarden A. MicroRNA-mediated feedback and feed forward loops are recurrent network motifs in mammals. *Mol Cell*. 2007; 26: 753-767.

25. Gaughwin PM, Ciesla M, Lahiri N, Tabrizi SJ, Brundin P, Bjorkqvist M. Hsa-miR-34b is a plasma-stable microRNA that is elevated in pre-manifest Huntington's disease. *Hum Mol Genet*. 2011; 20:2225-2237.

26. Saunders LR, Sharma AD, Tawney J, Nakagawa M, Okita K, Yamanaka S, Willenbring H, Verdin E. miRNAs regulate SIRT1 expression during mouse embryonic stem cell differentiation and in adult mouse tissues. *Aging* (Albany NY). 2010; 2:415-431.

27. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, Mari B, Barbry P, Mosnier JF, Hebuterne X, Harel-Bellan A, Mograbi B, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet*. 2011; 43: 242-245.

28. Yamakuchi M, Ferlito M, Lowenstein CJ. miR-34a repression of SIRT1 regulates apoptosis. *Proc Natl Acad Sci U S A*. 2008; 105: 13421-13426.

29. Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell*. 2007; 26: 745-52.

30. Simpson RJ, Jensen SS, Lim JW. Proteomic profiling of exosomes: current perspectives. *Proteomics*. 2008; 8: 4083-4099.

31. Guessous F, Zhang Y, Kofman A, Catania A, Li Y, Schiff D, Purow B, Abounader R. MicroRNA-34a is tumor suppressive in brain tumors and glioma stem cells. *Cell cycle*. 2010; 9: 1031-1036.

32. Hermeking H. MiR-34a and p53. *Cell Cycle*. 2009; 8: 1308.

33. Yamakuchi M, Lowenstein CJ. MiR-34, SIRT1 and p53: the feedback loop. *Cell Cycle*. 2009; 8: 712-715.

34. Burnett C, Valentini S, Cabreiro F, Goss M, Somogyvari M, Piper MD, Hoddinott M, Sutphin GL, Leko V, McElwee JJ,

Vazquez-Manrique RP, Orfila AM, Ackerman D, et al. Absence of effects of Sir2 overexpression on lifespan in *C. elegans* and *Drosophila*. *Nature*. 2011; 477: 482-485.

35. Herranz D, Serrano M. Impact of Sirt1 on mammalian aging. *Aging* (Albany NY). 2010; 2: 315-316.

36. Lee J, Kemper JK. Controlling SIRT1 expression by microRNAs in health and metabolic disease. *Aging* (Albany NY). 2010; 2: 527-534.

37. Ramadori G, Coppari R. Does hypothalamic SIRT1 regulate aging? *Aging* (Albany NY). 2011; 3: 325-328.

38. Poyurovsky MV, Prives C. P53 and aging: A fresh look at an old paradigm. *Aging* (Albany NY). 2010; 2: 380-382.

39. Biteau B, Jasper H. It's all about balance: p53 and aging. *Aging* (Albany NY). 2009; 1: 884-886.

40. Roemer K. Are the conspicuous interdependences of fecundity, longevity and cognitive abilities in humans caused in part by p53? *Cell Cycle*. 2010; 9: 3438-3441.

41. Ohyagi Y, Asahara H, Chui DH, Tsuruta Y, Sakae N, Miyoshi K, Yamada T, Kikuchi H, Taniwaki T, Murai H, Ikezoe K, Furuya H, Kawarabayashi T, et al. Intracellular Abeta42 activates p53 promoter: a pathway to neurodegeneration in Alzheimer's disease. *FASEB J*. 19: 255-257.

42. Schipper HM, Maes OC, Chertkow HM, Wang E. MicroRNA expression in Alzheimer blood mononuclear cells. *Gene Regul Syst Bio*. 2007; 1: 263-274.

43. Breitner JC, Baker LD, Montine TJ, Meinert CL, Lyketsos CG, Ashe KH, Brandt J, Craft S, Evans DE, Green RC, Ismail MS, Martin BK, Mullan MJ, et al. Extended results of the Alzheimer's disease anti-inflammatory prevention trial. *Alzheimers Dement*. 2011; 7: 402-411.

44. Griffin WS, Liu L, Li Y, Mrak RE, Barger SW. Interleukin-1 mediates Alzheimer and Lewy body pathologies. *J Neuroinflammation*. 2006; 3: 5.

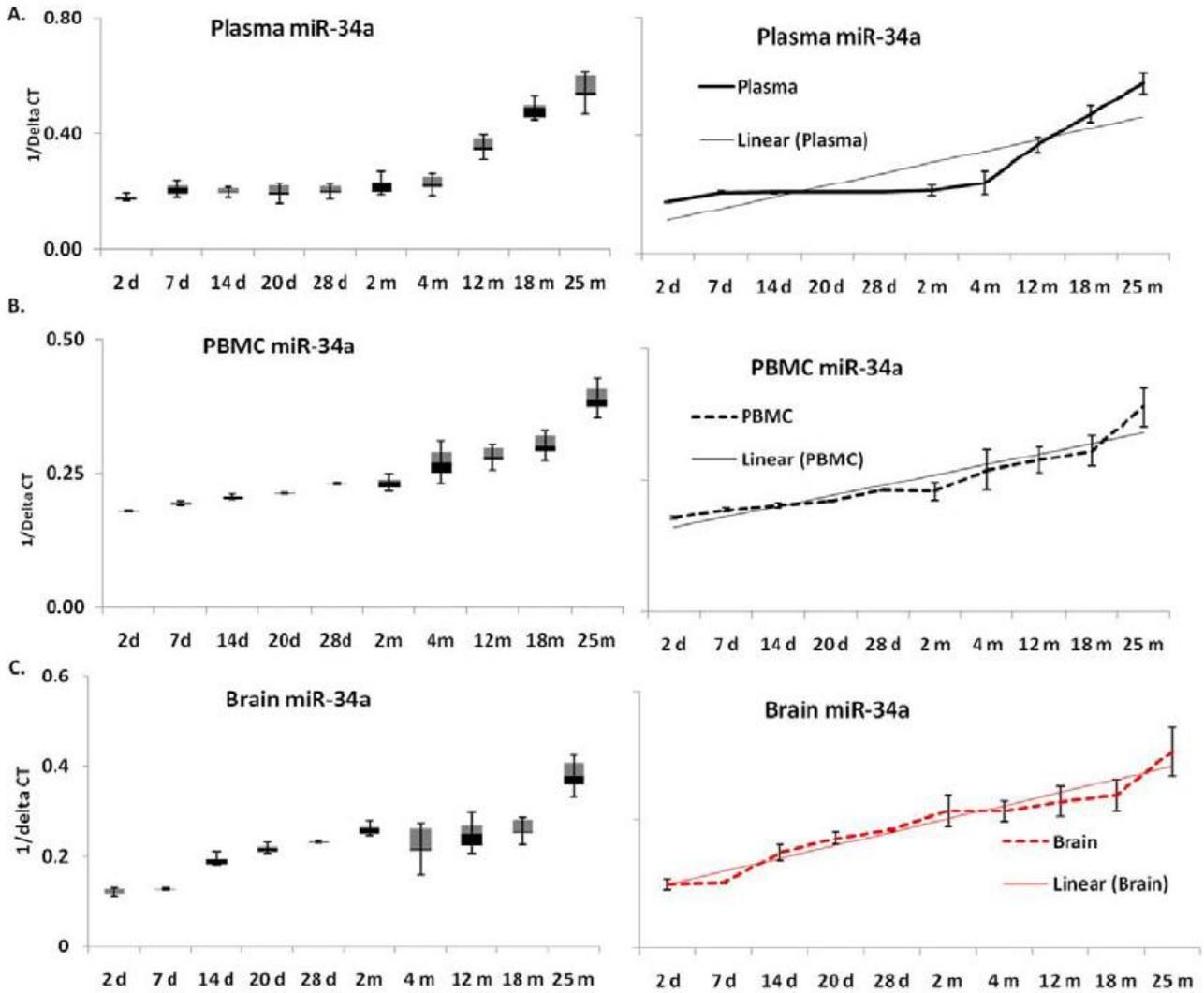
45. Bhaumik D, Scott GK, Schokrpur S, Patil CK, Orjalo AV, Rodier F, Lithgow GJ, Campisi J. MicroRNAs miR-146a/b negatively modulate the senescence-associated inflammatory mediators IL-6 and IL-8. *Aging* (Albany NY). 2009; 1: 402-411.

46. Delay C, Hébert SS. MicroRNAs and Alzheimer's disease mouse models: Current insights and future research avenues. *Int J Alzheimers Dis*. 2011; 2011: 894938.

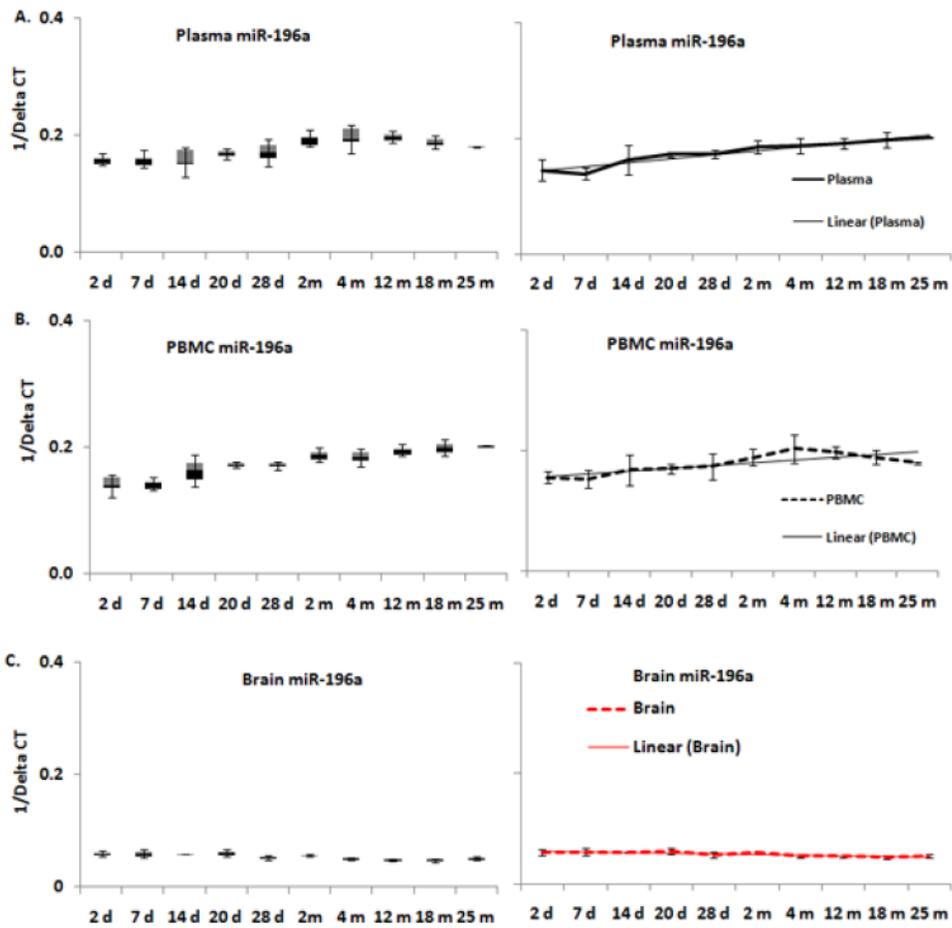
47. Provost P. Interpretation and applicability of microRNA data to the context of Alzheimer's and age-related diseases. *Aging* (Albany NY). 2010; 2: 166-169.

48. Pendyala G, Trauger SA, Siuzdak G, Fox HS. Quantitative plasma proteomic profiling identifies the vitamin E binding protein afamin as a potential pathogenic factor in SIV induced CNS disease. *J Proteome Res*. 2010; 9, 352-358.

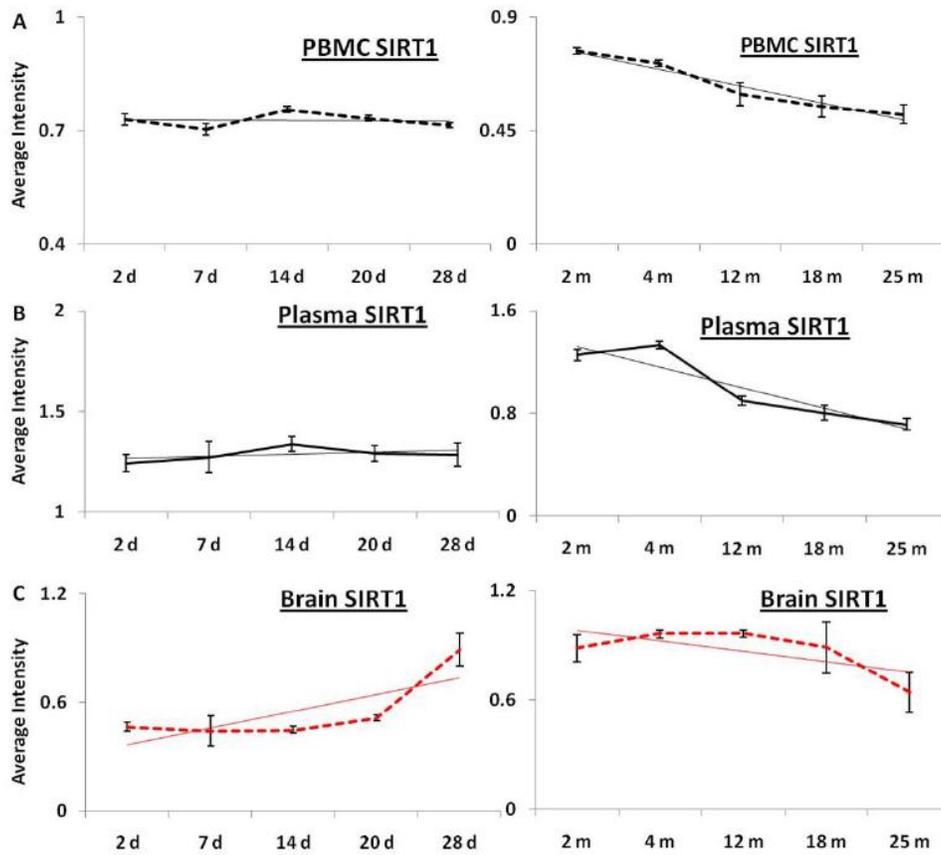
**SUPPLEMENTAL DATA**



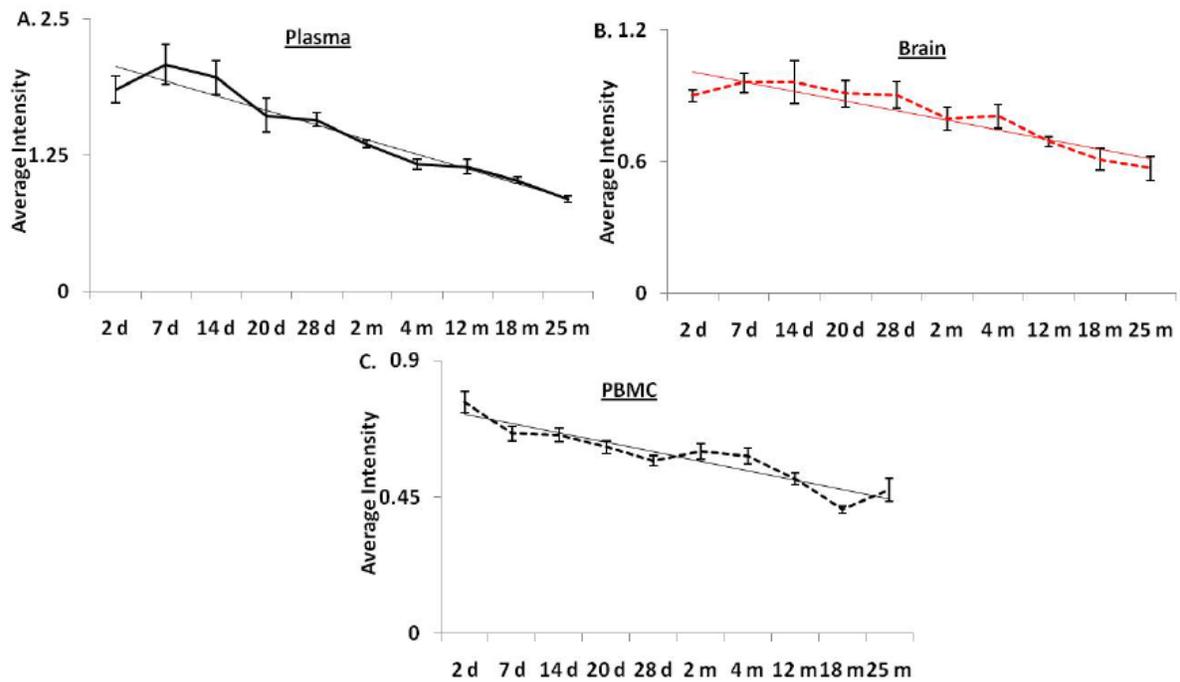
**S-Figure 1: Age-dependent levels of miRNA-34a expression in various tissues in C57/B6 mice** Panels (A-B-C) present box plot graphs of miR-34a expression in plasma, PBMCs and brain samples, along with graphs of the 1/delta CT trend. (n = 3; three different animals were used from each age group selected for the study.)



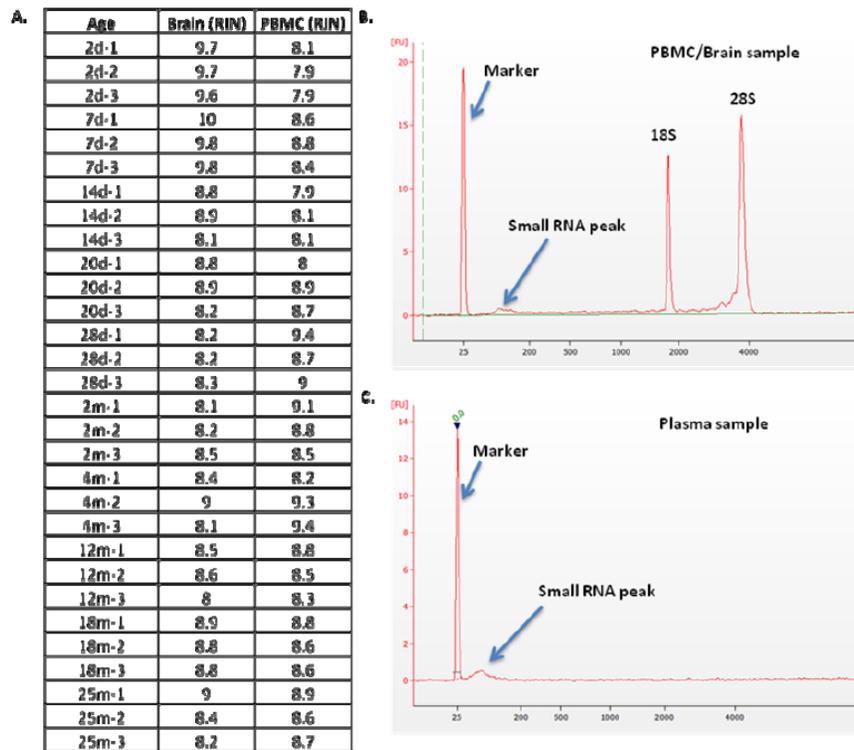
**S-Figure 2: Age-dependent expression levels of miRNA-196a in various tissues of C57/B6 mice** Panels (A-B-C) present box plot graphs of miR-196a expression in PBMCs, plasma and brain samples, along with graphs of the 1/delta CT trend. (n = 3; three different biological samples from each age group.)



**S-Figure 3: Graphical presentation of age-dependent expression levels of miR-34a target gene SIRT1 in plasma, PBMCs and brain samples of C57/B6 mice** Panels (A, B, C) show line graphs representing average intensity of SIRT1 expression, and trend lines depicting expression trends in various age groups, separated into different panels, with neonatal 2 to 28 days (d) in panel (a), and early adult from 2 months (m) to old age of 25 months in panel (b) in PBMCs, plasma and brain samples. (All graphs represent Mean ± SD; n = 3; three different biological samples from each age group.)



**S-Figure 4: Age-dependent expression levels of miR-34a target gene Bcl-2 in blood and tissue samples of C57/B6 mice** Panels (A, B, C) show average intensities of Bcl-2 expression, and trend lines depicting expression in various age groups, in plasma, brain and PBMCs samples. All graphs represent Mean  $\pm$  SD; n = 3; three different biological samples from each age group.



**S-Figure 5: RNA integrity analysis, using Agilent 2100 Bioanalyzer** Panels (A) Table shows RNA Integrity Number (RIN) for PBMCs and brain samples across the age groups considered; (B) shows the representative graphical representation of 28s, 18s, and small RNA bands observed during RNA integrity analysis for PBMCs and brain samples; (C) shows the representative graphical representation of small RNA bands as a small peak, and absence of 28s and 18s RNA peaks, as expected, in plasma samples during RNA integrity analysis.

miR-34a	PBMC	Plasma	Brain	PBMC st. dev	Plasma st. dev	Brain st. dev
2 d	0.17831	0.175008	0.124781	0.001964711	0.013204477	0.009879
7 d	0.193852	0.205179	0.128194	0.003947466	0.029856154	0.002109
14 d	0.20292	0.20589	0.186576	0.004920559	0.017730736	0.014666
20 d	0.210534	0.206923	0.214163	0.001528952	0.034964016	0.012307
28 d	0.230595	0.207074	0.231095	0.001104613	0.025660036	0.002859
2 m	0.229277	0.212698	0.267935	0.016294321	0.040063722	0.03081
4 m	0.269929	0.237729	0.266976	0.039299384	0.039417281	0.020578
12 m	0.28858	0.36711	0.2851	0.024423387	0.044062006	0.029361
18 m	0.305962	0.470866	0.298624	0.028231812	0.042045036	0.030545
25 m	0.3895	0.575449	0.383631	0.036379734	0.072269062	0.047753
miR-196a	PBMC	Plasma	Brain	PBMC st. dev	Plasma st. dev	Brain st. dev
2 d	0.155237	0.145426	0.058357	0.009934149	0.017700297	0.005374
7 d	0.15287	0.139138	0.058657	0.015062261	0.010709666	0.007212
14 d	0.167582	0.162178	0.058033	0.025016863	0.025719909	0.00064
20 d	0.169772	0.172685	0.060152	0.008497829	0.004363298	0.005829
28 d	0.173446	0.172879	0.052983	0.022425045	0.006633064	0.004827
2 m	0.188968	0.184865	0.056971	0.014384059	0.011332817	0.001917
4 m	0.202892	0.18698	0.050238	0.023577881	0.013130434	0.002125
12 m	0.197019	0.191407	0.050349	0.01035898	0.009434624	0.002197
18 m	0.188218	0.197933	0.049506	0.01177371	0.013458207	0.003197
25 m	0.179569	0.202233	0.050491	0.001811847	0.001184624	0.003156

**S-Table-1: Statistical analysis of miR-34a and miR-196a expression in blood and tissue samples of C57/B6 mice.**  
 Tabular presentation of mean standard deviation calculation for miRs-34a and 196a in PBMCs, plasma and brain samples. (n = 3, three different biological samples from each age group)

## Recent developments in the use of $\gamma$ -H2AX as a quantitative DNA double-strand break biomarker

Christophe E. Redon, Asako J. Nakamura, Olga A. Martin, Palak R. Parekh, Urbain S. Weyemi and William M. Bonner

Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA

**Key words:** DNA damage, DNA repair,  $\gamma$ -H2AX, double-strand break, biomarker, cancer, senescent cells

**Received:** 2/3/11; **Accepted:** 2/10/11; **Published:** 2/11/11

**Correspondence with** William Bonner, PhD to bonnerw@mail.nih.gov

© Redon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

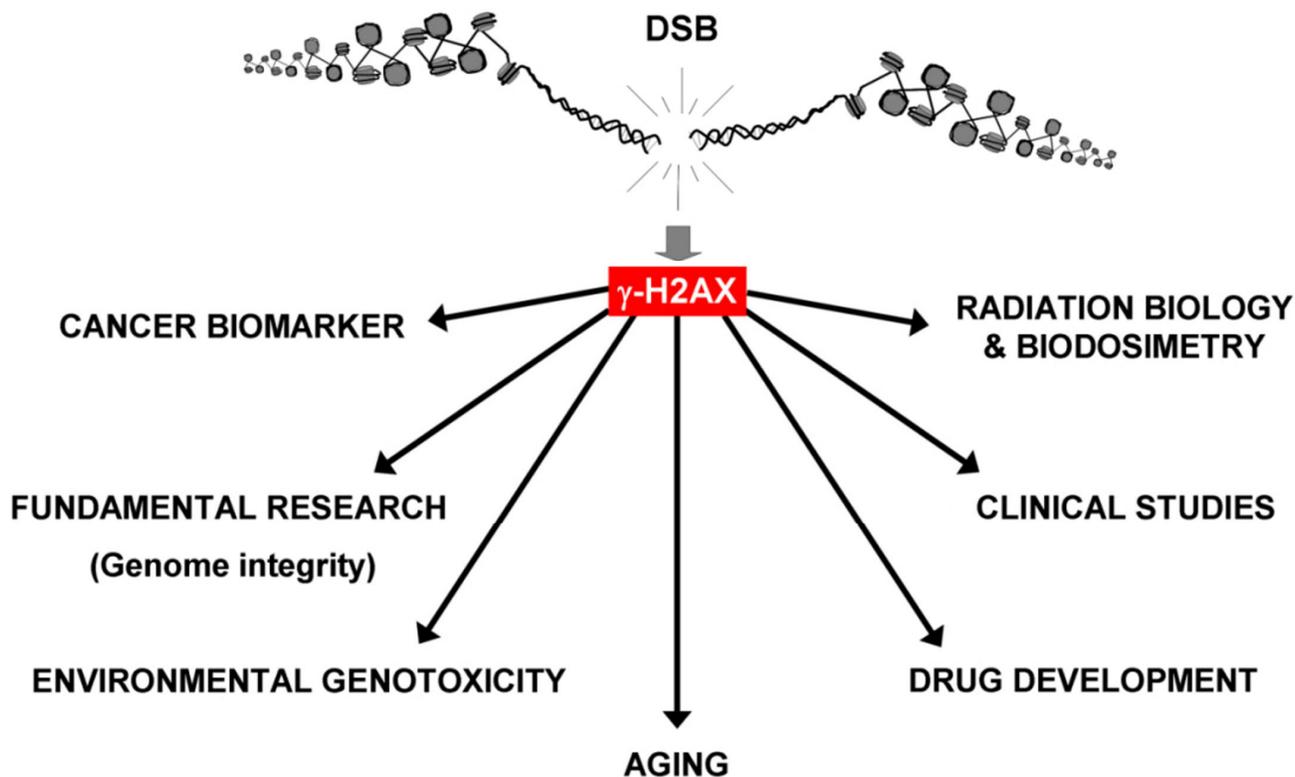
**Abstract:** The past year has seen considerable developments in the use of the DNA double-strand breaks (DSBs) to evaluate genome alterations in cells undergoing a variety of genotoxic stresses *in vitro* and *in vivo*. When the  $\gamma$ -H2AX foci which mark the DSBs are stained, individual breaks are detectable, making the assay suitable for situations requiring great sensitivity. While the methods for the detection of  $\gamma$ -H2AX foci are still evolving, particularly for *in vivo* detection, the basic assay has proven to be useful in several diverse areas of research. We will highlight recent developments of the assay in four areas: radiation biodosimetry, the evaluation or validation of new cancer drugs in clinical studies, chronic inflammation, and environmental genotoxicity.

### Background

The creation of a double-strand break (DSB) in eukaryotic cells is generally accompanied by the formation of hundreds of histone  $\gamma$ -H2AX (H2AX-S139PO4 in humans) molecules in the chromatin flanking the DSB site [1]. Antibodies to  $\gamma$ -H2AX allow the visualization of a “focus” at the DSB site. The foci also serve as sites for accumulation of other proteins involved in DSB repair, leading to the suggestion that the foci have roles in signal amplification and the accumulation of DNA repair factors that, in turn, facilitate chromatin remodeling, cell cycle checkpoint functioning, sister chromatid-dependent recombinational repair and chromatin anchoring to prevent the dissociation of broken ends [reviewed in [1-5]]. However, although  $\gamma$ -H2AX appears to be a principal player in the DNA damage response and necessary for the initial rapid phase of DSB repair, mice lacking H2AX, while hypersensitive to ionizing radiation, are still capable of DNA damage signaling and repair [6]. The viability of the H2AX-null mouse indicates that H2AX is not essential for homologous recombination or non-homologous end joining itself. However, these

mice suffer from two major deficits—a lack of class switch recombination during immune system development and a lack of sperm production in males. The former process is known to involve DSBs and genomic rearrangements. In the testes of H2AX-null males, autosome pairing and synapsis appear to take place normally, but the X and Y chromosomes fail to form a condensed sex body [1,6].

Potential practical uses of  $\gamma$ -H2AX foci formation in cells and tissues have been apparent from soon after its discovery. The amplified response makes it possible to easily visualize individual DSBs in cell nuclei, making  $\gamma$ -H2AX foci staining more sensitive than other methods of DSB detection [7]. The realization that DSBs, whether alone or as one type of a spectrum of DNA lesions, are involved in many processes that disturb cellular homeostasis has led to broadening use of  $\gamma$ -H2AX foci detection beyond basic research. It has been used as a biomarker for aging and cancer, and a biodosimeter for drug development, radiation exposure and for clinical trials for cancer chemo- and radiotherapy. Finally, other emerging uses for  $\gamma$ -H2AX include detection of toxic environmental agents and the



**Figure 1. Applications for  $\gamma$ -H2AX detection.** Because of its sensitivity, the  $\gamma$ -H2AX assay is now utilized in many research areas “from benchtop to bedside” by researchers and clinicians. In addition to being widely used for fundamental research (study of genome stability, DNA repair, etc.) in the last decade,  $\gamma$ -H2AX was identified as a biomarker for cancer (and premalignant lesions) and used to better understand aging. Additionally,  $\gamma$ -H2AX has been developed for radiation biology and biodosimetry for drug development and clinical studies (chemotherapy, the impact of chronic inflammation and diabetes on genome integrity). Finally,  $\gamma$ -H2AX measurement is an efficient and sensitive genotoxic assay for environmental studies.

detection of chronic inflammation (Figure 1). In this report we highlight recent advances in four areas utilizing  $\gamma$ -H2AX detection.

**Radiation biodosimetry.** Since ionizing radiation induces DSBs among a broad spectrum of DNA lesions, assessment of the biologic response to radiation exposure is a straightforward application in the use of  $\gamma$ -H2AX as a biodosimeter. Detection of radiation-induced DSBs *in vivo* using the  $\gamma$ -H2AX assay has been utilized as a tool for dose estimation in the clinic for localized irradiation with both high (radiotherapy) and low doses (X-ray examination, computed tomography (CT) scan, etc) [see [7-9] as examples].  $\gamma$ -H2AX may be used to improve the conditions for patients undergoing radiation treatments, for example as a tool to estimate individual radiosensitivity, scattering or abscopal effects in normal tissues (discussed below). Kuefner *et al.* (2010) followed  $\gamma$ -H2AX foci formation during different cardiac CT protocols and showed that it is important to carefully adapt these protocols to avoid unnecessary

X-ray-induced DSBs [7,10]. While the  $\gamma$ -H2AX assay gave results which supported the exposures estimated by physical modeling, it takes into account various biological factors not available in physical modeling, giving more confidence to these measurements.

However, while studies reported the use of  $\gamma$ -H2AX foci induction following exposure to therapeutic doses of ionizing radiation [9,11-15], how the assay would perform at higher doses, particularly in humans, remained unclear. Recently, the opportunity arose to evaluate  $\gamma$ -H2AX biodosimetry in a study using non-human primates subjected to total-body irradiation in the non-lethal to lethal dose ranges [16]. Using realistic scenarios for accidental exposures, the authors showed that  $\gamma$ -H2AX analysis in lymphocytes and plucked hair follicles (eyebrows and whiskers) may be useful for estimation of radiation dose at times at least 4-days post-exposure at doses of 3.5 Gy and above. In addition, the Center for High-Throughput Minimally Invasive Radiation Biodosimetry has developed a fully auto-

mated high-throughput system, the RABIT (Rapid Automated Biodosimetry Tool) to respond to major radiological accidents. The RABIT is able to perform the  $\gamma$ -H2AX analysis for radiation biodosimetry of up to 30,000 blood samples a day and is intended to fully automate the  $\gamma$ -H2AX assay, from the isolation of human blood lymphocytes to the immunolabeling of  $\gamma$ -H2AX and image acquisition [17,18].

**Drug biodosimeter.** While ionizing radiation and a few cancer drugs create DSBs directly in the DNA, many drugs induce DSB formation indirectly through interference with DNA replication and transcription [1,19,20]. A well characterized example involves the drug camptothecin that traps topoisomerase I (top1) in DNA complexes which in replicating cells often result in DSB formation as replication forks collide with the trapped top1 complexes [1,21].

While chemotherapeutic agents are given with a particular type of tumor in mind, every cell in the body may be exposed to the agent. Measuring the amount of DNA damage in a patient's tissues soon after drug administration would allow researchers and clinicians to determine the efficacy of a drug to create DNA damage and genomic instability in the cells of a particular patient. Such information could help "personalize" the doses and delivery of a single drug or combinations of drugs to individual patients in terms of DNA damage efficiencies, and further correlate of these results with data on tumor regression and ultimate patient survival. This may result in optimized protocols that improve patient outcomes.

The practicality of  $\gamma$ -H2AX as a reproducible pharmacodynamic marker of top 1 inhibitor activity has been evaluated with an assay developed and validated in two laboratories [22]. Using three structurally related indenoisoquinoline Top1 inhibitors in human xenograft mouse models, the assay gave significant responses in tumor biopsies and in skin snips at the mouse equivalents of clinically relevant doses. One advantage of this assay is that results on drug activities were obtained four hours after administration, more quickly than waiting for visible tumor responses. On the basis of this assay, two of the three compounds were selected for further clinical evaluation.

However, the specificity of many cancer drugs for replicating cells creates a problem concerning appropriate tissues to sample for measuring  $\gamma$ -H2AX foci formation. Sampling the tumor may be the most direct means to measure the efficiency of a cancer therapy, but collecting tumor biopsies is often risky and invasive for

the patient. However, tumor cells are shed by many tumors into the peripheral blood of the patients. Monitoring  $\gamma$ -H2AX levels in a patient's circulating tumor cells (CTCs) following cancer treatment has been evaluated and may be a promising technique for following the pharmacodynamic effects of anticancer therapies [23].

The lymphocytes present in peripheral blood have been utilized to detect  $\gamma$ -H2AX formation during cancer treatment, but these terminally differentiated cells may respond poorly to anticancer drugs that interfere with DNA metabolism. However, recent work has shown that plucked hair bulbs, which contain replicating cells and can be obtained non-invasively, may be utilized to monitor DSB formation *in vivo* after drug administration [24,25]. The indenoquinoline study showed that hair follicles in skin snips, used instead of plucked hair bulbs in athymic nude mice, responded similarly to the tumor, suggesting that plucked hair bulbs may be an appropriate surrogate tissue [22]. However, while both the tumor and hair bulb contain replicating cells, those of the tumor may have genetic alterations that make their responses to a particular agent different than that of normal replicating cells, both in terms of  $\gamma$ -H2AX foci formation and cell survival. This is a question requiring further research—how does the response of a surrogate tissue correlate with the response of the tumor and the patient and can it be predictive of that response.

**Distant DNA damage and chronic inflammation.** The high sensitivity of the  $\gamma$ -H2AX foci assay has enabled researchers to measure low levels of DNA damage. Intercellular communication among cells in a culture or organism, where some of the cells have been damaged, has been found to result in the presence of low levels of DNA damage in cells peripheral to the damaged cells. For example, the radiation-induced bystander effect refers to the situation where a larger fraction of the cell population dies compared to the fraction hit by ionizing particles [26]. The effect can be demonstrated when a few cells in a culture are targeted with alpha particles, when an irradiated culture is mixed with an unirradiated one, and when the media conditioned on an irradiated culture is transferred to an unirradiated culture [27]. This last method indicates that substances released into the media from the irradiated culture are inducing the effects in the recipient normal, bystander culture. Bystander cells have been shown to exhibit greater numbers of chromosomal aberrations, micronuclei and  $\gamma$ -H2AX foci as well as increased mortality [28]. While the incidence of these defects is often just a few-fold elevated over the control values, the increase can be measured with  $\gamma$ -H2AX foci.

Cells, not exposed to genotoxic agents, but which are aberrant in some way may also induce a bystander effect in normal cells.  $\gamma$ -H2AX levels are elevated in cancer and aging cells in which it marks both DSBs and abnormal telomeres [29]. Media conditioned on cultures of both aging and tumor cells were found to induce elevated levels of  $\gamma$ -H2AX foci in normal cells. These results suggest that cells may constantly be releasing factors into their surroundings which affect other cells, at least in culture [30].

Recently, a tumor-induced bystander effect has been demonstrated in mice [31]. Syngeneic tumors were implanted subcutaneously in normal mice, and after two weeks, various organs of the mice were analyzed for DNA DSBs and another type of serious DNA lesions, oxidative clustered DNA lesions (OCDLs). Elevated levels of  $\gamma$ -H2AX foci were found in tissues of the gastrointestinal tract and skin including hair follicles; elevated levels of OCDLs were more widely distributed, present in lung and ovary as well as the gastrointestinal tissues and skin. A possible explanation for these results is that DSB formation is more common in proliferating cells where replication forks may encounter single strand breaks with a resulting DSB, while OCDL formation is less dependent on DNA metabolism. Gastrointestinal tract tissues and hair follicles contain higher fractions of proliferating cells compared to other tissues like ovary and lung.

In contrast to cell culture, this distant bystander effect in mice was found to be dependent on the immune system. Serum taken from tumor-bearing mice exhibited elevated levels of several chemokines involved in macrophage activation, and F4/80+ macrophages were found in the affected distant tissues of tumor-bearing mice compared to their control cohorts. Furthermore, when this experiment was repeated in mice lacking the gene for one of these chemokines, monocyte chemoattractant protein 1 (MCP-1/CCL2), levels of DNA damage in distant tissues were not elevated. These results strengthen a connection between chronic inflammation and the presence of elevated levels of DNA damage induced by the production/release in tissues of oxidative molecules (i.e. by activated macrophages).

While these results were obtained in a mouse model system, similar processes appear to be occurring in humans. There is a relationship between chronic inflammation and obesity [32], characterized by increased secretion of proinflammatory cytokines such as interleukin-(IL)-6, IL-8 and MCP-1/CCL2, abundant infiltration of activated macrophages into adipose tissue and obesity-induced insulin resistance [33]. Furthermore, adipose tissue biopsies of obese individuals

revealed that both subcutaneous adipose tissue and visceral adipose tissue exhibited  $\gamma$ -H2AX foci but that the visceral tissue exhibited 3-fold more foci [34]. Finally, in one study with obese children, peripheral blood lymphocytes were found to exhibit a greater than 5-fold increase in foci incidence in overweight children and 8-fold in obese children [35]. The normal cohort exhibited  $0.0034 \pm 0.0006$  foci per cell (fpc), overweight  $0.019 \pm 0.0039$  and obese  $0.0274 \pm 0.0029$  fpc. Micro-nuclear frequencies were also elevated in stimulated lymphocytes from the overweight and obese children compared to normal. These values for  $\gamma$ -H2AX foci incidence correspond to one per 294 cells in control individuals versus one per 37 in obese individuals. As with the mouse study, the elevated numbers of foci found were small, equivalent to approximately the number of foci observed 30 minutes after exposure to 2 mGy of ionizing radiation. However, the differences between the obese subjects and normal controls were substantially significant, attesting to the potential value of  $\gamma$ -H2AX foci measurements at low DSB levels.

In addition, there is a documented correlation between obesity and overall cancer incidence, as well as the incidence of other health issues. If increased levels of foci and DSB damage in lymphocytes and adipose tissue is indicative of increased damage levels elsewhere, these findings provide a possible molecular mechanism by which obesity may lead to increased cancer risk.

**Environmental genotoxicity and high-throughput assays.** DSB detection by  $\gamma$ -H2AX foci has been used in studies of the effects of personal exposure to environmental agents such as air pollution, food toxins and industrial chemicals, all of which have the potential of large scale impacts on human populations as well as populations of other organisms. With the  $\gamma$ -H2AX assay researchers are able to monitor small amounts of genotoxicity *in vivo*, enabling them to evaluate the effects of environmental conditions on a population. As one practical example, a study with women in rural India exposed to high amounts of smoke from cooking with biomass fuel in poorly ventilated kitchens revealed a 4-fold elevation in the fraction of airway epithelial cells exhibiting  $\gamma$ -H2AX foci and 5-fold in peripheral blood lymphocytes compared to the ones using cleaner fuel liquefied petroleum gas [36]. These authors also found the elevation of other DNA damage markers. Such studies could provide evidence for governmental action to limit human exposure to certain environmental agents.

With the demonstrated utility of  $\gamma$ -H2AX foci measurements *in vivo*, new tools for high throughput

$\gamma$ -H2AX screenings are being developed. Audebert *et al.* reported the development of an *in vitro* genotoxicity assay system using  $\gamma$ -H2AX measurement to evaluate the effects of food poisons such as benzo(a)pyrene, fluoranthene, and 3-methylcholanthrene [37]. In addition to the automated high throughput RABIT system discussed above [17,18], one high-content screening assay [38] and two enzyme-linked immunosorbent assays (ELISA) that can accurately quantify  $\gamma$ -H2AX level, one for screening genotoxic molecules [39] and one for clinical trials [25] have been developed. These tools will greatly increase the capacity to identify genotoxic compounds, not only to humans, but also in other organisms. Since the H2AX phosphorylation site is highly conserved throughout eukaryotes, the  $\gamma$ -H2AX assay may find greatly expanded utility.

Finally, however, it should be kept in mind that while a  $\gamma$ -H2AX focus may indicate the presence of a DNA DSB in the vast majority of cases there are situations where one may be present without the other. The presence of  $\gamma$ -H2AX foci in the absence of DNA damage has been demonstrated by tethering single repair factors to chromatin [40]. Also, in normal senescent cell cultures,  $\gamma$ -H2AX may also be an indicator of mTOR-dependent senescent phenotype independent of DNA damage and of a classic DNA damage response [41]. In contrast, the presence of DSBs without  $\gamma$ -H2AX foci formation has been demonstrated in mouse kidney cells when bathed in high salt [42]. Break repair is hindered while the osmolality is high, but when it is returned to normal,  $\gamma$ -H2AX foci form and the breaks are repaired. Understanding these exceptional situations may help clarify the overall relationship between  $\gamma$ -H2AX foci and DSBs.

## CONCLUSIONS

The studies presented here demonstrate the wide range of genotoxic events *in vitro* and *in vivo* that are amenable to DNA damage measurements utilizing the  $\gamma$ -H2AX assay. In some cases there is substantial induction of  $\gamma$ -H2AX foci with exposure to high levels of ionizing radiation and cancer drugs, while in others the induction is of smaller magnitude, such as with diagnostic CT scans, environmental toxicity and chronic inflammation.  $\gamma$ -H2AX studies indicate a connection between DNA damage and chronic inflammation, but is the DNA damage observed during chronic inflammation just another indicator of a biological system that is not at optimum homeostasis, or does it have a causal roles on the long-term deleterious effects of chronic

inflammation? Also, while the effects of large-scale DNA damage as indicated by strong  $\gamma$ -H2AX signals are readily apparent, it is not yet clear what a several-fold elevation of DNA damage markers including  $\gamma$ -H2AX foci mean to the organism either in the short-term or long-term. In the case of drug biodosimetry, the relationships of  $\gamma$ -H2AX responses of surrogate tissues relative to the  $\gamma$ -H2AX responses of tumors as well as the crucial biological responses of tumors, regression and disappearance, are beginning to be studied. These are all important questions that remain to be answered.

$\gamma$ -H2AX studies are beginning to give useful input for decision making. Two examples discussed here are the comparison of various CT scanning protocols [10] and the choice of indenoisoquinolines for clinical evaluation [22]. While such studies to date have concentrated on human subjects, the almost universal conservation of DNA damage induced H2AX phosphorylation throughout eukaryotic evolution including plants suggests that future research may include many more subjects.

## ACKNOWLEDGEMENTS

This work was supported by the Intramural Research Program of the National Cancer Institute, NIH.

## REFERENCES

1. Bonner WM, Redon CE, Dickey JS, Nakamura AJ, Sedelnikova OA, Solier S, Pommier Y. gammaH2AX and cancer. *Nat Rev Cancer*. 2008; 8:957-967.
2. van Attikum H and Gasser SM. ATP-dependent chromatin remodeling and DNA double-strand break repair. *Cell Cycle*. 2005; 4:1011-1014.
3. Thiriet C and Hayes JJ. Chromatin in need of a fix: phosphorylation of H2AX connects chromatin to DNA repair. *Mol Cell*. 2005; 18:617-622.
4. Rothkamm K and Horn S. gamma-H2AX as protein biomarker for radiation exposure. *Ann Ist Super Sanita*. 2009; 45:265-271.
5. Yuan J, Adamski R and Chen J. Focus on histone variant H2AX: to be or not to be. *FEBS Lett*. 2010; 584:3717-3724.
6. Celeste A, Petersen S, Romanienko PJ, Fernandez-Capetillo O, Chen HT, Sedelnikova OA, Reina-San-Martin B, Coppola V, Meffre E, Difilippantonio MJ, Redon C, Pilch DR, Olaru A, Eckhaus M, Camerini-Otero RD, Tessarollo L, Livak F, Manova K, Bonner WM, Nussenzweig MC, Nussenzweig A. Genomic instability in mice lacking histone H2AX. *Science*. 2002; 296:922-927.
7. Rothkamm K and Lobrich M. Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses. *Proc Natl Acad Sci U S A*. 2003; 100:5057-5062.

8. Qvarnstrom OF, Simonsson M, Johansson KA, Nyman J, Turesson I. DNA double strand break quantification in skin biopsies. *Radiother Oncol.* 2004; 72:311-317.
9. Loblrich M, Rief N, Kuhne M, Heckmann M, Fleckenstein J, Rube C, Uder M. In vivo formation and repair of DNA double-strand breaks after computed tomography examinations. *Proc Natl Acad Sci U S A.* 2005; 102:8984-8989.
10. Kuefner MA, Grudzinski S, Hamann J, Achenbach S, Lell M, Anders K, Schwab SA, Haberle L, Loblrich M, Uder M. Effect of CT scan protocols on x-ray-induced DNA double-strand breaks in blood lymphocytes of patients undergoing coronary CT angiography. *Eur Radiol.* 2010; 20:2917-2924.
11. Kuefner MA, Grudzinski S, Schwab SA, Wiederseiner M, Heckmann M, Bautz W, Loblrich M, Uder M. DNA double-strand breaks and their repair in blood lymphocytes of patients undergoing angiographic procedures. *Invest Radiol.* 2009; 44:440-446.
12. Sak A, Grehl S, Erichsen P, Engelhard M, Grannass A, Levegrun S, Pottgen C, Groneberg M, Stuschke M. gamma-H2AX foci formation in peripheral blood lymphocytes of tumor patients after local radiotherapy to different sites of the body: dependence on the dose-distribution, irradiated site and time from start of treatment. *Int J Radiat Biol.* 2007; 83:639-652.
13. Beels L, Bacher K, De Wolf D, Werbrouck J, Thierens H. gamma-H2AX foci as a biomarker for patient X-ray exposure in pediatric cardiac catheterization: are we underestimating radiation risks? *Circulation.* 2009; 120:1903-1909.
14. Rothkamm K, Balroop S, Shekhdar J, Fernie P, Goh V. Leukocyte DNA damage after multi-detector row CT: a quantitative biomarker of low-level radiation exposure. *Radiology.* 2007; 242:244-251.
15. Geisel D, Heverhagen JT, Kalinowski M, Wagner HJ. DNA double-strand breaks after percutaneous transluminal angioplasty. *Radiology.* 2008; 248:852-859.
16. Redon CE, Nakamura AJ, Gouliava K, Rahman A, Blakely WF, Bonner WM. The use of gamma-H2AX as a biodosimeter for total-body radiation exposure in non-human primates. *PLoS One.* 2010; 5:e15544.
17. Garty G, Chen Y, Salerno A, Turner H, Zhang J, Lyulko O, Bertucci A, Xu Y, Wang H, Simaan N, Randers-Pehrson G, Yao YL, Amundson SA, Brenner DJ. The RABIT: a rapid automated biodosimetry tool for radiological triage. *Health Phys.* 2010; 98:209-217.
18. Turner HC, Brenner DJ, Chen Y, Bertucci A, Zhang J, Wang H, Lyulko OV, Xu Y, Shuryak I, Schaefer J, Simaan N, Randers-Pehrson G, Yao YL, Amundson SA, Garty G. Adapting the gamma-H2AX Assay for Automated Processing in Human Lymphocytes. 1. Technological Aspects. *Radiat Res.* 2010; Dec 28, epub ahead of print
19. Sordet O, Redon CE, Guirouilh-Barbat J, Smith S, Solier S, Douarre C, Conti C, Nakamura AJ, Das BB, Nicolas E, Kohn KW, Bonner WM, Pommier Y. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Rep.* 2009; 10:887-893.
20. Sedelnikova OA, Redon CE, Dickey JS, Nakamura AJ, Georgakilas AG, Bonner WM. Role of oxidatively induced DNA lesions in human pathogenesis. *Mutat Res.* 2010.
21. Pommier Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer.* 2006; 6:789-802.
22. Kinders RJ, Hollingshead M, Lawrence S, Ji J, Tabb B, Bonner WM, Pommier Y, Rubinstein L, Evrard YA, Parchment RE, Tomaszewski J, Doroshow JH. Development of a validated immunofluorescence assay for gammaH2AX as a pharmacodynamic marker of topoisomerase I inhibitor activity. *Clin Cancer Res.* 2010; 16:5447-5457.
23. Wang LH, Pfister TD, Parchment RE, Kummar S, Rubinstein L, Evrard YA, Gutierrez ME, Murgo AJ, Tomaszewski JE, Doroshow JH, Kinders RJ. Monitoring drug-induced gammaH2AX as a pharmacodynamic biomarker in individual circulating tumor cells. *Clin Cancer Res.* 2010; 16:1073-1084.
24. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ, Ashworth A, Carmichael J, Kaye SB, Schellens JH, de Bono JS. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med.* 2009; 361:123-134.
25. Redon CE, Nakamura AJ, Zhang YW, Ji JJ, Bonner WM, Kinders RJ, Parchment RE, Doroshow JH, Pommier Y. Histone gammaH2AX and poly(ADP-ribose) as clinical pharmacodynamic biomarkers. *Clin Cancer Res.* 2010; 16:4532-4542.
26. Nagasawa H and Little JB. Induction of sister chromatid exchanges by extremely low doses of alpha-particles. *Cancer Res.* 1992; 52:6394-6396.
27. Sokolov MV, Smilenov LB, Hall EJ, Panyutin IG, Bonner WM, Sedelnikova OA. Ionizing radiation induces DNA double-strand breaks in bystander primary human fibroblasts. *Oncogene.* 2005; 24:7257-7265.
28. Prise KM and O'Sullivan JM. Radiation-induced bystander signalling in cancer therapy. *Nat Rev Cancer.* 2009; 9:351-360.
29. Nakamura AJ, Redon CE, Bonner WM, Sedelnikova OA. Telomere-dependent and telomere-independent origins of endogenous DNA damage in tumor cells. *Aging.* 2009; 1:212-218.
30. Dickey JS, Baird BJ, Redon CE, Sokolov MV, Sedelnikova OA, Bonner WM. Intercellular Communication of Cellular Stress Monitored by {gamma}-H2AX Induction. *Carcinogenesis.* 2009; 30:1686-1695.
31. Redon CE, Dickey JS, Nakamura AJ, Kareva IG, Naf D, Newshean S, Kryston TB, Bonner WM, Georgakilas AG, Sedelnikova OA. Tumors induce complex DNA damage in distant proliferative tissues in vivo. *Proc Natl Acad Sci U S A.* 2010; 107:17992-17997.
32. Gustafson B. Adipose tissue, inflammation and atherosclerosis. *J Atheroscler Thromb.* 2010; 17:332-341.
33. Bassaganya-Riera J, Misyak S, Guri AJ, Hontecillas R. PPAR gamma is highly expressed in F4/80(hi) adipose tissue macrophages and dampens adipose-tissue inflammation. *Cell Immunol.* 2009; 258:138-146.

34. Villaret A, Galitzky J, Decaunes P, Esteve D, Marques MA, Sengenès C, Chiotasso P, Tchkonina T, Lafontan M, Kirkland JL, Bouloumie A. Adipose tissue endothelial cells from obese human subjects: differences among depots in angiogenic, metabolic, and inflammatory gene expression and cellular senescence. *Diabetes*. 2010; 59:2755-2763.
35. Scarpato R, Verola C, Fabiani B, Bianchi V, Saggese G, Federico G. Nuclear damage in peripheral lymphocytes of obese and overweight Italian children as evaluated by the  $\gamma$ -H2AX focus assay and micronucleus test. *FASEB J*. 2011; 25:685-693.
36. Mondal NK, Mukherjee B, Das D, Ray MR. Micronucleus formation, DNA damage and repair in premenopausal women chronically exposed to high level of indoor air pollution from biomass fuel use in rural India. *Mutat Res*. 2010; 697:47-54.
37. Audebert M, Riu A, Jacques C, Hillenweck A, Jamin EL, Zalko D, Cravedi JP. Use of the  $\gamma$ H2AX assay for assessing the genotoxicity of polycyclic aromatic hydrocarbons in human cell lines. *Toxicol Lett*. 2010; 199: 182-192.
38. Yip KW, Cuddy M, Pinilla C, Giulianotti M, Heynen-Genel S, Matsuzawa SI, Reed JC. A High-Content Screening (HCS) Assay for the Identification of Chemical Inducers of PML Oncogenic Domains (PODs). *J Biomol Screen*. 2011; 16:251-258.
39. Matsuzaki K, Harada A, Takeiri A, Tanaka K, Mishima M. Whole cell-ELISA to measure the  $\gamma$ H2AX response of six aneugens and eight DNA-damaging chemicals. *Mutat Res*. 2010; 700:71-79.
40. Soutoglou E and Misteli T. Activation of the cellular DNA damage response in the absence of DNA lesions. *Science*. 2008; 320:1507-1510.
41. Pospelova TV, Demidenko ZN, Bukreeva EI, Pospelov VA, Gudkov AV, Blagosklonny MV. Pseudo-DNA damage response in senescent cells. *Cell Cycle*. 2009; 8:4112-4118.
42. Dmitrieva NI and Burg MB. Analysis of DNA breaks, DNA damage response, and apoptosis produced by high NaCl. *Am J Physiol Renal Physiol*. 2008; 295:F1678-1688.

# Notch1/2/3/4 are prognostic biomarker and correlated with immune infiltrates in gastric cancer

Jian Hu<sup>1,\*</sup>, Jiangong Yu<sup>2,\*</sup>, Jun Gan<sup>1,\*</sup>, Ning Song<sup>3</sup>, Liubin Shi<sup>1</sup>, Jie Liu<sup>2</sup>, Ziqiang Zhang<sup>3</sup>, Jianjun Du<sup>1</sup>

<sup>1</sup>Department of General Surgery, Huashan Hospital, Fudan University, Shanghai 200040, China

<sup>2</sup>Department of Digestive Diseases of Huashan Hospital and Institute of Biomedical Sciences, Fudan University, Shanghai 200040, China

<sup>3</sup>Department of General Surgery, Huashan Hospital North, Fudan University, Shanghai 200040, China

\*Equal contribution

**Correspondence to:** Jianjun Du, Ziqiang Zhang, Jie Liu; **email:** [dujjp@hotmail.com](mailto:dujjp@hotmail.com), [zzqblue@hotmail.com](mailto:zzqblue@hotmail.com), [jieliu@fudan.edu.cn](mailto:jieliu@fudan.edu.cn)

**Keywords:** Notch, gastric cancer, prognosis, immune infiltrates

**Received:** September 11, 2019 **Accepted:** January 12, 2020

**Published:** February 6, 2020

**Copyright:** Hu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Notch refers to a set of genes encoding a family of transmembrane receptors in mammalian cells. Previous evidence indicated that Notch receptors were implicated in the onset of gastric cancer. However, there is little evidence on the different genetic expression patterns of the four Notch receptors and their values for patient prognosis. Most recently, we investigated the transcriptional data of Notch receptors and related patient survival in patients with GC based on several databases, including ONCOMINE, GEPIA, Kaplan–Meier Plotter, cBioPortal and TIMER. According to our findings, gastric cancer tissues, compared with adjacent normal tissues presented a higher level of expression of Notch1/2/3. We also performed a survival analysis and noted that gastric cancer patients with high transcription levels of Notch1/2/3/4 had a low relapse-free survival. In gastric cancer patients, higher levels of infiltration in their CD4+ T cells, macrophages, neutrophils, and dendritic cells were positive associated with the expression of Notch receptors. Notch expression had significant association with diverse immune marker sets in gastric cancer. Overall, this study provides evidence that Notch1/2/3/4 could become the potential targets for precision treatment and new biomarkers in the prognosis of gastric cancer.

## INTRODUCTION

Gastric cancer (GC) is a very common disease worldwide and has the second highest mortality rate among all cancers. In the past decade, researchers have found that the deregulated expression of specific genes can increase the risk of GC. According to microarray analysis, previous studies revealed that in GC tissues the expression of specific genes is different from that in adjacent normal tissue. Although significant progress has been achieved in GC diagnosis and treatment, the five-year survival of patients is still unsatisfactory [1]. Researchers have identified epigenetic and genetic

alterations as some of the main factors inducing GC. However, the underlying molecular pathogenic mechanisms on molecular level are still obscure. Hence, it is important to identify prognostic markers and potential drug targets to enhance prognosis and individualized treatments.

The Notch signaling pathway is highly conserved among various species. In mammals, four type I transmembrane Notch receptors (Notch 1-4) are synthesized, all of them with unique roles during the generation of immunocytes [2, 3]. Notch signaling also exerts important function in the development and

tissue homeostasis of various organ [4, 5]. Provided the importance of Notch signaling in regulating cellular behavior, it is perhaps not surprising that Notch also has an important role in many types of cancer, particularly due to its importance in the regulation of stem and progenitor cells. Several mechanisms such as epigenetic regulation, posttranslational modifications, gene overexpression and mutations, may lead to the dysregulation of the Notch pathway [2]. Interestingly, Notch activity is associated with oncogenic and tumor-suppressive functions [6, 7]. It is involved in cell survival, cell death pathways, proliferation and growth arrest, as well as cell differentiation into terminally differentiated cells versus cancer cell “stemness” [8]. These functions provide evidence of a context-dependent nature of Notch-induced cellular reactions.

The onset of GC can be described as the result of interactions between a series of factors concerning genetics, epigenetics and the external environment, which jointly lead to the deregulation of the signaling pathways that may induce the onset of cancer [9, 10]. Further, there has been a general belief that it is the dysfunctional oncogenic pathways that induce the onset of GC, which may include the epidermal growth factor receptor (EGFR), Notch, Hedgehog, nuclear factor- $\kappa$ B and Wnt/ $\beta$ -catenin pathway [11]. Among these pathways, Notch signaling is involved in direct cell-cell communication, cell differentiation, proliferation and apoptosis [12].

The balance between immune effector cells in the tumor microenvironment helps the malignant cells escape from the immune response. Tumor infiltrations of tumor associated macrophages, neutrophils, regulatory T cells are correlated with poor prognosis [13–15]. Tumor infiltrating CD8<sup>+</sup> cytotoxic T lymphocytes and DCs are generally associated with favorable outcome of GC [16, 17], although some subsets of these immune cells have inverse prognosis prediction values. High ratios of Foxp3<sup>+</sup>/CD4<sup>+</sup> and Foxp3<sup>+</sup>/CD8<sup>+</sup> in tumors are associated with a poor prognosis [18, 19]; whereas high Th1/Th2 ratio in tumors predicts a good prognosis [20].

In previous studies, researchers have investigated the consequences of dysregulation of the Notch pathway and how it relates to clinicopathological features and prognosis in human GC. Nevertheless, the role of Notch family members in the development and progression of GC remains unknown. This study aims to address this question through in-depth analysis of the mutational activation and expression of Notch family members and their link with prognosis and immune infiltrates in GC patients.

## RESULTS

### Transcriptional levels of Notch in patients with GC

Using the ONCOMINE databases, a comparative analysis investigating transcription levels of Notch receptors was performed on cancer tissues and adjacent normal tissues (Figure 1A). According to the information from five datasets, a significant upregulation of Notch3 mRNA expression was detected in GC patients. In Chen’s dataset [21], the expression of Notch3 in gastric adenocarcinoma was 1.594 and 1.871 times respectively of that in the samples of normal tissue (Table 1). In Wang’s dataset [5], the expression of Notch3 in GC tissue was 2.549 times of that in normal tissue. In DErrico’s dataset [22] the expression of Notch3 in gastric intestinal type adenocarcinoma was 2.869 times of that in normal tissues. In Cho’s dataset [23] the expression of Notch3 in GC tissue was 1.630 times of that in normal tissue. Chen [21] showed that the expression of Notch1 was also higher in cancer tissues, since the expression in gastric intestinal type adenocarcinoma and diffuse gastric adenocarcinoma was 1.920 and 1.733 times respectively of that in patients with normal gastric tissue, respectively (Table 1). Based on the dataset of DErrico, the expression of Notch1 in gastric intestinal type adenocarcinoma was also 2.247 times of that in normal tissue [22]. In comparison with normal tissue, GC exhibited 1.625 [5] and 1.832 [22] times expression levels of Notch2, respectively.

### Expression levels of Notch in GC patients

The comparative investigation was conducted between the Notch mRNA expression in normal and gastric adenocarcinoma tissues based on the information provided by the GEPIA (Gene Expression Profiling Interactive Analysis) dataset. Based on the findings, in comparison with normal gastric tissues, gastric adenocarcinoma tissues presented higher expression levels of Notch1/2/3 (Figure 1B). The expression of Notch1, Notch2, Notch3 and Notch4 proteins, which were examined by immunohistochemistry, were higher in the gastric adenocarcinoma tissues than that in the normal gastric tissues (Figure 1C).

### Association of Notch1/2/3/4 mRNA expression with the prognosis of GC patients

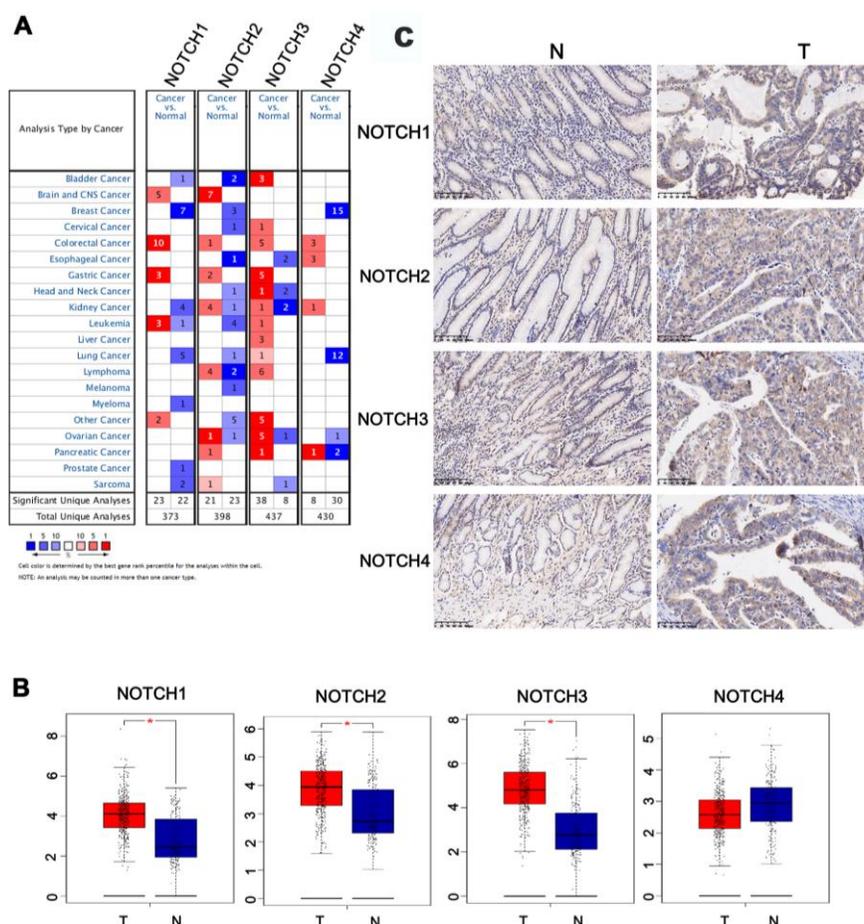
In this study we also conducted further research to investigate how the survival of GC patients was affected by Notch. Kaplan–Meier Plotter tools were applied to analyze the correlation between Notch mRNA levels and the survival of patients with GC in 882 gastric tumors. According to the results of analysis, all the

sampled GC patients were negative correlated with mRNA expression of all the four types of Notch in terms of post-progression survival (PPS), progression-free survival (FP), and overall survival (OS) ( $p < 0.05$ ) (Figure 2A–2D). We further validated these results through the analysis of samples from gastric adenocarcinoma patients with immunohistochemistry (Figure 1C) and survival analysis. The patients with higher Notch1/2/3/4 expression levels had worse overall survival than those with lower Notch1/2/3/4 expression levels (Figure 3A–3D). Generally, lower Notch mRNA expression levels in GC patients indicated higher PPS, FP, and OS.

### Cellular functions, pathways and frequently altered neighbor genes predicted to be affected by changes in Notch receptors in patients with GC

In this study, the cBioPortal online tool was used for analysis of Notch alterations, correlation, and networks. Notch receptors were altered in 223 samples out of 478

patients with stomach adenocarcinoma (47%). Two or more alterations were detected in 51 samples (10%) (Figure 4A). Subsequently, a Notch network was built up using the 50 most frequently altered neighbor genes. According to the results, there was a close correlation between the alterations of Notch and cell cycle-related genes, including E2F1, E2F3, E2F4 and E2F5, as well as histone acetylation-related genes, including HDAC2, HDAC4 and HDAC6 (Figure 4B). The analysis of mRNA was used to assess the correlation between Notch receptors with Pearson's test. Results showed the following Notch members to be significantly positively correlated: Notch1 with Notch2, Notch3, and Notch4; Notch2 with Notch1, Notch3, and Notch4; Notch3 with Notch1, Notch2 and Notch4; Notch4 with Notch1, Notch2 and Notch3 (Figure 4C). Furthermore, a Kaplan-Meier plot and log-rank test demonstrated no significant effect of Notch genetic alterations on OS (Figure 4D,  $p=0.156$ ) but were associated with disease free survival (DFS) (Figure 4D,  $p=0.0318$ ) of gastric adenocarcinoma patients.



**Figure 1. The expression of Notch in different tissues.** (A) The transcription levels of Notch receptors in different types of cancers. (B) The expression of Notch receptors in GC. (C) The expression of Notch protein in GC. Students't-test was used in comparative analysis of Notch expression. T: tumor, N: normal tissue, number (T) = 415, number (N) = 211.

**Table 1. The significant changes of Notch expression in mRNA level between different types of GC and stomach tissues.**

	Type of GC VS. Stomach	Fold Change	P value	t-test	Ref
NOTCH1	Gastric intestinal type adenocarcinoma vs. Normal	2.247	1.04E-13	9.844	DErrico [14]
	Diffuse gastric adenocarcinoma vs. Normal	1.733	6.14E-8	6.700	Chen [13]
	Gastric intestinal type adenocarcinoma vs. Normal	1.920	9.17E-11	7.739	Chen [13]
NOTCH2	Gastric cancer vs. Normal	1.625	5.76E-5	5.128	Wang [5]
	Gastric intestinal type adenocarcinoma vs. Normal	1.832	3.15E-8	6.617	DErrico [14]
NOTCH3	Gastric cancer vs. Normal	2.549	1.42E-6	6.232	Wang [5]
	Diffuse gastric adenocarcinoma vs. Normal	1.871	4.15E-8	7.496	Chen [13]
	Gastric intestinal type adenocarcinoma vs. Normal	1.594	4.15E-10	7.175	Chen [13]
	Gastric intestinal type adenocarcinoma vs. Normal	2.869	1.14E-9	7.723	DErrico [14]
	Diffuse gastric adenocarcinoma vs. Normal	1.630	7.15E-5	4.142	Cho [15]
NOTCH4	NA	NA	NA	NA	NA

Ref: reference

Based on the analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID), we predicted a significant correlation between Notch alterations and functional activities. The function of target host genes were predicted via GO enrichment analysis in terms of three aspects, namely, biological processes, cellular components, and molecular functions. Enrichment analysis was applied to investigate the biological functions of these target genes. According to the enriched gene ontology (GO) cell component (CC) analysis, there was significant enrichment for genes related to the Golgi membrane, endoplasmic reticulum membrane, and receptor complex (Supplementary Figure 1A). For biological processes (BP), there was an enrichment in receptor activity, calcium ion binding and enzyme binding (Supplementary Figure 1B). In addition, in terms of molecular function (MF) results were mainly associated with the Notch signaling pathway and transcription initiation from RNA polymerase II promoter (Supplementary Figure 1C). The KEGG disease terms were mainly enriched in cancers, cancers of the digestive system and musculoskeletal diseases (Supplementary Figure 1D). We also observed significant enrichment in KEGG pathways (Supplementary Figure 1E). Among these

pathways, cell cycle, TGF- $\beta$  signaling pathway, Notch signaling pathways, microRNA in cancer, and Wnt signaling pathway were involved in the tumorigenesis and pathogenesis of GC (Supplementary Figures 2A and 2B).

#### **Notch expression is correlated with immune infiltration levels in gastric cancer**

Tumor-infiltrating lymphocytes have been used to predict sentinel lymph node status and survival in cancers [24]. Hence, we explored the correlation between the levels of immune infiltration and the expression of Notch in gastric adenocarcinoma patients based on TIMER. Interestingly, we found high levels of Notch mRNA expression to be associated with high immune infiltration in gastric adenocarcinoma. Notch1 mRNA expression level was significantly positively correlated with infiltrating levels of CD4+ T cells ( $r = 0.262$ ,  $P = 3.87e-07$ ) (Figure 5A). Notch2 mRNA expression level was significantly positively correlated with infiltrating levels of CD4+ T cells ( $r = 0.394$ ,  $P = 5.04e-15$ ), macrophages ( $r = 0.53$ ,  $P = 3.07e-28$ ), neutrophils ( $r = 0.205$ ,  $P = 7.24e-05$ ) and dendritic cells (DCs) ( $r = 0.362$ ,  $P = 6.62e-13$ ) (Figure 5B). Notch3 mRNA expression level was significantly positively correlated with infiltrating levels of CD4+ T cells ( $r =$

0.378,  $P = 7.41e-14$ ) and macrophages ( $r = 0.354$ ,  $P = 2.17e-12$ ) (Figure 5C). Similarly, Notch4 mRNA expression was positively correlated with infiltrating levels of CD4+ T cells ( $r = 0.434$ ,  $P = 2.82e-18$ ) and macrophages ( $r = 0.342$ ,  $P = 1.30e-11$ ) (Figure 5D). In addition, this study did not find significant correlations

between the expression of Notch and infiltrating levels of B cells and CD8+ T cells in gastric adenocarcinoma. These findings strongly indicated that Notch played an important role in immune infiltration in gastric adenocarcinoma, particularly for CD4+ T cells and macrophages.

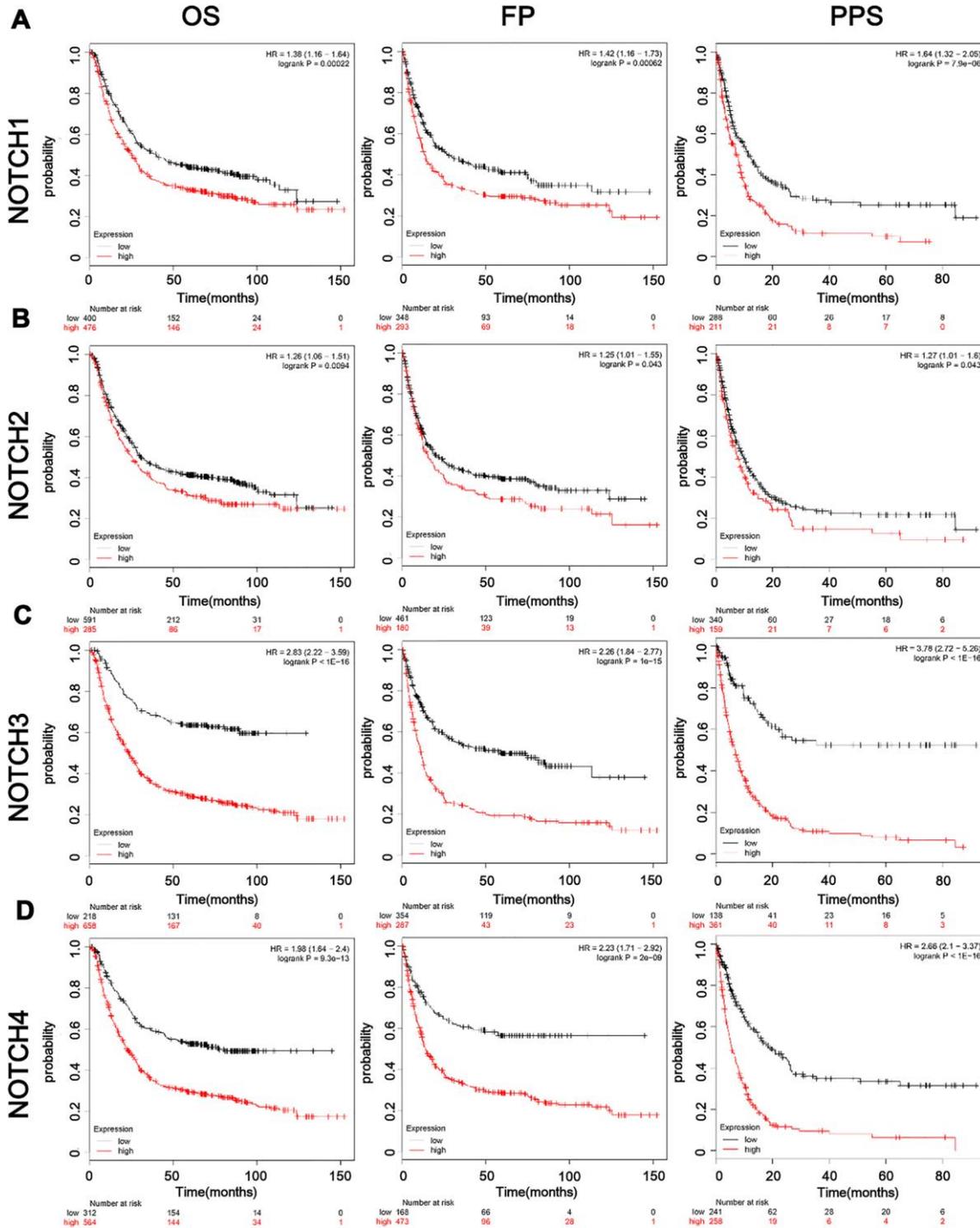
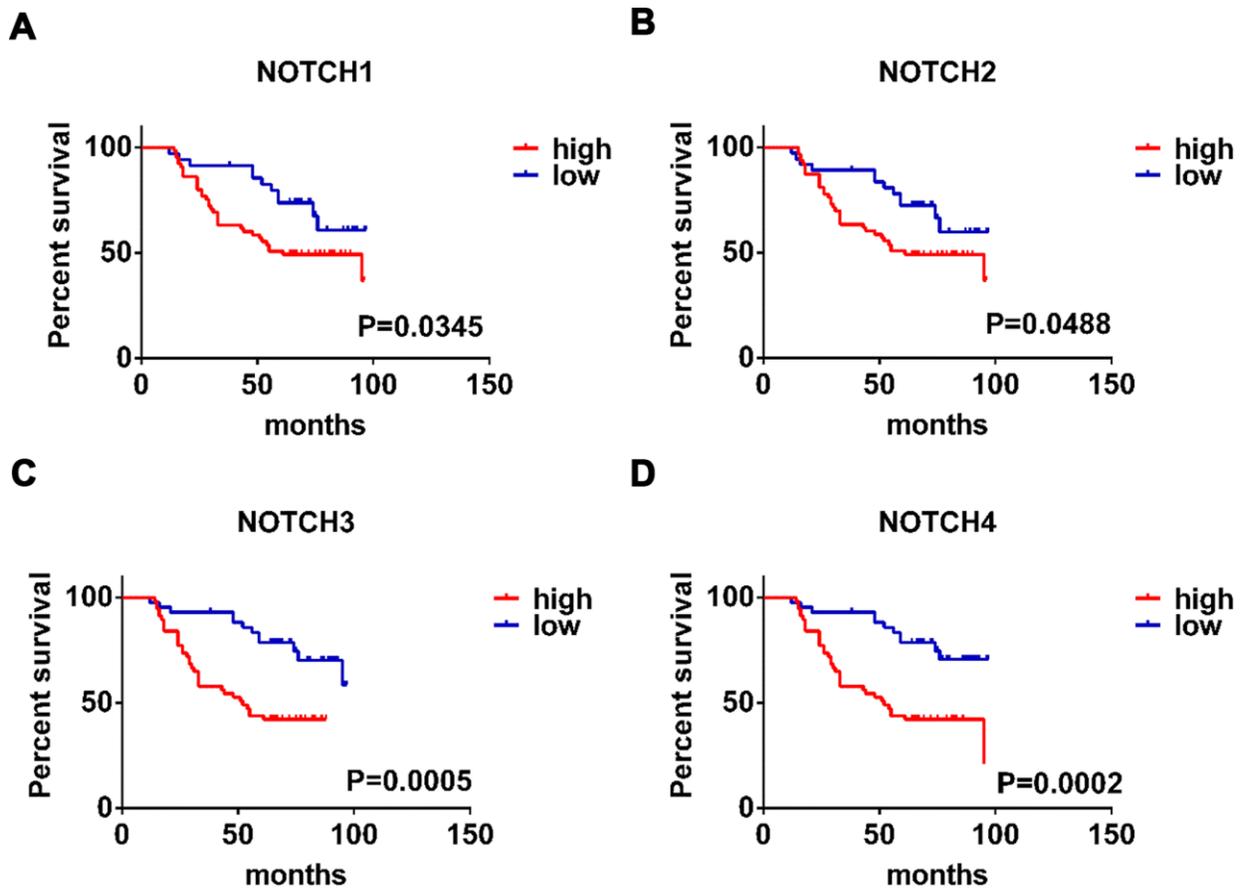


Figure 2. The prognostic value of mRNA level of Notch receptors in GC patients. (A–D). Logrank test was used in analysis of OS/FP/PPS.

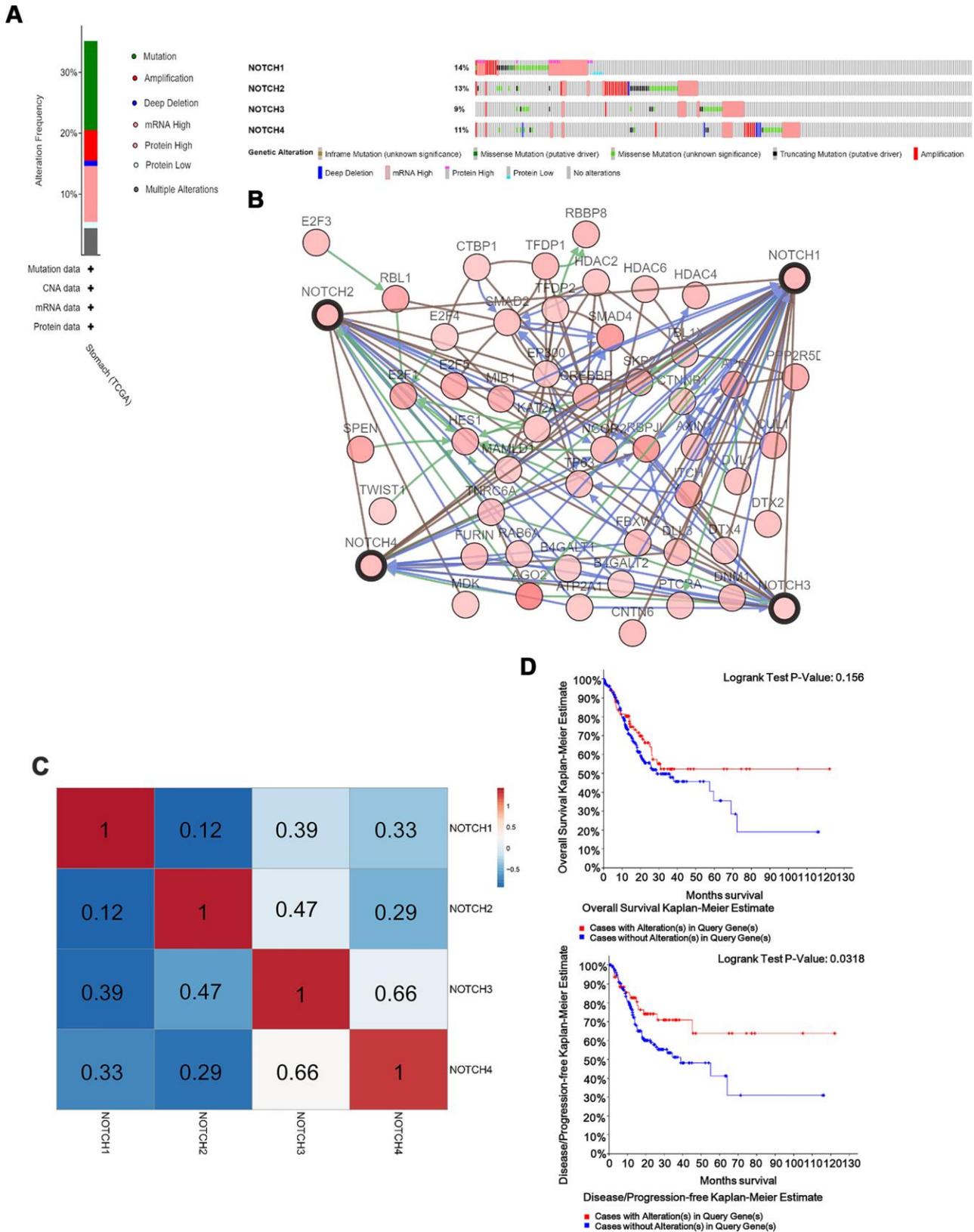
### Correlation analysis between Notch expression and immune markers

In order to detect the relationship between Notch and the diverse immune infiltrating cells, we concentrated particularly on the correlations between Notch and immune markers of various immune cells in gastric adenocarcinoma using the TIMER databases. We analyzed different immune cells including tumor-associated macrophages (TAMs), neutrophils and DCs in gastric adenocarcinoma (Table 2). Meanwhile, Th1 cells, Th2 cells and Tregs were also analyzed. After adjusting the correlation for tumor purity, results demonstrated that the Notch mRNA expression level had significant correlations with TAMs, DCs, CD4+ T cells and neutrophils in gastric adenocarcinoma (Table 2). Specifically, we showed chemokine (C-C motif) ligand (CCL)-2, IL-10, CD163, CSF1R and FCGR2A of TAMs, ITGAX, CD1C, NRP1 and THBD of DCs, CCR7, ITGAM of neutrophils, STAT4 of Th1 cells, GATA3, CXCR4, CCR4 and CCR8 of Th2 cells, STAT5B and TGFB1 of Tregs to be significantly correlated with the expression of Notch2 in gastric

adenocarcinoma (Table 2). High Notch3/4 expression was associated with high levels of infiltration of macrophage and CD4+ T cells in gastric adenocarcinoma. TAM markers such as CCL2 and CSF1R, DCs markers such as ITGAX, NRP1 and THBD, neutrophils markers such as CCR7 and ITGAM, Th2 markers such as GATA3, CXCR4, CCR4 and CCR8 and Treg markers such as STAT5B and TGFB1 showed significant correlation with Notch3 expression. TAM markers such as CCL2 and CSF1R, DCs markers such as ITGAX, CD1C, NRP1 and THBD, neutrophils markers such as CCR7 and ITGAM, Th2 markers such as GATA3, CXCR4, CCR4, Treg markers such as FOXP3, STAT5B, TGFB1 were also significantly correlated the expression of Notch4. Based on all the findings above, we provide evidence that infiltration of TAMs/Th2/Tregs is increased by expression of Notch3/4. Additionally, Notch1 had a positive correlation with CD4, GATA3 and STAT5B and a negative correlation with CD59 in gastric adenocarcinoma. Overall, these results further confirm the critical role of Notch in inhibiting immune activities in the gastric adenocarcinoma microenvironment.



**Figure 3. The prognostic value of expression level of Notch receptors in gastric adenocarcinoma patients. (A–D).** Logrank test was used in analysis of overall survival. The numbers of high Notch expression group and low Notch expression group are 50, respectively.



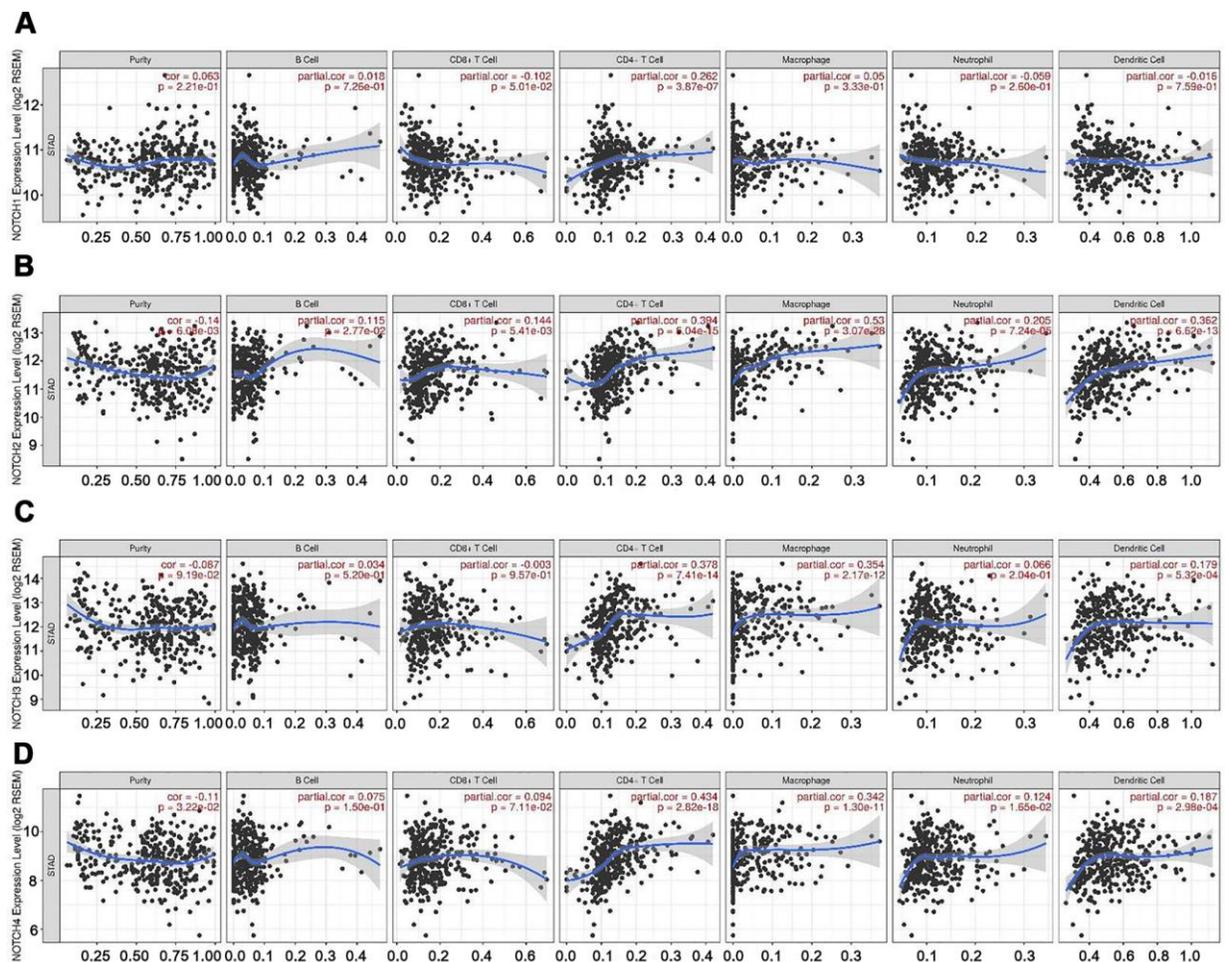
**Figure 4. Notch genes expression and mutation analysis in GC.** (A) Notch1, Notch2, Notch3 and Notch4 mutation rates were 14%, 13%, 9% and 11%, respectively. (B) Network for 4 Notch receptors and the 50 most frequently altered neighbor genes. (C) Genetic alterations in Notch receptor were correlated with longer DFS in gastric adenocarcinoma patients. (D) The relationship between genetic alterations in Notch and OS/DFS. Logrank test was used in analysis of OS/DFS.

## DISCUSSION

The importance and effect of the Notch signaling pathway on cellular activities and in cancer with either oncogenic or tumor-suppressive functions have been widely recognized [2, 25, 26]. Notch signaling functions as juxtacrine signaling between cells. This type of signaling allows Notch to regulate heterotypic interactions between stroma and tumor. These interactions are known to be of importance in various aspects of tumor biology, such as angiogenesis, cancer stem cell maintenance, immune infiltration, and resistance to therapy. Despite the already confirmed effects of Notch signaling on GC, the functions of different Notch family members in GC remain to be elucidated. In this study, we conducted detailed analyses on various members of Notch referring to

mutation, expression and prognostic values in GC patients.

Results from our study showed that over-expression of mRNA and protein were found in Notch family members. Also, higher mRNA expressions of Notch1/2/3/4 were significantly associated with shorter OS in GC patients. Moreover, a high Notch mutation rate (47%) was observed in gastric adenocarcinoma patients. Further, genetic alteration of Notch receptors was associated with shorter DFS in gastric adenocarcinoma patients. Finally, the functions and pathways affected by mutations in Notch receptors as well as the most significant 50 frequently altered neighbor genes in gastric adenocarcinoma patients were analyzed. Results showed that cell cycle-related genes, including E2F1, E2F3, E2F4 and E2F5, and histone



**Figure 5. Correlation of Notch expression with immune infiltration level in GC. (A)** Notch1 expression had significant positive correlations with infiltrating levels of CD4+ T cells (n = 415). **(B)** Notch2 expression had significant correlations with infiltrating levels of CD4+ T cells, macrophages, neutrophils, and dendritic cells in GC (n = 415). **(C)** Notch3 expression was significantly related to infiltrating levels of CD4+ T cells and macrophages in GC but no significant correlation with infiltrating level of B cells (n = 415). **(D)** Notch4 expression had significant positive correlations with infiltrating levels of CD4+ T cells and macrophages in GC (n = 415). Spearman's correlation was used to analyze correlation between Notch receptors and immune cells.

**Table 2. Correlation analysis between Notch and related markers of immune cells.**

Description	Gene marker	NOTCH1		NOTCH2		NOTCH3		NOTCH4	
		Cor	P	Cor	P	Cor	P	Cor	P
TAM	CCL2	0.052	0.313	0.264	***	0.239	***	0.27	***
	IL10	0.081	0.117	0.27	***	0.171	***	0.199	***
	CD163	0.097	0.058	0.454	***	0.193	***	0.163	**
	VSIG4	0.019	0.711	0.402	***	0.177	***	0.099	0.053
	CSF1R	0.066	0.198	0.466	***	0.251	***	0.235	***
	FCGR2A	-0.012	0.22	0.337	***	0.168	**	0.125	*
	FCER2	0.141	**	0.249	***	0.197	***	0.327	***
DCs	ITGAX	0.111	*	0.327	***	0.252	***	0.278	***
	CD1C	0.031	0.55	0.396	***	0.182	***	0.287	***
	NRP1	0.135	**	0.483	***	0.461	***	0.507	***
	THBD	0.164	**	0.453	***	0.518	***	0.598	***
Neutrophils	CCR7	0.15	**	0.324	***	0.209	***	0.338	***
	ITGAM	0.105	*	0.479	***	0.286	***	0.28	***
	CD59	-0.323	***	0.311	***	0.152	**	0.153	**
Th1	STAT4	0.047	0.363	0.316	***	0.075	0.146	0.152	**
	TBX21	0.149	**	0.185	***	0.088	0.088	0.148	**
	CD4	0.324	***	0.178	***	0.324	***	0.197	***
Th2	GATA3	0.22	***	0.173	***	0.22	***	0.2	***
	CXCR4	0.097	0.059	0.375	***	0.213	***	0.288	***
	CCR4	0.169	***	0.408	***	0.24	***	0.3	***
	CCR8	0.137	**	0.347	***	0.216	***	0.198	***
Treg	FOXP3	0.136	**	0.173	***	0.176	***	0.204	***
	STAT5B	0.351	***	0.584	***	0.399	***	0.414	***
	TGFB1	0.152	**	0.373	***	0.532	***	0.384	***

\*P<0.05, \*\*P<0.01, \*\*\*P<0.001

acetylation-related genes, including HDAC2, HDAC4 and HDAC6 were significantly affected by mutations in Notch receptors. Results also showed the following to be remarkably regulated by Notch mutation in GC: biological processes such as receptor activity, calcium ion binding and enzyme binding; cellular components such as Golgi membrane, endoplasmic reticulum membrane, and receptor complex; molecular functions such as Notch signaling pathway and transcription initiation from RNA polymerase II promoter; KEGG disease terms such as cancers, cancers of the digestive system and musculoskeletal diseases; KEGG pathways such as cell cycle, Notch signaling pathways and TGF- $\beta$  signaling pathway.

Of note, Notch1 played a critical role in regulating the senescence secretome in fibroblasts [27], which was perhaps a part of its function in the regulation of stromal activation in the process of tumorigenesis. In breast cancer models, Notch1 was induced by fibroblast-derived CCL2 to maintain a stem cell phenotype and had a possible oncogenic role [17, 28]. Importantly, Notch1 activation enabled primary

melanoma cells to acquire metastatic capabilities [29]. In gastric cancer, the AKT1/NF- $\kappa$ B/Notch1/PTEN axis had a significant role in the development of chemoresistance [30]. Notch1 activation also showed correlation with GC progression and was defined as an independent prognostic factor [31]. In B cell lymphoma, Jag1 induced the expression of FGF4 which in turn activated Notch2 in lymphoma cells [32]. Notch ligand Jag1 interacted with Notch3 to regulate the resistance [33]. Moreover, Notch3 triggered by stromal cell-derived exosomes activated antiviral signaling depending on STAT1 in cancer cells [33]. Induction of Notch3 by CAFs led to an increase in proliferation of cancer stem cells [34]. Meanwhile, the renewal of cancer stem cell could be induced by IL-6 through Notch3 in breast cancer [35]. Notch4 had a causative role in the tumorigenesis [36] and might affect the development of fibroblasts [37]. Recent data also showed that Notch is involved in liver glucose and lipid homeostasis [38, 39].

Another important aspect of this study was that Notch expression was correlated with diverse immune

infiltration levels in gastric adenocarcinoma. Notch1 mRNA expression level was significantly positively correlated with infiltrating levels of CD4+ T cells. Notch2 mRNA expression level was significantly positively correlated with infiltrating levels of CD4+ T cells, macrophages, neutrophils and DCs. Notch3 mRNA expression level was significantly positively correlated with infiltrating levels of CD4+ T cells and macrophages. Moreover, the correlation between Notch expression and the marker genes of immune cells imply the role of Notch in regulating tumor immunology in gastric adenocarcinoma. The M2 macrophage markers such as CD163, VSIG4, and CSF1R showed correlations with Notch2 expression. These results indicated the potential role of Notch2 in regulation of TAMs polarization. Furthermore, our results revealed that Notch had the potential to activate Tregs. The increase in Notch expression positively correlates with the expression of Treg markers (FOXP3, CCR8, STAT5B in Table 2). In addition, there were significant correlations between Notch expression and several markers of T helper cells (Th1, Th2), DCs and neutrophils in gastric adenocarcinoma. These correlations could be indicative of a potential mechanism that Notch regulated immune cells in gastric adenocarcinoma. Together these findings suggested that Notch play a significant role in recruitment and regulation of immune infiltrating cells in gastric adenocarcinoma.

In this study, a systemic analysis was performed on Notch receptors expression, mutation and GC patients' prognosis, which provided a further understanding of the biomolecular properties of GC. Our results demonstrated that the high expression of Notch1/2/3/4 in GC tissues might exert a significant function in GC tumorigenesis. High Notch1/2/3/4 expression could also act as molecular markers to categorized high-risk subgroups of GC patients. According to the present study, Notch1/2/3/4 could be potential therapeutic targets for GC, and transcription levels of Notch1/2/3/4 could be potential prognostic markers overall promoting GC survival and prognostic accuracy.

## MATERIALS AND METHODS

### ONCOMINE analysis

Using the online cancer microarray database, ONCOMINE (<https://www.oncomine.org/>), we analyzed the transcription levels of Notch in various cancers tissues. A students' t-test was performed to conduct a comparative analysis of the different Notch expression in normal tissues versus cancer tissue samples. The cut-off value for a significant fold change was 1.5, while the cut-off value for significant p value was 0.0001.

### GEPIA dataset

GEPIA (<http://gepia.cancer-pku.cn/>), a web server for analyzing the sequencing expression data of RNA based on 9,736 tumors and 8,587 normal samples from the cancer genome atlas (TCGA) and the GTEx projects [40], was used for analysis of differential expression.

### The Kaplan-Meier plotter

Kaplan-Meier Plotter (<https://kmplot.com/>) was applied to the assessment the prognostic value of Notch. This online database provides gene expression data and information on patients' survival performance based on 882 clinical GC cases [41]. All the sampled patients were categorized into groups based on median expression of Notch (high vs. low expression) to analyze the overall survival (OS), progression-free survival (FP), and post progression survival (PPS) of GC patients using a Kaplan-Meier survival plot. We only chose the best JetSet probe set of Notch to obtain Kaplan-Meier plots.

### TCGA data and cBioPortal

TCGA provides sequencing and pathology data for over 30 different cancers [42]. We used the dataset for stomach adenocarcinoma, which included 478 cases with pathology reports to further analyze Notch expression using cBioPortal (<https://www.cbioportal.org/>) [43]. From the genomic profiles, we obtained data regarding protein expression Z-scores (RPPA), mRNA expression z-scores, putative copy-number alterations (CNA) from GISTIC and mutations. Co-expression and network analysis were conducted following online instructions of cBioPortal.

### Immunohistochemistry

Tissue samples were fixed in 4% paraformaldehyde and embedded in paraffin. Anti-Notch1 (1:200; CST, Cat#: 3608, Danvers, MA, USA), anti-Notch2 (1:200; CST, Cat#: 5732), anti-Notch3 (1:200; Abcam, Cat#: ab23426, Cambridge, MA, USA) and anti-Notch4 (1:200; Abcam, Cat#: ab222400) were used as primary antibodies. Immunohistochemistry was performed according to the method previously described [44–46]. Gastric cancer tissues and adjacent normal gastric tissues were obtained from the Huashan Hospital (Shanghai, China).

### Immune infiltrates analysis in TIMER database

TIMER is a comprehensive tool established for analyzing immune infiltrates across different types of

cancer (<https://cistrome.shinyapps.io/timer/>) [47]. We analyzed Notch expression in different cancer types and the correlation of Notch expression with the abundance of immune infiltrates. Meanwhile, correlations between Notch expression and gene markers of tumor-infiltrating immune cells were also explored.

## AUTHOR CONTRIBUTIONS

Conceptualization, Ziqiang Zhang; Investigation, Jian Hu, Jianghong Yu; Writing – Original Draft, Jian Hu, Jianghong Yu; Writing – Review & Editing, Ziqiang Zhang, Jun Gan, Ning Song, Liubin Shi, Jie Liu, Jianjun Du; Visualization, Ziqiang Zhang; Supervision, Jie Liu, Jianjun Du; Funding Acquisition, Jianjun Du.

## CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

## FUNDING

This study was supported by the National Natural Science Foundation of China (grant No. 81270440).

## REFERENCES

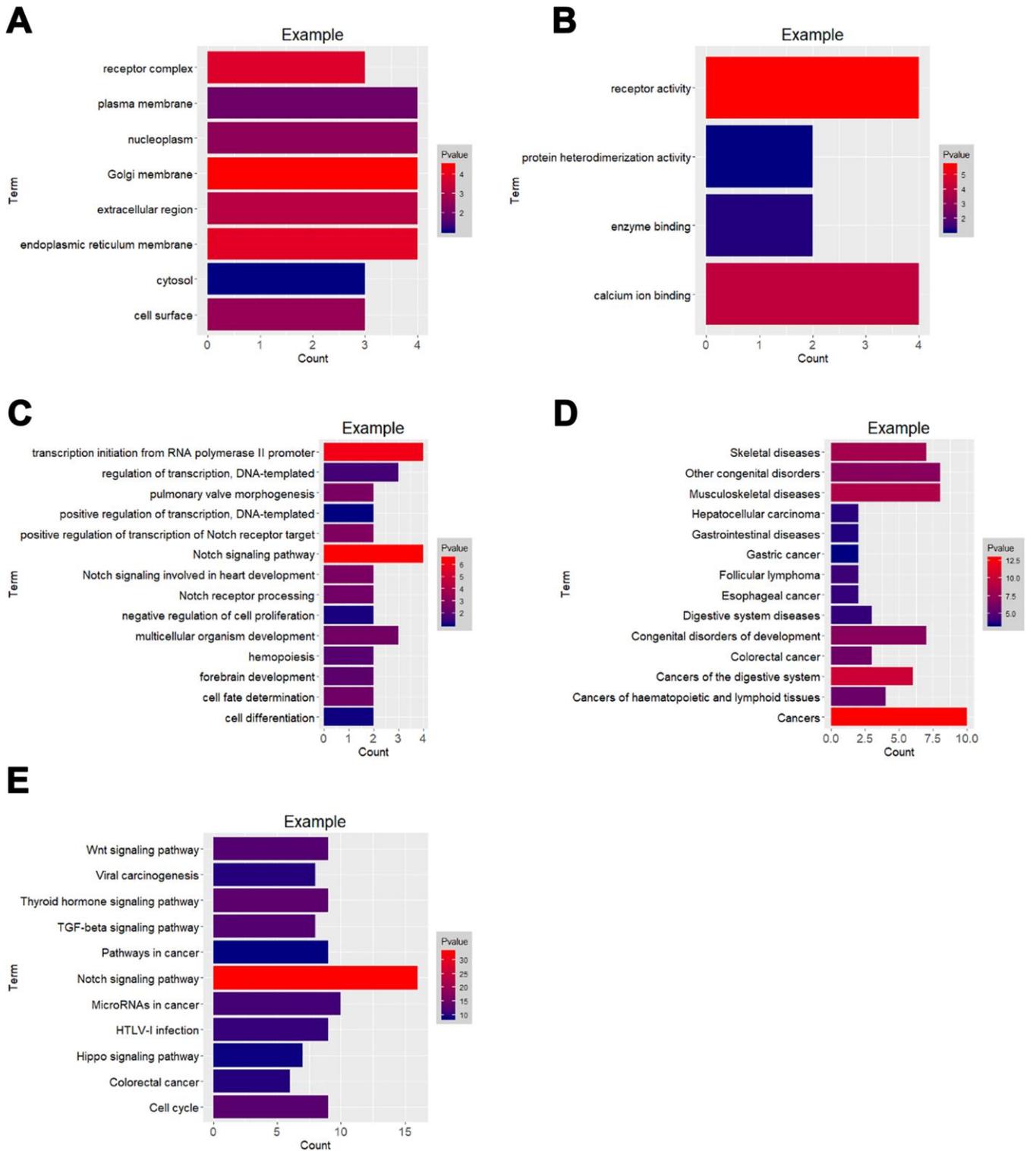
1. Luebeck EG, Curtius K, Jeon J, Hazelton WD. Impact of tumor progression on cancer incidence curves. *Cancer Res.* 2013; 73:1086–96. <https://doi.org/10.1158/0008-5472.CAN-12-2198> PMID:23054397
2. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a centennial for notch signaling. *Cancer Cell.* 2014; 25:318–34. <https://doi.org/10.1016/j.ccr.2014.02.018> PMID:24651013
3. Espinoza I, Miele L. Notch inhibitors for cancer treatment. *Pharmacol Ther.* 2013; 139:95–110. <https://doi.org/10.1016/j.pharmthera.2013.02.003> PMID:23458608
4. Wilson A, Radtke F. Multiple functions of Notch signaling in self-renewing organs and cancer. *FEBS Lett.* 2006; 580:2860–68. <https://doi.org/10.1016/j.febslet.2006.03.024> PMID:16574107
5. Wang Q, Wen YG, Li DP, Xia J, Zhou CZ, Yan DW, Tang HM, Peng ZH. Upregulated INHBA expression is associated with poor survival in gastric cancer. *Med Oncol.* 2012; 29:77–83. <https://doi.org/10.1007/s12032-010-9766-y> PMID:21132402
6. Koch U, Radtke F. Notch and cancer: a double-edged sword. *Cell Mol Life Sci.* 2007; 64:2746–62. <https://doi.org/10.1007/s00018-007-7164-1> PMID:17687513
7. Radtke F, Raj K. The role of Notch in tumorigenesis: oncogene or tumour suppressor? *Nat Rev Cancer.* 2003; 3:756–67. <https://doi.org/10.1038/nrc1186> PMID:14570040
8. South AP, Cho RJ, Aster JC. The double-edged sword of Notch signaling in cancer. *Semin Cell Dev Biol.* 2012; 23:458–64. <https://doi.org/10.1016/j.semcdb.2012.01.017> PMID:22309843
9. Brose K, Tessier-Lavigne M. Slit proteins: key regulators of axon guidance, axonal branching, and cell migration. *Curr Opin Neurobiol.* 2000; 10:95–102. [https://doi.org/10.1016/S0959-4388\(99\)00066-5](https://doi.org/10.1016/S0959-4388(99)00066-5) PMID:10679444
10. Houghton J, Wang TC. Helicobacter pylori and gastric cancer: a new paradigm for inflammation-associated epithelial cancers. *Gastroenterology.* 2005; 128:1567–78. <https://doi.org/10.1053/j.gastro.2005.03.037> PMID:15887152
11. Wu WK, Cho CH, Lee CW, Fan D, Wu K, Yu J, Sung JJ. Dysregulation of cellular signaling in gastric cancer. *Cancer Lett.* 2010; 295:144–53. <https://doi.org/10.1016/j.canlet.2010.04.025> PMID:20488613
12. Previs RA, Coleman RL, Harris AL, Sood AK. Molecular pathways: translational and therapeutic implications of the Notch signaling pathway in cancer. *Clin Cancer Res.* 2015; 21:955–61. <https://doi.org/10.1158/1078-0432.CCR-14-0809> PMID:25388163
13. Zhang QW, Liu L, Gong CY, Shi HS, Zeng YH, Wang XZ, Zhao YW, Wei YQ. Prognostic significance of tumor-associated macrophages in solid tumor: a meta-analysis of the literature. *PLoS One.* 2012; 7:e50946. <https://doi.org/10.1371/journal.pone.0050946> PMID:23284651
14. Kashimura S, Saze Z, Terashima M, Soeta N, Ohtani S, Osuka F, Kogure M, Gotoh M. CD83(+) dendritic cells and Foxp3(+) regulatory T cells in primary lesions and regional lymph nodes are inversely correlated with prognosis of gastric cancer. *Gastric Cancer.* 2012; 15:144–53. <https://doi.org/10.1007/s10120-011-0090-9> PMID:22083420
15. Zhao JJ, Pan K, Wang W, Chen JG, Wu YH, Lv L, Li JJ, Chen YB, Wang DD, Pan QZ, Li XD, Xia JC. The prognostic value of tumor-infiltrating neutrophils in

- gastric adenocarcinoma after resection. *PLoS One*. 2012; 7:e33655.  
<https://doi.org/10.1371/journal.pone.0033655>  
 PMID:[22442706](https://pubmed.ncbi.nlm.nih.gov/22442706/)
16. Lee HE, Chae SW, Lee YJ, Kim MA, Lee HS, Lee BL, Kim WH. Prognostic implications of type and density of tumour-infiltrating lymphocytes in gastric cancer. *Br J Cancer*. 2008; 99:1704–11.  
<https://doi.org/10.1038/sj.bjc.6604738>  
 PMID:[18941457](https://pubmed.ncbi.nlm.nih.gov/18941457/)
  17. Tsukayama S, Omura K, Yoshida K, Tanaka Y, Watanabe G. Prognostic value of CD83-positive mature dendritic cells and their relation to vascular endothelial growth factor in advanced human gastric cancer. *Oncol Rep*. 2005; 14:369–75.  
<https://doi.org/10.3892/or.14.2.369> PMID:[16012717](https://pubmed.ncbi.nlm.nih.gov/16012717/)
  18. Kim HI, Kim H, Cho HW, Kim SY, Song KJ, Hyung WJ, Park CG, Kim CB. The ratio of intra-tumoral regulatory T cells (Foxp3+)/helper T cells (CD4+) is a prognostic factor and associated with recurrence pattern in gastric cardia cancer. *J Surg Oncol*. 2011; 104:728–33.  
<https://doi.org/10.1002/jso.22038> PMID:[21792941](https://pubmed.ncbi.nlm.nih.gov/21792941/)
  19. Shen Z, Zhou S, Wang Y, Li RL, Zhong C, Liang C, Sun Y. Higher intratumoral infiltrated Foxp3+ Treg numbers and Foxp3+/CD8+ ratio are associated with adverse prognosis in resectable gastric cancer. *J Cancer Res Clin Oncol*. 2010; 136:1585–95.  
<https://doi.org/10.1007/s00432-010-0816-9>  
 PMID:[20221835](https://pubmed.ncbi.nlm.nih.gov/20221835/)
  20. Ubukata H, Motohashi G, Tabuchi T, Nagata H, Konishi S, Tabuchi T. Evaluations of interferon- $\gamma$ /interleukin-4 ratio and neutrophil/lymphocyte ratio as prognostic indicators in gastric cancer patients. *J Surg Oncol*. 2010; 102:742–47.  
<https://doi.org/10.1002/jso.21725> PMID:[20872813](https://pubmed.ncbi.nlm.nih.gov/20872813/)
  21. Chen X, Leung SY, Yuen ST, Chu KM, Ji J, Li R, Chan AS, Law S, Troyanskaya OG, Wong J, So S, Botstein D, Brown PO. Variation in gene expression patterns in human gastric cancers. *Mol Biol Cell*. 2003; 14:3208–15.  
<https://doi.org/10.1091/mbc.e02-12-0833>  
 PMID:[12925757](https://pubmed.ncbi.nlm.nih.gov/12925757/)
  22. D'Errico M, de Rinaldis E, Blasi MF, Viti V, Falchetti M, Calcagnile A, Sera F, Saieva C, Ottini L, Palli D, Palombo F, Giuliani A, Dogliotti E. Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. *Eur J Cancer*. 2009; 45:461–69.  
<https://doi.org/10.1016/j.ejca.2008.10.032>  
 PMID:[19081245](https://pubmed.ncbi.nlm.nih.gov/19081245/)
  23. Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM, Kim SB, Kim H, Hong SW, Park YN, Noh SH, Park ES, Chu IS, Hong WK, Ajani JA, Lee JS. Gene expression signature-based prognostic risk score in gastric cancer. *Clin Cancer Res*. 2011; 17:1850–7.  
<https://doi.org/10.1158/1078-0432.CCR-10-2180>  
 PMID:[21447720](https://pubmed.ncbi.nlm.nih.gov/21447720/)
  24. Azimi F, Scolyer RA, Rumcheva P, Moncrieff M, Murali R, McCarthy SW, Saw RP, Thompson JF. Tumour-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J Clin Oncol*. 2012; 30:2678–83.  
<https://doi.org/10.1200/JCO.2011.37.8539>  
 PMID:[22711850](https://pubmed.ncbi.nlm.nih.gov/22711850/)
  25. Aster JC, Pear WS, Blacklow SC. The Varied Roles of Notch in Cancer. *Annu Rev Pathol*. 2017; 12:245–75.  
<https://doi.org/10.1146/annurev-pathol-052016-100127> PMID:[27959635](https://pubmed.ncbi.nlm.nih.gov/27959635/)
  26. Nowell CS, Radtke F. Notch as a tumour suppressor. *Nat Rev Cancer*. 2017; 17:145–59.  
<https://doi.org/10.1038/nrc.2016.145> PMID:[28154375](https://pubmed.ncbi.nlm.nih.gov/28154375/)
  27. Hoare M, Ito Y, Kang TW, Weekes MP, Matheson NJ, Patten DA, Shetty S, Parry AJ, Menon S, Salama R, Antrobus R, Tomimatsu K, Howat W, et al. NOTCH1 mediates a switch between two distinct secretomes during senescence. *Nat Cell Biol*. 2016; 18:979–92.  
<https://doi.org/10.1038/ncb3397> PMID:[27525720](https://pubmed.ncbi.nlm.nih.gov/27525720/)
  28. Weijzen S, Rizzo P, Braid M, Vaishnav R, Jonkheer SM, Zlobin A, Osborne BA, Gottipati S, Aster JC, Hahn WC, Rudolf M, Siziopikou K, Kast WM, Miele L. Activation of Notch-1 signaling maintains the neoplastic phenotype in human Ras-transformed cells. *Nat Med*. 2002; 8:979–86.  
<https://doi.org/10.1038/nm754> PMID:[12185362](https://pubmed.ncbi.nlm.nih.gov/12185362/)
  29. Balint K, Xiao M, Pinnix CC, Soma A, Veres I, Juhasz I, Brown EJ, Capobianco AJ, Herlyn M, Liu ZJ. Activation of Notch1 signaling is required for beta-catenin-mediated human primary melanoma progression. *J Clin Invest*. 2005; 115:3166–76.  
<https://doi.org/10.1172/JCI25001> PMID:[16239965](https://pubmed.ncbi.nlm.nih.gov/16239965/)
  30. Zhou W, Fu XQ, Zhang LL, Zhang J, Huang X, Lu XH, Shen L, Liu BN, Liu J, Luo HS, Yu JP, Yu HG. The AKT1/NF-kappaB/Notch1/PTEN axis has an important role in chemoresistance of gastric cancer cells. *Cell Death Dis*. 2013; 4:e847.  
<https://doi.org/10.1038/cddis.2013.375>  
 PMID:[24113181](https://pubmed.ncbi.nlm.nih.gov/24113181/)
  31. Zhang H, Wang X, Xu J, Sun Y. Notch1 activation is a poor prognostic factor in patients with gastric cancer. *Br J Cancer*. 2014; 110:2283–90.  
<https://doi.org/10.1038/bjc.2014.135> PMID:[24642622](https://pubmed.ncbi.nlm.nih.gov/24642622/)
  32. Cao Z, Ding BS, Guo P, Lee SB, Butler JM, Casey SC, Simons M, Tam W, Felsher DW, Shido K, Rafii A, Scandura JM, Rafii S. Angiocrine factors deployed by

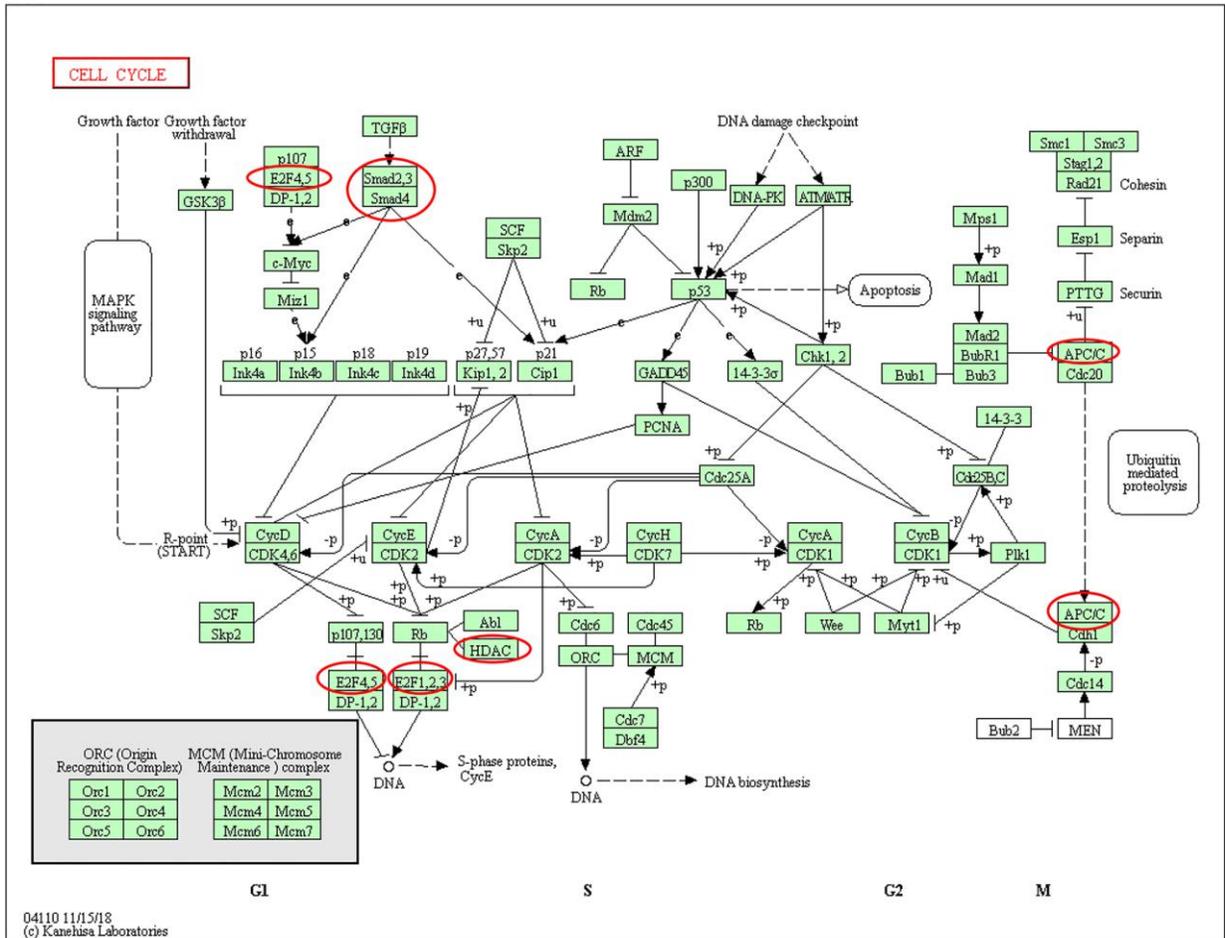
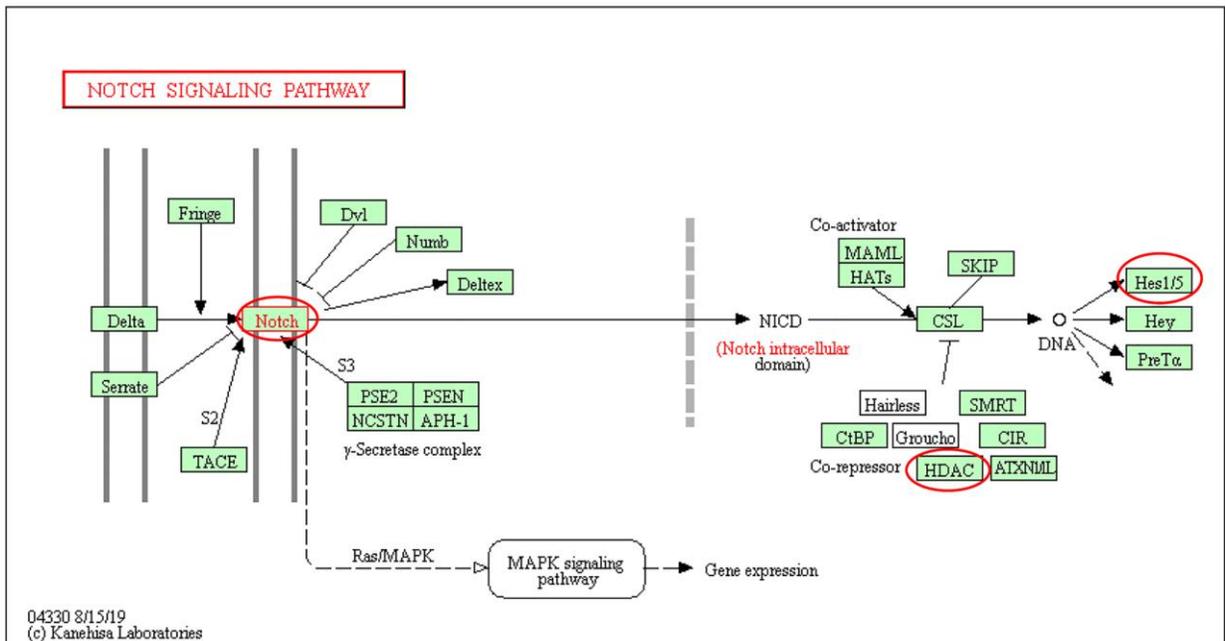
- tumor vascular niche induce B cell lymphoma invasiveness and chemoresistance. *Cancer Cell*. 2014; 25:350–65.  
<https://doi.org/10.1016/j.ccr.2014.02.005>  
 PMID:[24651014](https://pubmed.ncbi.nlm.nih.gov/24651014/)
33. Boelens MC, Wu TJ, Nabet BY, Xu B, Qiu Y, Yoon T, Azzam DJ, Twyman-Saint Victor C, Wiemann BZ, Ishwaran H, Ter Brugge PJ, Jonkers J, Slingerland J, Minn AJ. Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. *Cell*. 2014; 159:499–513.  
<https://doi.org/10.1016/j.cell.2014.09.051>  
 PMID:[25417103](https://pubmed.ncbi.nlm.nih.gov/25417103/)
  34. Liu C, Liu L, Chen X, Cheng J, Zhang H, Zhang C, Shan J, Shen J, Qian C. LSD1 Stimulates Cancer-Associated Fibroblasts to Drive Notch3-Dependent Self-Renewal of Liver Cancer Stem-like Cells. *Cancer Res*. 2018; 78:938–49.  
<https://doi.org/10.1158/0008-5472.CAN-17-1236>  
 PMID:[29259010](https://pubmed.ncbi.nlm.nih.gov/29259010/)
  35. Sansone P, Storci G, Tavolari S, Guarnieri T, Giovannini C, Taffurelli M, Ceccarelli C, Santini D, Paterini P, Marcu KB, Chieco P, Bonafè M. IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. *J Clin Invest*. 2007; 117:3988–4002.  
 PMID:[18060036](https://pubmed.ncbi.nlm.nih.gov/18060036/)
  36. Raafat A, Bargo S, Anver MR, Callahan R. Mammary development and tumorigenesis in mice expressing a truncated human Notch4/Int3 intracellular domain (h-Int3sh). *Oncogene*. 2004; 23:9401–07.  
<https://doi.org/10.1038/sj.onc.1208187>  
 PMID:[15531924](https://pubmed.ncbi.nlm.nih.gov/15531924/)
  37. Hu MS, Borrelli MR, Januszyk M, Luan A, Malhotra S, Walmsley GG, Hong WX, Tevlin R, Gurtner GC, Longaker MT, Lorenz HP. Pathway Analysis of Gene Expression of E14 Versus E18 Fetal Fibroblasts. *Adv Wound Care (New Rochelle)*. 2018; 7:1–10.  
<https://doi.org/10.1089/wound.2017.0763>  
 PMID:[29344429](https://pubmed.ncbi.nlm.nih.gov/29344429/)
  38. Pajvani UB, Shawber CJ, Samuel VT, Birkenfeld AL, Shulman GI, Kitajewski J, Accili D. Inhibition of Notch signaling ameliorates insulin resistance in a FoxO1-dependent manner. *Nat Med*. 2011; 17:961–67.  
<https://doi.org/10.1038/nm.2378> PMID:[21804540](https://pubmed.ncbi.nlm.nih.gov/21804540/)
  39. Pajvani UB, Qiang L, Kangsamaksin T, Kitajewski J, Ginsberg HN, Accili D. Inhibition of Notch uncouples Akt activation from hepatic lipid accumulation by decreasing mTORc1 stability. *Nat Med*. 2013; 19:1054–60.  
<https://doi.org/10.1038/nm.3259> PMID:[23832089](https://pubmed.ncbi.nlm.nih.gov/23832089/)
  40. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017; 45:W98–102.  
<https://doi.org/10.1093/nar/gkx247> PMID:[28407145](https://pubmed.ncbi.nlm.nih.gov/28407145/)
  41. Szász AM, Lánckzy A, Nagy Á, Förster S, Hark K, Green JE, Boussioutas A, Busuttill R, Szabó A, Gyórfy B. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget*. 2016; 7:49322–33.  
<https://doi.org/10.18632/oncotarget.10337>  
 PMID:[27384994](https://pubmed.ncbi.nlm.nih.gov/27384994/)
  42. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015; 19:A68–77.  
<https://doi.org/10.5114/wo.2014.47136>  
 PMID:[25691825](https://pubmed.ncbi.nlm.nih.gov/25691825/)
  43. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013; 6:p11.  
<https://doi.org/10.1126/scisignal.2004088>  
 PMID:[23550210](https://pubmed.ncbi.nlm.nih.gov/23550210/)
  44. Wise LM, Stuart GS, Real NC, Fleming SB, Mercer AA. VEGF Receptor-2 Activation Mediated by VEGF-E Limits Scar Tissue Formation Following Cutaneous Injury. *Adv Wound Care (New Rochelle)*. 2018; 7:283–97.  
<https://doi.org/10.1089/wound.2016.0721>  
 PMID:[30087804](https://pubmed.ncbi.nlm.nih.gov/30087804/)
  45. McLaughlin PJ, Cain JD, Titunick MB, Sassani JW, Zagon IS. Topical Naltrexone Is a Safe and Effective Alternative to Standard Treatment of Diabetic Wounds. *Adv Wound Care (New Rochelle)*. 2017; 6:279–88.  
<https://doi.org/10.1089/wound.2016.0725>  
 PMID:[28894635](https://pubmed.ncbi.nlm.nih.gov/28894635/)
  46. Lei J, Priddy LB, Lim JJ, Masee M, Koob TJ. Identification of Extracellular Matrix Components and Biological Factors in Micronized Dehydrated Human Amnion/Chorion Membrane. *Adv Wound Care (New Rochelle)*. 2017; 6:43–53.  
<https://doi.org/10.1089/wound.2016.0699>  
 PMID:[28224047](https://pubmed.ncbi.nlm.nih.gov/28224047/)
  47. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res*. 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
 PMID:[29092952](https://pubmed.ncbi.nlm.nih.gov/29092952/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. The functions of Notch and genes significant associated with Notch alterations were predicted. (A) Cellular components. (B) Biological processes. (C) Molecular functions. (D) KEGG disease. (E) KEGG pathway.

**A****B**

**Supplementary Figure 2. Cell cycle and Notch signaling pathway regulated by the Notch alteration in GC. The (A) cell cycle and (B) Notch signaling pathway regulated by the Notch alteration in GC (cBioPortal) are shown.**

# Identification of biomarkers related to CD8<sup>+</sup> T cell infiltration with gene co-expression network in clear cell renal cell carcinoma

Jiaying Lin<sup>1,\*</sup>, Meng Yu<sup>2,\*</sup>, Xiao Xu<sup>3</sup>, Yutao Wang<sup>1</sup>, Haotian Xing<sup>1</sup>, Jun An<sup>1</sup>, Jieping Yang<sup>1</sup>, Chaozhi Tang<sup>1</sup>, Dan Sun<sup>1</sup>, Yuyan Zhu<sup>1</sup>

<sup>1</sup>Department of Urology, The First Hospital of China Medical University, Shenyang 110001, Liaoning, China

<sup>2</sup>Department of Reproductive Biology and Transgenic Animal, China Medical University, Shenyang 110001, Liaoning, China

<sup>3</sup>Department of Pediatric Intensive Care Unit, The Shengjing Hospital of China Medical University, Shenyang 110001, Liaoning, China

\*Equal contribution

**Correspondence to:** Yuyan Zhu; email: [yzhu@cmu.edu.cn](mailto:yzhu@cmu.edu.cn)

**Keywords:** clear cell renal cell carcinoma (ccRCC), CIBERSORT, weighted gene co-expression network analysis (WGCNA), CD8<sup>+</sup> T cells, CCL5

**Received:** October 1, 2019

**Accepted:** February 4, 2020

**Published:** February 20, 2020

**Copyright:** Lin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Clear cell renal cell carcinoma (ccRCC) is an extremely common kind of kidney cancer in adults. Immunotherapy and targeted therapy are particularly effective at treating ccRCC. In this study, weighted gene co-expression network analysis and a deconvolution algorithm that quantifies the cellular composition of immune cells were used to analyze ccRCC expression data from the Gene Expression Omnibus database, and identify modules related to CD8<sup>+</sup> T cells. Ten hub genes (*LCK*, *CD2*, *CD3D*, *CD3G*, *IRF1*, *IFNG*, *CCR5*, *CD8A*, *CCL5*, and *CXCL9*) were identified by co-expression network and protein-protein interactions network analysis. Datasets obtained from The Cancer Genome Atlas were analyzed and the data revealed that the hub genes were meaningfully up-regulated in tumor tissues and correlated with promotion of tumor progression. After Kaplan-Meier analysis and Oncomine meta-analysis, *CCL5* was selected as a prognostic biomarker. Finally, the experimental results show that reduced expression of *CCL5* decreased cell proliferation and invasion in the ccRCC cell line. Various analyses were performed and verified, *CCL5* is a potential biomarker and therapeutic target which related to CD8<sup>+</sup> T cell infiltration in ccRCC.

## INTRODUCTION

Renal cell carcinoma (RCC) is one of the most common malignant tumors, ranking seventh among male malignant tumors and tenth among female malignant tumors [1]. RCC accounts for 80% of all kidney cancer, and clear cell renal cell carcinoma is the most common subtype of renal cell carcinoma [2]. Smoking, obesity, and high blood pressure increase the risk of kidney cancer [3]. In recent years, immune checkpoint inhibitors have become the standard for first-line treatment of renal cell carcinoma [4–6]. However, no

specific molecular markers have been for immunotherapy of renal cell carcinoma [2]. Therefore, the exploration of immune-related molecular markers is an important focus of renal cell carcinoma research.

RCC are prone to immune infiltration, and the characteristics of tumor microenvironment strongly alter the response to immunotherapy [7]. CD8<sup>+</sup> T cells contribute to tumor adaptive immunity. Among the immune cells of ccRCC, CD8<sup>+</sup> T cells account for the largest proportion [8]. In most solid tumors, highly infiltrating CD8<sup>+</sup> T cells are beneficial to tumor

treatment [9–11], but high infiltration of CD8<sup>+</sup> T cells in RCC is related to bad prognosis [12]. Many studies have explored the immune-related biomarkers of renal cell carcinoma, but the findings cannot be directly applied to actual clinical work. Previous studies have reported that low expression of *CD40* is associated with poor prognosis in patients with RCC [13]. Co-expression of *PD-1* and *Tim-3* was reported to correlate with poor overall survival [14]. However, their studies included less than 50 samples, so the result should be validated in a larger cohort. *CD19* was identified as a surface marker of B cells and can predict the prognosis of metastatic renal cell carcinoma [15]. However, CD8<sup>+</sup> T cells are more important in tumor adaptive immunity, so it is unclear if *CD19* can guide the immunotherapy of renal cell carcinoma. Therefore, the identification of biomarkers related to CD8<sup>+</sup> T cell infiltration will facilitate the monitoring of RCC immunotherapy response and the exploration of immune infiltration mechanism.

With the rapid development of bioinformatics technology, many tools have been developed to find biomarkers [16]. Weighted gene co-expression network analysis (WGCNA) is an effective tool that can be used to mine related patterns between genes to identify relevant modules and hub genes for cancer [17]. This algorithm has been widely used to find biomarkers at the transcriptional level [18, 19]. Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) is another bioinformatics tool for analysis of gene expression data. This tool quantifies the cellular composition of immune cells using a deconvolution algorithm [20]. This algorithm has been successfully used to approximate the level of immune cell infiltration in various cancers, such as prostate cancer [21] and kidney cancer [8].

To explore the effect of the tumor microenvironment and identify potential biomarkers of ccRCC, WGCNA was performed using ccRCC gene expression data. The T-cell compositions of samples were calculated using the CIBERSORT algorithm. We then identified important modules and hub genes related to CD8<sup>+</sup> T cell infiltration levels, and the immune and clinical features of these genes were verified by database analysis. Prognostic biomarkers were then identified and verified. This is the first utilization of WGCNA to identify CD8<sup>+</sup> T cell-related biomarkers of ccRCC.

## RESULTS

### RNA expression data

The research strategy is presented in Figure 1.

We obtained RNA expression data for 265 ccRCC samples Gene Expression Omnibus (GEO) database. All data for tumor samples in the dataset were obtained, and 4411 genes with Coefficient of variation values greater than 0.1 were selected for additional analysis (Supplementary Table 1).

### Evaluation of tumor-infiltrating immune cells (TIICs)

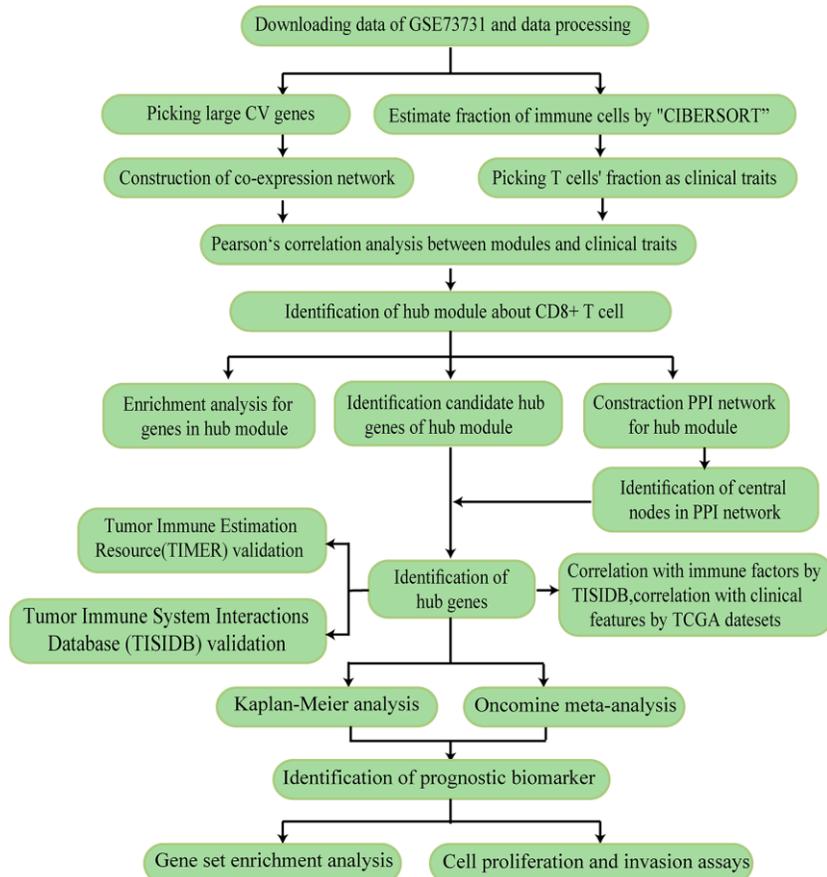
CIBERSORT is an analytical algorithm that analyzes RNA expression data to assess the abundance of different cell subtypes for each sample. The fractions of 22 TIICs were calculated by using the R package “CIBERSORT”. Then, the fractions of seven subtypes of T cells in every sample were selected as trait data of WGCNA (Supplementary Table 2).

### Gene co-expression network of ccRCC

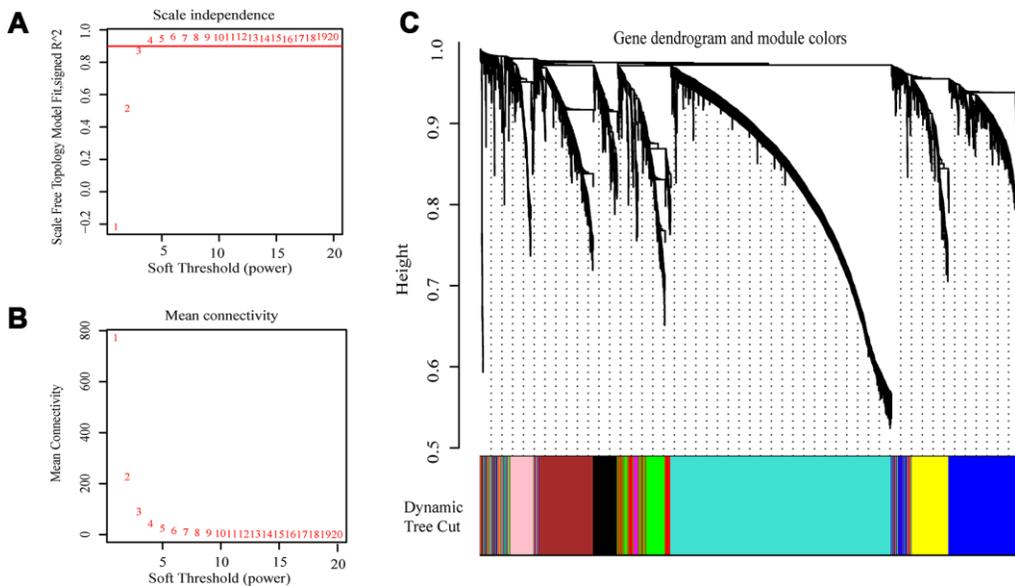
The expression values of the 4411 genes were used to construct a co-expression network using the R package “WGCNA”. We calculated average linkage and Pearson’s correlation values to cluster the samples of GSE73731 (Supplementary Figure 1). To build a scale-free network, we picked  $\beta = 3$  (scale free  $R^2 = 0.8723676$ ) as the soft-thresholding power (Figure 2A, 2B). A hierarchical clustering tree was constructed using dynamic hybrid cutting. Each leaf on the tree represents a single gene, and genes with similar expression data are close together and form a branch of the tree, representing a gene module. Nine modules were generated (Figure 2C).

### Identification of hub modules and enrichment analysis

Among the nine module, the green module was highly correlated to T cells CD8 (CD8<sup>+</sup> T cells) ( $R^2=0.5$ ,  $P=2e-18$ ), T cells CD4 memory activated ( $R^2=0.45$ ,  $P=1e-14$ ), and T cells gamma delta ( $R^2=0.62$ ,  $P=3e-29$ ), and the yellow module showed higher correlation with activated T cells CD4 memory activated ( $R^2=0.5$ ,  $P=4e-18$ ; Figure 3A). The correlation between other modules and T cells was less than 0.5. We were interested specifically in the CD8<sup>+</sup> T cells, so focused on the green module that showed correlation with CD8<sup>+</sup> T cells was identified as a hub module. Genes included in this module were next analyzed using the web tool “Matascape” for pathway and process enrichment analysis. The 20 highest enrichment terms were all immune-related terms (Figure 3B), and the three most highly enriched terms were Lymphocyte Activation, Adaptive Immune Response, and Cytokine-mediated Signaling Pathway (Supplementary Table 3).



**Figure 1. The workflow of the study.**

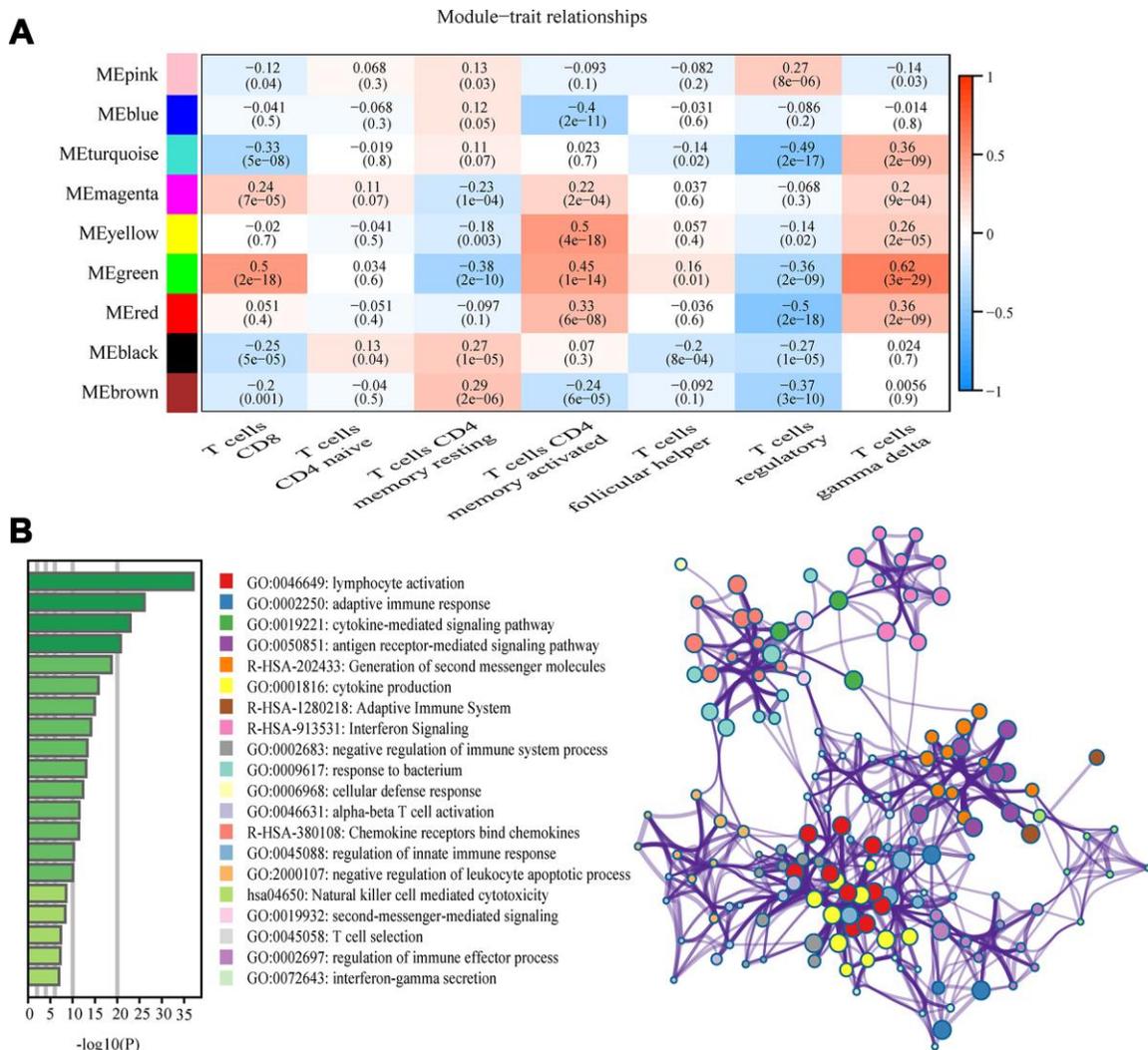


**Figure 2. Selection of the appropriate beta value to construct a hierarchical cluster number. (A)** Analyze the scale-free fit index of the 1-20 soft threshold power ( $\beta$ ). **(B)** Analyze the average connectivity of 1-20 soft threshold power. **(C)** Genes are grouped into various modules by hierarchical clustering, and different colors represent different modules.

## Identification and validation of hub genes

The highly connected genes of the module were investigated as potential key factors related to CD8<sup>+</sup> T cell infiltration level. According to the cut-off standard (Module-Membership > 0.8 and Gene-Significance > 0.5), 30 genes were selected as candidate hub genes (Figure 4B). From the protein-protein interactions (PPI) network, cut-offs of reliability > 0.7 and connectivity > 15 (node/edge) were applied to identify 30 genes as central nodes, and we visualized these results using Cytoscape (Figure 4A). Ten genes were selected in both analyses designated as hub genes (*LCK*, *CD2*, *CD3D*, *CD3G*, *IRF1*, *IFNG*, *CCR5*, *CD8A*, *CCL5* and *CXCL9*) (Figure 4C). To investigate the relationship between these hub genes and CD8<sup>+</sup> T cells, we analyzed the expression data for these genes in the TIMER database.

The results showed positive correlation of the expression values of the 10 genes with the infiltration levels of CD8<sup>+</sup> T cells (correlation coefficient of at least 0.75 for all genes except *CXCL9*) (Figure 5A). As an example, we show a scatter plot of *CCL5* expression and CD8<sup>+</sup> T cell infiltration level in Figure 5B. We next queried the TISIDB database to obtain the Spearman correlation values between abundance of tumor-infiltrating lymphocytes and gene expression. The results show a positive correlation between hub genes and tumor-infiltrating lymphocytes. The correlation values were highest for Activated CD8<sup>+</sup> T cell (Act CD8) and Effector memory CD8<sup>+</sup> T cell (Tem CD8) (Figure 5C). These analyses verified the identified hub genes as strongly associated with the CD8<sup>+</sup> T cell infiltration level and playing significant roles in the immune microenvironment.

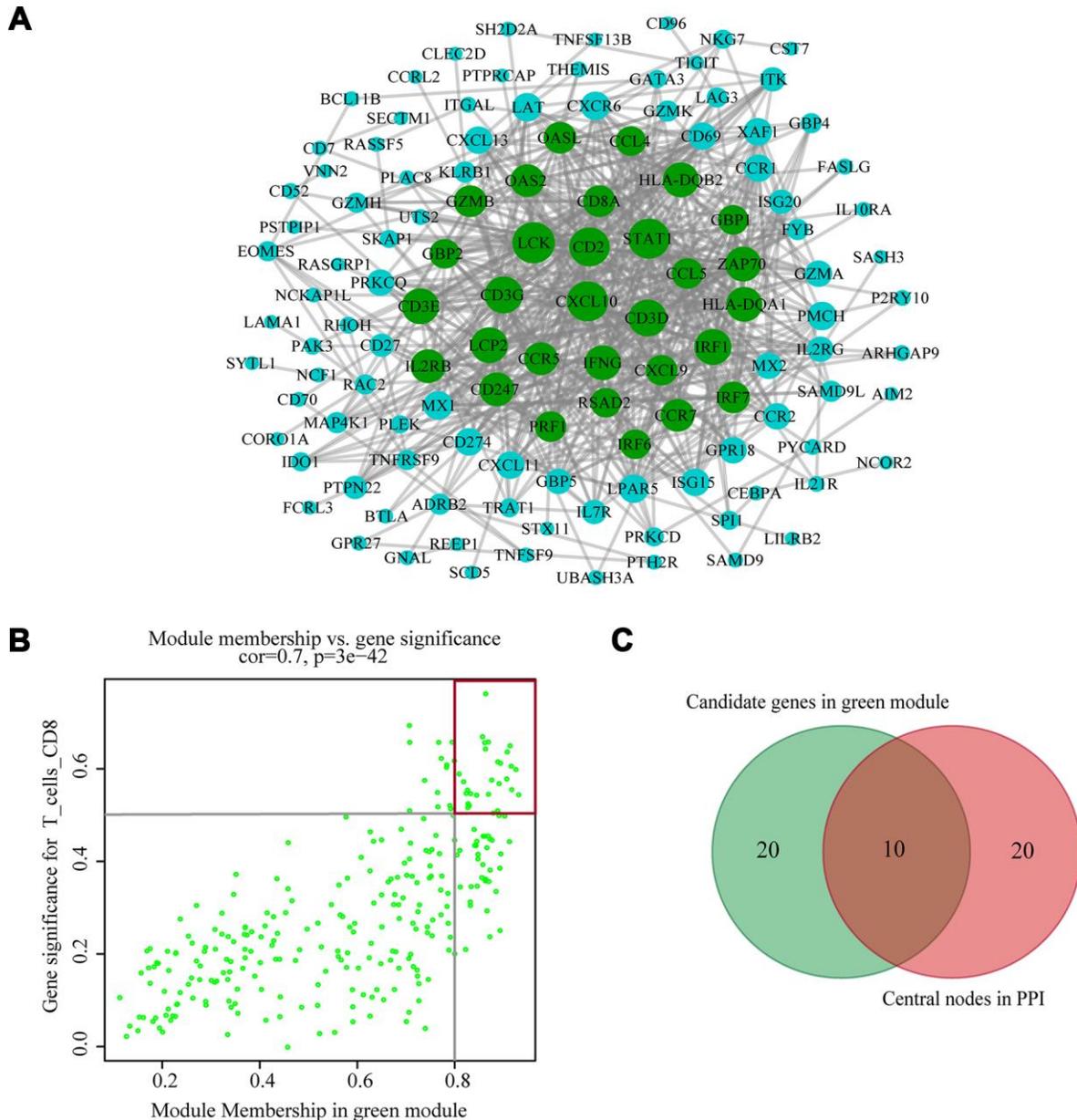


**Figure 3. Key modules and feature notes.** (A) Heatmap shows correlations of module eigengenes with T-cell infiltration. (B) The first 20 enriched terms are shown as a bar chart on the left. The network diagram on the right is constructed with each enrichment term as a node and the similarity of the node as the edge. Nodes with the same cluster ID are the same color.

## Determination of immune and clinical characteristics

We searched for the Spearman correlations of expression of these 10 hub genes with the expression of immune factors in TISIDB database, including immune-inhibitory factors, immune-stimulatory factors, chemokines, and receptors. These correlations are presented as heat maps (Supplementary Figure 2). We identified 38 immune-related factors with average correlations with

the 10 hub genes greater than 0.5. We constructed an immune infiltration interaction network based on the 10 hub genes and the 38 immune-related factors to explore the infiltration mechanism of CD8<sup>+</sup> T cells using the STRING database. The results were imported into Cytoscape for visualization (Figure 5D). We next obtained the expression levels of the 10 genes of ccRCC from the The Cancer Genome Atlas (TCGA). The expression levels of these genes were higher in tumor tissues than in normal tissues ( $P < 0.05$ ) based on



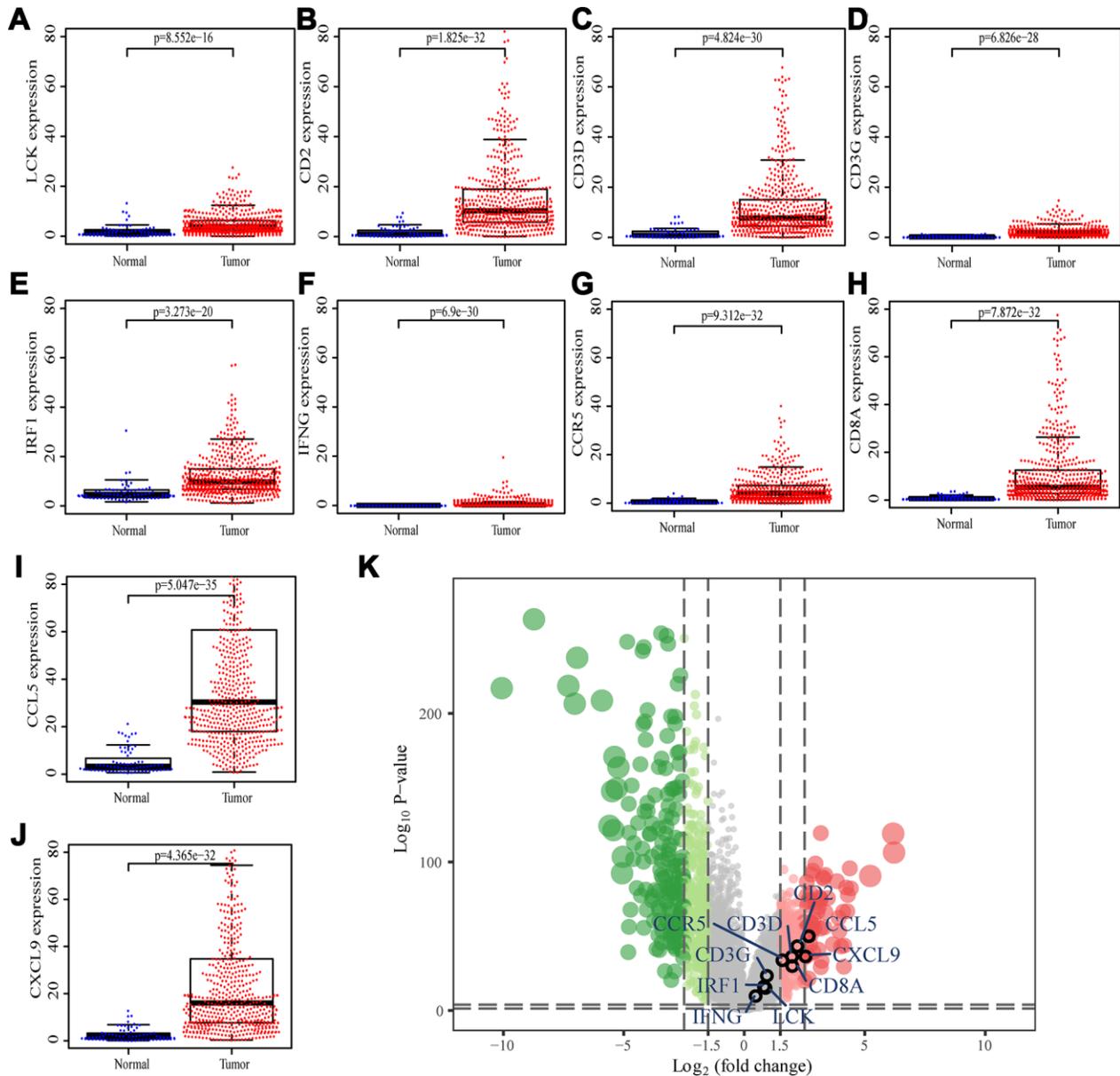
**Figure 4. Identification of hub genes.** (A) PPI network of genes from the green module. The higher the number of connected nodes, the larger the size of the node. The green nodes represent a central node with more than 15 connections. (B) A scatter plot of the genes in the green module. Each green dot represents a gene, and dots within the red box indicate genes of Module Membership > 0.8 and Gene Significance > 0.5. (C) Hub genes were selected based on overlap between PPI and co-expression networks.



## Identification of prognostic biomarker

We analyzed 10 hub genes by Kaplan-Meier analysis, and only the results of *CCL5* and *IFNG* were statistically significant ( $p < 0.05$ ). For these two genes, the survival prognosis of patients with high expression was poor (Figure 8A, 8B). To validate differential expression of 10 genes in tumor and normal tissues, we used OncoPrint to perform a meta-analysis of five analyses using four data sets, all of which included both tumor tissues and normal

controls for 10 genes (Figure 8C). Data for *IFNG* were not included in these data sets, so we obtained meta-analytical results for the other nine genes. *LCK*, *CD2*, *CD3D*, *CCR5*, *CCL5*, and *CXCL9* have Median Rank values less than 1000, *CD3G* is 3007.0, *IRF1* is 1445.0, and *CD8A* is 1577.0. The results show these genes exhibit significant overexpression in tumor tissues which was consistent with the TCGA datasets analyses. Through Kaplan-Meier and OncoPrint meta-analysis, we selected *CCL5* as a prognostic biomarker for further analysis.

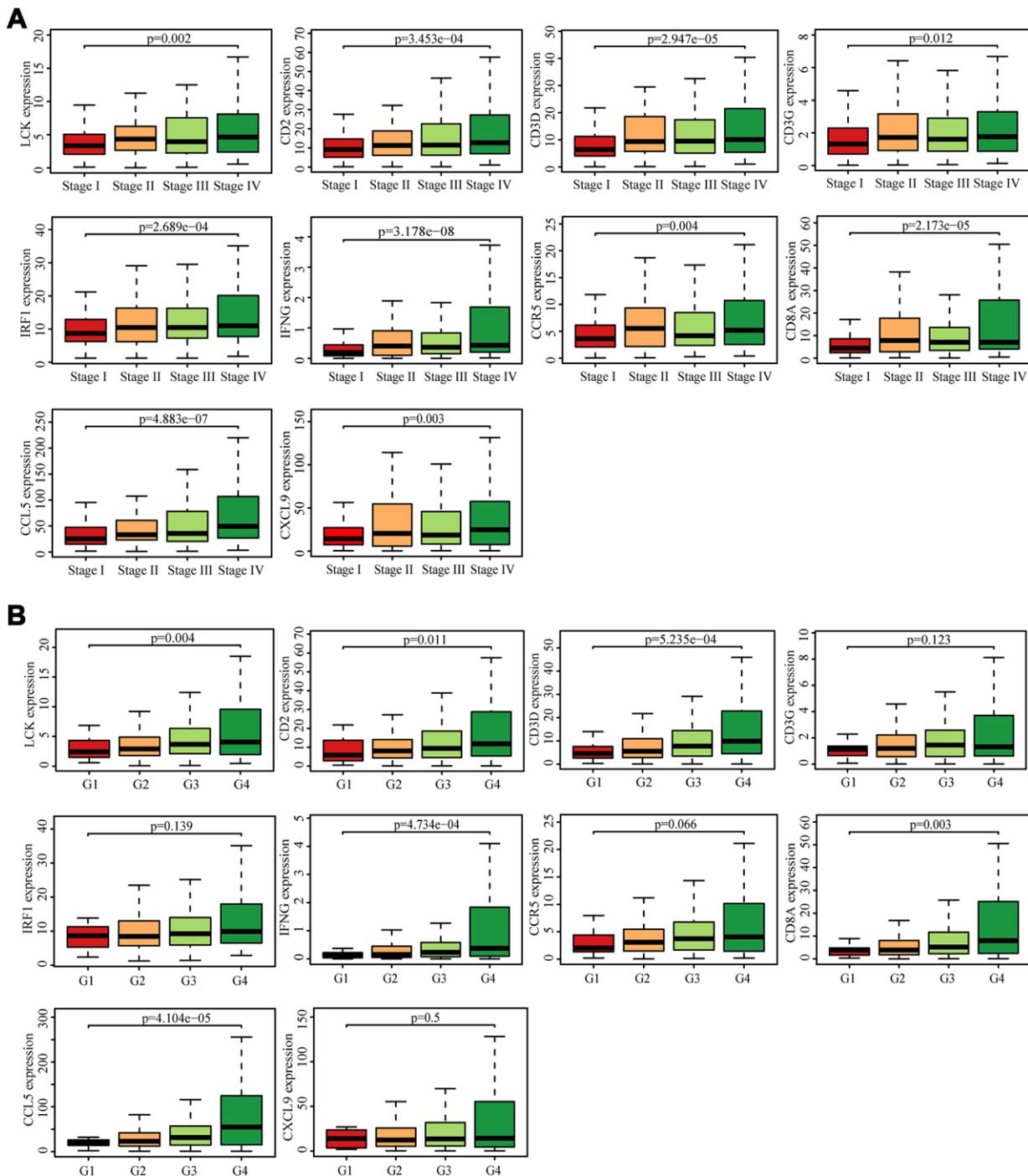


**Figure 6. Differential expression of the hub genes in transcriptional data of TCGA.** (A) *LCK*, blue dots represent normal tissue and red dots represent tumor tissue. The y-axis shows the expression value of the gene. (B) *CD2*. (C) *CD3D*. (D) *CD3G*. (E) *IRF1*. (F) *IFNG*. (G) *CCR5*. (H) *CD8A*. (I) *CCL5*. (J) *CXCL9*. (K) The volcano plot of differentially expressed genes. Red dots indicate overexpression genes, green dots indicate low expression genes, and black circles represent hub genes.

## Gene set enrichment analysis of *CCL5*

According to *CCL5* expression median value, ccRCC samples from TCGA were divided into high expression group and low expression group for pathway gene set enrichment analysis. The enrichment results showed that immune-related pathways were enriched in the high

expression group, with a total of 23 pathways statistically significantly enriched ( $p\text{-value} < 0.05$ ,  $q\text{-value} < 0.05$ ). The three most enriched pathways were “Antigen procession and presentation”, “Cell adhesion molecules cams”, and “Autoimmune thyroid disease” (Figure 9A, 9B, Supplementary Table 5). There were no significantly enriched pathways for the low expression group.



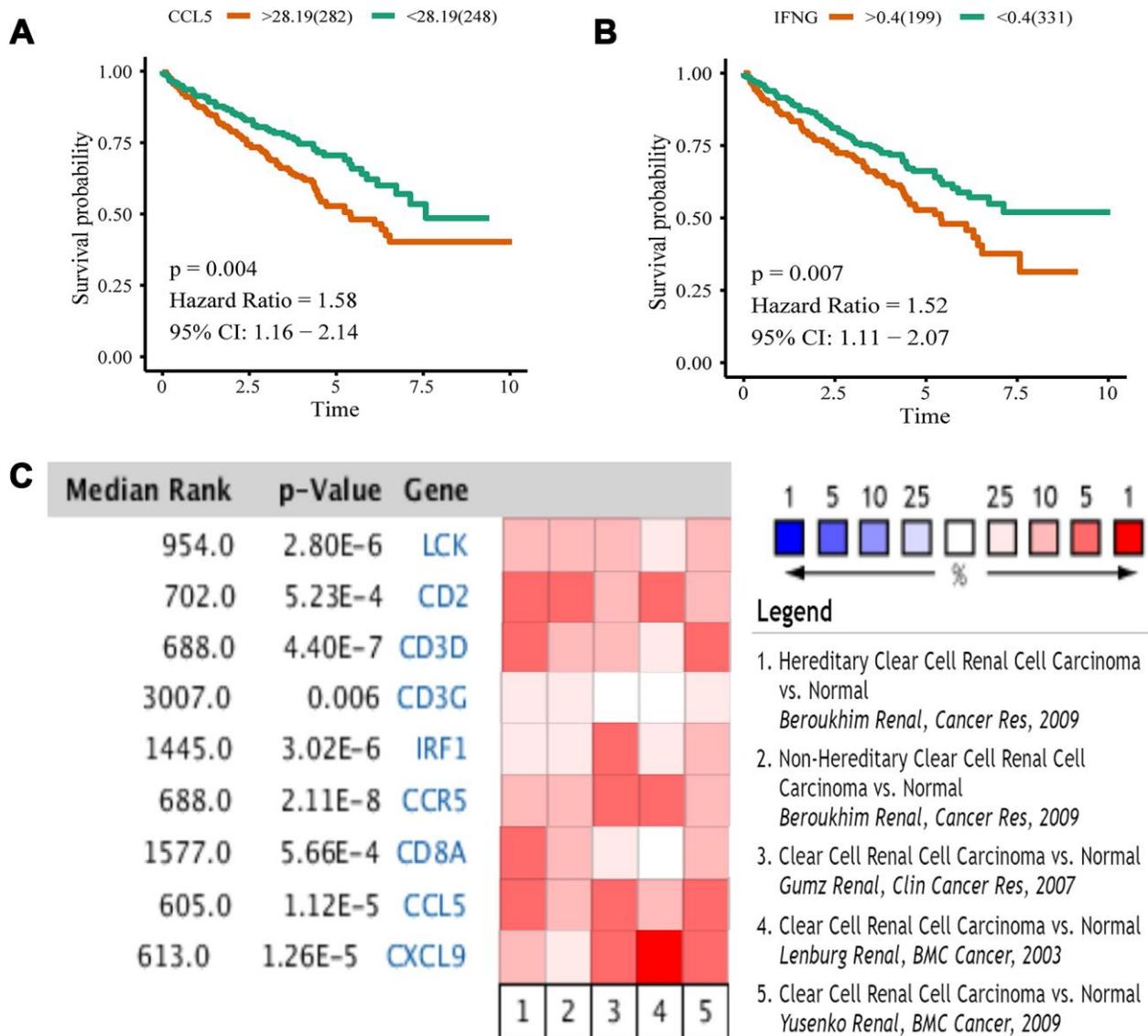
**Figure 7. Analysis of hub genes and clinical indicators in the TCGA dataset. (A)** Box plot of the hub genes for different pathological stages. **(B)** Box plot of the hub genes for different tumor grades.

## Knockdown of *CCL5* significantly inhibited cell proliferation and invasion

Because *CCL5* is overexpressed in ccRCC and is related to poor prognosis, we next conducted functional experiments to explore the potential biological function of *CCL5*. First of all, we used small interfering RNAs to lower expression levels of *CCL5* in the renal cell carcinoma cell line 769-P (Figure 9C) and evaluated the cell proliferation ability by Cell Counting Kit-8 (CCK-8). The results showed decreased cell line proliferation ability after *CCL5* knockdown (Figure 9D). Finally, the invasion ability of the cell line decreased significantly after *CCL5* knockdown (Figure 9E).

## DISCUSSION

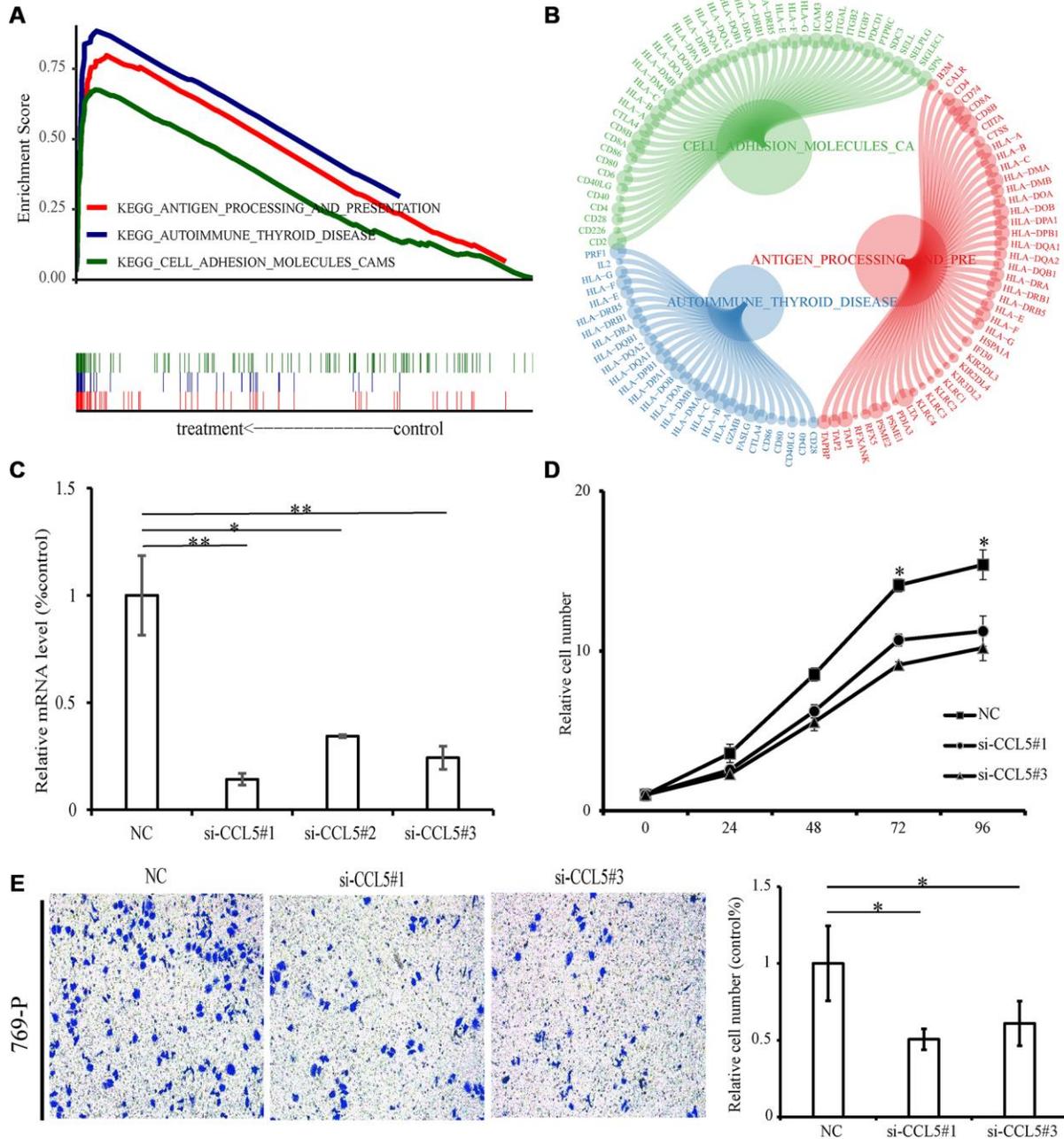
Clear cell renal cell carcinoma is the main histological subtype of RCC, and has relatively poor prognosis [22]. The successful application of immune checkpoint inhibitors in ccRCC has increased interest in exploring the potential targeting of specific immune-related factors for immunotherapy [7]. CD8<sup>+</sup> T cells play a central role in immunotherapy. In this study, we identified 10 hub genes whose expression correlated with CD8<sup>+</sup> T cell infiltration level, suggesting a potential mechanism through which these genes promote the progress of ccRCC. Of the identified 10 genes, *CCL5* was identified as a potential prognostic biomarker and target.



**Figure 8. Kaplan-Meier analysis and OncoPrint meta-analysis.** (A) The overall survival analysis of *CCL5*. (B) The overall survival analysis of *IFNG*. (C) A meta-analysis of gene expression from OncoPrint datasets. Colored squares represent the median of genes (relative to normal tissue) in five analyses. Red represents overexpression, blue represents low expression. This P value gives the average rank analysis.

Significant progress has been made in the study of the molecular basis of ccRCC by using animal tumor models, *in-vitro* cell lines, and clinical samples. However, the complexity of the ccRCC micro-environment requires further analysis and larger datasets. We used gene expression matrix to construct the co-expression network and calculate the infiltration

level of T cells, and correlations were determined to identify the genes most related to CD8<sup>+</sup> T cells. The gene enrichment analysis of the selected module shows that it is a highly immune-related module. The most highly connected genes in the co-expression network and the protein-protein network were considered hub genes (*LCK*, *CD2*, *CD3D*, *CD3G*, *IRF1*, *IFNG*, *CCR5*,



**Figure 9. GSEA and experiment of CCL5.** (A) The above picture shows the enrichment fractional broken line of the three pathways, and the lines in the lower figure correspond to the genes of each pathway. (B) The circle diagram shows three enrichment pathways and the core genes that play a role in the enrichment process. The larger the circle corresponding to each gene, the larger the rank metric score value. (C) The relative mRNA levels of normal control and si-CCL5 of 769-P cells. (D) The results of CCK-8 assay showed decreased proliferation ability of 769-P cells when treated with si-CCL5. (E) The invasion assay results showed decreased invasion ability of 769-P cells treated with si-CCL5.

*CD8A*, *CCL5* and *CXCL9*). Querying the relationship between these 10 genes and immune cells in the TIMER/TISIDB database revealed positively correlated expression of these genes with immune cells, particularly CD8<sup>+</sup> T cells. We used hub genes and related immune factors to construct a CD8<sup>+</sup> T cell infiltration network using the TISIDB and STRING databases as a strategy to explore the immune mechanism of ccRCC. The TCGA datasets were used to observe the differential expression and clinical characteristics of the selected 10 genes. The results showed overexpression of the 10 genes in tumor tissue, indicating potential use as biomarkers. Expression correlated with increase of tumor stage and grade are especially important for potential prognostic factors. These 10 hub genes can explain why the prognosis of ccRCC is poor in the case of highly infiltrated CD8<sup>+</sup> T cells. We performed Kaplan-Meier analysis for the 10 hub genes. Only the results of *CCL5* and *IFNG* were statistically significant, and the prognosis was poor when these genes were highly expressed. The Oncomine database was used to perform a meta-analysis, and the difference between cancer tissue and normal tissue of *CCL5* expression was the most significant. Combined with these two analyses, *CCL5* was selected as the best potential biomarker for the detection and prediction prognosis of renal cell carcinoma.

Chemokine ligand 5 (*CCL5*) belongs to the CC chemokine family, which acts mainly by binding to its corresponding chemokine receptor. Several studies have focused on the effect of *CCL5* on tumors, finding that *CCL5* can significantly promote tumor growth, metastasis [23], angiogenesis [24, 25], and immune escape [26, 27]. *CCL5* is expressed not only in immune cells, but also in tumor cells [28]. The expression of *CCL5* in breast cancer cells promotes the proliferation and invasion of breast cancer cells through an autocrine pathway [29, 30]. *CCL5* also promotes the progression of melanoma [31] and pleomorphic glioma [32]. However, there has been no experimental study of the effect of *CCL5* on the proliferation and invasion of ccRCC cells. The experimental results show that si-*CCL5* can decrease the proliferation and invasion ability of ccRCC cells, indicating that *CCL5* could be utilized as a therapeutic target.

In short, this study is the first attempt to use WGCNA and CIBERSORT algorithms to identify potential CD8<sup>+</sup> T cell related biomarkers of ccRCC. Ten hub genes were identified which were overexpression in tumor and promoting tumor progression. Through multiple verification of bioinformatics and experiments, *CCL5* was identified as a potential biomarker and target for clear cell renal cell carcinoma immunotherapy. However, this study has certain limitations. Additional sample data are needed to verify these results and the specific mechanism of *CCL5* in ccRCC requires further investigation.

## MATERIALS AND METHODS

### Gene expression data and processing

We downloaded ccRCC RNA expression data from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), which contains data related to 265 samples [33]. The dataset of GSE73731 was obtained using the platform Affymetrix Human Genome U133 Plus 2.0 Array (HG U133 Plus 2.0). We used the R package “limma” [34] to normalize the RNA-sequencing data. Small variation of gene expression data often represents noise, so we used Coefficient of Variation values to select the most variant genes, which were then used to construct the network.

### Evaluation of tumor-infiltrating immune cells

In this study, we used the R package “CIBERSORT” to estimate the fraction of immune cells of GSE73731 samples. Specifically, the CIBERSORT algorithm was used to calculate the fractions of the 22 types of TIICs [20]. CIBERSORT is considered better to previous deconvolution methods for analysis of unknown mixture content and noise. This algorithm can be used to statistically estimate the relative proportions of cell sub-populations from complex tissue expression profiles, making it a useful tool to estimate the abundances of special cells in mixed tissue.

### Co-expression network construction

Expression values of 4411 genes were used to construct a weight co-expression network using the R package “WGCNA” [17]. First, based on the Pearson’s correlation value between paired genes, the expression levels of individual transcripts were converted into a similarity matrix. Next, the similarity matrix was transformed to an adjacency matrix, as calculated by  $amn = |cmn|^\beta$  ( $cmn$  = Pearson’s correlation between paired genes;  $amn$  = adjacency between paired genes). Parameter  $\beta$  can improve strong correlations and decrease weak correlations between genes. The adjacency matrix was then converted into a topological overlap matrix when the power of  $\beta = 3$ . To categorize genes with similar expression patterns into different modules, we applied a dynamic hybrid cutting method, using a bottom-up algorithm with a module minimum size cutoff of 30.

### Construct module trait relationships

Module eigengenes were used to perform component analysis of each module. We calculated the correlation between module eigengenes and the infiltration level of T cells to determine the significance of modules by

Pearson test. An individual module was considered significantly correlated with T cells when  $p < 0.05$ . We selected the interest T cell subtype and module with the highest correlation coefficient and defined that as a hub module.

### Pathway and process enrichment analysis

To determine the function of genes in the identified hub module, we used the web tool “Metascape” (<http://metascape.org>) for pathway and process enrichment analysis [35]. The tool displays the first 20 enriched terms as a bar graph. In order to further explore the relationship between terms, terms with similarity greater than 0.3 are connected by edges and presented as a network graph.

### Identification of hub genes

We selected candidate hub genes based on the modular connectivity and clinical traits relationship of each gene in the hub module. Module connectivity is defined as the absolute value of the Pearson’s correlation between genes (Module Membership). Clinical trait relationship is defined as the absolute value of the Pearson’s correlation between each gene and the trait (Gene Significance). We set the Module-Membership  $> 0.8$  and the Gene-Significance  $> 0.5$  for candidate hub genes. Meanwhile we selected all genes in the hub module and used the Search Tool for the Retrieval of Interacting Genes (STRING; <https://string-db.org/>) database to construct PPI network and looked for central nodes [36]. Genes with node connectivity  $> 15$  were considered central nodes. We used Cytoscape to present the network (<https://cytoscape.org/>) [37]. We did Venn analysis to compare candidate hub genes and central nodes in the PPI network using the online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

### Validation of hub genes

We used two immune-associated databases that are based on TCGA to validate these hub genes, as described below. First, we obtained the content of CD8<sup>+</sup> T cells in each sample of ccRCC based on data in the Tumor Immune Estimation Resource (TIMER; <https://cistrome.shinyapps.io/timer/>) [38]. Spearman correlations between the infiltration level of CD8<sup>+</sup> T cells and the expression of hub genes were calculated, and the results were compared using the R package “ggstatsplot”. Second, the Tumor Immune System Interactions Database (TISIDB; <http://cis.hku.hk/TISIDB>) [39] was then searched to determine Spearman correlations between the hub genes and TIICs. We present these results as a heat map constructed using the R package “pheatmap”.

### Immune and clinical characteristic identification

Spearman correlations between hub genes and different immune factors were obtained from the TISIDB database, which includes immune-inhibitory and immune-stimulatory factors, chemokines, and receptors. We then constructed a heat map by using the R package “heatmap”. Immune factors related to hub genes with average correlation coefficients greater than 0.5 were picked to construct a network using STRING and Cytoscape [37]. To explore the clinical characteristic of hub genes, RNA expression data and clinical data of ccRCC were obtained from The Cancer Genome Atlas (TCGA; <https://cancergenome.nih.gov/>). The R packages “limma” and “beeswarm” were used to construct a scatter differential diagram of the data. The statistical significance of differences in expression between normal and tumor samples was analyzed using the Wilcoxon signed-rank test. The differences of all coding genes were analyzed by R package “limma”, and the volcano map was drawn using the R package “ggplot2”. Finally, a boxplot was constructed to display the relationship between genes and clinical features, and the statistical significances were analyzed by the Kruskal-Wallis test.

### Kaplan-Meier analysis and Oncomine meta-analysis

We found best separation through the R package “survminer” to divide the patients into high and low expression groups, such grouping minimizes the p value of the survival curve. Then took Kaplan-Meier analysis for the groups by R package “survival”. To validate the expression patterns of the hub genes, four independent microarray datasets were used from the Oncomine Cancer Microarray database (Oncomine, <https://www.oncomine.org>) to perform meta-analysis [40–43].

### Gene set enrichment analysis (GSEA)

GSEA is a computational method used to determine whether a set of basically defined gene sets exhibits statistically significant differences between two biological states [44]. According to the median value of gene expression, the samples were divided into two groups, and “c2.cp.kegg.v7.0.symbols” gene set enrichment analysis was carried out, with p-value  $< 0.05$  and q-value  $< 0.05$  as indicative of statistical significance. The enrichment pathway was visualized using the R packages “ggplot2” and “clusterProfiler”. In the circle diagram, we only show the core genes of the enrichment process, and the size of the dots represent the Rank metric score, which indicates the amount of sequencing value of the gene.

## Cell culture and siRNA-PTEN construction

The 769-P cells were cultivated in DMEM with 10% of fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin. The cells were cultured at 37 °C, with 5% CO<sub>2</sub>. The CCL5-310 siRNA sequences were as follows: 5'-GCUGAACAAAGGGAAGCUUTT-3' 5'-AAGCUU GCCCUUGUUCAGCTT-3'. The CCL5-286 siRNA sequences were: 5'-GCAGGAUUUCCUGUAUG ACTT-3' and 5'-GUCAUACAGGAAAUCCUGCTT-3'. The CCL5-240 siRNA sequences were: 5'-UC GUCCACAGGUCAAGGAUTT-3' 5'-AUCCUUGAC CUGUGGACGATT-3'

## RNA extraction and quantitative RT-PCR

RNA samples were extracted from cells with the FastPure Cell/Tissue Total RNA Isolation kit (Vazyme Biotech) and they were reverse-transcribed by HiScript III RT SuperMix for qPCR (Vazyme Biotech). To determine the relative transcript level, PCR was quantified in real-time using a LifeECO PCR machine (BIOER Technology Co, Ltd). SYBR Green was used as the fluorophore. The CCL5 primers sequences were as follows: 5'-CTC ATTGCTACTACTGCCCTCTGCGCTCCTGC-3' and 5'GCTCATCTCCAAAGAGTTGATGTACTC -3'. The PCR parameters were: 95 °C 5min, followed by 50 cycles of 30 s and 1 min at 60 °C. Each sample was measured in three independent reactions. The threshold values of each sample/primer were determined, and the average error and standard error were calculated. Melting curve analysis was performed, and the mRNA expression levels were normalized against that of β-actin.

## Cell proliferation analysis and invasion experiment

769-P cells were cultivated in five 96-well cell culture plates (1500 cells/well) respectively for 5 days. A volume of 10 µl of CCK-8 solution was added to each well in a plate at an interval of 24 h. The cells were cultured for 3 h in 37 °C, and then a photometry test was performed at wavelength of 450 nm. The 769-P cells were then transferred into the top layer holes in a basement membrane plate with wells containing the medium without serum. The lower cavity was filled with 12% fetal calf serum medium to allow chemical attraction. The medium was removed after 12-hour incubation and the implant in the upper cavity was discarded. Cells invading the lower cavity were fixed with 700 ml of 4% PFA and stained with crystal for 1 h to visualize. The cells in the lower cavity were counted and normalized to the number in control conditions to measure the relative invasion capacity.

## Abbreviations

ccRCC: Clear cell renal cell carcinoma; WGCNA: Weighted gene co-expression network analysis; CIBERSORT: Cell-type identification by estimating relative subsets of RNA transcripts; TIICs: Tumor-infiltrating immune cells; PPI: Protein-protein interactions. GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; GSEA: Gene set enrichment analysis.

## ACKNOWLEDGMENTS

We want to thank TCGA, GEO, TIMER, and TISDIB for free use.

## CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

## FUNDING

This work was supported in part by the National Natural Science Foundation of China (grants 81672523, 81472404, 81472403, 81272834, and 31000572), the 2018 Support Plan for innovative talents in Colleges and Universities of Liaoning Province, the 2018 “million talents Project” funded Project of Liaoning Province, and the 2019 Key R & D projects of Shenyang.

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016; 66:7–30. <https://doi.org/10.3322/caac.21332> PMID:[26742998](https://pubmed.ncbi.nlm.nih.gov/26742998/)
2. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, Grünwald V, Gillessen S, Horwich A, and ESMO Guidelines Committee. Electronic address: [clinicalguidelines@esmo.org](mailto:clinicalguidelines@esmo.org). Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2019; 30:706–20. <https://doi.org/10.1093/annonc/mdz056> PMID:[31987338](https://pubmed.ncbi.nlm.nih.gov/31987338/)
3. Tahbaz R, Schmid M, Merseburger AS. Prevention of kidney cancer incidence and recurrence: lifestyle, medication and nutrition. *Curr Opin Urol.* 2018; 28:62–79. <https://doi.org/10.1097/MOU.0000000000000454> PMID:[29059103](https://pubmed.ncbi.nlm.nih.gov/29059103/)
4. Considine B, Hurwitz ME. Current Status and Future Directions of Immunotherapy in Renal Cell Carcinoma. *Curr Oncol Rep.* 2019; 21:34.

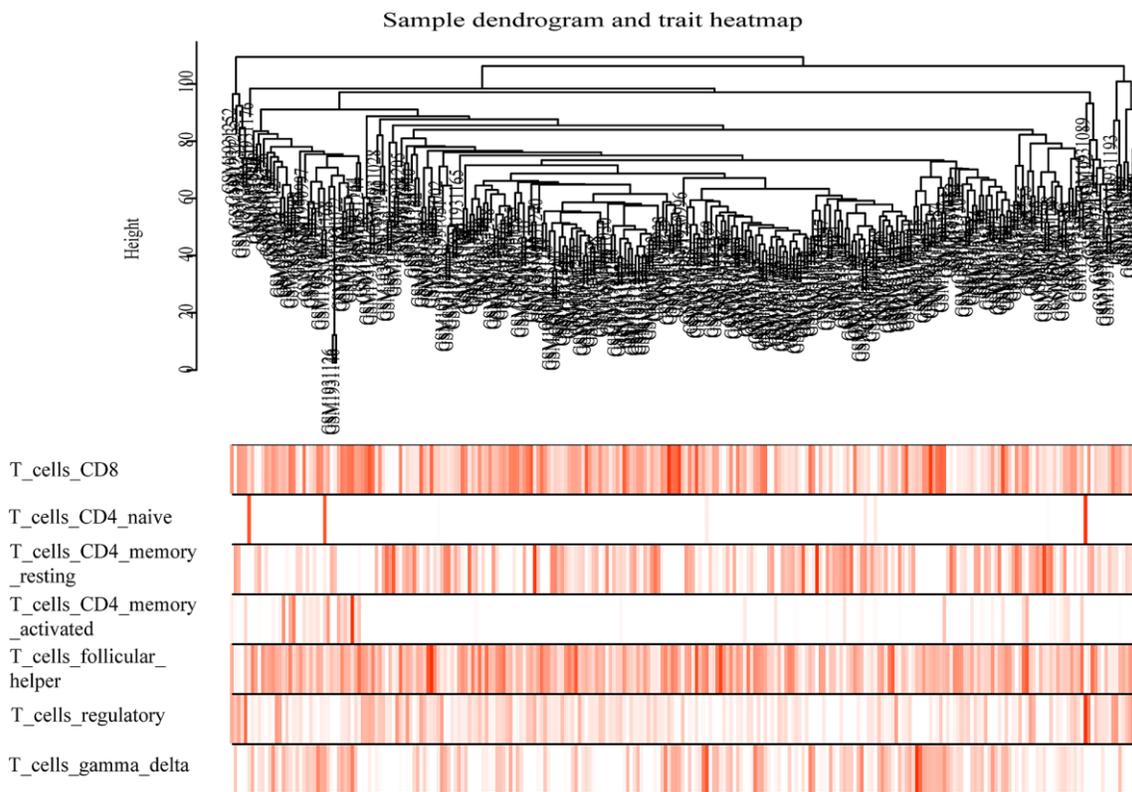
- <https://doi.org/10.1007/s11912-019-0779-1>  
PMID:[30848378](https://pubmed.ncbi.nlm.nih.gov/30848378/)
5. Motzer RJ, Rini BI, McDermott DF, Arén Frontera O, Hammers HJ, Carducci MA, Salman P, Escudier B, Beuselinck B, Amin A, Porta C, George S, Neiman V, et al, and CheckMate 214 investigators. Nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial. *Lancet Oncol.* 2019; 20:1370–85.  
[https://doi.org/10.1016/S1470-2045\(19\)30413-9](https://doi.org/10.1016/S1470-2045(19)30413-9)  
PMID:[31427204](https://pubmed.ncbi.nlm.nih.gov/31427204/)
  6. Grimm MO, Bex A, De Santis M, Ljungberg B, Catto JW, Rouprêt M, Hussain SA, Bellmunt J, Powles T, Wirth M, Van Poppel H. Safe Use of Immune Checkpoint Inhibitors in the Multidisciplinary Management of Urological Cancer: The European Association of Urology Position in 2019. *Eur Urol.* 2019; 76:368–80.  
<https://doi.org/10.1016/j.eururo.2019.05.041>  
PMID:[31235192](https://pubmed.ncbi.nlm.nih.gov/31235192/)
  7. Vuong L, Kotecha RR, Voss MH, Hakimi AA. Tumor Microenvironment Dynamics in Clear-Cell Renal Cell Carcinoma. *Cancer Discov.* 2019; 9:1349–57.  
<https://doi.org/10.1158/2159-8290.CD-19-0499>  
PMID:[31527133](https://pubmed.ncbi.nlm.nih.gov/31527133/)
  8. Zhang S, Zhang E, Long J, Hu Z, Peng J, Liu L, Tang F, Li L, Ouyang Y, Zeng Z. Immune infiltration in renal cell carcinoma. *Cancer Sci.* 2019; 110:1564–72.  
<https://doi.org/10.1111/cas.13996>  
PMID:[30861269](https://pubmed.ncbi.nlm.nih.gov/30861269/)
  9. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoué F, Bruneval P, Cugnenc PH, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* 2006; 313:1960–64.  
<https://doi.org/10.1126/science.1129139>  
PMID:[17008531](https://pubmed.ncbi.nlm.nih.gov/17008531/)
  10. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol.* 2011; 29:1949–55.  
<https://doi.org/10.1200/JCO.2010.30.5037>  
PMID:[21483002](https://pubmed.ncbi.nlm.nih.gov/21483002/)
  11. Sharma P, Shen Y, Wen S, Yamada S, Jungbluth AA, Gnjatic S, Bajorin DF, Reuter VE, Herr H, Old LJ, Sato E. CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc Natl Acad Sci USA.* 2007; 104:3967–72.  
<https://doi.org/10.1073/pnas.0611618104>  
PMID:[17360461](https://pubmed.ncbi.nlm.nih.gov/17360461/)
  12. Drake CG, Stein MN. The Immunobiology of Kidney Cancer. *J Clin Oncol.* 2018. [Epub ahead of print]  
<https://doi.org/10.1200/JCO.2018.79.2648>  
PMID:[30372396](https://pubmed.ncbi.nlm.nih.gov/30372396/)
  13. Weiss JM, Gregory Alvord W, Quiñones OA, Stauffer JK, Wiltrout RH. CD40 expression in renal cell carcinoma is associated with tumor apoptosis, CD8(+) T cell frequency and patient survival. *Hum Immunol.* 2014; 75:614–20.  
<https://doi.org/10.1016/j.humimm.2014.04.018>  
PMID:[24801648](https://pubmed.ncbi.nlm.nih.gov/24801648/)
  14. Granier C, Dariane C, Combe P, Verkarre V, Urien S, Badoual C, Roussel H, Mandavit M, Ravel P, Sibony M, Biard L, Radulescu C, Vinatier E, et al. Tim-3 Expression on Tumor-Infiltrating PD-1<sup>+</sup>CD8<sup>+</sup> T Cells Correlates with Poor Clinical Outcome in Renal Cell Carcinoma. *Cancer Res.* 2017; 77:1075–82.  
<https://doi.org/10.1158/0008-5472.CAN-16-0274>  
PMID:[27872087](https://pubmed.ncbi.nlm.nih.gov/27872087/)
  15. Lin Z, Liu L, Xia Y, Chen X, Xiong Y, Qu Y, Wang J, Bai Q, Guo J, Xu J. Tumor infiltrating CD19<sup>+</sup> B lymphocytes predict prognostic and therapeutic benefits in metastatic renal cell carcinoma patients treated with tyrosine kinase inhibitors. *Oncoimmunology.* 2018; 7:e1477461.  
<https://doi.org/10.1080/2162402X.2018.1477461>  
PMID:[30288343](https://pubmed.ncbi.nlm.nih.gov/30288343/)
  16. Lin T, Gu J, Qu K, Zhang X, Ma X, Miao R, Xiang X, Fu Y, Niu W, She J, Liu C. A new risk score based on twelve hepatocellular carcinoma-specific gene expression can predict the patients' prognosis. *Aging (Albany NY).* 2018; 10:2480–97.  
<https://doi.org/10.18632/aging.101563>  
PMID:[30243023](https://pubmed.ncbi.nlm.nih.gov/30243023/)
  17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559.  
<https://doi.org/10.1186/1471-2105-9-559>  
PMID:[19114008](https://pubmed.ncbi.nlm.nih.gov/19114008/)
  18. Niemira M, Collin F, Szalkowska A, Bielska A, Chwialkowska K, Reszec J, Niklinski J, Kwasniewski M, Kretowski A. Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (WGCNA). *Cancers (Basel).* 2019; 12:E37.  
<https://doi.org/10.3390/cancers12010037>  
PMID:[31877723](https://pubmed.ncbi.nlm.nih.gov/31877723/)
  19. Liu H, Sun Y, Tian H, Xiao X, Zhang J, Wang Y, Yu F. Characterization of long non-coding RNA and messenger RNA profiles in laryngeal cancer by weighted gene co-expression network analysis. *Aging*

- (Albany NY). 2019; 11:10074–99.  
<https://doi.org/10.18632/aging.102419>  
PMID:[31739287](https://pubmed.ncbi.nlm.nih.gov/31739287/)
20. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12:453–57.  
<https://doi.org/10.1038/nmeth.3337> PMID:[25822800](https://pubmed.ncbi.nlm.nih.gov/25822800/)
  21. Zhao SG, Lehrer J, Chang SL, Das R, Erho N, Liu Y, Sjöström M, Den RB, Freedland SJ, Klein EA, Karnes RJ, Schaeffer EM, Xu M, et al. The Immune Landscape of Prostate Cancer and Nomination of PD-L2 as a Potential Therapeutic Target. *J Natl Cancer Inst*. 2019; 111:301–10.  
<https://doi.org/10.1093/jnci/djy141> PMID:[30321406](https://pubmed.ncbi.nlm.nih.gov/30321406/)
  22. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng DY, Larkin J, Ficarra V. Renal cell carcinoma. *Nat Rev Dis Primers*. 2017; 3:17009.  
<https://doi.org/10.1038/nrdp.2017.9> PMID:[28276433](https://pubmed.ncbi.nlm.nih.gov/28276433/)
  23. Lin S, Wan S, Sun L, Hu J, Fang D, Zhao R, Yuan S, Zhang L. Chemokine C-C motif receptor 5 and C-C motif ligand 5 promote cancer cell migration under hypoxia. *Cancer Sci*. 2012; 103:904–12.  
<https://doi.org/10.1111/j.1349-7006.2012.02259.x>  
PMID:[22380870](https://pubmed.ncbi.nlm.nih.gov/22380870/)
  24. Spring H, Schüler T, Arnold B, Hämmerling GJ, Ganss R. Chemokines direct endothelial progenitors into tumor neovessels. *Proc Natl Acad Sci USA*. 2005; 102:18111–16.  
<https://doi.org/10.1073/pnas.0507158102>  
PMID:[16326806](https://pubmed.ncbi.nlm.nih.gov/16326806/)
  25. Tang S, Xiang T, Huang S, Zhou J, Wang Z, Xie R, Long H, Zhu B. Ovarian cancer stem-like cells differentiate into endothelial cells and participate in tumor angiogenesis through autocrine CCL5 signaling. *Cancer Lett*. 2016; 376:137–47.  
<https://doi.org/10.1016/j.canlet.2016.03.034>  
PMID:[27033454](https://pubmed.ncbi.nlm.nih.gov/27033454/)
  26. Kershaw MH, Westwood JA, Darcy PK. Gene-engineered T cells for cancer therapy. *Nat Rev Cancer*. 2013; 13:525–41.  
<https://doi.org/10.1038/nrc3565>  
PMID:[23880905](https://pubmed.ncbi.nlm.nih.gov/23880905/)
  27. Frame MC, Patel H, Serrels B, Lietha D, Eck MJ. The FERM domain: organizing the structure and function of FAK. *Nat Rev Mol Cell Biol*. 2010; 11:802–14.  
<https://doi.org/10.1038/nrm2996> PMID:[20966971](https://pubmed.ncbi.nlm.nih.gov/20966971/)
  28. Levy JA. The unexpected pleiotropic activities of RANTES. *J Immunol*. 2009; 182:3945–6.  
<https://doi.org/10.4049/jimmunol.0990015>  
PMID:[19299688](https://pubmed.ncbi.nlm.nih.gov/19299688/)
  29. Jiao X, Katiyar S, Willmarth NE, Liu M, Ma X, Flomenberg N, Lisanti MP, Pestell RG. c-Jun induces mammary epithelial cellular invasion and breast cancer stem cell expansion. *J Biol Chem*. 2010; 285:8218–26.  
<https://doi.org/10.1074/jbc.M110.100792>  
PMID:[20053993](https://pubmed.ncbi.nlm.nih.gov/20053993/)
  30. Azenshtein E, Luboshits G, Shina S, Neumark E, Shahbazian D, Weil M, Wigler N, Keydar I, Ben-Baruch A. The CC chemokine RANTES in breast carcinoma progression: regulation of expression and potential mechanisms of promalignant activity. *Cancer Res*. 2002; 62:1093–102.  
<https://doi.org/10.1186/bcr377> PMID:[11861388](https://pubmed.ncbi.nlm.nih.gov/11861388/)
  31. Mrowietz U, Schwenk U, Maune S, Bartels J, Küpper M, Fichtner I, Schröder JM, Schadendorf D. The chemokine RANTES is secreted by human melanoma cells and is associated with enhanced tumour formation in nude mice. *Br J Cancer*. 1999; 79:1025–31.  
<https://doi.org/10.1038/sj.bjc.6690164>  
PMID:[10098731](https://pubmed.ncbi.nlm.nih.gov/10098731/)
  32. Kranjc MK, Novak M, Pestell RG, Lah TT. Cytokine CCL5 and receptor CCR5 axis in glioblastoma multiforme. *Radiol Oncol*. 2019; 53:397–406.  
<https://doi.org/10.2478/raon-2019-0057>  
PMID:[31747383](https://pubmed.ncbi.nlm.nih.gov/31747383/)
  33. Wei X, Choudhury Y, Lim WK, Anema J, Kahnoski RJ, Lane B, Ludlow J, Takahashi M, Kanayama HO, Beldegrun A, Kim HL, Rogers C, Nicol D, et al. Recognizing the Continuous Nature of Expression Heterogeneity and Clinical Outcomes in Clear Cell Renal Cell Carcinoma. *Sci Rep*. 2017; 7:7342.  
<https://doi.org/10.1038/s41598-017-07191-y>  
PMID:[28779136](https://pubmed.ncbi.nlm.nih.gov/28779136/)
  34. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47.  
<https://doi.org/10.1093/nar/gkv007> PMID:[25605792](https://pubmed.ncbi.nlm.nih.gov/25605792/)
  35. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019; 10:1523.  
<https://doi.org/10.1038/s41467-019-09234-6>  
PMID:[30944313](https://pubmed.ncbi.nlm.nih.gov/30944313/)
  36. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019; 47:D607–13.  
<https://doi.org/10.1093/nar/gky1131> PMID:[30476243](https://pubmed.ncbi.nlm.nih.gov/30476243/)

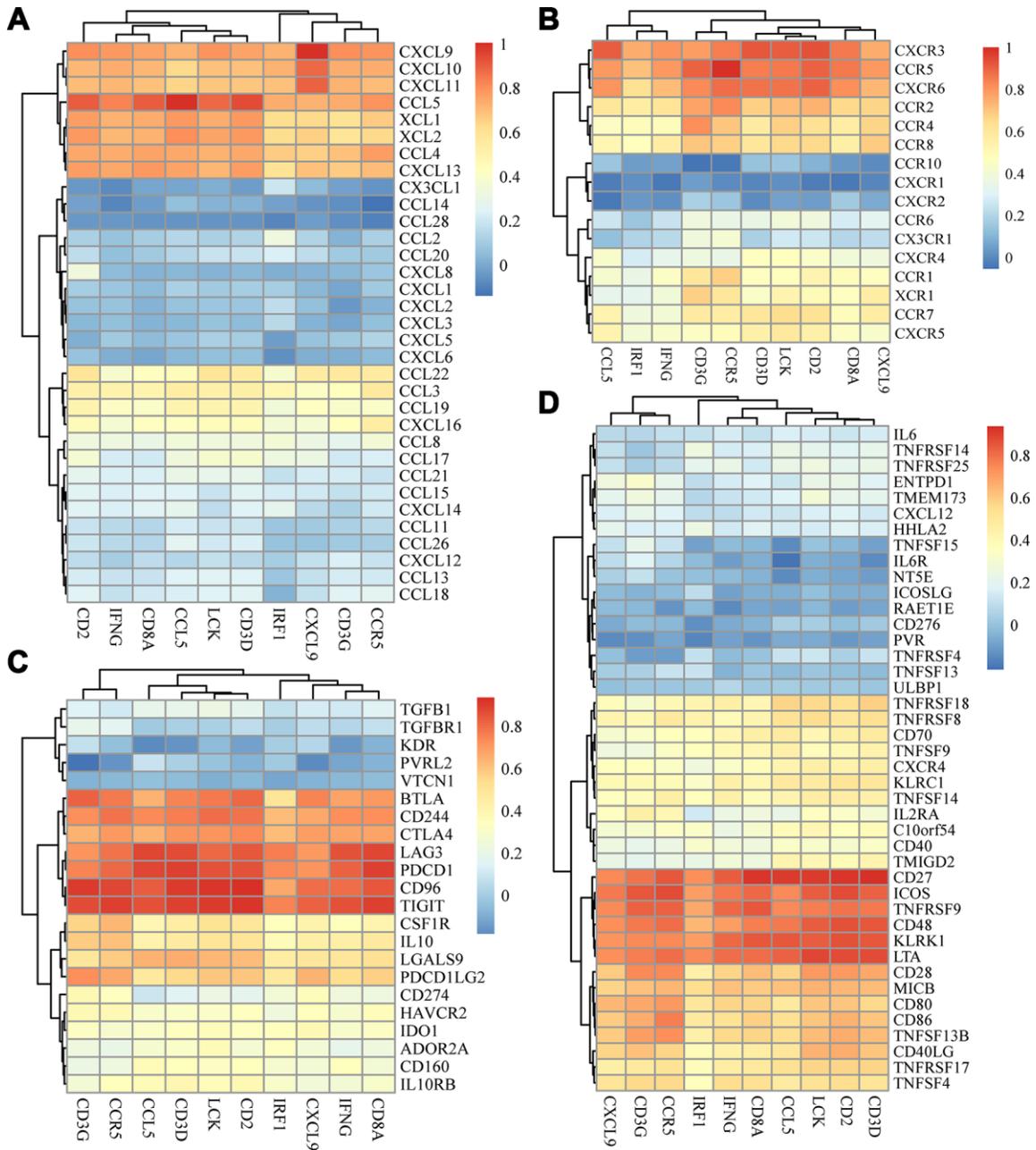
37. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–504. <https://doi.org/10.1101/gr.1239303> PMID:14597658
38. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017; 77:e108–10. <https://doi.org/10.1158/0008-5472.CAN-17-0307> PMID:29092952
39. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SS, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics.* 2019; 35:4200–02. <https://doi.org/10.1093/bioinformatics/btz210> PMID:30903160
40. Lenburg ME, Liou LS, Gerry NP, Frampton GM, Cohen HT, Christman MF. Previously unidentified changes in renal cell carcinoma gene expression identified by parametric analysis of microarray data. *BMC Cancer.* 2003; 3:31. <https://doi.org/10.1186/1471-2407-3-31> PMID:14641932
41. Yusenko MV, Kuiper RP, Boethe T, Ljungberg B, van Kessel AG, Kovacs G. High-resolution DNA copy number and gene expression analyses distinguish chromophobe renal cell carcinomas and renal oncocytomas. *BMC Cancer.* 2009; 9:152. <https://doi.org/10.1186/1471-2407-9-152> PMID:19445733
42. Gumz ML, Zou H, Kreinest PA, Childs AC, Belmonte LS, LeGrand SN, Wu KJ, Luxon BA, Sinha M, Parker AS, Sun LZ, Ahlquist DA, Wood CG, Copland JA. Secreted frizzled-related protein 1 loss contributes to tumor phenotype of clear cell renal cell carcinoma. *Clin Cancer Res.* 2007; 13:4740–9. <https://doi.org/10.1158/1078-0432.CCR-07-0143> PMID:17699851
43. Beroukhim R, Brunet JP, Di Napoli A, Mertz KD, Seeley A, Pires MM, Linhart D, Worrell RA, Moch H, Rubin MA, Sellers WR, Meyerson M, Linehan WM, et al. Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res.* 2009; 69:4674–81. <https://doi.org/10.1158/0008-5472.CAN-09-0146> PMID:19470766
44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID:16199517

## SUPPLEMENTARY MATERIALS

### Supplementary Figures



**Supplementary Figure 1. Sample clustering (GSE73731): sample dendrogram and trait indicator.** In the heat map, the darker the color, the higher the degree of infiltration of the seven kinds of T cells.



**Supplementary Figure 2. Heat map of the correlation between immune factors and hub genes. (A) Chemokines. (B) Receptors. (C) Immune-inhibitory factors. (D) Immune-stimulatory factors.**

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–3.

**Supplementary Table 1. Gene expression matrix for constructing gene co-expression network.**

**Supplementary Table 2. The fractions of six subtypes of T cells in every sample.**

**Supplementary Table 3. Genes of green module and Matascape enrichment analysis result.**

**Supplementary Table 4. The result of differential analysis of gene expression between normal tissue and tumor tissue by R package “limma”.**

Gene	logFC	AveExpr	t	P.Value	adj.P.Val	B
CCL5	2.761844	4.760074	16.69318	8.34E-52	1.82E-50	106.7216
CD2	2.28705	3.233462	15.32089	4.60E-45	7.75E-44	91.25291
CXCL9	2.625227	3.859447	13.8642	3.42E-38	4.27E-37	75.49697
CD3D	2.063852	2.945952	13.82764	5.04E-38	6.23E-37	75.11169
CCR5	1.667775	2.207914	13.2961	1.32E-35	1.44E-34	69.57093
CD8A	2.076911	2.644136	12.42316	9.51E-32	8.71E-31	60.73685
CD3G	1.022994	1.294262	10.84224	3.62E-25	2.40E-24	45.68733
IRF1	0.899948	3.35192	8.877243	7.60E-18	3.42E-17	28.98526
LCK	0.966669	2.169819	8.614633	6.00E-17	2.55E-16	26.94321
IFNG	0.555027	0.53599	6.865693	1.63E-11	5.02E-11	14.62135

Please browse Full Text version to see the data of Supplementary Tables 5.

**Supplementary Table 5. The results of gene set enrichment analysis.**

# A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment

Rongzhi Huang<sup>1,\*</sup>, Min Mao<sup>1,\*</sup>, Yunxin Lu<sup>1,\*</sup>, Qingliang Yu<sup>1,\*</sup>, Liang Liao<sup>1,2</sup>

<sup>1</sup>The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, The Guangxi Zhuang Autonomous Region, China

<sup>2</sup>Department of Traumatic Orthopedics and Hand Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, The Guangxi Zhuang Autonomous Region, China

\*Equal contribution

**Correspondence to:** Liang Liao; **email:** [237586233@qq.com](mailto:237586233@qq.com)

**Keywords:** melanoma, immune-related genes, classifier, overall survival, microenvironment

**Received:** August 7, 2019

**Accepted:** March 29, 2020

**Published:** April 20, 2020

**Copyright:** Huang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Background:** Melanoma is a cancer of the skin with potential to spread to other organs and is responsible for most deaths due to skin cancer. It is imperative to identify immune biomarkers for early melanoma diagnosis and treatment.

**Results:** 63 immune-related genes of the total 1039 unique IRGs retrieved were associated with overall survival of melanoma. A multi-IRGs classifier constructed using eight IRGs showed a powerful predictive ability. The classifier had better predictive power compared with the current clinical data. GSEA analysis showed multiple signaling differences between high and low risk score group. Furthermore, biomarker was associated with multiple immune cells and immune infiltration in tumor microenvironment.

**Conclusions:** The immune-related genes prognosis biomarker is an effective potential prognostic classifier in the immunotherapies and surveillance of melanoma.

**Methods:** Melanoma samples of genes were retrieved from TCGA and GEO databases while the immune-related genes (IRGs) were retrieved from the ImmPort database. WGCNA, Cox regression analysis and LASSO analysis were used to classify melanoma prognosis. ESTIMATE and CIBERSORT algorithms were used to explore the relationship between risk score and tumor immune microenvironment. GSEA analysis was performed to explore the biological signaling pathway.

## INTRODUCTION

Melanoma is a life-threatening malignancy with high metastasis and mortality rates [1, 2]. Approximately 232,000 new melanoma patients were diagnosed in 2011 and with 55,000 deaths recorded in the same year [3]. High mortality rates result from poor prognosis leading to late diagnosis. Therefore, there is need to come up with approaches for early diagnosis [4–6].

The TNM stage is an effective approach for detection of the cancer stage, is invaluable in cancer prognosis and

informs on the right therapy approaches [7]. However, differences in the overall survival associated with TNM stage method are observed [8]. Current studies on tumors have revealed the clinical limitations of TNM stage method [9, 10]. Therefore, there is a need to explore new melanoma markers to guide the clinical treatment and improve melanoma prognosis. Gene-based biomarkers have become more popular with the advances in human gene sequencing [11, 12].

Most immune system components are implicated in the initiation and progression of melanoma [13, 14]. In

tumor immunity, tumor cells act as antigens while immune cells and leukocytes infiltrates the tumor tissue function through chemotaxis for immune defense [13]. Immune escape also is an important factor in tumorigenesis [15, 16]. Currently, a myriad of new immunotherapy are used in melanoma and including PD-1, PD-L1 and CTLA-4 inhibitors [17, 18]. However, these approaches are effective only on a few patients while the majority of the patients have limited or no response to the therapy especially during melanoma progression [19, 20]. Therefore, comprehensive analyses of the correlation between immune genes and overall survival in melanoma are important in exploring the potential prognostic value of immune genes and new biomarkers.

In this study, our aim was to construct a novel immune-related genes biomarker for use in immunotherapies and melanoma prognosis. Comprehensive bioinformatics analyses were performed to explore underlying mechanisms of the biomarker. This study provides information for subsequent personalized diagnosis and treatment of melanoma.

## RESULTS

### Identification of survival-related modules by WGCNA

WGCNA analysis was carried out on 950 overlapping IRGs (Figure 1). The soft-thresholding power in WGCNA was determined based on a scale-free R2 ( $R^2 = 0.95$ ). Six modules were identified based on the average linkage hierarchical clustering and the

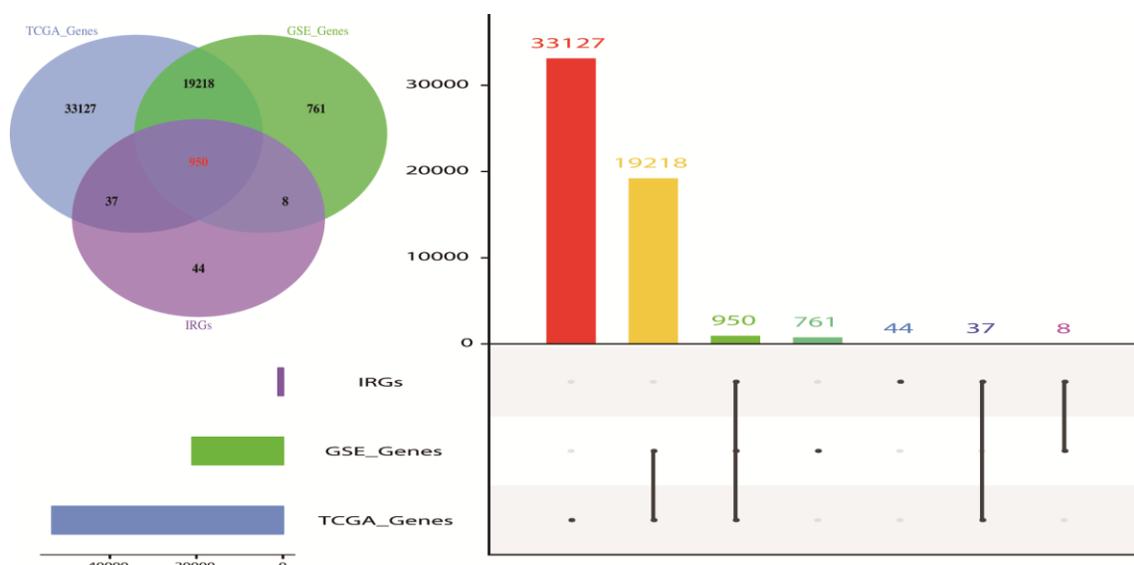
soft-thresholding power. The red module showed the highest correlation with the overall survival of melanoma. Additionally, the blue module was highly correlated with the overall survival of melanoma. The red module contained 22 IRGs while the blue module contained 138 IRGs (Figure 2). Data for these two modules were selected for further analysis.

### Construction of prognostic classifier based on IRGs

63 IRGs of the red and blue modules were identified as survival related IRGs of melanoma with the criterion of  $P < 0.01$  (Supplementary File 1). LASSO analysis identified eight IRGs (PSME1, CDC42, CMTM6, HLA-DQB1, HLA-C, CXCR6, CD8B, TNFSF13) which were included in the classifier (Figure 3). The coefficients of the eight IRGs are shown in Table 1 and the expression levels are shown in Figure 4. The high-RS group showed a poor overall survival rate compared with low-RS group based the Kaplan-Meier analysis (Figure 5B). Time-dependent ROC curves showed that the classifier had a strong predictive ability in GSE dataset (Figure 5A). In the training cohort, the AUC was 0.679 in 1 year, 0.743 in 3 years and 0.740 in 5 years (Figure 5A).

### Verification of the prognostic classifier in TCGA cohort

We used the TCGA cohort to validate the predictive ability of the classifier. Kaplan-Meier analysis showed that the high-RS group had a poor overall survival ( $P < 0.0001$ , Figure 5D). Time-dependent ROC curves

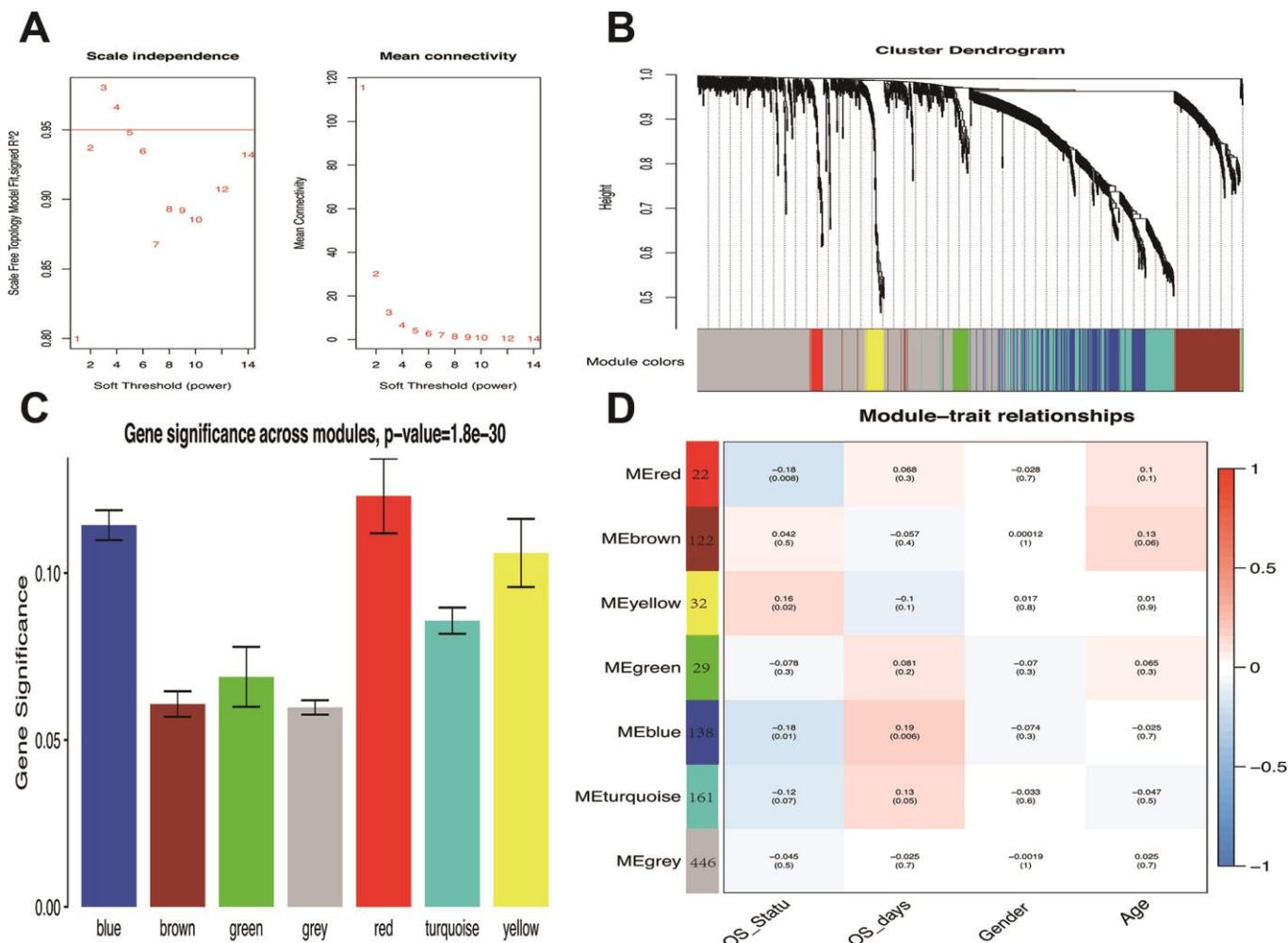


**Figure 1. Venn diagram and Histogram was used to visualize common IRGs shared between GEO dataset, TCGA dataset and IRGs.** 950 IRGs overlapped in the three datasets. The value used represented the number of gene symbol covered from the ensemble IDs and probe IDs. The number of genes annotated are presented on the y-axis.

showed that the classifier had a good accuracy with 0.642 in 1 year, 0.636 in 3 years and 0.645 in 5 years (Figure 5C). Moreover, the classifier had better predictive power and accuracy compared with other clinical features (Figure 5E, 5F). In Addition, the classifier was an independent factor in multivariate Cox analysis. Results of univariate and multivariate analyses in prognostic factors and overall survival were showed in Table 2.

### Immune infiltration score between high and low RS group

Kaplan-Meier analysis showed that different immune scores had differential overall survival in melanoma samples (Figure 6A, 6B). The immune score showed a significant difference between high and low-RS group (Figure 6C, 6D).



**Figure 2. Weighted melanoma gene co-expression network.** (A) The scale-free fit index for soft-thresholding powers. The soft-thresholding power in the WGCNA was determined based on a scale-free R<sup>2</sup> (R<sup>2</sup> = 0.95). The left panel presents the relationship between the soft-threshold and scale-free R<sup>2</sup>. The right panel presents the relationship between the soft-threshold and mean connectivity. (B) A dendrogram of the differentially expressed genes clustered based on different metrics. Each branch in the figure represents one gene, and every color below represents one co-expression module. (C) Distribution of average gene significance and errors in the modules associated with overall survival of melanoma patients. Based on the average linkage hierarchical clustering and the soft-thresholding power, six modules were identified. To determine the significance of each module, gene significance (GS) was calculated to measure the correlation between genes and sample traits. GS was defined as the log<sub>10</sub> conversion of the p-value in the linear regression between gene expression and clinical data (GS = lg P). The red and blue module showed high correlation with the survival of melanoma patients. (D) A heatmap showing the correlation between the gene module and clinical traits. The red module contained 22 IRGs while the blue module contained 138 IRGs. The correlation coefficient in each cell represented the correlation between gene module and the clinical traits, which decreased in size from red to blue. The blue module showed the highest positive correlation with the survival while the red module showed the highest negative correlation with the survival.

**Table 1. The IRGs in the prognostic classifier associated with OS in the GSE dataset.**

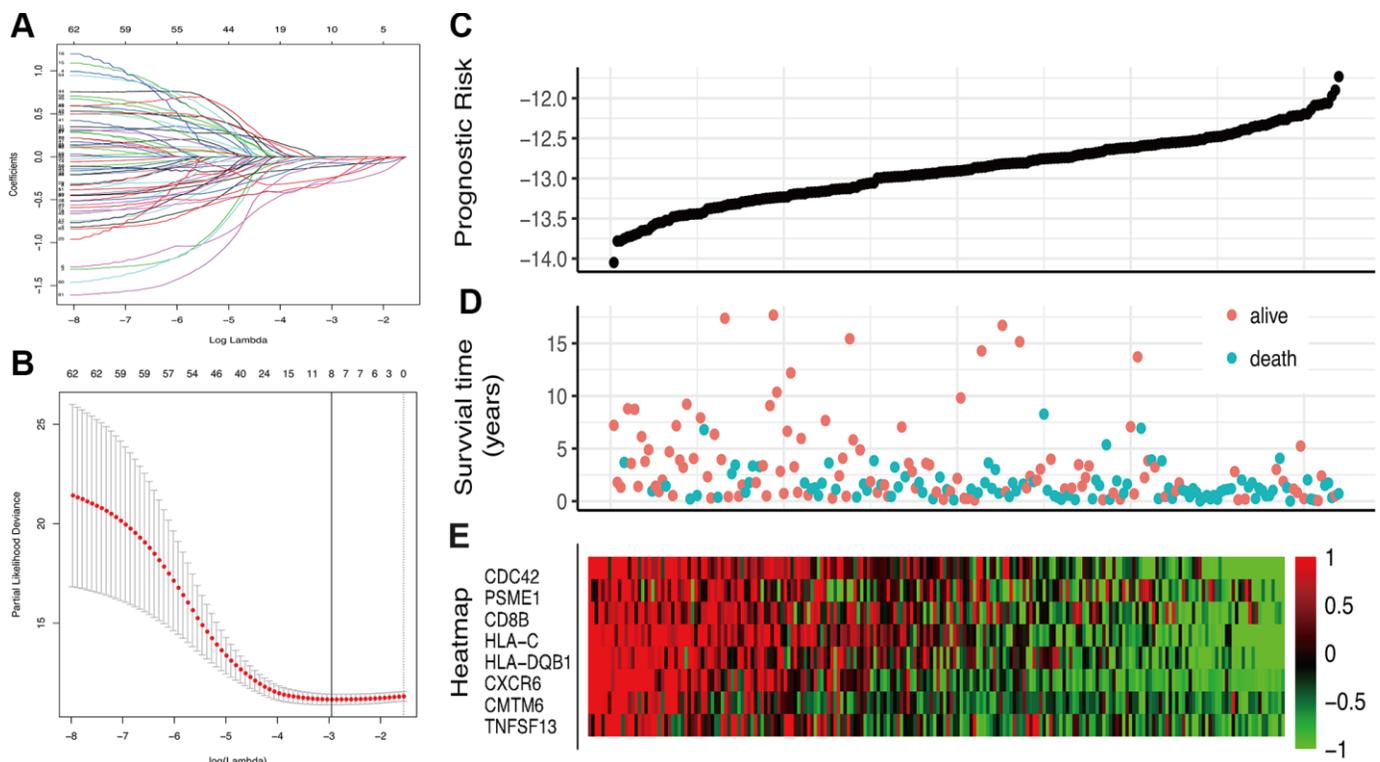
Symbol	Univariate Cox regression analysis			LASSO coefficient
	HR	95%CI	P Value	
PSME1	0.416	0.285-0.608	5.854205e-06	-0.30396287
CDC42	0.428	0.248-0.74	0.00236537	-0.24399092
CMTM6	0.364	0.218-0.608	0.0001131757	-0.23548175
HLA-DQB1	0.692	0.592-0.809	3.711835e-06	-0.07311844
HLA-C	0.595	0.466-0.759	2.920363e-05	-0.10691953
CXCR6	0.509	0.363-0.713	8.635839e-05	-0.03143482
CD8B	0.248	0.108-0.566	0.0009273984	-0.05032655
TNFSF13	0.172	0.055-0.54	0.002576346	-0.25872281

### Immune cell subtypes between high and low RS group

The 22 immune cell proportions of melanoma are shown in Figure 7A, 7B. Macrophages M0, Macrophages M2 and T cells CD8 accounted for a large proportion of melanoma immune cell infiltration. High and low RS groups showed differential immune cells expression (Figure 7C, 7D).

### GSEA analysis

GSEA analysis showed 14 significant KEGG pathways associated with risk score, including Rap1 signaling pathway, Ras signaling pathway, Herpes simplex virus 1 infection, Regulation of actin cytoskeleton, MAPK signaling pathway, Neuroactive ligand-receptor interaction, Human cytomegalovirus infection, Human T-cell leukemia virus 1 infection, Human



**Figure 3. Construction of the IRGs prognostic classifier. (A, B)** Determination of the number of factors by the LASSO analysis. **(C)** The distribution of RS. **(D)** The survival duration and status of patients. **(E)** A heatmap of IRGs in the classifier.

**Table 2. Univariate and multivariate analyses of prognostic factors and overall survival of melanoma patients in TCGA cohort.**

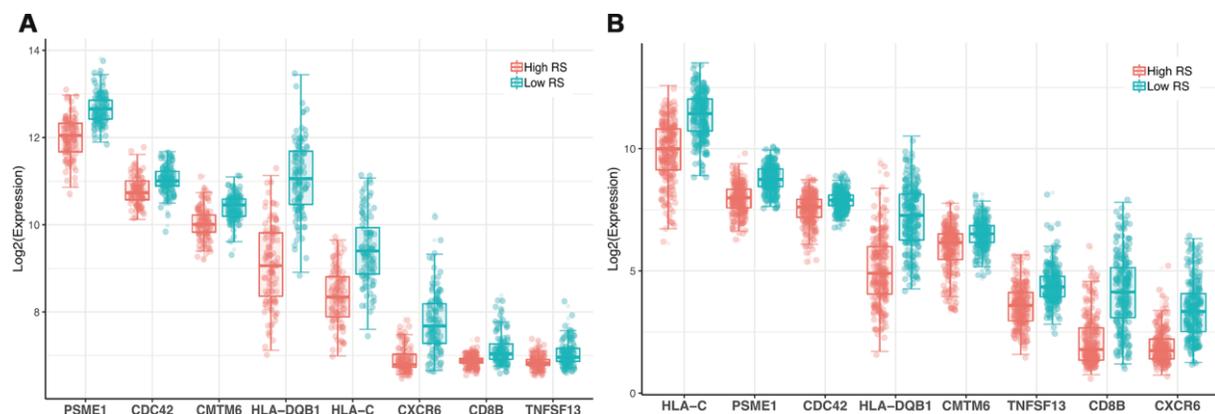
Characteristics	Univariate Cox regression analysis			multivariate Cox regression analysis		
	HR	95%CI	P Value	HR	95%CI	P Value
Age	1.025	1.015-1.035	3.63e-07	1.021	0.01-4.06	4.83e-05
Gender	0.877	0.655-1.175	3.79e-01	1.088	0.16-0.54	0.592
Local invasion	0.988	0.955-1.021	4.65e-01	0.987	0.02--0.66	0.511
Lymph node metastasis	1.087	1.032-1.145	1.74e-03	1.092	0.03-3.3	0.00096
Distant metastasis	1.161	0.887-1.52	2.78e-01	1.429	0.14-2.55	0.0107
TNM stage	1.000	0.964-1.038	9.80e-01	0.982	0.02--0.75	0.455
Multi-IRGs Classify	1.588	1.315-1.919	1.61e-06	1.704	0.1-5.13	2.94e-07

immunodeficiency virus 1 infection, Kaposi sarcoma-associated herpesvirus infection, Chemokine signaling pathway, Epstein-Barr virus infection, Tuberculosis and Cytokine-cytokine receptor interaction (Figure 8).

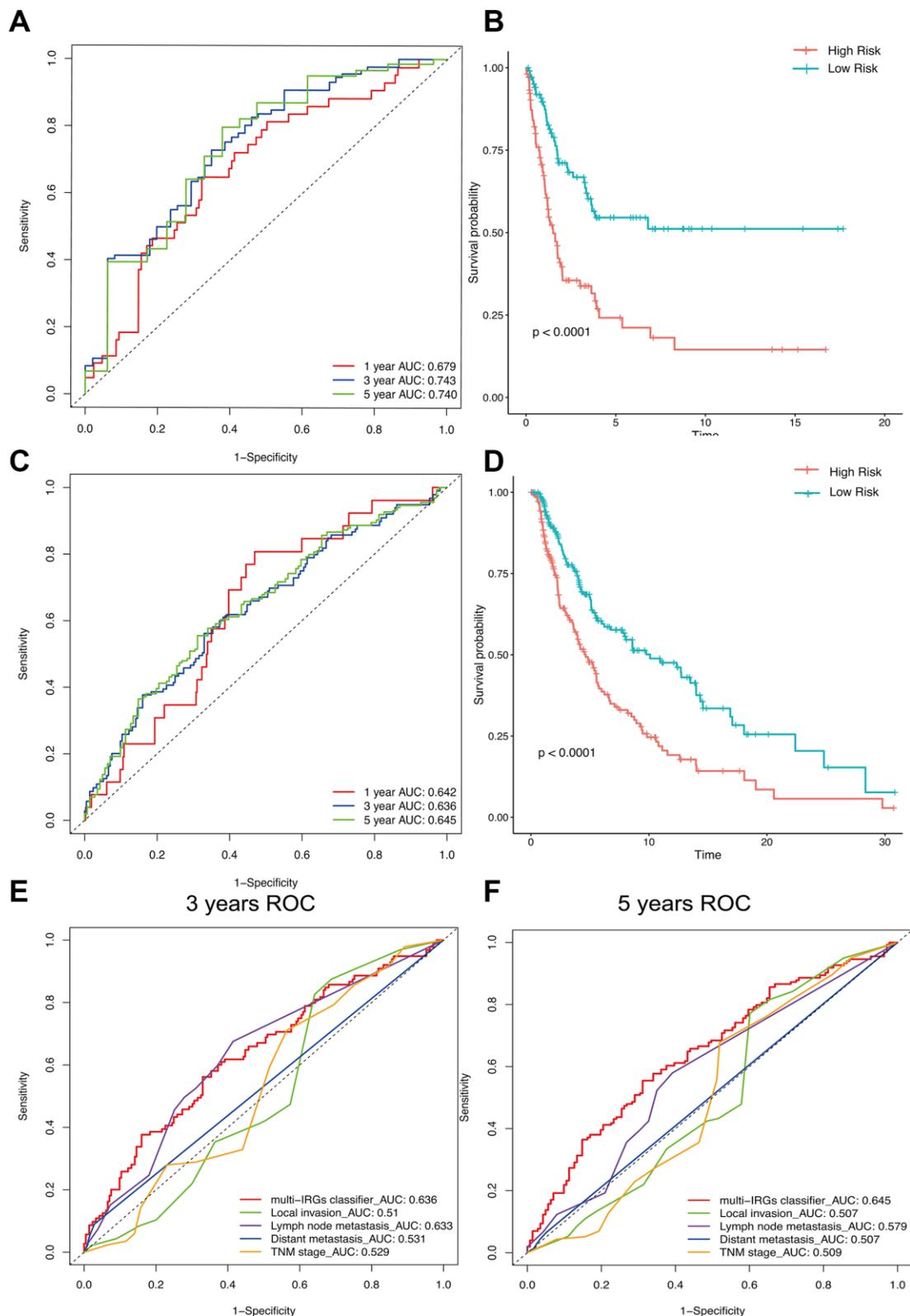
## DISCUSSION

Melanoma is a fatal skin cancer that affects many people worldwide each year [21]. Currently, immunotherapy is a successful treatment option for melanoma [22]. Notably, many researchers demonstrates the role of the immune cells on tumor cells [23, 24]. Moreover, immune components in melanoma tissue can be used to evaluate therapeutic efficacy and melanoma prognosis in patients [25]. In this study, 63 IRGs were found to be associated with melanoma prognosis, of which eight IRGs were adopted to construct a classifier. The classifier showed reliable predictive value and accuracy. In addition, we explored the relationship between RS and the prognosis value in melanoma. The findings showed differences in immune cell infiltration and multiple signaling pathways between high and low-RS group.

The PSME1, CDC42, CMTM6, HLA-DQB1, HLA-C, CXCR6, CD8B and TNFSF13 RGs were used in the classifier. These IRGs were reported to be associated with tumor prognosis in previous studies. Cell division cycle 42 (CDC42) protein, a member of Rho GTPases, activates multiple cellular processes by regulating actin cytoskeleton [26]. In addition, CDC42 facilitates the invasion and migration of melanoma cells [27–29]. Therefore, CDC42 inhibitors have been effective in melanoma treatment [30, 31]. CMTM6 is a ubiquitously expressed protein encoded by two distinct gene clusters located on chromosome 16 and chromosome 3 [32]. It enhances PD-L1 expression and anti-tumor immunity. Therefore, CMTM6 is a potential biomarker and therapeutic target for melanoma patients [33–35]. Among the HLA class I antigens, HLA-C locus recognizes the inhibitory killer cells and suppresses the functions of NK cells in melanoma patients [36–38]. Furthermore, the frequency of HLA-DQB1\*0301 and HLA-DQB1\*0303 alleles are highly expressed in melanoma patients [39]. Moreover, melanoma patients with DQB1\*0301 allele have thicker primary tumor and are more likely to have local or distant metastatic



**Figure 4. Expression profile of 8 genes. (A) GSE dataset (B) TCGA dataset.**



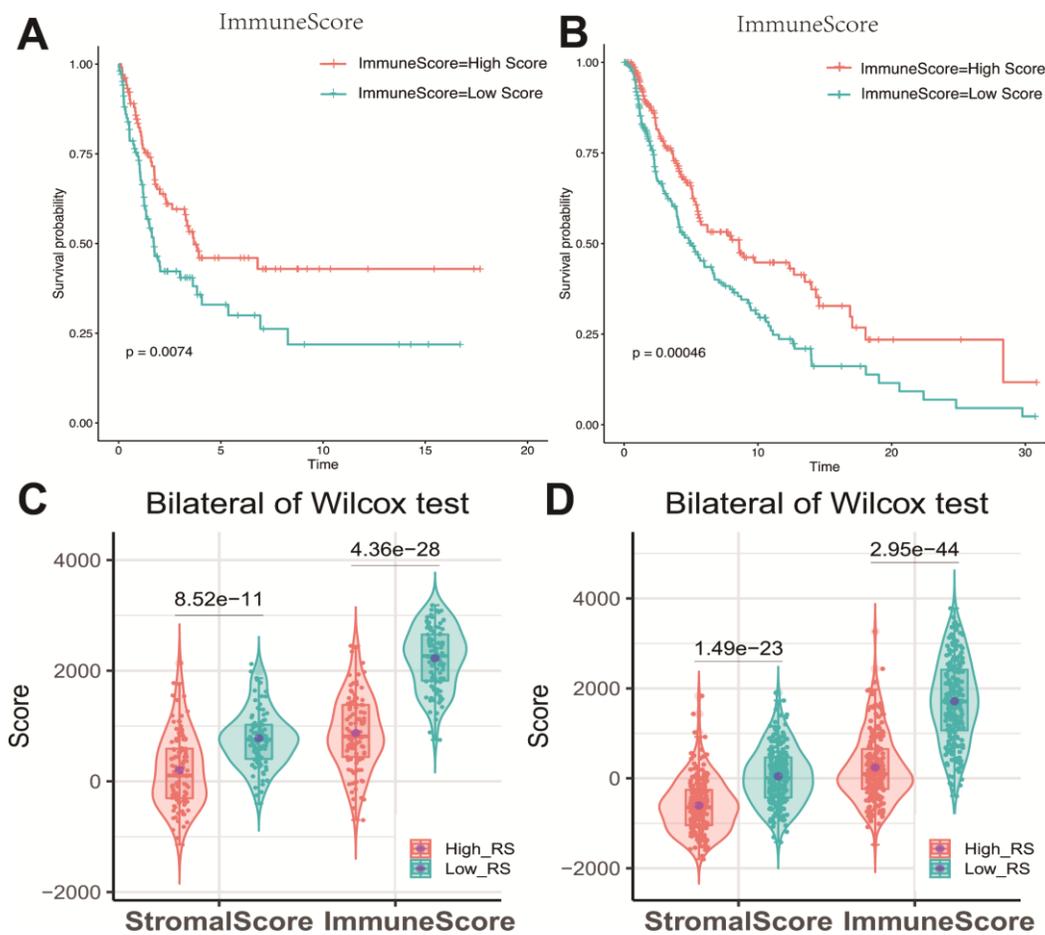
**Figure 5. The distribution of time-dependent ROC curves and Kaplan-Meier survival based on the integrated classifier in the training and independent validation sets.** ROC, receiver operator characteristic. AUC, the area under the curve. **(A)** ROC curve for the GSE cohort. **(B)** KM curve of the GSE cohort. **(C)** ROC curve of the TCGA cohort. **(D)** KM curve of the TCGA cohort. **(E)** 3-years correlation ROC curve in the TCGA cohort for the comparison of the classifier prognostic accuracy and clinical characteristics. **(F)** 5-years correlation ROC curve in the TCGA cohort for the comparison of the classifier prognostic accuracy and clinical characteristics.

disease [40]. Besides, the chemokine co-receptor CXCR6 was identified as a new biomarker associated with asymmetric self-renewal of tissue-specific stem cells. CXCR6+ cells cause rapid increase in tumor mass compared with CXCR6- cells [41]. TNFSF13, a member of the TNF superfamily, was reported to indicate the proliferative or survival state in tumor cells [42]. The multi-IRGs classifier established in this study showed high predictive value and accuracy through various analyses.

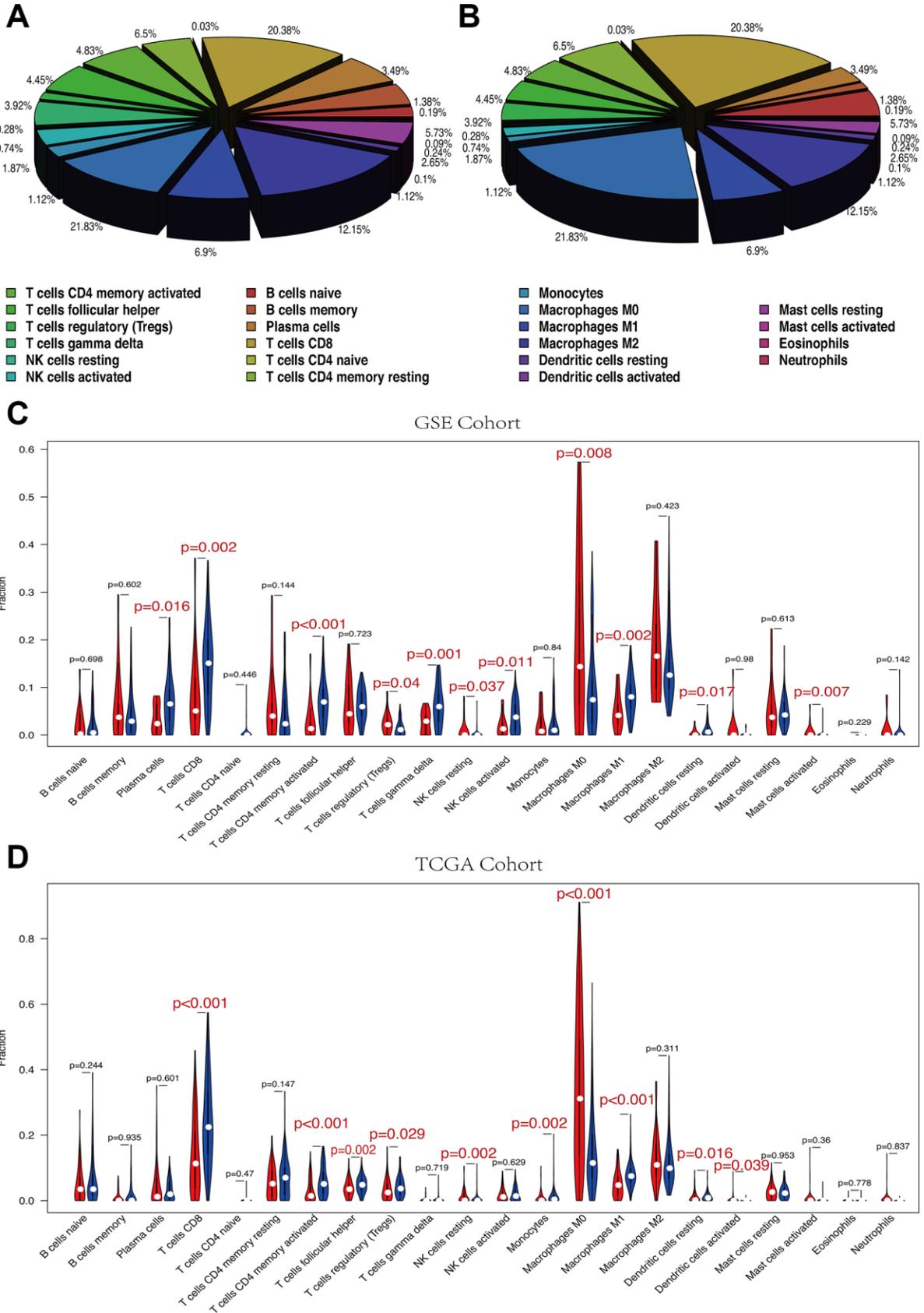
The degree of immune infiltration significantly affected melanoma survival. Previous studies demonstrate that immune cells in the tumor microenvironment can be used in the prognostic assessment of multiple tumors, such as glioblastoma, breast cancer, and melanoma [43–45]. In this study, the expression of eight genes affected immune infiltration scores. Patients with higher immune scores had better prognosis. This finding implies that prognosis value of risk score is associated with melanoma immune system.

To further explore the immune and risk score, we used the CIBERSORT algorithms to calculate the immune cell subtype in R platform. Our result showed that the two risk score groups expressed differential immune cell subtypes. Ali et al. demonstrated that imbalance in immune cell component ratio is highly correlated with poor prognosis and low survival in cancer patients [46, 47]. A previous study reported that CD8+ T cells produces granulocyte and perforin to kill tumor cells [48]. In our study, the immune cells found in melanoma mainly comprised macrophages M0, macrophages M2 and T cells CD8. In this study, T cell CD8 levels were low whereas M0 and M2 macrophages levels were high in the high-risk group. This implies that imbalance of T cell CD8 and M0, M2 macrophage ratio may reduce the survival rate of patients in the high-risk group. High expression of CD8+T cells may improve the prognosis of melanoma patients as well as reduce the risk factors.

GSEA analysis showed differences in 14 important signaling pathways between high and low RS groups.



**Figure 6.** (A) Impact of immune score on overall survival in melanoma based on KM analysis. (A) GSE cohort. (B) TCGA cohort. (C, D) Association with immune score, stromal score and risk score. The high-RS group showed lower immune score and stromal score comparing with low-RS group. (C) GSE cohort. (D) TCGA cohort.



**Figure 7.** (A, B) The mean proportion of 22 immune cells in GSE cohort. Macrophages M0, Macrophages M2 and T cells CD8 account for a large proportion of melanoma immune cell infiltration. (A) GSE cohort. (B) TCGA cohort. (C, D) Violin plot showing the relationship between risk score with immune score and stromal score. Red color represents high-RS group while blue color represents low-RS group. Differential immune cell type expression was observed between the high and low-RS groups. (C) GSE cohort. (D) TCGA cohort.

Inhibition of MAPK signaling pathway improved melanoma immune microenvironment by enhancing the melanoma antigen expression and down-regulating immunosuppressive cytokines [49, 50]. Additionally, chemokine signaling pathway participates in tumor growth. Some chemokines, such as CCR10 and CXCR3, have been shown to play an important role in the proliferation and metastasis of melanoma cells [51, 52].

In this study, LASSO regression analysis was used to establish a novel classifier using multiple IRGs and the classifier was verified using an independent cohort. Currently, few studies have used ESTIMATE and CIBERSORT algorithms to explore immune infiltration in melanoma. In this study, we use these algorithms to explore immune infiltration in melanoma using the R

software. These preliminary results could provide a perspective for exploring the role of immune infiltration in melanoma. However, this study has the following limitations. First, the reliability of our molecular mechanism analysis results is limited due to lack of *in vitro* or *in vivo* experiments. Second, this study was a retrospective study, therefore, prospective study should be carried out to validate the findings of our study.

In conclusion, we successfully constructed a multi-IRGs classifier with the powerful predictive function. Differences in the overall survival of high and low risk groups are implicated in immune infiltration, tumor microenvironment and the interaction of multiple signaling pathways. This study provides additional information on the analysis of melanoma pathogenesis and clinical treatment.

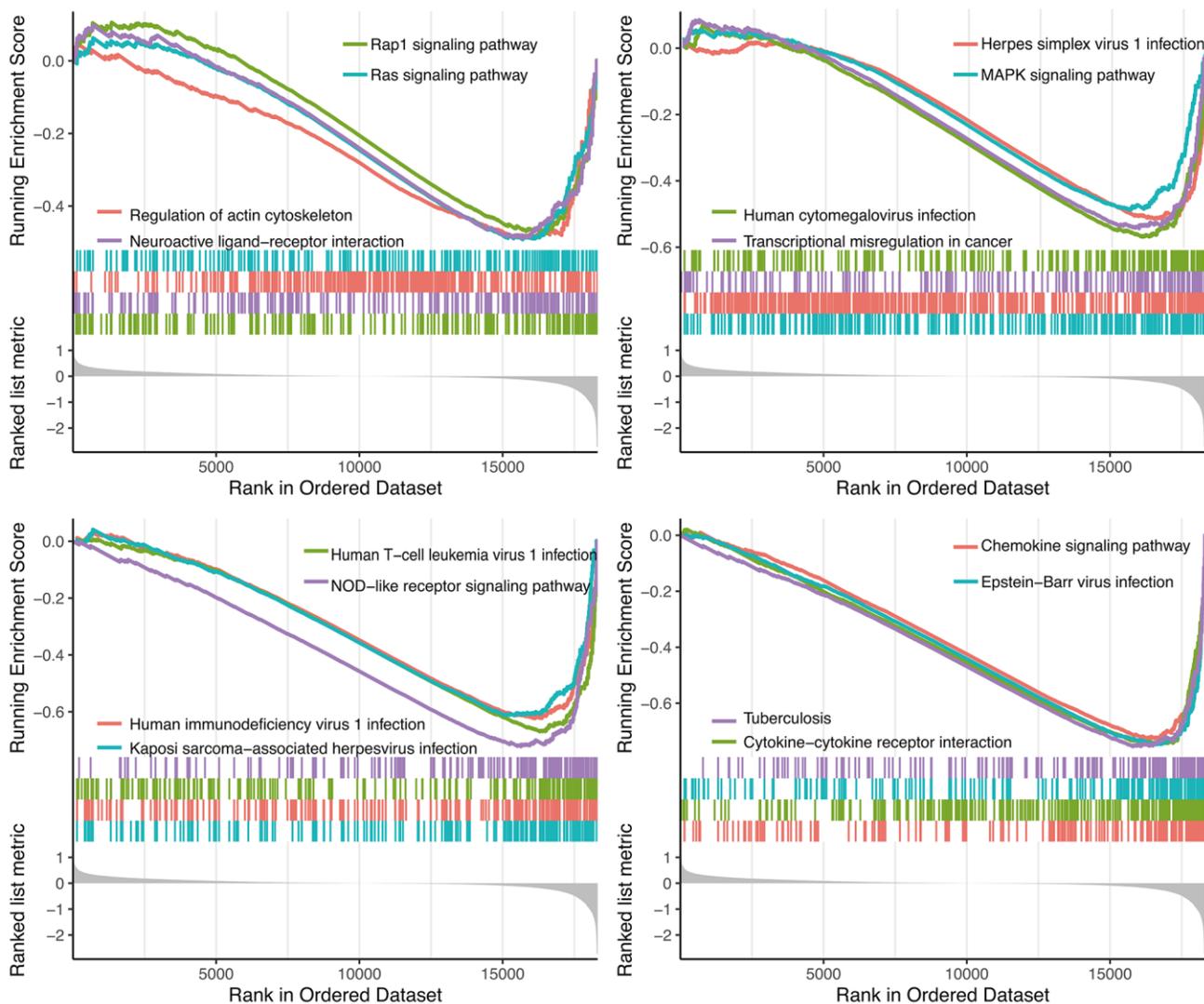


Figure 8. GSEA analysis.

## MATERIALS AND METHODS

### Data Processing

GSE65904 gene expression profiles were retrieved from the gene expression omnibus database (GEO: <https://www.ncbi.nlm.nih.gov/geo/>). In this study, the samples with no follow-up information or follow-up time less than 1 day were excluded. 210 melanoma samples were retrieved for subsequent analysis. Further, the probe IDs were converted to gene symbols using the illuminaHumanv4.db R package. The probe IDs with the highest mean value were reversed when more than one probe had a matched gene symbol. The GEO expression file was converted into log<sub>2</sub> (expression) for further analysis. Additionally, the RNA-TPM data and clinical data of melanoma samples were retrieved for external validation analysis using the TCGA biolinks R package. Samples with no follow-up information or follow-up time less than 1 day were excluded. The expression file of patients with the highest mean value was reversed when more than one expression file had matched patients. 428 melanoma samples in TCGA were used for analysis.

### Immune-related gene extraction

Immune-related genes (IRGs) data were retrieved from the ImmPort database (<https://immport.niaid.nih.gov>) (Supplementary File 2). Overlapping immune-related genes from the GEO dataset, TCGA dataset and IRGs were selected for further analysis.

### Weighted gene co-expression network analysis

GEO expression file was used for weighted gene co-expression network analysis (WGCNA) using WGCNA package. WGCNA was used to explore the relationship between the clinical features with expression modules. Module eigengenes (MEs) were defined as the first principal component of each gene module and adopted as the representative of all genes in each module. Gene significance (GS), as the mediator p-value ( $GS = \lg P$ ) for each gene, represented the degree of linear correlation between gene expression of the module and clinical features. Survival-related modules were defined according to  $P < 0.01$  and the higher GS value was extracted for further analysis.

### LASSO analysis

Univariate Cox regression analysis was performed to explore the impact of each gene on overall survival. The

IRGs of survival-related modules with  $P < 0.01$  were identified as survival-related IRGs and integrated into the Least Absolute Shrinkage and Selection Operator (LASSO) regression for identification of prognostic risk signatures. The risk score (RS) of each sample was calculated using the formula: risk score =  $\sum \text{expgene}_i * \beta_i$ .

The Kaplan-Meier curve analysis was further conducted to evaluate the relationship between the risk score and overall survival. The median value was used as the cut-off. Univariate and multivariate Cox regression analysis were performed to study the relationship between the index and the clinical features. To validate the accuracy and predictive ability of the signature, it was included in the TCGA dataset. The area under the curve (AUC) of the ROC curve was calculated and compared to examine the classifier performance using time ROC R package.

### Comparison of the degree of immune cell infiltration between high and low RS groups

To explore the relationship between risk score and melanoma prognosis, we analyzed the relationship between risk score and tumor microenvironment. The tumor microenvironment comprises a variety of cell types, including immune cells, mesenchymal cells, endothelial cells, inflammatory mediators, and extracellular matrix (ECM) molecules [53]. We used the ESTIMATE algorithm to determine the immune score of each sample using R software and further compared the difference in degree of immune cell infiltration between high and low-risk groups by Wilcoxon test.

### Comparison of 22 immune cell subtypes between high RS and low RS groups

To explore the differences of immune cell subtypes, CIBERSORT package was used to assess the proportions of 22 immune cell subtypes based on expression file [54]. The perm was set at 1000. Samples with  $P < 0.05$  in CIBERSORT analysis result were used in further analysis. Mann-Whitney U test was used to compare differences in immune cell subtypes in the high RS and low RS groups.

### Gene Set Enrichment analysis (GSEA)

To identify signaling pathway that are differentially activated between the high RS and low RS groups, we selected an ordered list of genes through limma R package and conducted Gene Set Enrichment Analysis (GSEA) with adjusted  $p < 0.05$  using the cluster flier R package.

## Statistical analysis

All analyses were carried out by R version 3.5.2 and corresponding packages. Kaplan-Meier analysis was further conducted to evaluate the relationship between immune score and overall survival using the survminer R package. The median value was set as the cut-off. The glmnet R package was used for LASSO analysis.

## Availability of data and materials

The GSE65904 gene expression profiles were retrieved from GEO (<https://www.ncbi.nlm.nih.gov/geo>). The TCGA data were retrieved from GDC data portal (<https://portal.gdc.cancer.gov/>). The immune-related genes (IRGs) data were retrieved from the ImmPort database (<https://import.niaid.nih.gov>). The R software (<https://www.r-project.org/>) was used for all statistical analyses.

## Abbreviations

IRGs: Immune-related genes; TME: Tumor microenvironment; GEO: Gene Expression Omnibus; TCGA: The cancer genome atlas project; LASSO: Least Absolute Shrinkage and Selection Operator; MEs: Module eigengenes; GS: Gene significance; ROC: Receiver operating characteristic curve; AUC: Area under the curve; RS: Risk score; OS: Overall survival.

## AUTHOR CONTRIBUTIONS

Data curation, Rongzhi Huang and Min Mao; Methodology, Rongzhi Huang; Software, Min Mao and Yunxin Lu; Verification, Yunxin Lu and Qingliang Yu; Visualization, Qingliang Yu and Liang Liao; Writing – original draft, Qingliang Yu and Yunxin Lu; Writing – review and editing, Liang Liao.

## ACKNOWLEDGMENTS

The authors thank TCGA and GEO for sharing the melanoma data.

## CONFLICTS OF INTEREST

All authors have read and approved submission of the manuscript. There authors declare no conflict of interest in relation to the submission.

## FUNDING

This work was supported by the funding by Guangxi health commission appropriate technology promotion project (S2018064) and Guangxi university young and

middle-aged teachers' basic research ability promotion project (2019KY0114).

## REFERENCES

1. Rastrelli M, Tropea S, Rossi CR, Alaibac M. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo*. 2014; 28:1005–11. PMID:[25398793](https://pubmed.ncbi.nlm.nih.gov/25398793/)
2. Wei D. A multigene support vector machine predictor for metastasis of cutaneous melanoma. *Mol Med Rep*. 2018; 17:2907–14. <https://doi.org/10.3892/mmr.2017.8219> PMID:[29257259](https://pubmed.ncbi.nlm.nih.gov/29257259/)
3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136:E359–86. <https://doi.org/10.1002/ijc.29210> PMID:[25220842](https://pubmed.ncbi.nlm.nih.gov/25220842/)
4. Cho YR, Chiang MP. Epidemiology, staging (new system), and prognosis of cutaneous melanoma. *Clin Plast Surg*. 2010; 37:47–53. <https://doi.org/10.1016/j.cps.2009.07.001> PMID:[19914457](https://pubmed.ncbi.nlm.nih.gov/19914457/)
5. Avilés-Izquierdo JA, Molina-López I, Rodríguez-Lomba E, Marquez-Rodas I, Suarez-Fernandez R, Lazaro-Ochaita P. Who detects melanoma? Impact of detection patterns on characteristics and prognosis of patients with melanoma. *J Am Acad Dermatol*. 2016; 75:967–74. <https://doi.org/10.1016/j.jaad.2016.07.009> PMID:[27645105](https://pubmed.ncbi.nlm.nih.gov/27645105/)
6. Lee CS, Thomas CM, Ng KE. An Overview of the Changing Landscape of Treatment for Advanced Melanoma. *Pharmacotherapy*. 2017; 37:319–33. <https://doi.org/10.1002/phar.1895> PMID:[28052356](https://pubmed.ncbi.nlm.nih.gov/28052356/)
7. Bertero L, Massa F, Metovic J, Zanetti R, Castellano I, Ricardi U, Papotti M, Cassoni P. Eighth Edition of the UICC Classification of Malignant Tumours: an overview of the changes in the pathological TNM classification criteria-What has changed and why? *Virchows Arch*. 2018; 472:519–31. <https://doi.org/10.1007/s00428-017-2276-y> PMID:[29209757](https://pubmed.ncbi.nlm.nih.gov/29209757/)
8. Perakis SO, Thomas JE, Pichler M. Non-coding RNAs Enabling Prognostic Stratification and Prediction of Therapeutic Response in Colorectal Cancer Patients. *Adv Exp Med Biol*. 2016; 937:183–204. [https://doi.org/10.1007/978-3-319-42059-2\\_10](https://doi.org/10.1007/978-3-319-42059-2_10) PMID:[27573901](https://pubmed.ncbi.nlm.nih.gov/27573901/)

9. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010; 17:1471–4.  
<https://doi.org/10.1245/s10434-010-0985-4>  
PMID:[20180029](https://pubmed.ncbi.nlm.nih.gov/20180029/)
10. Galon J, Pagès F, Marincola FM, Thurin M, Trinchieri G, Fox BA, Gajewski TF, Ascierto PA. The immune score as a new possible approach for the classification of cancer. *J Transl Med*. 2012; 10:1.  
<https://doi.org/10.1186/1479-5876-10-1>  
PMID:[22214470](https://pubmed.ncbi.nlm.nih.gov/22214470/)
11. Gu X, Li B, Jiang M, Fang M, Ji J, Wang A, Wang M, Jiang X, Gao C. RNA sequencing reveals differentially expressed genes as potential diagnostic and prognostic indicators of gallbladder carcinoma. *Oncotarget*. 2015; 6:20661–71.  
<https://doi.org/10.18632/oncotarget.3861>  
PMID:[25970782](https://pubmed.ncbi.nlm.nih.gov/25970782/)
12. Liu Y, Jing R, Xu J, Liu K, Xue J, Wen Z, Li M. Comparative analysis of oncogenes identified by microarray and RNA-sequencing as biomarkers for clinical prognosis. *Biomark Med*. 2015; 9:1067–78.  
<https://doi.org/10.2217/bmm.15.97>  
PMID:[26501374](https://pubmed.ncbi.nlm.nih.gov/26501374/)
13. Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol*. 2013; 25:261–67.  
<https://doi.org/10.1016/j.coi.2013.03.004>  
PMID:[23579076](https://pubmed.ncbi.nlm.nih.gov/23579076/)
14. Buonaguro FM, Pauza CD, Tornesello ML, Hainaut P, Franco R, Tommasino M. Cancer Diagnostic and Predictive Biomarkers 2016. *Biomed Res Int*. 2017; 2017:7362721.  
<https://doi.org/10.1155/2017/7362721>  
PMID:[28698877](https://pubmed.ncbi.nlm.nih.gov/28698877/)
15. Liu Y, Cao X. Immunosuppressive cells in tumor immune escape and metastasis. *J Mol Med (Berl)*. 2016; 94:509–22.  
<https://doi.org/10.1007/s00109-015-1376-x>  
PMID:[26689709](https://pubmed.ncbi.nlm.nih.gov/26689709/)
16. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, Birkbak NJ, Veeriah S, Van Loo P, Herrero J, Swanton C; TRACERx Consortium. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*. 2017; 171:1259–71.e11.  
<https://doi.org/10.1016/j.cell.2017.10.001>  
PMID:[29107330](https://pubmed.ncbi.nlm.nih.gov/29107330/)
17. Becht E, Giraldo NA, Dieu-Nosjean MC, Sautès-Fridman C, Fridman WH. Cancer immune contexture and immunotherapy. *Curr Opin Immunol*. 2016; 39:7–13.  
<https://doi.org/10.1016/j.coi.2015.11.009>  
PMID:[26708937](https://pubmed.ncbi.nlm.nih.gov/26708937/)
18. Silva IP, Long GV. Systemic therapy in advanced melanoma: integrating targeted therapy and immunotherapy into clinical practice. *Curr Opin Oncol*. 2017; 29:484–92.  
<https://doi.org/10.1097/CCO.0000000000000405>  
PMID:[28914644](https://pubmed.ncbi.nlm.nih.gov/28914644/)
19. Braun DA, Burke KP, Van Allen EM. Genomic Approaches to Understanding Response and Resistance to Immunotherapy. *Clin Cancer Res*. 2016; 22:5642–50.  
<https://doi.org/10.1158/1078-0432.CCR-16-0066>  
PMID:[27698000](https://pubmed.ncbi.nlm.nih.gov/27698000/)
20. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017; 541:321–30.  
<https://doi.org/10.1038/nature21349>  
PMID:[28102259](https://pubmed.ncbi.nlm.nih.gov/28102259/)
21. Owens B. Melanoma. *Nature*. 2014; 515:S109.  
<https://doi.org/10.1038/515S109a>  
PMID:[25407704](https://pubmed.ncbi.nlm.nih.gov/25407704/)
22. Valpione S, Campana LG. Immunotherapy for advanced melanoma: future directions. *Immunotherapy*. 2016; 8:199–209.  
<https://doi.org/10.2217/imt.15.111>  
PMID:[26809078](https://pubmed.ncbi.nlm.nih.gov/26809078/)
23. Giavina-Bianchi MH, Giavina-Bianchi PF, Festa C. Melanoma: tumor microenvironment and new treatments. *An Bras Dermatol*. 2017; 92:156–66.  
<https://doi.org/10.1590/abd1806-4841.20176183>  
PMID:[28538872](https://pubmed.ncbi.nlm.nih.gov/28538872/)
24. Fischer GM, Vashisht Gopal YN, McQuade JL, Peng W, DeBerardinis RJ, Davies MA. Metabolic strategies of melanoma cells: Mechanisms, interactions with the tumor microenvironment, and therapeutic implications. *Pigment Cell Melanoma Res*. 2018; 31:11–30.  
<https://doi.org/10.1111/pcmr.12661>  
PMID:[29049843](https://pubmed.ncbi.nlm.nih.gov/29049843/)
25. Ladányi A. Prognostic and predictive significance of immune cells infiltrating cutaneous melanoma. *Pigment Cell Melanoma Res*. 2015; 28:490–500.  
<https://doi.org/10.1111/pcmr.12371>  
PMID:[25818762](https://pubmed.ncbi.nlm.nih.gov/25818762/)
26. Hall A. Rho family GTPases. *Biochem Soc Trans*. 2012; 40:1378–82.  
<https://doi.org/10.1042/BST20120103>  
PMID:[23176484](https://pubmed.ncbi.nlm.nih.gov/23176484/)
27. Tucci MG, Lucarini G, Brancorsini D, Zizzi A, Pugnali A, Giacchetti A, Ricotti G, Biagini G. Involvement of E-

- cadherin, beta-catenin, Cdc42 and CXCR4 in the progression and prognosis of cutaneous melanoma. *Br J Dermatol.* 2007; 157:1212–16.  
<https://doi.org/10.1111/j.1365-2133.2007.08246.x>  
PMID:17970806
28. Gadea G, Sanz-Moreno V, Self A, Godi A, Marshall CJ. DOCK10-mediated Cdc42 activation is necessary for amoeboid invasion of melanoma cells. *Curr Biol.* 2008; 18:1456–65.  
<https://doi.org/10.1016/j.cub.2008.08.053>  
PMID:18835169
29. Vaysse A, Fang S, Brossard M, Wei Q, Chen WV, Mohamdi H, Vincent-Fetita L, Margaritte-Jeannin P, Lavielle N, Maubec E, Lathrop M, Avril MF, Amos CI, et al. A comprehensive genome-wide analysis of melanoma Breslow thickness identifies interaction between CDC42 and SCIN genetic variants. *Int J Cancer.* 2016; 139:2012–20.  
<https://doi.org/10.1002/ijc.30245> PMID:27347659
30. Lin CM, Lin YL, Ho SY, Chen PR, Tsai YH, Chung CH, Hwang CH, Tsai NM, Tzou SC, Ke CY, Chang J, Chan YL, Wang YS, et al. The inhibitory effect of 7,7'-dimethoxyagastisflavone on the metastasis of melanoma cells via the suppression of F-actin polymerization. *Oncotarget.* 2016; 8:60046–59.  
<https://doi.org/10.18632/oncotarget.10960>  
PMID:28947953
31. Maldonado MD, Dharmawardhane S. Targeting Rac and Cdc42 GTPases in Cancer. *Cancer Res.* 2018; 78:3101–11.  
<https://doi.org/10.1158/0008-5472.CAN-18-0619>  
PMID:29858187
32. Han W, Ding P, Xu M, Wang L, Rui M, Shi S, Liu Y, Zheng Y, Chen Y, Yang T, Ma D. Identification of eight genes encoding chemokine-like factor superfamily members 1-8 (CKLFSF1-8) by in silico cloning and experimental validation. *Genomics.* 2003; 81:609–17.  
[https://doi.org/10.1016/S0888-7543\(03\)00095-8](https://doi.org/10.1016/S0888-7543(03)00095-8)  
PMID:12782130
33. Burr ML, Sparbier CE, Chan YC, Williamson JC, Woods K, Beavis PA, Lam EY, Henderson MA, Bell CC, Stolzenburg S, Gilan O, Bloor S, Noori T, et al. CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature.* 2017; 549:101–05.  
<https://doi.org/10.1038/nature23643>  
PMID:28813417
34. Mezzadra R, Sun C, Jae LT, Gomez-Eerland R, de Vries E, Wu W, Logtenberg ME, Slagter M, Rozeman EA, Hofland I, Broeks A, Horlings HM, Wessels LF, et al. Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. *Nature.* 2017; 549:106–10.  
<https://doi.org/10.1038/nature23669>  
PMID:28813410
35. Mameessier E, Birnbaum DJ, Finetti P, Birnbaum D, Bertucci F. CMTM6 stabilizes PD-L1 expression and refines its prognostic value in tumors. *Ann Transl Med.* 2018; 6:54.  
<https://doi.org/10.21037/atm.2017.11.26>  
PMID:29610746
36. Campillo JA, Martínez-Escribano JA, Muro M, Moya-Quiles R, Marín LA, Montes-Ares O, Guerra N, Sánchez-Pedreño P, Frías JF, Lozano JA, García-Alonso AM, Alvarez-López MR. HLA class I and class II frequencies in patients with cutaneous malignant melanoma from southeastern Spain: the role of HLA-C in disease prognosis. *Immunogenetics.* 2006; 57:926–33.  
<https://doi.org/10.1007/s00251-005-0065-2>  
PMID:16365741
37. Konjević G, Mirjacić Martinović K, Jurisić V, Babović N, Spuzić I. Biomarkers of suppressed natural killer (NK) cell function in metastatic melanoma: decreased NKG2D and increased CD158a receptors on CD3-CD16+ NK cells. *Biomarkers.* 2009; 14:258–70.  
<https://doi.org/10.1080/13547500902814658>  
PMID:19489688
38. Kandilarova SM, Paschen A, Mihaylova A, Ivanova M, Schadendorf D, Naumova E. The Influence of HLA and KIR Genes on Malignant Melanoma Development and Progression. *Arch Immunol Ther Exp (Warsz).* 2016 (Suppl 1); 64:73–81.  
<https://doi.org/10.1007/s00005-016-0437-3>  
PMID:28083606
39. Bateman AC, Turner SJ, Theaker JM, Howell WM. HLA-DQB1\*0303 and \*0301 alleles influence susceptibility to and prognosis in cutaneous malignant melanoma in the British Caucasian population. *Tissue Antigens.* 1998; 52:67–73.  
<https://doi.org/10.1111/j.1399-0039.1998.tb03025.x>  
PMID:9714476
40. Lee JE, Reveille JD, Ross MI, Platsoucas CD. HLA-DQB1\*0301 association with increased cutaneous melanoma risk. *Int J Cancer.* 1994; 59:510–13.  
<https://doi.org/10.1002/ijc.2910590413>  
PMID:7960221
41. Taghizadeh R, Noh M, Huh YH, Ciusani E, Sigalotti L, Maio M, Arosio B, Nicotra MR, Natali P, Sherley JL, La Porta CA. CXCR6, a newly defined biomarker of tissue-specific stem cell asymmetric self-renewal, identifies more aggressive human melanoma cancer stem cells. *PLoS One.* 2010; 5:e15183.  
<https://doi.org/10.1371/journal.pone.0015183>  
PMID:21203549
42. Mhaweck-Fauceglia P, Kaya G, Sauter G, McKee T, Donze O, Schwaller J, Huard B. The source of APRIL up-regulation in human solid tumor lesions. *J Leukoc Biol.* 2006; 80:697–704.

- <https://doi.org/10.1189/jlb.1105655>  
PMID:[16793914](https://pubmed.ncbi.nlm.nih.gov/16793914/)
43. Fortis SP, Mahaira LG, Anastasopoulou EA, Voutsas IF, Perez SA, Baxevanis CN. Immune profiling of melanoma tumors reflecting aggressiveness in a preclinical model. *Cancer Immunol Immunother*. 2017; 66:1631–42.  
<https://doi.org/10.1007/s00262-017-2056-1>  
PMID:[28871365](https://pubmed.ncbi.nlm.nih.gov/28871365/)
44. Priedigkeit N, Watters RJ, Lucas PC, Basudan A, Bhargava R, Horne W, Kolls JK, Fang Z, Rosenzweig MQ, Brufsky AM, Weiss KR, Oesterreich S, Lee AV. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. *JCI Insight*. 2017; 2:95703.  
<https://doi.org/10.1172/jci.insight.95703>  
PMID:[28878133](https://pubmed.ncbi.nlm.nih.gov/28878133/)
45. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY)*. 2018; 10:592–605.  
<https://doi.org/10.18632/aging.101415>  
PMID:[29676997](https://pubmed.ncbi.nlm.nih.gov/29676997/)
46. Ali HR, Chlon L, Pharoah PD, Markowitz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med*. 2016; 13:e1002194.  
<https://doi.org/10.1371/journal.pmed.1002194>  
PMID:[27959923](https://pubmed.ncbi.nlm.nih.gov/27959923/)
47. Bense RD, Sotiriou C, Piccart-Gebhart MJ, Haanen JB, van Vugt MA, de Vries EG, Schröder CP, Fehrmann RS. Relevance of Tumor-Infiltrating Immune Cell Composition and Functionality for Disease Outcome in Breast Cancer. *J Natl Cancer Inst*. 2016; 109: djw192.  
<https://doi.org/10.1093/jnci/djw192>  
PMID:[27737921](https://pubmed.ncbi.nlm.nih.gov/27737921/)
48. Tsukumo SI, Yasutomo K. Regulation of CD8<sup>+</sup> T Cells and Antitumor Immunity by Notch Signaling. *Front Immunol*. 2018; 9:101.  
<https://doi.org/10.3389/fimmu.2018.00101>  
PMID:[29441071](https://pubmed.ncbi.nlm.nih.gov/29441071/)
49. Deken MA, Gadiot J, Jordanova ES, Lacroix R, van Gool M, Kroon P, Pineda C, Geukes Foppen MH, Scolyer R, Song JY, Verbrugge I, Hoeller C, Dummer R, et al. Targeting the MAPK and PI3K pathways in combination with PD1 blockade in melanoma. *Oncoimmunology*. 2016; 5:e1238557.  
<https://doi.org/10.1080/2162402X.2016.1238557>  
PMID:[28123875](https://pubmed.ncbi.nlm.nih.gov/28123875/)
50. Kakavand H, Rawson RV, Pupo GM, Yang JY, Menzies AM, Carlino MS, Kefford RF, Howle JR, Saw RP, Thompson JF, Wilmott JS, Long GV, Scolyer RA, Rizos H. PD-L1 Expression and Immune Escape in Melanoma Resistance to MAPK Inhibitors. *Clin Cancer Res*. 2017; 23:6054–61.  
<https://doi.org/10.1158/1078-0432.CCR-16-1688>  
PMID:[28724663](https://pubmed.ncbi.nlm.nih.gov/28724663/)
51. Murakami T, Cardones AR, Finkelstein SE, Restifo NP, Klaunberg BA, Nestle FO, Castillo SS, Dennis PA, Hwang ST. Immune evasion by murine melanoma mediated through CC chemokine receptor-10. *J Exp Med*. 2003; 198:1337–47.  
<https://doi.org/10.1084/jem.20030593>  
PMID:[14581607](https://pubmed.ncbi.nlm.nih.gov/14581607/)
52. Jenkins MH, Brinckerhoff CE, Mullins DW. CXCR3 signaling in BRAFWT melanoma increases IL-8 expression and tumorigenicity. *PLoS One*. 2015; 10:e0121140.  
<https://doi.org/10.1371/journal.pone.0121140>  
PMID:[25798946](https://pubmed.ncbi.nlm.nih.gov/25798946/)
53. Runa F, Hamalian S, Meade K, Shisgal P, Gray PC, Kelber JA. Tumor microenvironment heterogeneity: challenges and opportunities. *Curr Mol Biol Rep*. 2017; 3:218–29.  
<https://doi.org/10.1007/s40610-017-0073-7>  
PMID:[29430386](https://pubmed.ncbi.nlm.nih.gov/29430386/)
54. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12:453–57.  
<https://doi.org/10.1038/nmeth.3337>  
PMID:[25822800](https://pubmed.ncbi.nlm.nih.gov/25822800/)

## SUPPLEMENTARY MATERIALS

### Supplementary Files

Please browse Full Text version to see the data of Supplementary Files 1 and 2.

**Supplementary File 1.** 63 IRGs of the two modules were identified as survival-related IRGs of melanoma with the criterion of  $P < 0.01$ .

**Supplementary File 2.** The downloaded data of Immune-related genes (IRGs).

# PTX3 modulates the immunoflogosis in tumor microenvironment and is a prognostic factor for patients with clear cell renal cell carcinoma

Giuseppe Stefano Netti<sup>1,\*</sup>, Giuseppe Lucarelli<sup>2,\*</sup>, Federica Spadaccino<sup>1</sup>, Giuseppe Castellano<sup>3</sup>, Margherita Gigante<sup>1</sup>, Chiara Divella<sup>3</sup>, Maria Teresa Rocchetti<sup>1</sup>, Federica Rascio<sup>4</sup>, Vito Mancini<sup>5</sup>, Giovanni Stallone<sup>4</sup>, Giuseppe Carrieri<sup>5</sup>, Loreto Gesualdo<sup>3</sup>, Michele Battaglia<sup>2</sup>, Elena Ranieri<sup>1</sup>

<sup>1</sup>Clinical Pathology Unit and Center of Molecular Medicine, Department of Medical and Surgical Sciences, University of Foggia, Viale Luigi Pinto 71122, Foggia, Italy

<sup>2</sup>Urology and Renal Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari "Aldo Moro", Bari 70124, Italy

<sup>3</sup>Nephrology Dialysis and Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari "Aldo Moro", Bari 70124, Italy

<sup>4</sup>Nephrology Dialysis and Transplantation Unit, Department of Medical and Surgical Sciences, University of Foggia, Viale Luigi Pinto 71122, Foggia, Italy

<sup>5</sup>Urology and Renal Transplantation Unit, Department of Medical and Surgical Sciences, University of Foggia, Viale Luigi Pinto 71122, Foggia, Italy

\*Equal contribution

**Correspondence to:** Elena Ranieri; **email:** [elena.ranieri@unifg.it](mailto:elena.ranieri@unifg.it)

**Keywords:** renal cell carcinoma, pentraxin 3, complement system, biomarker

**Received:** February 17, 2020

**Accepted:** March 31, 2020

**Published:** April 28, 2020

**Copyright:** Netti et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Pentraxin-3 (PTX3) belongs to the pentraxine family, innate immune regulators involved in angiogenesis, proliferation and immune escape in cancer. Here, we evaluated PTX3 tissue expression and serum levels as biomarkers of clear cell renal cell carcinoma (ccRCC) and analyzed the possible role of complement system activation on tumor site. A 10-year retrospective cohort study including patients undergoing nephrectomy for ccRCC was also performed. PTX3 expression was elevated in both neoplastic renal cell lines and tissues, while it was absent in both normal renal proximal tubular cells (HK2) and normal renal tissues. Analysis of complement system activation on tumor tissues showed the co-expression of PTX3 with C1q, C3aR, C5R1 and CD59, but not with C5b-9 terminal complex. RCC patients showed higher serum PTX3 levels as compared to non-neoplastic patients ( $p < 0.0001$ ). Higher PTX3 serum levels were observed in patients with higher Fuhrman grade ( $p < 0.01$ ), lymph node ( $p < 0.0001$ ), and visceral metastases ( $p < 0.001$ ). Patients with higher PTX3 levels also showed significantly lower survival rates ( $p = 0.002$ ). Our results suggest that expression of PTX3 can affect the immunoflogosis in the ccRCC microenvironment, by activating the classical pathway of CS (C1q) and releasing pro-angiogenic factors (C3a, C5a). The up-regulation of CD59 also inhibits the complement-mediated cellular lysis.

## INTRODUCTION

Renal cell carcinoma (RCC) is the most common type of renal neoplasia and accounts for about 3% of all adult

malignancies in western countries [1]. Recent estimates have calculated that in 2020, in the United States 73,750 new cases will be diagnosed and 14,830 patients will die of RCC [2]. Due to usually asymptomatic clinical

course, the diagnosis of most cases of renal cancer is often incidental, following diagnostic tests performed for other clinical conditions, and not rarely shows neoplasms in advanced clinical stage [1]. Furthermore, RCC is a chemo- and radio-resistant neoplasia, therefore the current therapeutic strategies are ultimately based on the surgical approach [3].

RCC encompasses a heterogeneous group of cancers derived from renal tubular epithelial cells [4]. The most common renal cancer type in adults is adenocarcinoma: it accounts almost 90% and in a small percentage (2%) it may also be bilateral. Clear cell renal cell carcinoma (ccRCC) is the most frequently diagnosed subtype and causes the most clinically severe phenotype. In addition, up to 30% of the patients present metastatic disease at diagnosis, and around 20-30% of subjects undergoing surgery will suffer recurrence. In this scenario, early diagnosis is crucial for improving the survival rate of these patients, and the introduction of high-throughput omics technologies has led not only to a detailed molecular characterization of RCC, but also to the identification of novel biomarkers [5–13].

In the last few years, several studies have showed that cigarette smoking, obesity, hypertension, diabetes and End Stage Renal Disease, represent common risk factors for this tumor [14–18]. Moreover, an in-depth understanding of the molecular basis of RCC has led to introduction in clinical practice of novel targeted therapies, including anti-angiogenic agents (sorafenib, sunitinib, pazopanib, axitinib, and bevacizumab), mTOR (temsirolimus and everolimus), and immune checkpoint inhibitors (nivolumab). However, these drugs yield partial responses in a minority of patients, with no evidence of complete responses [3, 19].

Modifications of the tumor microenvironment represent a growing field of investigation in order to highlight potential mechanisms of tumor progression and resistance to targeted therapies [20].

In particular during the oncogenesis, the reactive immunoflogosis seems to play a crucial role to counteract the development of neoplastic cells. On the other hand, its persistence may paradoxically promote the cancer progression, by enriching the tumor microenvironment with pro-inflammatory cytokines and growth factors that can lead to uncontrolled proliferative response [21–23]. In this setting Pentraxin-3 (PTX-3) might play a crucial role. PTX3 is an opsonin belonging to the pentraxin superfamily, which acts as pattern recognition molecule (PRM) of the immune system. This molecule is able to recognize microbial fractions and cellular debris, to promote

phagocytosis (opsonization), to activate the complement system and to modulate the inflammation process [24].

The similarity of PTX3 with C-reactive protein (CRP), the most widely used inflammation biomarker, has led to investigate the role of PTX3 in several infectious and inflammatory disorders. PTX3 acts as an acute phase protein and its plasma levels increase rapidly (peak at 6-8 hours) from very low baseline values in healthy subjects, up to higher serum levels in inflammatory conditions. PTX3 can be produced directly within the site of inflammation from both resident and infiltrating cell types and acts with a paracrine effect, unlike other short pentraxins (including CRP), which are produced by the liver and released in the blood [23].

Recent studies have shown how PTX3 can influence the pathogenesis of different cancer types, but plays an ambivalent role, i.e. acting as a tumor suppressor or pro-oncogenic factor in relation to the neoplastic type and the tumor microenvironment. In some cases PTX3 acts as tumor suppressor, inhibiting the proliferation and angiogenesis of FGF2-mediated tumor cells, as well as the epithelial-to-mesenchymal transition (EMT) and its metastatic potential. In other cancer types the high intratumoral expression of PTX3 is associated with poor prognosis [25, 26].

Moreover, PTX3 shares with the other short classic pentraxins the ability to modulate the complement system, through a direct interaction with the key molecules involved in the activation and/or regulation of the complement system cascade. PTX3 mediates the activation of the classical pathway and lectin pathway because of binding with C1q and MBL, and is able to affect the alternative pathway via CFH binding. The complement system, as part of the innate immune system, enhances the ability of antibodies and phagocytic cells to clear microbes and damaged cells from an organism, promotes inflammation, and attacks the pathogen's cell membrane. In the inflammatory context, complement system releases very strong pro-inflammatory molecules, such as the anaphylatoxins C3a and C5a. However, recent evidence show that complement system may promote cell proliferation and regeneration, thus suggesting a possible relationship between complement system activation and cancer [27].

In the present study, we evaluated PTX3 tissue expression and serum levels as biomarkers of clear cell renal cell carcinoma (ccRCC) and analyzed the possible role of complement system activation on tumor site.

## RESULTS

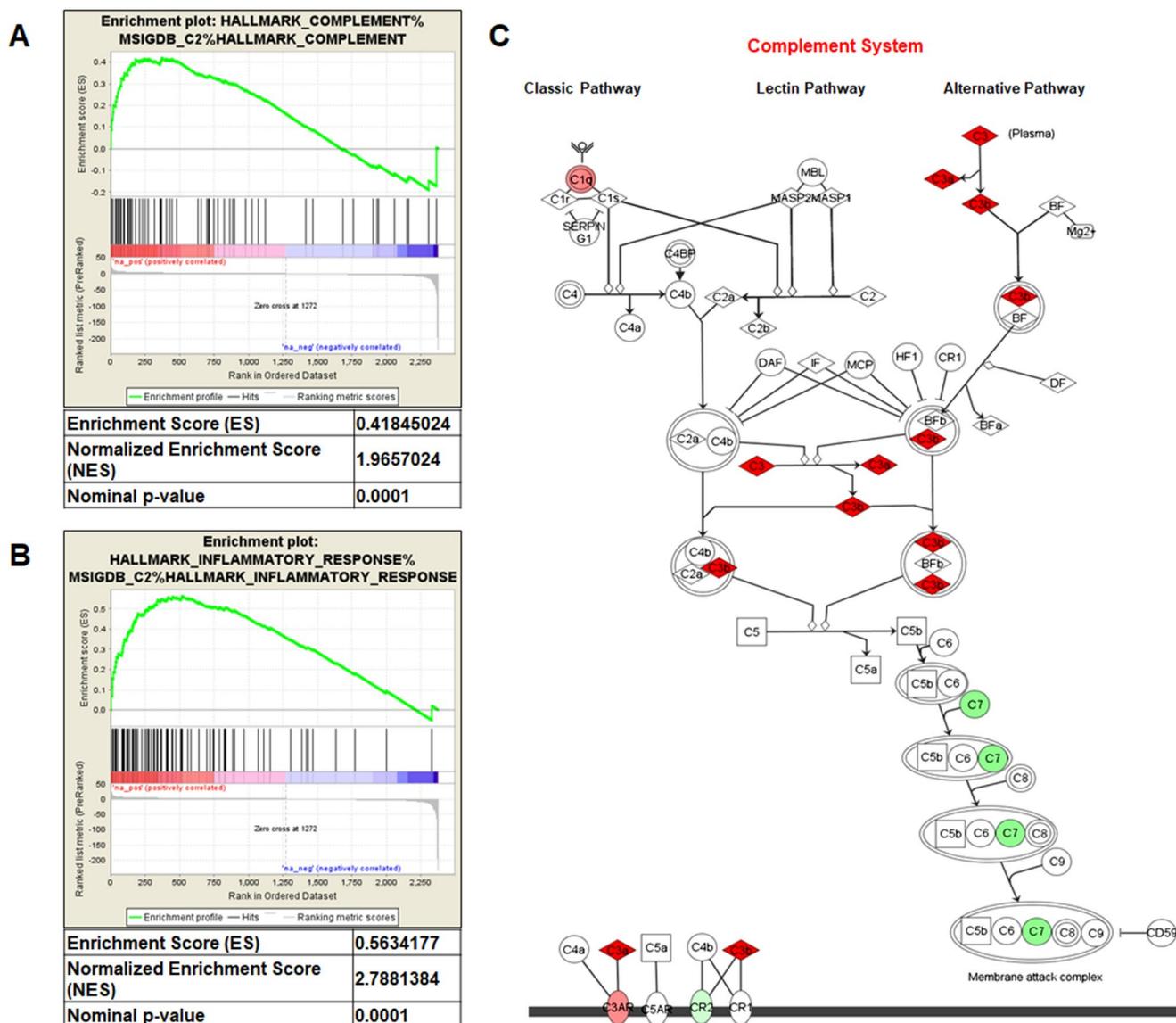
### Gene set enrichment analysis (GSEA) and microarray analysis

Gene Set Enrichment Analysis (GSEA) of the GSE47032 dataset showed that ccRCC featured multiple enriched gene sets depicting inflammatory response (NES=2.78;  $p=0.0001$ ; Figure 1A) and complement activation (NES=1.96;  $p=0.0001$ ; Figure 1B). Moreover, microarray analysis using IPA, showed that the canonical molecules associated with complement system were significantly modulated (Figure 1C).

### PTX3 protein expression in RCC cell lines and renal tissues from ccRCC patients

PTX3 protein expression was analyzed both in neoplastic and in normal renal proximal tubular epithelial cells (PTEC) by confocal microscopy. Noteworthy, three different renal cancer cell lines showed significantly higher PTX3 expression, as compared by PTEC (Figure 2A, 2B). This observation was strengthened by quantification of specific fluorescence ( $p<0.05$ ; Figure 2C).

Then, we analyzed the tissue expression of PTX3 in tumor tissues of 30 consecutive patients who underwent



**Figure 1.** Gene Set Enrichment Analysis (GSEA) of the GSE47032 dataset (A and B). Complement system pathway from Ingenuity Pathway Analysis (C). Genes in green and red are respectively under- and over-expressed in the ccRCC-gene signature.

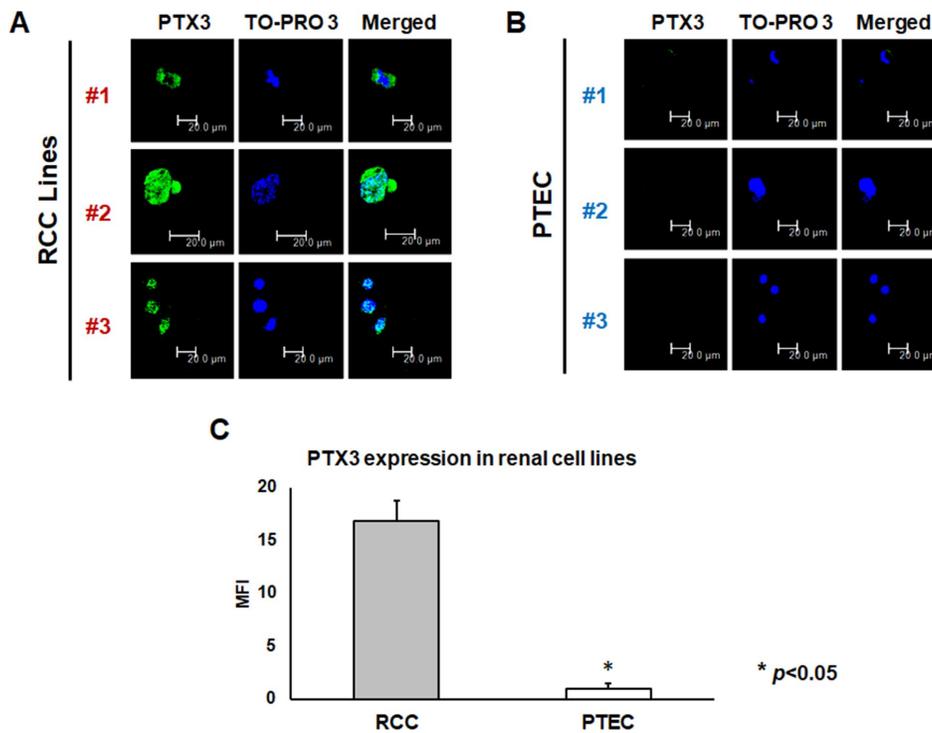
radical nephrectomy for ccRCC. A control group of 10 subject who underwent renal biopsy in suspicion of chronic nephropathy, but with a normal renal histology, were also analyzed. The main clinical and histologic features of the entire study population of 30 patients with renal clear cell carcinoma subjected to PTX3 tissue expression analyses are summarized in Supplementary Table 1.

PTX3 tissue expression was significantly higher in ccRCC patients (Figure 3A–3C), while in normal

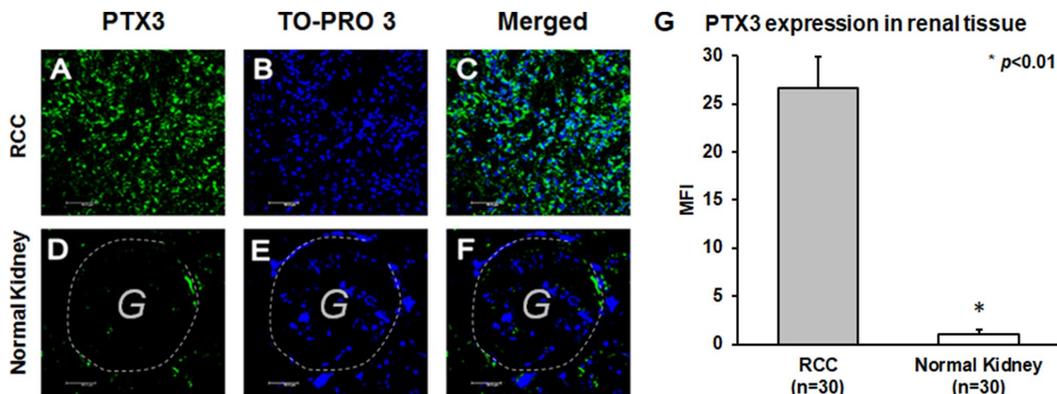
kidney it was virtually absent (Figure 3D–3F). This observation was strengthened by quantification of specific fluorescence ( $p < 0.01$ ; Figure 3G).

### Complement system activation in renal tissues from ccRCC patients

We then investigated the activation of the complement cascade in renal cancer and in normal renal tissue. Since PTX3 can activate the complement system through the classic pathway, we evaluated the deposition of C1q



**Figure 2.** PTX3 expression in neoplastic (A) and proximal tubular epithelial cells (PTEC) (B) by confocal microscopy and quantification of specific fluorescence (C).



**Figure 3.** PTX3 expression in neoplastic (A–C) and normal renal tissues (D–F) by confocal microscopy and quantification of specific fluorescence (G).

Interestingly, C1q deposition was extensively present in ccRCC tissue samples and co-localized with PTX3 (Figure 4A–4D). On the other hand, the deposition of MBL, the first protein in the lectin pathway of complement cascade activation, was absent in ccRCC (data not shown).

To validate the complete activation of the complement cascade, we next evaluated the tissue deposition of the terminal complement complex, C5b-9. Surprisingly, the increased activation of the classical pathway of the complement cascade did not correspond to an increased deposition of the terminal complement complex. Indeed, C5b-9 specific immunofluorescence was completely absent in the renal cancer tissue (Figure 4E–4H).

The complement system is characterized by several regulatory proteins that can inhibit the activation of the enzymatic cascade at different levels [26]. CD59 is one of such inhibitors that can prevent C5b-9 assembly [28]. Interestingly, an increased CD59 expression has been reported in several neoplasia [29], although no information is available on the level of CD59 expression within neoplastic renal tissue. Noteworthy, CD59 protein expression was markedly increased in RCC (Figure 4I–4L). Since anaphylotoxins were suggested as possible soluble mediators modulating both cancer cell proliferation and neoplastic angiogenesis [30], we investigated the protein expression of C3a and C5a receptors. The expression of both trans-membrane proteins was markedly increased in ccRCC tissues (Figure 4M–4P and 4Q–4T). Remarkably, the expression of CD59, C3aR and C5aR co-localized with PTX3 expression within the renal cancer tissue samples (Figures 4L, 4P and 4T, respectively).

### **PTX3 serum levels as biomarkers of ccRCC**

In the attempt to validate the role of PTX3 as a potential biomarker of renal cancer, we retrospectively investigated baseline serum levels of PTX3 in a cohort of 168 consecutive patients undergoing nephrectomy for ccRCC.

The main clinical features of the entire study population of 168 RCC patients, as well as the pathological features of cancers diagnosed are summarized in Table 1.

At time of nephrectomy, PTX3 serum levels were significantly higher in patients with ccRCC as compared with non-neoplastic patients ( $p < 0.0001$ ; Figure 5A).

Moreover, at time of diagnosis significantly higher PTX3 serum levels were observed in patients with

higher Fuhrman grade (G3-4 vs G1-2  $p < 0.01$ ; Figure 5B), with lymph node involvement (N1 vs N0  $p < 0.0001$ ; Figure 5C), and with visceral metastases (M1 vs M0  $p < 0.001$ ; Figure 5D).

A ROC curve analysis was carried out to further validate the association of PTX3 serum levels with the cancer-specific survival. The analysis showed that PTX3 serum levels were significantly associated with ccRCC-specific survival (AUC:0.83,  $p < 0.0001$ ) and identified a cutoff value of 165.0 pg/mL with an 86% (95%CI: 73.3-94.2) specificity and a 70.7% (95%CI: 60.7-79.4) sensitivity.

Survival analysis was performed after the assignment of all patients to two groups according to the operational cut-off of PTX3. ccRCC patients with baseline serum PTX3 levels  $< 165.0$  pg/ml showed significantly higher 10-year rate of overall survival, as compared with ccRCC patients with serum PTX3 levels  $> 165.0$  pg/ml (73.7% vs. 48.4%,  $p = 0.002$ ; Figure 6).

To estimate the relative risk for cancer-specific survival (CSS) and progression-free survival in ccRCC patients showing serum PTX3 level above or below the operational cut-off, Cox regression analyses were performed using cancer-related death and recurrence as dependent variables, and T stage (T3-4 vs. T1-2), lymph node invasion (N+ vs N0), metastatic disease (M+ vs M0), Fuhrman grade (G1-2 vs. G3-4), tumor necrosis and size, and PTX3 serum levels as covariates (Table 2).

KaplanMeier survival curves for cancer-specific survival (CSS) and progression-free survival (PFS), stratified by PTX3 serum levels, are shown in Figure 6. Both CSS and PFS were significantly decreased in patients with high levels of PTX3. Univariate analysis for the predefined variables showed that the pathological stage, presence of nodal and visceral metastases, Fuhrman grade, presence of necrosis, tumor size, and high levels of PTX3, were significantly associated with the risk of death (Table 2) and progression (Table 3). At multivariate analysis by Cox regression modeling, the pathological stage, presence of nodal and visceral metastases, Fuhrman grade, and high serum levels of PTX3, were independent adverse prognostic factors for CSS and PFS (Tables 2 and 3).

## **DISCUSSION**

The results of the present study suggest that serum concentration of PTX3, a recognized modulator of complement system cascade, might represent a reliable non-invasive biomarker for the diagnosis and prognosis of renal cancer. Moreover, we show that an increased

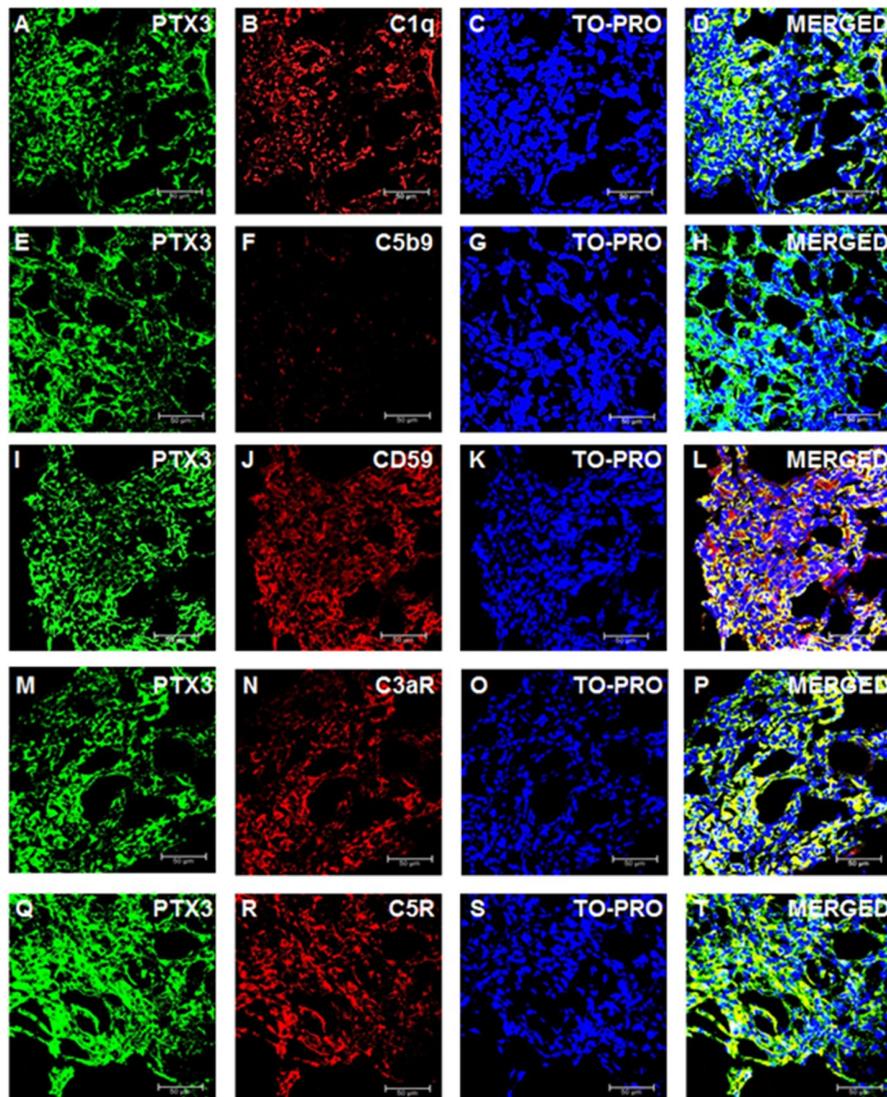
production of PTX3 within the renal tumour may modulate the immunoflogosis in the ccRCC micro-environment, by partially activating the classical pathway of complement system (C1q) and releasing pro-angiogenic factors (C3a, C5a), but inhibiting the complement-mediated cellular lysis due to local up-regulation of CD59.

We showed for the first time an increased expression of PTX3 on both renal cancer cells and ccRCC-derived tissues. Several reports analyzing PTX3 overexpressing cells have suggested that this long pentraxin may exert a pro-tumorigenic role by promoting tumor cell migration and invasion (cervical cancer, head and neck tumors) or proliferation (glioma), epithelial-to-mesenchymal transition (hepatocellular carcinoma) and macrophage chemotaxis [31–34]. PTX3 was also found to be an

oncogenic phosphoinositide 3-kinase signaling critical target, involved in promotion of stem cell-like traits in basal-like breast cancers [35].

In previous observations, our research group demonstrated that an increased expression of PTX3 was related to diagnosis of prostatic cancer and to an increased risk of prostate cancer development, if assessed in patients undergoing prostate biopsy [36].

Our observations are in line with previous evidences, which indicate that PTX3 could modulate the tumor microenvironment and could be a local or systemic marker of cancer-related inflammation. Increased PTX3 gene expression was reported in ovarian cancer with stromal signature [37], in aggressive breast cancer with distant bone metastases [38–40], in prostate



**Figure 4. Complement system factors' expression and co-localization with PTX3 in renal clear cell carcinoma.** Intra-tumoral expression of PTX3 (green) and co-localization with C1q (A-D), C5b9 (E-H), CD59 (I-L), C3aR (M-P), C5R (Q-T).

**Table 1. Clinical and pathological characteristics of patients who underwent radical or partial nephrectomy for ccRCC.**

Variable	n=168
Age (years)	
median (range)	62 (26-85)
Gender, n (%)	
Male	110 (65.5%)
Female	58 (34.5%)
Dimensions (cm)	
median (range)	5.0 (3 -12)
Pathological stage (TNM/AJCC), n (%)	
pT1	104 (62%)
pT2	24 (14%)
pT3	35 (21%)
pT4	5 (3%)
pN+	34 (20.2%)
cM+	30 (17.8%)
Fuhrman grade, n (%)	
G1-2	101 (60%)
G3-4	67 (40%)

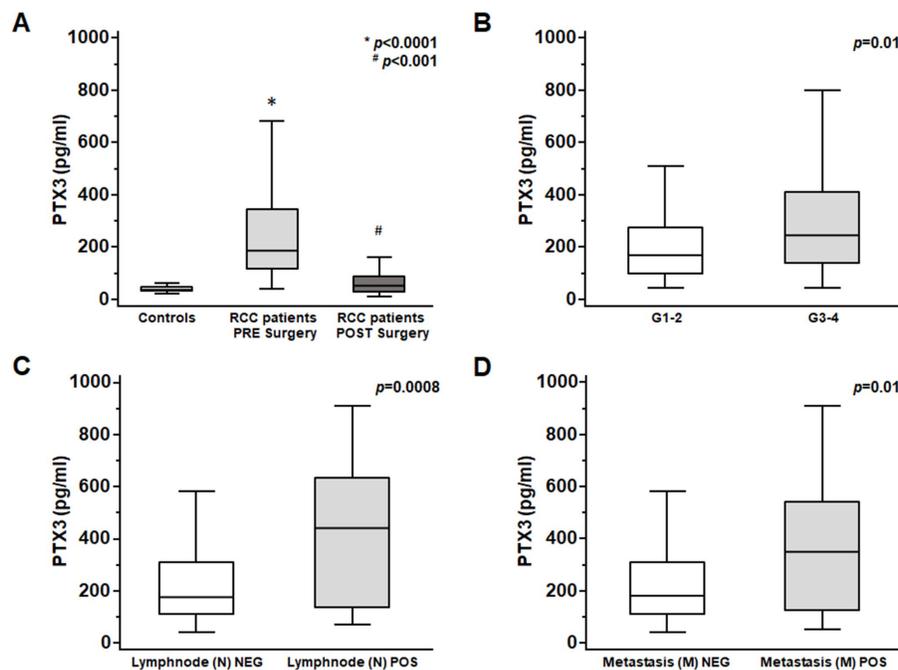
Abbreviations: TNM/AJCC: Tumor size, Lymph Nodes affected, Metastases/American Joint Committee on Cancer.

cancer [41], in glioblastoma [42], in anaplastic thyroid carcinoma [43], and in soft tissue liposarcoma [44].

In this context, the close relationship between PTX3 overexpression, tumour microenvironment modulation and complement cascade activation may represent an interesting pathogenic mechanisms, although experi-

mental data on the activation of the complement cascade in renal cancer are limited. PTX3 has been shown to bind C1q as well as MBL inducing the activation of the complement cascade [45].

In our setting, we demonstrated a clear co-localization with C1q, but not with MBL suggesting a local



**Figure 5.** PTX3 serum levels at baseline in patients with renal clear cell carcinoma before and after surgery ( A) and at different Fuhrman grading ( B), lymphnode involvement ( C) and metastasis staging ( D).

activation of the complement system through the classical pathway. Interestingly, the activation of the complement system in our setting did not lead to the formation of the terminal complement complex with the subsequent lysis of the neoplastic cells. As in most neoplasia, transformed cells may activate several mechanisms to escape complement-dependent lysis. In particular, the production of CD59 or protectin, one of the main inhibitors of C5b-9 assembly ubiquitously expressed at low levels in normal conditions is significantly increased in many tumors [46]. Indeed, we observed a clear overexpression of this protein in the renal cancer tissues. Thus our observation strongly supports the hypothesis that the inhibition of the complement cascade may play a key role in the escape from immunosurveillance of neoplastic cells and might

represent a crucial step in the development of clinically evident neoplastic disease. The factors affecting CD59 expression in this setting remain to be clarified, although the inflammatory milieu may represent the answer also for this event, since Bjorge et al. reported that two key pro-inflammatory cytokines, interleukin-1 (IL-1) and tumor necrosis factor alpha, induce the expression of CD59 in human colonic adenocarcinoma cells [47].

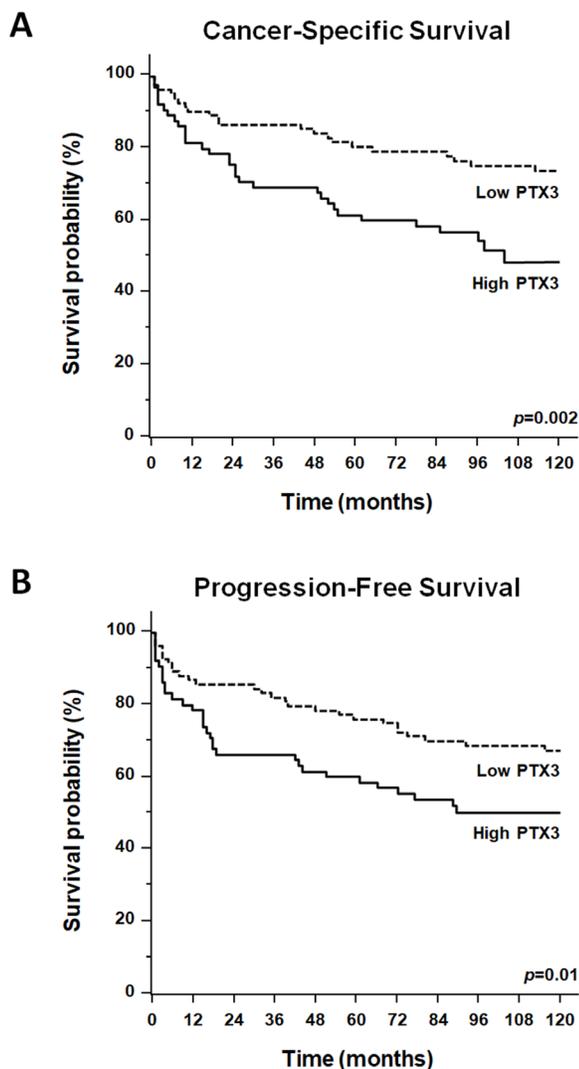
More intriguingly, in our study, the increased expression of PTX3 is associated with a partial activation of the classical pathway of the Complement system with an overexpression of C1q and the receptors of C3a and C5a. In renal cancer, as well as in other tumors, the Complement system seems to emerge as a major regulator of cancer immunity. Complement effectors such as C1q, anaphylatoxins C3a and C5a, and their receptors C3aR and C5aR1, have been associated with tolerogenic cell death and inhibition of antitumor T-cell responses through the recruitment and/or activation of immunosuppressive cell subpopulations such as myeloid-derived suppressor cells (MDSCs), regulatory T cells (Tregs), or M2 tumor-associated macrophages (TAMs) [48–50].

Moreover, the receptor of anaphylatoxin C5a (C5aR) has been strongly detected in metastatic renal cell carcinoma and seems to play a crucial role in cell invasion via the ERK and PI3 kinase pathways [51]. In another tumor, C3aR expression seems to contribute to melanoma carcinogenesis through the inhibition of neutrophils and CD4+T cell response [52]. In recent observations, the anaphylatoxins C3a and C5a seems to significantly contribute to cancer-related inflammation, recruiting myeloid suppressor cells, and promoting IL-1 $\beta$  and IL-17 response in neutrophils, thus enhancing colon carcinogenesis [53–56].

In our setting, both C3a and C5a receptors were dramatically up-regulated in the renal cancer tissue, supporting the hypothesis that the two soluble modulators available *in situ* after the activation of the complement cascade may play a direct or indirect effect on resident cells to sustain carcinogenesis.

The analysis of PTX3 serum levels before nephrectomy revealed that their levels were significantly higher in patients with ccRCC as compared with non-neoplastic patients. Intriguingly, after nephrectomy PTX3 levels significantly lowered, thus strengthening the relationship between intra-tumor PTX3 production and PTX3 serum levels.

When compared to histologic and clinical grading, the basal levels of PTX3 resulted significantly higher in



**Figure 6.** Kaplan-Meier estimate of 12-year cancer-specific survival (CSS: **A**) and progression-free survival (PFS: **B**) of ccRCC patients according to different PTX3 serum levels at baseline.

**Table 2. Univariate and multivariate analyses for cancer-specific survival.**

Variable	Category	Univariate analysis				Multivariate analysis			
		HR	CI 95%		p value	HR	CI 95%		p value
			Lower	Higher			Lower	Higher	
T stage	T3-4 vs T1-2	2.09	1.59	2.75	<b>0.0001</b>	1.56	1.14	2.13	<b>0.004</b>
N stage	N+ vs N0	3.49	1.84	6.60	<b>0.001</b>	1.25	1.01	2.68	<b>0.02</b>
M stage	M+ vs M0	6.15	3.44	10.97	<b>0.0001</b>	4.77	2.05	9.12	<b>0.003</b>
Grade	G3-4 vs G1-2	2.24	1.15	6.86	<b>0.01</b>	1.41	1.05	2.08	<b>0.02</b>
Necrosis	Yes vs No	2.06	1.18	3.86	<b>0.01</b>	-	-	-	-
Tumor size	Continuous	1.48	1.12	2.26	<b>0.01</b>	-	-	-	-
PTX3	>165.0 vs <165.0 pg/ml	2.41	1.21	6.31	<b>0.001</b>	1.86	1.05	2.86	<b>0.01</b>

CI: confidence interval; HR: hazard ratio.

**Table 3. Univariate and multivariate analyses for progression-free survival.**

Variable	Category	Univariate analysis				Multivariate analysis			
		HR	CI 95%		p value	HR	CI 95%		p value
			Lower	Higher			Lower	Higher	
T stage	T3-4 vs T1-2	2.27	1.76	2.93	<b>0.0001</b>	1.60	1.19	2.16	<b>0.001</b>
N stage	N+ vs N0	2.38	1.96	7.24	<b>0.001</b>	1.22	1.02	2.74	<b>0.03</b>
M stage	M+ vs M0	7.19	4.33	12.87	<b>0.0001</b>	5.26	2.44	11.42	<b>0.001</b>
Grade	G3-4 vs G1-2	2.05	1.54	2.73	<b>0.01</b>	1.51	1.09	2.11	<b>0.01</b>
Necrosis	Yes vs No	1.95	1.08	3.34	<b>0.01</b>	-	-	-	-
Tumor size	Continuous	1.34	1.02	2.96	<b>0.01</b>	-	-	-	-
PTX3	>165.0 vs <165.0 pg/ml	2.01	1.31	7.41	<b>0.01</b>	1.56	1.03	2.95	<b>0.01</b>

CI: confidence interval; HR: hazard ratio.

patients with higher Furhman grading (G3-4) and with both lymph nodes positive distant metastases already present at time of diagnosis, thus suggesting a possible role of PTX not only as diagnostic marker but also as disease severity parameter.

Lastly, if related to patient survival, higher PTX3 serum levels at time of nephrectomy were associated with a significantly lower long-term survival, and shorter time to progression as shown by the Kaplan Meyer curves and confirmed by Cox regression analysis. Data from the cancer genome atlas (TCGA) clear cell renal cell carcinoma patient cohort (KIRC), confirmed our findings showing a reduced survival in patients with high expression levels of PTX3 (Supplementary Figure 1).

Our data seems to be in line with previous studies in other clinical settings. Increased circulating levels of PTX3 were observed in myeloproliferative neoplasms [57], lung cancers [58, 59], soft tissue sarcomas [60], gliomas [61], pancreatic and hepatocellular carcinomas [62, 63]. Moreover, high PTX3 levels were associated with advanced clinical stage and poor overall survival of patients with pancreatic carcinoma [61].

Taken together, our data support the potential role of serum PTX3 as a diagnostic and prognostic marker of ccRCC. Moreover, the strong involvement of complement system in the ccRCC microenvironment strongly support the idea that PTX3 up-regulation modulates the effector routes associated with the cancer-immunity cycle, providing the rationale for new therapeutic combinations aimed to enhance the antitumor efficacy of anti-PD-1/PD-L1 checkpoint inhibitors in this neoplasia.

Our study limits are the monocentric retrospective analysis and the rather limited number of cases. However, further prospective multicenter studies are warranted to confirm our observations.

Taken together, our results suggest that expression of PTX3 can modulate the immunoflogosis in the ccRCC microenvironment, by activating the classical pathway of complement system (C1q) and releasing pro-angiogenic factors (C3a, C5a). The up-regulation of CD59 also inhibits the complement-mediated cellular lysis. Moreover, the finding of elevated serum PTX3 levels in the ccRCC patient before nephrectomy suggests its potential role as biomarker of ccRCC diagnosis and prognosis.

## MATERIALS AND METHODS

### Gene set enrichment analysis (GSEA)

Clear cell-RCC transcriptome data derived from exon array analysis of 20 total samples (10 ccRCC tumor sample and their matched non-tumor kidney tissues samples) were used. Exon array data are deposited in GEO at Series accession number GSE47032. GSEA [64] was used to determine which pathways were statistically enriched across the renal cancer dataset. The normalized enrichment score (NES) was used to evaluate the extent and direction of enrichment of each pathway.

### Analysis of biological networks

Pathway analysis was performed using Ingenuity Pathway Analysis (IPA; Qiagen). The data were obtained from the GSE47032 array and the gene IDs and fold-changes were imported into IPA software. Gene symbols were mapped to their corresponding gene object in the Ingenuity Pathways Knowledge Base (IKB). The networks identified are presented in maps showing interactions between genes. Genes are represented as nodes in the networks. The intensity of the node color indicates the degree of up- or down-regulation (upregulation in red, downregulation in green). Canonical pathway analysis was used to identify the signaling pathways, which were most significant in the analyzed data set.

### Cell lines

Three different tumor renal cell lines (RCC-SHAW, RCCBA85#21, primary RCC cells) were tested and cultured in a Roswell Park Memorial Institute medium (RPMI) (Sigma Aldrich, Saint Louis, MO USA), supplemented with 10% fetal bovine serum (FBS), 2 mM l-glutamine, and 100 U/ml penicillin–streptomycin (all from Sigma-Aldrich), and incubated for 48 hours at 37 °C, 5% CO<sub>2</sub>, as previously described [65]. As control lines, human HK-2 cells, a proximal tubular epithelial cell line of human origin, were grown in Dulbecco's modified Eagle's medium (DMEM)–F12 medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS), 2 mM l-glutamine, and 100 U/ml penicillin–streptomycin (all from Sigma-Aldrich) at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub> [64].

### Study population and tissue collection

30 primary renal tumors were collected from patients who underwent nephrectomy for ccRCC at Urology Unit of University Hospital “Ospedali Riuniti” of Foggia. Detailed clinical and pathological characteristics of the

patients are summarized in Supplementary Table 1. A control group of 10 subject underwent renal biopsy at Nephrology Unit of University Hospital “Ospedali Riuniti” of Foggia in suspicion of chronic nephropathy, but with a normal renal histology, were also analyzed.

Two pathologists confirmed the presence of ccRCC in the neoplastic tissues and excluded tumor cells in the healthy specimens. Tumor and normal tissues were collected, frozen at –80°C according to a standard procedure and stored.

In addition, serum samples were collected from 168 patients who underwent radical or partial nephrectomy for ccRCC and 40 volunteers with no evidence of malignancy at Urology Unit of University Hospital “Policlinico” of Bari. Serum samples were obtained from each patient at the time of nephrectomy and stored at –30°C. Detailed clinical and pathological characteristics of the patients are summarized in Table 1. All patients were preoperatively staged by thoraco-abdominal Computed Tomography or Magnetic Resonance Imaging. Tumor staging was reassigned according to the seventh edition of the AJCC-UICC TNM classification. The 2016 World Health Organization and Fuhrman classifications were used to attribute histological type and nuclear grade, respectively. Written informed consent to take part was given by all participants. The protocol for the research project has been approved by the local Ethics Committee (Decision n. 152/CE/2014 of September 03, 2014; Ethical Committee at the University Hospital “Ospedali Riuniti” of Foggia) and conforms to the guidelines laid down by the Regional Ethics Committee on human experimentation and to the provisions of the Declaration of Helsinki in 1995.

### PTX-3 serum level assessment

PTX-3 serum levels were tested on serum samples drawn at the time of nephrectomy in the whole study population. Circulating PTX3 was measured was assayed using a commercially available ELISA Kit, according to the manufacturer's instructions (R&D Systems, Minneapolis, MN), as previously described [66, 67].

### Indirect immunofluorescence and confocal laser scanning microscopy

A double-label immunofluorescence was performed to evaluate the expression of PTX-3, C1q, MBL, C3aR, C5R1, C5b-9 and CD59 and their eventual co-localization. To this purpose we employed the following primary antibodies: rat monoclonal IgG2a anti-PTX-3 antibody (clone MNB4, Abcam, Cambridge UK),

mouse monoclonal IgG2b anti-C1q (clone JL-1; Abcam); rabbit monoclonal IgG anti-Mannose Binding Lectin (anti-MBL) (clone EPSISR5; Abcam); rabbit polyclonal IgG anti-C3aR (Abcam); mouse monoclonal IgG2a anti-C5R1/CD88 (clone P12/1; Abcam); mouse monoclonal IgG2a anti-C5b-9 (clone aE11; Abcam); rabbit polyclonal IgG anti-CD59 (Sigma-Merck KGaA, Darmstadt, Germany).

Frozen tissue sections were incubated at 4°C overnight with a mixture of primary antibodies diluted 1:100 in PBS pH 7.4. The immune complexes were detected using the Alexa-Fluor 488 goat anti-rat and 546 goat anti-mouse IgG and 546 goat anti-rabbit IgG (all from Alexa, Thermo Fisher, Waltham, MA).

After washing in PBS (3x5') the sections and the negative control were incubated 1h at room temperature with goat anti-rat IgG 488 and goat anti-mouse IgG 546 or goat anti-rabbit IgG 546, as appropriate. All secondary antibodies were used at a dilution of 1:250.

To stain the nuclei, after washing in PBS pH 7.4 (3x5') samples were incubated with TO-PRO diluted 1:5000 in PBS pH 7.4 (Invitrogen-Molecular Probe, Thermo Fisher, Waltham, MA). The slides were mounted in Gel Mount (Sigma) and sealed.

Specific fluorescence was evaluated by confocal microscopy using the Leica TCS SP5 (Leica, Wetzlar, Germany) equipped with argon-krypton (488 nm), green-neon (543 nm), and helium-neon (633 nm) lasers. Fluorescence quantification was performed as previously described [68, 69].

### Statistical analysis

Statistical analysis was performed as described elsewhere [69, 70, 71]. In detail, statistical calculations were performed with MedCalc 9.2.0.1 (MedCalc software, Mariakerke, Belgium) and PASW 18 software (PASW 18, SPSS, Chicago, Ill, USA). Comparisons of median protein values between different groups were evaluated by Mann–Whitney U test. Receiver Operating Characteristic (ROC) curve analysis was performed to identify the PTX3 cut-off for survival stratification.

In the cancer-specific survival (CSS) analysis, patients who died of RCC unrelated causes or were lost to follow-up were censored. Progression-free survival (PFS) was calculated from the date of surgery to the date of disease recurrence. Estimates of CSS and PFS were calculated according to the Kaplan–Meier method and compared with the log-rank test. Univariate and multivariate analyses were performed using the Cox proportional hazards regression model to identify the

most significant variables for predicting CSS and PFS. A backward selection procedure was performed with removal criterion  $P > 0.10$  based on likelihood ratio tests. A P-value of  $< 0.05$  was considered statistically significant.

### Ethics approval

The present study involving human participants was approved by the local ethical committee (Decision n. 152/CE/2014 of September 03, 2014; Ethical Committee at the University Hospital “Ospedali Riuniti” of Foggia). All procedures performed in the present study were in accordance with the ethical standards of the Declaration of Helsinki and all the enrolled patients provided an informed consent to participate in the present study.

### AUTHOR CONTRIBUTIONS

Giuseppe Stefano Netti and Giuseppe Lucarelli conceived and designed the study, analyzed the data and drafted the manuscript; Federica Spadaccino, Margherita Gigante, Chiara Divella and Federica Rascio performed the experiments and interpreted results of experiments; Maria Teresa Rocchetti and Vito Mancini collected the clinical data and helped to interpret the results; Giuseppe Castellano and Giovanni Stallone analyzed the data, interpreted results and prepared the figures; Giuseppe Carrieri, Loreto Gesualdo and Michele Battaglia helped to draft the manuscript; Elena Ranieri edited and revised manuscript and approved the final version of manuscript

### CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### REFERENCES

1. Cohen HT, McGovern FJ. Renal-cell carcinoma. *N Engl J Med.* 2005; 353:2477–90. <https://doi.org/10.1056/NEJMra043172> PMID:[16339096](https://pubmed.ncbi.nlm.nih.gov/16339096/)
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:[31912902](https://pubmed.ncbi.nlm.nih.gov/31912902/)
3. Battaglia M, Lucarelli G. The role of renal surgery in the era of targeted therapy: the urologist's perspective. *Urologia.* 2015; 82:137–38. <https://doi.org/10.5301/uro.5000105> PMID:[25704789](https://pubmed.ncbi.nlm.nih.gov/25704789/)

4. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng DY, Larkin J, Ficarra V. Renal cell carcinoma. *Nat Rev Dis Primers*. 2017; 3:17009. <https://doi.org/10.1038/nrdp.2017.9> PMID:[28276433](https://pubmed.ncbi.nlm.nih.gov/28276433/)
5. de Martino M, Gigante M, Cormio L, Praticchizzo C, Cavalcanti E, Gigante M, Ariano V, Netti GS, Montemurno E, Mancini V, Battaglia M, Gesualdo L, Carrieri G, Ranieri E. JAK3 in clear cell renal cell carcinoma: mutational screening and clinical implications. *Urol Oncol*. 2013; 31:930–37. <https://doi.org/10.1016/j.urolonc.2011.07.001> PMID:[21868263](https://pubmed.ncbi.nlm.nih.gov/21868263/)
6. Lucarelli G, Ditunno P, Bettocchi C, Vavallo A, Rutigliano M, Galleggiante V, Larocca AM, Castellano G, Gesualdo L, Grandaliano G, Selvaggi FP, Battaglia M. Diagnostic and prognostic role of preoperative circulating CA 15-3, CA 125, and beta-2 microglobulin in renal cell carcinoma. *Dis Markers*. 2014; 2014:689795. <https://doi.org/10.1155/2014/689795> PMID:[24692843](https://pubmed.ncbi.nlm.nih.gov/24692843/)
7. Gigante M, Lucarelli G, Divella C, Netti GS, Pontrelli P, Cafiero C, Grandaliano G, Castellano G, Rutigliano M, Stallone G, Bettocchi C, Ditunno P, Gesualdo L, et al. Soluble Serum  $\alpha$ Klotho Is a Potential Predictive Marker of Disease Progression in Clear Cell Renal Cell Carcinoma. *Medicine (Baltimore)*. 2015; 94:e1917. <https://doi.org/10.1097/MD.0000000000001917> PMID:[26559258](https://pubmed.ncbi.nlm.nih.gov/26559258/)
8. Lucarelli G, Galleggiante V, Rutigliano M, Sanguedolce F, Cagiano S, Bufo P, Lastilla G, Maiorano E, Ribatti D, Giglio A, Serino G, Vavallo A, Bettocchi C, et al. Metabolomic profile of glycolysis and the pentose phosphate pathway identifies the central role of glucose-6-phosphate dehydrogenase in clear cell-renal cell carcinoma. *Oncotarget*. 2015; 6:13371–86. <https://doi.org/10.18632/oncotarget.3823> PMID:[25945836](https://pubmed.ncbi.nlm.nih.gov/25945836/)
9. Lucarelli G, Rutigliano M, Sanguedolce F, Galleggiante V, Giglio A, Cagiano S, Bufo P, Maiorano E, Ribatti D, Ranieri E, Gigante M, Gesualdo L, Ferro M, et al. Increased Expression of the Autocrine Motility Factor is Associated With Poor Prognosis in Patients With Clear Cell-Renal Cell Carcinoma. *Medicine (Baltimore)*. 2015; 94:e2117. <https://doi.org/10.1097/MD.0000000000002117> PMID:[26579829](https://pubmed.ncbi.nlm.nih.gov/26579829/)
10. Papale M, Vocino G, Lucarelli G, Rutigliano M, Gigante M, Rocchetti MT, Pesce F, Sanguedolce F, Bufo P, Battaglia M, Stallone G, Grandaliano G, Carrieri G, et al. Urinary RKIP/p-RKIP is a potential diagnostic and prognostic marker of clear cell renal cell carcinoma. *Oncotarget*. 2017; 8:40412–24. <https://doi.org/10.18632/oncotarget.16341> PMID:[28418894](https://pubmed.ncbi.nlm.nih.gov/28418894/)
11. Lucarelli G, Rutigliano M, Ferro M, Giglio A, Intini A, Triggiano F, Palazzo S, Gigante M, Castellano G, Ranieri E, Buonerba C, Terracciano D, Sanguedolce F, et al. Activation of the kynurenine pathway predicts poor outcome in patients with clear cell renal cell carcinoma. *Urol Oncol*. 2017; 35:461.e15–27. <https://doi.org/10.1016/j.urolonc.2017.02.011> PMID:[28359744](https://pubmed.ncbi.nlm.nih.gov/28359744/)
12. Lucarelli G, Loizzo D, Franzin R, Battaglia S, Ferro M, Cantiello F, Castellano G, Bettocchi C, Ditunno P, Battaglia M. Metabolomic insights into pathophysiological mechanisms and biomarker discovery in clear cell renal cell carcinoma. *Expert Rev Mol Diagn*. 2019; 19:397–407. <https://doi.org/10.1080/14737159.2019.1607729> PMID:[30983433](https://pubmed.ncbi.nlm.nih.gov/30983433/)
13. Lucarelli G, Rutigliano M, Sallustio F, Ribatti D, Giglio A, Lepore Signorile M, Grossi V, Sanese P, Napoli A, Maiorano E, Bianchi C, Perego RA, Ferro M, et al. Integrated multi-omics characterization reveals a distinctive metabolic signature and the role of NDUFA4L2 in promoting angiogenesis, chemoresistance, and mitochondrial dysfunction in clear cell renal cell carcinoma. *Aging (Albany NY)*. 2018; 10:3957–85. <https://doi.org/10.18632/aging.101685> PMID:[30538212](https://pubmed.ncbi.nlm.nih.gov/30538212/)
14. Hunt JD, van der Hel OL, McMillan GP, Boffetta P, Brennan P. Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. *Int J Cancer*. 2005; 114:101–08. <https://doi.org/10.1002/ijc.20618> PMID:[15523697](https://pubmed.ncbi.nlm.nih.gov/15523697/)
15. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008; 371:569–78. [https://doi.org/10.1016/S0140-6736\(08\)60269-X](https://doi.org/10.1016/S0140-6736(08)60269-X) PMID:[18280327](https://pubmed.ncbi.nlm.nih.gov/18280327/)
16. Weikert S, Boeing H, Pischon T, Weikert C, Olsen A, Tjønneland A, Overvad K, Becker N, Linseisen J, Trichopoulou A, Mountokalakis T, Trichopoulos D, Sieri S, et al. Blood pressure and risk of renal cell carcinoma in the European prospective investigation into cancer and nutrition. *Am J Epidemiol*. 2008; 167:438–46. <https://doi.org/10.1093/aje/kwm321> PMID:[18048375](https://pubmed.ncbi.nlm.nih.gov/18048375/)
17. Vavallo A, Simone S, Lucarelli G, Rutigliano M, Galleggiante V, Grandaliano G, Gesualdo L, Campagna M, Cariello M, Ranieri E, Pertosa G, Lastilla G, Selvaggi FP, et al. Pre-existing type 2 diabetes mellitus is an

- independent risk factor for mortality and progression in patients with renal cell carcinoma. *Medicine* (Baltimore). 2014; 93:e183.  
<https://doi.org/10.1097/MD.0000000000000183>  
PMID:[25501064](https://pubmed.ncbi.nlm.nih.gov/25501064/)
18. Breda A, Lucarelli G, Rodriguez-Faba O, Guirado L, Facundo C, Bettocchi C, Gesualdo L, Castellano G, Grandaliano G, Battaglia M, Palou J, Ditunno P, Villavicencio H. Clinical and pathological outcomes of renal cell carcinoma (RCC) in native kidneys of patients with end-stage renal disease: a long-term comparative retrospective study with RCC diagnosed in the general population. *World J Urol*. 2015; 33:1–7.  
<https://doi.org/10.1007/s00345-014-1248-y>  
PMID:[24504760](https://pubmed.ncbi.nlm.nih.gov/24504760/)
  19. Di Lorenzo G, De Placido S, Pagliuca M, Ferro M, Lucarelli G, Rossetti S, Bosso D, Puglia L, Pignataro P, Ascione I, De Cobelli O, Caraglia M, Aieta M, et al. The evolving role of monoclonal antibodies in the treatment of patients with advanced renal cell carcinoma: a systematic review. *Expert Opin Biol Ther*. 2016; 16:1387–401.  
<https://doi.org/10.1080/14712598.2016.1216964>  
PMID:[27463642](https://pubmed.ncbi.nlm.nih.gov/27463642/)
  20. Tamma R, Rutigliano M, Lucarelli G, Annese T, Ruggieri S, Cascardi E, Napoli A, Battaglia M, Ribatti D. Microvascular density, macrophages, and mast cells in human clear cell renal carcinoma with and without bevacizumab treatment. *Urol Oncol*. 2019; 37:355.e11–19.  
<https://doi.org/10.1016/j.urolonc.2019.01.025>  
PMID:[30738745](https://pubmed.ncbi.nlm.nih.gov/30738745/)
  21. Markiewski MM, Lambris JD. Is complement good or bad for cancer patients? A new perspective on an old dilemma. *Trends Immunol*. 2009; 30:286–92.  
<https://doi.org/10.1016/j.it.2009.04.002>  
PMID:[19428302](https://pubmed.ncbi.nlm.nih.gov/19428302/)
  22. Loveland BE, Cebon J. Cancer exploiting complement: a clue or an exception? *Nat Immunol*. 2008; 9:1205–06.  
<https://doi.org/10.1038/ni1108-1205>  
PMID:[18936777](https://pubmed.ncbi.nlm.nih.gov/18936777/)
  23. Ostrand-Rosenberg S. Immune surveillance: a balance between protumor and antitumor immunity. *Curr Opin Genet Dev*. 2008; 18:11–18.  
<https://doi.org/10.1016/j.gde.2007.12.007>  
PMID:[18308558](https://pubmed.ncbi.nlm.nih.gov/18308558/)
  24. Garlanda C, Bottazzi B, Magrini E, Inforzato A, Mantovani A. PTX3, a humoral pattern recognition molecule, in innate immunity, tissue repair, and cancer. *Physiol Rev*. 2018; 98:623–39.  
<https://doi.org/10.1152/physrev.00016.2017>  
PMID:[29412047](https://pubmed.ncbi.nlm.nih.gov/29412047/)
  25. Ronca R, Di Salle E, Giacomini A, Leali D, Alessi P, Coltrini D, Ravelli C, Matarazzo S, Ribatti D, Vermi W, Presta M. Long pentraxin-3 inhibits epithelial-mesenchymal transition in melanoma cells. *Mol Cancer Ther*. 2013; 12:2760–71.  
<https://doi.org/10.1158/1535-7163.MCT-13-0487>  
PMID:[24130051](https://pubmed.ncbi.nlm.nih.gov/24130051/)
  26. Giacomini A, Ghedini GC, Presta M, Ronca R. Long pentraxin 3: A novel multifaceted player in cancer. *Biochim Biophys Acta Rev Cancer*. 2018; 1869:53–63.  
<https://doi.org/10.1016/j.bbcan.2017.11.004>  
PMID:[29175552](https://pubmed.ncbi.nlm.nih.gov/29175552/)
  27. Rutkowski MJ, Sughrue ME, Kane AJ, Mills SA, Parsa AT. Cancer and the complement cascade. *Mol Cancer Res*. 2010; 8:1453–65.  
<https://doi.org/10.1158/1541-7786.MCR-10-0225>  
PMID:[20870736](https://pubmed.ncbi.nlm.nih.gov/20870736/)
  28. Mamidi S, Höne S, Kirschfink M. The complement system in cancer: ambivalence between tumour destruction and promotion. *Immunobiology*. 2017; 222:45–54.  
<https://doi.org/10.1016/j.imbio.2015.11.008>  
PMID:[26686908](https://pubmed.ncbi.nlm.nih.gov/26686908/)
  29. Macor P, Tripodo C, Zorzet S, Piovan E, Bossi F, Marzari R, Amadori A, Tedesco F. In vivo targeting of human neutralizing antibodies against CD55 and CD59 to lymphoma cells increases the antitumor activity of rituximab. *Cancer Res*. 2007; 67:10556–63.  
<https://doi.org/10.1158/0008-5472.CAN-07-1811>  
PMID:[17975000](https://pubmed.ncbi.nlm.nih.gov/17975000/)
  30. Ajona D, Ortiz-Espinosa S, Pio R. Complement anaphylatoxins C3a and C5a: emerging roles in cancer progression and treatment. *Semin Cell Dev Biol*. 2017; 17:30375–86.  
<https://doi.org/10.1016/j.semcdb.2017.11.023>  
PMID:[29155219](https://pubmed.ncbi.nlm.nih.gov/29155219/)
  31. Chang WC, Wu SL, Huang WC, Hsu JY, Chan SH, Wang JM, Tsai JP, Chen BK. PTX3 gene activation in EGF-induced head and neck cancer cell metastasis. *Oncotarget*. 2015; 6:7741–57.  
<https://doi.org/10.18632/oncotarget.3482>  
PMID:[25797258](https://pubmed.ncbi.nlm.nih.gov/25797258/)
  32. Song T, Wang C, Guo C, Liu Q, Zheng X. Pentraxin 3 overexpression accelerated tumor metastasis and indicated poor prognosis in hepatocellular carcinoma via driving epithelial-mesenchymal transition. *J Cancer*. 2018; 9:2650–58.  
<https://doi.org/10.7150/jca.25188>  
PMID:[30087705](https://pubmed.ncbi.nlm.nih.gov/30087705/)
  33. Choi B, Lee EJ, Park YS, Kim SM, Kim EY, Song Y, Kang SW, Rhu MH, Chang EJ. Pentraxin-3 silencing suppresses gastric cancer-related inflammation by

- inhibiting chemotactic migration of macrophages. *Anticancer Res.* 2015; 35:2663–68.  
PMID:[25964543](https://pubmed.ncbi.nlm.nih.gov/25964543/)
34. Ying TH, Lee CH, Chiou HL, Yang SF, Lin CL, Hung CH, Tsai JP, Hsieh YH. Knockdown of Pentraxin 3 suppresses tumorigenicity and metastasis of human cervical cancer cells. *Sci Rep.* 2016; 6:29385.  
<https://doi.org/10.1038/srep29385>  
PMID:[27377307](https://pubmed.ncbi.nlm.nih.gov/27377307/)
  35. Thomas C, Henry W, Cuiffo BG, Collmann AY, Marangoni E, Benhamo V, Bhasin MK, Fan C, Fuhrmann L, Baldwin AS, Perou C, Vincent-Salomon A, Toker A, Karnoub AE. Pentraxin-3 is a PI3K signaling target that promotes stem cell-like traits in basal-like breast cancers. *Sci Signal.* 2017; 10:eaah4674.  
<https://doi.org/10.1126/scisignal.aah4674>  
PMID:[28223411](https://pubmed.ncbi.nlm.nih.gov/28223411/)
  36. Stallone G, Cormio L, Netti GS, Infante B, Selvaggio O, Fino GD, Ranieri E, Bruno F, Prattichizzo C, Sanguedolce F, Tortorella S, Bufo P, Grandaliano G, Carrieri G. Pentraxin 3: a novel biomarker for predicting progression from prostatic inflammation to prostate cancer. *Cancer Res.* 2014; 74:4230–38.  
<https://doi.org/10.1158/0008-5472.CAN-14-0369>  
PMID:[24950910](https://pubmed.ncbi.nlm.nih.gov/24950910/)
  37. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, Johnson DS, Trivett MK, Etemadmoghadam D, Locandro B, Traficante N, Fereday S, Hung JA, et al, and Australian Ovarian Cancer Study Group. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res.* 2008; 14:5198–208.  
<https://doi.org/10.1158/1078-0432.CCR-08-0196>  
PMID:[18698038](https://pubmed.ncbi.nlm.nih.gov/18698038/)
  38. Choi B, Lee EJ, Song DH, Yoon SC, Chung YH, Jang Y, Kim SM, Song Y, Kang SW, Yoon SY, Chang EJ. Elevated Pentraxin 3 in bone metastatic breast cancer is correlated with osteolytic function. *Oncotarget.* 2014; 5:481–92.  
<https://doi.org/10.18632/oncotarget.1664>  
PMID:[24457902](https://pubmed.ncbi.nlm.nih.gov/24457902/)
  39. Chen WY, Wu F, You ZY, Zhang ZM, Guo YL, Zhong LX. Analyzing the differentially expressed genes and pathway cross-talk in aggressive breast cancer. *J Obstet Gynaecol Res.* 2015; 41:132–40.  
<https://doi.org/10.1111/jog.12495>  
PMID:[25227254](https://pubmed.ncbi.nlm.nih.gov/25227254/)
  40. Player A, Abraham N, Burrell K, Bengone IO, Harris A, Nunez L, Willaims T, Kwende S, Walls W. Identification of candidate genes associated with triple negative breast cancer. *Genes Cancer.* 2017; 8:659–72.  
<https://doi.org/10.18632/genesandcancer.147>  
PMID:[28966727](https://pubmed.ncbi.nlm.nih.gov/28966727/)
  41. Ravenna L, Sale P, Di Vito M, Russo A, Salvatori L, Tafani M, Mari E, Sentinelli S, Petrangeli E, Gallucci M, Di Silverio F, Russo MA. Up-regulation of the inflammatory-reparative phenotype in human prostate carcinoma. *Prostate.* 2009; 69:1245–55.  
<https://doi.org/10.1002/pros.20966> PMID:[19444819](https://pubmed.ncbi.nlm.nih.gov/19444819/)
  42. Tafani M, Di Vito M, Frati A, Pellegrini L, De Santis E, Sette G, Eramo A, Sale P, Mari E, Santoro A, Raco A, Salvati M, De Maria R, Russo MA. Pro-inflammatory gene expression in solid glioblastoma microenvironment and in hypoxic stem cells from human glioblastoma. *J Neuroinflammation.* 2011; 8:32.  
<https://doi.org/10.1186/1742-2094-8-32>  
PMID:[21489226](https://pubmed.ncbi.nlm.nih.gov/21489226/)
  43. Espinal-Enrriquez J, Muñoz-Montero S, Imaz-Rosshandler I, Huerta-Verde A, Mejía C, Hernández-Lemus E. Genome-wide expression analysis suggests a crucial role of dysregulation of matrix metalloproteinases pathway in undifferentiated thyroid carcinoma. *BMC Genomics.* 2015; 16:207.  
<https://doi.org/10.1186/s12864-015-1372-0>  
PMID:[25887408](https://pubmed.ncbi.nlm.nih.gov/25887408/)
  44. Willeke F, Assad A, Findeisen P, Schromm E, Grobholz R, von Gerstenbergk B, Mantovani A, Peri S, Friess HH, Post S, von Knebel Doeberitz M, Schwarzbach MH. Overexpression of a member of the pentraxin family (PTX3) in human soft tissue liposarcoma. *Eur J Cancer.* 2006; 42:2639–46.  
<https://doi.org/10.1016/j.ejca.2006.05.035>  
PMID:[16959485](https://pubmed.ncbi.nlm.nih.gov/16959485/)
  45. Bonavita E, Gentile S, Rubino M, Maina V, Papait R, Kunderfranco P, Greco C, Feruglio F, Molgora M, Laface I, Tartari S, Doni A, Pasqualini F, et al. PTX3 is an extrinsic oncosuppressor regulating complement-dependent inflammation in cancer. *Cell.* 2015; 160:700–14.  
<https://doi.org/10.1016/j.cell.2015.01.004>  
PMID:[25679762](https://pubmed.ncbi.nlm.nih.gov/25679762/)
  46. Li B, Lin H, Fan J, Lan J, Zhong Y, Yang Y, Li H, Wang Z. CD59 is overexpressed in human lung cancer and regulates apoptosis of human lung cancer cells. *Int J Oncol.* 2013; 43:850–58.  
<https://doi.org/10.3892/ijco.2013.2007> PMID:[23835643](https://pubmed.ncbi.nlm.nih.gov/23835643/)
  47. Bjørge L, Jensen TS, Matre R. Characterisation of the complement-regulatory proteins decay-accelerating factor (DAF, CD55) and membrane cofactor protein (MCP, CD46) on a human colonic adenocarcinoma cell line. *Cancer Immunol Immunother.* 1996; 42:185–92.  
<https://doi.org/10.1007/s002620050269>  
PMID:[8640847](https://pubmed.ncbi.nlm.nih.gov/8640847/)
  48. Pio R, Ajona D, Ortiz-Espinosa S, Mantovani A, Lambris JD. Complementing the Cancer-Immunity Cycle. *Front Immunol.* 2019; 10:774.

<https://doi.org/10.3389/fimmu.2019.00774>

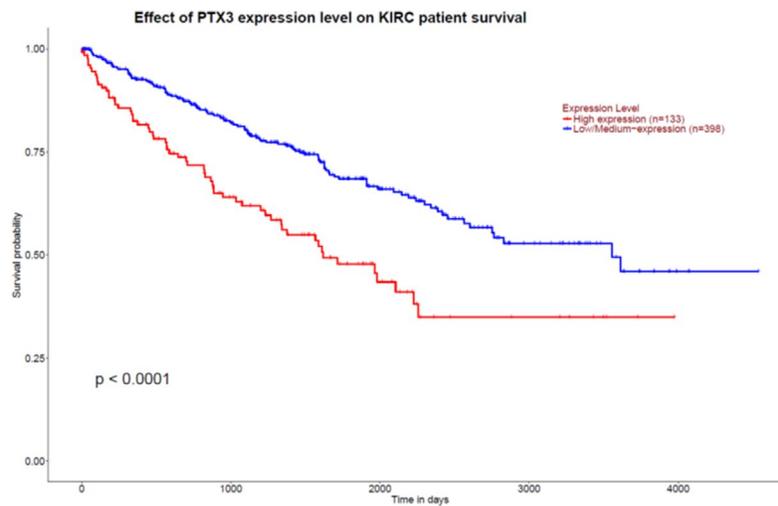
PMID:[31031765](https://pubmed.ncbi.nlm.nih.gov/31031765/)

49. Gigante M, Mandic M, Wesa AK, Cavalcanti E, Dambrosio M, Mancini V, Battaglia M, Gesualdo L, Storkus WJ, Ranieri E. Interferon-alpha (IFN-alpha)-conditioned DC preferentially stimulate type-1 and limit Treg-type in vitro T-cell responses from RCC patients. *J Immunother*. 2008; 31:254–62.  
<https://doi.org/10.1097/CJI.0b013e318167b023>  
PMID:[18317362](https://pubmed.ncbi.nlm.nih.gov/18317362/)
50. Montinaro V, Serra L, Perissutti S, Ranieri E, Tedesco F, Schena FP. Biosynthesis of C3 by human mesangial cells. Modulation by proinflammatory cytokines. *Kidney Int*. 1995; 47:829–36.  
<https://doi.org/10.1038/ki.1995.125>  
PMID:[7538612](https://pubmed.ncbi.nlm.nih.gov/7538612/)
51. Maeda Y, Kawano Y, Wada Y, Yatsuda J, Motoshima T, Murakami Y, Kikuchi K, Imamura T, Eto M. C5aR is frequently expressed in metastatic renal cell carcinoma and plays a crucial role in cell invasion via the ERK and PI3 kinase pathways. *Oncol Rep*. 2015; 33:1844–50.  
<https://doi.org/10.3892/or.2015.3800>  
PMID:[25682807](https://pubmed.ncbi.nlm.nih.gov/25682807/)
52. Nabizadeh JA, Manthey HD, Steyn FJ, Chen W, Widiapradja A, Md Akhir FN, Boyle GM, Taylor SM, Woodruff TM, Rolfe BE. The Complement C3a Receptor Contributes to Melanoma Tumorigenesis by Inhibiting Neutrophil and CD4+ T Cell Responses. *J Immunol*. 2016; 196:4783–92.  
<https://doi.org/10.4049/jimmunol.1600210>  
PMID:[27183625](https://pubmed.ncbi.nlm.nih.gov/27183625/)
53. Markiewski MM, DeAngelis RA, Benencia F, Ricklin-Lichtsteiner SK, Koutoulaki A, Gerard C, Coukos G, Lambris JD. Modulation of the antitumor immune response by complement. *Nat Immunol*. 2008; 9:1225–35.  
<https://doi.org/10.1038/ni.1655>  
PMID:[18820683](https://pubmed.ncbi.nlm.nih.gov/18820683/)
54. Ning C, Li YY, Wang Y, Han GC, Wang RX, Xiao H, Li XY, Hou CM, Ma YF, Sheng DS, Shen BF, Feng JN, Guo RF, et al. Complement activation promotes colitis-associated carcinogenesis through activating intestinal IL-1 $\beta$ /IL-17A axis. *Mucosal Immunol*. 2015; 8:1275–84.  
<https://doi.org/10.1038/mi.2015.18>  
PMID:[25736459](https://pubmed.ncbi.nlm.nih.gov/25736459/)
55. Guglietta S, Rescigno M. Hypercoagulation and complement: connected players in tumor development and metastases. *Semin Immunol*. 2016; 28:578–86.  
<https://doi.org/10.1016/j.smim.2016.10.011>  
PMID:[27876232](https://pubmed.ncbi.nlm.nih.gov/27876232/)
56. Medler TR, Murugan D, Horton W, Kumar S, Cotechini T, Forsyth AM, Leyshock P, Leitenberger JJ, Kulesz-Martin M, Margolin AA, Werb Z, Coussens LM. Complement C5a fosters squamous carcinogenesis and limits T cell response to chemotherapy. *Cancer Cell*. 2018; 34:561–578.e6.  
<https://doi.org/10.1016/j.ccell.2018.09.003>  
PMID:[30300579](https://pubmed.ncbi.nlm.nih.gov/30300579/)
57. Barbui T, Carobbio A, Finazzi G, Vannucchi AM, Barosi G, Antonioli E, Guglielmelli P, Pancrazzi A, Salmoiraghi S, Zilio P, Ottomano C, Marchioli R, Cuccovillo I, et al, and AGIMM and IIC Investigators. Inflammation and thrombosis in essential thrombocythemia and polycythemia vera: different role of C-reactive protein and pentraxin 3. *Haematologica*. 2011; 96:315–18.  
<https://doi.org/10.3324/haematol.2010.031070>  
PMID:[21173097](https://pubmed.ncbi.nlm.nih.gov/21173097/)
58. Diamandis EP, Goodglick L, Planque C, Thornquist MD. Pentraxin-3 is a novel biomarker of lung carcinoma. *Clin Cancer Res*. 2011; 17:2395–99.  
<https://doi.org/10.1158/1078-0432.CCR-10-3024>  
PMID:[21257721](https://pubmed.ncbi.nlm.nih.gov/21257721/)
59. Infante M, Allavena P, Garlanda C, Nebuloni M, Morengi E, Rahal D, Roncalli M, Cavuto S, Pesce S, Monari M, Valaperta S, Montanelli A, Solomon D, et al. Prognostic and diagnostic potential of local and circulating levels of pentraxin 3 in lung cancer patients. *Int J Cancer*. 2016; 138:983–91.  
<https://doi.org/10.1002/ijc.29822>  
PMID:[26348770](https://pubmed.ncbi.nlm.nih.gov/26348770/)
60. Germano G, Frapolli R, Simone M, Tavecchio M, Erba E, Pesce S, Pasqualini F, Grosso F, Sanfilippo R, Casali PG, Gronchi A, Virdis E, Tarantino E, et al. Antitumor and anti-inflammatory effects of trabectedin on human myxoid liposarcoma cells. *Cancer Res*. 2010; 70:2235–44.  
<https://doi.org/10.1158/0008-5472.CAN-09-2335>  
PMID:[20215499](https://pubmed.ncbi.nlm.nih.gov/20215499/)
61. Locatelli M, Ferrero S, Martinelli Boneschi F, Boiocchi L, Zavanone M, Maria Gaini S, Bello L, Valentino S, Barbati E, Nebuloni M, Mantovani A, Garlanda C. The long pentraxin PTX3 as a correlate of cancer-related inflammation and prognosis of malignancy in gliomas. *J Neuroimmunol*. 2013; 260:99–106.  
<https://doi.org/10.1016/j.jneuroim.2013.04.009>  
PMID:[23664694](https://pubmed.ncbi.nlm.nih.gov/23664694/)
62. Kondo S, Ueno H, Hosoi H, Hashimoto J, Morizane C, Koizumi F, Tamura K, Okusaka T. Clinical impact of pentraxin family expression on prognosis of pancreatic carcinoma. *Br J Cancer*. 2013; 109:739–46.  
<https://doi.org/10.1038/bjc.2013.348> PMID:[23828517](https://pubmed.ncbi.nlm.nih.gov/23828517/)
63. Carmo RF, Aroucha D, Vasconcelos LR, Pereira LM, Moura P, Cavalcanti MS. Genetic variation in PTX3 and plasma levels associated with hepatocellular carcinoma in patients with HCV. *J Viral Hepat*. 2016; 23:116–22.

- <https://doi.org/10.1111/jvh.12472>  
PMID:[26400151](https://pubmed.ncbi.nlm.nih.gov/26400151/)
64. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50.  
<https://doi.org/10.1073/pnas.0506580102>  
PMID:[16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
65. Netti GS, Prattichizzo C, Montemurno E, Simone S, Cafiero C, Rascio F, Stallone G, Ranieri E, Grandaliano G, Gesualdo L. Exposure to low- vs iso-osmolar contrast agents reduces NADPH-dependent reactive oxygen species generation in a cellular model of renal injury. *Free Radic Biol Med*. 2014; 68:35–42.  
<https://doi.org/10.1016/j.freeradbiomed.2013.11.016>  
PMID:[24300339](https://pubmed.ncbi.nlm.nih.gov/24300339/)
66. Santangelo L, Gigante M, Netti GS, Diella S, Puteo F, Carbone V, Grandaliano G, Giordano M, Gesualdo L. A novel SMARCAL1 mutation associated with a mild phenotype of Schimke immuno-osseous dysplasia (SIOD). *BMC Nephrol*. 2014; 15:41.  
<https://doi.org/10.1186/1471-2369-15-41>  
PMID:[24589093](https://pubmed.ncbi.nlm.nih.gov/24589093/)
67. Gigante M, d’Altilia M, Montemurno E, Diella S, Bruno F, Netti GS, Ranieri E, Stallone G, Infante B, Grandaliano G, Gesualdo L. Branchio-Oto-Renal Syndrome (BOR) associated with focal glomerulosclerosis in a patient with a novel EYA1 splice site mutation. *BMC Nephrol*. 2013; 14:60.  
<https://doi.org/10.1186/1471-2369-14-60>  
PMID:[23506628](https://pubmed.ncbi.nlm.nih.gov/23506628/)
68. Stallone G, Matteo M, Netti GS, Infante B, Di Lorenzo A, Prattichizzo C, Carlucci S, Trezza F, Gesualdo L, Greco P, Grandaliano G. Semaphorin 3F expression is reduced in pregnancy complicated by preeclampsia. An observational clinical study. *PLoS One*. 2017; 12:e0174400.  
<https://doi.org/10.1371/journal.pone.0174400>  
PMID:[28350837](https://pubmed.ncbi.nlm.nih.gov/28350837/)
69. Netti GS, Sangregorio F, Spadaccino F, Staffieri F, Crovace A, Infante B, Maiorano A, Godeas G, Castellano G, Di Palma AM, Prattichizzo C, Cotoia A, Mirabella L, et al. LPS removal reduces CD80-mediated albuminuria in critically ill patients with Gram-negative sepsis. *Am J Physiol Renal Physiol*. 2019; 316:F723–31.  
<https://doi.org/10.1152/ajprenal.00491.2018>  
PMID:[30672713](https://pubmed.ncbi.nlm.nih.gov/30672713/)
70. Netti GS, Infante B, Spadaccino F, Godeas G, Corallo MG, Prisciandaro C, Croce L, Rotondi M, Gesualdo L, Stallone G, Grandaliano G, Ranieri E. Serum Levels of BAFF and APRIL Predict Clinical Response in Anti-PLA2R-Positive Primary Membranous Nephropathy. *J Immunol Res*. 2019; 2019:8483650.  
<https://doi.org/10.1155/2019/8483650>  
PMID:[31781684](https://pubmed.ncbi.nlm.nih.gov/31781684/)
71. Netti GS, Rotondi M, Di Lorenzo A, Papantonio D, Teri A, Schirone M, Spadaccino F, Croce L, Infante B, Perulli R, Coperchini F, Rocchetti MT, Iannelli G, et al. Nocturnal haemodialysis is associated with a reduced occurrence of low triiodothyronine serum levels in haemodialysed patients. *Clin Kidney J*. 2020.  
<https://doi.org/10.1093/ckj/sfaa003>

## SUPPLEMENTARY MATERIALS

### Supplementary Figure



**Supplementary Figure 1. Kaplan-Meier survival curves, stratified by the expression level (FPKM) of PTX3 in the tumor tissue at the time of diagnosis.** Data from the cancer genome atlas (TCGA) clear cell renal cell carcinoma patient cohort (KIRC).

## Supplementary Table

**Supplementary Table 1. Clinical and pathological characteristics of patients with renal clear cell carcinoma subjected to PTX3 tissue expression analysis.**

<b>Clinical characteristics of RCC patients</b>		
Patients, n		30
Age, n		59.1±11.1
Female Gender, n (%)		13 (43.3%)
Diabetes Mellitus, n (%)		11 (36.7%)
C reactive protein (CRP), mg/dL		4.5±1.5
CKD-EPI eGFR, ml/min/1.73 m <sup>2</sup>		94.5±10.5
<b>Histologic characterization of of RCC patients</b>		
Furhman Grading	G 1	5 (16.7%)
	G 2	18 (60.0%)
	G 3	4 (13.3%)
	G 4	3 (10.0%)
TNM/AJCC Staging	pT1a	5 (16.7%)
	pT1b	13 (43.3%)
	pT2a	5 (10.0%)
	pT2b	3 (10.0%)
	pT3a	4 13.3%
	pT3b-c	0 (0.0%)
	pT4	0 (0.0%)

Abbreviations: CKD-EPI, Chronic Kidney Disease EPIdemiology Collaboration; eGFR, estimated Glomerular Filtration Rate; TNM/AJCC: Tumor size, Lymph Nodes affected, Metastases/American Joint Committee on Cancer.

Values are expressed as mean ± standard deviation, or number of cases and (percentage).

# Aberrantly glycosylated integrin $\alpha 3\beta 1$ is a unique urinary biomarker for the diagnosis of bladder cancer

Di Jin<sup>1</sup>, Ruiyun Zhang<sup>1</sup>, Haige Chen<sup>1</sup>, Chong Li<sup>2,3</sup>

<sup>1</sup>Department of Urology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

<sup>2</sup>Core Facility for Protein Research, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup>Beijing Jianlan Institute of Medicine, Beijing 100190, China

**Correspondence to:** Chong Li, Haige Chen; **email:** [lichong@moon.ibp.ac.cn](mailto:lichong@moon.ibp.ac.cn), [kirbyhaige@aliyun.com](mailto:kirbyhaige@aliyun.com)

**Keywords:** bladder cancer, urinary biomarker, aberrantly glycosylated integrin  $\alpha 3\beta 1$ , diagnosis

**Received:** February 20, 2020

**Accepted:** April 28, 2020

**Published:** June 13, 2020

**Copyright:** Jin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Bladder cancer (BC) is the most common malignancy of the urinary tract. We developed a new and ELISA kit for detecting aberrantly glycosylated integrin  $\alpha 3\beta 1$  (AG31) in human urine. We analysed urine samples (n=408) of patients with BC, renal cell carcinoma (RCC), prostate cancer (PC), cystitis, nephritis, and prostatitis from two centres in China. The subjects in the validation groups (n=2317) were recruited from other centres in China between July 2012 and September 2013. Receiver operating characteristic (ROC) curves were used to determine diagnostic accuracy. AG31 levels in urine samples were significantly higher in patients with BC than in any of the control subjects. Moreover, elevated levels of AG31 in urine could distinguish BC from benign inflammatory diseases. Finally, the urinary AG31 test was much more sensitive and specific than the NMP22 test. Therefore, the urinary AG31 test will provide an ideal and assay for the detection of BCs.

## INTRODUCTION

Bladder cancer (BC) is the most common malignancy of the urinary tract and ranks as the fourth most common cancer in men [1]. Tumours of the urinary bladder present either as non-muscle-invasive bladder cancer (NMIBC) or as muscle-invasive bladder cancer (MIBC). Approximately 75% of BC patients present with NMIBC at first diagnosis (Ta, T1 or tumour in situ (Tis)), and 25% present with muscle-invasive disease with a high risk of death from distant metastasis [2, 3]. Approximately 70% of NMIBCs recur, and the tumours of as many as 10% to 20% of patients will eventually progress to muscle-invasive cancer after treatment [4]. Different interventions against NMIBC change the biological and clinical behaviour of the disease [5]. Thus, early diagnosis and monitoring of the progression of BC are critical for successful treatment.

Cystoscopy and voided urine cytology are the most commonly used methods for the diagnosis and monitoring of BC recurrence and progression. Cystoscopy, the gold standard for the detection of BC, allows direct visualization and biopsy of the bladder urothelium. However, cystoscopy is invasive and relatively expensive, which limits its use. Cytologic testing of voided urine is the most commonly utilized non-invasive method for detecting BC. Voided urine cytology has good sensitivity for high-grade BC, but its sensitivity for the detection of low-grade tumours is only 4% to 31% [6]. Over the last decade, some urine-based assays, including one involving nuclear mitotic apparatus protein 22 (NMP22), a marker that has been marketed to diagnose BC, have been developed, but no markers have reached widespread use due to their low specificities [7–10]. Thus, cost-effective and non-invasive tools for the early diagnosis and lifelong surveillance of BC are urgently needed.

Integrins, a large family of cell membrane receptors, are involved in a variety of processes, including cell proliferation, migration, and cell extracellular matrix adhesion [11]. Integrin  $\alpha3\beta1$  acts as a high-affinity receptor for laminin, fibronectin, and collagen, whose interactions play critical roles in organogenesis and the maintenance of epithelial tissues [12, 13]. It has been reported that glycosylated integrin  $\alpha3\beta1$  appears in BC cells [14]. Notably, aberrant glycosylation has been implicated in the tumourigenesis of several tumour types [15, 16]. We previously generated an antibody, BCMab1, that specifically recognizes the aberrantly glycosylated integrin  $\alpha3\beta1$  (AG31) epitope on the membranes of BC cells [17]. We demonstrated that AG31 is specifically expressed in BC tissues but not in normal or other tumour tissues. AG31-mediated signalling triggers FAK activation in the tumourigenesis of BC. Furthermore, AG31 expression levels in tumour tissues are positively correlated with clinical severity and the prognosis of BC patients [17]. In this large-scale, multicentre validation study, we developed an ELISA-based assay to quantitatively detect AG31 levels in the voided urine of patients with BC or other urologic tumours. Our data show that the urinary AG31 test is a promising assay for the detection of BC.

## RESULTS

### Study groups

We recruited 2725 participants overall: 408 in the test groups and 2317 in the validation groups (Figure 1). Of all the participants, 1314 had BC, and their demographic characteristics, modes of presentation, tumour stages and tumour grades are summarized in Table 1. The test population was predominantly male (78.8%), and 46.7% were over 63 years of age. Of those with a recorded presentation, 10.3% had visible haematuria, and 89.7% had non-visible haematuria. A total of 72.8% of patients had no recurrence, 14.7% experienced recurrence, and 12.5% were lost to follow-up. A total of 1130 BC patients were enrolled into the validation groups, and their clinicopathological characteristics were well matched to those of the test groups (Table 1); these two groups were well matched for age. Additionally, the patients with non-BC urologic conditions were also matched for age. Cut-off points were set as follows: 65 years of age for RCC, PC and cystitis patients, 47 years of age for nephritis patients, and 57 years of age for prostatitis patients (Supplementary Table 1).

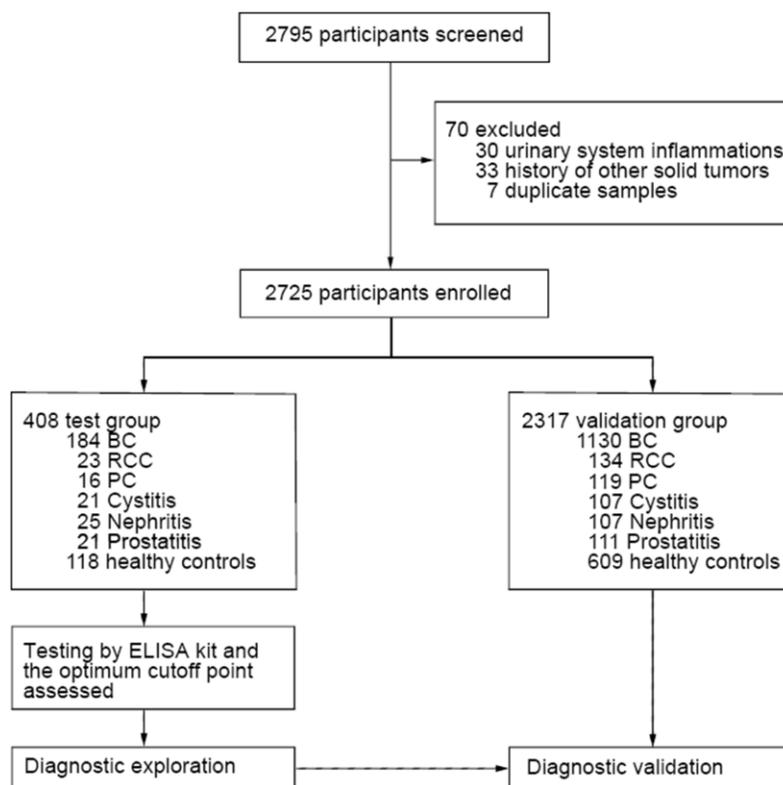


Figure 1. Patient selection process and classification. Overall patient selection process and their classification based on types of diseases.

**Table 1. Demographic and clinicopathologic characteristics of patients with bladder cancer in the test and validation groups.**

Variable	Test (n=184)		Validation (n=1130)		p value
	No.	Percentage (%)	No.	Percentage (%)	
Age (y)					
	≤63	92	50.0	490	43.4
	>63	86	46.7	609	53.9
	Missing	6	3.3	31	2.7
Gender					
	Female	32	17.4	234	20.7
	Male	145	78.8	868	76.8
	Missing	7	3.8	28	2.5
Pathological stage					
	Ta+ T1	134	72.8	924	81.8
	T2+T3+T4	50	27.2	206	18.2
Pathological grade					
	G1	69	37.5	279	24.7
	G2+G3	114	62.0	851	75.3
	Missing	1	0.5		
Hematuria					
	Visible	19	10.3	360	31.9
	Not-visible	165	89.7	770	68.1
Tumor recurrence					
	Yes	27	14.7	188	16.6
	No	134	72.8	887	78.5
	Missing	23	12.5	55	4.9

### Urinary AG31 is a sensitive marker for the detection of bladder cancer

To detect AG31 levels in voided urine samples, an ELISA-based assay kit was developed by our laboratory (Supplementary Figure 1A). We generated several monoclonal antibodies that could recognize the AG31 molecule on BC tumours as previously described [17]. For establishment of the ELISA, we screened out an antibody termed BCMab3 as the capture antibody, which could strongly bind to AG31. The BCMab1 antibody also specifically recognizes AG31 on BC tumours [17], and BCMab1-conjugated horseradish peroxidase (HRP) served as the detection antibody. This ELISA kit could be used to accurately measure urinary AG31 levels (Supplementary Figure 1B). Relative light units (RLUs) were used to indicate the test levels of AG31 concentrations in urine.

In the test group, urinary AG31 concentrations in the BC patients were significantly higher than those in the healthy controls (median 6525, interquartile range (IQR) 3600-26,585; mean 54,768, standard deviation (SD) 276,559 [BC patients] vs median 576, IQR 378-988; mean 859, SD 819 [healthy controls];  $p < 0.0001$ ) (Figure 2A, and Supplementary Table 2). Notably, the urinary AG31 levels in patients with RCC, PC, cystitis,

nephritis, or prostatitis were comparable to those in healthy individuals (Figure 2A and Supplementary Table 2). The ROC curves were plotted for urinary AG31 in BC patients versus different test groups. For BC patients from all test groups, the area under the curve (AUC) of AG31 was 0.9567 (95% CI 0.9337-0.9797), with a sensitivity of 90.76% and specificity of 91.52% (Figure 1C, and Table 2). The optimum cut-off value was set to 1991 (Supplementary Figure 1C). The four ROC curves of AG31 levels between the BC patients and different test groups showed that the AUCs were greater than 95% (Figure 2C, and Table 2). These results were confirmed by the corresponding validation group tests (Figure 2B, 2D). Predictive values and likelihood ratios for AG31 in the diagnosis of BC are shown in Table 2. Altogether, the urinary AG31 test can distinguish BC patients from patients with other urologic tumours and benign inflammatory diseases.

### Urinary AG31 levels are well correlated with the clinicopathologic features of bladder cancer

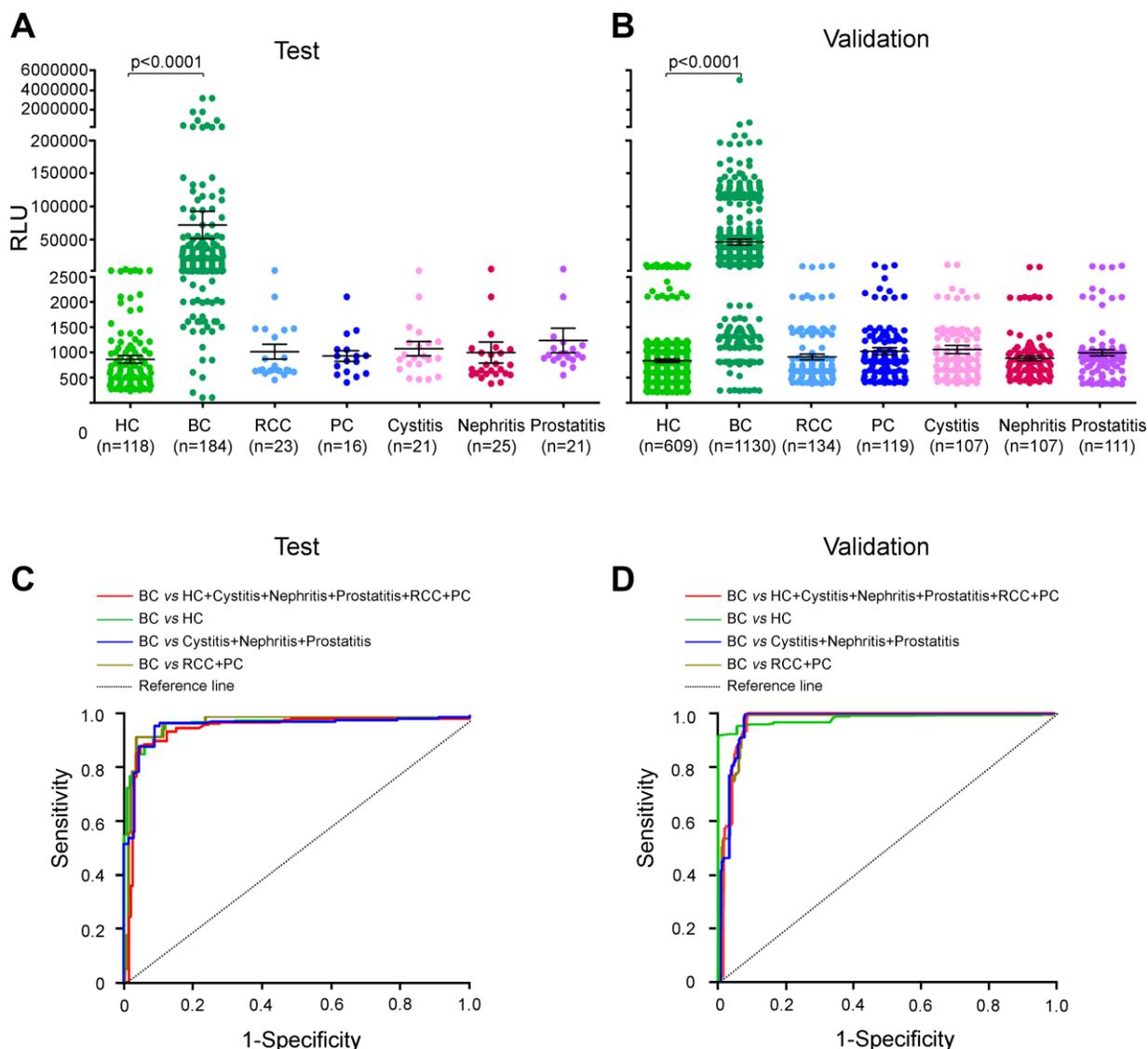
In the test group, the urinary AG31 levels of the BC patients with different disease stages were much higher than those of the healthy controls (Figure 3A, and Supplementary Table 2). Moreover, urinary AG31 levels were significantly increased in patients with high-

stage BC. These observations were further confirmed in the validation group (Figure 3B and Supplementary Table 2). Furthermore, the AUCs for the AG31 levels of the BC patients with the different disease stages were greater than 92%, with sensitivities over 89% and specificities over 90% (Supplementary Figure 2, and Supplementary Table 3). Similarly, urinary AG31 levels increased with advancing pathological grade (Figure 3C, D, Supplementary Figure 3, Supplementary Tables 2, 3). Correlation analysis showed that urinary AG31 levels were positively correlated with tumour stage and grade (both  $p < 0.01$ ). Taken together, these data indicate

that AG31 levels are well correlated with the clinicopathologic features of BCs.

### Urinary AG31 test is diagnostically accurate for NMIBC patients

Approximately 70% of BC patients with NMIBC (Tis+Ta+T1) will experience one or more recurrences after transurethral resection (TUR), and the BCs 10%-20% of patients will eventually progress to MIBC (3,26). Regarding recurrence and progression, patients with NMIBC are classified as low/intermediate- and



**Figure 2. Urinary AG31 levels are elevated in bladder cancers in the test and validation groups.** (A) Urinary AG31 levels for test groups. (B) Urinary AG31 levels for validation groups. Black horizontal lines are means, and error bars are SEs. Urinary AG31 levels were measured with RLU (relative light unit). HC, healthy control; BC, bladder cancer; RCC, renal cell carcinoma; PC, prostate cancer. (C) ROC curves for all patients with bladder cancer versus all controls in the test groups. (D) ROC curves for all patients with bladder cancer versus all controls in the validation groups. Jagged curves denote ROC curves; diagonal lines represent reference lines. HC, healthy control; BC, bladder cancer; RCC, renal cell carcinoma; PC, prostate cancer.

**Table 2. Urinary AG31 test for the diagnosis of bladder cancers.**

Test						Validation							
AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Positive LR	Negative LR	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Positive LR	Negative LR
BC vs HC+Cystitis+Nephritis+Prostatitis+RCC+PC													
0.9567 (0.9337- 0.9797)	0.9076	0.9152	0.8978	0.9234	10.7002	0.1010	0.9760 (0.9694- 0.9826)	0.9230	0.9292	0.9255	0.9269	13.0430	0.0829
BC vs HC													
0.9597 (0.9368- 0.9826)	0.9076	0.9153	0.9435	0.8640	10.7098	0.1009	0.9801 (0.9741- 0.9862)	0.9230	0.9442	0.9684	0.8686	16.5327	0.0815
BC vs Cystitis+Nephritis+Prostatitis													
0.9500 (0.9200- 0.9801)	0.9076	0.9104	0.9653	0.7821	10.1350	0.1015	0.9714 (0.9636- 0.9791)	0.9230	0.9108	0.9729	0.7728	10.3441	0.0845
BC vs RCC+PC													
0.9592 (0.9328- 0.9856)	0.9076	0.9231	0.9824	0.6792	11.7989	0.1001	0.9720 (0.9644- 0.9796)	0.9230	0.9170	0.9803	0.7273	11.1201	0.0840

HC, healthy control; BC, bladder cancer; RCC, renal cell carcinoma; PC, prostate cancer; AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; CI, confidence interval.

high-risk groups (5,27). We next wanted to determine whether the urinary AG31 test could distinguish NMIBC patients. Of all the BC patients recruited, 1058 were classified as having stage Ta or T1 disease. The mean AG31 level of the Ta-stage patients was 13.49 times higher than that of the healthy controls in the test group (Figure 3A and Supplementary Table 2). In contrast, the mean AG31 level of the T1-stage patients was 32.11 times higher than that of the healthy controls. For the patients with NMIBC, ROC curves showed that the AUC of the AG31 levels was 0.9524 (95% CI 0.9248-0.9800), with a sensitivity of 91.04% and specificity of 91.35% in the test group compared with the patients with benign inflammatory diseases and the healthy controls (Figure 4A, and Table 3). When the NMIBC patients were compared only with the healthy controls, the AUC of the AG31 levels was 0.9574 (95% CI 0.9310-0.9838), with a sensitivity of 91.04% and specificity of 91.53% in the test group. Similar results were obtained in the validation groups (Figure 4B, and Table 3). Importantly, the diagnostic accuracy of the AG31 test was not affected by age, sex, or haematuria in BC patients (Supplementary Table 4). Altogether, these findings indicate that the urinary AG31 test is diagnostically accurate for NMIBC patients.

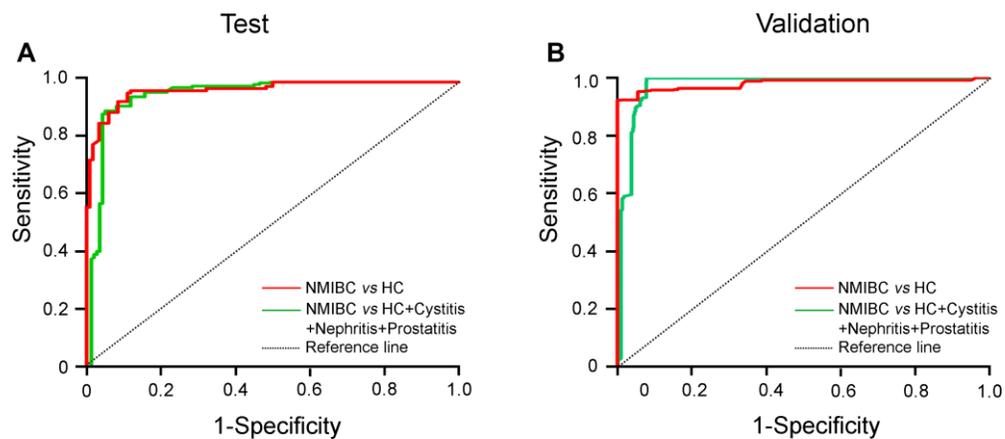
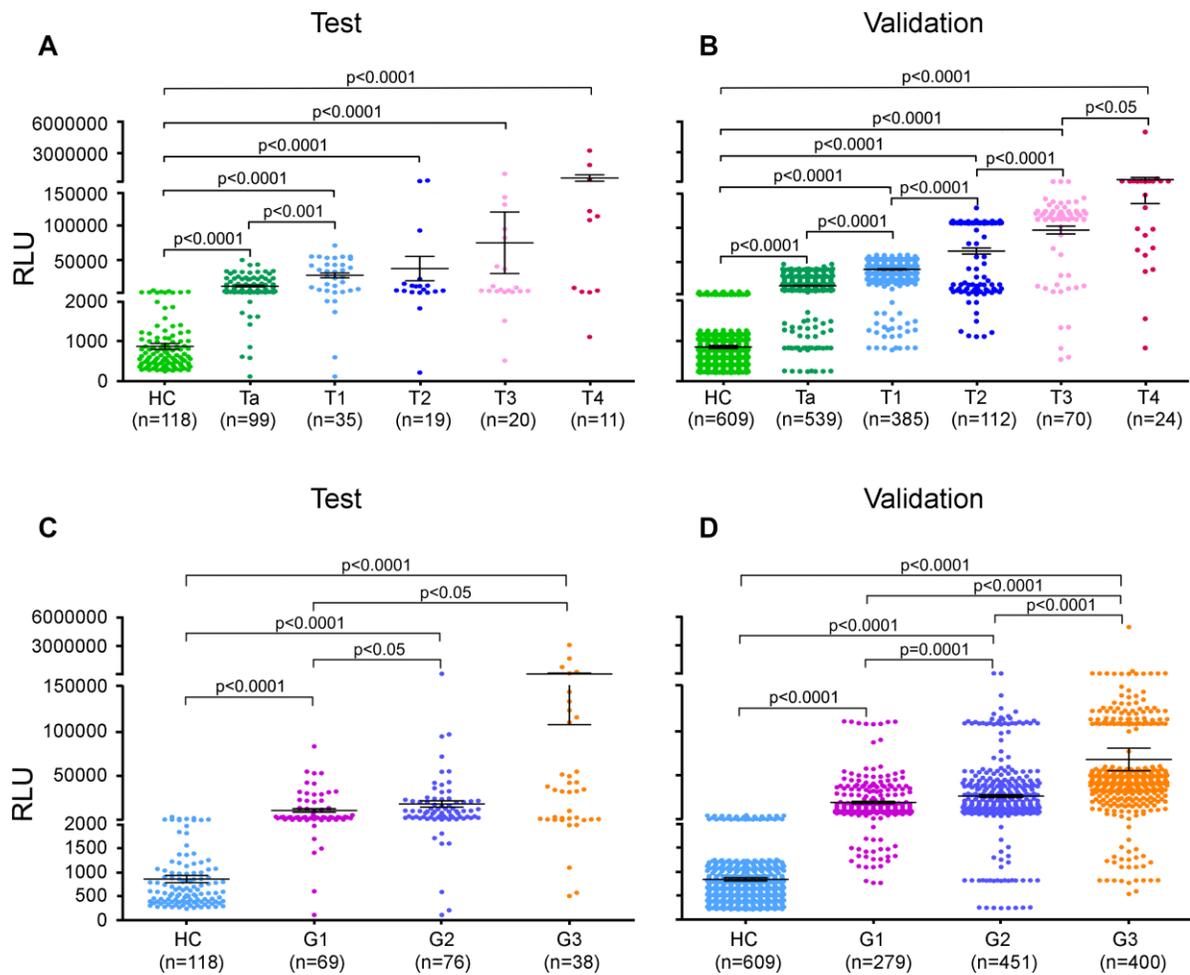
### The urinary AG31 test is much more sensitive and specific than the NMP22 test

NMP22, a member of the nuclear matrix protein (NMP) family, is much more prevalent in malignant urothelial cells than in normal cells. Given that NMP22 is released into urine upon cell apoptosis, NMP22 is

significantly elevated in the urine of BC patients compared to that of healthy individuals. The NMP22 test kit was approved by the U.S. Food and Drug Administration (FDA) for use in surveillance of BC. Given that the NMP22 test was used as a biomarker for the auxiliary diagnosis of BC, we applied the urinary NMP22 test and compared the results to those of the AG31 test. To further assess the diagnostic accuracy of the AG31 assay, we measured AG31 and NMP22 levels in freshly voided urine samples from BC patients (n=53) and healthy individuals (n=54). The AUC of the AG31 levels was 99.65% (95% CI: 0.9906-1.0020), with a sensitivity and specificity for the detection of BC of 90.57% and 98.15%, respectively (Supplementary Figure 4A, 4B). However, the AUC of NMP22 levels was 74.11% (95% CI: 0.6476-0.8346), and the sensitivity and specificity for the detection of BC were 47.17% and 87.04%, respectively. Discordances between the AG31 test and NMP22 test were assessed using the nonparametric McNemar test, and the *p* value was 0.01. Importantly, the predictive values and likelihood ratios for AG31 were much better than those for NMP22 (Supplementary Figure 4C). In summary, the AG31 test in urine has a better diagnostic accuracy for BC than the NMP22 test.

## DISCUSSION

We previously generated a monoclonal antibody, BCMab1, that specifically recognizes the AG31 antigen on BC tumours [17]. AG31 expression levels in tumour tissues are well correlated with the clinical severity and prognosis of BC patients. In our large-scale, multicentre



**Table 3. Results for measurement of urinary AG31 in the diagnosis of non-muscle invasive bladder cancer (NMIBC) patients.**

Test						Validation							
AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Positive LR	Negative LR	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Positive LR	Negative LR
NMIBC vs HC													
0.9574 (0.9310-0.9838)	0.9104	0.9153	0.9242	0.9000	10.7433	0.0978	0.9790 (0.9720-0.9860)	0.9232	0.9442	0.9617	0.8901	16.5354	0.0814
NMIBC vs HC+Cystitis+Nephritis+Prostatitis													
0.9524 (0.9248-0.9800)	0.9104	0.9135	0.8841	0.9337	10.5271	0.0980	0.9758 (0.9684-0.9833)	0.9232	0.9325	0.9312	0.9246	13.6862	0.0824

HC, healthy control; NMIBC, non-muscle invasive bladder cancer; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; CI, confidence interval.

study, we developed an ELISA kit based on BCMab1 and measured AG31 levels in the voided urine of BC patients. We show that urinary AG31 levels are substantially elevated in BC patients. The cut-off value was chosen to be 1991, and both the sensitivity and specificity of AG31 for the diagnosis of BC were over 90%. Moreover, the AG31 test could distinguish BC patients from healthy controls. Thus, urinary AG31 is a sensitive and specific biomarker for the detection of BC.

The diagnosis of BC currently depends on cystoscopy and urine cytology. Both examination methods have disadvantages and limitations. Cystoscopy is invasive, expensive, and associated with post-cystoscopy pain and/or risk of urinary infections [26, 27]. Cystoscopy is prone to missing flat lesions, such as those in patients with early Tis-stage BC, whereas urine cytology has a tendency to miss well-differentiated low-grade lesions [6, 28]. Furthermore, both methods rely on observer expertise, thus limiting their clinical applications. As such, there is an urgent need for a better, simpler, and cheaper diagnostic test in the diagnosis and surveillance of BC in patients. Over the last decade, several urine-based assays have been developed and become commercially available, including the NMP22, ImmnoCyt, and BTA stat tests [6, 29]. More recently, other associated protein biomarkers have been reported to be useful as urinary markers for the detection of BCs [7, 30, 31]. Given that these markers are not specific to BC, their specificities for the diagnosis of BC are usually low. Therefore, no markers have reached widespread use to date. Here, we utilized BCMab1, which recognizes the BC-specific membrane protein AG31, to develop an ELISA-based kit for the detection of AG31 released in urine.

Early diagnosis and vigilant surveillance of recurrences will immediately provide an ideal therapeutic strategy

to treat BC patients. Our data show that the AG31 test has high sensitivity and specificity for detecting BCs, including early-stage BCs. Among the NMIBC patients, 638 were first diagnosed with Ta-stage BC. The AG31 test can detect Ta patients with 90.9% sensitivity and 91.5% specificity compared to healthy controls. Importantly, the AG31 test can also detect low-grade tumours with over 90% sensitivity and specificity. Thus, the urinary AG31 test is an ideal examination tool for the early diagnosis of BCs.

Due to the lack of disease-specific symptoms, the diagnosis and follow-up of BC remains a major challenge. The most common presenting symptom of BC is gross painless haematuria, usually accompanied by unexplained urinary frequency, urgency, or irritative voiding symptoms [4]. These symptoms are quite similar to those of other benign urinary infections or malignancies. To date, no biomarker has been validated for the differential diagnosis of these urinary diseases. Here, we show that the urinary AG31 test solely and specifically detects BCs but not other benign urinary infections (cystitis, nephritis or prostatitis) or other urinary malignancies, such as RCC and PC. More importantly, haematuria does not influence the detection sensitivity or specificity of AG31 for BCs.

For diagnostic evaluation of the urinary AG31 test in this study, we set the cut-off value to 1991. At this cut-off point, the sensitivity and specificity for the detection of BC are 90.76% and 91.52%, respectively. If the urinary AG31 detection kit is used for routine health screening, the cut-off point can be chosen to be 6027, in which the specificity for BC diagnosis is 100% (sensitivity is 51.63%). At this level, almost no healthy individual is predicted to have a false positive result. In contrast, if the urinary AG31 detection kit is applied to monitor therapeutic response, the cut-off point can be

set to 1090, whose sensitivity for BC diagnosis is 96.74% (specificity is 77.68%). Therefore, the urinary AG31 test will serve as an ideal test tool for routine screening of BC patients.

The NMP22 test kit, approved by the FDA, is available for clinical application. Although urinary NMP22 is elevated in BC, dead and dying urothelial cells in other malignancies or inflammatory conditions can also release NMP22, thus decreasing its specificity [32, 33]. Several studies have reported that the sensitivity of NMP22 ranges from 33% to 100%, and the specificity ranges from 40% to 93% [34]. We measured NMP22 and AG31 levels in voided urine samples from BC patients and healthy controls. The sensitivity of the urinary NMP22 test is only 47.17%, which is much lower than that of the AG31 test. In conclusion, the urinary AG31 test is a sensitive and specific diagnostic test for BC that detects BCs of all stages and grades. Thus, the AG31 test will act as a promising urinary marker for the detection of BC.

## **MATERIALS AND METHODS**

### **Primary antibodies**

BCMab1 and BCMab3 monoclonal antibodies, which specifically recognize the AG31 epitope on the membranes of BC cells, were generated by our group [17]. BCMab3 was used as the capture antibody, and BCMab1 served as the detecting antibody. BCMab1 was conjugated to HRP by linkage of 2 imine-carbon (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide, EDC) as described previously [18, 19].

### **Study population**

From January to June 2012, we enrolled consecutive patients with BC from the Peking University People's Hospital (Beijing, China) and Cancer Institute and Hospital, Chinese Academy of Medical Science (Beijing, China) for constitute our test group. During the same time period, we also recruited consecutive patients with RCC, prostate cancer (PC), cystitis, nephritis, or prostatitis and healthy control subjects from The First Affiliated Hospital of Shenzhen University (Shenzhen, China). From July 2012 to September 2013, validation groups comprising patients with BC, RCC, PC, cystitis, nephritis, or prostatitis and healthy control subjects were recruited from The First Affiliated Hospital of Zhejiang University (Hangzhou, China), The Second Affiliated Hospital of Kunming Medical University (Kunming, China), Renji Hospital Affiliated to Shanghai Jiaotong University School of Medicine (Shanghai, China), and The First Affiliated Hospital of Jilin University (Changchun, China).

The presence of BC was confirmed by cystoscopy, together with histopathological information obtained after subsequent surgical interventions. Tumours were graded according to the WHO criteria [20, 21], and tumour stages were defined according to the tumour-node-metastasis (TNM) staging system [22]. BC patients in this study were subdivided into five stages (Ta, T1, T2, T3, and T4) and three pathological grades (G1, G2, and G3). For the purpose of this study, we classified TNM stage Tis tumours and Ta tumours including stage Ta, Ta and T1 tumours as NMIBC. The diagnoses of cystitis and nephritis were based on symptoms, physical examination and the results of the urine culture, according to the guidelines of the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases [23]. The diagnosis of prostatitis was based on clinical, laboratory, and imaging evidence (X-ray, ultrasonography, CT, or MRI) [24, 25]. The presence of RCC was confirmed by symptoms, imaging studies (CT or MRI), laboratory data, and renal tumour biopsy. The diagnosis of PC was based on digital rectal examination, the prostate-specific antigen test, imaging evidence (transrectal ultrasonography, CT, MRI, or ECT), prostate biopsy, and pathohistological examination of radical prostatectomy specimens, according to the clinical guidelines of the European Association of Urology. The healthy control subjects were eligible volunteers with no diseases of the urinary system, no viral hepatitis and no malignant diseases. Individuals who had a history of other solid tumours were excluded from the study.

For the test groups, we enrolled patients with BC, RCC, PC, cystitis, nephritis, or prostatitis, along with healthy control subjects, from January to June 2012. For the validation groups, patients with BC, RCC, PC, cystitis, nephritis, or prostatitis and healthy control subjects were recruited from July 2012 to September 2013. These two groups were matched for age ( $\leq 63$  years versus  $>63$  years) and sex for comparisons of AG31 values. In this study, all urine samples of the patients with BC, RCC, PC, cystitis, nephritis or prostatitis were collected before treatment. Data collection and analysis were performed by three independent researchers (YD, CL, and QZ). Approval for the study was obtained from the institutional ethics review committee of each study centre. Written informed consent was obtained from all participants, according to the committees' regulations.

### **Measurement of urine samples for the AG31 test**

Each well of a 96-well Nunc-Immuno microtiter plate with a MaxiSorp surface (Nalge Nunc, Penfield, NY, USA) was coated with 100  $\mu\text{L}$  of a 4  $\mu\text{g}/\text{mL}$  solution of BCMab3 in 50 mM carbonate buffer (pH 9.5), and

plates were incubated at room temperature overnight. The wells were then blocked with 120  $\mu$ L of 10 mM phosphate buffer (pH 7.4) containing 1% bovine serum albumin and 1% gelatine at 37 °C for 2 h. The plates were washed four times with PBS containing 0.05% Tween-20 and then dried at room temperature. These plates were kept at 4 °C. For the urinary AG31 test, 50  $\mu$ L of urine samples and 50  $\mu$ L of HRP-BCMab1 (1 ng/mL) were added to each well and incubated at 37 °C for 1 h. After the plates were completely washed, they were dried at room temperature. A total of 100  $\mu$ L of freshly prepared substrate solution was added to each well and stirred. Chemiluminescence intensity was measured with a chemiluminescence apparatus (Hamamatsu Photonics, Beijing, China).

### NMP22 measurement

A commercial ELISA kit (Alere Scarborough Inc., Scarborough, ME, USA) was used for auxiliary diagnosis of BC as a biomarker. According to the manufacturer's protocol for the NMP22 test, urinary NMP22 levels were measured for comparison.

### Statistical analysis

Statistical analyses were conducted with SPSS for Windows (version 16.0) and GraphPad Prism statistical software (version 5.01). The Mann-Whitney U test was used to assess differences between two independent groups, such as the levels of urinary AG31 between the patients with BC and the healthy controls. The differences in clinicopathological stages or grades were assessed using the Kruskal-Wallis H test, which was used to analyse multiple variables. Receiver operating characteristic (ROC) analysis was used to characterize marker performance, and ROC curves were constructed to assess sensitivity, specificity, and the respective areas under the curve (AUCs) with 95% confidence intervals. In the ROC curves, the true positive rate (sensitivity) is plotted against the false positive rate (1-specificity) for different cut-off points of a parameter. We chose the optimum cut-off value for diagnosis by maximizing the sum of sensitivity and specificity and minimizing the overall error (square root of the sum  $[1-\text{sensitivity}]^2 + [1-\text{specificity}]^2$ ) and by minimizing the distance of the cut-off value to the top-left corner of the ROC curve. The correlations between the levels of AG31 in urine and clinicopathological characteristics were analysed using Pearson's chi-square test ( $\chi^2$ ) or Fisher's exact test. The bivariate correlation and the Spearman correlation coefficient were used to assess the magnitude of the correlation between urinary AG31 levels and tumour stages or grades. Discordances between the AG31 test and NMP22 test were assessed using the nonparametric McNemar test. In all statistical

analyses, a *p* value of 0.05 or less was considered statistically significant.

### CONFLICTS OF INTEREST

We declare that we have no conflicts of interest.

### FUNDING

This work was supported by the National Natural Science Foundation of China (81672956, 81972390, 81672514 and 81902562), Beijing Science and Technology Project (Z181100003818003), Beijing-Tianjin-Hebei Basic Research Cooperation Project (19JCZDJC65800(Z)), Shanghai Natural Science Foundation (16ZR1420300, 18410720400, 19431907400), Renji Hospital Research Funding Projects (RJZZ18-020, PYIII-17-017, PY2018-IIC-02), Shanghai Jiao Tong University School of Medicine Research Funding Projects (TM201708) and Foundation of Shanghai Hospital Development Center (SHDC12015125).

### REFERENCES

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin.* 2014; 64:9–29. <https://doi.org/10.3322/caac.21208> PMID:[24399786](https://pubmed.ncbi.nlm.nih.gov/24399786/)
2. Babjuk M, Burger M, Zigeuner R, Shariat SF, van Rhijn BW, Compérat E, Sylvester RJ, Kaasinen E, Böhle A, Palou Redorta J, Rouprêt M, and European Association of Urology. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2013. *Eur Urol.* 2013; 64:639–53. <https://doi.org/10.1016/j.eururo.2013.06.003> PMID:[23827737](https://pubmed.ncbi.nlm.nih.gov/23827737/)
3. Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer.* 2015; 15:25–41. <https://doi.org/10.1038/nrc3817> PMID:[25533674](https://pubmed.ncbi.nlm.nih.gov/25533674/)
4. Kaufman DS, Shipley WU, Feldman AS. Bladder cancer. *Lancet.* 2009; 374:239–49. [https://doi.org/10.1016/S0140-6736\(09\)60491-8](https://doi.org/10.1016/S0140-6736(09)60491-8) PMID:[19520422](https://pubmed.ncbi.nlm.nih.gov/19520422/)
5. Babjuk M. What are the limitations of endoscopic management of urothelial carcinoma of the upper urinary tract? *Eur Urol.* 2011; 60:961–63. <https://doi.org/10.1016/j.eururo.2011.08.021> PMID:[21862204](https://pubmed.ncbi.nlm.nih.gov/21862204/)
6. Xylinas E, Kluth LA, Lotan Y, Daneshmand S, Rieken M, Karakiewicz PI, Shariat SF. Blood- and tissue-based biomarkers for prediction of outcomes in

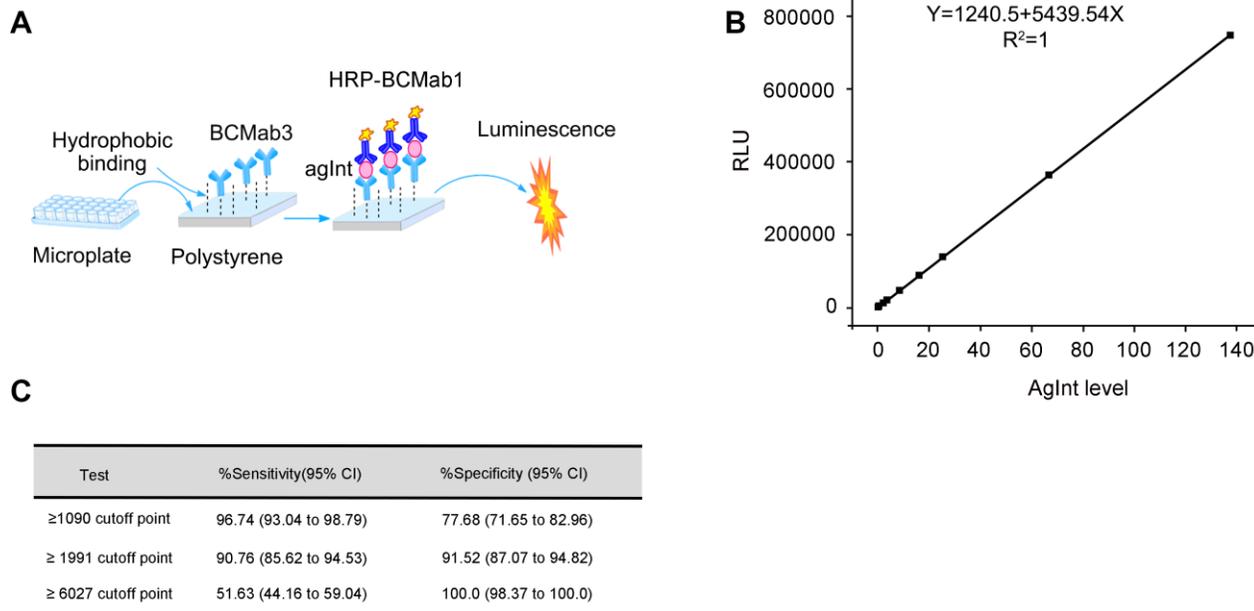
- urothelial carcinoma of the bladder. *Urol Oncol*. 2014; 32:230–42.  
<https://doi.org/10.1016/j.urolonc.2013.06.009>  
 PMID:24332648
7. Kandimalla R, Masius R, Beukers W, Bangma CH, Orntoft TF, Dyrskjot L, van Leeuwen N, Lingsma H, van Tilborg AA, Zwarthoff EC. A 3-plex methylation assay combined with the FGFR3 mutation assay sensitively detects recurrent bladder cancer in voided urine. *Clin Cancer Res*. 2013; 19:4760–69.  
<https://doi.org/10.1158/1078-0432.CCR-12-3276>  
 PMID:23842048
  8. Kinde I, Munari E, Faraj SF, Hruban RH, Schoenberg M, Bivalacqua T, Allaf M, Springer S, Wang Y, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N, Netto GJ. TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine. *Cancer Res*. 2013; 73:7162–67.  
<https://doi.org/10.1158/0008-5472.CAN-13-2498>  
 PMID:24121487
  9. Margulis V, Lotan Y, Karakiewicz PI, Fradet Y, Ashfaq R, Capitanio U, Montorsi F, Bastian PJ, Nielsen ME, Müller SC, Rigaud J, Heukamp LC, Netto G, et al. Multi-institutional validation of the predictive value of ki-67 labeling index in patients with urinary bladder cancer. *J Natl Cancer Inst*. 2009; 101:114–19.  
<https://doi.org/10.1093/jnci/djn451>  
 PMID:19141773
  10. Stoeber K, Swinn R, Prevost AT, de Clive-Lowe P, Halsall I, Dilworth SM, Marr J, Turner WH, Bullock N, Doble A, Hales CN, Williams GH. Diagnosis of genito-urinary tract cancer by detection of minichromosome maintenance 5 protein in urine sediments. *J Natl Cancer Inst*. 2002; 94:1071–79.  
<https://doi.org/10.1093/jnci/94.14.1071>  
 PMID:12122098
  11. Winograd-Katz SE, Fässler R, Geiger B, Legate KR. The integrin adhesome: from genes and proteins to human disease. *Nat Rev Mol Cell Biol*. 2014; 15:273–88.  
<https://doi.org/10.1038/nrm3769>  
 PMID:24651544
  12. Sachs N, Sonnenberg A. Cell-matrix adhesion of podocytes in physiology and disease. *Nat Rev Nephrol*. 2013; 9:200–10.  
<https://doi.org/10.1038/nrneph.2012.291>  
 PMID:23338211
  13. Liu Y, Chattopadhyay N, Qin S, Szekeres C, Vasylyeva T, Mahoney ZX, Taglienti M, Bates CM, Chapman HA, Miner JH, Kreidberg JA. Coordinate integrin and c-met signaling regulate Wnt gene expression during epithelial morphogenesis. *Development*. 2009; 136:843–53.  
<https://doi.org/10.1242/dev.027805>  
 PMID:19176588
  14. Lityńska A, Przybyło M, Ksiazek D, Laidler P. Differences of alpha3beta1 integrin glycans from different human bladder cell lines. *Acta Biochim Pol*. 2000; 47:427–34.  
[https://doi.org/10.18388/abp.2000\\_4022](https://doi.org/10.18388/abp.2000_4022)  
 PMID:11051207
  15. Brooks CL, Schietinger A, Borisova SN, Kufer P, Okon M, Hirama T, Mackenzie CR, Wang LX, Schreiber H, Evans SV. Antibody recognition of a unique tumor-specific glycopeptide antigen. *Proc Natl Acad Sci USA*. 2010; 107:10056–61.  
<https://doi.org/10.1073/pnas.0915176107>  
 PMID:20479270
  16. Park JH, Nishidate T, Kijima K, Ohashi T, Takegawa K, Fujikane T, Hirata K, Nakamura Y, Katagiri T. Critical roles of mucin 1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 in mammary carcinogenesis. *Cancer Res*. 2010; 70:2759–69.  
<https://doi.org/10.1158/0008-5472.CAN-09-3911>  
 PMID:20215525
  17. Li C, Yang Z, Du Y, Tang H, Chen J, Hu D, Fan Z. BCMab1, a monoclonal antibody against aberrantly glycosylated integrin  $\alpha 3 \beta 1$ , has potent antitumor activity of bladder cancer in vivo. *Clin Cancer Res*. 2014; 20:4001–13.  
<https://doi.org/10.1158/1078-0432.CCR-13-3397>  
 PMID:25002124
  18. Watanabe H, Satake A, Kido Y, Tsuji A. Monoclonal-based enzyme-linked immunosorbent assay and immunochromatographic assay for enrofloxacin in biological matrices. *Analyst*. 2002; 127:98–103.  
<https://doi.org/10.1039/b109427k> PMID:11827405
  19. Yu F, Yu S, Yu L, Li Y, Wu Y, Zhang H, Qu L, Harrington Pde B. Determination of residual enrofloxacin in food samples by a sensitive method of chemiluminescence enzyme immunoassay. *Food Chem*. 2014; 149:71–75.  
<https://doi.org/10.1016/j.foodchem.2013.10.024>  
 PMID:24295678
  20. Mostofi FK, Sobin LH, Torloni H, and World Health Organization. Histological typing of urinary bladder tumours. F. K. Mostofi in collaboration with L. H. Sobin, H. Torloni and pathologists in fourteen countries. Geneva: World Health Organization; 1973.  
<https://apps.who.int/iris/handle/10665/41533>
  21. Chen Z, Ding W, Xu K, Tan J, Sun C, Gou Y, Tong S, Xia G, Fang Z, Ding Q. The 1973 WHO classification is more suitable than the 2004 WHO classification for predicting prognosis in non-muscle-invasive bladder cancer. *PLoS One*. 2012; 7:e47199.  
<https://doi.org/10.1371/journal.pone.0047199>

PMID:[23082147](#)

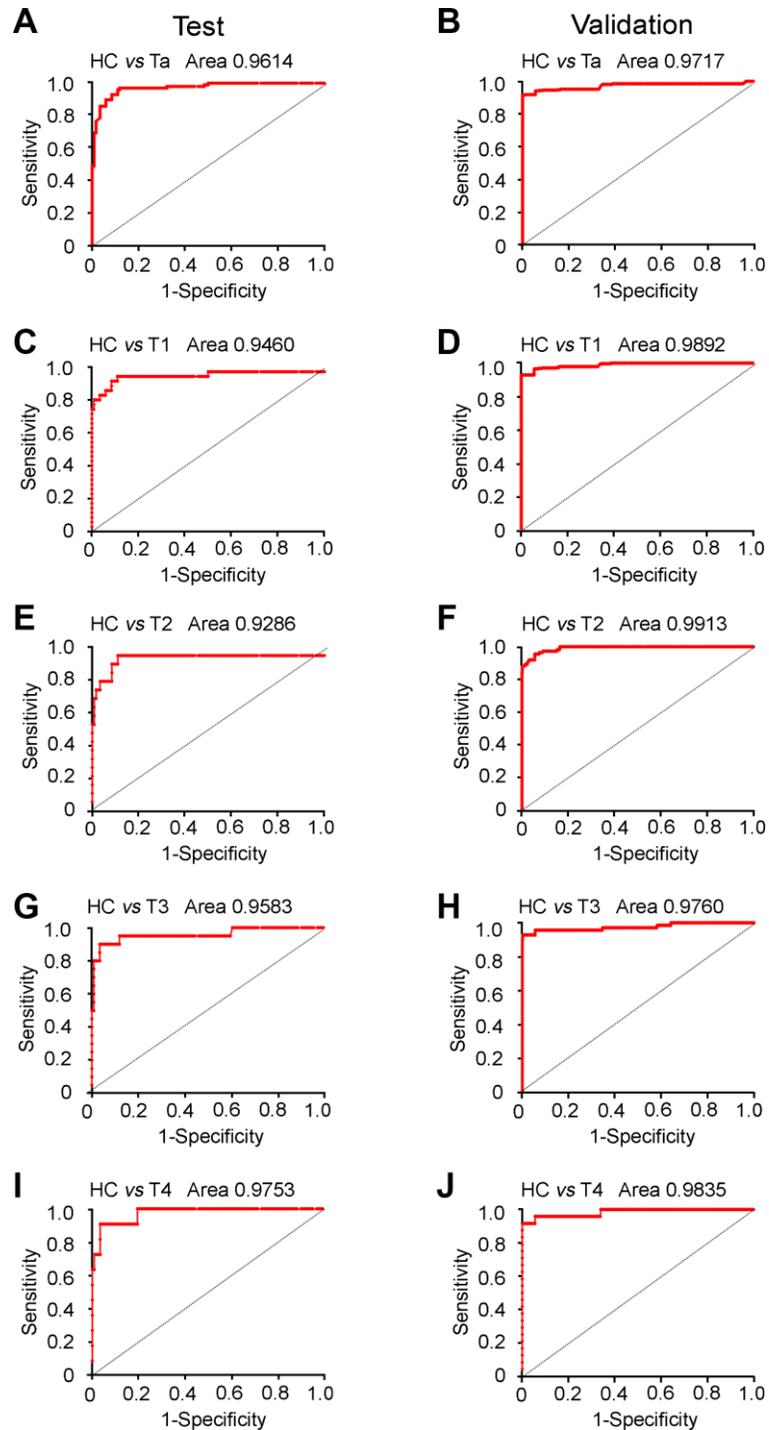
22. American Joint Committee on Cancer. AJCC cancer staging handbook: from the AJCC cancer staging manual. 6th ed. New York: Springer; 2002.
23. Gupta K, Hooton TM, Naber KG, Wullt B, Colgan R, Miller LG, Moran GJ, Nicolle LE, Raz R, Schaeffer AJ, Soper DE, and Infectious Diseases Society of America, and European Society for Microbiology and Infectious Diseases. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: a 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clin Infect Dis*. 2011; 52:e103–20.  
<https://doi.org/10.1093/cid/ciq257>  
PMID:[21292654](#)
24. Schaeffer AJ, Knauss JS, Landis JR, Probert KJ, Alexander RB, Litwin MS, Nickel JC, O’Leary MP, Nadler RB, Pontari MA, Shoskes DA, Zeitlin SI, Fowler JE Jr, et al, and Chronic Prostatitis Collaborative Research Network Study Group. Leukocyte and bacterial counts do not correlate with severity of symptoms in men with chronic prostatitis: the national institutes of health chronic prostatitis cohort study. *J Urol*. 2002; 168:1048–53.  
<https://doi.org/10.1097/01.ju.0000024762.69326.df>  
PMID:[12187220](#)
25. Müller A, Mulhall JP. Sexual dysfunction in the patient with prostatitis. *Curr Opin Urol*. 2005; 15:404–09.  
<https://doi.org/10.1097/01.mou.0000186847.25259.28>  
PMID:[16205492](#)
26. Cheung G, Sahai A, Billia M, Dasgupta P, Khan MS. Recent advances in the diagnosis and treatment of bladder cancer. *BMC Med*. 2013; 11:13.  
<https://doi.org/10.1186/1741-7015-11-13>  
PMID:[23327481](#)
27. Mitra AP, Cote RJ. Molecular pathogenesis and diagnostics of bladder cancer. *Annu Rev Pathol*. 2009; 4:251–85.  
<https://doi.org/10.1146/annurev.pathol.4.110807.092230> PMID:[18840072](#)
28. Tilki D, Burger M, Dalbagni G, Grossman HB, Hakenberg OW, Palou J, Reich O, Rouprêt M, Shariat SF, Zlotta AR. Urine markers for detection and surveillance of non-muscle-invasive bladder cancer. *Eur Urol*. 2011; 60:484–92.  
<https://doi.org/10.1016/j.eururo.2011.05.053>  
PMID:[21684071](#)
29. Yossepowitch O, Herr HW, Donat SM. Use of urinary biomarkers for bladder cancer surveillance: patient perspectives. *J Urol*. 2007; 177:1277–82.  
<https://doi.org/10.1016/j.juro.2006.11.066>  
PMID:[17382711](#)
30. Ardelt P, Grünemay N, Strehl A, Jilg C, Miernik A, Kneitz B, Butt E. LASP-1, a novel urinary marker for detection of bladder cancer. *Urol Oncol*. 2013; 31:1591–98.  
<https://doi.org/10.1016/j.urolonc.2012.02.002>  
PMID:[22481019](#)
31. Kelly JD, Dudderidge TJ, Wollenschlaeger A, Okoturo O, Burling K, Tulloch F, Halsall I, Prevost T, Prevost AT, Vasconcelos JC, Robson W, Leung HY, Vasdev N, et al. Bladder cancer diagnosis and identification of clinically significant disease by combined urinary detection of Mcm5 and nuclear matrix protein 22. *PLoS One*. 2012; 7:e40305.  
<https://doi.org/10.1371/journal.pone.0040305>  
PMID:[22792272](#)
32. Shariat SF, Karakiewicz PI, Palapattu GS, Lotan Y, Rogers CG, Amiel GE, Vazina A, Gupta A, Bastian PJ, Sagalowsky AI, Schoenberg MP, Lerner SP. Outcomes of radical cystectomy for transitional cell carcinoma of the bladder: a contemporary series from the bladder cancer research consortium. *J Urol*. 2006; 176:2414–22.  
<https://doi.org/10.1016/j.juro.2006.08.004>  
PMID:[17085118](#)
33. Barbieri CE, Cha EK, Chromecki TF, Dunning A, Lotan Y, Svatek RS, Scherr DS, Karakiewicz PI, Sun M, Mazumdar M, Shariat SF. Decision curve analysis assessing the clinical benefit of NMP22 in the detection of bladder cancer: secondary analysis of a prospective trial. *BJU Int*. 2012; 109:685–90.  
<https://doi.org/10.1111/j.1464-410X.2011.010419.x>  
PMID:[21851550](#)
34. Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths TR, N’Dow J, Nabi G, Cook J, Vale L. Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer. *Health Technol Assess*. 2010; 14:1–331.  
<https://doi.org/10.3310/hta14040>  
PMID:[20082749](#)

**SUPPLEMENTARY MATERIALS**

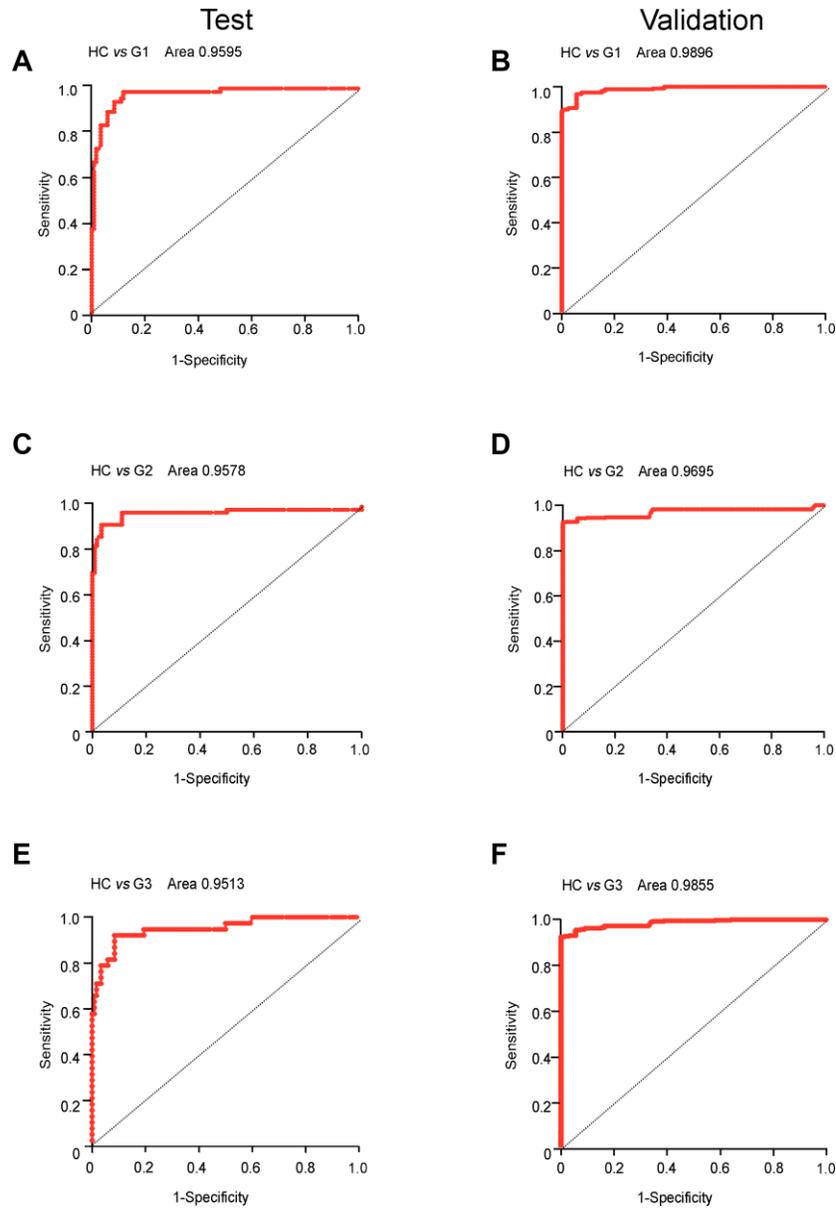
**Supplementary Figures**



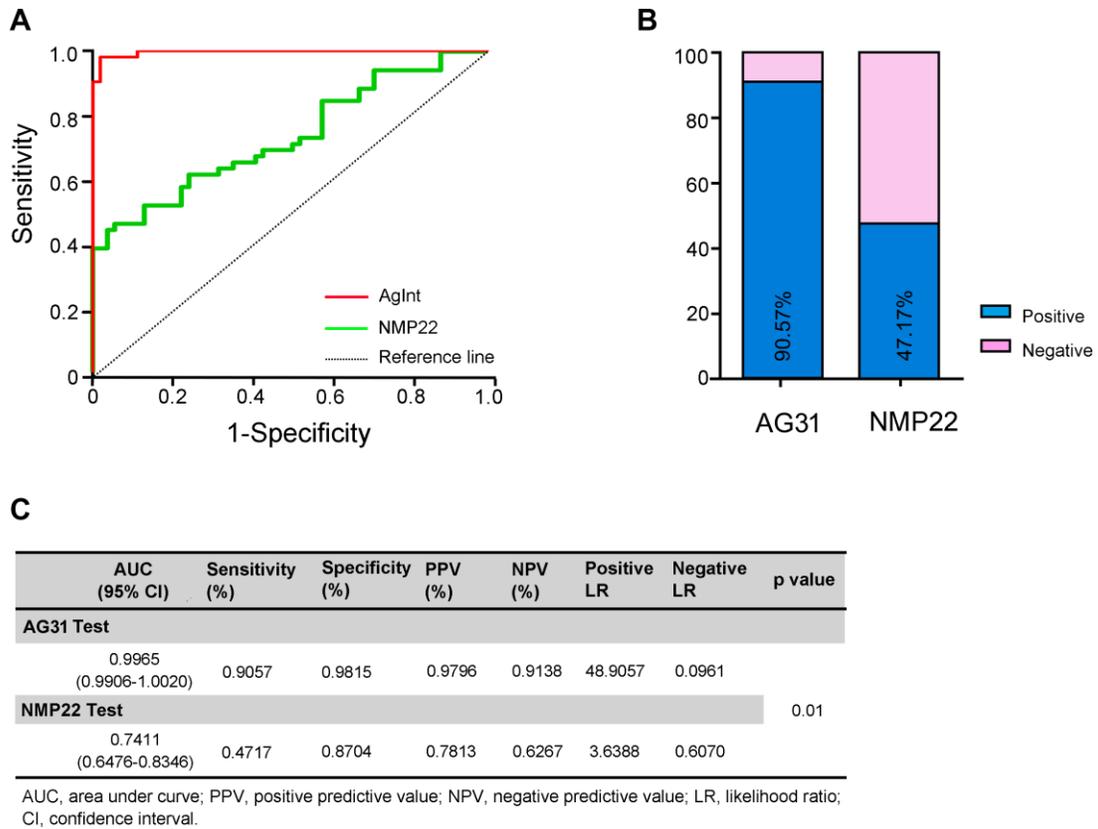
**Supplementary Figure 1. Development of the urinary AG31 test kit.** (A) Strategy of the ELISA-based AG31 test assay. (B) Calibration curve of the AG31 test with the standards. AG31: aberrantly glycosylated integrin  $\alpha3\beta1$ ; BCMab1: one monoclonal antibody against AG31; BCMab3: another monoclonal antibody against AG31; HRP: horseradish peroxidase; RLU: relative light unit. (C) The optimum cut-off value was set to 1991.



**Supplementary Figure 2. Diagnostic capabilities of urinary AG31 test in discriminating bladder cancer patients with different stages versus healthy controls.** (A, B) Receiver operating characteristic (ROC) curve for patients with Ta stage versus healthy controls in the test group (A), and in the validation group (B). (C, D) ROC curve for patients with T1 stage versus healthy controls in the test group (C), and in the validation group (D). (E, F) ROC curve for patients with T2 stage versus healthy controls in the test group (E), and in the validation group (F). (G, H) ROC curve for patients with T3 stage versus healthy controls in the test group (G), and in the validation group (H). (I, J) ROC curve for patients with T4 stage versus healthy controls in the test group (I), and in the validation group (J). A jagged curve denotes a ROC curve; a diagonal line represents a reference line. HC, healthy controls.



**Supplementary Figure 3. Diagnostic capabilities of urinary AG31 test in discriminating bladder cancer patients with different grades versus healthy controls.** (A, B) Receiver operating characteristics (ROC) curve for BC patients with G1 versus healthy controls in the test group (A), and in the validation group (B). (C, D) ROC curve for BC patients with G2 versus healthy controls in the test group (C), and in the validation group (D). (E, F) ROC curve for BC patients with G3 versus healthy controls in the test group (E), and in the validation group (F). A jagged curve denotes a ROC curve; a diagonal line represents a reference line. HC, healthy controls.



**Supplementary Figure 4. The urinary AG31 test is more sensitive and specific than the NMP22 test.** (A) ROC curves for AG31 and NMP22 for the patients with bladder cancer versus healthy controls in the test group. Jagged curves denote ROC curves; the diagonal line represents a reference line. (B) The rates of positive results for AG31 or NMP22 in all tested bladder cancer patients in the test group. (C) Comparison of the AG31 test with NMP22 test for the diagnosis of bladder cancer.

## Supplementary Tables

**Supplementary Table 1A. Age distribution of patients with renal cell carcinoma (RCC) in the test and validation groups.**

Variable	Test (n=23)		Validation (n=134)		p value
	No.	Percentage (%)	No.	Percentage (%)	
Age (years)					
	≤65	14	60.9	66	49.3
	>65	9	39.1	68	50.7

**Supplementary Table 1B. Age distribution of patients with prostate cancer (PC) in the test and validation groups.**

Variable	Test (n=16)		Validation (n=119)		p value
	No.	Percentage (%)	No.	Percentage (%)	
Age (years)					
	≤65	12	75.0	65	54.6
	>65	4	25.0	54	45.4

**Supplementary Table 1C. Age distribution of patients with cystitis in the test and validation groups.**

Variable	Test (n=21)		Validation (n=107)		p value
	No.	Percentage (%)	No.	Percentage (%)	
Age (years)					
	≤65	9	42.9	56	52.3
	>65	12	57.1	51	47.7

**Supplementary Table 1D. Age distribution of patients with nephritis in the test and validation groups.**

Variable	Test (n=25)		Validation (n=107)		p value
	No.	Percentage (%)	No.	Percentage (%)	
Age (years)					
	≤47	12	48.0	56	52.3
	>47	13	52.0	51	47.7

**Supplementary Table 1E. Age distribution of patients with prostatitis in the test and validation groups.**

Variable	Test (n=21)		Validation (n=111)		p value
	No.	Percentage (%)	No.	Percentage (%)	
Age (years)					
	≤57	10	47.6	59	53.2
	>57	11	52.4	52	46.8

**Supplementary Table 2. Urinary AG31 levels in different measurement groups in both test and validation groups.**

Group	Test					Validation				
	No.	Median (IQR)	Mean (SD)	times	p value	No.	Median (IQR)	Mean (SD)	times	p value
HC	118	576 (378-988)	859 (819)			609	660 (440-909)	840 (858)		
BC	184	6525 (3600-26585)	54768 (276559)			1130	23342 (11459-45637)	40281 (150259)		
RCC	23	648 (630-1440)	1011 (703)			134	667 (601-902)	916 (664)		
PC	16	875 (586-1054)	925 (427)			119	844 (611-1240)	1031 (777)		
Cystitis	21	885 (715-1201)	1070 (652)			107	863 (634-1375)	1064 (846)		
Nephritis	25	663 (569-1022)	995 (1048)			107	755 (560-990)	886 (494)		
Prostatitis	21	955 (866-1109)	1234 (1114)			111	914 (702-1045)	998 (656)		
Ta	99	5409 (3487-19089)	11584 (11522)	13.49		539.	12262 (10593-17900)	15250 (9247)	18.15	
T1	35	26860 (5326-43584)	27585 (21465)	32.11		385.	42489 (32469-50096)	39146 (14676)	46.60	
T2	19	6609 (2814-15289)	37709 (78583)	43.90		112.	106507 (10817-107541)	65989 (47413)	78.56	
T3	20	5933 (4183-72474)	75671 (203268)	88.09		70.	114721 (84494-124044)	96689 (47844)	115.11	
T4	11	109344 (3340-411814)	521252 (1022000)	606.81		24.	155164 (68320-189943)	336146 (982861)	400.17	
G1	69	5115 (3278-15313)	12667 (16439)	14.75		279	12153 (10437-27527)	20825 (20877)	24.79	
G2	76	11611 (4740-22854)	19828 (30457)	23.08		451	16621 (11065-35386)	27772 (29740)	33.06	
G3	38	33621 (2845-110717)	202477 (589592)	235.71		400	44916 (32455-56397)	67956 (247740)	80.90	

HC, healthy controls; BC, bladder cancer; RCC, renal cell carcinoma; PC, prostate cancer; IQR, interquartile range; SD, standard definition.

**Supplementary Table 3. Results for measurement of urinary AG31 in the diagnosis of bladder cancer with different stages and grades *versus* healthy controls.**

	Test					Validation								
	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Positive LR	Negative LR	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Positive LR	Negative LR
HC vs Ta	0.9614 (0.9341- 0.9887)	0.9091	0.9153	0.9000	0.9231	10.7273	0.0993	0.9717 (0.9606- 0.9829)	0.9184	0.9442	0.9357	0.9289	16.4496	0.0865
HC vs T1	0.9460 (0.8846- 1.0070)	0.9143	0.9153	0.7619	0.9730	10.7886	0.0937	0.9892 (0.9838- 0.9945)	0.9299	0.9442	0.9133	0.9551	16.6556	0.0743
HC vs T2	0.9286 (0.8283- 1.0290)	0.8947	0.9153	0.6296	0.9818	10.5579	0.1150	0.9913 (0.9855- 0.9971)	0.9196	0.9442	0.7518	0.9846	16.4724	0.0851
HC vs T3	0.9583 (0.9000- 1.0170)	0.9500	0.9153	0.6552	0.9908	11.2100	0.0546	0.9760 (0.9504- 1.0020)	0.9286	0.9442	0.6566	0.9914	16.6324	0.0757
HC vs T4	0.9753 (0.9401- 1.0110)	0.9091	0.9153	0.5000	0.9908	10.7273	0.0993	0.9835 (0.9561- 1.0110)	0.9167	0.9442	0.3929	0.9965	16.4191	0.0883
HC vs G1	0.9595 (0.9254- 0.9936)	0.9130	0.9153	0.8630	0.9474	10.7739	0.0950	0.9896 (0.9840- 0.9951)	0.9068	0.9442	0.8815	0.9567	16.2426	0.0987
HC vs G2	0.9578 (0.9192- 0.9964)	0.9079	0.9153	0.8734	0.9391	10.7132	0.1006	0.9695 (0.9565- 0.9826)	0.9268	0.9442	0.9248	0.9457	16.6011	0.0775
HC vs G3	0.9513 (0.9093- 0.9933)	0.9211	0.9153	0.7778	0.9730	10.8684	0.0863	0.9855 (0.9786- 0.9924)	0.9300	0.9442	0.9163	0.9536	16.6579	0.0741

HC, healthy control; AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; CI, confidence interval.

**Supplementary Table 4. Correlation between AG31 levels and clinicopathologic characteristics of bladder cancer patients in both test and validation groups.**

Variable	Test (n=184)		Validation (n=1130)	
	Negative (Percentage)	Positive (Percentage)	Negative (Percentage)	Positive (Percentage)
	17	167	87	1043
Age (years)				
≤63	9 (9.78%)	83 (90.22%)	31 (6.33%)	459 (93.67%)
>63	7 (8.14%)	79 (91.86%)	41 (6.73%)	568 (93.27%)
Missing	1	5	15	16
Gender				
Female	4 (12.50%)	28 (87.50%)	13 (5.56%)	221 (94.44%)
Male	12 (8.28%)	133 (91.72%)	61 (7.03%)	807 (92.97%)
Missing	1	6	13	15
Pathological stage				
Ta+T1	12 (8.96%)	122 (91.04%)	71 (7.68%)	853 (92.32%)
T2+T3+T4	5 (10.00%)	45 (90.00%)	16 (7.77%)	190 (92.23%)
Pathological grade				
G1	6 (8.70%)	63 (91.30%)	26 (9.32%)	253 (90.68%)
G2+G3	10 (8.77%)	104 (91.23%)	61 (7.17%)	790 (92.83%)
Missing	1			
Hematuria				
Visible	4 (21.05%)	15 (78.95%)	26 (7.22%)	334 (92.78%)
Non-visible	13 (7.88%)	152 (92.12%)	61 (7.92%)	709 (92.08%)
Tumor recurrence				
Yes	2 (7.41%)	25 (92.59%)	23 (12.23%)	165 (87.77%)
No	13 (9.70%)	121 (90.30%)	44 (4.96%)	843 (95.04%)
Missing	2	21	20	35

\*: Fisher exact test. Chi-square tests for all the other analysis.

# SNAI1 is a prognostic biomarker and correlated with immune infiltrates in gastrointestinal cancers

Jiaying Fang<sup>2</sup>, Zan Ding<sup>1</sup>

<sup>1</sup>The Institute of Metabolic Diseases, Baoan Central Hospital of Shenzhen, The Fifth Affiliated Hospital of Shenzhen University, Shenzhen 518102, Guangdong, China

<sup>2</sup>Medical Department, Affiliated Huadu Hospital, Southern Medical University (People's Hospital of Huadu), Shenzhen 518102, Guangdong, China

**Correspondence to:** Zan Ding; email: [dingzandz@126.com](mailto:dingzandz@126.com), <https://orcid.org/0000-0002-4051-8037>

**Keywords:** gastrointestinal cancers, EMT, immune infiltration, biomarker

**Received:** February 17, 2020

**Accepted:** June 4, 2020

**Published:** August 21, 2020

**Copyright:** Fang and Ding. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Epithelial-mesenchymal transition (EMT)-related genes play an important role in immunosuppression. However, the correlations of EMT-related genes to prognosis and tumor-infiltrating lymphocytes in different cancers remain unclear. TCGA, GEO databases were used to analyze the expression, prognosis, and immune infiltration of EMT markers in cancer. RT-qPCR, immunohistochemistry, and western blot were used to analysis the expression and prognosis of SNAI1 in gastrointestinal cancers. High SNAI1 expression was closely related with poorer overall survival in gastrointestinal cancers in TCGA cohort. High SNAI1 expression was closely related with poorer overall survival in gastrointestinal cancers, and was validated in GEO database. Simultaneously, high expression of SNAI1 correlates with clinical relevance of gastric cancer. Moreover, SNAI1 expression was associated with tumor-infiltrating immune cells in gastrointestinal cancers. In addition, RT-qPCR, immunohistochemistry, and western blot showed SNAI1 expression was higher in gastrointestinal cancers compared to the normal tissues. Finally, high SNAI1 expression was closely related with poorer overall survival and correlates with clinical relevance of gastrointestinal cancers in an independent validation cohort. In summary, the results approaches to suggest that SNAI1 can be used as a prognostic biomarker for determining prognosis and immune infiltration in gastrointestinal cancers.

## INTRODUCTION

Invasion and metastasis are the important characteristics of gastrointestinal (GI) cancers, and leads to a poor prognosis [1, 2]. Surgery, radiotherapy, and chemotherapy are the predominant treatments for GI [3, 4]. Although comprehensive treatment may cure some patients with early-stage GI, most patients are diagnosed with advanced stage [5]. With the rapid development of medical immunology and molecular biology techniques, immunotherapy as a new treatment method has received extensive attention in the field of cancer therapy [6]. Immunotherapy involves destroying

tumor cells by activating and training the patient's immune system to recognize tumor cells as targets [7, 8]. However, only 10% to 20% of the population can benefit from immunotherapy [9, 10]. Due to the heterogeneity of tumors, the current biomarkers for predicting prognosis have certain limitations. Therefore, this field requires new biomarkers as prognostic indicators to effectively enhance prognosis and individualized treatment.

Epithelial-mesenchymal transition (EMT) is a multistage process in which epithelial cells develop into mesenchymal-like cells with a large number of distinct genetic and epigenetic alterations [11]. EMT also occurs

in cancer, which endows invasive, metastatic, and immunosuppressive properties upon cancer cells that favor successful colonization of distal target organs [12]. The process of EMT is regulated by a variety of cytokines and transcription factors, among which the changes of migration and invasion are the most important characteristics of EMT, and the changes of EMT-related genes expression are the key to the formation of EMT [13]. These findings suggest the EMT-related genes play an important role in cancer progression, invasion, and metastasis.

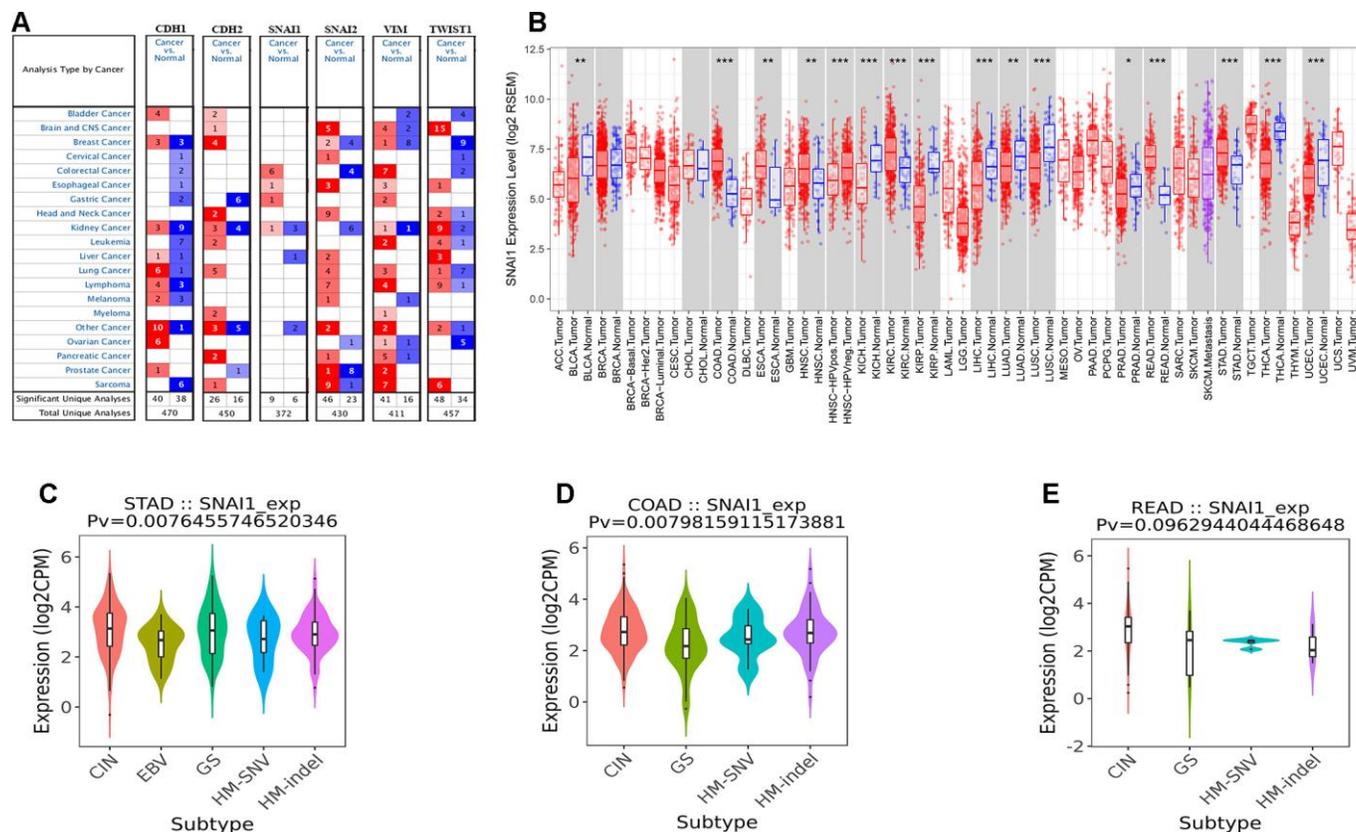
This study focused on six key EMT-related genes (CDH1 encoding E-cadherin, CDH2 encoding N-cadherin, EMT-induced transcription factor: SNAI1, SNAI2, TWIST1, and VIM) [14]. Firstly, we detected the expression of EMT-related genes in human cancers. Secondly, we comprehensively analyze EMT-related genes correlation with prognosis of cancer and was validated in GEO database. Moreover, we detected the relationship of SNAI1 and tumor-infiltrating immune cells in gastrointestinal cancers. Finally, we used

RT-qPCR, immunohistochemistry, and western blot to detect the expression of SNAI1 in gastrointestinal cancers and paired adjacent normal tissue, and SNAI1 was validated as a marker in an independent gastrointestinal cancers validation sample cohort.

## RESULTS

### The six EMT-related genes expression in human cancers

To analysis the expression of EMT-related genes in human cancers, the Oncomine database was used to analysis the EMT-related genes mRNA levels in human cancers. Specifically, as shown in Figure 1A, the SNAI1 expression was higher in colorectal cancer, esophageal cancer, gastric cancer, and kidney cancer. To further analysis EMT markers in human cancers, we compare the expression level of EMT-related genes in TCGA dataset (Supplementary Figure 1A–1E). Specifically, as shown in Figure 1B, compared with normal tissues, the SNAI1 expression was significantly higher in bladder



**Figure 1. The six EMT markers expression levels in different types of human cancers. (A)** Increased or decreased six EMT markers in datasets of different cancers compared with normal tissues in the Oncomine database. **(B)** SNAI1 expression levels in different tumor types from TCGA database were determined by TIMER (\*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001). **(C–E)** Correlation of SNAI1 expression and immune subtypes (wound healing, IFN-gamma dominant, inflammatory, lymphocyte depleted, TGF-β dominant) in COAD (colon adenocarcinoma), READ (rectum adenocarcinoma), and STAD (stomach adenocarcinoma).

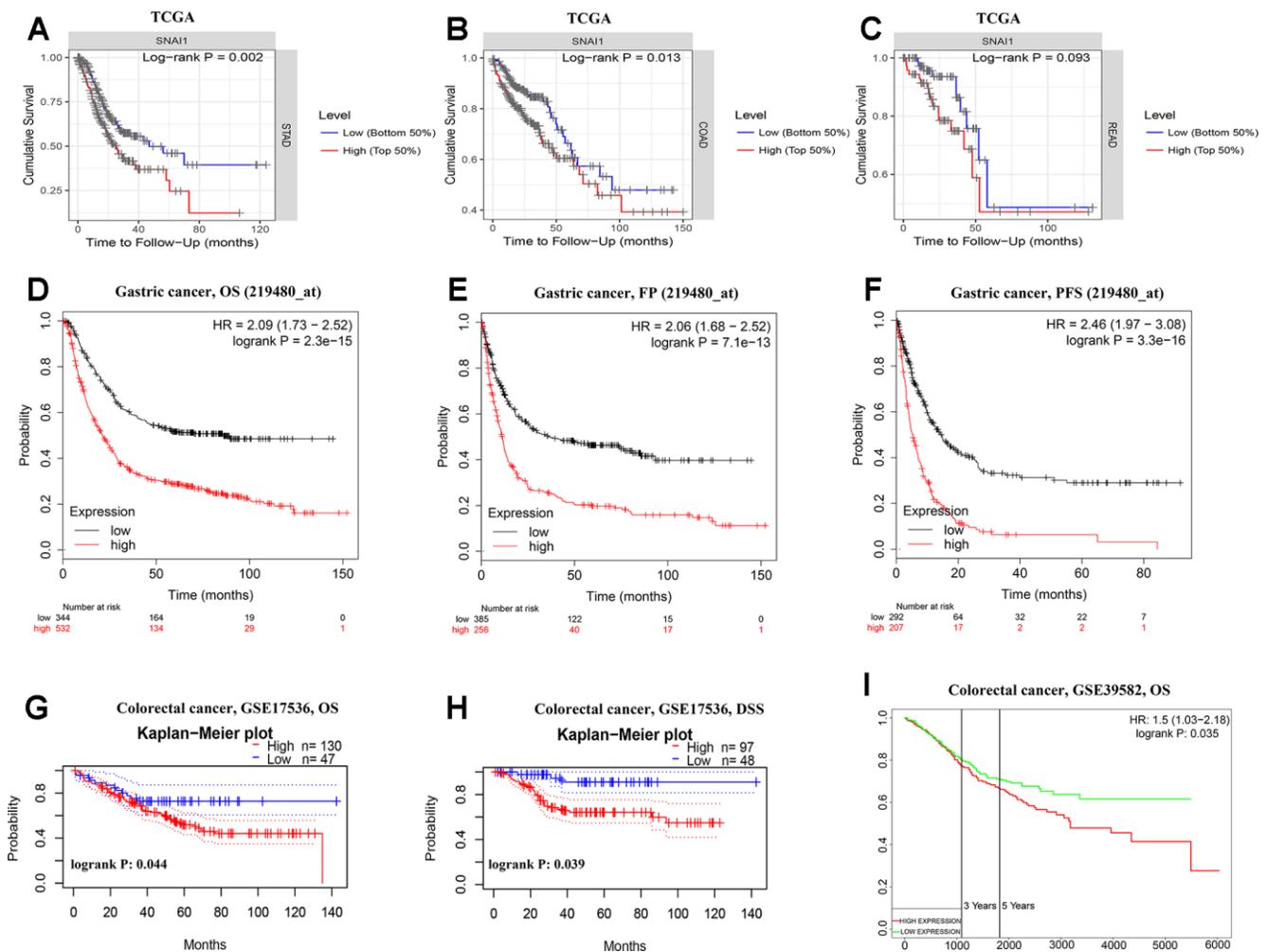
urothelial carcinoma (BLCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD). However, SNAI1 expression was significantly lower in kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), Prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), corpus Endometrial Carcinoma (UCEC).

### Survival analysis of six EMT-related genes in cancers

Next, to inspect whether six EMT-related genes were related with prognosis in cancer patients, GEPIA site was used to analysis the prognosis of genes in cancers by

using the TCGA dataset (Supplementary Figures 2–7). Notably, high SNAI1 expression levels was closely related with poorer prognosis of overall survival (OS) in STAD ( $p=0.002$ ), COAD ( $p=0.013$ ), ESCA ( $p=0.031$ ), KIRP ( $p=0.045$ ), LGG ( $p<0.001$ ), LUAD ( $p=0.026$ ), LUSC ( $p<0.001$ ), MESO ( $p=0.007$ ), OV ( $p=0.048$ ), and THYM ( $p=0.001$ ) (Figure 2A, 2B and Supplementary Figure 4). Simultaneously, high SNAI1 expression levels was moderate closely related with poorer prognosis of OS in READ ( $p=0.093$ ) (Figure 2C). Therefore, we focus on SNAI1 expression and prognosis in gastrointestinal cancers.

Next, to inspect whether SNAI1 expression was related with subtype of gastrointestinal cancers, we divide gastrointestinal cancers into five subtypes (CIN, EBV, HM-SNV, HM-indel) [15]. We found high SNAI1 expression was significantly related with subtype of



**Figure 2. Kaplan-Meier survival curves comparing the high and low expression of SNAI1 in gastrointestinal cancers in the TCGA dataset and GEO dataset. (A–C)** Survival curves of overall survival in COAD (colon adenocarcinoma), READ (rectum adenocarcinoma), and STAD (stomach adenocarcinoma) in TCGA cohorts. **(D–F)** Survival curves of OS, FP, and PFS in six STAD cohorts (GSE29013, GSE31210, GSE31908, GSE43580, GSE50081, GSE8894). **(G–I)** Survival curves of OS and DSS in two CRC cohorts (GSE17536, GSE39582).

STAD ( $p=0.008$ ) and COAD ( $p=0.008$ ), and moderate closely related with READ ( $p=0.096$ ) (Figure 1C–1E). To validate the prognostic potential of SNAI1 in STAD, we used Kaplan-Meier plotter database to validate the prognostic potential of SNAI1 in STAD. Interestingly, a cohorts including GSE62254, GSE14210, GSE15459, GSE22377, GSE29272, and GSE51105 indicated that high SNAI1 expression was closely related with poorer prognosis in STAD (OS HR = 2.09, 95% CI =1.73 to 2.52,  $P<0.001$ ; FP HR = 2.06, 95% CI =1.68 to 2.52,  $P<0.001$ ; PFS HR = 2.46, 95% CI = 1.97 to 3.08,  $P<0.001$ ) (Figure 2D–2F). Moreover, two cohorts including GSE17536, GSE39582 indicated that high SNAI1 expression was closely related with poorer prognosis in CRC (OS HR = 1.50, 95% CI =1.03 to 2.18,  $P=0.035$ ) (Figure 2G–2I).

### High expression of SNAI1 correlates with clinical relevance of STAD

Next, we examined the association between the SNAI1 expression and clinical relevance of STAD. As shown in Table 1, high SNAI1 expression was closely related with poorer prognosis in female (OS HR = 2.24,  $P<0.001$ ; PFS HR =1.91,  $P = 0.002$ ) and male (OS HR = 2.27,  $P <0.001$ ; PFS HR =2.43,  $P<0.001$ ). Moreover, high SNAI1 expression was closely related with poorer OS and PFS in stage 1 (OS HR = 3.63,  $P=0.011$ ; PFS HR =2.88,  $P = 0.023$ ), stage 2 (OS HR = 2.03,  $P=0.021$ ; PFS HR =1.93,  $P = 0.030$ ), stage 3 (OS HR = 1.98,  $P<0.001$ ; PFS HR =2.06,  $P <0.001$ ), and stage 4 (OS HR = 2.06,  $P<0.001$ ; PFS HR =2.30,  $P <0.001$ ) of STAD, and poorer OS and PFS in TNM stage. Furthermore, high SNAI1 expression was closely related with poorer prognosis in the lauren classification (OS HR = 1.89,  $P<0.001$ ; PFS HR =2.33,  $P <0.001$ ), moderate differentiation (OS HR = 1.94,  $P=0.046$ ; PFS HR =2.27,  $P =0.011$ ), negative (OS HR = 1.98,  $P<0.001$ ; PFS HR =2.02,  $P <0.001$ ) and positive (OS HR = 2.14,  $P<0.001$ ; PFS HR =2.36,  $P <0.001$ ) HER-2 status (Table 1). These results suggest that high SNAI1 expression can impact the prognosis in STAD with lymph node metastasis, and SNAI1 is an independent prognostic marker but can also predict the clinicopathological features of STAD.

### SNAI1 expression was correlated with tumor-infiltrating lymphocytes (TILs)

TILs are an independent predictor in cancers [16, 17]. Therefore, TIMER database was used to infer the relations between abundance of TILs and expression of SNAI1. The relationship between SNAI1 expression and TILs in different types of cancer was shown in Supplementary Figure 8. Specifically, SNAI1 expression was negative closely related with infiltrating levels

of tumor purity and B cell in STAD, and positively closely related with macrophages in STAD. Moreover, SNAI1 expression was negative closely related with infiltrating levels of tumor purity and B cell in COAD, and positively closely related with CD4+T cells, macrophages, dendritic cells, and neutrophils in COAD. Finally, SNAI1 expression was negative closely related with infiltrating levels of B cell and CD8+ T cells in READ, and positively closely related with CD4+T cells and dendritic cells in READ (Figure 3A–3C).

Next, to examine which TILs was related with gastrointestinal cancers, TISIDB database was used to infer the relations between abundance of 28 TILs and expression of SNAI1. The landscape of relationship between SNAI1 expression and TILs in different types of cancer was shown in Figure 4A. The relations between abundance of 28 TIL types and expression of SNAI1 in gastrointestinal cancers was significant correlated (Figure 4B–4R and Supplementary Figure 9). Next, we detected the associations between SNAI1 expression and immune subtypes in gastrointestinal cancers, and we divided the cells into six immunophenotypes (C1: wound healing, C2: IFN-gamma dominant, C3: inflammatory, C4: lymphocyte depleted, C5: immunologically quiet, C6: TGF- $\beta$  dominant) [20]. Specifically, SNAI1 expression was correlated with immune subtypes (wound healing, IFN-gamma dominant, inflammatory, lymphocyte depleted, TGF- $\beta$  dominant) in STAD and COAD, but not in READ (Figure 4S–4U). These results suggest that SNAI1 expression was associated with tumor-infiltrating immune cells in gastrointestinal cancers.

### Validation SNAI1 in gastrointestinal cancers

To further validate SNAI1 in gastrointestinal cancers, RT-qPCR was used to detect the SNAI1 mRNA expression in STAD, CRC, and paired adjacent normal tissue (PANT). Compared with the PANT group, the SNAI1 mRNA level was significantly higher in the CRC and GC group (Figure 6A, 6C). Next, we measured the SNAI1 protein level by immunohistochemistry and western blot, and the result showed compared with the PANT group, the SNAI1 level was significantly higher in the CRC and GC group (Figures 5A, 5B and 6A, 6C). Additionally, Kaplan–Meier and Cox’s proportional hazards regression model survival analysis revealed that patients with high expression levels of SNAI1 had shorter overall survival in CRC (HR = 1.71, 95% CI = 1.28 to 2.94,  $P = 0.023$ ) and GC (HR = 1.68, 95% CI = 1.23 to 2.57,  $P = 0.022$ ) (Figure 6B, 6D). Simultaneously, high expression of SNAI1 correlates with clinical relevance of CRC and GC (Table 2). To verify the relationship between SNAI1 and the diverse immune infiltrating cells, we focused on the correlations between SNAI1 and

**Table 1. Correlation of SNAI1 mRNA expression and clinical prognosis in gastric cancer with different clinicopathological factors by Kaplan-Meier plotter (GSE62254, GSE14210, GSE15459, GSE22377, GSE29272, GSE51105).**

Clinicopathological characteristics	Overall survival (n = 882)			Progression-free survival (n = 646)		
	N	Hazard ratio	P	N	Hazard ratio	P
<b>SEX</b>						
Female	244	2.24 (1.46–3.35)	1.0e-04	201	1.91 (1.25–2.91)	0.002
Male	567	2.27 (1.81–2.83)	1.5e-13	438	2.43 (1.90–3.11)	2.7e-13
<b>STAGE</b>						
1	69	3.63 (1.26–10.46)	0.011	60	2.88 (0.78–10.63)	0.023
2	145	2.03 (1.10–3.74)	0.021	131	1.93 (1.05–3.55)	0.030
3	319	1.98 (1.42–2.76)	3.6e-05	186	2.06 (1.38–3.07)	3.2e-05
4	152	2.06 (1.40–3.03)	2.0e-04	141	2.30 (1.54–3.43)	2.9e-05
<b>STAGE T</b>						
2	253	2.18 (1.42–3.34)	2.6e-05	239	1.93(1.27–2.92)	0.002
3	208	2.06 (1.40–3.04)	1.7e-04	204	1.80 (1.23–2.63)	0.002
4	39	2.40 (1.03–5.58)	0.036	39	5.66 (2.22–14.41)	5.5e-05
<b>STAGE N</b>						
0	76	2.80 (1.17–6.70)	0.016	72	2.63 (1.11–6.24)	0.023
1	232	2.45 (1.61–3.71)	1.3e-05	222	2.12 (1.44–3.14)	1.2e-05
2	129	2.67 (1.66–4.31)	2.7e-05	125	2.44 (1.54–3.84)	7.9e-05
3	76	2.76 (1.59–4.79)	1.8e-05	76	2.38 (1.38–4.10)	0.001
1+2+3	437	2.15 (1.65–2.80)	7.9e-09	423	1.90 (1.47–2.45)	5.0e-07
<b>STAGE M</b>						
0	459	2.09 (1.58–2.76)	1.5e-07	443	1.85 (1.41–2.41)	4.6e-06
1	58	2.13 (1.11–4.08)	0.020	56	1.81 (0.97–3.36)	0.059
<b>LAUREN CLASSIFICATION</b>						
Intestinal	336	2.78 (2.02–3.82)	5.0e-11	263	2.32 (1.61–3.34)	3.3e-06
Diffuse	248	1.89 (1.34–2.65)	2.0e-04	231	2.33 (1.47–3.69)	2.0e-04
<b>DIFFERENTIATION</b>						
Poor	166	1.25 (0.84–1.87)	0.277	121	1.60 (1.02–2.53)	0.041
Moderate	67	1.94 (1.00–3.75)	0.046	67	2.27 (1.19–4.34)	0.011
<b>PERFORATION</b>						
No	169	1.70 (1.13–2.58)	0.011	58	2.35 (1.20–4.62)	0.011
<b>TREATMENT</b>						
Surgery alone	393	1.71 (1.28–2.28)	2.5e-05	375	1.51 (1.14–2.00)	0.004
5 FU based adjuvant	158	1.66 (1.13–2.45)	0.009	153	1.68 (1.14–2.48)	0.008
Other adjuvant	80	2.55 (1.02–6.40)	0.039	80	2.34 (1.01–5.43)	0.042
<b>HER2 STATUS</b>						
Negative	641	1.98 (1.58–2.49)	2.0e-09	408	2.02 (1.54–2.65)	2.2e-07
Positive	425	2.14 (1.57–2.91)	7.7e-07	233	2.36 (1.65–3.39)	1.4e-06

immune marker sets of various immune cells in gastrointestinal cancers. We analyzed the correlations between SNAI1 expression and immune marker genes of different immune cells, included CD8+ T cell, T cell (general), B cell, monocyte, tumor-associated macrophage (TAM), M1 macrophage, M2 macrophage,

neutrophils, natural killer cell, dendritic cell, T helper cell (Th) 1, Th2, follicular helper T cell (Tfh), Th17, regulatory T cell (Treg), and T cell exhaustion. It is only when SNAI1 is related to all the markers of an immune cell that we think that SNAI1 is related to immune cells. The results revealed the SNAI1 expression level was

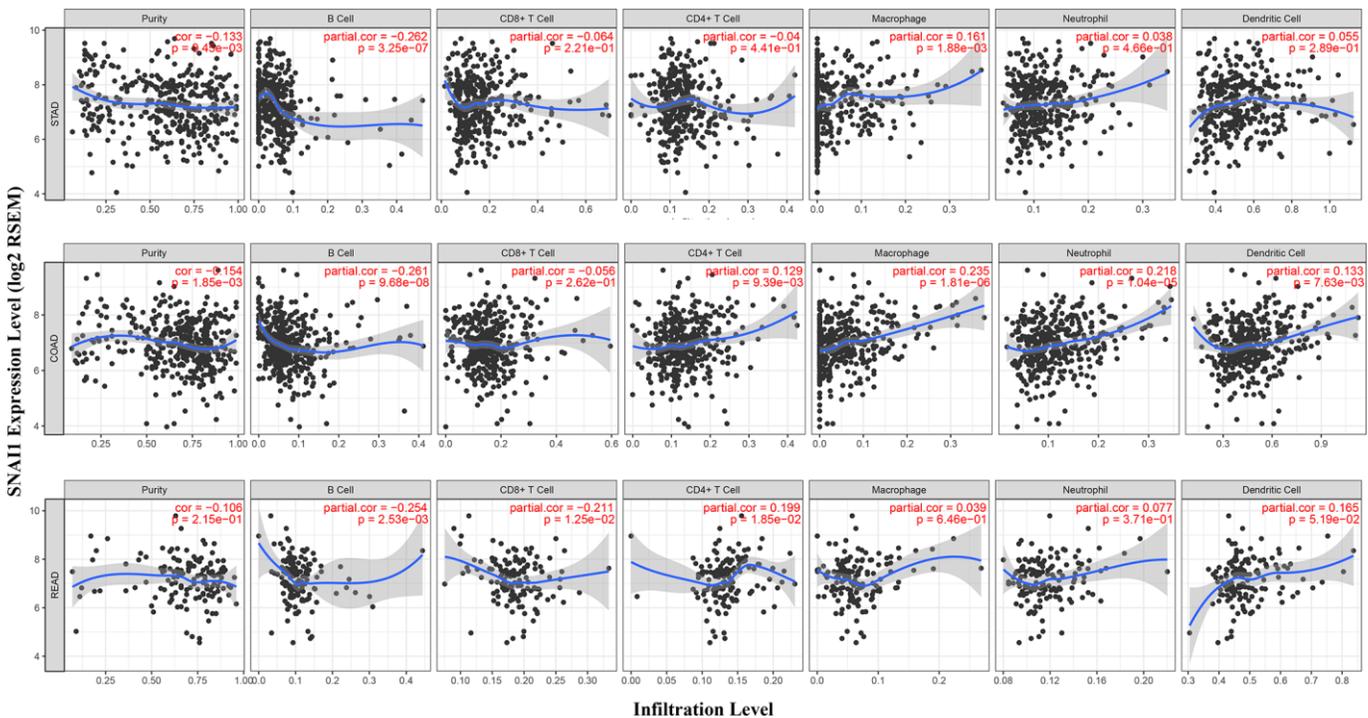
significantly correlated with M2 macrophage in STAD, and was significantly correlated with monocyte, TAM, M2 macrophage, dendritic cell, Th1, and treg in CRC (Table 3). Correlation results between SNAI1 and TILs are similar to those in TCGA database (Figure 4B–4R and Supplementary Figure 9). These findings suggest that SNAI1 may regulate M2 macrophage in gastrointestinal cancers.

## DISCUSSION

Due to advances in treatment, the mortality rate of tumors has been declining in recent years, a large part of which is due to immunotherapy. Immunotherapy represented by anti-PD-1/PD-L1 monoclonal antibody drugs and CAR-T cell therapy has attracted much attention, and encouraging results have continued. Both of them are essentially the ability of human autoimmune system to recruit and activate human core immune guardian-T cells to identify and clear cancer cells through antigen-antibody response [18]. However, not every patient responds to this treatment, especially in gastrointestinal cancers [19]. Therefore, there is an

urgent need to clarify and identify new immune-related therapeutic targets. High throughput technology has been widely employed to investigate gene expression in numerous tumors, providing a novel method to identify significant genes and explore on tumor progression and initiation.

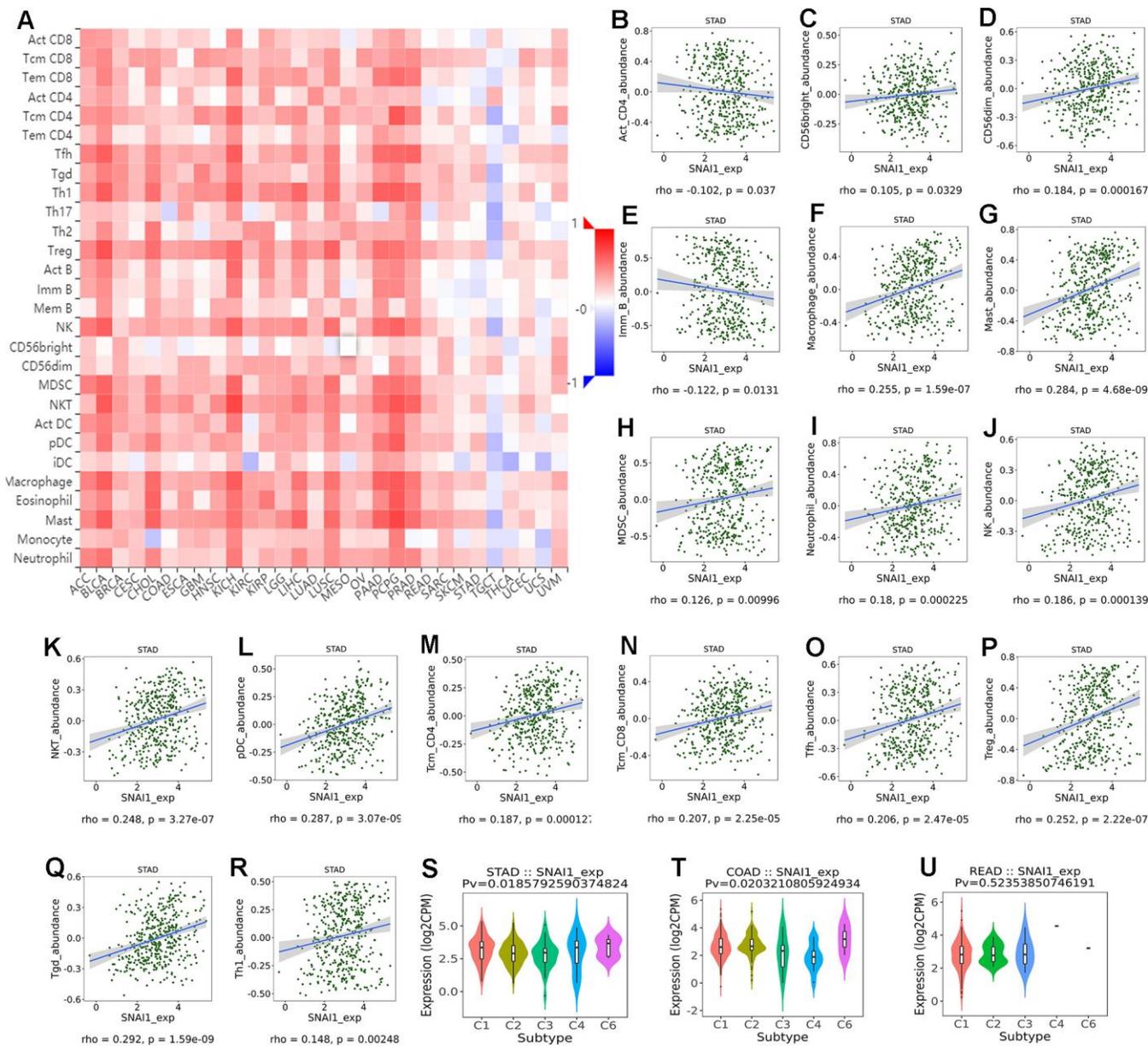
Here, we report that the expression and prognosis of six key EMT-related genes in cancer, and found high SNAI1 expression was closely related with poorer overall survival in gastrointestinal cancers, and was validated in GEO database. Simultaneously, high expression of SNAI1 correlates with clinical relevance of gastric cancer. Moreover, SNAI1 expression was associated with tumor-infiltrating immune cells in gastrointestinal cancers. In addition, RT-qPCR, immunohistochemistry, and western blot showed SNAI1 expression was higher in gastrointestinal cancers compared to the normal tissues. Finally, high SNAI1 expression was closely related with poorer overall survival and correlates with clinical relevance of gastrointestinal cancers in an independent validation cohort.



**Figure 3. Correlation of SNAI1 expression with immune infiltration level in COAD (colon adenocarcinoma), READ (rectum adenocarcinoma), and STAD (stomach adenocarcinoma).** (A) SNAI1 expression is significantly negatively related to tumor purity and infiltrating levels of B cells and has significant positive correlations with infiltrating levels of macrophages in STAD, other than CD8+ T cells, CD4+ T cell, neutrophils, and dendritic cells. (B) SNAI1 expression is significantly negatively related to tumor purity and infiltrating levels of B cells and has significant positive correlations with infiltrating levels of CD4+ T cell, macrophages, neutrophils, and dendritic cells in COAD, other than CD8+ T cells. (C) SNAI1 expression has no significant correlations with tumor purity, macrophages, and neutrophils in READ.

E-cad is a  $\text{Ca}^{2+}$ -dependent transmembrane glycoprotein closely related to intercellular adhesion [20]. Studies have found that the expression of E-cad is down-regulated or even completely lost in the malignant progression of epithelial tumors, which leads to the weakening of adhesion between tumor cells and the transformation from benign and non-invasive to malignant and invasive phenotypes [21]. In the study of clinical samples of human gastric cancer, Rosivatz et al.

found that the down-regulation of E-cad expression was closely related to the up-regulation of transcription factors Snail, Twist, and SIP1 [22], while Wang et al. also found high expression of transcription factors such as Snail, Twist and Slug in immortalized gastric epithelial cell line Ges-1 and human gastric cancer cell lines MGC-803, BGC-823 and SGC-7901, which were negative for E-cad expression [23]. These transcription factors can bind to the E-box element in the promoter



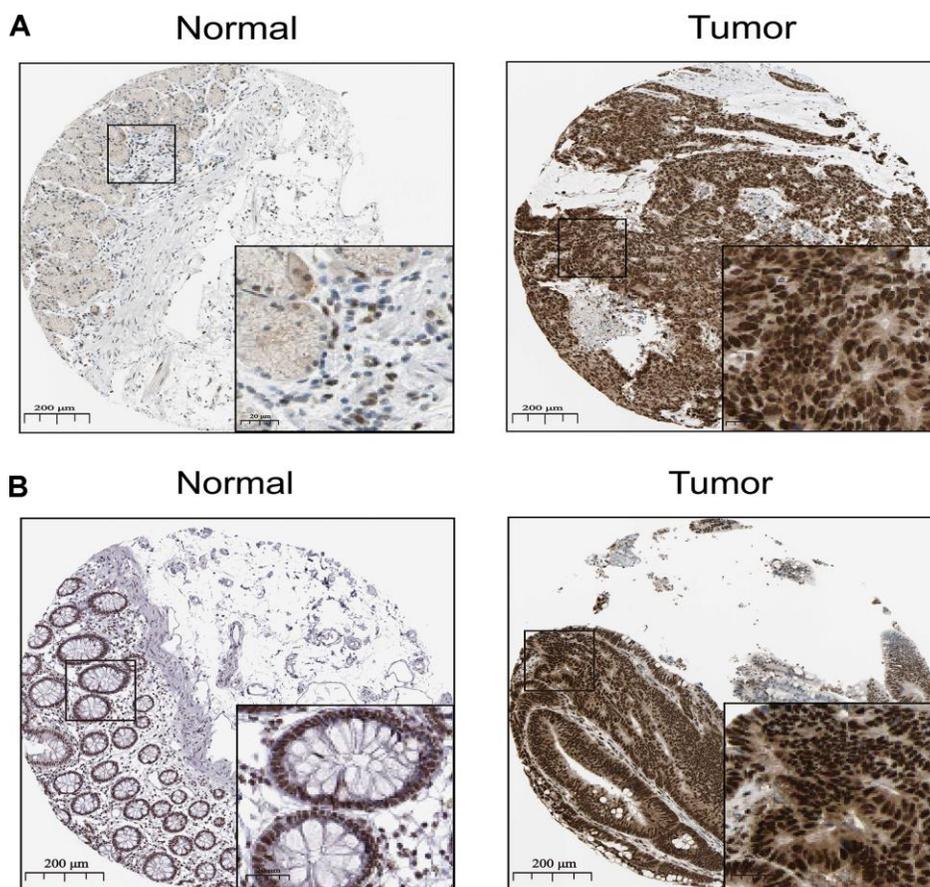
**Figure 4. Correlation of SNAI1 expression with immune cells in cancer.** (A) The landscape of relationship between SNAI1 expression and TILs in different types of cancer (red is positive correlated and blue is negative correlated). (B–R) SNAI1 expression was positively closely related with infiltrating levels of CD56bright, CD56dim, macrophage, mast, MDSC, neutrophil, NK, NKT, pDC, tcm\_CD4, tcm\_CD8, Tfh, Tgd, Th1, Treg in stomach cancer, and was negatively correlated with infiltrating levels of act\_CD4, Imm\_B in stomach cancer. (S–U) Correlation of SNAI1 expression and immune subtypes (wound healing, IFN-gamma dominant, inflammatory, lymphocyte depleted, TGF- $\beta$  dominant) in COAD (colon adenocarcinoma), READ (rectum adenocarcinoma), and STAD (stomach adenocarcinoma).

sequence of E-cad gene and inhibit the transcription of E-cad gene. In this study high SNAI1 expression was found in 8 types of cancer (Figure 1B).

Snail, a zinc finger protein, was initially thought to affect EMT, mainly by inhibiting the expression of E-cad, but it was later found that its mechanism of promoting EMT was also related to its down-regulation of epithelial cell characteristic markers (such as claudins, occludins and cytokeratins) and up-regulation of stromal cell characteristic markers (such as fibronectin and vitronectin) [24]. SNAI1 has a certain value in evaluating the disease progression and survival prognosis of patients with gastric cancer. Its expression level in gastric cancer is significantly correlated with tumor size, degree of differentiation, clinical stage, lymph node and distant metastasis, and the overall survival rate of gastric cancer patients with high expression of SNAI1 is significantly lower than that of patients with low expression of SNAI1 [25]. In cell experiment, it was also found that up-regulation and inhibition of SNAI1 expression could enhance and inhibit the migration and invasion ability of gastric

cancer cells in vitro, respectively [26, 27]. In addition, studies have found that cyclooxygenase-2 COX-2) can regulate the expression of E-cad in gastric cancer by affecting nuclear factor  $\kappa$ B and SNAI1 pathway, which is also one of the mechanisms of COX-2 regulating the invasion and metastasis of gastric cancer [28]. All these suggest that snail plays an important role in the invasion and metastasis of gastrointestinal cancers.

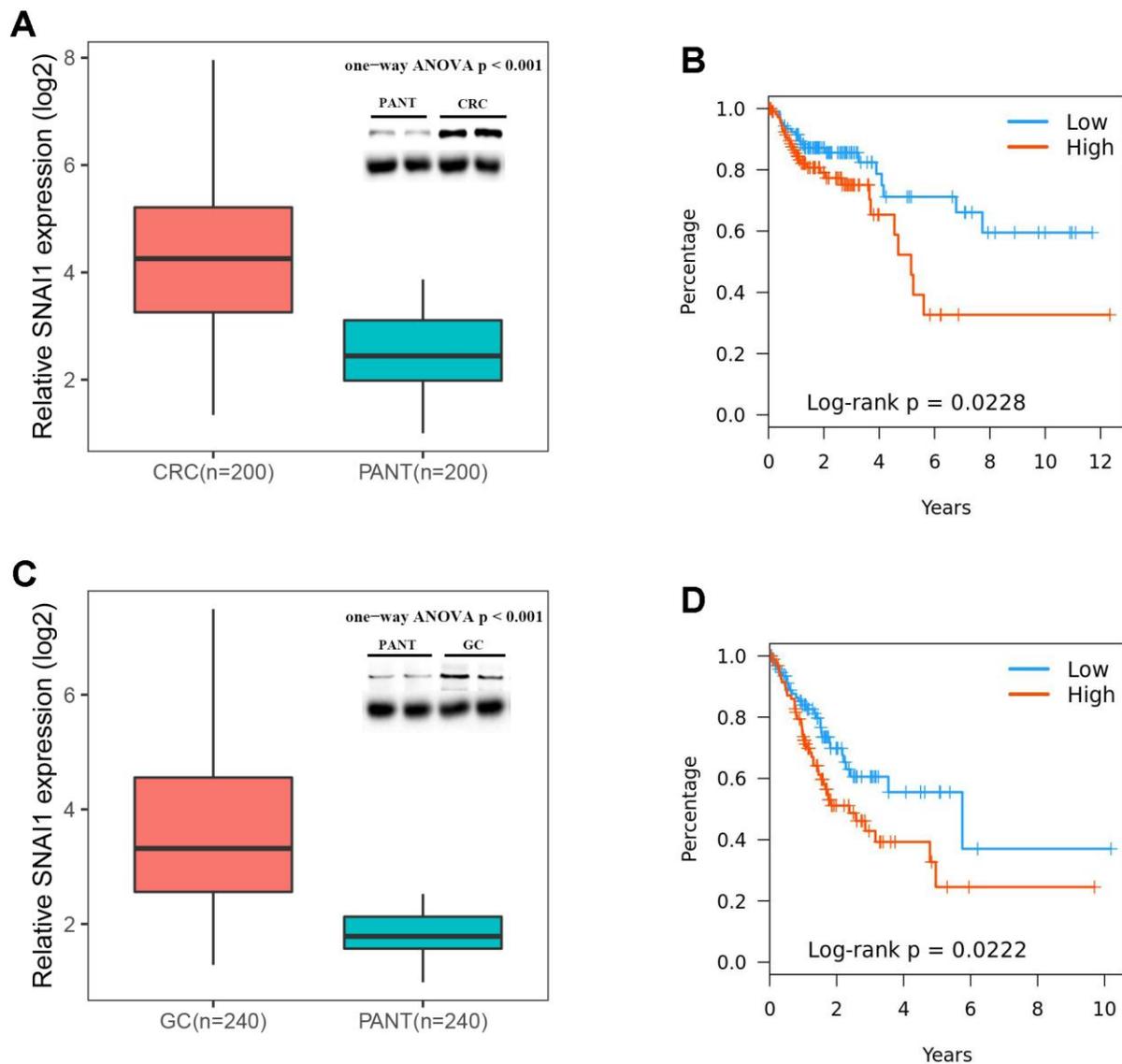
EMT gives cells the ability to transfer and invade, including stem cell characteristics, reduces apoptosis and aging, and promotes immunosuppression [29]. NF- $\kappa$ B up-regulates the transcription of SNAI1 gene by binding to the promoter region of EMT gene [30]. Related studies have shown that NF- $\kappa$ B can also prevent the apoptosis of epithelial cells with deteriorating tendency in inflammation-related tumors [31]. In addition, the synergistic effect of NF- $\kappa$ B and STAT3 also promotes the generation and metastasis of inflammation-related tumors. STAT3 can not only activate NF- $\kappa$ B, but also induce EMT by regulating SNAI1 to inhibit the activity of E-cadherin promoter [32]. Simultaneously, it is believed that the occurrence



**Figure 5. The distribution of SNAI1 in cancer.** (A, B) Representative IHC images of SNAI1 expression in normal stomach tissues, stomach cancer tissues(A), normal tissues, and colorectal cancer tissues (B).

of EMT in gastrointestinal cancer cells is the result of the interaction between the acting factors in the tumor microenvironment and gastrointestinal cancer cells. The acting factors transfer the extracellular signal into the cell by binding to the specific receptors on the cell surface, and activate the nuclear transcription factors of SNAI1 through intracellular TGF- $\beta$ , Wnt, phosphatidylinositol 3-kinase/protein kinase B and other signal transduction pathways. Regulate the expression of downstream genes and the transformation of epithelial cells into stroma, and eventually mediate the transformation of normal epithelial cells into gastrointestinal cancer cells [33].

GI cancers microenvironment plays a critical role in controlling the cancer cell fate, treatment and prognosis. However, the role of EMT in reshaping tumor microenvironment (TME) is unclear. Tumor-associated macrophage is the main type of host immune cells in TME. They regulate tumor colonization and progression by regulating tumor invasion, local tumor immunity and angiogenesis. Hsieh et al. reported that EMT transcription factor Snail directly activates the transcription of miR-21, resulting in miR-21-rich tumor-derived exosome. The exosomes containing miR-21 were phagocytized by CD14+ human monocytes, thus inhibiting the expression of M1 markers and increasing



**Figure 6. Expressions of SNAI1 is positive correlated in CRC (colorectal cancer) and GC (gastric cancer).** (A) SNAI1 mRNA and protein levels are shown for the CRC and paired adjacent normal tissue (PANT). (B) Kaplan–Meier analysis of overall survival based on SNAI1 mRNA levels in 200 cases of CRC patients. (C) SNAI1 mRNA and protein levels are shown for the GC and paired adjacent normal tissue (PANT). (D) Kaplan–Meier analysis of overall survival based on SNAI1 mRNA levels in 240 cases of GC patients.

**Table 2. Univariate and multivariate analyses of clinicopathological characteristics, SNAI1 with overall survival in independent validation cohort.**

Parameters	Univariate analysis <sup>a</sup>		Multivariate analysis <sup>b</sup>	
	HR (95% CI)	P value	HR (95% CI)	P value
<b>CRC (n=200)</b>				
Gender(male vs. female)	1.024 (0.681–1.547)	0.925	NA	NA
Age (y, >60 vs ≤60)	1.192 (0.835–1.927)	0.346	NA	NA
TNM stage (III and IV vs. I and II)	<i>2.101(1.218–3.709)</i>	<i>0.002</i>	<i>1.914(1.165–3.226)</i>	<i>0.009</i>
Pathologic M stage (M1 vs. M0)	<i>3.127(1.852–4.467)</i>	<i>0.003</i>	<i>2.519(1.248–3.814)</i>	<i>0.024</i>
Vascular invasion (negative vs. positive)	<i>2.300(1.113–3.671)</i>	<i>0.014</i>	<i>1.562(0.779–3.530)</i>	0.136
SNAI1 (≥median vs. < median)	<i>1.708 (1.276–2.937)</i>	<i>0.023</i>	<i>1.449 (1.091–2.279)</i>	<i>0.038</i>
<b>GC (n=240)</b>				
Gender(male vs. female)	1.447 (0.919–2.152)	0.453	NA	NA
Age (y, >60 vs ≤60)	1.237 (0.797–2.294)	0.674	NA	NA
TNM stage (III and IV vs. I and II)	<i>1.669(1.192–2.547)</i>	<i>&lt;0.001</i>	<i>1.436(1.023–2.067)</i>	<i>0.007</i>
Pathologic M stage (M1 vs. M0)	<i>1.514 (1.038–2.027)</i>	<i>&lt;0.001</i>	<i>1.336 (1.023–1.933)</i>	<i>0.012</i>
Vascular invasion (negative vs. positive)	<i>1.742 (1.228–2.368)</i>	<i>&lt;0.001</i>	<i>1.502 (1.043–2.388)</i>	<i>0.008</i>
SNAI1 (≥median vs. < median)	<i>1.680 (1.234–2.574)</i>	<i>0.022</i>	<i>1.581 (1.126–2.256)</i>	<i>0.038</i>

<sup>a</sup>The data were subjected to Cox’s proportional hazards regression model. Bold italics indicate statistically significant values (P < 0.05),

<sup>b</sup>Multivariate analysis used stepwise addition and removal of clinical covariates found to be associated with survival in univariate models (P < 0.05) and final models include only those covariates that were significantly associated with survival (Wald statistic, P < 0.05). Bold italics indicate statistically significant values (P < 0.05).

**Table 3. Correlation analysis between SNAI1 expression and markers expression of immune cells.**

Immune cells	Gene markers	STAD		CRC	
		Cor	P	Cor	P
<b>CD8+ T cell</b>	CD8A	-0.084	0.091	0.171	*
	CD8B	0.163	***	0.065	0.321
<b>T cell (general)</b>	CD3D	-0.122	0.017	0.154	0.014
	CD3E	-0.089	0.071	0.210	***
	CD2	-0.089	0.074	0.221	***
<b>B cell</b>	CD19	-0.072	0.152	0.111	0.060
	CD79A	-0.094	0.059	0.148	0.025
<b>Monocyte</b>	CD86	0.124	0.018	0.500	***
	CD115 (CSF1R)	0.192	***	0.473	***
<b>TAM</b>	CCL2	0.300	***	0.573	***
	CD68	0.132	0.038	0.451	***
	IL10	-0.028	0.580	0.510	***
<b>M1 Macrophage</b>	INOS (NOS2)	-0.056	0.260	-0.141	0.019
	IRF5	0.052	0.300	0.170	*
	COX2(PTGS2)	0.200	***	0.231	***
<b>M2 Macrophage</b>	CD163	0.243	***	0.423	***
	VSIG4	0.222	***	0.470	***
	MS4A4A	0.141	***	0.480	***
<b>Neutrophils</b>	CD66b (CEACAM8)	0.140	***	0.052	0.392
	CD11b (ITGAM)	0.089	0.071	0.481	***
	CCR7	-0.024	0.632	0.200	***

<b>Natural killer cell</b>	KIR2DL1	-0.025	0.621	0.172	*
	KIR2DL3	-0.013	0.800	0.160	*
	KIR2DL4	-0.046	0.354	0.110	0.064
	KIR3DL1	-0.058	0.244	0.086	0.161
	KIR3DL2	-0.052	0.288	0.158	*
	KIR3DL3	0.002	0.971	0.053	0.382
	KIR2DS4	0.091	0.134	0.185	0.113
<b>Dendritic cell</b>	HLA-DPB1	-0.054	0.276	0.300	***
	HLA-DQB1	-0.082	0.098	0.183	**
	HLA-DRA	-0.091	0.068	0.322	***
	HLA-DPA1	-0.053	0.284	0.293	***
	BDCA-1(CD1C)	-0.018	0.722	0.211	***
	BDCA-4(NRP1)	0.366	***	0.561	***
	CD11c (ITGAX)	0.143	*	0.454	***
<b>Th1</b>	T-bet (TBX21)	-0.053	0.281	0.221	***
	STAT4	-0.034	0.500	0.264	***
	STAT1	0.029	0.562	0.264	***
	IFN- $\gamma$ (IFNG)	-0.074	0.133	0.233	***
	TNF- $\alpha$ (TNF)	0.243	***	0.389	***
<b>Th2</b>	GATA3	0.017	0.742	0.367	***
	STAT6	-0.045	0.364	-0.036	0.557
	STAT5A	0.074	0.133	0.242	***
<b>Tfh</b>	IL13	0.098	0.049	0.240	***
	BCL6	0.194	***	0.499	***
<b>Th17</b>	IL21	-0.110	0.032	0.123	0.053
	STAT3	0.200	***	0.261	***
<b>Treg</b>	IL17A	0.017	0.738	-0.153	0.015
	FOXP3	0.058	0.244	0.364	***
	CCR8	0.063	0.223	0.383	***
	STAT5B	0.144	*	0.345	***
	TGF $\beta$ (TGFB1)	0.444	***	0.566	***
<b>T cell exhaustion</b>	PD-1 (PDCD1)	-0.003	0.959	0.205	***
	LAG3	-0.066	0.183	0.072	0.232
	CTLA4	0.010	0.831	0.300	***
	TIM-3 (HAVCR2)	0.087	0.081	0.474	***
	GZMB	-0.052	0.304	-0.02	0.757

STAD: Stomach adenocarcinoma; CRC: Colorectal cancer; TAM: Tumor-associated macrophage; Th: T helper cell; Tfh: Follicular helper T cell; Treg: Regulatory T cell; Cor, R value of Spearman's correlation; \*P < 0.01; \*\*P < 0.001; \*\*\*P < 0.0001.

the expression of M2 markers [34]. Faget et al. reported that high expression of SNAI1 reduces T cell homing, changes angiogenesis, leads to hypoxia, and blocks anti-PD1 immunotherapy. Therefore, SNAI1 accelerates the progression of the disease and increases the infiltration of neutrophils in the tumor to maintain the harmful tumor microenvironment [35]. At the same time, CD47 is the direct target of SNAI1. By using CD47 to block the genetic targeting of SNAI1, the phagocytosis of EMT-activated cells can be rescued [36]. Hence, the important aspect of our study is to emphasize the role of SNAI1 in immune cell infiltration and immune escape in GI cancers. Specifically, SNAI1 expression was negative

closely related with infiltrating levels of tumor purity and B cell in STAD, and positively closely related with macrophages in STAD. Moreover, SNAI1 expression was negative closely related with infiltrating levels of tumor purity and B cell in COAD, and positively closely related with CD4+T cells, macrophages, dendritic cells, and neutrophils in COAD. Finally, SNAI1 expression was negative closely related with infiltrating levels of B cell and CD8+ T cells in READ, and positively closely related with CD4+T cells and dendritic cells in READ. Therefore, the cross-talk between SNAI1 and tumor microenvironment might be an important mechanism for the development and progression of GI cancers.

Nevertheless, more functional and mechanism experiments are needed for further verification.

In summary, we applied integrated bioinformatics approaches to suggest that high SNAI1 was closely related with prognosis and immune infiltrating levels in gastrointestinal cancer. Therefore, SNAI1 can be used as a prognostic biomarker for determining prognosis and immune infiltration in gastrointestinal cancer, which might provide a novel direction to explore the pathogenesis of gastrointestinal cancer

## MATERIALS AND METHODS

### Data source and processing

Using TIMER (Tumor Immune Estimation Resource, <https://cistrome.shinyapps.io/timer/>) site to analysis the expression of EMT markers in cancers, the mRNA profiling information is from TCGA (The Cancer Genome Atlas, <https://cancergenome.nih.gov/>) database [37]. We also Oncomine database (<https://www.oncomine.org/resource/login.html>) to analysis the expression of SNAI1 in cancers [38]. Bayes test was used to select EMT markers with a change  $\geq 2$  fold and a *P* value cutoff of 0.05 was defined as statistically significant.

### Survival analysis

GEPIA (Gene Expression Profiling Interactive Analysis, <http://gepia.cancer-pku.cn/>) site was used to analysis the prognosis of EMT markers in cancers by using the TCGA dataset [39]. A Cox *P*-value  $< 0.05$  was defined as statistically significant. Moreover, we also used Kaplan-Meier plotter database (<http://kmplot.com/analysis/index.php?p=service&cancer=gastric>) to validate the prognosis of SNAI1 in gastrointestinal cancers [40]. Kaplan-Meier plotter can assess the effect of 54,675 genes on survival using 10,461 cancer samples. These samples include 5,143 breast, 1,816 ovarian, 2,437 lung, and 1,065 gastric cancer samples on the HGU133 Plus 2.0 array with a mean follow-up of 69, 40, 49, and 33 months, respectively. The hazard ratio (HR) with 95% confidence intervals (CI) and log-rank *p*-value were estimated.

### Immune infiltration

To revealed the immune infiltration of SNAI1 in gastrointestinal cancers, we used TISIDB (tumor-immune system interactions and drugbank, <http://cis.hku.hk/TISIDB/index.php>) database to infer the relations between abundance of tumor-infiltrating lymphocytes (TILs) and expression of SNAI1. The immune-related signatures of 28 TIL types from

Charoentong's study, which can be viewed in the download page. The relative abundance of TILs was inferred by using genomic variation analysis based on gene expression profiles [41].

### Patients

A total of 440 gastrointestinal cancers patients (200 CRC patients, 240 GC patients) from January 2009 to December 2018, were consecutively recruited in analyses, with clinical information. The clinicopathological characteristics were showed in Table 4.

### Immunohistochemistry

200 colorectal cancer tissues, 240 gastric cancer tissues and corresponding adjacent tissues were collected to explore the expression of SNAI1 in the tissue samples by using immunohistochemical staining (IHC). IHC staining was performed according to the manufacturer's instructions. Tissue samples were fixed in 10% formalin embedded in paraffin and cut into slices (4 $\mu$ m). The frozen tissue sections were dewaxed in xylene and rehydrated in descending grades of ethanol. Endogenous peroxidase activity was blocked by 0.3% H<sub>2</sub>O<sub>2</sub>. After heat-induced antigen retrieval, the sections were blocked with 10% rabbit serum for 10min. Subsequently, the sections were incubated with primary antibodies against SNAI1 (1:200) overnight at 4°C, and incubated with an anti-rabbit secondary antibody (1:2,000) for 30 min at 37°C. After that, the sections were visualized with DAB for 10 min at 37°C and counterstained with hematoxylin for 30 sec at 37°C, followed by dehydration with gradient ethanol and sealing with neutral gum.

### Quantitative reverse transcription polymerase chain reaction (qRT-PCR) assays

Total RNA from cells or tissues was isolated using TRIzol (Invitrogen, Canada) reagent, the specific operation is carried out with reference to the instructions for the operation of the kit. RNA (1  $\mu$ g) was converted into cDNA using the RevertAid First Strand cDNA Synthesis Kit (Takara, China). qRT-PCR was performed using SYBR Green Mixture (Takara, China) in the ABI StepOne-Plus System (ABI7500, USA). Target gene expression was normalized against GAPDH. The primer sequences are 5'-CACCTCCAGACCCACTCAGAT-3' (sense) and 5'-CCTGAGTGGGGTGGGAGCTTCC-3' (antisense)

### Western blot

Using the kit to measure protein concentration. An equal amount of protein is added to the gel. The protein

**Table 4. Clinicopathologic parameters of patients with CRC and GC in validation cohort.**

Clinicopathologic	CRC (n=200)	GC (n=240)
<b>Gender</b>		
Male	122	169
Female	78	71
<b>Age</b>		
60 years or younger	62	125
Older than 60 years	138	115
<b>TNM stage</b>		
I	29	28
II	83	77
III	65	103
IV	23	37
<b>Pathologic M stage</b>		
M0	167	203
M1	23	32
<b>Lymphovascular invasion</b>		
Negative	163	194
Positive	37	46

was resolved on SDS-PAGE under denatured reducing conditions and transferred to nitrocellulose membranes. The membranes were blocked with 5% non-fat dried milk at room temperature for 30 min and incubated with primary antibodies. The membranes were washed and incubated with horseradish peroxidase-conjugated secondary antibody. Scanning with the Odyssey two-color infrared fluorescence imaging system.

### Statistical analyses

The expression of EMT-related genes in cancer was using TIMER and Oncomine database. The survival curve was generated by GEPIA, PrognoScan, and Kaplan Meier diagrams. One-Way ANOVA was used to compare the expression level of SNAI1 in tumor and normal tissues. To control FDR, the *p*-value was set to less than 0.05. If any independent value is lost, the whole sample is excluded from statistical analysis. Above all, statistical analysis was performed in R version 3.5. *p*-values < 0.05 were considered statistically significant.

### AUTHOR CONTRIBUTIONS

Jiaying Fang conducted the qRT-PCR assays, Western blotting assays, and provided the tissue samples and the clinical data. Zan Ding and Jiaying Fang designed the research, analyzed the data and wrote the paper. All authors read and approved the final manuscript.

### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

### FUNDING

This work was received no funding

### REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:31912902
2. Mehlen P, Puisieux A. Metastasis: a question of life or death. *Nat Rev Cancer.* 2006; 6:449–58. <https://doi.org/10.1038/nrc1886> PMID:16723991
3. Schmidt B, Lee HJ, Ryeom S, Yoon SS. Combining bevacizumab with radiation or chemoradiation for solid tumors: a review of the scientific rationale, and clinical trials. *Curr Angiogenesis.* 2012; 1:169–79. <https://doi.org/10.2174/2211552811201030169> PMID:24977113
4. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin.* 2019; 69:363–85. <https://doi.org/10.3322/caac.21565> PMID:31184787
5. Saragoni L, Morgagni P, Gardini A, Marfisi C, Vittimberga G, Garcea D, Scarpi E. Early gastric cancer: diagnosis, staging, and clinical impact. Evaluation of 530 patients. New elements for an updated definition and classification. *Gastric Cancer.* 2013; 16:549–54. <https://doi.org/10.1007/s10120-013-0233-2> PMID:23423491

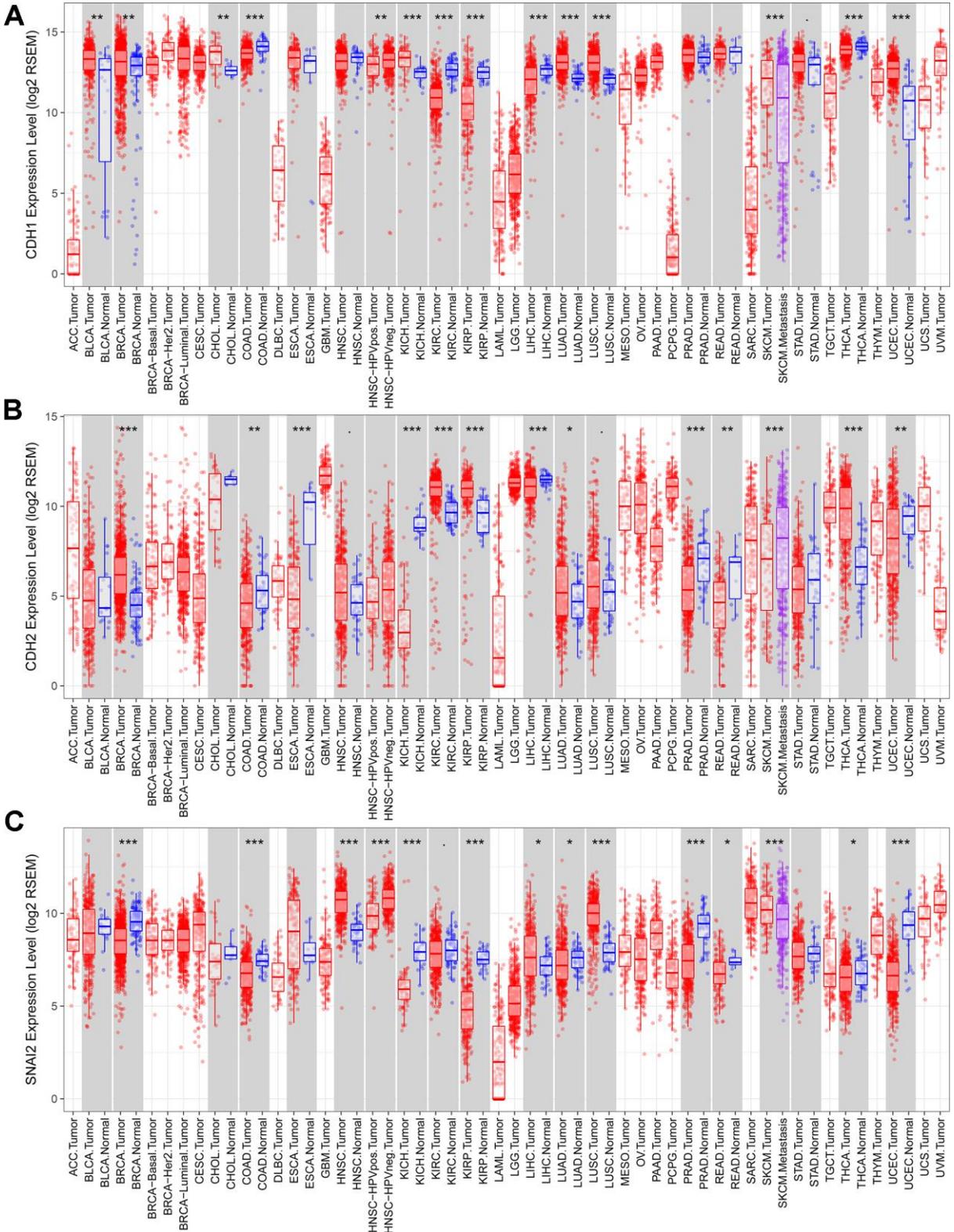
6. Procaccio L, Schirripa M, Fassan M, Vecchione L, Bergamo F, Prete AA, Intini R, Manai C, Dadduzio V, Boscolo A, Zagonel V, Lonardi S. Immunotherapy in gastrointestinal cancers. *Biomed Res Int.* 2017; 2017:4346576.  
<https://doi.org/10.1155/2017/4346576>  
PMID:[28758114](https://pubmed.ncbi.nlm.nih.gov/28758114/)
7. Barbee MS, Ogunniyi A, Horvat TZ, Dang TO. Current status and future directions of the immune checkpoint inhibitors ipilimumab, pembrolizumab, and nivolumab in oncology. *Ann Pharmacother.* 2015; 49:907–37.  
<https://doi.org/10.1177/1060028015586218>  
PMID:[25991832](https://pubmed.ncbi.nlm.nih.gov/25991832/)
8. Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L, Carcereny E, Ahn MJ, Felip E, et al, and KEYNOTE-001 Investigators. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med.* 2015; 372:2018–28.  
<https://doi.org/10.1056/NEJMoa1501824>  
PMID:[25891174](https://pubmed.ncbi.nlm.nih.gov/25891174/)
9. Rudd CE. A new perspective in cancer immunotherapy: PD-1 on myeloid cells takes center stage in orchestrating immune checkpoint blockade. *Sci Immunol.* 2020; 5:eaaz8128.  
<https://doi.org/10.1126/sciimmunol.aaz8128>  
PMID:[31901075](https://pubmed.ncbi.nlm.nih.gov/31901075/)
10. Grosser R, Cherkassky L, Chintala N, Adusumilli PS. Combination immunotherapy with CAR T cells and checkpoint blockade for the treatment of solid tumors. *Cancer Cell.* 2019; 36:471–82.  
<https://doi.org/10.1016/j.ccell.2019.09.006>  
PMID:[31715131](https://pubmed.ncbi.nlm.nih.gov/31715131/)
11. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol.* 2014; 15:178–96.  
<https://doi.org/10.1038/nrm3758>  
PMID:[24556840](https://pubmed.ncbi.nlm.nih.gov/24556840/)
12. Zhang Y, Weinberg RA. Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. *Front Med.* 2018; 12:361–73.  
<https://doi.org/10.1007/s11684-018-0656-6>  
PMID:[30043221](https://pubmed.ncbi.nlm.nih.gov/30043221/)
13. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal.* 2014; 7:re8.  
<https://doi.org/10.1126/scisignal.2005189>  
PMID:[25249658](https://pubmed.ncbi.nlm.nih.gov/25249658/)
14. Xu H, Xu WH, Ren F, Wang J, Wang HK, Cao DL, Shi GH, Qu YY, Zhang HL, Ye DW. Prognostic value of epithelial-mesenchymal transition markers in clear cell renal cell carcinoma. *Aging (Albany NY).* 2020; 12:866–83.  
<https://doi.org/10.18632/aging.102660> PMID:[31915310](https://pubmed.ncbi.nlm.nih.gov/31915310/)
15. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, Seoane JA, Farshidfar F, Bowlby R, Islam M, Kim J, Chatila W, Akbani R, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell.* 2018; 33:721–735.e8.  
<https://doi.org/10.1016/j.ccell.2018.03.010>  
PMID:[29622466](https://pubmed.ncbi.nlm.nih.gov/29622466/)
16. Azimi F, Scolyer RA, Rumcheva P, Moncrieff M, Murali R, McCarthy SW, Saw RP, Thompson JF. Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J Clin Oncol.* 2012; 30:2678–83.  
<https://doi.org/10.1200/JCO.2011.37.8539>  
PMID:[22711850](https://pubmed.ncbi.nlm.nih.gov/22711850/)
17. Ohtani H. Focus on TILs: prognostic significance of tumor infiltrating lymphocytes in human colorectal cancer. *Cancer Immun.* 2007; 7:4.  
PMID:[17311363](https://pubmed.ncbi.nlm.nih.gov/17311363/)
18. Muro K, Chung HC, Shankaran V, Geva R, Catenacci D, Gupta S, Eder JP, Golan T, Le DT, Burtness B, McRee AJ, Lin CC, Pathiraja K, et al. Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncol.* 2016; 17:717–26.  
[https://doi.org/10.1016/S1470-2045\(16\)00175-3](https://doi.org/10.1016/S1470-2045(16)00175-3)  
PMID:[27157491](https://pubmed.ncbi.nlm.nih.gov/27157491/)
19. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, Wong F, Azad NS, Rucki AA, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* 2017; 357:409–13.  
<https://doi.org/10.1126/science.aan6733>  
PMID:[28596308](https://pubmed.ncbi.nlm.nih.gov/28596308/)
20. Hummel S, Veltman K, Cichon C, Sonnenborn U, Schmidt MA. Differential targeting of the e-cadherin/ $\beta$ -catenin complex by gram-positive probiotic lactobacilli improves epithelial barrier function. *Appl Environ Microbiol.* 2012; 78:1140–47.  
<https://doi.org/10.1128/AEM.06983-11>  
PMID:[22179242](https://pubmed.ncbi.nlm.nih.gov/22179242/)
21. Bex G, van Roy F. Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol.* 2009; 1:a003129.  
<https://doi.org/10.1101/cshperspect.a003129>  
PMID:[20457567](https://pubmed.ncbi.nlm.nih.gov/20457567/)
22. Rosivatz E, Becker I, Specht K, Fricke E, Luber B, Busch R, Höfler H, Becker KF. Differential expression of the epithelial-mesenchymal transition regulators snail, SIP1, and twist in gastric cancer. *Am J Pathol.* 2002; 161:1881–91.

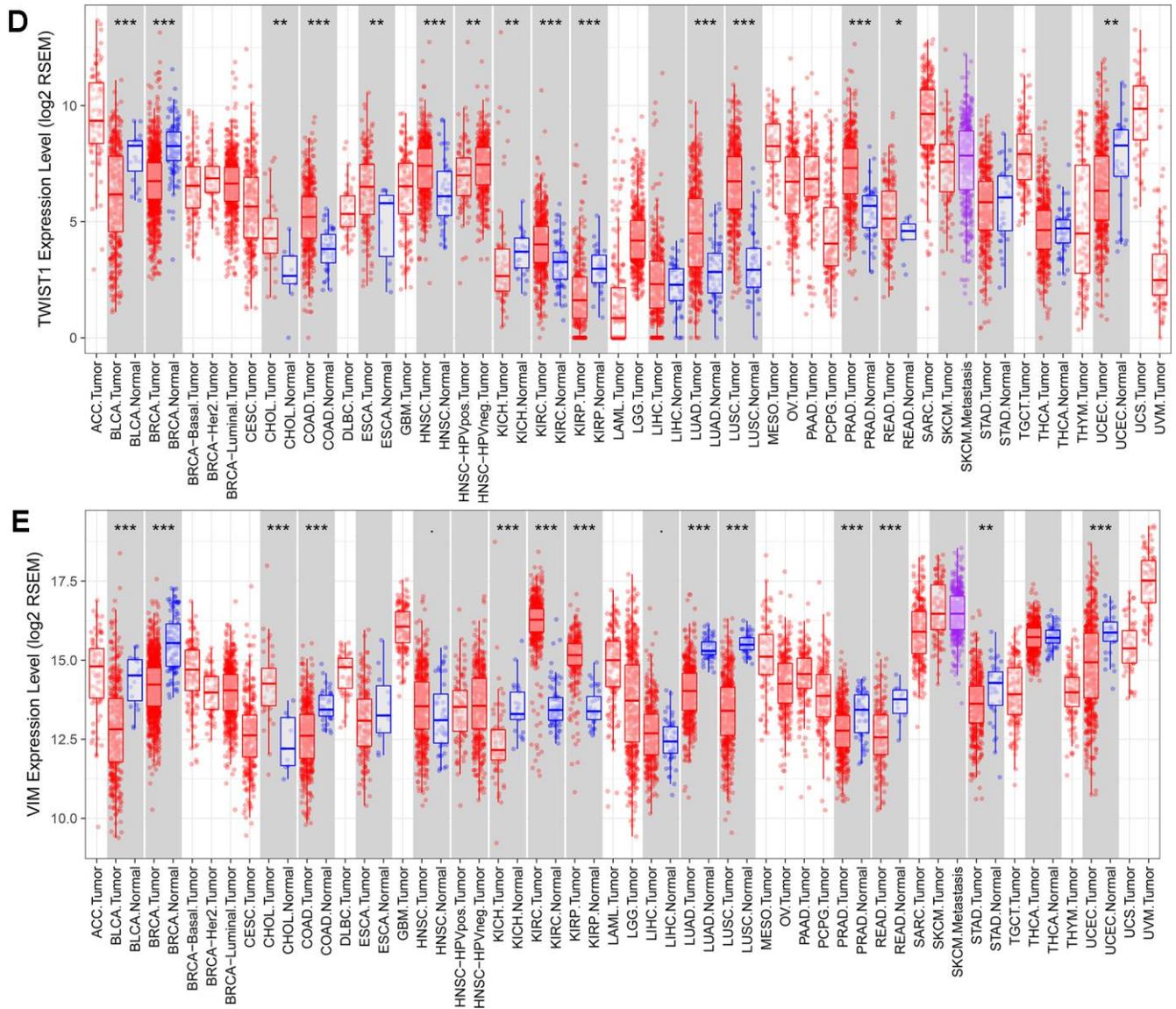
- [https://doi.org/10.1016/S0002-9440\(10\)64464-1](https://doi.org/10.1016/S0002-9440(10)64464-1)  
PMID:[12414534](https://pubmed.ncbi.nlm.nih.gov/12414534/)
23. Wang BJ, Zhang ZQ, Ke Y. Conversion of cadherin isoforms in cultured human gastric carcinoma cells. *World J Gastroenterol*. 2006; 12:966–70.  
<https://doi.org/10.3748/wjg.v12.i6.966>  
PMID:[16521229](https://pubmed.ncbi.nlm.nih.gov/16521229/)
24. Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest*. 2009; 119:1429–37.  
<https://doi.org/10.1172/JCI36183>  
PMID:[19487819](https://pubmed.ncbi.nlm.nih.gov/19487819/)
25. He H, Chen W, Wang X, Wang C, Liu F, Shen Z, Xu J, Gu J, Sun Y. Snail is an independent prognostic predictor for progression and patient survival of gastric cancer. *Cancer Sci*. 2012; 103:1296–303.  
<https://doi.org/10.1111/j.1349-7006.2012.02295.x>  
PMID:[22471696](https://pubmed.ncbi.nlm.nih.gov/22471696/)
26. Shin NR, Jeong EH, Choi CI, Moon HJ, Kwon CH, Chu IS, Kim GH, Jeon TY, Kim DH, Lee JH, Park DY. Overexpression of snail is associated with lymph node metastasis and poor prognosis in patients with gastric cancer. *BMC Cancer*. 2012; 12:521.  
<https://doi.org/10.1186/1471-2407-12-521>  
PMID:[23151184](https://pubmed.ncbi.nlm.nih.gov/23151184/)
27. Guo HM, Zhang XQ, Xu CH, Zou XP. Inhibition of invasion and metastasis of gastric cancer cells through snail targeting artificial microRNA interference. *Asian Pac J Cancer Prev*. 2011; 12:3433–38.  
PMID:[22471493](https://pubmed.ncbi.nlm.nih.gov/22471493/)
28. Chen Z, Liu M, Liu X, Huang S, Li L, Song B, Li H, Ren Q, Hu Z, Zhou Y, Qiao L. COX-2 regulates e-cadherin expression through the NF- $\kappa$ B/snail signaling pathway in gastric cancer. *Int J Mol Med*. 2013; 32:93–100.  
<https://doi.org/10.3892/ijmm.2013.1376>  
PMID:[23670240](https://pubmed.ncbi.nlm.nih.gov/23670240/)
29. Dongre A, Rashidian M, Reinhardt F, Bagnato A, Keckesova Z, Ploegh HL, Weinberg RA. Epithelial-to-mesenchymal transition contributes to immunosuppression in breast carcinomas. *Cancer Res*. 2017; 77:3982–89.  
<https://doi.org/10.1158/0008-5472.CAN-16-3292>  
PMID:[28428275](https://pubmed.ncbi.nlm.nih.gov/28428275/)
30. Liu J, Wu Z, Han D, Wei C, Liang Y, Jiang T, Chen L, Sha M, Cao Y, Huang F, Geng X, Yu J, Shen Y, et al. Mesencephalic Astrocyte-Derived Neurotrophic Factor Inhibits Liver Cancer Through Small Ubiquitin-Related Modifier (SUMO)ylation-Related Suppression of NF- $\kappa$ B/Snail Signaling Pathway and Epithelial-Mesenchymal Transition. *Hepatology*. 2020; 71:1262–1278.  
<https://doi.org/10.1002/hep.30917>  
PMID:[31469428](https://pubmed.ncbi.nlm.nih.gov/31469428/)
31. Greten FR, Eckmann L, Greten TF, Park JM, Li ZW, Egan LJ, Kagnoff MF, Karin M. IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell*. 2004; 118:285–96.  
<https://doi.org/10.1016/j.cell.2004.07.013>  
PMID:[15294155](https://pubmed.ncbi.nlm.nih.gov/15294155/)
32. Xiong H, Hong J, Du W, Lin YW, Ren LL, Wang YC, Su WY, Wang JL, Cui Y, Wang ZH, Fang JY. Roles of STAT3 and ZEB1 proteins in e-cadherin down-regulation and human colorectal cancer epithelial-mesenchymal transition. *J Biol Chem*. 2012; 287:5819–32.  
<https://doi.org/10.1074/jbc.M111.295964>  
PMID:[22205702](https://pubmed.ncbi.nlm.nih.gov/22205702/)
33. Bates RC, Mercurio AM. Tumor necrosis factor-alpha stimulates the epithelial-to-mesenchymal transition of human colonic organoids. *Mol Biol Cell*. 2003; 14:1790–800.  
<https://doi.org/10.1091/mbc.e02-09-0583>  
PMID:[12802055](https://pubmed.ncbi.nlm.nih.gov/12802055/)
34. Hsieh CH, Tai SK, Yang MH. Snail-overexpressing cancer cells promote M2-like polarization of tumor-associated macrophages by delivering MiR-21-abundant exosomes. *Neoplasia*. 2018; 20:775–88.  
<https://doi.org/10.1016/j.neo.2018.06.004>  
PMID:[29981499](https://pubmed.ncbi.nlm.nih.gov/29981499/)
35. Faget J, Groeneveld S, Boivin G, Sankar M, Zangger N, Garcia M, Guex N, Zlobec I, Steiner L, Piersigilli A, Xenarios I, Meylan E. Neutrophils and snail orchestrate the establishment of a pro-tumor microenvironment in lung cancer. *Cell Rep*. 2017; 21:3190–204.  
<https://doi.org/10.1016/j.celrep.2017.11.052>  
PMID:[29241546](https://pubmed.ncbi.nlm.nih.gov/29241546/)
36. Noman MZ, Van Moer K, Marani V, Gemmill RM, Tranchevent LC, Azuaje F, Muller A, Chouaib S, Thiery JP, Berchem G, Janji B. CD47 is a direct target of SNAI1 and ZEB1 and its blockade activates the phagocytosis of breast cancer cells undergoing EMT. *Oncoimmunology*. 2018; 7:e1345415.  
<https://doi.org/10.1080/2162402X.2017.1345415>  
PMID:[29632713](https://pubmed.ncbi.nlm.nih.gov/29632713/)
37. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
PMID:[29092952](https://pubmed.ncbi.nlm.nih.gov/29092952/)
38. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincaid-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007; 9:166–80.  
<https://doi.org/10.1593/neo.07112> PMID:[17356713](https://pubmed.ncbi.nlm.nih.gov/17356713/)

39. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45:W98–102.  
<https://doi.org/10.1093/nar/gkx247>  
PMID:[28407145](https://pubmed.ncbi.nlm.nih.gov/28407145/)
40. Szász AM, Lánckzy A, Nagy Á, Förster S, Hark K, Green JE, Boussioutas A, Busuttill R, Szabó A, Gyórfy B. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget.* 2016; 7:49322–33.  
<https://doi.org/10.18632/oncotarget.10337>  
PMID:[27384994](https://pubmed.ncbi.nlm.nih.gov/27384994/)
41. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SS, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics.* 2019; 35:4200–02.  
<https://doi.org/10.1093/bioinformatics/btz210>  
PMID:[30903160](https://pubmed.ncbi.nlm.nih.gov/30903160/)

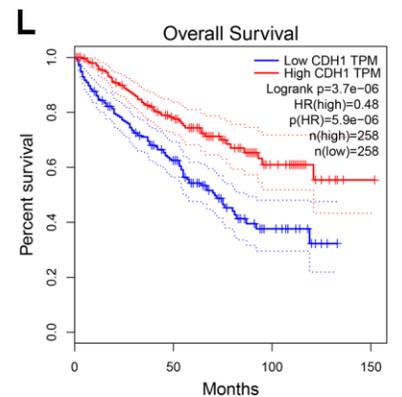
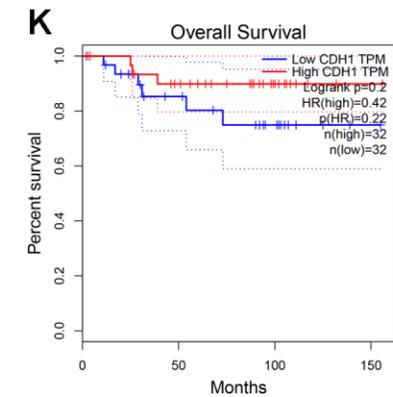
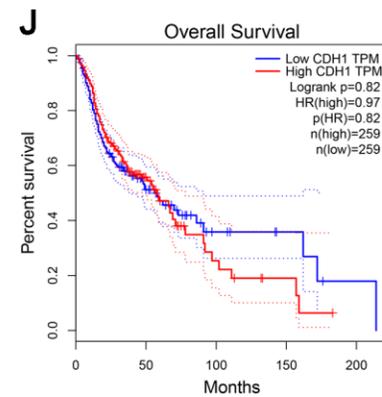
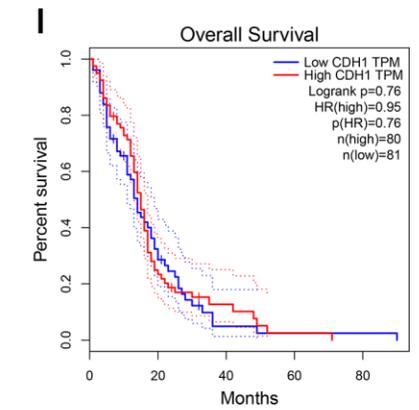
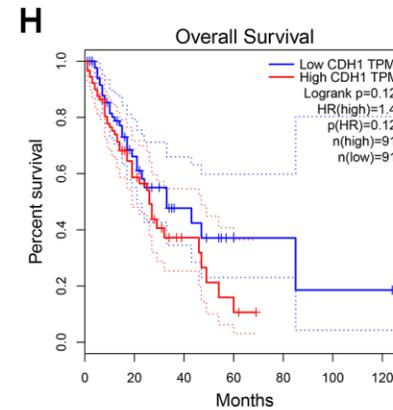
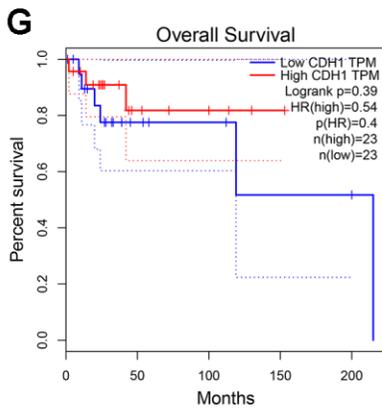
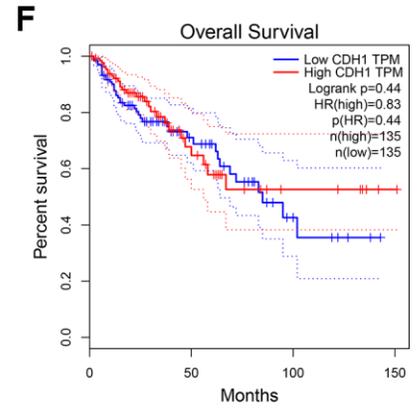
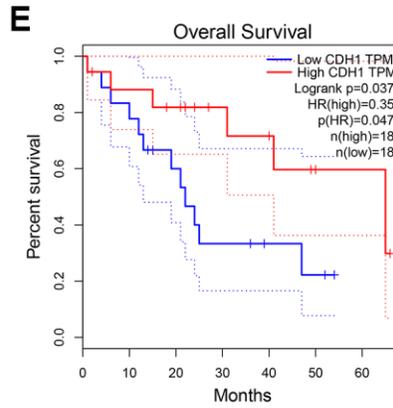
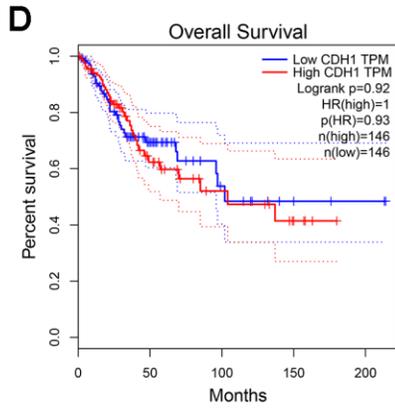
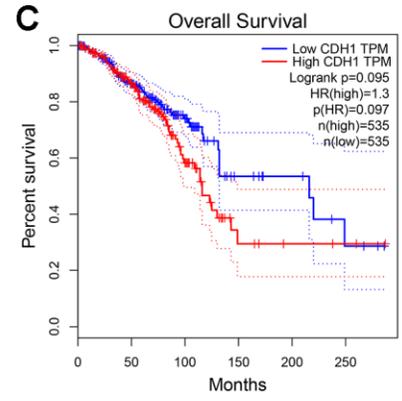
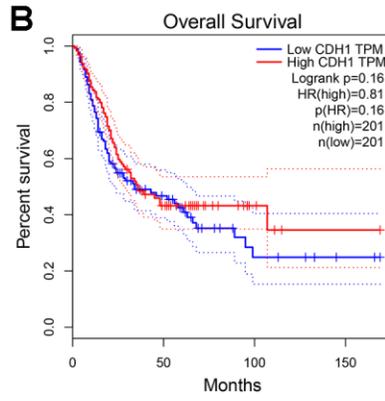
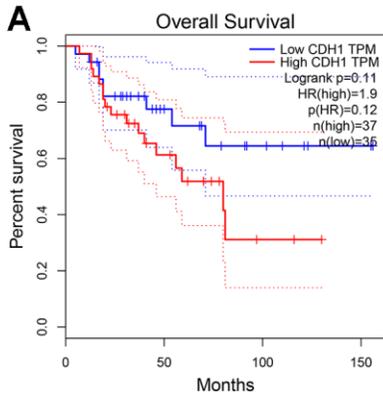
SUPPLEMENTARY MATERIALS

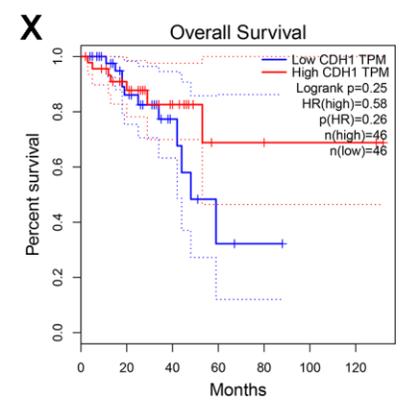
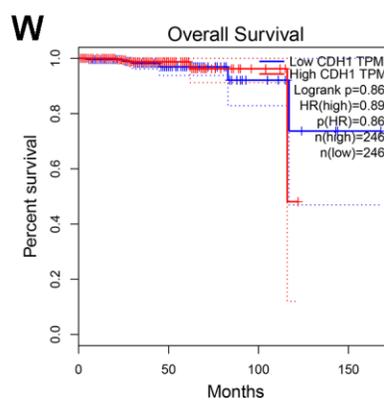
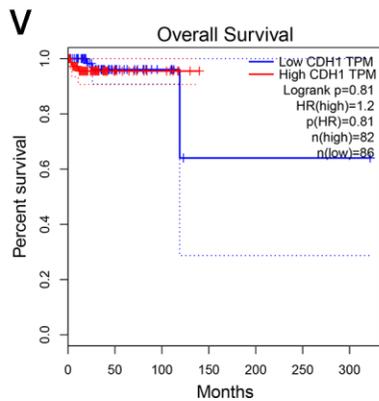
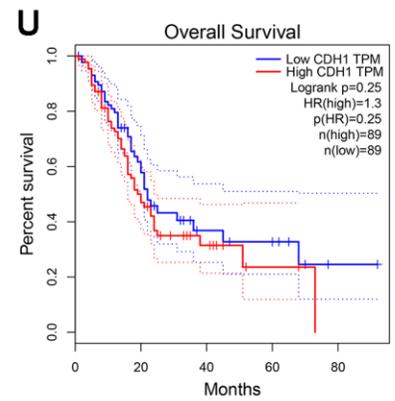
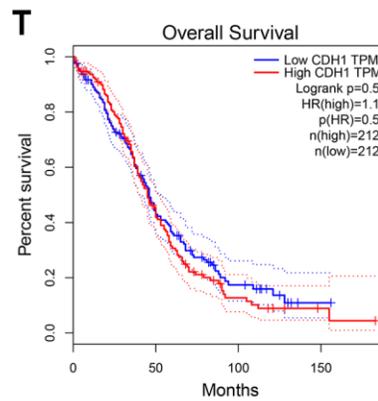
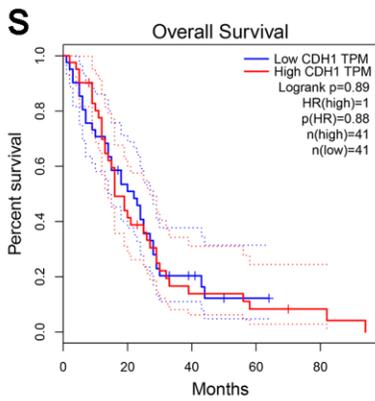
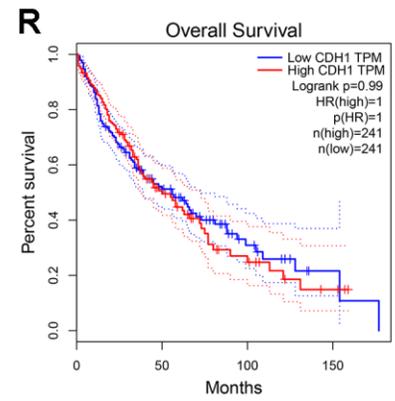
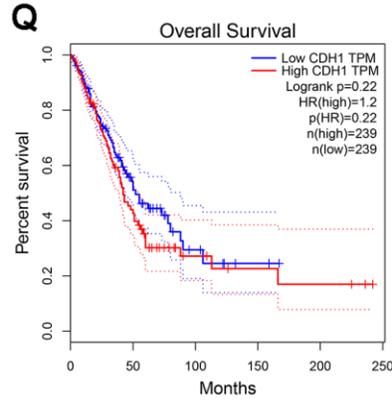
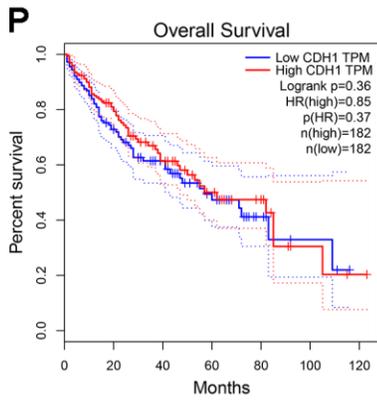
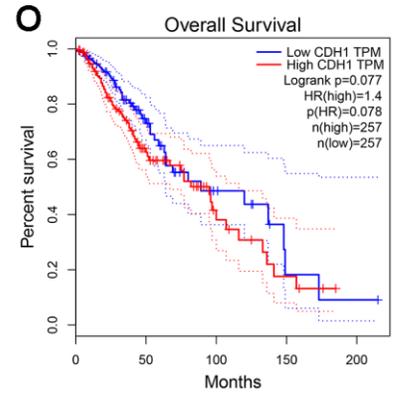
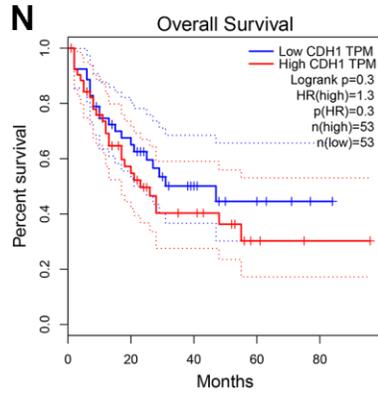
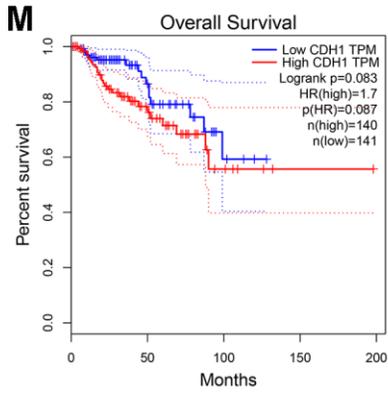
Supplementary Figures

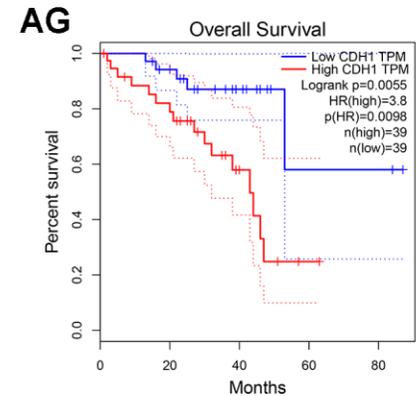
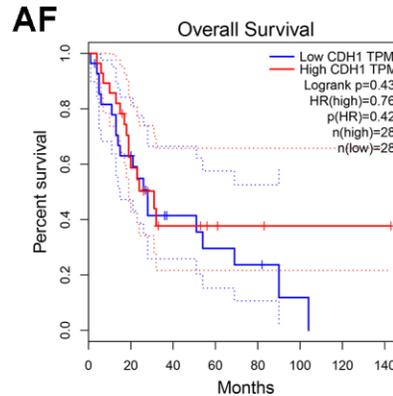
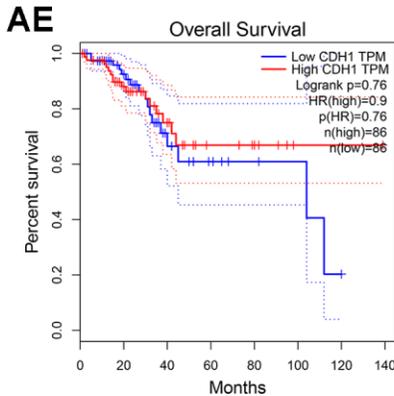
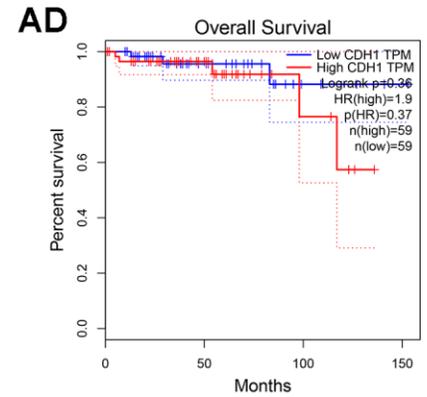
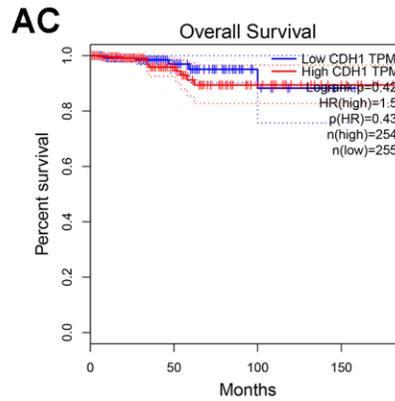
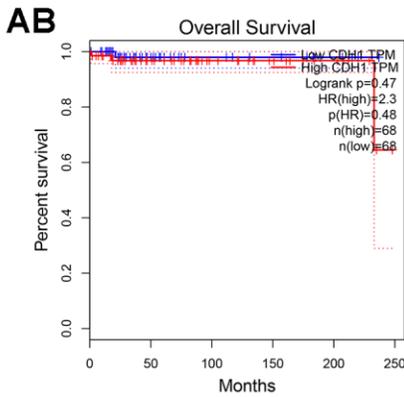
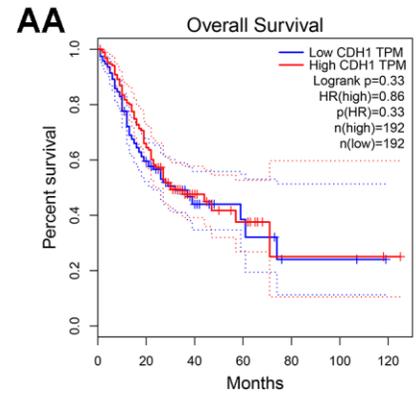
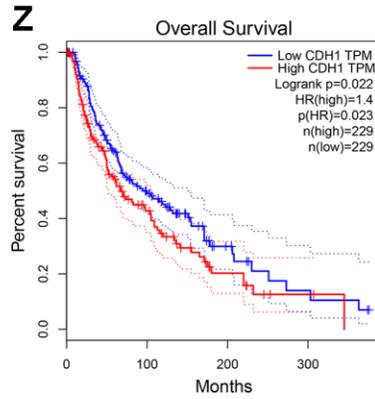
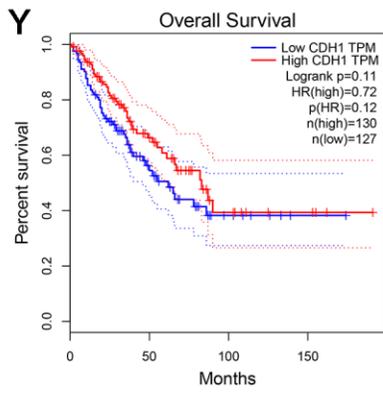




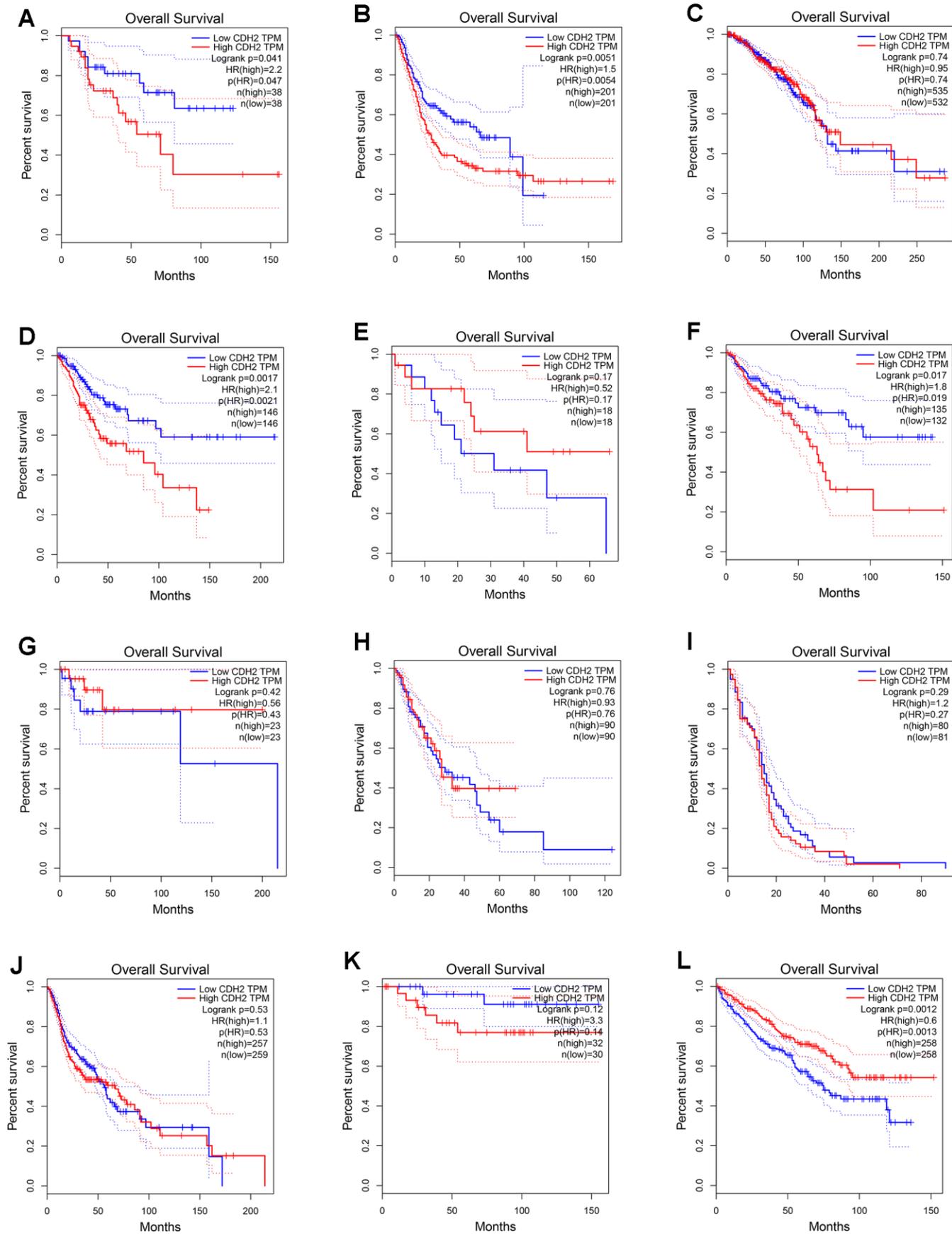
Supplementary Figure 1. The five EMT markers expression levels in different types of human cancers in TCGA dataset.

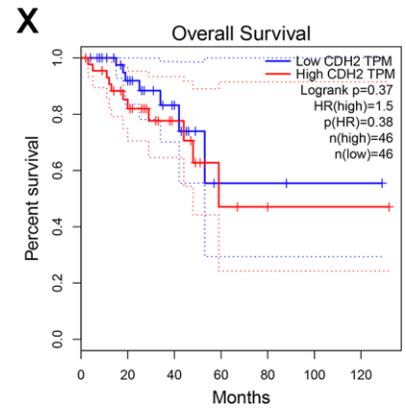
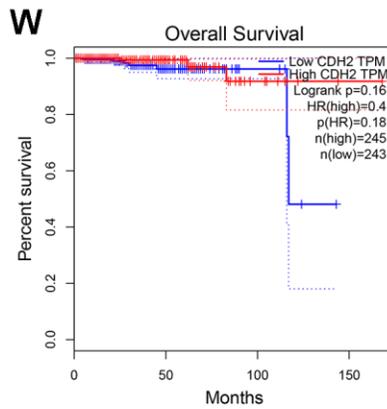
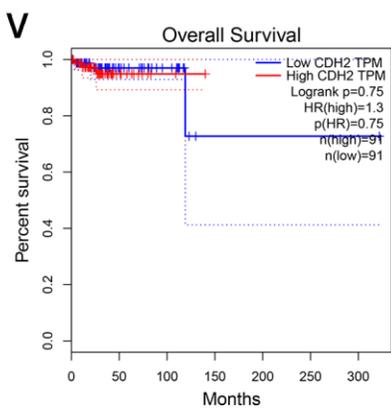
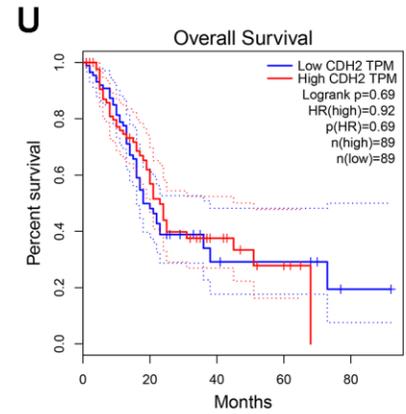
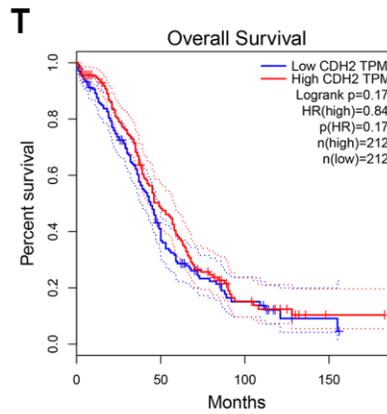
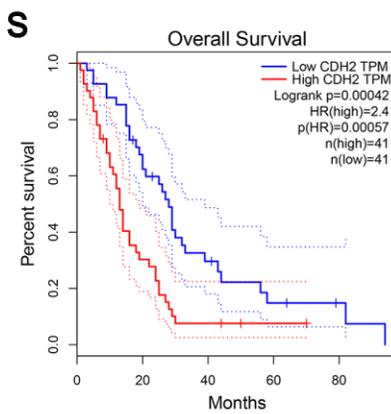
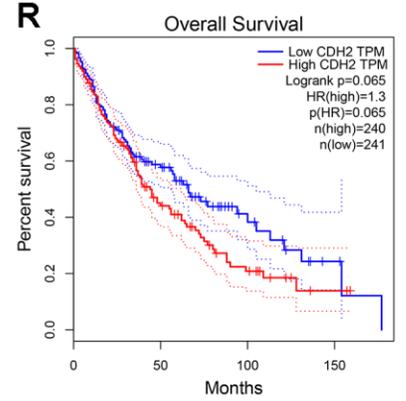
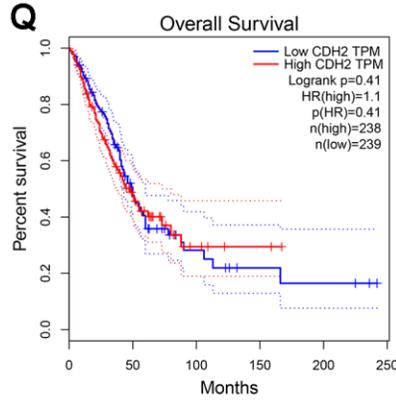
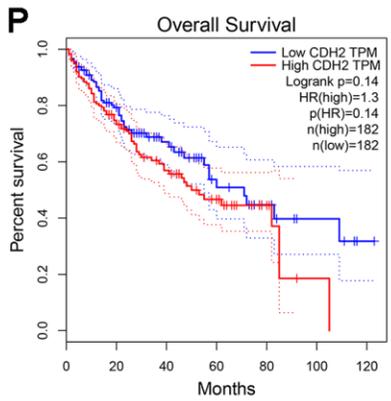
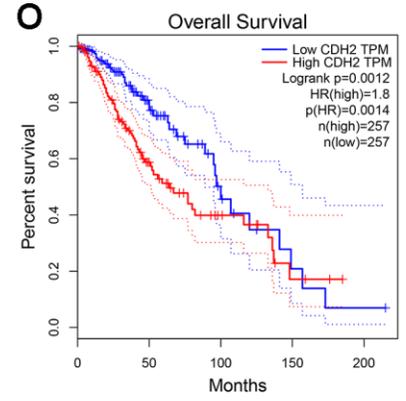
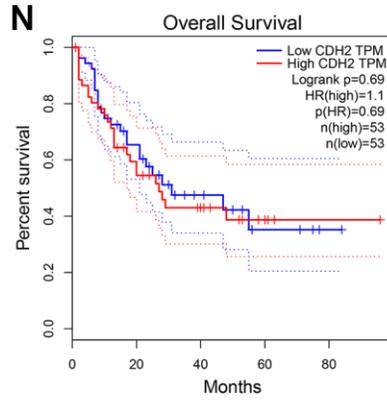
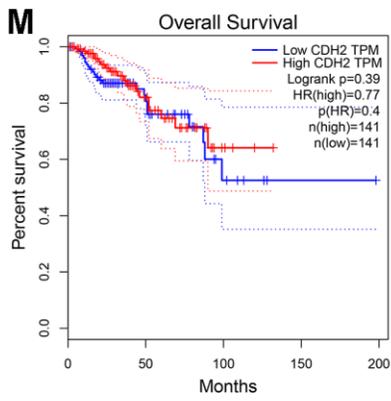


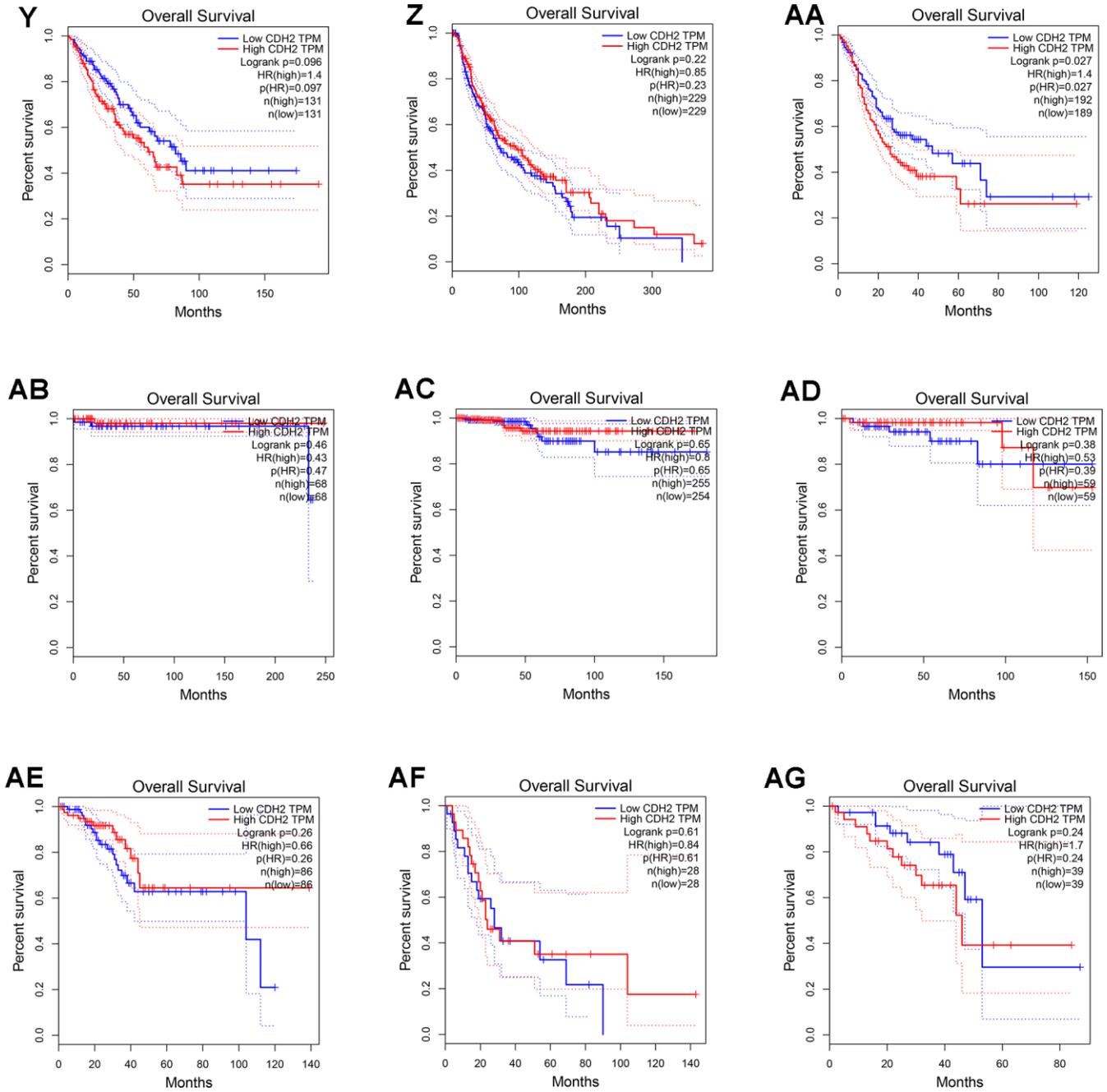




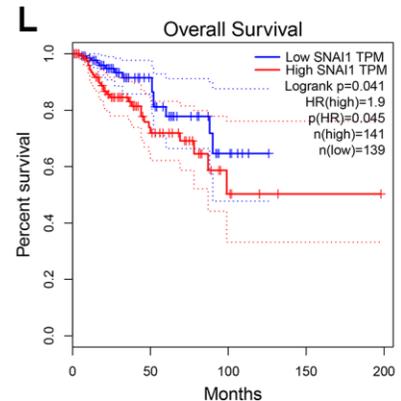
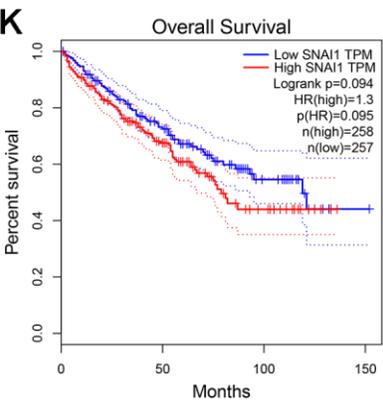
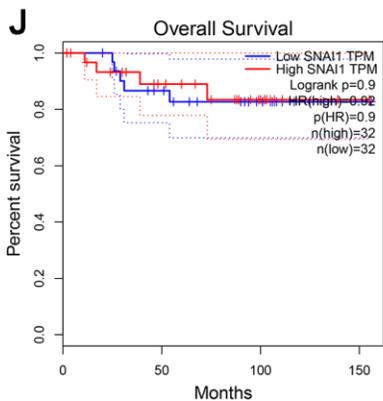
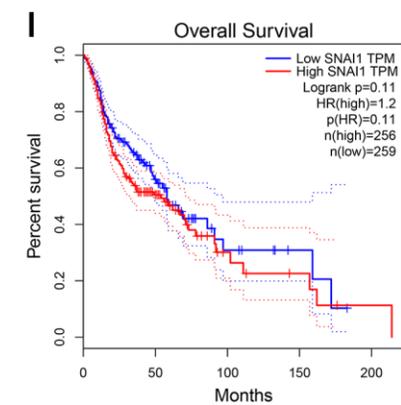
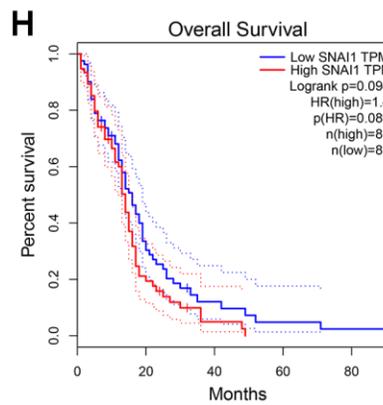
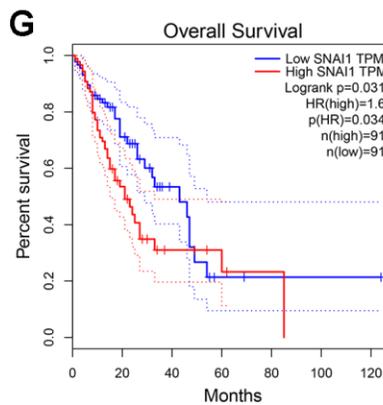
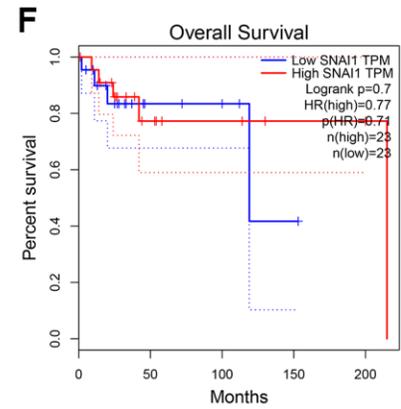
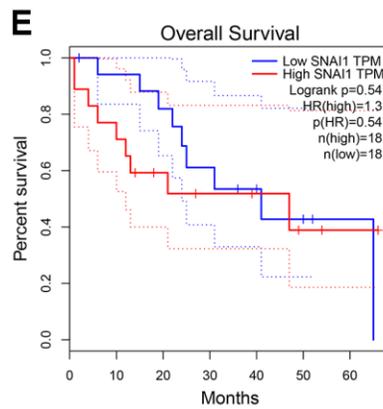
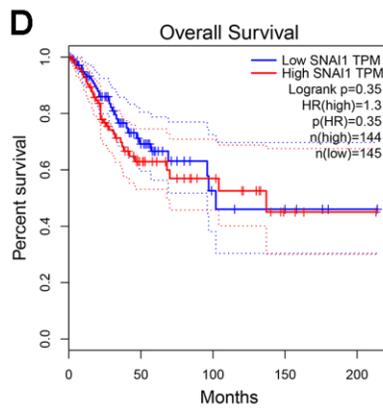
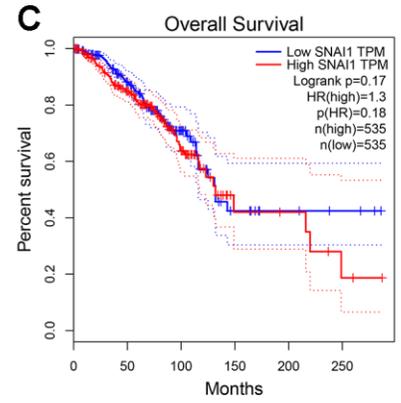
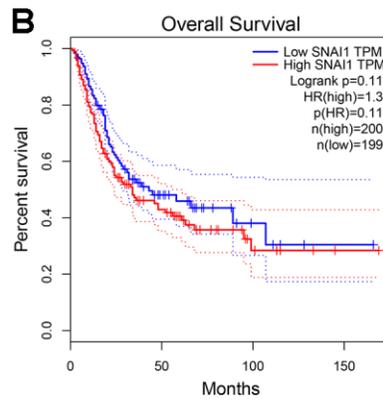
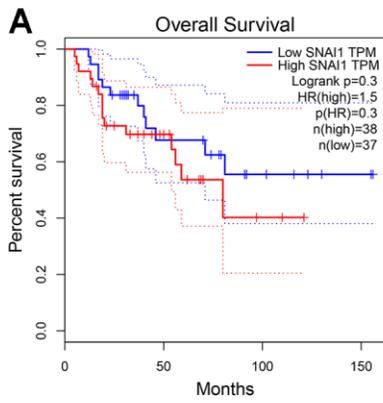
Supplementary Figure 2. Kaplan-Meier survival curves comparing the high and low expression of CDH1 in different types of human cancers in TCGA dataset.

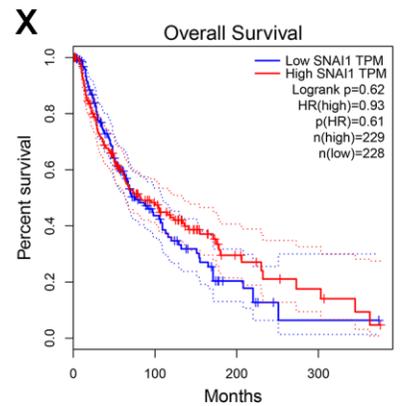
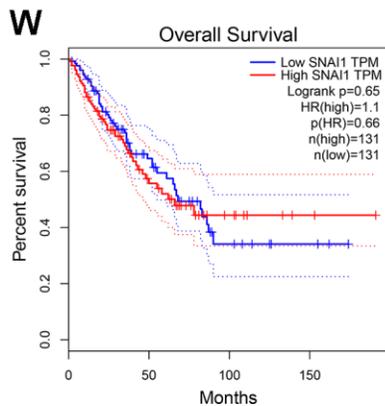
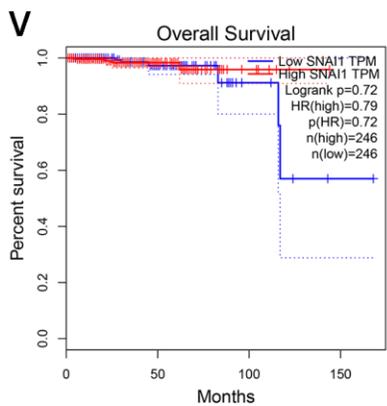
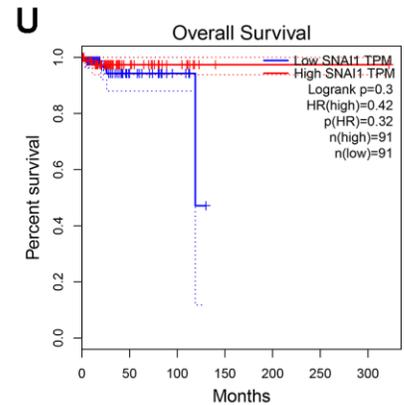
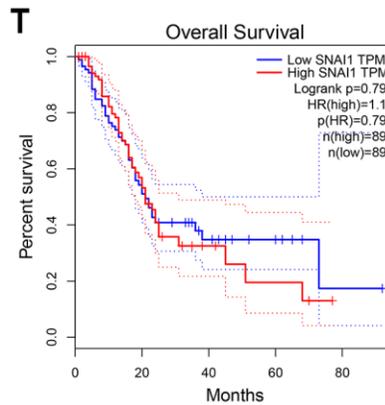
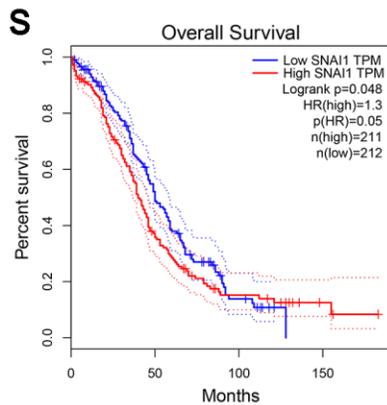
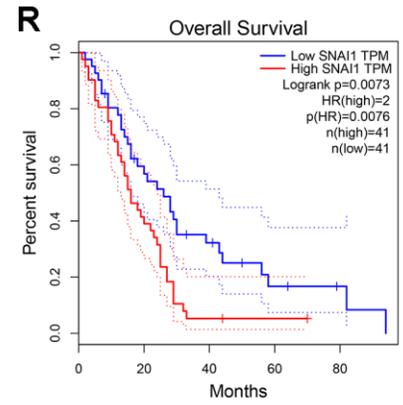
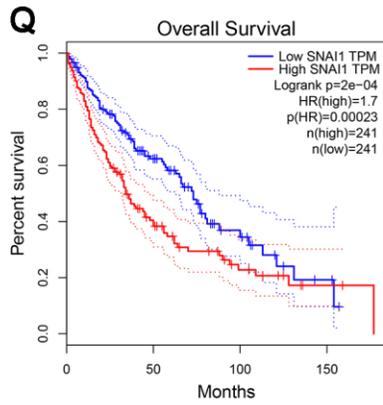
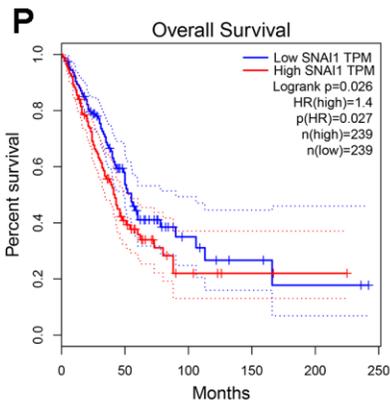
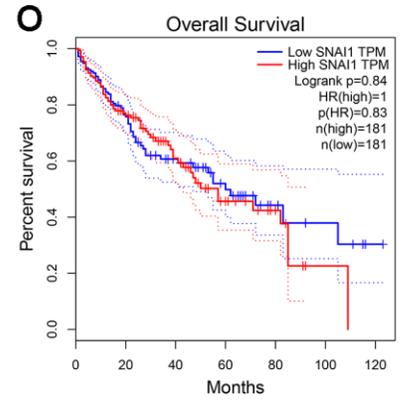
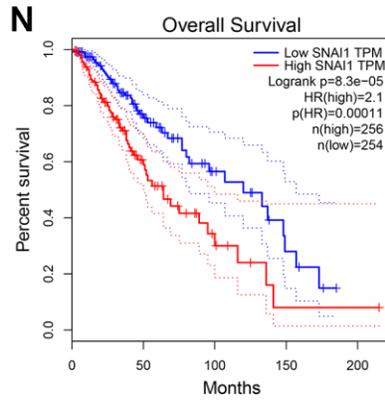
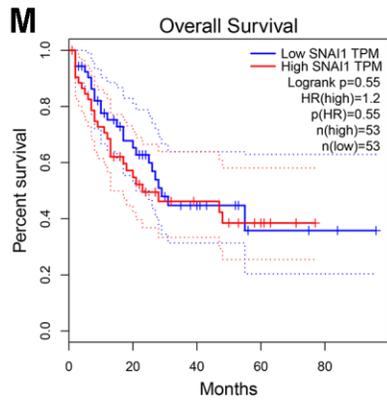


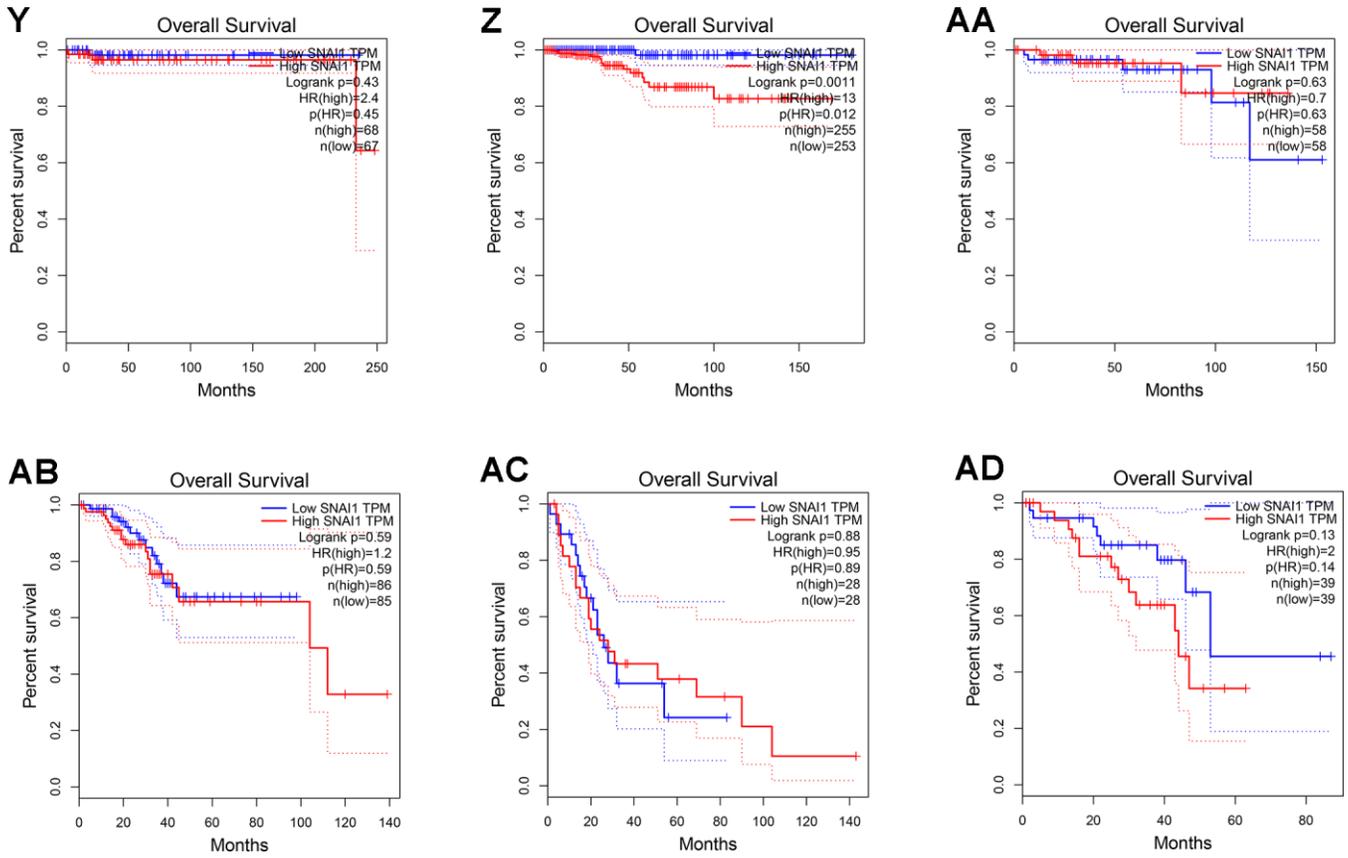




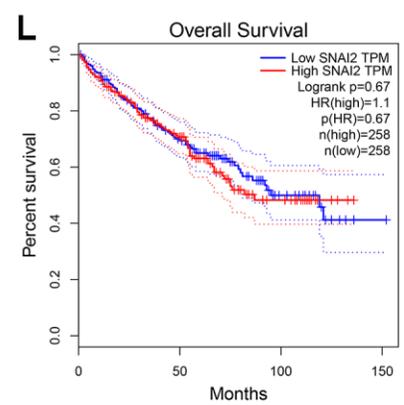
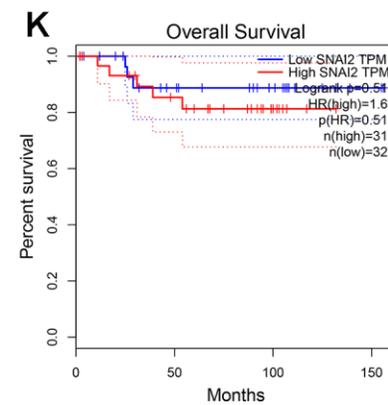
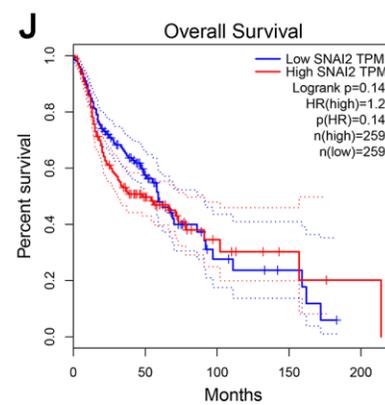
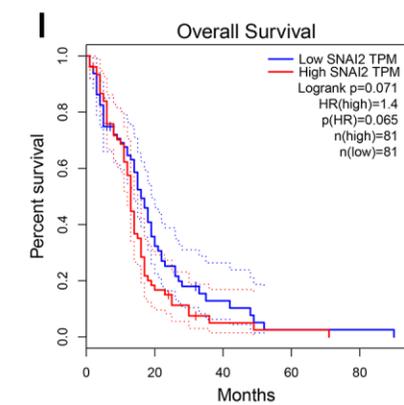
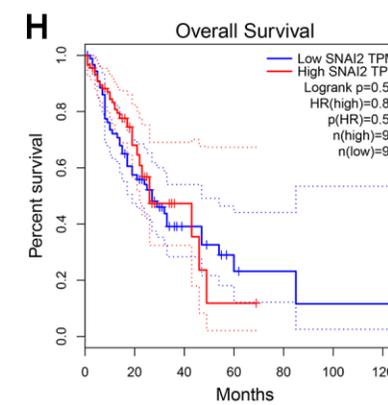
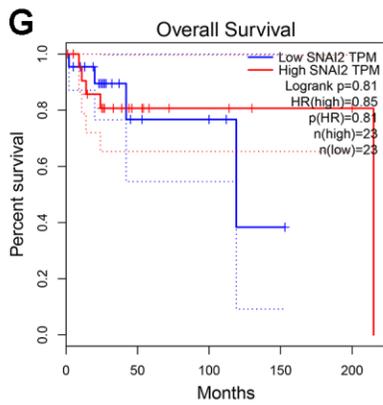
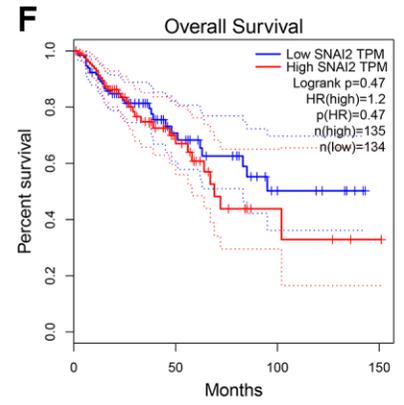
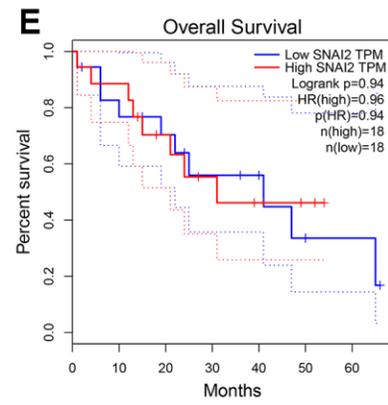
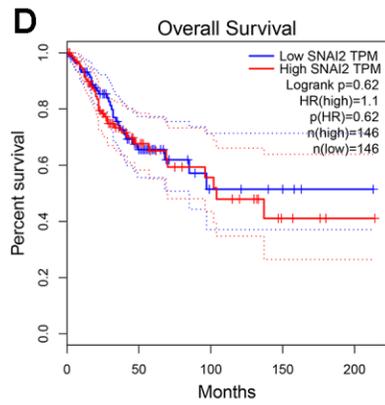
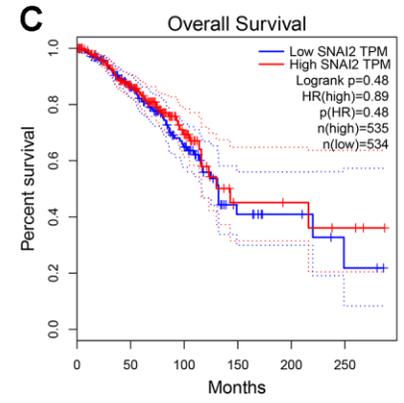
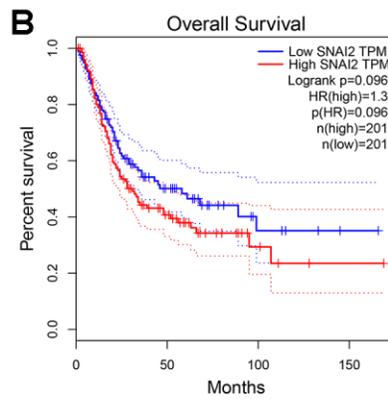
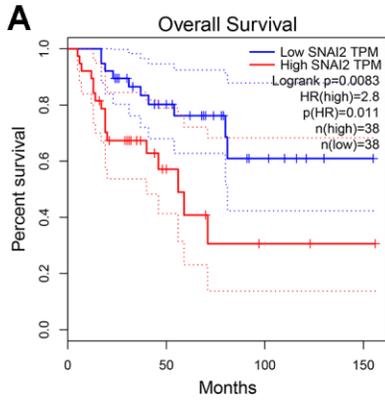
Supplementary Figure 3. Kaplan-Meier survival curves comparing the high and low expression of CDH2 in different types of human cancers in TCGA dataset.

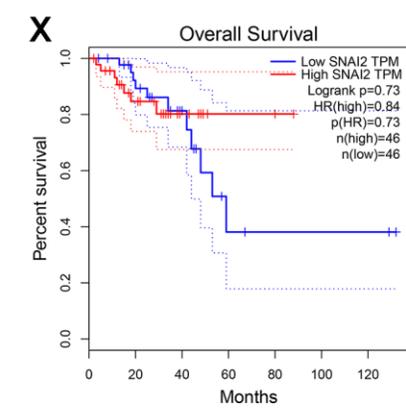
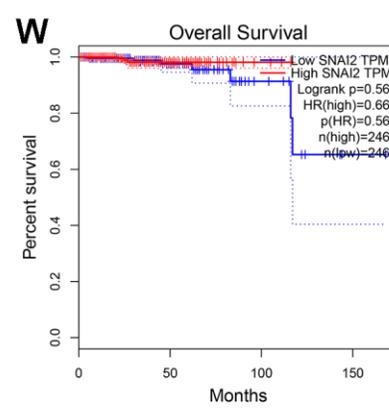
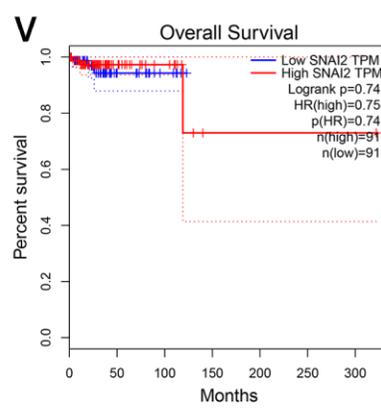
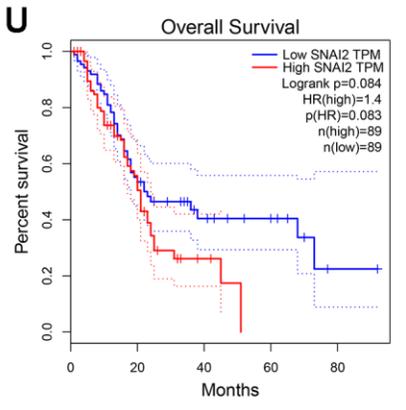
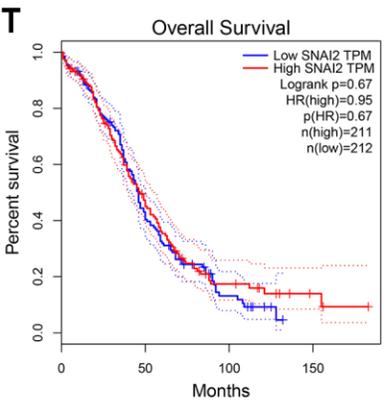
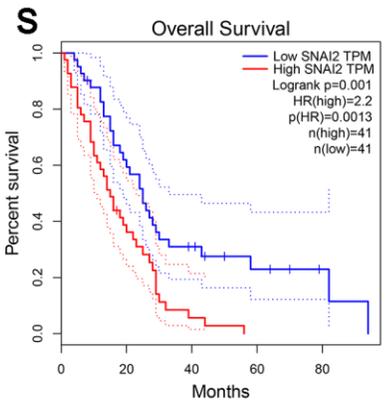
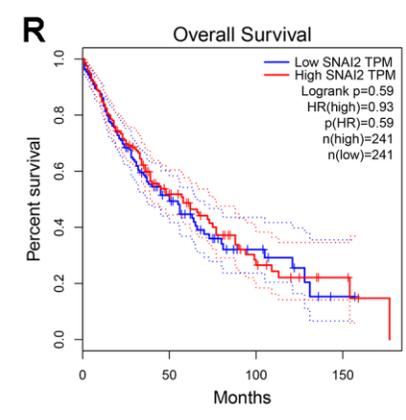
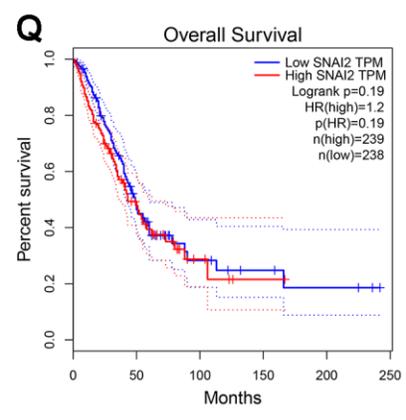
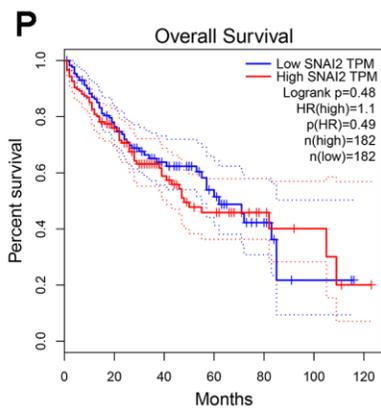
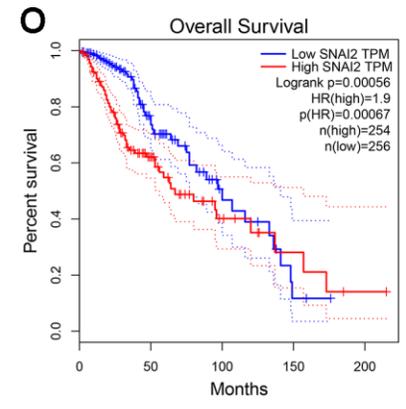
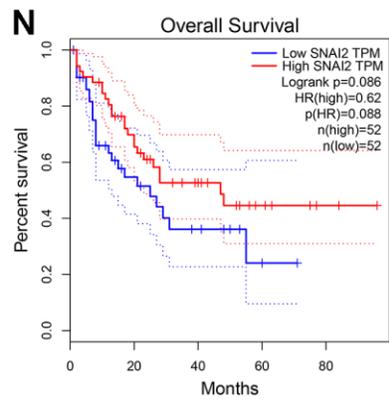
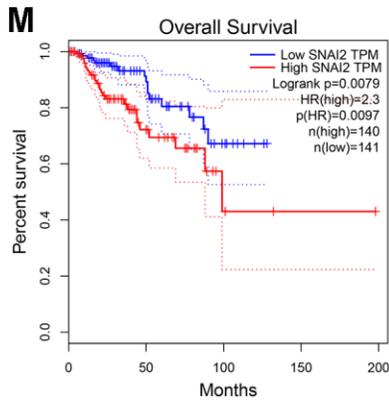


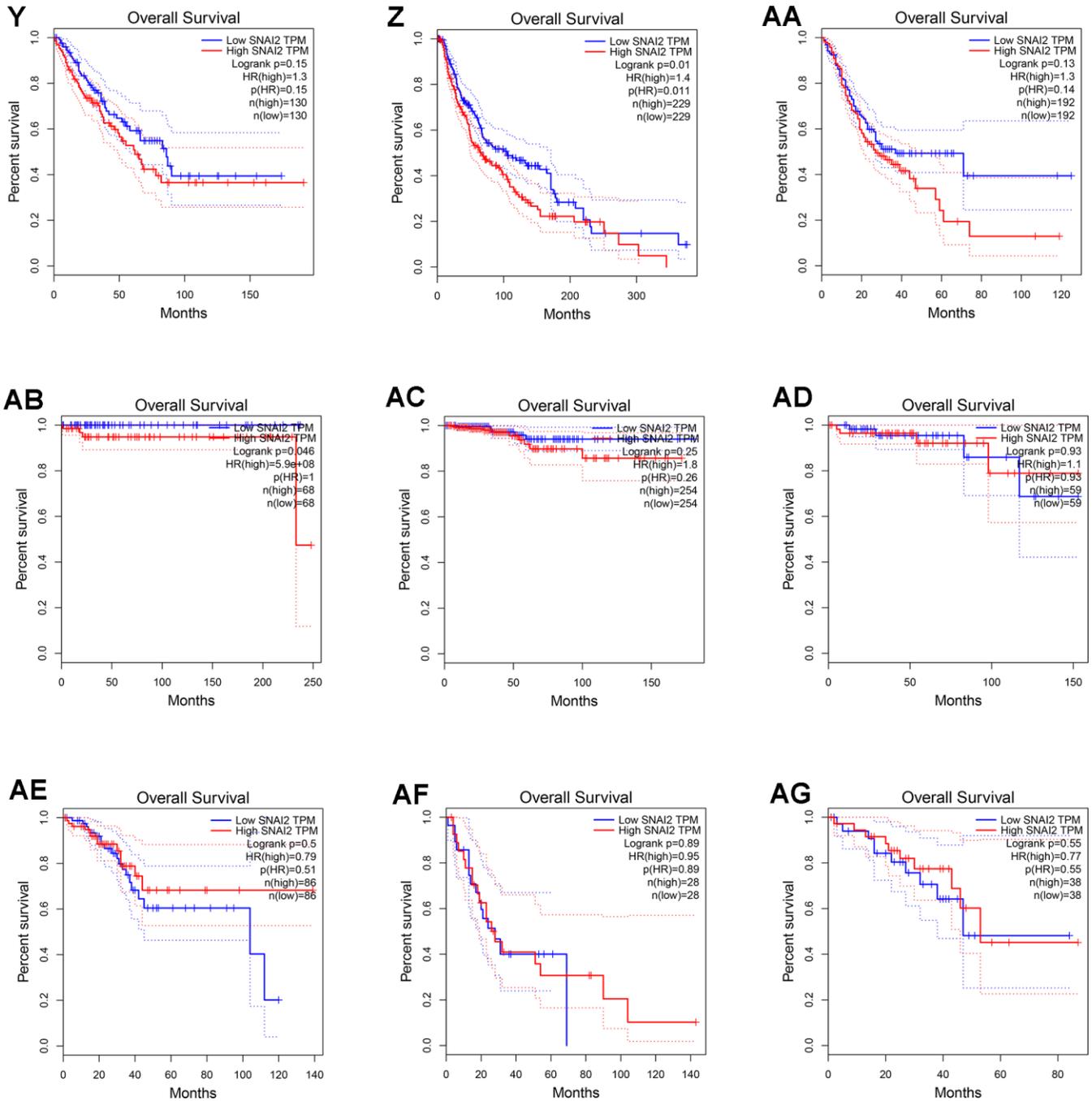




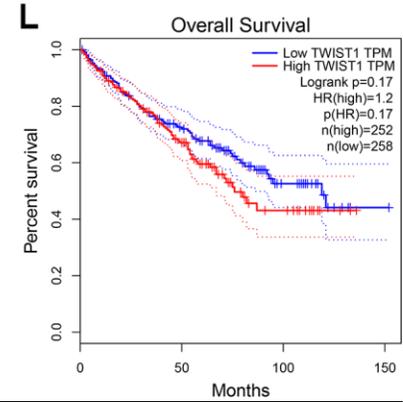
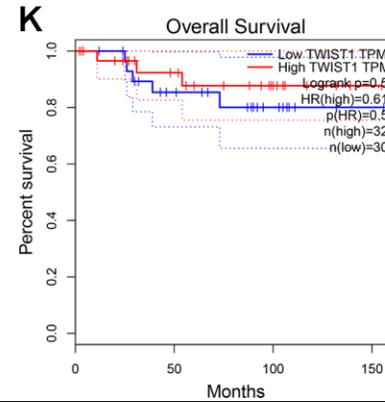
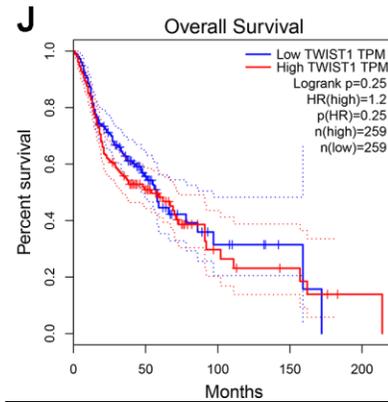
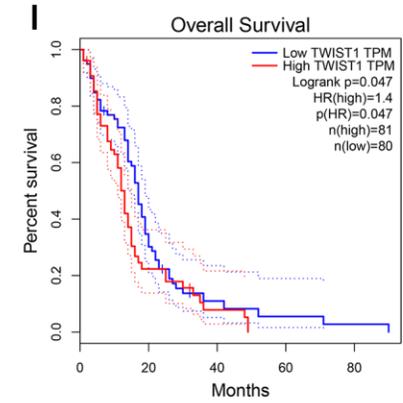
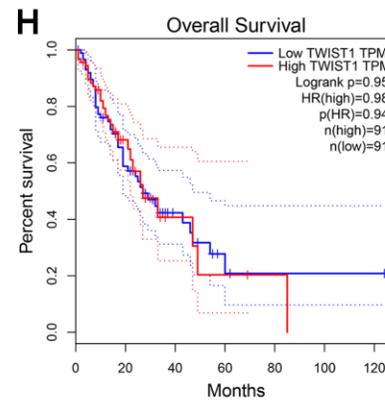
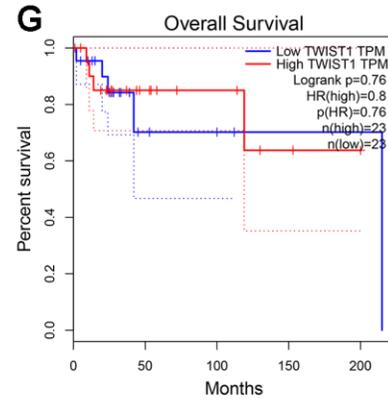
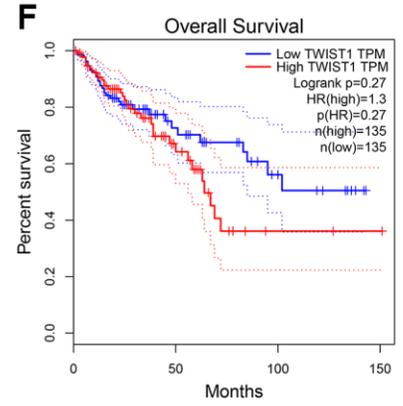
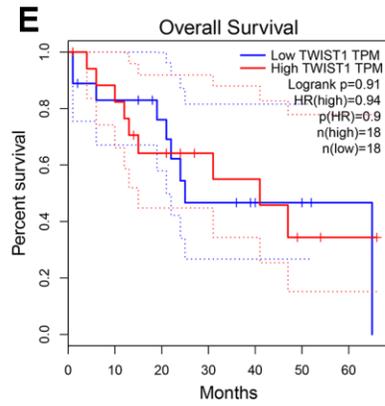
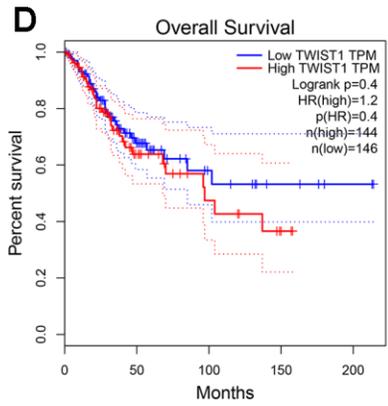
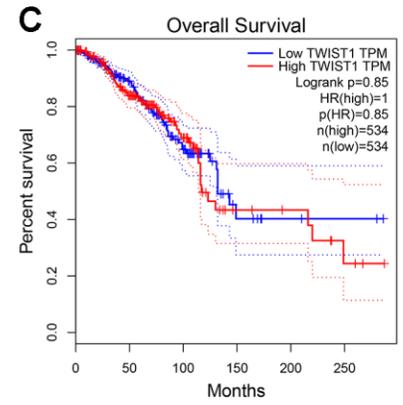
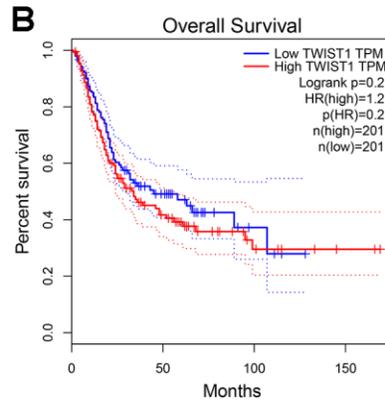
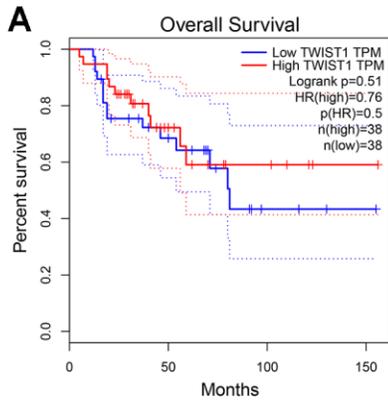
Supplementary Figure 4. Kaplan-Meier survival curves comparing the high and low expression of SNAI1 in different types of human cancers in TCGA dataset.

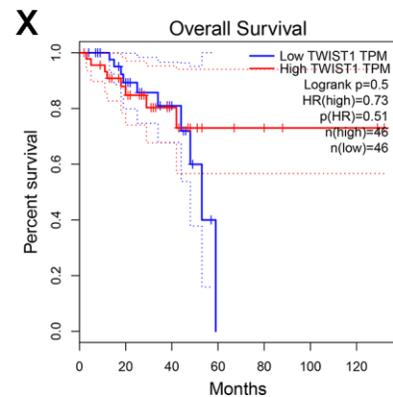
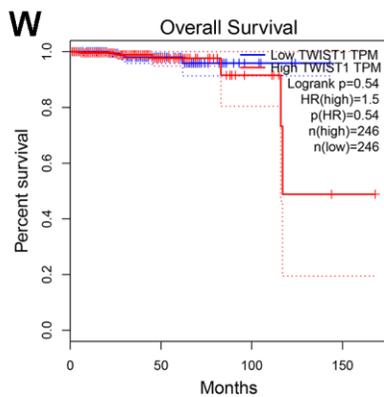
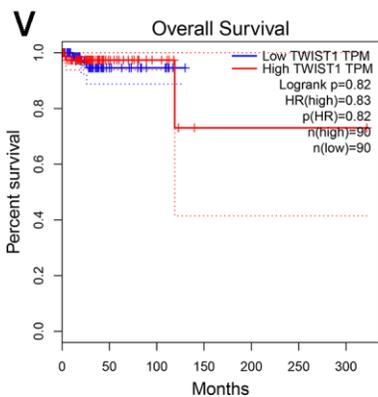
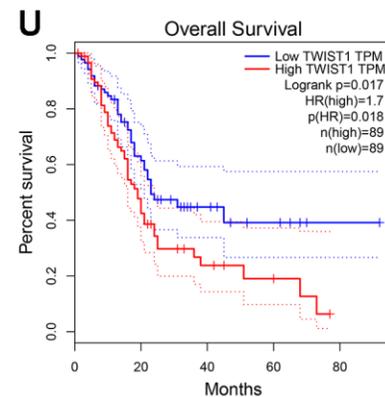
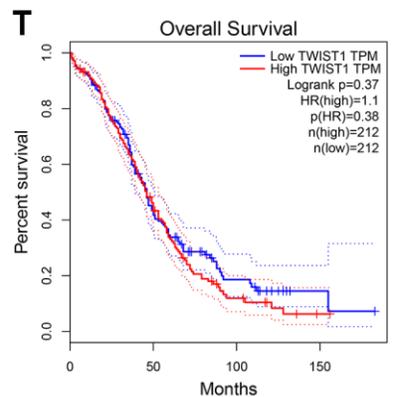
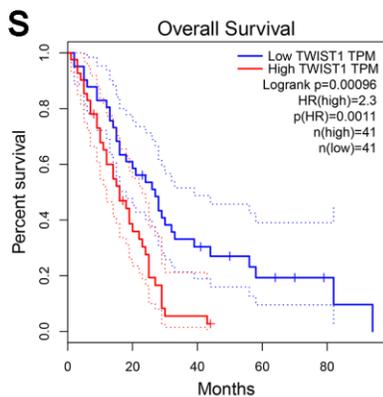
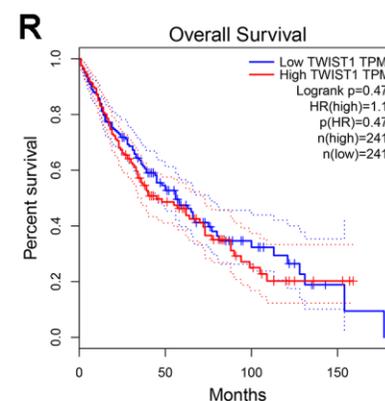
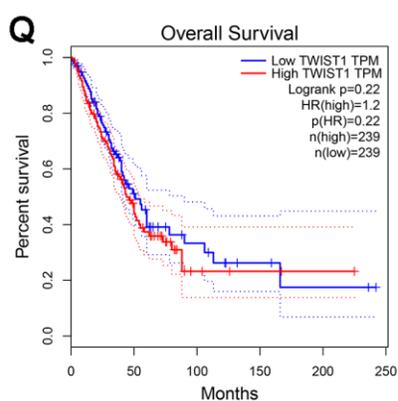
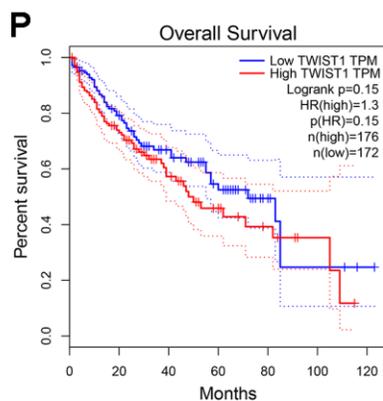
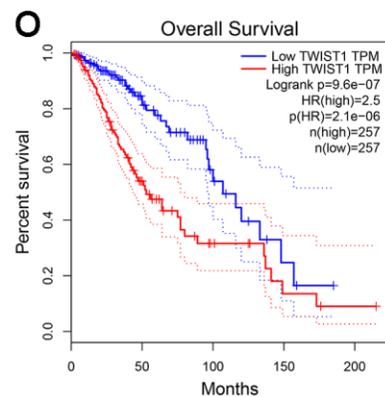
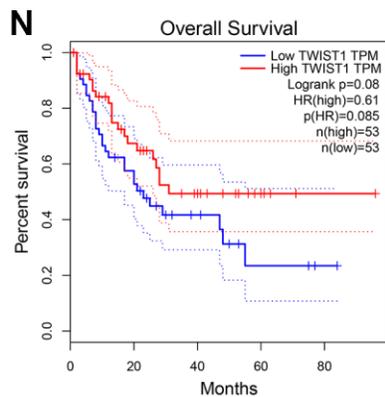
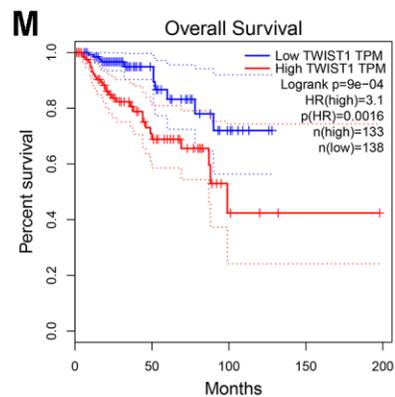


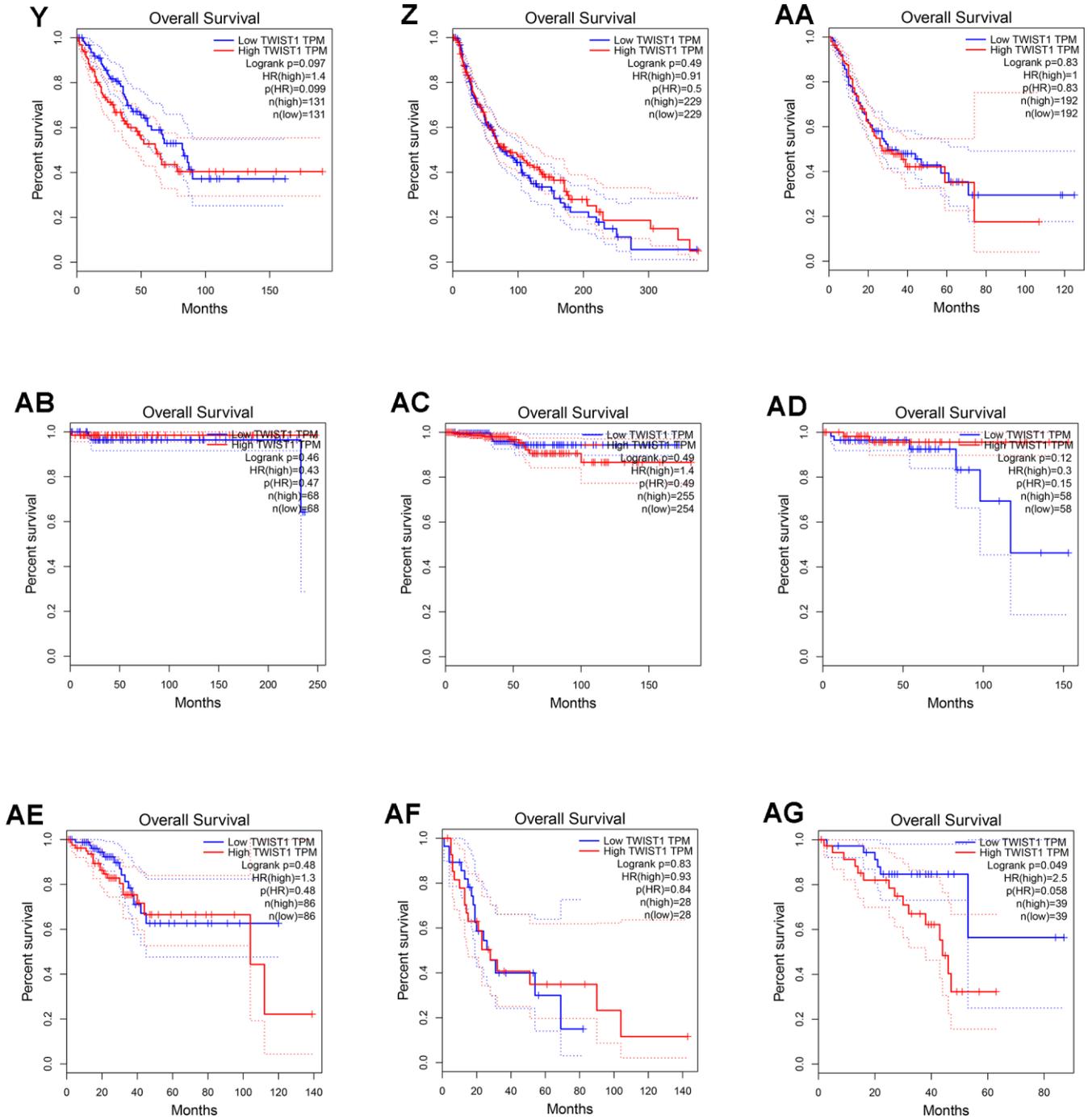




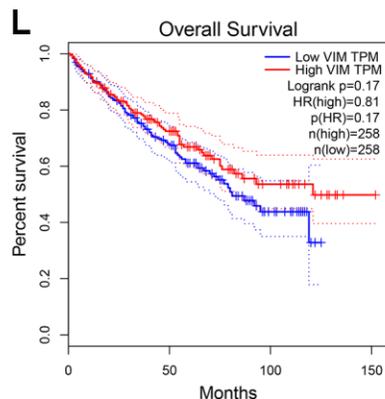
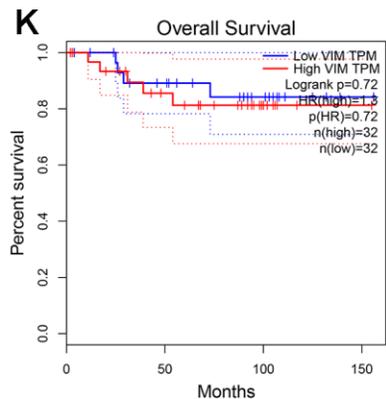
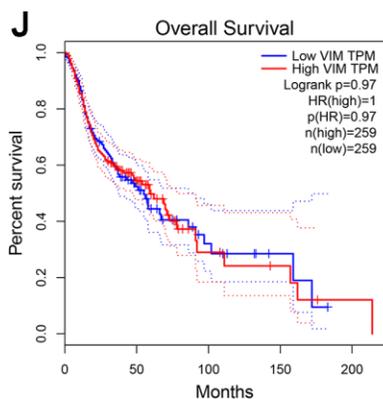
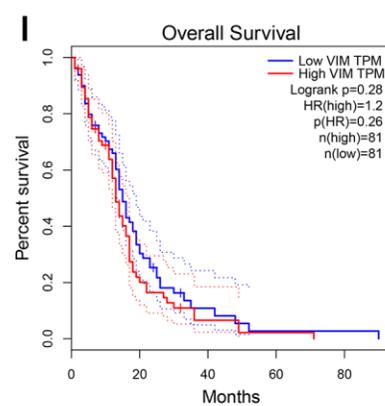
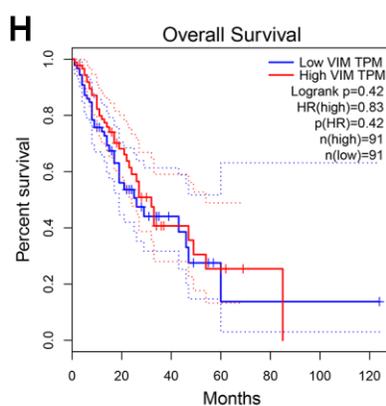
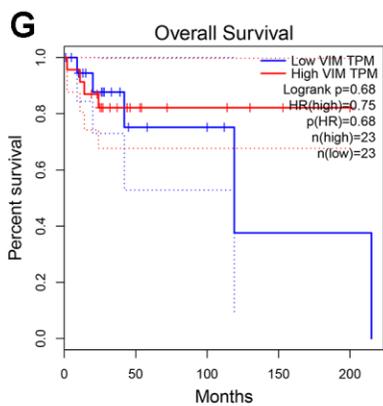
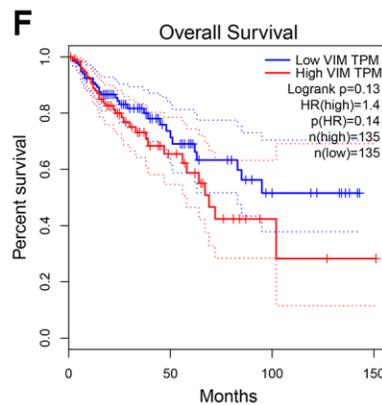
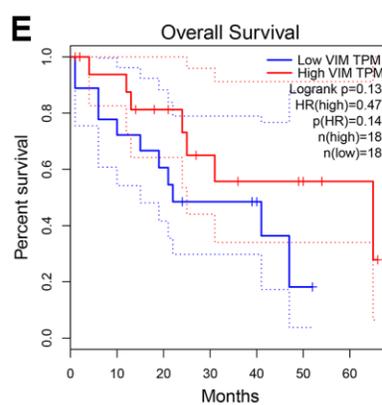
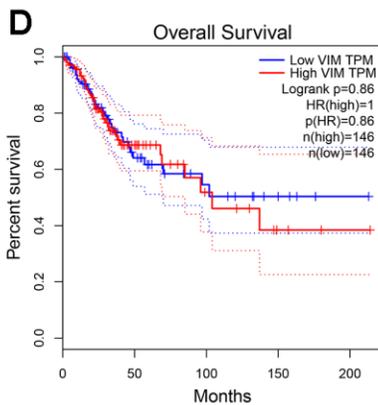
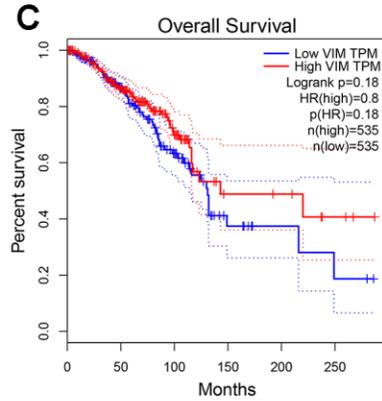
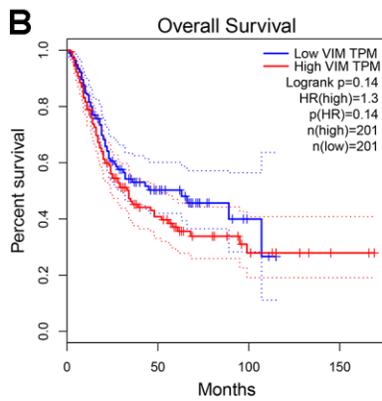
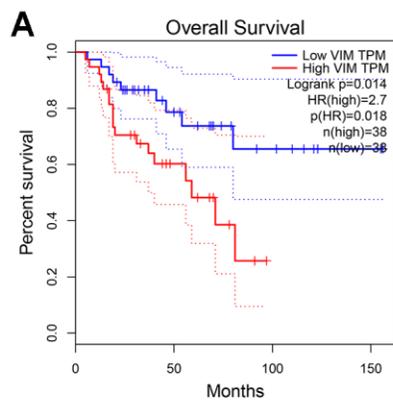
Supplementary Figure 5. Kaplan-Meier survival curves comparing the high and low expression of SNAI2 in different types of human cancers in TCGA dataset.

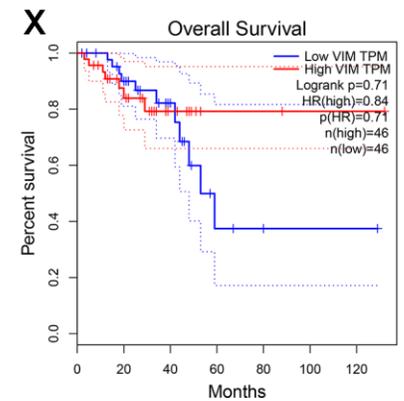
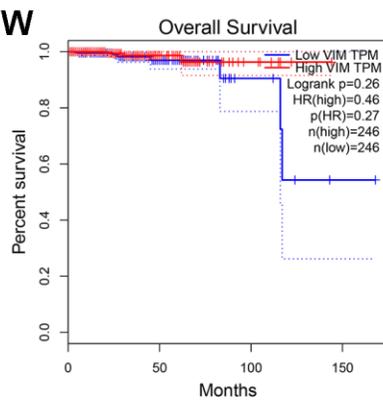
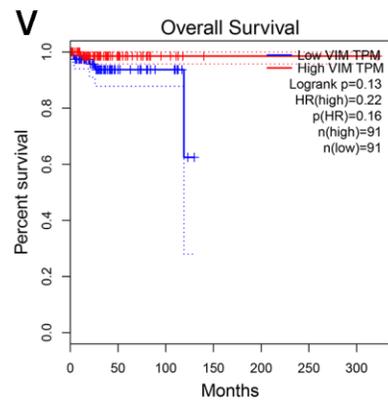
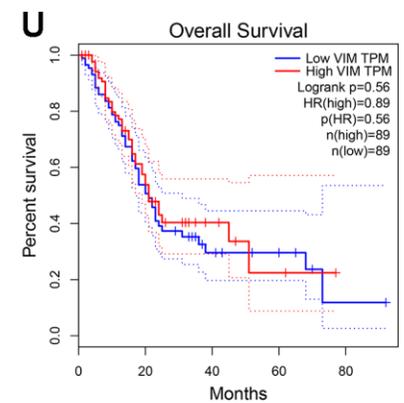
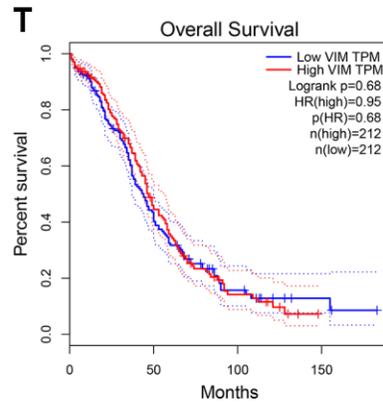
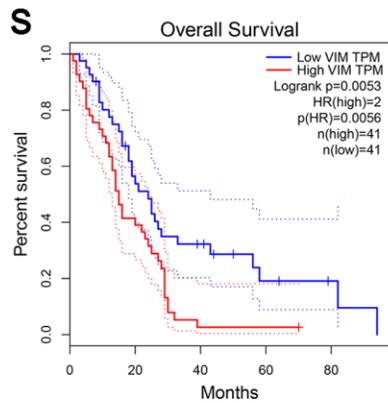
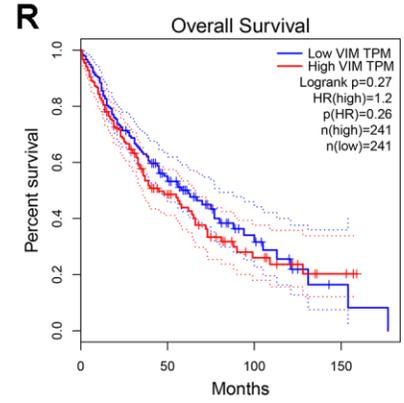
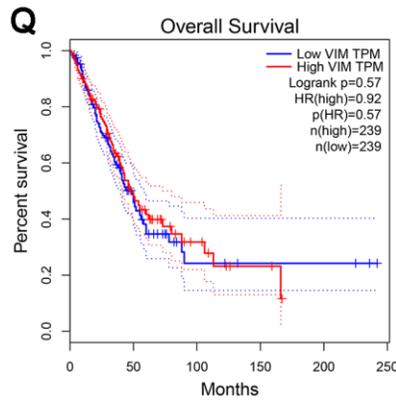
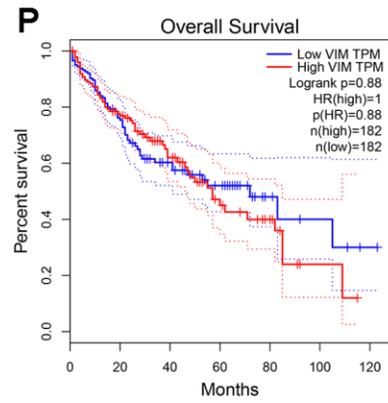
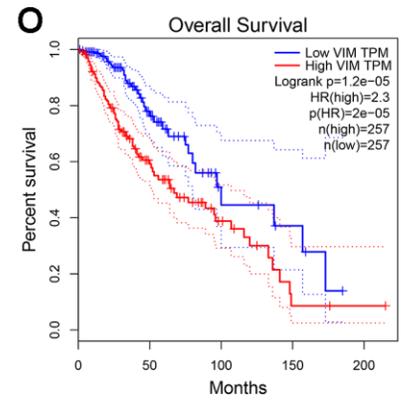
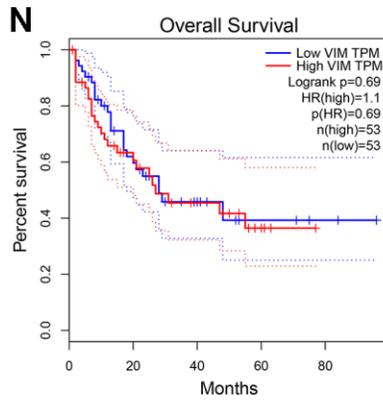
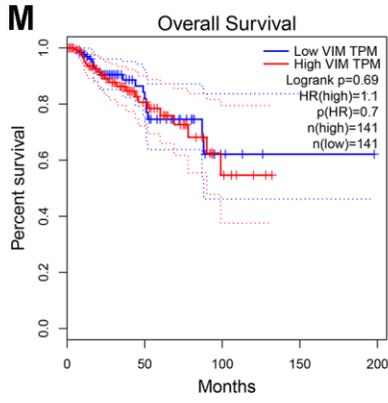


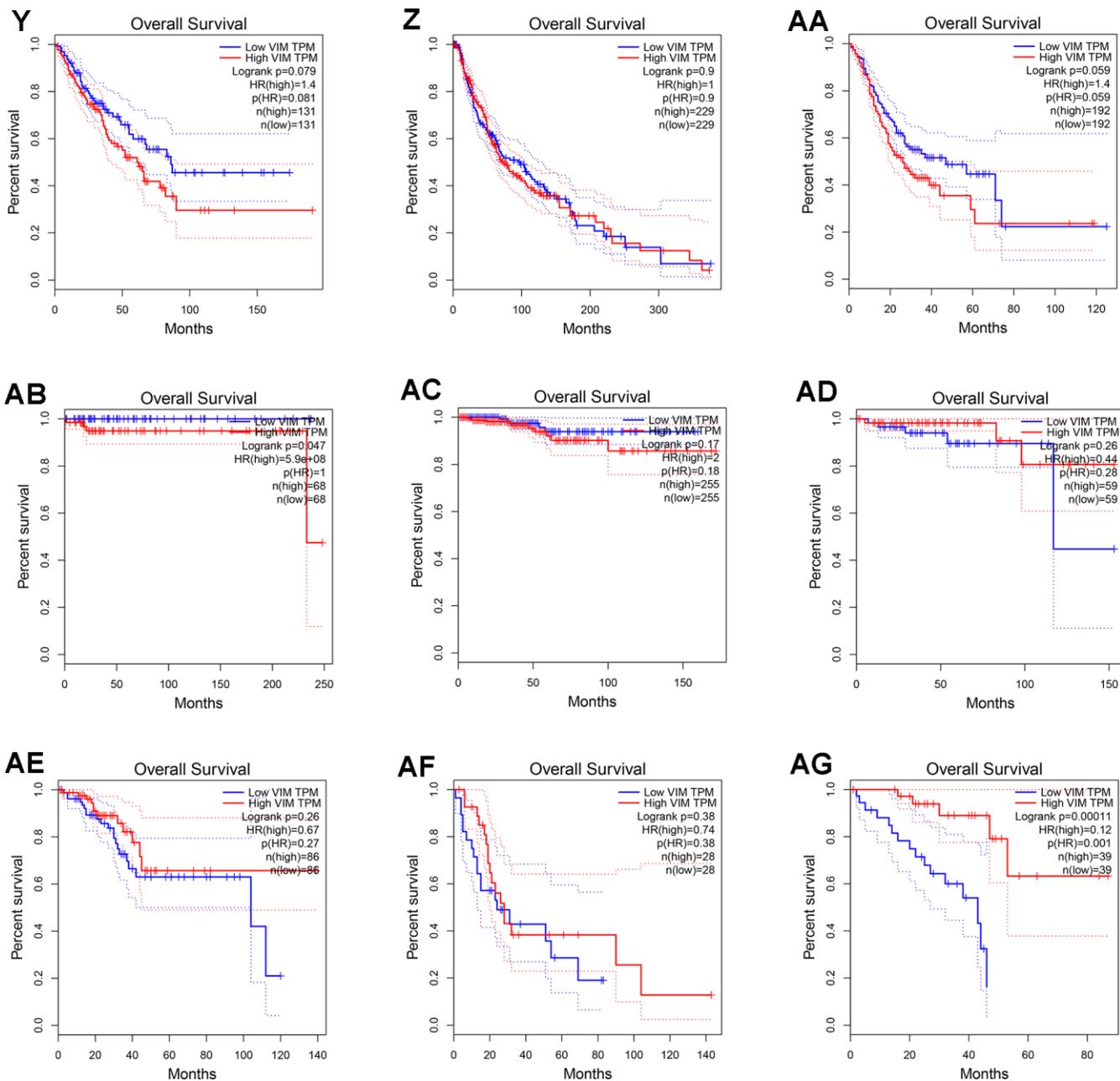




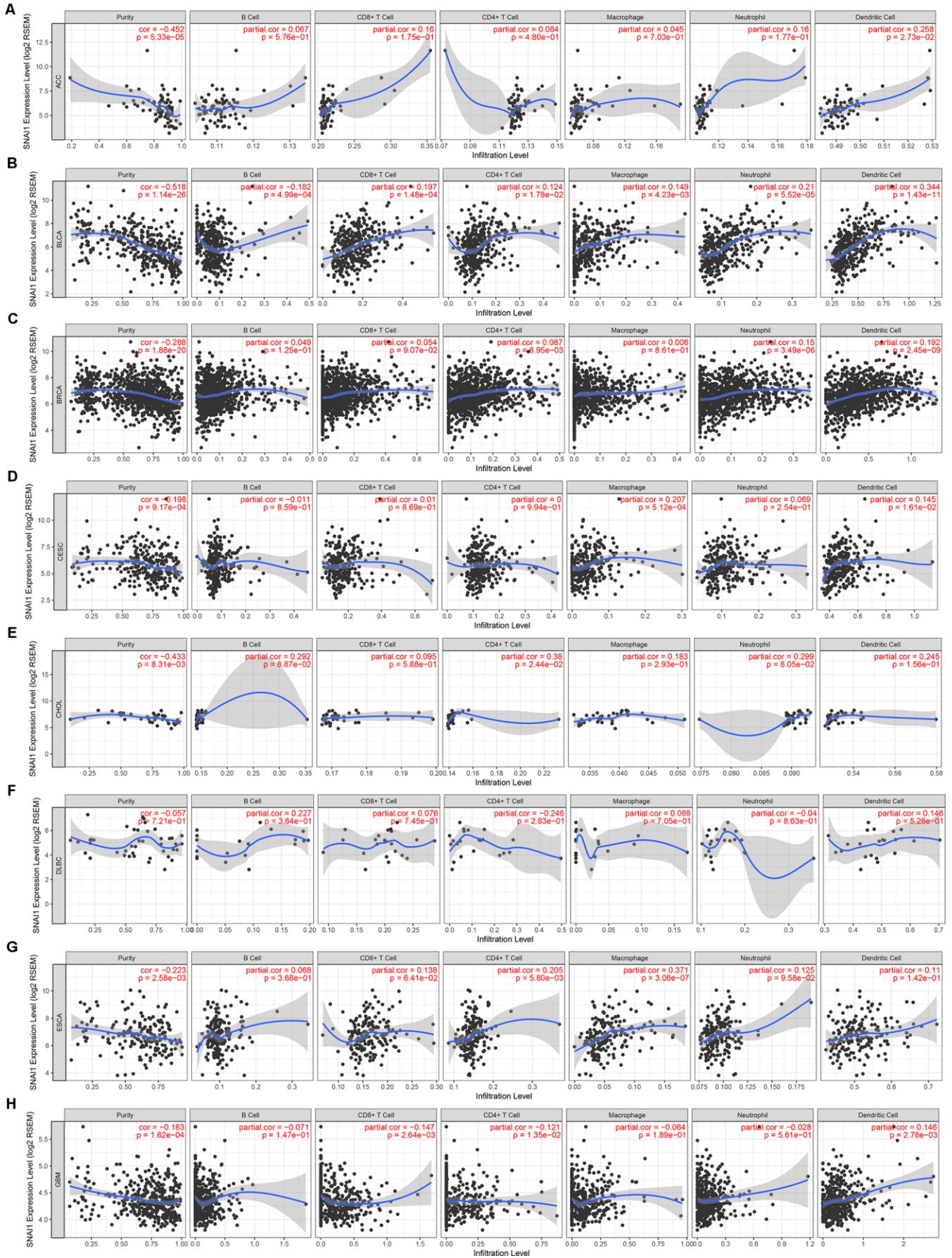
Supplementary Figure 6. Kaplan-Meier survival curves comparing the high and low expression of TWIST1 in different types of human cancers in TCGA dataset.

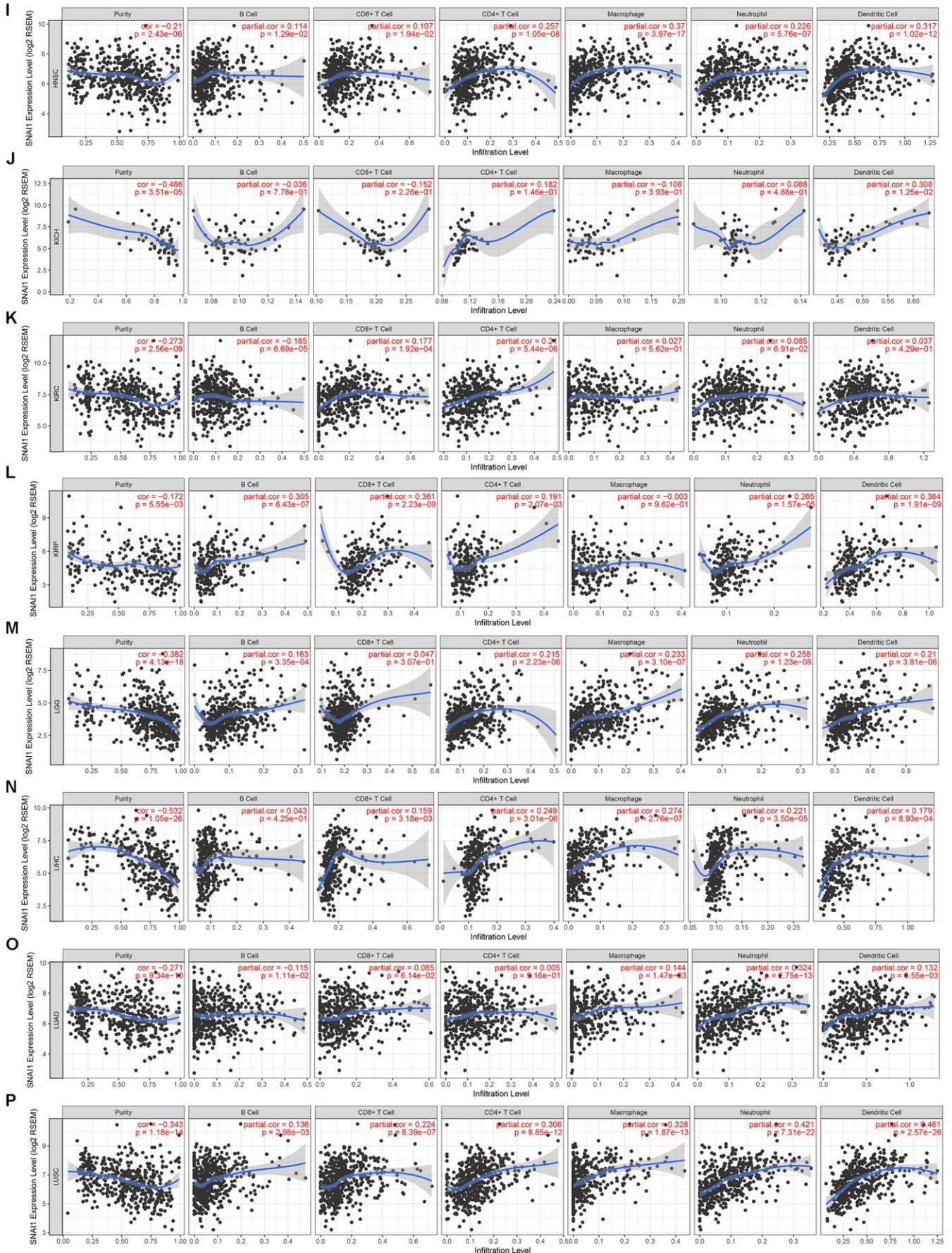


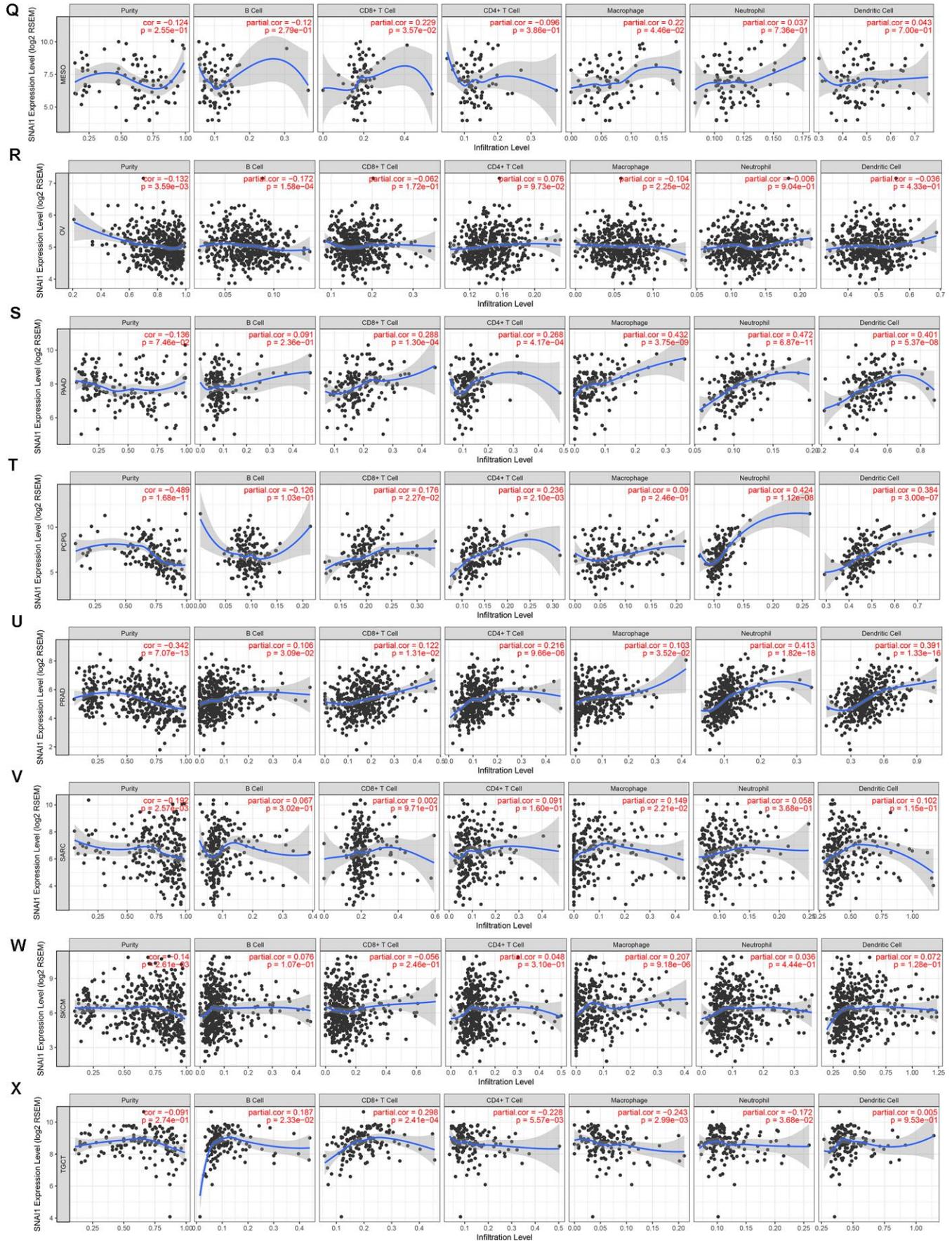


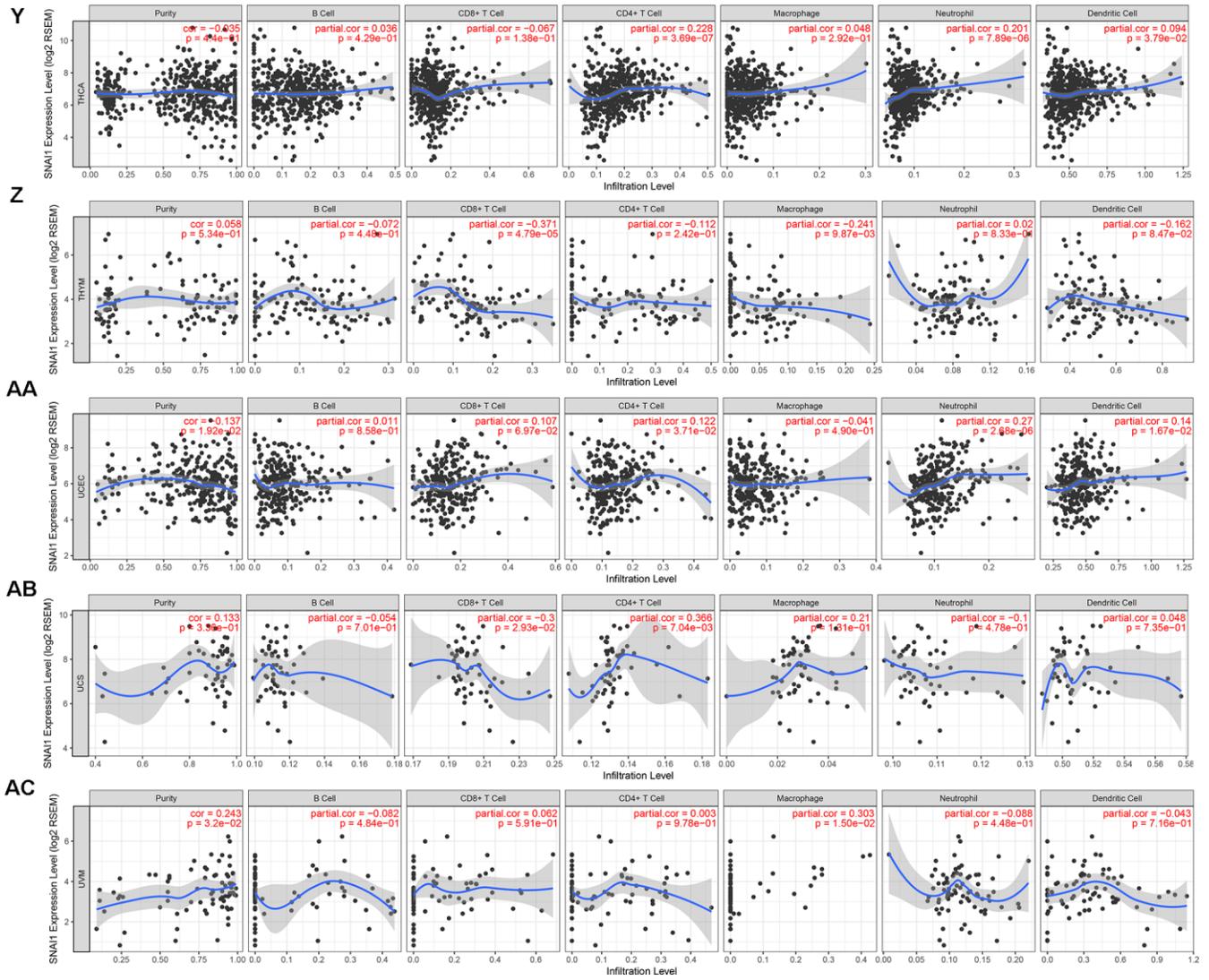


Supplementary Figure 7. Kaplan-Meier survival curves comparing the high and low expression of VIM in different types of human cancers in TCGA dataset.

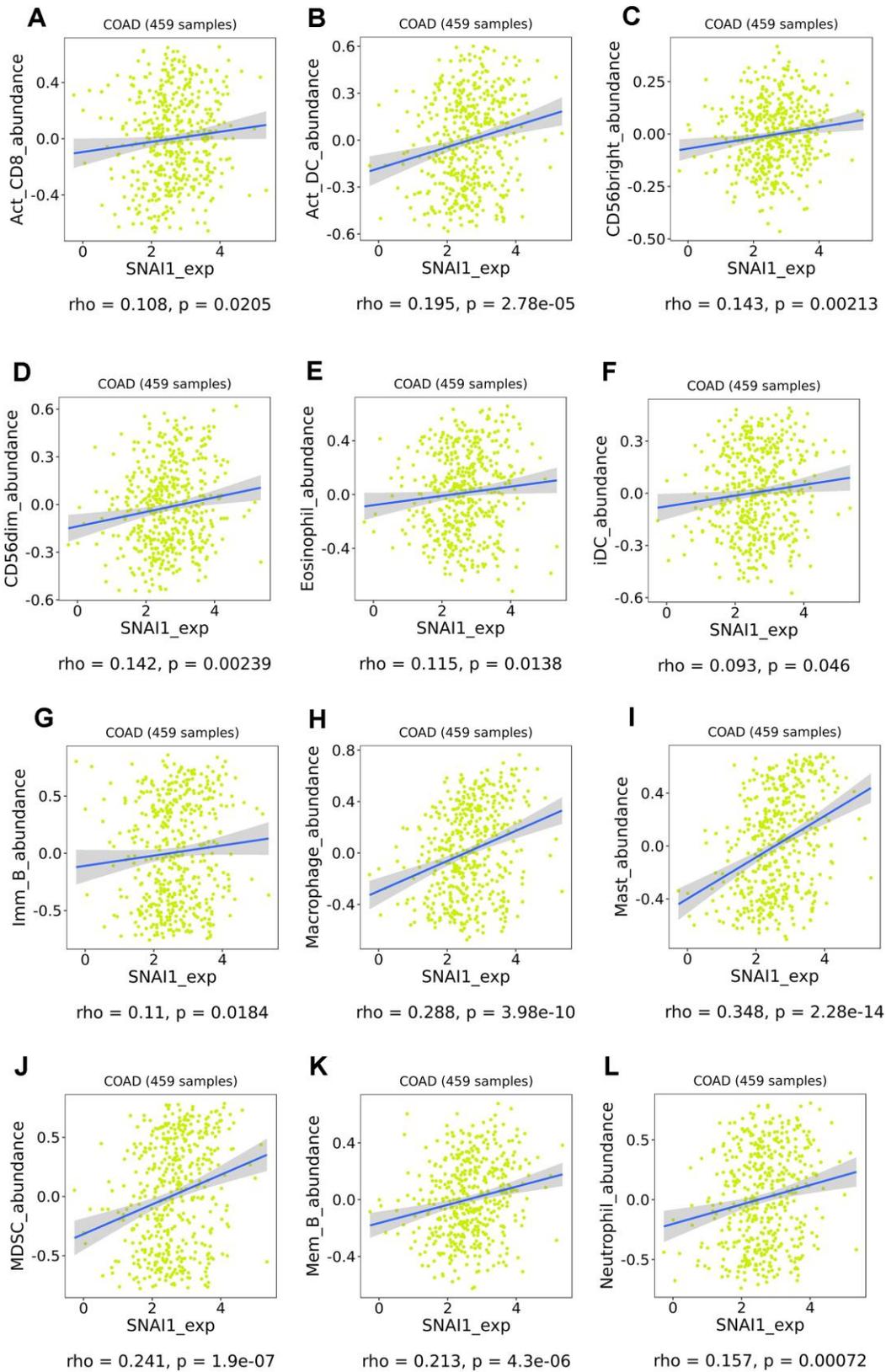


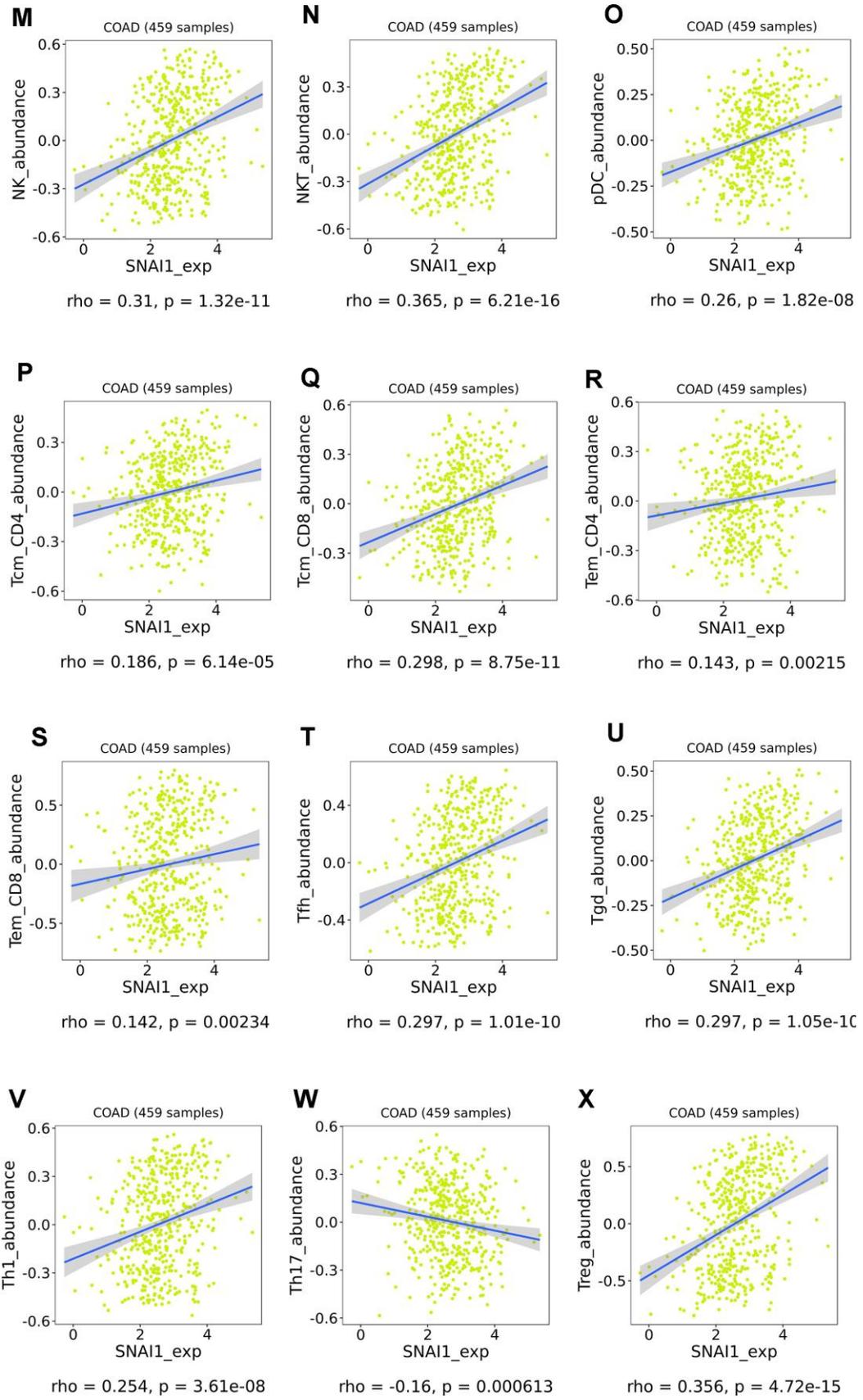






Supplementary Figure 8. Correlation of SNAI1 expression with tumor-infiltrating lymphocytes in different types of human cancers in TCGA dataset.





Supplementary Figure 9. Correlation of SNAI1 expression with 28 TILs in different types of human cancers in TCGA dataset.

## TILRR (FREM1 isoform 2) is a prognostic biomarker correlated with immune infiltration in breast cancer

Xiao-Yi Xu<sup>1,7,\*</sup>, Wen-Jing Guo<sup>1,\*</sup>, Shi-Hua Pan<sup>1</sup>, Ying Zhang<sup>1,3</sup>, Feng-Lin Gao<sup>1,3</sup>, Jiang-Tao Wang<sup>6</sup>, Sheng Zhang<sup>1,3</sup>, He-Ying Li<sup>1</sup>, Ren Wang<sup>5,7</sup>, Xiao Zhang<sup>1,2,3,4</sup>

<sup>1</sup>CAS Key Laboratory of Regenerative Biology, Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, and Guangzhou Medical University, Guangzhou, Guangdong, China

<sup>2</sup>Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou 510530, Guangdong, China

<sup>3</sup>Guangdong Provincial Key Laboratory of Biocomputing, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, Guangdong, China

<sup>4</sup>Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, Guangdong, China

<sup>5</sup>Affiliated Cancer Hospital and Institute, Guangzhou Medical University, Guangzhou 511436, Guangdong, China

<sup>6</sup>Department of Pathology, First People Hospital, Changde 415003, Hunan, China

<sup>7</sup>State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center, Guangzhou 510060, Guangdong, China

\*Equal contribution

**Correspondence to:** Xiao Zhang, Ren Wang; **email:** [zhang\\_xiao@gibh.ac.cn](mailto:zhang_xiao@gibh.ac.cn), [wangren@sysucc.org.cn](mailto:wangren@sysucc.org.cn)

**Keywords:** *TILRR*, immune cell infiltrating, prognostic biomarker, breast cancer

**Received:** January 24, 2020

**Accepted:** July 7, 2020

**Published:** October 8, 2020

**Copyright:** © 2020 Xu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

In atherosclerosis, upregulated *TILRR* (FREM1 isoform 2) expression increases immune cell infiltration. We hypothesized that *TILRR* expression is also correlated with cancer progression. By analyzing data from OncoPrint and the Tumor Immune Estimation Resource, we found that *TILRR* mRNA expression was significantly lower in breast cancer tissue than adjacent normal tissue. Kaplan-Meier survival analysis and immunohistochemical staining revealed shortened overall survival and disease-free survival in patients with low *TILRR* expression. *TILRR* transcript expression was positively correlated with immune score, immune cell biomarkers and the expression of *CXCL10* and *CXCL11*. *TILRR* expression was also positively correlated with CD8+ and CD4+ T-cell infiltration. These correlations were verified using the ESTIMATE algorithm, gene set enrichment analysis and Q-PCR. We concluded that impaired *TILRR* expression is correlated with breast cancer prognosis and immune cell infiltration.

### INTRODUCTION

The product of the FRAS1-related extracellular matrix 1 (*FREM1*) gene was first identified as a secretory protein excreted by mesenchymal cells that play a critical role in the development of multiple organs [1]. Multiple *FREM1* transcripts can be found in the mammalian

system [2]. In a previous study, *TILRR* (Toll-like/IL-1 receptor regulator) was identified as the IL-1R co-receptor, a 715-amino acid heparan sulfate glycoprotein encoded within the gene for the extracellular matrix protein *FREM1* [2]. Hence, in the National Center for Biotechnology database, *TILRR* is annotated as *FREM1* (isoform 2).

TILRR binds to the cell membrane through a C-terminal lectin domain, and partners with IL-1R1 as its co-receptor to enhance ligand binding. Overexpressed TILRR interacts with IL-1R1 via its TIR domain. This association potentiates the recruitment of the MyD88 adapter protein, and the signal amplification enhances activation of NF- $\kappa$ B and pro-inflammatory genes [3].

Recently, it was reported that TILRR upregulates pro-inflammatory gene expression in the progression of atherosclerosis [4]. Because TILRR induces immune cell infiltration, we wondered if TILRR expression might contribute to tumor progression. We hypothesized that the influence of TILRR on pro-inflammatory gene expression might have a prognostic value in cancer treatment.

First, we examined the correlation between *FREMI* expression and the prognosis of cancer patients. Subsequently, we investigated the expression of TILRR in tumor cells within different tumor microenvironments. The findings shed light on the crucial role of TILRR in breast cancers. The protein may be useful as a prognostic biomarker.

## RESULTS

### ***FREMI* mRNA expression levels in human breast cancer tumors**

To investigate the role of *FREMI* gene expression in cancer, we determined the *FREMI* mRNA levels in tumor and normal adjacent tissues of 20 cancer types. Data obtained from Oncomine was analyzed using the following threshold criteria: 2-fold change,  $P$  value  $< 0.0001$  and a gene rank of 10%. Downregulation of *FREMI* gene expression was found in breast cancer, ovarian cancer and pancreatic cancer tissues. In breast cancer tissues, 9 out of 43 samples met the threshold criteria in 4 out of 10 datasets. In the other two cancer types, only a single event reached threshold criteria (Figure 1A).

Further evaluation of *FREMI* expression in multiple malignancies was carried out using the gene expression data from The Cancer Genome Atlas (TCGA). Differential *FREMI* expression between the tumor and adjacent normal tissues is illustrated in Figure 1B. In most types of cancer, *FREMI* expression is significantly lower ( $P$  value  $< 0.001$ ) than in adjacent normal tissues (Figure 1B).

*FREMI* expression in breast cancer was further investigated using TCGA and Gene Expression Omnibus (GEO) data (Figure 1C and 1D).

*FREMI* expression in tumor tissues was 1.25-fold (GEO) and 4.18-fold (TCGA) lower than in normal adjacent tissues. These findings were confirmed in paired patient biopsy samples via Q-PCR ( $N = 36$ ,  $P = 0.0336$ ) (Figure 1D and 1E). In tumor biopsy samples, *FREMI* expression was reduced by as much as 6.16-fold ( $P < 0.001$ ). In 17 samples (47.2%), expression was decreased more than 2-fold (Figure 1E). We expanded our analysis of *FREMI* expression to various subtypes of breast cancer (Figure 1G and 1H). This analysis compared normal tissue ( $n = 120/n = 75$ ) with tissues from the following breast cancer subtypes: Basal-like ( $n = 168/n = 75$ ), *HER2*-enriched ( $n = 78/n = 24$ ), Luminal-like A ( $n = 493/n = 89$ ) and Luminal-like B ( $n = 194/n = 49$ ). These data showed decreased *FREMI* expression in Basal-like (6.07-fold/1.11-fold with mean), *HER2*-enriched [6.79-fold/1.09-fold with mean ( $P = 0.0279$ )], Luminal-like A (3.46-fold/1.08-fold with mean) and Luminal-like B (7.54-fold/1.11-fold with mean) tissues. Hence, *FREMI* expression is downregulated in breast cancer tissue.

### ***FREMI* transcription level is correlated with survival and progression in breast cancer patients**

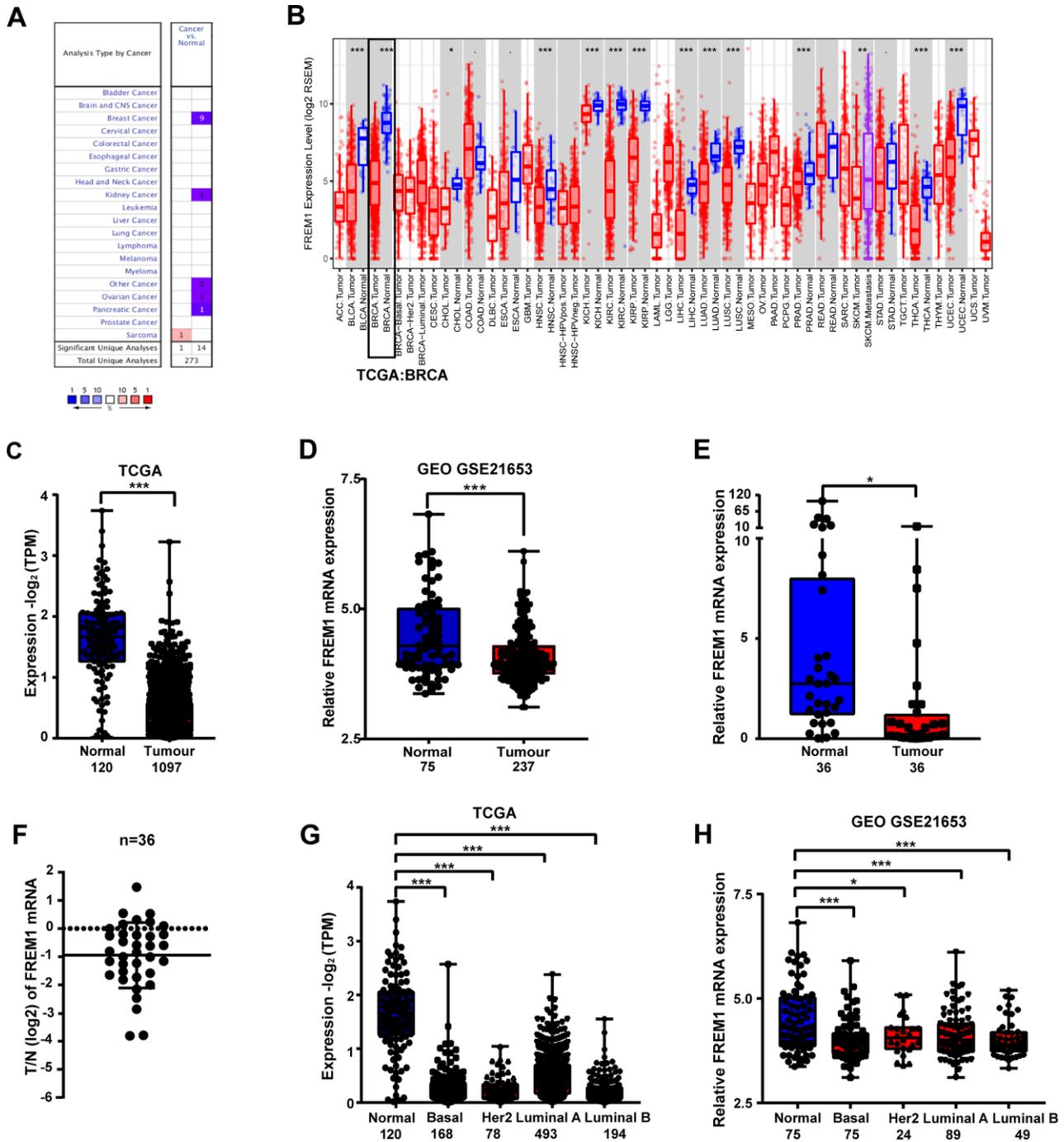
We further analyzed TCGA data to determine if there is a relationship between *FREMI* expression and overall survival (OS) or disease-free survival (DFS) in breast cancer patients. Patients were divided into high- or low-level groups over the median value of *FREMI* expression in breast cancer tissues. Patients with high *FREMI* expression in their tumors experienced a prolonged OS and DFS compared with those who with low *FREMI* expression (Figures 2A and 2B). Additionally, KM analysis revealed that *HER2*-positive *BRCA* patients in the low-level *FREMI* expression group generally demonstrated shorter DFS (Figures 2C and 2D). To confirm the clinical significance of downregulated *FREMI* expression in breast cancer patients, the correlation between survival rate and *FREMI* transcription was determined. Survival analysis of GEO data indicated that low-level *FREMI* expression was correlated with reduced OS and DFS (Figure 2E and 2F), and the same scenario was shown in *HER2*-positive *BRCA* patients (Figure 2G and 2H). This data was consistent with the results obtained from TCGA cohort. These results suggest that reduced *FREMI* transcription influences breast cancer tumor progression and is associated with shortened patient survival.

### **Downregulation of the *FREMI* protein in primary human breast cancer tissues correlates with disease progression**

To determine the clinical significance of the *FREMI* protein, we performed representative

immunohistochemical (IHC) staining in primary human breast cancer tumor and adjacent normal tissue (N = 47). *FREM1* staining in human primary breast cancer was scored as follows: +++, high; ++, moderate;

+, weak; -, negative. *FREM1* expression was significantly decreased in tumors, as indicated by lower IHC staining scores ( $P < 0.001$ ) (Figure 3A and 3B). Among the 47 breast cancer samples, *FREM1* staining



**Figure 1. *FREM1* expression levels in different types of human cancers.** (A) Expression levels of *FREM1* in different types of cancer compared with normal tissues. Data is from the Oncomine database. (B) Expression levels of *FREM1* in different tumor types from the TCGA database were analyzed by TIMER. (C, D) *FREM1* expression levels analyzed by TIMER in normal and tumor samples from the TCGA and GEO databases. (E, F) *FREM1* expression levels were analyzed by Q-PCR in paired normal and tumor tissues. (G, H) *FREM1* expression levels in normal and subtypes of breast cancer samples from the TCGA and GEO databases were analyzed by TIMER. (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).

was high in 4 samples (8.5%), moderate in 22 samples (46.8%) and weak or undetectable in 21 samples (44.7%) (Figure 3C and 3D). The AOD (average optical density) quantified data set was plotted in Supplementary Figure 1.

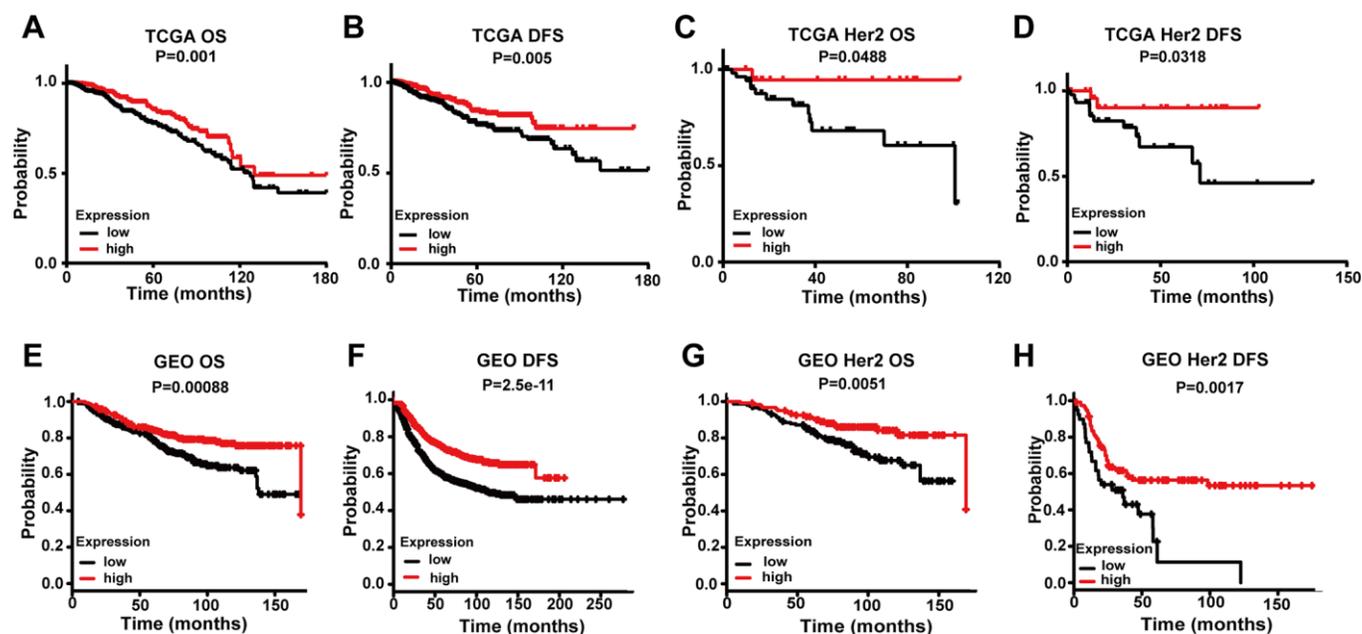
From the clinical information provided with the tissue chip, the survival rate was analyzed. High *FREM1* expression correlated with a prolonged OS and DFS compared with low *FREM1* expression, but the statistical significance of the DFS prolongation was weak ( $P = 0.0509$ ) (Figure 3E and 3F). When data was re-analyzed only using *HER2*-enriched samples, the correlation between *FREM1* expression and survival rate was robustly significant in both OS ( $P = 0.0006$ ) and DFS ( $P = 0.0035$ ) (Figure 3G and 3H). Thus, *FREM1* downregulation in breast cancer tissue, especially in *HER2*-enriched samples, was correlated with shorter survival times. This finding from IHC staining is consistent with our database analysis.

### ***TILRR* is the clinically relevant isoform of *FREM1* in breast cancer**

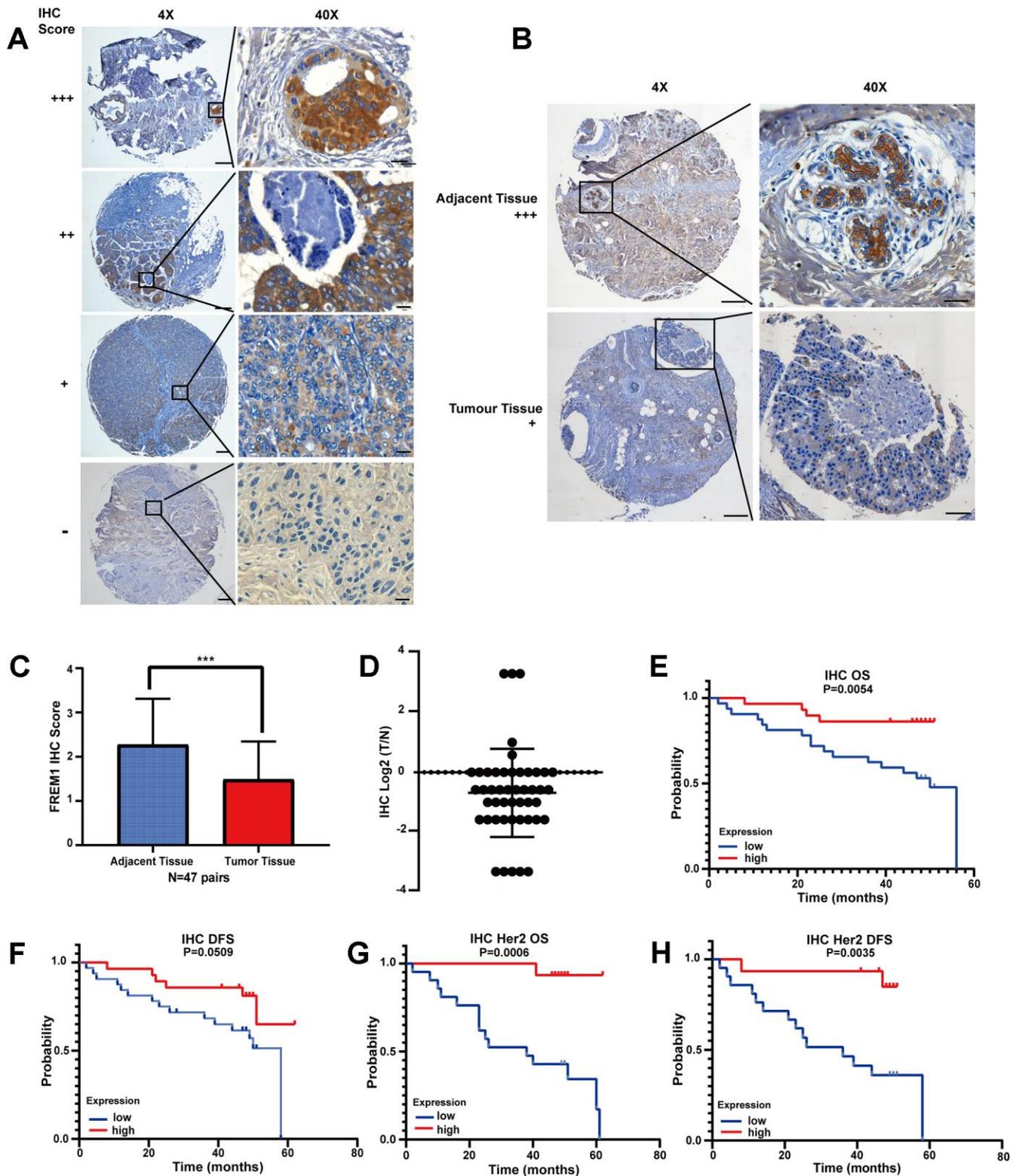
Previously, *TILRR* has been identified as isoform 2 of the *FREM1* gene [3]. We wanted to determine if *TILRR* is the dominate isoform that is downregulated in breast cancer. In the Ensembl and Gene Expression Profiling Interactive Analysis 2 (GEPIA2) databases, three

different *FREM1* isoforms are described. In these isoforms, ENST00000422223.6 and ENST00000380880.3 encode a protein with 2179 amino acids, and the much shorter version of ENST00000380894.5 encodes a protein with 715 amino acids. ENST00000380894.5 is the *FREM1* isoform 2 transcript, which we referred to previously as *TILRR* (Figure 4A). The correlation between breast cancer survival rate and *FREM1* usage was investigated using the GEPIA2 database. In Figure 4B, the violin-plot and bar-plot panels present the expression distribution and usage of each *FREM1* isoform. *TILRR*, or *FREM1* isoform 2, shows convincing cancer specificity with strong usage and a significantly low hazard ratio (HR).

Using a *TILRR*-specific primer (Figure 4C), *TILRR* expression in 30 paired breast cancer samples was investigated using Q-PCR. The data showed a 4.39-fold decrease ( $P = 0.0274$ ) in expression in tumor tissue compared with normal tissue (Figure 4D and 4E), which confirmed that *TILRR* is downregulated in breast cancer tissues. The protein expression identity of *FREM1* and *TILRR* was verified using four paired breast cancer tissues (sample numbers: 160, 172, 109 and 164) (Figure 4F and 4G). The *FREM1* commercial polyclonal antibody blotting pattern was identical to that of the *TILRR* polyclonal antibody, which detected a 75 KDa peptide with an approximately 4.76- and 9.09-fold



**Figure 2. Comparing the low and high expression levels of *FREM1* by Kaplan-Meier survival analysis in breast cancer and the *HER2* subtype. (A–D) Survival curves of OS and DFS in breast cancer and *HER2* from the TCGA database. (E–H) Survival curves of OS and DFS in breast cancer and *HER2* from the GEO database. OS, overall survival; DFS, disease-free survival.**



**Figure 3. FREM1 expression is reduced in tumor tissues, which is associated with poor survival.** (A) Representative IHC staining of FREM1 in human primary tumor and adjacent non-tumor tissue. Magnifications: 4X; boxed area is 40X. Scale bar: 4X, 200  $\mu$ m; 40X, 20  $\mu$ m. (B) FREM1 level is higher in adjacent tissues than in tumor tissues. Magnifications: 4X; and boxed area is 40X. Scale bar: 4X, 200  $\mu$ m; 40X, 20  $\mu$ m. (C, D) Analysis of FREM1 IHC staining scores in adjacent non-tumor (n = 48) and tumor tissue (n = 56). (\* $P$  < 0.05, \*\* $P$  < 0.01, \*\*\* $P$  < 0.001). (E–H) Survival curves of OS and DFS in breast cancer and *HER2* with low and high *FREM1* expression. Median survival time of the high-expression group versus low-expression group.

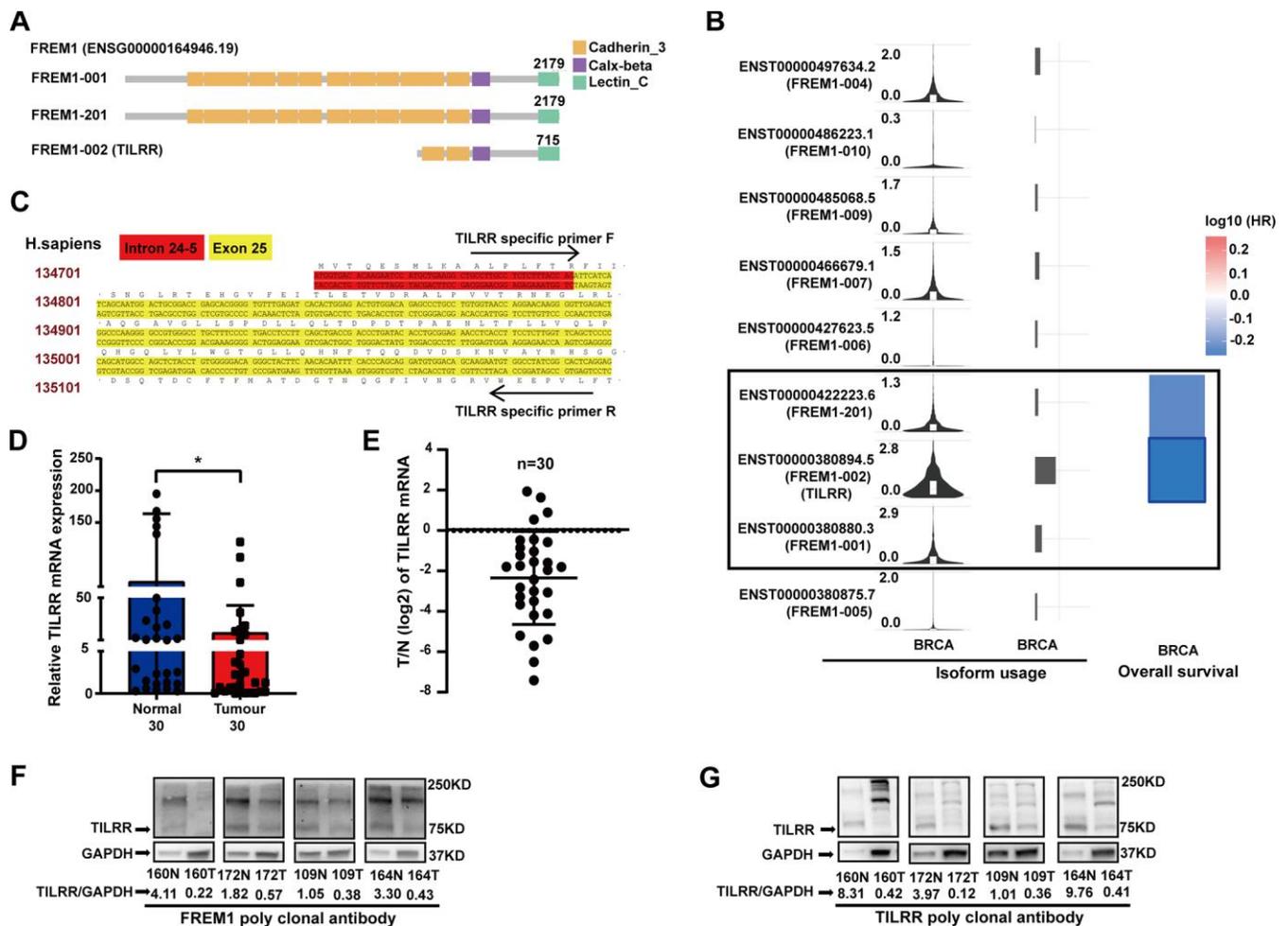
reduction in expression in tumor tissues. A similar finding was discovered using TILRR monoclonal probes (data not shown). These results confirm that downregulation of *TILRR* is associated with breast cancer prognosis.

### Analysis of *TILRR* transcript level associated with tumor-infiltrating lymphocytes

As shown above, *TILRR* is the dominant isoform of *FREM1* expressed in breast cancer tissues. A prior study found that TILRR is related to monocyte infiltration in atherosclerosis plaque development [4]. Moreover, the lymphocyte-specific immune (LYM) recruitment metagene signature is related to tumor infiltration by lymphocytes, and it is associated with a favorable prognosis in breast cancer. To investigate the

role of TILRR in the mediation of immune cell infiltration, the correlation between *TILRR* expression and LYM was evaluated in TCGA breast cancer data. The LYM metagene sets *PTPRC* (*CD45*), *CD53*, *LCP2* (*SLP-76*), *LAPTM5*, *DOCK2*, *IL10RA*, *CYBB*, *CD48*, *ITGB2* (*LFA-1*) and *EVI2B* were positively correlated with expression of *FREM1* (Spearman  $r = 0.42$   $P = 3.5e-47$ ; Pearson  $r = 0.3$   $P = 0$ ) (Figure 5A).

The estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) algorithm can calculate the gene expression signature in tumor cells and normal, tumor-associated epithelial and stromal cells, immune cells and vascular cells [5]. Stromal cells are thought to have essential roles in tumor growth, disease progression and drug resistance, and ESTIMATE



**Figure 4. *TILRR*, the *FREM1* isoform 2 transcript, is clinically relevant in breast cancer.** (A) The three *FREM1* isoforms. (B) *TILRR*, the gene encoding isoform 2, shows convincing cancer specificity with strong usage and significantly low hazard ratio. (C) *TILRR*-specific primer of human DNA. (D, E) *TILRR* expression levels were analyzed by Q-PCR in paired normal and tumor tissues (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). (F, G) *TILRR* immunoblotting (indicated by the *FREM1* or *TILRR* antibody) in paired normal and tumor tissues. The band intensities are normalized to GAPDH.

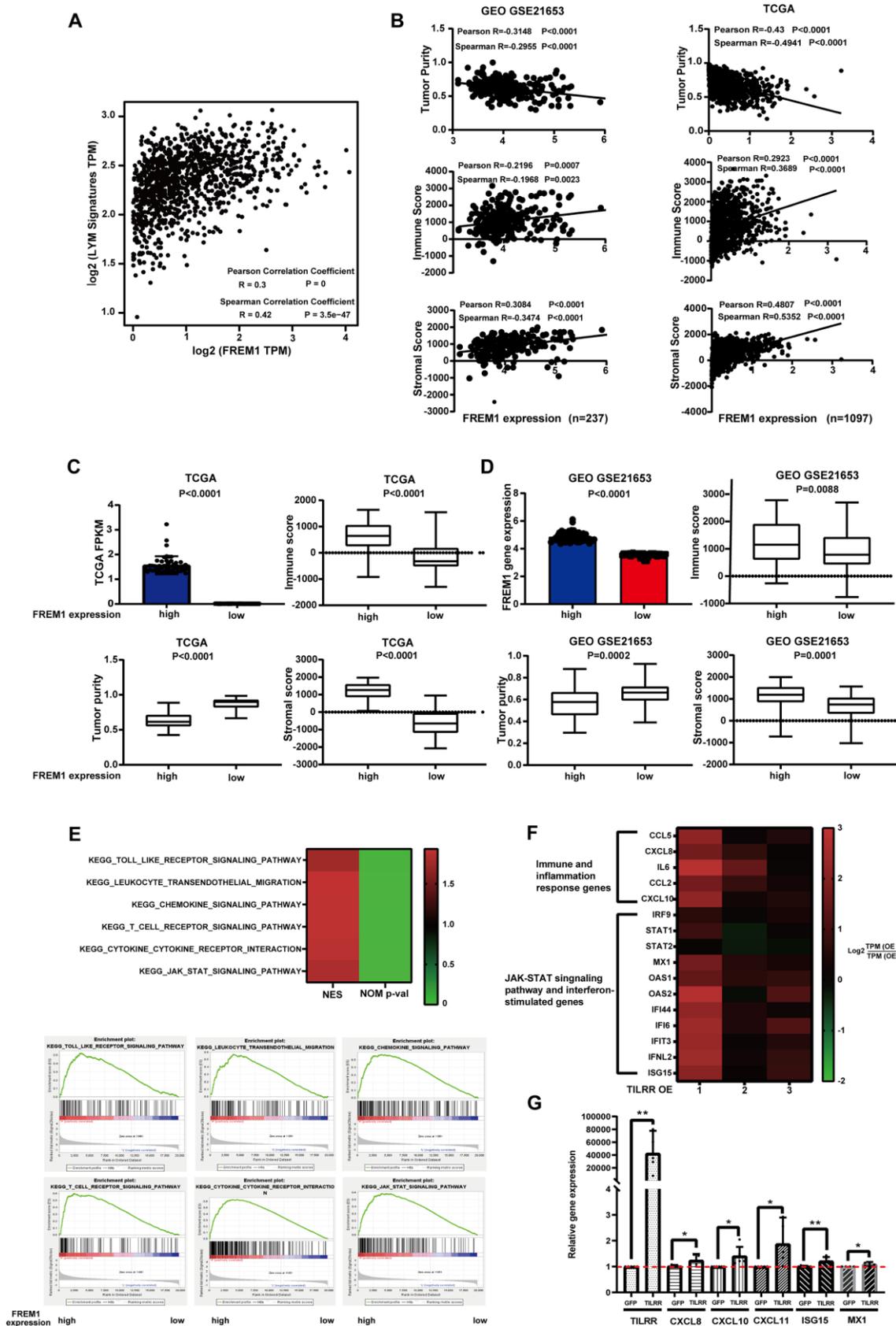
immune scores could serve as an indicator for immunotherapy response [5]. Tumor purity and the expression of *TILRR* were analyzed using the ESTIMATE algorithm to investigate in GEO (N = 237) and TCGA (N = 1097) data. The results showed that the expression of *TILRR* was negatively correlated with tumor purity, but positively correlated with immune score and stromal score (Figure 5B). This correlation was confirmed with the 50 highest and 50 lowest *TILRR*-expressing tumor tissue samples (Figure 5C and 5D).

To understand the underlying mechanism of *FREMI* in breast cancer, we analyzed differentially expressed data from the 50 highest and 50 lowest *FREMI*-expressing tumor tissue samples. The gene expression enrichment of several signal pathways was analyzed using gene set enrichment analysis (GSEA). *TILRR* is positively correlated with Toll-like receptor signaling and cytokine-to-cytokine receptor interactions (Figure 5E), which is consistent with our previous finding [3]. We also found that leukocyte transendothelial migration, chemokine signaling, T-cell receptor signaling and JAK-STAT signaling are associated with the migration and infiltration of immune cells (Figure 5E). To verify these pathways, RNA-seq analysis of the BT474 cancer cell line overexpressing *TILRR* was performed using an empty GFP construct as a control. The analysis showed that *TILRR* overexpression not only potentiates *IL-6* and *CXCL8* immune- and inflammation-induced gene expression, but it also induces the expression of other cytokine genes such as *CCL5/2* and chemokine genes such as *CXCL10* (Figure 5F). JAK-STAT signaling pathways and interferon-stimulated genes (further detailed analysis will be published elsewhere) were also upregulated (Figure 5F). This upregulation of pro-inflammatory secretory factors and pathway activation indicated that enhanced *TILRR* expression might cause long distal immune recruitment. Q-PCR was performed to confirm RNA-seq detection among expression of *CXCL8*, *CXCL10*, *CXCL11*, *MX1* and the interferon stimulation gene, *ISG15* (Figure 5G).

Furthermore, we analyzed the 50 highest *TILRR*-expressing and 50 lowest *TILRR*-expressing tumor tissue samples (TCGA database) to determine the relationship between the expression of *CXCL10* and *CXCL11* and the expression of *TILRR* in cancer tissues. The results showed that high expression of *TILRR* positively correlates with *CXCL10* and *CXCL11* expression (Supplementary Figure 2). These findings strengthened the hypothesis that *TILRR* expression is associated with breast cancer progression and prognosis, likely through signaling pathways that regulate the distal recruitment of immune cell infiltration.

## Expression of *TILRR* is correlated with immune cell infiltration in breast cancer

Increasing evidence suggests that tumor-infiltrating immune cells can be an indicator in the clinical analysis of tumor samples [6, 7]. Gene expression profiling of heterogeneous cell populations in cancer tissue, including tumor-infiltrating lymphocytes, serves as an independent predictor of survival in prognostic cancer models [8, 9]. Thus, we used the TIMER database to evaluate the correlation between *FREMI* mRNA expression and six different infiltrating immune cell types (B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils and dendritic cells) in different subtypes of breast cancer. Data showed that the *TILRR* transcription level was inversely related to the purity of tumor tissue in breast cancer subtypes (Figure 6A–6D, first line on the left). The low heterogeneity in tumor tissue correlated with high expression of *FREMI* in all but one subtype, which is consistent with the positive association between *TILRR* expression and level of immune cell infiltration (Figure 6). The exception was macrophage cells of the *HER2* subtype ( $r = -0.107$ ,  $P = 4.23e-01$ ). The infiltration levels of CD8+ and CD4+ cells were significantly positively correlated with *TILRR* transcription. The level of CD8+ T-cell infiltration with *TILRR* expression in *BRCA* ( $r = 0.379$ ,  $P = 16e-34$ ), *BRCA*-Basal ( $r = 0.275$ ,  $P = 2.10e-03$ ), and *BRCA*-*HER2* ( $r = 0.406$ ,  $P = 1.2e-03$ ), and *BRCA*-Luminal ( $r = 0.48$ ,  $P = 7.89e-28$ ) were all significantly strong, respectively, which is similar to the CD4+ T-cell infiltration levels (*BRCA*:  $r = 0.369$ ,  $P = 2.07e-32$ ; Basal:  $r = 0.181$ ,  $P = 4.57e-02$ ; *HER2*:  $r = 0.595$ ,  $P = 8.37e-07$ ; luminal:  $r = 0.42$ ,  $P = 2.44-24e$ ). On the other hand, in the tumor tissue, the infiltration level from B cells, macrophages, neutrophils and dendritic cells correlated more weakly with *TILRR* transcription in subtypes of breast cancer (Figure 6). To verify the expression of *TILRR* and the infiltration of immune cells, the 50 highest *TILRR*-expressing and 50 lowest *TILRR*-expressing tumor tissue samples were selected (TCGA database) to analyze the expression of marker genes for CD8+, CD4+ and T cells (general). The results showed that *CD8A*, *CD8B* was expressed in tissues with high *TILRR* expression. (Figure 6E). Analysis of GEO data generated similar results (Supplementary Figure 3). To further verify this result, we selected the 5 samples with the highest *TILRR* expression and the 5 with the lowest *TILRR* expression patients shown in Figure 4D to evaluate the expression of marker genes in CD4- and CD8-expressing cells. Among these 10 samples, the expression of *CD8A*, *CD8B* and *CD4* in *TILRR* high- and low-expressing samples were positively correlated (Figure 6F). We also investigated the correlation between *TILRR* expression and biomarkers from other immune cells. Strong



**Figure 5.** FREM1 mRNA levels were associated with tumor infiltrating lymphocytes. (A) The average expression of the LYM metagene signature [*PTPRC* (*CD45*), *CD53*, *LCP2* (*SLP-76*), *LAPTM5*, *DOCK2*, *IL10RA*, *CYBB* and *CD48*, *ITGB2* (*LFA-1*) and *EVI2B*] in breast

cancers from the TCGA database relative to the *FREM1* mRNA level. (B–D) Tumor purity, immune score and stromal score were analyzed using the ESTIMATE algorithm from GEO (N = 237) and TCGA (N = 1097) database. (E) GSEA analysis was used to demonstrate the correlation between *FREM1* expression and the KEGG enriched pathway. (F) Immune and inflammation response genes, the JAK-STAT signaling pathway and interferon-stimulated genes were analyzed by RNA-seq in BT474 cancer cells. (G) *CXCL8*, *CXCL10*, *CXCL11* and the interferon stimulation genes, *ISG15* and *MX1*, were analyzed by Q-PCR in BT474 cancer cells.

associations were found between NK cells, Treg cells, T-cell exhaustion and *TILRR* expression (Supplementary Table 1). Consequently, it is likely that *TILRR* expression in breast cancer tissue is involved in immune cell recruitment.

## DISCUSSION

In the present study, we observed that *TILRR* expression level was associated with the survival of patients in breast cancer. Database analysis revealed that *TILRR* mRNA expression was significantly lower in breast cancer tissue and correlated with shorter OS and DFS. The TCGA and GEO data analysis were confirmed by IHC staining and Q-PCR. *TILRR* transcript expression is correlated with immune score, immune cell biomarkers and LYM metagene signature, which was verified using the ESTIMATE algorithm and GEPIA2 database analysis. This consistent association between increased *TILRR* mRNA levels and a favorable prognosis for the patient is shown in Figure 4. Two of five patients with higher *TILRR* mRNA levels had a favorable prognosis (Figure 6F). However, in the group with lower *TILRR* mRNA, 3 out of 5 patients were dead due to cancer (data not shown). To our knowledge, this is the first study to describe *TILRR* expression in cancer tissue.

In this study, we found that *TILRR* expression is profoundly downregulated in breast cancer and correlated with disease-specific survival. The PyMT mouse, a genetically engineered mouse model that is widely used to study human breast cancer, gene profiling and expression analysis has illustrated the effects of *TILRR* downregulation [10]. In *PyMT/Il1a*<sup>-/-</sup> and *PyMT/Il1r1*<sup>-/-</sup> mice, IL-1R1 signaling suppresses mammary tumor cell proliferation early in tumorigenesis and facilitates breast cancer outgrowth with pulmonary metastasis [11]. IL-1Ra is overexpressed in multiple cancers, including multiple myeloma, leukemia, cervical, ovarian, colorectal, pancreatic and breast cancer, but is downregulated in others [11–15]. This is consistent with the idea that *TILRR* enhances IL-1 alpha affinity binding to IL-1R1, as IL-1R1 is a low-copy, high-affinity receptor [3, 16]. Interestingly, the IL-1R1 signaling pathway has been reported to promote tumor growth, angiogenesis and metastasis in some contexts [17], while stimulating anti-tumor immunity or directly suppressing tumorigenesis in others [11]. *TILRR* activates oncogene *RAS* upstream of *TRAF6* in the IL-1R1-mediated pathway, and *TILRR* overexpression

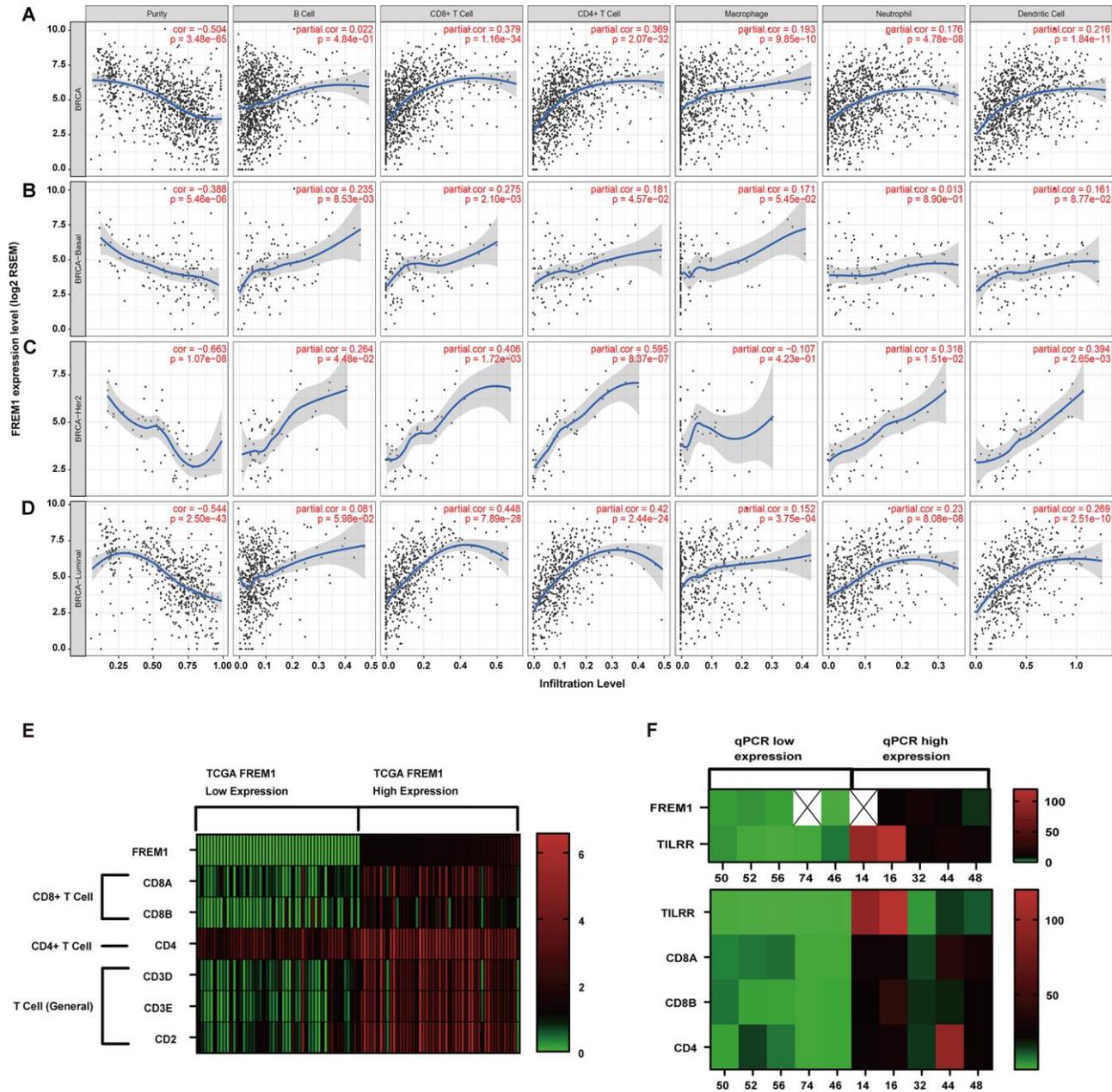
enhances AKT phosphorylation and HeLa-cell survival via the *TILRR* R425 site [18]. Previously it was reported that under *LDLR*<sup>-/-</sup> or *APOE*<sup>-/-</sup> genetic conditions, monocyte activation and infiltration were reduced in atherosclerosis and in the lung of *TILRR* KO mice or mice injected with *TILRR*-blocking antibody [4]. As reported, under *LDLR*<sup>-/-</sup> conditions, mouse models of hypercholesterolemia developed a smaller tumor. These mice were characterized by increased LDLR expression, as well as shorter OS and decreased DFS [19]. Interestingly, *TILRR* expression level is low in healthy tissue; however, it is remarkably enhanced in atherosclerotic plaques with a high level of immune cell infiltration [4].

Another important aspect of this study is the correlation between *TILRR* expression and the level of immune infiltration. Through data analysis, we observed associations between *TILRR* mRNA levels and immune score, immune cell biomarkers, the LYM metagene signature and levels of infiltrating immune cells. In previous studies, *TILRR* expression was correlated with monocyte infiltration; notably, *TILRR*<sup>-/-</sup> mice showed less recruitment of immune cells in atherosclerosis plaques [4]. ESTIMATE calculation and GSEA analysis of the top 50 highest *TILRR*-expressing samples compared with the top 50 lowest samples showed that *TILRR* transcription played a role in immune cell infiltration, migration and activation. On the other hand, RNA-seq and Q-PCR analysis showed that *TILRR* potentiates *CXCL10* and *CXCL11* chemokine expression in the BT474 cancer cell line. In response to specific chemokines, immune cells can regulate immune responses by migrating into the tumor microenvironment. It has been reported that the tumor production of *CXCL9* and *CXCL10* was repressed by enhancement of H3K27me3 and DNMT1-mediated DNA methylation. Moreover, EZH2 and DNMT1 are negatively associated with tumor-infiltrating CD8+ T cells [20]. In the tumor microenvironment, *CXCL11* upregulation enhanced CD8+ T-cell recruitment [21]. Collectively, we speculate that, in the tumor microenvironment, *TILRR* can enhance immune infiltration by regulating the *CXCL10* and *CXCL11* chemokines.

It has been confirmed that IL-18 potentiates IFN- $\gamma$ -induced *CXCL9*, *CXCL10*, and *CXCL11* mRNA expression and secretion by activating the NF- $\kappa$ B and JAK-STAT signaling pathways [22]. Our previous

studies showed that *TILRR* overexpression can increase the activation of the NF- $\kappa$ B signaling pathway, which agree with our RNA-seq data in this study. Interestingly, we also observed IFN- $\gamma$ -induced gene

expression; however, whether the *TILRR* effect on the upregulation of *CXCL-10* and *CXCL-11* was through an indirect cytokine chain reaction of multiple cell types remains unknown. We found in these tissues a positive



**Figure 6. Correlation of *FREM1* expression with immune infiltration level in the subtypes of breast cancer.** (A) *FREM1* expression is negatively related to tumor purity and has significant positive correlations with infiltrating levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils and dendritic cells in *BRCA*. (B) *FREM1* expression is negatively related to tumor purity and has positive correlations with infiltrating levels of B cells and CD8+ T cells, but not CD4+ T cells, macrophages, neutrophils and dendritic cells in *BRCA-Basal*. (C) *FREM1* expression is significantly negatively related to tumor purity and has significant positive correlations with infiltrating levels of CD8+ T cells, CD4+ T cells and dendritic cells, but not B cells, macrophages and neutrophils in *BRCA-HER2*. (D) *FREM1* expression is negatively related to tumor purity and has significant positive correlations with infiltrating levels of CD8+ T cells, CD4+ T cells, macrophages, neutrophils and dendritic cells, but not B cells in *BRCA-Luminal*. (E) The top 50 highest *TILRR*-expressing and top 50 lowest *TILRR*-expressing tumor tissue samples were selected (TCGA database) to analyze the expression of marker genes in CD8+, CD4+ and T cells (general). (F) *CD8A*, *CD8B* and *CD4* expression levels were analyzed by Q-PCR in the 5 highest *TILRR* expressing and the 5 lowest *TILRR* expressions samples.

correlation between *TILRR* upregulation and IFN- $\gamma$ -related signaling pathways. These results proved that the *TILRR* transcription level could indicate lymphocyte infiltration in breast cancer.

*TILRR*-related immune cell recruitment in breast cancer showed the strongest correlation with CD8+ positive cells and the T cell-related pathway. In cancer treatment, inhibiting immune checkpoint mediators, such as CTLA-4 and PD-1, has achieved noteworthy clinical outcomes in several malignancies [23–26]. Cardiovascular disease is also the consequence of targeted cancer therapies and chemotherapies in several clinical settings [27, 28]. *TILRR* expression related to chemokine secretion and NF- $\kappa$ B activation is well documented in cardiovascular disease [4]. Therefore, we conclude that *TILRR* might recruit immune cells through a similar cellular mechanism in breast cancer and atherosclerosis. We believe that *TILRR* could play different roles in different organs but may share a similar mechanism in breast cancer and atherosclerosis.

## MATERIALS AND METHODS

### *FREMI* gene expression of *BRCA* in the TCGA and GEO databases

The expression level of the *FREMI* gene in various types of cancers was identified in the Oncomine database (<https://www.oncomine.org/resource/login.html>) [29]. The threshold was determined according to the following values: *P* value of 0.001, fold change of 2. Differential expression module of TIMER (<https://cistrome.shinyapps.io/timer/>) was used to analyze the *FREMI* differential expression between tumor and adjacent normal tissues of various TCGA tumors [30, 31]. TCGA breast cancer data was downloaded from the Xena browser (<https://xenabrowser.net/datapages/>) [32]. GSE21653 data was downloaded from the GEO database, which contained 266 early cancer patients [33, 34].

### Immune cell infiltration analysis

The correlation of *FREMI* expression level with a LYM metagene signature [*PTPRC* (*CD45*), *CD53*, *LCP2* (*SLP76*), *LAPTM5*, *DOCK2*, *IL10RA*, *CYBB*, *CD48*, *ITGB2* (*LFA-1*) and *EVI2B*] was explored via Gene Expression Profile Interactive Analysis (GEPIA2; <http://gepia2.cancer-pku.cn/#index>) [35–37]. The stromal score, immune score and tumor purity of breast cancer (*BRCA*) tumor samples from the GSE21653 database (N = 237) and the TCGA database (N = 1097) were calculated by using the R 3.6.0 ESTIMATE package (1.0.13) [38]. The gene module of TIMER was used to evaluate the correlation of *FREMI* expression

with immune cell infiltration (including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells) in *BRCA*, which determined the purity-corrected partial Spearman's correlation and statistical significance. The correlation module of TIMER was used to calculate the Spearman's correlation and statistical significance between *FREMI* and immune cell marker genes. The top 50 highest and lowest *FREMI*-expressing TCGA tumor samples were used to conduct GSEA and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations analysis [39].

### RNA isolation and real-time PCR

Total mRNA of breast cancer tissue (SYSUCC, Sun Yat-sen University Cancer Center, Guangzhou, China) was extracted using TRIzol reagent (Invitrogen, #1556018) according to the manufacturer's protocols. The quantitative, real-time PCR using SsoAdvanced<sup>TM</sup> Universal SYBR Green Supermix (Bio-Rad, #1725274) was performed in an ABI StepOnePlus (ABI, #1725274). *FREMI* primer sequences: forward primer, 5'-AGAGCCCTGCCTGTGGTAAC-3'; reverse primer, 5'-GAAGGGGAATGCAAGAGTGTGATA-3'. *TILRR*-specific primer 5'-GCCTTGCCCTCTTTACCAGAT-3'; reverse primer, 5'-GAGTGCCGATAGGCACAT-3'. Relative gene expression was normalized to glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) (forward primer, 5'-GGAGCGAGATCCCTCCAAAAT-3'; reverse primer, 5'-GGCTGTTGTCA TACTTCTCATGG-3') expression and was analyzed using the  $2^{-\Delta\Delta CT}$  method.

### Immunohistochemistry (IHC)

Immunohistochemistry was performed on tissue section microarrays (Zhuolibiotech, #ZL-Brc Sur122). The staining procedure included heat-induced epitope retrieval using 0.1 M sodium citrate (pH 9.0, heated by microwave to 90–95 °C, 3 times for 5 minutes each), incubation with primary antibody at 4 °C overnight. Signal detection was performed using an IHC detection kit (Gene Tech, #GK500710). Microscopy of the immunostaining included an initial pre-screen at low power (4X) to identify regions with a technically optimal staining result. Subsequently, detailed analysis at high-power (40X) was performed to evaluate the staining according to routine algorithms employed in tumor diagnostics. Scoring of *FREMI* staining were evaluated using the semi-quantitative immunostaining score (ISS) method by pathologist. The immunostaining score was defined as 0 – 3 (range: +++/3, high; ++/2, moderate; +/1, weak; 0, negative.) [40]. The median score was used as cut off for classification of patients into high- and low-expression groups. Semi-quantitative analysis of the IHC images was conducted by Image-J,

by which integral optical density (IOD) and the area were collected. Then, average optical density (AOD) was calculated as IOD/area, which represented the staining intensity [41].

### Western blot

Total proteins were extracted from breast cancer and adjacent cancer tissue (SYSUCC, Sun Yat-sen University Cancer Center, Guangzhou, China) by using RIPA buffer (Beyotime, #P0013B) and passed the tissue through a 21-gauge needle more than 30 times. The cell lysate was centrifuged at 4 °C and 17000 g for 15 minutes to eliminate large aggregates. A BCA Protein Assay Kit (Tiangen, #PA115) was used to evaluate the cell lysate protein concentrations. Approximately 25-50 µg total protein, denatured with the standard SDS-sample buffer per lane, was loaded for electrophoresis on 10% pre-cast mini-polyacrylamide gels (GenScript SurePAGE, Bis-Tris, 10 cm x 8 cm gels). The gel containing proteins was then transferred to a PVDF membrane (BioRad). The membrane was blocked with 5% milk in 1×TBST for 2 hours, and incubated with the FREM1 primary antibody (1:1000 rabbit polyclonal, Proteintech, #13086-1-AP) and *TILRR* (1:2000 rabbit polyclonal custom, Genescript) at 4 °C overnight. Before incubating the membrane with a secondary antibody (1:10000 dilution, Jackson ImmunoResearch, #111-035-003) for 1 hour, it was washed with 1×PBST 5 times at room temperature. The blots were detected by chemiluminescence (Bio-Rad). The band intensity was quantified by ImageJ (<https://imagej.nih.gov/ij/download.html>).

### Cell culture

Cells were cultured (BT474, purchased from Procell Life Science and Technology Co., Ltd. CL-0040) at the density of  $1 \times 10^6$ /well in RPMI 1640 Medium (Gibco, #C11875500BT) with 20% FBS and insulin. Cells were incubated at 37 °C and 5% CO<sub>2</sub>. The day after seeding, cells were transfected with HA-GFP or HA-*TILRR*-T2A-GFP plasmid using the Lipofectamine 3000 Transfection Reagent (Invitrogen, #L3000-015). Cells were incubated at 37 °C and 5% CO<sub>2</sub> for 24 hours, at which time culture medium was replaced. Forty-eight hours after transfection, cells were detached by 0.25% trypsin for 3 minutes, and then collect the cell for cell sorting (Flow Cytometer, Aria II, BD).

### RNA-seq

Trizol lysate was used to lyse selective cells and extract RNA from the cells. To generate sequencing libraries, 1 µg RNA per sample was used (NEBNext Ultra™ RNA Library Prep Kit, Illumina; NE, USA). Library quality

was assessed on the Agilent Bioanalyzer 2100 system. Sequencing was performed on the Illumina Novaseq platform, and 150 bp paired-end reads were generated. Raw reads were aligned to a reference genome (UCSC GRCh38/hg38) using Bowtie 2 (2.2.5). Gene expression was quantified using RSEM v1.1.22.

### Statistical analysis

OS and DFS curves were generated by Kaplan-Meier survival analysis using SPSS 17.0 software. The results generated in OncoPrint are displayed with *P* values, fold changes, and ranks. The results of the Kaplan-Meier plots, PrognScan, and GEPIA2 are displayed with HR and *P* or Cox *P* values from a log-rank test. The correlation of gene expression was evaluated by Spearman's correlation and statistical significance, and the strength of the correlation was determined using the following guide for the absolute value: 0.00–0.19, very weak; 0.20–0.39, weak; 0.40–0.59, moderate; 0.60–0.79, strong; 0.80–1.0, very strong. *P* values < 0.05 were considered as statistically significant.

### AUTHOR CONTRIBUTIONS

Wen-Jing Guo, Ren Wang and Xiao Zhang conceived and planned the experiments, and wrote the manuscript; Xiao-Yi Xu, Wen-Jing Guo, Shi-Hua Pan, Sheng Zhang and He-Ying Li carried out the experiments; Ying Zhang, Feng-Lin Gao and Ren Wang carried out the bioinformatics analysis; Xiao-Yi Xu, Jiang-Tao Wang and Xiao Zhang performed the data analysis.

### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

### FUNDING

This work was supported by funds from the following grants: Engineering Project of Research and Development of Automatic Stem Cell Production Equipment (Y9284101, Xiao Zhang); Research and Development of Automatic Mesenchymal Stem Cell Production Equipment (Y9902001, Xiao Zhang); High-Throughput of Synthetic Biology Platform Development Project (O0392001, Xiao Zhang).

### REFERENCES

1. Beck TF, Shchelochkov OA, Yu Z, Kim BJ, Hernández-García A, Zaveri HP, Bishop C, Overbeek PA, Stockton DW, Justice MJ, Scott DA. Novel *frem1*-related mouse phenotypes and evidence of genetic interactions with

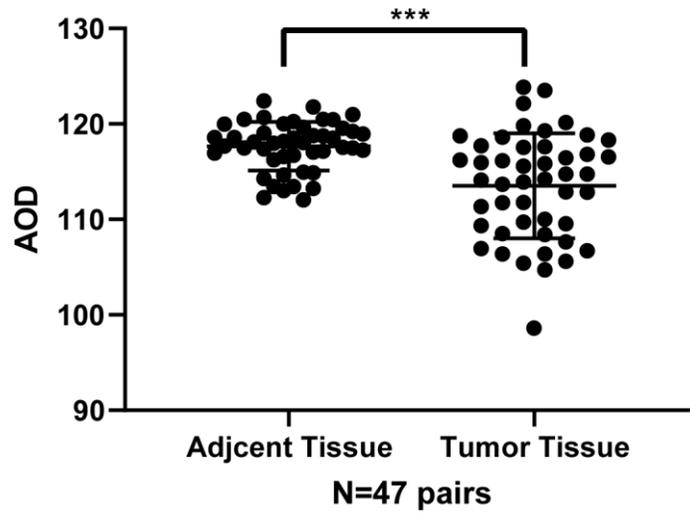
- gata4 and slit3. *PLoS One*. 2013; 8:e58830.  
<https://doi.org/10.1371/journal.pone.0058830>  
PMID:[23536828](https://pubmed.ncbi.nlm.nih.gov/23536828/)
2. Hudson RC, Gray C, Kiss-Toth E, Chico TJ, Qvarnstrom EE. Bioinformatics analysis of the FREM1 gene-evolutionary development of the IL-1R1 co-receptor, TILRR. *Biology (Basel)*. 2012; 1:484–94.  
<https://doi.org/10.3390/biology1030484>  
PMID:[24832504](https://pubmed.ncbi.nlm.nih.gov/24832504/)
  3. Zhang X, Shephard F, Kim HB, Palmer IR, McHarg S, Fowler GJ, O'Neill LA, Kiss-Toth E, Qvarnstrom EE. TILRR, a novel IL-1RI co-receptor, potentiates MyD88 recruitment to control ras-dependent amplification of NF-kappaB. *J Biol Chem*. 2010; 285:7222–32.  
<https://doi.org/10.1074/jbc.M109.073429>  
PMID:[19940113](https://pubmed.ncbi.nlm.nih.gov/19940113/)
  4. Smith SA, Samokhin AO, Alfadi M, Murphy EC, Rhodes D, Holcombe WM, Kiss-Toth E, Storey RF, Yee SP, Francis SE, Qvarnstrom EE. The IL-1RI co-receptor TILRR (FREM1 isoform 2) controls aberrant inflammatory responses and development of vascular disease. *JACC Basic Transl Sci*. 2017; 2:398–414.  
<https://doi.org/10.1016/j.jacbts.2017.03.014>  
PMID:[28920098](https://pubmed.ncbi.nlm.nih.gov/28920098/)
  5. Jung H, Kim HS, Kim JY, Sun JM, Ahn JS, Ahn MJ, Park K, Esteller M, Lee SH, Choi JK. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun*. 2019; 10:4278.  
<https://doi.org/10.1038/s41467-019-12159-9>  
PMID:[31537801](https://pubmed.ncbi.nlm.nih.gov/31537801/)
  6. Ino Y, Yamazaki-Itoh R, Shimada K, Iwasaki M, Kosuge T, Kanai Y, Hiraoka N. Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. *Br J Cancer*. 2013; 108:914–23.  
<https://doi.org/10.1038/bjc.2013.32>  
PMID:[23385730](https://pubmed.ncbi.nlm.nih.gov/23385730/)
  7. Zhang SC, Hu ZQ, Long JH, Zhu GM, Wang Y, Jia Y, Zhou J, Ouyang Y, Zeng Z. Clinical implications of tumor-infiltrating immune cells in breast cancer. *J Cancer*. 2019; 10:6175–84.  
<https://doi.org/10.7150/jca.35901>  
PMID:[31762828](https://pubmed.ncbi.nlm.nih.gov/31762828/)
  8. Noshio K, Baba Y, Tanaka N, Shima K, Hayashi M, Meyerhardt JA, Giovannucci E, Dranoff G, Fuchs CS, Ogino S. Tumour-infiltrating t-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *J Pathol*. 2010; 222:350–66.  
<https://doi.org/10.1002/path.2774>  
PMID:[20927778](https://pubmed.ncbi.nlm.nih.gov/20927778/)
  9. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, Budczies J, Huober J, Klauschen F, Furlanetto J, Schmitt WD, Blohmer JU, Karn T, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol*. 2018; 19:40–50.  
[https://doi.org/10.1016/S1470-2045\(17\)30904-X](https://doi.org/10.1016/S1470-2045(17)30904-X)  
PMID:[29233559](https://pubmed.ncbi.nlm.nih.gov/29233559/)
  10. Cai Y, Nogales-Cadenas R, Zhang Q, Lin JR, Zhang W, O'Brien K, Montagna C, Zhang ZD. Transcriptomic dynamics of breast cancer progression in the MMTV-PyMT mouse model. *BMC Genomics*. 2017; 18:185.  
<https://doi.org/10.1186/s12864-017-3563-3>  
PMID:[28212608](https://pubmed.ncbi.nlm.nih.gov/28212608/)
  11. Dagenais M, Dupaul-Chicoine J, Douglas T, Champagne C, Morizot A, Saleh M. The interleukin (IL)-1R1 pathway is a critical negative regulator of PyMT-mediated mammary tumorigenesis and pulmonary metastasis. *Oncoimmunology*. 2017; 6:e1287247.  
<https://doi.org/10.1080/2162402X.2017.1287247>  
PMID:[28405519](https://pubmed.ncbi.nlm.nih.gov/28405519/)
  12. Camargo MC, Mera R, Correa P, Peek RM Jr, Fontham ET, Goodman KJ, Piazuelo MB, Sicinski L, Zabaleta J, Schneider BG. Interleukin-1beta and interleukin-1 receptor antagonist gene polymorphisms and gastric cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006; 15:1674–87.  
<https://doi.org/10.1158/1055-9965.EPI-06-0189>  
PMID:[16985030](https://pubmed.ncbi.nlm.nih.gov/16985030/)
  13. Gong Z, Ma J, Su H, Guo T, Cai H, Chen Q, Zhao X, Qi J, Du J. Interleukin-1 receptor antagonist inhibits angiogenesis in gastric cancer. *Int J Clin Oncol*. 2018; 23:659–70.  
<https://doi.org/10.1007/s10147-018-1242-2>  
PMID:[29344744](https://pubmed.ncbi.nlm.nih.gov/29344744/)
  14. Sehoul J, Mustea A, Koensgen D, Chen FC, Lichtenegger W. Interleukin-1 receptor antagonist gene polymorphism is associated with increased risk of epithelial ovarian cancer. *Ann Oncol*. 2003; 14:1501–04.  
<https://doi.org/10.1093/annonc/mdg413>  
PMID:[14504049](https://pubmed.ncbi.nlm.nih.gov/14504049/)
  15. Lewis AM, Varghese S, Xu H, Alexander HR. Interleukin-1 and cancer progression: the emerging role of interleukin-1 receptor antagonist as a novel therapeutic agent in cancer treatment. *J Transl Med*. 2006; 4:48.  
<https://doi.org/10.1186/1479-5876-4-48>  
PMID:[17096856](https://pubmed.ncbi.nlm.nih.gov/17096856/)
  16. Vallés S, Tsoi C, Huang WY, Wyllie D, Carlotti F, Askari JA, Humphries MJ, Dower SK, Qvarnström EE. Recruitment of a heparan sulfate subunit to the

- interleukin-1 receptor complex. Regulation by fibronectin attachment. *J Biol Chem.* 1999; 274:20103–09.  
<https://doi.org/10.1074/jbc.274.29.20103>  
 PMID:10400621
17. Guo B, Fu S, Zhang J, Liu B, Li Z. Targeting inflammasome/IL-1 pathways for cancer immunotherapy. *Sci Rep.* 2016; 6:36107.  
<https://doi.org/10.1038/srep36107>  
 PMID:27786298
  18. Zhang X, Pino GM, Shephard F, Kiss-Toth E, Qwarnstrom EE. Distinct control of MyD88 adapter-dependent and Akt kinase-regulated responses by the interleukin (IL)-1RI co-receptor, TILRR. *J Biol Chem.* 2012; 287:12348–52.  
<https://doi.org/10.1074/jbc.C111.321711>  
 PMID:22262840
  19. Gallagher EJ, Zelenko Z, Neel BA, Antoniou IM, Rajan L, Kase N, LeRoith D. Elevated tumor LDLR expression accelerates LDL cholesterol-mediated breast cancer growth in mouse models of hyperlipidemia. *Oncogene.* 2017; 36:6462–71.  
<https://doi.org/10.1038/onc.2017.247> PMID:28759039
  20. Peng D, Kryczek I, Nagarsheth N, Zhao L, Wei S, Wang W, Sun Y, Zhao E, Vatan L, Szeliga W, Kotarski J, Tarkowski R, Dou Y, et al. Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature.* 2015; 527:249–53.  
<https://doi.org/10.1038/nature15520> PMID:26503055
  21. Barreira da Silva R, Laird ME, Yatim N, Fiette L, Ingersoll MA, Albert ML. Dipeptidylpeptidase 4 inhibition enhances lymphocyte trafficking, improving both naturally occurring tumor immunity and immunotherapy. *Nat Immunol.* 2015; 16:850–58.  
<https://doi.org/10.1038/ni.3201> PMID:26075911
  22. Kanda N, Shimizu T, Tada Y, Watanabe S. IL-18 enhances IFN-gamma-induced production of CXCL9, CXCL10, and CXCL11 in human keratinocytes. *Eur J Immunol.* 2007; 37:338–50.  
<https://doi.org/10.1002/eji.200636420>  
 PMID:17274000
  23. Buchbinder EI, Desai A. CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition. *Am J Clin Oncol.* 2016; 39:98–106.  
<https://doi.org/10.1097/COC.0000000000000239>  
 PMID:26558876
  24. Hassel JC, Heinzerling L, Aberle J, Bähr O, Eigentler TK, Grimm MO, Grünwald V, Leipe J, Reinmuth N, Tietze JK, Trojan J, Zimmer L, Gutzmer R. Combined immune checkpoint blockade (anti-PD-1/anti-CTLA-4): evaluation and management of adverse drug reactions. *Cancer Treat Rev.* 2017; 57:36–49.  
<https://doi.org/10.1016/j.ctrv.2017.05.003>  
 PMID:28550712
  25. Seidel JA, Otsuka A, Kabashima K. anti-PD-1 and anti-CTLA-4 therapies in cancer: mechanisms of action, efficacy, and limitations. *Front Oncol.* 2018; 8:86.  
<https://doi.org/10.3389/fonc.2018.00086>  
 PMID:29644214
  26. Qin S, Xu L, Yi M, Yu S, Wu K, Luo S. Novel immune checkpoint targets: moving beyond PD-1 and CTLA-4. *Mol Cancer.* 2019; 18:155.  
<https://doi.org/10.1186/s12943-019-1091-2>  
 PMID:31690319
  27. Tajiri K, Aonuma K, Sekine I. Cardiovascular toxic effects of targeted cancer therapy. *Jpn J Clin Oncol.* 2017; 47:779–85.  
<https://doi.org/10.1093/jjco/hyx071> PMID:28531278
  28. Beyer AM, Bonini MG, Moslehi J. Cancer therapy-induced cardiovascular toxicity: old/new problems and old drugs. *Am J Physiol Heart Circ Physiol.* 2019; 317:H164–67.  
<https://doi.org/10.1152/ajpheart.00277.2019>  
 PMID:31172808
  29. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia.* 2007; 9:166–80.  
<https://doi.org/10.1593/neo.07112>  
 PMID:17356713
  30. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
 PMID:29092952
  31. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016; 17:174.  
<https://doi.org/10.1186/s13059-016-1028-7>  
 PMID:27549193
  32. Goldman M, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo YH, Rogers D, Brooks AN, Zhu JC, Haussler D. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *BioRxiv.* 2018; 1–16.  
<https://doi.org/10.1101/326470>
  33. Sabatier R, Finetti P, Cervera N, Lambaudie E, Esterni B, Mamessier E, Tallet A, Chabannon C, Extra JM, Jacquemier J, Viens P, Birnbaum D, Bertucci F. A gene

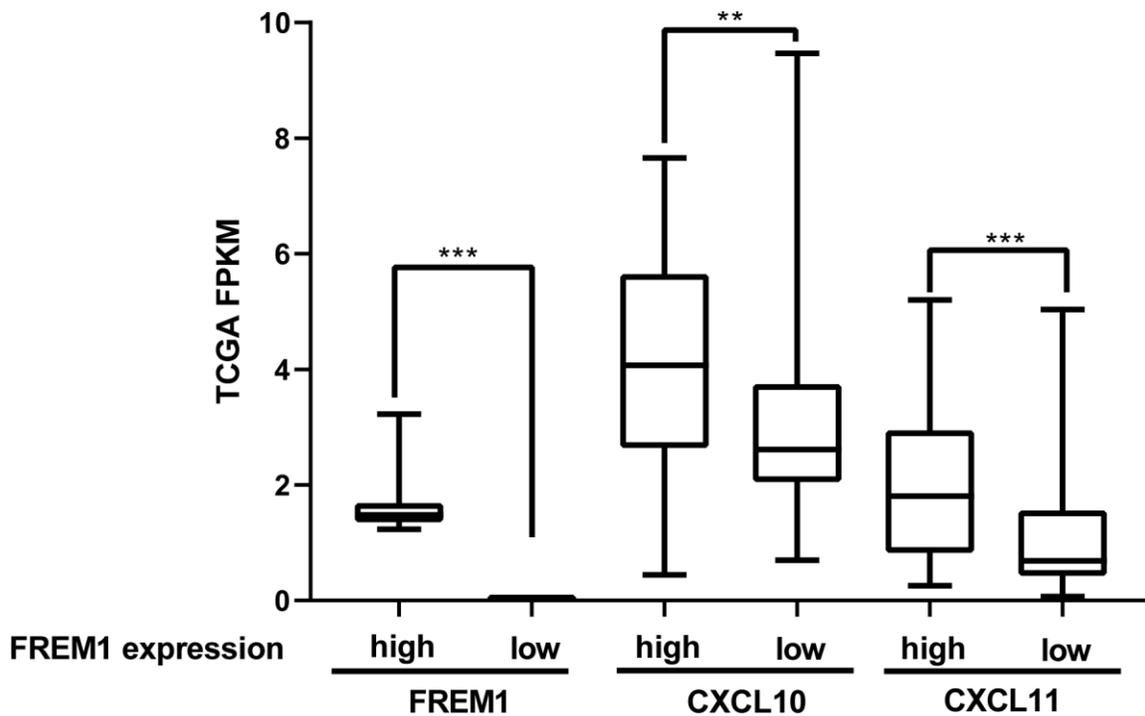
- expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res Treat.* 2011; 126:407–20.  
<https://doi.org/10.1007/s10549-010-0897-9>  
PMID:[20490655](https://pubmed.ncbi.nlm.nih.gov/20490655/)
34. Sabatier R, Finetti P, Adelaide J, Guille A, Borg JP, Chaffanet M, Lane L, Birnbaum D, Bertucci F. Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One.* 2011; 6:e27656.  
<https://doi.org/10.1371/journal.pone.0027656>  
PMID:[22110708](https://pubmed.ncbi.nlm.nih.gov/22110708/)
35. Cheng WY, Ou Yang TH, Anastassiou D. Development of a prognostic model for breast cancer survival in an open challenge environment. *Sci Transl Med.* 2013; 5:181ra50.  
<https://doi.org/10.1126/scitranslmed.3005974>  
PMID:[23596202](https://pubmed.ncbi.nlm.nih.gov/23596202/)
36. McCarthy N. Prognostic models: rising to the challenge. *Nat Rev Cancer.* 2013; 13:378.  
<https://doi.org/10.1038/nrc3530> PMID:[23640208](https://pubmed.ncbi.nlm.nih.gov/23640208/)
37. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019; 47:W556–60.  
<https://doi.org/10.1093/nar/gkz430> PMID:[31114875](https://pubmed.ncbi.nlm.nih.gov/31114875/)
38. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013; 4:2612.  
<https://doi.org/10.1038/ncomms3612>  
PMID:[24113773](https://pubmed.ncbi.nlm.nih.gov/24113773/)
39. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–50.  
<https://doi.org/10.1073/pnas.0506580102>  
PMID:[16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
40. Wang R, Yin C, Li XX, Yang XZ, Yang Y, Zhang MY, Wang HY, Zheng XF. Reduced SOD2 expression is associated with mortality of hepatocellular carcinoma patients in a mutant p53-dependent manner. *Aging (Albany NY).* 2016; 8:1184–200.  
<https://doi.org/10.18632/aging.100967>  
PMID:[27221200](https://pubmed.ncbi.nlm.nih.gov/27221200/)
41. Song L, Shen W, Zhang H, Wang Q, Wang Y, Zhou Z. Differential expression of androgen, estrogen, and progesterone receptors in benign prostatic hyperplasia. *Bosn J Basic Med Sci.* 2016; 16:201–08.  
<https://doi.org/10.17305/bjbm.2016.1209>  
PMID:[27294569](https://pubmed.ncbi.nlm.nih.gov/27294569/)

SUPPLEMENTARY MATERIALS

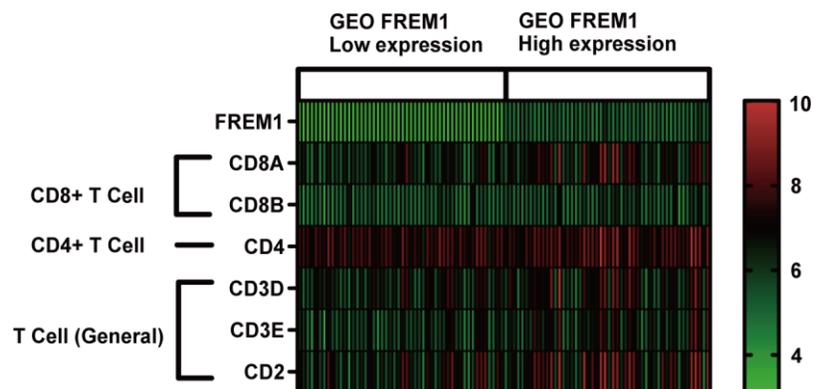
Supplementary Figures



Supplementary Figure 1. Semi-quantitative analysis of the IHC images was conducted by Image-J, and the integral optical density (IOD) and area were collected (N = 47 pairs).



Supplementary Figure 2. The top 50 highest *TILRR*-expressing and top 50 lowest *TILRR*-expressing tumor tissue samples were used (TCGA database) to analyze the expression levels of *CXCL10* and *CXCL11*.



Supplementary Figure 3. The top 50 highest *TILRR*-expressing and top 50 lowest *TILRR*-expressing tumor tissue samples were selected (GEO database) to analyze the expression of marker genes of CD8+, CD4+ and T cells (general).

### Supplementary Table

Please browse Full Text version to see the data of Supplementary Table 1.

**Supplementary Table 1. Correlation analysis between FREM1 and related genes and markers of immune cells in TIMER.**

# TRIM28 is a distinct prognostic biomarker that worsens the tumor immune microenvironment in lung adenocarcinoma

Jie Liu<sup>1</sup>, Xiao Han<sup>2</sup>, Lijuan Chen<sup>1</sup>, Dong Han<sup>3</sup>, Xiaoqian Mu<sup>1</sup>, Xiufeng Hu<sup>1</sup>, Hongbo Wu<sup>1</sup>, Huijuan Wu<sup>1</sup>, Wenjing Liu<sup>1</sup>, Yanqiu Zhao<sup>1</sup>

<sup>1</sup>Department of Internal Medicine, The Affiliated Tumor Hospital of Zhengzhou University, Zhengzhou, China

<sup>2</sup>Department of Obstetrics and Gynecology, The Third Affiliated Hospital of Zhengzhou University, Zhengzhou, China

<sup>3</sup>Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Correspondence to: Yanqiu Zhao; email: [zlyyliujie2791@zzu.edu.cn](mailto:zlyyliujie2791@zzu.edu.cn)

Keywords: TRIM28, tumor immune microenvironment, lung adenocarcinoma, prognosis

Received: April 20, 2020

Accepted: July 9, 2020

Published: October 22, 2020

Copyright: © 2020 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

The tumor immune microenvironment (TIME) is an important determinant of cancer prognosis and treatment efficacy. To identify immune-related prognostic biomarkers of lung adenocarcinoma, we used the ESTIMATE algorithm to calculate the immune and stromal scores of 517 lung adenocarcinoma patients from The Cancer Genome Atlas (TCGA). We detected 985 differentially expressed genes (DEGs) between patients with high and low immune and stromal scores, and we analyzed their functions and protein-protein interactions. *TRIM28* was upregulated in lung adenocarcinoma patients with low immune and stromal scores, and was associated with a poor prognosis. The TISIDB and TIMER databases indicated that *TRIM28* expression correlated negatively with immune infiltration. We then explored genes that were co-expressed with *TRIM28* in TCGA, and investigated DEGs based on *TRIM28* expression in GSE43580 and GSE7670. The 429 common DEGs from these analyses were functionally analyzed. We also performed a Gene Set Enrichment Analysis using TCGA data, and predicted substrates of *TRIM28* using UbiBrowser. The results indicated that *TRIM28* may negatively regulate the TIME by increasing the SUMOylation of IRF5 and IRF8. Correlation analyses and validations in two lung adenocarcinoma cell lines (PC9 and H1299) confirmed these findings. Thus, *TRIM28* may worsen the TIME and prognosis of lung adenocarcinoma.

## INTRODUCTION

Lung cancer is still the most common form of cancer and the leading cause of cancer death worldwide in both developing and developed regions [1]. Non-small-cell lung cancer (NSCLC) accounts for 85% of lung cancer cases, and can be classified as adenocarcinoma, squamous cell carcinoma or large cell carcinoma. Lung adenocarcinoma (LUAD) is one of the most common subtypes of NSCLC [2], and has been extensively studied in recent years due to the great success of molecular targeted therapy.

Immune evasion is acknowledged as a hallmark of tumors [3], and different immune cell types contribute to immune infiltration and immune evasion. Immunotherapies such as programmed cell death-1 (PD1) / programmed cell death ligand-1 (PD-L1) inhibitors have become standard-of-care treatment options for NSCLC patients. However, only a small subset (20-30%) of patients respond to such treatments [4-7]. At present, the prognosis of LUAD remains poor, and the overall five-year survival rate is < 15% due to local and distant recurrences [8].

Tumor-infiltrating lymphocytes and neutrophils are known to influence the prognosis of cancers and the

efficacy of antitumor therapies [9, 10]. The level of infiltrating stromal and immune cells in tumor samples can be predicted with the ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) algorithm, which calculates immune and stromal scores based on unique gene signatures. Two main gene signatures are used: one based on 141 stroma-related genes that reflect the presence of stroma in tumor tissues, and the other based on 141 immune-related genes that represent the infiltration of immune cells into tumor tissues [11]. The current knowledge about the link between the tumor immune microenvironment (TIME) and LUAD is insufficient. Therefore, there is an urgent need to better understand tumor-immune interactions and identify more precise prognostic predictors and molecular biomarkers for lung cancer.

Tripartite motif-containing (TRIM) proteins, which include a structurally conserved RING-finger domain, one or two B-box zinc finger domains and a coiled-coil domain, are considered to be significant regulators of carcinogenesis [12]. TRIM28 (also known as KAP1, TIF1 $\beta$  or KRIP1), one of the 60 members of the TRIM family, is a small ubiquitin-like modifier (SUMO) E3 ligase and a fundamental component of several macromolecular complexes [13–15]. TRIM28 is a poorly understood transcriptional co-factor with pleiotropic biological activities, including inducing gene silencing, promoting cellular proliferation and differentiation, promoting neoplastic transformation, inhibiting apoptosis, facilitating DNA repair, and guarding genomic integrity [16]. TRIM28 also promotes T cell activation, T cell tolerance, and the expression of various interleukins and other pro-inflammatory molecules [17–25]. The upregulation of TRIM28 predicts a poor prognosis in patients with gastric cancer [26], ovarian cancer [27], breast cancer [12] and colorectal cancer [28]. Lei et al. [29] found that the upregulation of TRIM28 promoted the growth of NSCLC and was a potential predictor of metastasis and prognosis in early-stage NSCLC patients. However, Chen et al. [30] reported that TRIM28 exerted anti-proliferative activity in lung cancer by repressing E2F family members that are critical for cell proliferation. Due to these contradictory observations, the prognostic value of TRIM28 in lung cancer remains unclear.

In this study, we conducted a comprehensive analysis of immune cell infiltration and gene expression in the TIME of LUAD based on the ESTIMATE algorithm, and then correlated these data with clinical and prognostic features. The results revealed the significant prognostic value of *TRIM28* expression and a potential mechanism whereby TRIM28 alters the TIME in LUAD.

## RESULTS

### The correlations among the immune and stromal scores, clinical features and prognoses of LUAD patients

The overall flowchart of this study is shown in Figure 1. In total, 517 LUAD patients with RNA sequencing data and clinical information in The Cancer Genome Atlas (TCGA) database were included (<http://www.cbioportal.org>, Firehose Legacy, Supplementary Table 1) [31]. Patients' immune and stromal scores were determined using the ESTIMATE algorithm based on gene expression data [11]. The detailed results are presented in Supplementary Table 2.

After comprehensively analyzing the stromal and immune scores, clinical information and RNA sequencing data, we found that both the stromal and immune scores were significantly lower in men ( $p = 0.009$ ,  $p = 0.005$ , respectively; Figure 2A and 2B), in patients with higher *TRIM28* expression ( $p < 0.001$ ,  $p < 0.001$ ; Figure 2C and 2D) and in patients with metastasis ( $p = 0.007$ ,  $p = 0.035$ ; Figure 2K and 2L) than in their respective counterpart groups. In addition, the immune scores were lower in patients in higher tumor-node-metastasis (TNM) stages and T stages ( $p = 0.036$ ,  $p = 0.005$ ; Figure 2F and 2H). However, the stromal scores did not correlate with the TNM stages or T stages ( $p = 0.107$ ,  $p = 0.286$ ; Figure 2E and 2G). The stromal and immune scores also did not correlate significantly with the lymph node metastasis status ( $p = 0.746$ ,  $p = 0.439$ ; Figure 2I and 2J).

Patients were then divided into two groups according to the median value of the stromal score or the immune score. Then, overall survival (OS) (Figure 2M and 2N) and disease-free survival (DFS) (Figure 2O and 2P) were compared between the respective groups. OS was worse in patients with lower stromal scores or immune scores than in those with higher scores ( $p = 0.059$ ,  $p = 0.009$ ). However, DFS did not differ significantly between the respective groups.

### Identification and functional annotation of differentially expressed genes

Next, we examined the differentially expressed genes (DEGs) between patients with high and low stromal scores, as shown in the heatmap in Figure 3A. We found that 1,401 genes were upregulated in the group with high stromal scores, while 448 genes were upregulated in the group with low stromal scores. We also evaluated the DEGs between patients with high and low immune scores (Figure 3B), and found that 1,278 genes were upregulated in the group with high immune scores, while 278 genes were upregulated in the group with low immune scores.

Then, using an online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>), we generated Venn diagrams to identify overlapping DEGs (Figure 3C and 3D). The results indicated that 872 genes were commonly upregulated in the groups with high stromal scores and high immune scores, while 113 genes were commonly upregulated in the groups with low stromal scores and low immune scores.

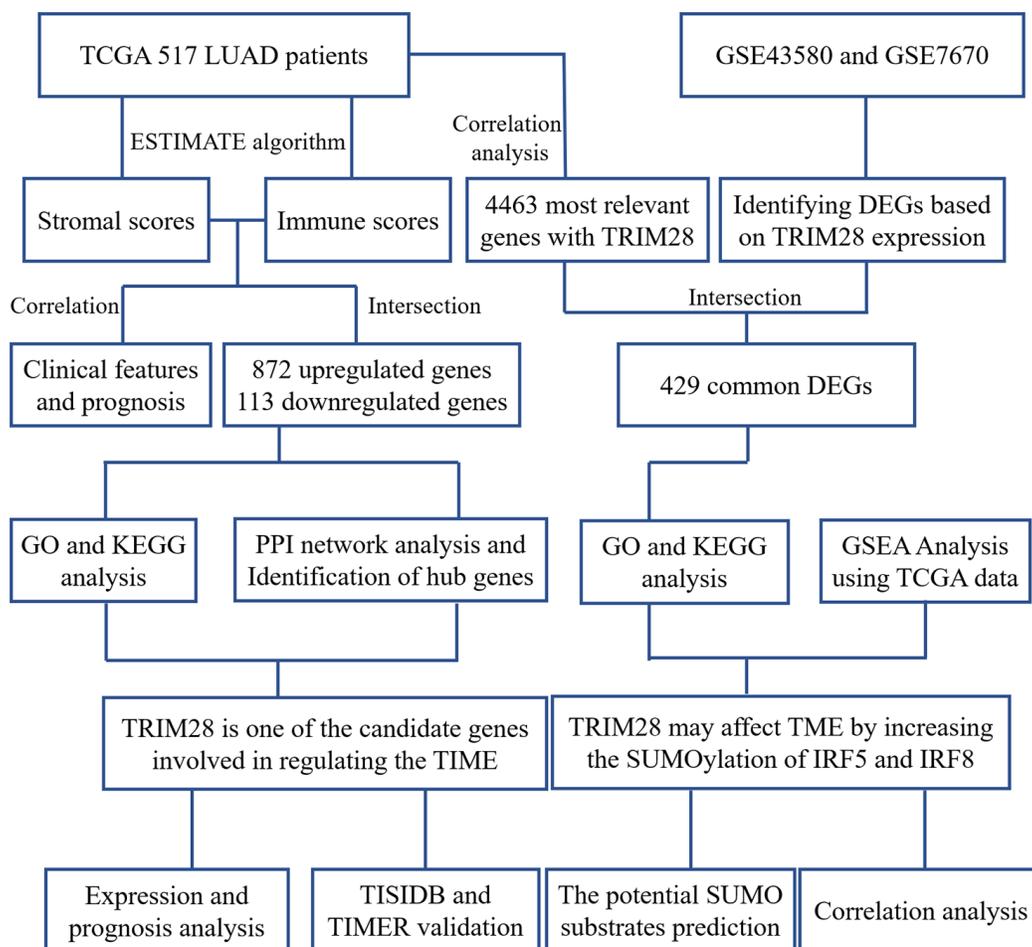
To explore the biological functions of the 985 DEGs, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses using Metascape [32]. The top 15 GO enrichment terms and KEGG enrichment terms are shown in Figure 3F and 3G, respectively. Most of the terms were related to immune regulation, including lymphocyte activation, regulation of cytokine production, interferon production, etc.

Considering the poor prognoses of patients with low stromal or immune scores, we then performed a protein-

protein interaction (PPI) analysis on the 113 genes that were commonly upregulated in patients with low stromal and immune scores (Figure 3E). The top 15 genes identified using the maximal clique centrality method were chosen as hub genes through the cytoHubba plugin: *ARHGEF16*, *ESRP1*, *TRIM28*, *RBBP8NL*, *CLDN7*, *RAB25*, *APIM2*, *EPCAM*, *LLGL2*, *KDF1*, *OVOL2*, *FAM83H*, *EPN3*, *CAMSAP3* and *SHMT1* (Figure 3H). Interestingly, other methods in the cytoHubba plugin also identified *TRIM28* as a crucial hub gene (Supplementary Table 3).

### The mRNA and protein levels of TRIM28 across cancer types

To determine whether *TRIM28* expression differed between tumor tissues and healthy tissues, we used the Oncomine database to analyze *TRIM28* mRNA levels in multiple cancer types. *TRIM28* expression was higher in bladder cancer, colorectal cancer, gastric cancer, head



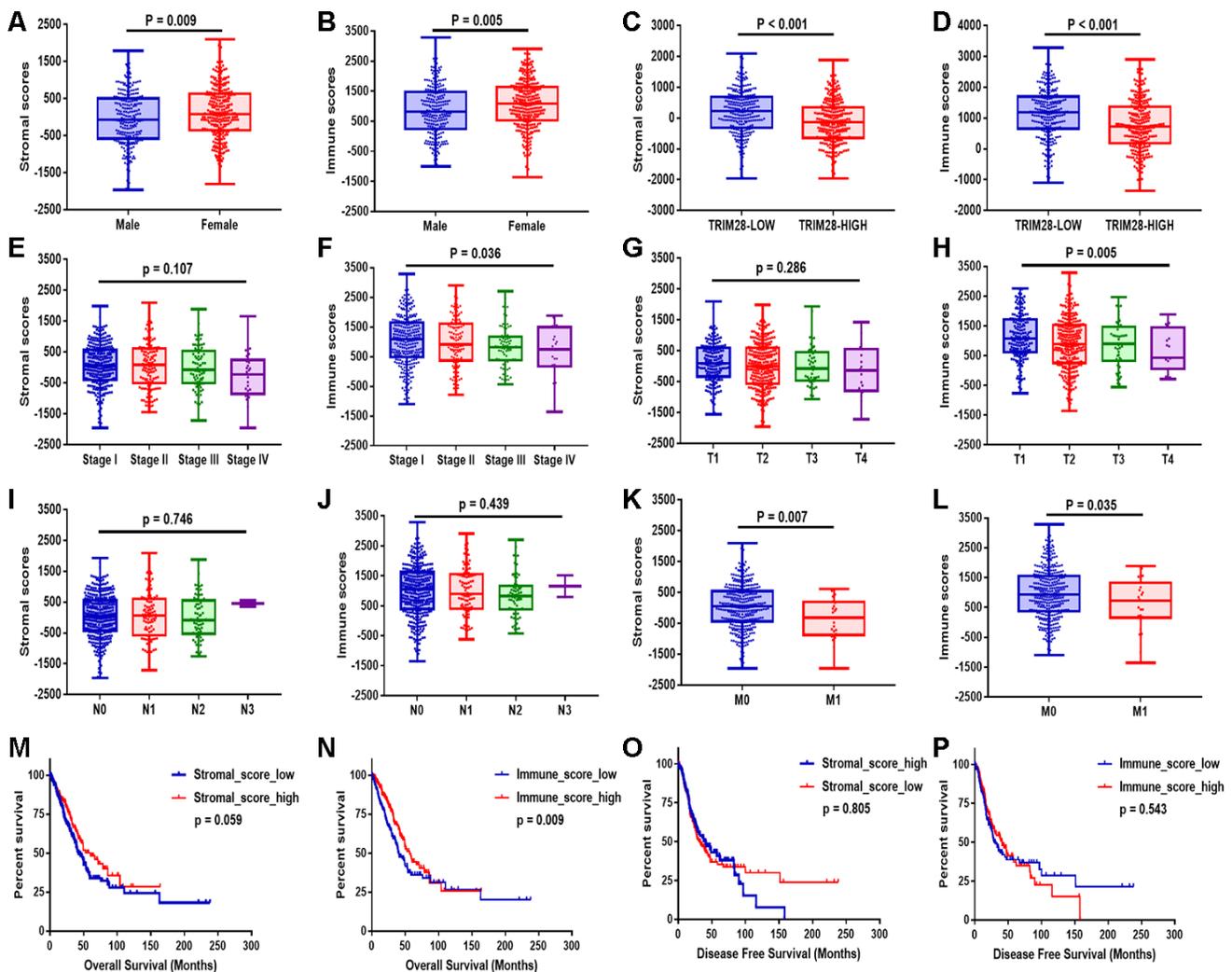
**Figure 1. Workflow of the present study.** TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; PPI, protein-protein interaction; GSE, Gene Expression Omnibus data series; DEGs, differentially expressed genes; TIME, tumor immune microenvironment.

and neck cancer, liver cancer, lung cancer and multiple myeloma than in healthy tissues (Figure 4A). We also examined *TRIM28* RNA levels in various tumor tissues and adjacent healthy tissues using the RNA sequencing data in TCGA (Figure 4B). *TRIM28* was significantly upregulated in most of the tumor tissues, including bladder urothelial carcinoma, breast invasive carcinoma, LUAD, lung squamous cell carcinoma, etc. However, *TRIM28* RNA levels were significantly lower in kidney renal papillary cell carcinoma than in adjacent healthy tissues. The significant increase in *TRIM28* expression in LUAD was further validated in four independent data sets, including GSE32863 [33], GSE7670 [34], GSE19188 [35] and the Beer Lund dataset [36] (Figure 4C–4F).

We also used the UALCAN cancer database to examine *TRIM28* protein levels in various cancer tissues. *TRIM28* protein expression was upregulated in breast cancer, colon cancer, ovarian cancer, clear cell renal cell carcinoma, uterine corpus endometrial carcinoma and LUAD (Figure 4G).

### Prognostic value of *TRIM28* across cancer types

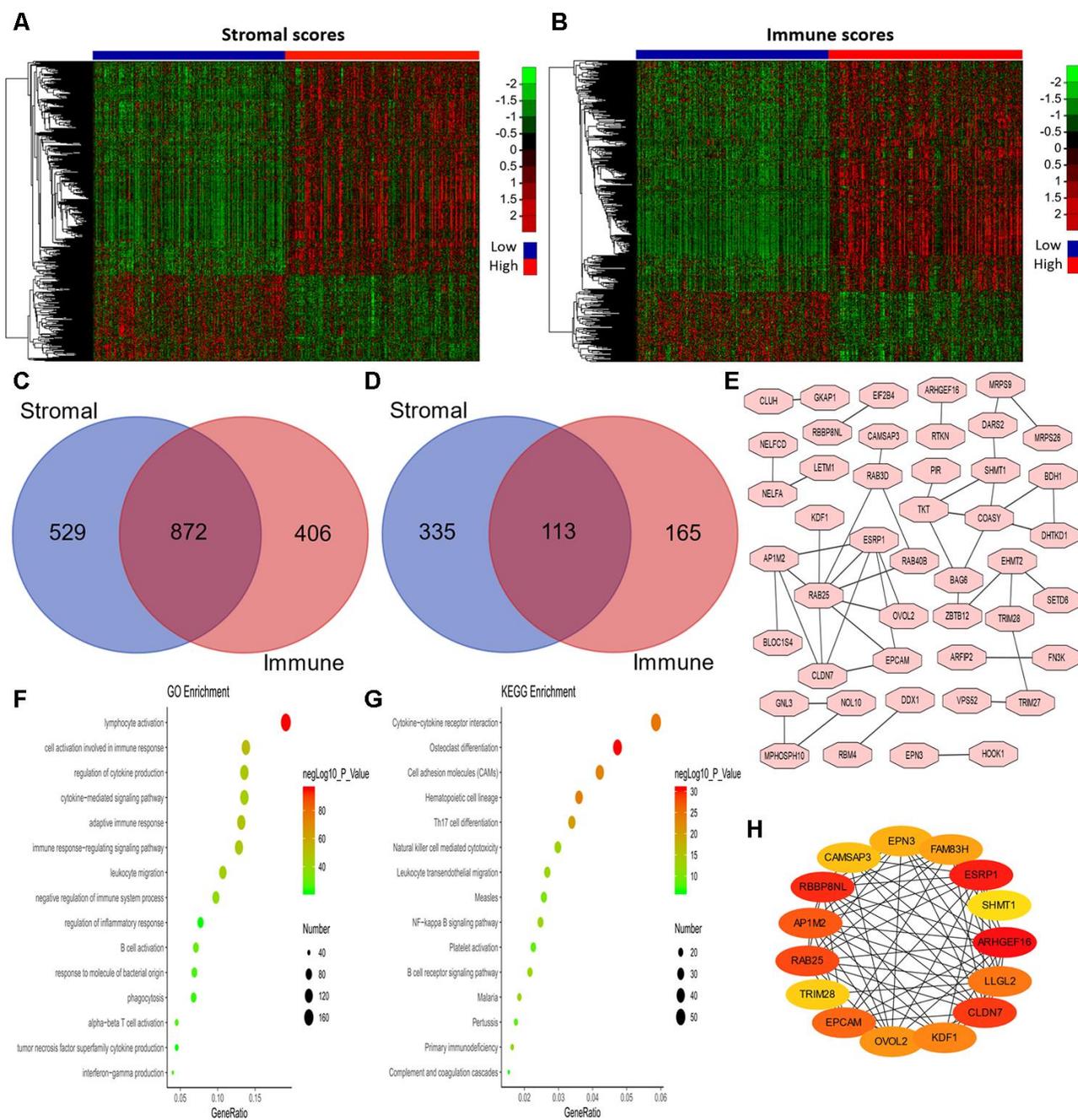
We then used the PrognoScan database to investigate whether *TRIM28* expression correlated with the prognosis of cancer patients. Notably, *TRIM28* expression significantly impacted the prognosis of seven cancer types, including breast, lung, ovarian, brain, skin, prostate and blood cancers (Figure 5A–5L).



**Figure 2. Stromal and immune scores were associated with the clinical characteristics and OS of LUAD patients.** (A–L) The stromal and immune score distributions between patients with different genders (A, B), TRIM28 levels (C, D), TNM stages (E, F), T classifications (G, H), lymph node metastasis statuses (I, J) and distant metastasis statuses (K, L). (M–P) Patients were then divided into two groups according to the median stromal score or immune score. OS (M, N) and DFS (O, P) analyses were performed between the respective groups. OS, overall survival; LUAD, lung adenocarcinoma.

In three cohorts (GSE4922-GPL96, GSE3494-GPL96 and GSE7378) [37–39] that respectively included 249 samples, 236 samples and 54 cases at different stages of breast cancer, higher *TRIM28* expression was

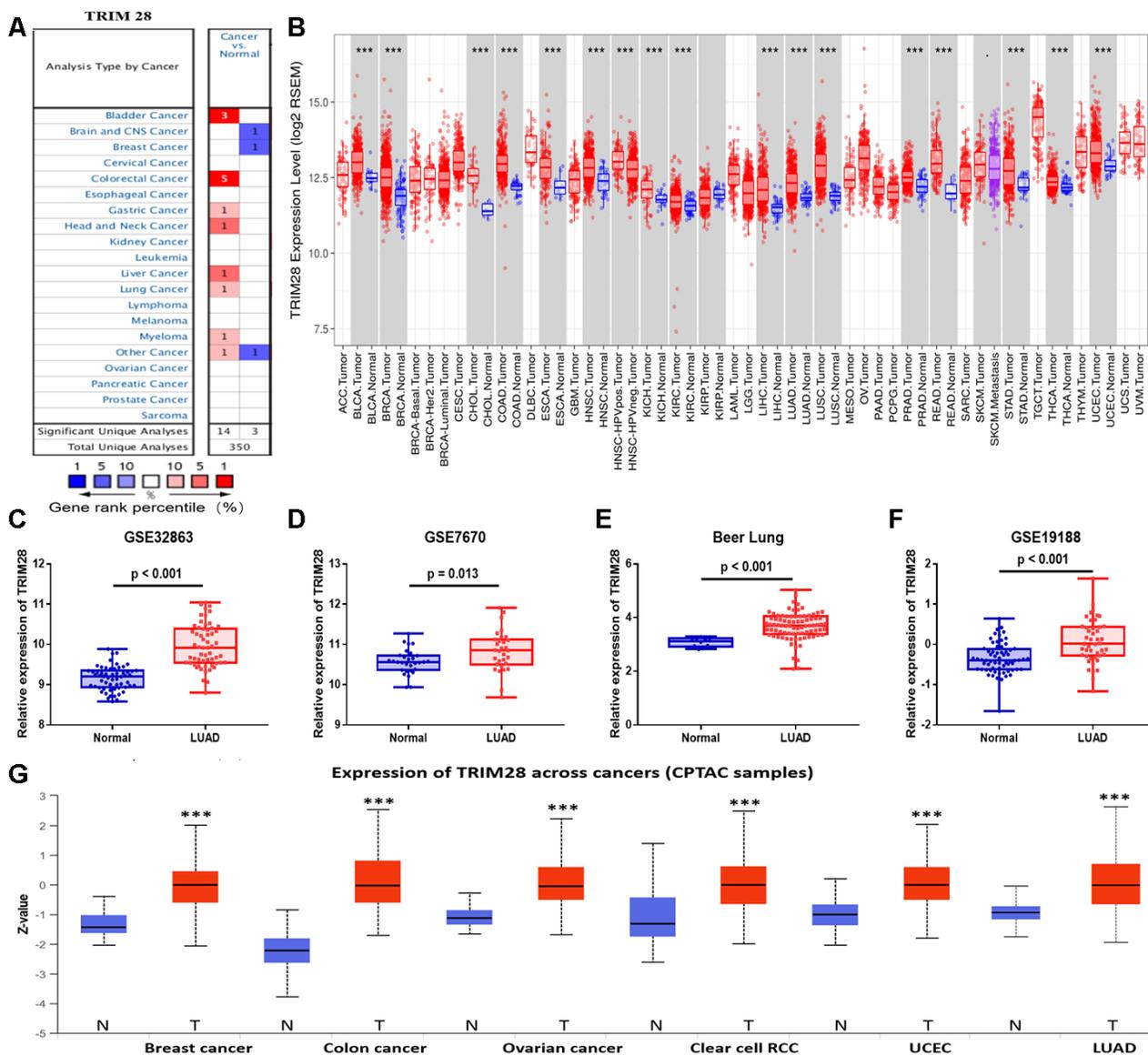
marginally associated with poorer DFS or disease-specific survival (DSS) (DFS hazard ratio [HR] = 3.62, 95% confidence interval [CI] = 2.03 to 6.44, Cox  $p < 0.001$ ; DSS HR = 3.76, 95% CI = 1.77 to 7.98, Cox



**Figure 3. Analysis of DEGs according to the immune and stromal scores in LUAD patients.** (A, B) Clustering Heatmap plot of the DEGs. The upper color bar represents the sample class; red represents the group with higher scores, while blue represents the group with lower scores. Genes with higher levels are shown in red, while those with lower levels are shown in green. (C, D) Venn diagrams showing the number of commonly upregulated (C) or downregulated (D) DEGs. (E) PPI analysis of downregulated DEGs via STRING. The interaction score was set to medium confidence (0.400). (F, G) The top 15 GO enrichment terms (F) and KEGG enrichment terms (G) for all DEGs, analyzed in Metascape. (H) The first 15 genes identified through the maximal clique centrality method were chosen as hub genes using the cytoHubba plugin. More red color represents more forward ranking. GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes; LUAD, lung adenocarcinoma; MMC, maximal clique centrality.

$p < 0.001$ ; DFS HR = 104.71, 95% CI = 6.95 to 1577.65, Cox  $p < 0.001$ ; Respectively; Figure 5A–5C). However, in two other cohorts (GSE9893 and GSE11121) [40, 41] that respectively included 155 and 200 samples at different stages of breast cancer, lower *TRIM28* expression was associated with poorer OS or distant metastasis-free survival (DMFS) (OS HR = 0.80, 95% CI = 0.68 to 0.95, Cox  $p = 0.008$ ; DMFS HR = 0.34, 95% CI = 0.13 to 0.89, Cox  $p = 0.028$ ; Respectively; Figure 5D, 5E). In one cohort (GSE31210) [42] that included 204 samples at different stages of

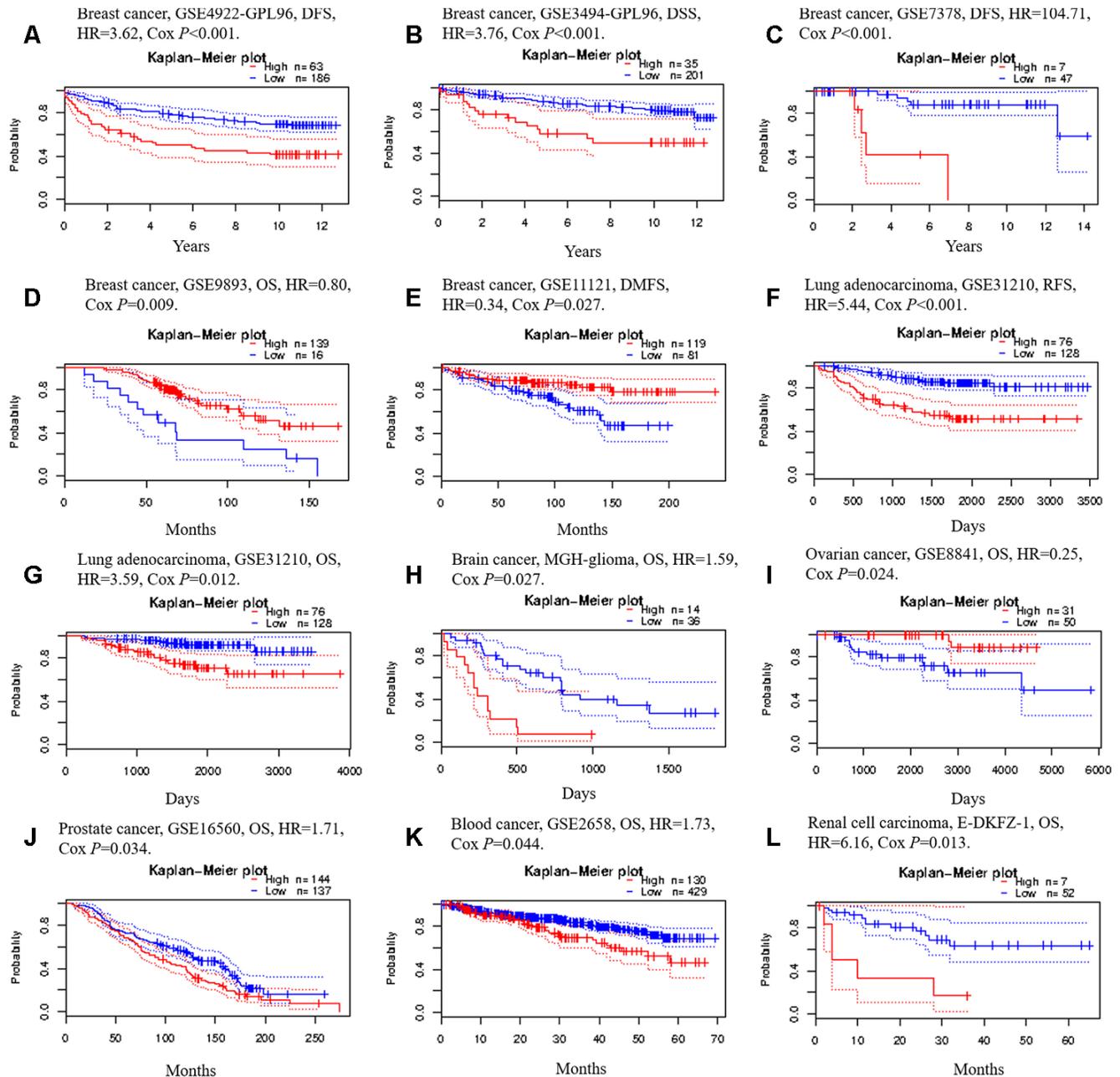
LUAD, higher *TRIM28* expression was marginally associated with poorer recurrence-free survival (RFS) and OS (RFS HR = 5.44, 95% CI = 2.73 to 10.87, Cox  $p < 0.001$ ; OS HR = 3.59, 95% CI = 1.33 to 9.68, Cox  $p = 0.012$ ; Figure 5F and 5G). We also observed the poor prognostic value of *TRIM28* in brain cancer, prostate cancer, blood cancer and renal cell carcinoma (Figure 5H, 5J–5L) and its good prognostic value in ovarian cancer (Figure 5I). These results suggested that *TRIM28* expression influences the prognosis of LUAD and other tumor types.



**Figure 4. TRIM28 levels in different human cancer types.** (A) Increased or reduced TRIM28 levels in different cancers compared with normal tissues in the Oncomine database. (B) Human TRIM28 levels in different tumor types from TCGA were determined using TIMER. (C–F) The significant increase in TRIM28 expression in LUAD was further validated in GSE32863 (C), GSE7670 (D), the Beer Lung dataset (E) and GSE19188 (F). (G) The protein expression of TRIM28 in various cancer tissues was detected using the UALCAN cancer database ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ). TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; GSE, Gene Expression Omnibus data series.

To further examine the prognostic potential of *TRIM28* in different cancers based on Affymetrix microarrays, we used the Kaplan-Meier plotter database. Higher *TRIM28* expression was associated with a poorer prognosis in lung cancer and gastric cancer. However, *TRIM28* expression had less of an impact on the

prognosis of ovarian cancer and breast cancer (Supplementary Figure 1). These results confirmed the significant prognostic value of *TRIM28* expression in lung cancer and gastric cancer. We also analyzed the prognostic potential of *TRIM28* in 33 different cancer types by using Gene Expression Profiling Interactive



**Figure 5. Kaplan-Meier survival curves generated from the PrognScan database for *TRIM28* expression in different tumor types. (A–E) DFS, DSS, DMFS and OS curves for five breast cancer cohorts (GSE4922-GPL96, GSE3494-GPL96, GSE7378, GSE9893 and GSE11121). (F, G) RFS and OS curves for lung cancer (GSE31210). (H) OS curve for brain cancer (MGH-glioma). (I) OS curve for ovarian cancer (GSE8841). (J) OS curve for prostate cancer (GSE16560). (K) OS curve for blood cancer (GSE2658). (L) OS curve for renal cell carcinoma. DFS, disease-free survival; DSS, disease-specific survival; DMFS, distant metastasis-free survival; OS, overall survival; RFS, recurrence-free survival; GSE, Gene Expression Omnibus data series; HR, hazard ratio.**

Analysis (GEPIA) to evaluate RNA sequencing data from TCGA. Higher *TRIM28* levels were associated with poorer OS in adrenocortical carcinoma, brain lower grade glioma, LUAD, mesothelioma, skin cutaneous melanoma, etc. (Supplementary Figure 2). Thus, although the prognostic value of *TRIM28* expression differed among different cancer types, the results from different databases all confirmed the prognostic value of *TRIM28* expression in LUAD.

Next, we explored the relationship between *TRIM28* expression and the clinical characteristics of lung cancer patients in the Kaplan-Meier plotter database. Over-expression of *TRIM28* was associated with worse OS and worse first progression (FP), regardless of gender and smoking history ( $p < 0.001$ ). Interestingly, when patients were analyzed according to the type of lung cancer, the upregulation of *TRIM28* was associated with worse OS and FP in LUAD (OS HR = 2.65,  $p < 0.001$ ; FP HR = 2.00,  $p < 0.001$ ), but was not associated with OS in lung squamous cell carcinoma (OS HR = 1.25,  $p = 0.095$ ). Moreover, higher *TRIM28* expression was associated with worse OS in stage 1, stage 2, stage N0 and stage M0, but was not associated with worse OS according to the grade, stage 3, stage T, stage N1 or stage N2 (Table 1). These results indicated that *TRIM28* expression had more significant prognostic value in LUAD patients than in lung squamous cell carcinoma patients, and had more significant prognostic value in early-stage than in late-stage LUAD patients.

### ***TRIM28* expression is associated with the immune infiltration level**

Numerous studies have demonstrated that the immune infiltrates in various human tumor types are associated with the prognosis and response to therapy [9, 10, 43]. We used the TISIDB and Tumor Immune Estimation Resource (TIMER) databases to assess whether *TRIM28* expression was associated with the level of immune infiltration across human tumors. *TRIM28* levels correlated negatively with the levels of 28 types of tumor-infiltrating lymphocytes across human tumors in the TISIDB database (Figure 6A). *TRIM28* levels also correlated negatively with the levels of central memory CD8+ T cells ( $R = -0.212$ ,  $p < 0.001$ ), macrophages ( $R = -0.353$ ,  $p < 0.001$ ), natural killer T cells ( $R = -0.313$ ;  $p < 0.001$ ), myeloid-derived suppressor cells ( $R = -0.3$ ,  $p < 0.001$ ), regulatory T cells ( $R = -0.323$ ,  $p < 0.001$ ) and neutrophils ( $R = -0.287$ ,  $p < 0.001$ ) (Figure 6B). We then analyzed the relationship between *TRIM28* expression and immune infiltration in 39 tumor types in the TIMER database. *TRIM28* levels had strong negative associations with the infiltrating levels of CD8+ T cells in 24 tumor types, CD4+ T cells in 14 tumor types, macrophages in 25 tumor types,

neutrophils in 27 tumor types and dendritic cells in 24 tumor types. In LUAD samples, *TRIM28* levels correlated negatively with the infiltrating levels of B cells ( $R = -0.141$ ,  $p = 1.89e-03$ ), CD8+ T cells ( $R = -0.234$ ,  $p = 1.67e-07$ ), macrophages ( $R = -0.277$ ,  $p = 5.75e-10$ ), neutrophils ( $R = -0.192$ ,  $p = 2.16e-05$ ) and dendritic cells ( $R = -0.296$ ,  $p = 2.67e-11$ ) (Figure 6C). These results strongly suggested that *TRIM28* inhibits immune infiltration in LUAD.

We then explored the effects of *TRIM28* levels, clinicopathological characteristics and immune infiltration levels on survival using a multivariate Cox proportional hazards model. We found that age ( $p = 0.028$ ), stage ( $p < 0.001$ ), infiltrating B cell levels ( $p = 0.014$ ) and *TRIM28* levels ( $p < 0.001$ ) were independent predictors of survival in LUAD. However, only age ( $p = 0.024$ ) and stage 3 ( $p = 0.011$ ) were independent predictors of survival in lung squamous cell carcinoma (Table 2).

### **Analysis of genes co-expressed with *TRIM28* in LUAD**

Next, we used LinkedOmics [44] to identify genes that were co-expressed with *TRIM28* based on mRNA sequencing data from LUAD patients in TCGA. We generated a volcano map of all the genes associated with *TRIM28*, and found that interferon regulatory factor 5 (*IRF5*) and *IRF8* levels correlated negatively with *TRIM28* levels (Figure 7A). We also downloaded two mRNA expression datasets (GSE43580 and GSE7670) from the Gene Expression Omnibus (GEO) [34, 45]. We divided the samples in each dataset into two groups according to *TRIM28* expression, and we analyzed the DEGs between patients with higher and lower *TRIM28* levels. The volcano graphs in Figures 7B and 7C display the DEGs in GSE43580 and GSE7670, respectively. We then used a Venn diagram to evaluate the overlapping DEGs from Figure 7A–7C (Figure 7D). There were 429 common DEGs, including *IRF5*, *IRF8*, *B2M*, *CD44*, *HLA-DRA*, *HLA-DRB1* and *HLA-E*.

We then performed GO and KEGG analyses of these DEGs using Metascape [32]. We identified the top 14 GO enrichment terms (Figure 7E), which included the regulation of cytokine production and the interferon-gamma signaling pathway. The top 13 KEGG enrichment terms (Figure 7F) included the Toll-like receptor signaling pathway and ubiquitin-induced proteolysis. The correlations among the top 10 enriched terms from the GO analysis are shown as a network in Figure 7G. We also performed a Gene Set Enrichment Analysis using TCGA data, and found that the interferon-gamma and Toll-like receptor signaling pathways were enriched (Figure 7H). Combined with

**Table 1. Correlation between TRIM28 mRNA expression and prognosis in lung cancer patients with different clinicopathological characteristics, determined via Kaplan-Meier plotter.**

Clinicopathological characteristics	OS (n = 1928)			FP (n = 646)		
	N	Hazard ratio	P-value	N	Hazard ratio	P-value
Sex						
Female	715	2.11(1.54-2.9)	<0.001	468	2.09(1.57-2.78)	<0.001
male	1387	1.48(1.25-1.75)	<0.001	514	1.87(1.44-2.44)	<0.001
Histology						
Adenocarcinoma	720	2.65(1.96-3.59)	<0.001	461	2.00(1.45-2.75)	<0.001
Squamous cell carcinoma	524	1.25(0.96-1.64)	0.095	141	2.15(1.27-3.64)	0.003
Grade						
I	201	0.81(0.56-1.16)	0.250	140	0.81(0.51-1.29)	0.380
II	310	1.34(0.96-1.87)	0.079	165	1.68(1.08-2.62)	0.020
III	77	0.64(0.28-1.47)	0.290	51	1.36(0.6-3.05)	0.460
Stage						
1	577	2.45(1.85-3.25)	<0.001	325	1.47(0.94-2.31)	0.091
2	244	2.2(1.45-3.35)	<0.001	130	0.79(0.46-1.35)	0.380
3	70	0.69(0.39-1.21)	0.200	19	-	-
4	4	-	-	0	-	-
Stage T						
1	437	1.37(0.97-1.92)	0.070	177	2.31(1.38-3.89)	0.001
2	589	1.26(0.99-1.6)	0.056	351	1.45(1.08-1.96)	0.014
3	81	0.8(0.47-1.34)	0.390	21	0.58(0.21-1.62)	0.290
4	46	1.46(0.72-2.96)	0.290	7	-	-
Stage N						
0	781	1.31(1.05-1.64)	0.015	374	1.83(1.3-2.56)	<0.001
1	252	1.29(0.94-1.76)	0.110	130	1.96(1.24-3.1)	0.004
2	111	1.18(0.78-1.78)	0.430	51	1.82(0.91-3.65)	0.087
Stage M						
0	681	1.61(1.3-1.99)	<0.001	195	1.56(0.94-2.59)	0.080
1	10	-	-	0	-	-
Smoking history						
Never smoked	205	4.1(2.17-7.73)	<0.001	193	2.72(1.66-4.45)	<0.001
Smoked	820	1.87(1.39-2.52)	<0.001	603	1.71(1.34-2.18)	<0.001

OS, overall survival; FP, first progression.

our previous results, these results may indicate that TRIM28 inhibits the interferon-gamma and Toll-like receptor signaling pathways by increasing the ubiquitination (degradation) of IRF5 and IRF8, ultimately suppressing immune infiltration.

### TRIM28 may worsen the TIME by increasing the SUMOylation of IRF5 and IRF8

To identify potential SUMO substrates of TRIM28, we queried TRIM28 as E3 in the web tool of UbiBrowser [46]. The 79 predicted substrates with middle-confidence interactions and 347 predicted substrates with low-confidence interactions are presented in Supplementary Table 4. Figure 8A displays some of the substrates, and Figure 8B displays the predicted binding regions of IRF5 and IRF8 to TRIM28. We then performed a correlation analysis, which indicated that TRIM28 expression correlated negatively with IRF5 and

IRF8 expression in TCGA ( $R = -0.210$ ,  $p < 0.001$ ;  $R = -0.302$ ,  $p < 0.001$ ; Figure 8C and 8D), GSE43580 ( $R = -0.371$ ,  $p < 0.001$ ;  $R = -0.420$ ,  $p < 0.001$ ; Figure 8E and 8F) and GSE7670 ( $R = -0.491$ ,  $p = 0.004$ ;  $R = -0.430$ ,  $p = 0.014$ ; Figure 8G and 8H), respectively. Moreover, TRIM28 expression exhibited a strong negative correlation with stromal scores and immune scores (Figure 8I and 8J), while IRF5 and IRF8 expression exhibited strong positive correlations with stromal scores (Figure 8K and 8M) and immune scores (Figure 8L and 8N), respectively.

To verify our hypothesis that TRIM28 downregulates IRF5 and IRF8, we knocked down TRIM28 in two different LUAD cell lines (PC9 and H1299). The knockdown efficiency was validated through Western blotting (Figure 8O). As expected, IRF5 and IRF8 levels increased significantly when TRIM28 was knocked down (Figure 8O).



**Table 2. A multivariate Cox proportional hazards model was used to explore the effects of TRIM28 expression, clinicopathological characteristics and immune infiltration levels on survival.**

LUAD (N=436)	Variable	Coef	HR	95%CI_l	95%CI_u	P-value
	Age	0.019	1.019	1.002	1.037	0.028
	Gender	-0.258	0.772	0.553	1.079	0.130
	Stage2	0.857	2.356	1.550	3.583	<0.001
	Stage3	1.065	2.901	1.900	4.430	<0.001
	Stage4	1.256	3.510	1.918	6.424	<0.001
	Purity	0.357	1.430	0.628	3.257	0.395
	B cell	-3.507	0.030	0.002	0.493	0.014
	CD8+Tcell	0.471	1.601	0.217	11.801	0.644
	CD4+Tcell	2.090	8.086	0.537	121.759	0.131
	Macrophage	0.782	2.187	0.138	34.629	0.579
	Neutrophil	-1.888	0.151	0.003	7.376	0.341
	Dendritic	0.219	1.245	0.297	5.208	0.765
	TRIM28	0.595	1.813	1.317	2.495	<0.001
LUSC (N=452)		Coef	HR	95%CI_l	95%CI_u	P-value
	Age	0.02	1.02	1.003	1.038	0.024
	Gender	0.303	1.353	0.957	1.913	0.087
	Stage2	0.107	1.113	0.792	1.564	0.539
	Stage3	0.495	1.641	1.121	2.402	0.011
	Stage4	0.982	2.67	0.949	7.507	0.063
	Purity	-0.128	0.88	0.422	1.834	0.733
	B cell	1.298	3.663	0.298	45.061	0.311
	CD8+Tcell	-1.727	0.178	0.028	1.124	0.066
	CD4+Tcell	0.705	2.023	0.153	26.753	0.593
	Macrophage	-0.339	0.713	0.061	8.326	0.787
	Neutrophil	0.980	2.665	0.099	71.525	0.559
	Dendritic	0.722	2.059	0.489	8.662	0.325
	TRIM28	0.000	1.000	0.774	1.292	1.000

Coef, regression coefficient; HR, hazard ratio; 95%CI\_l, 95% confidence interval lower limit; 95%CI\_u, 95% confidence interval upper limit.

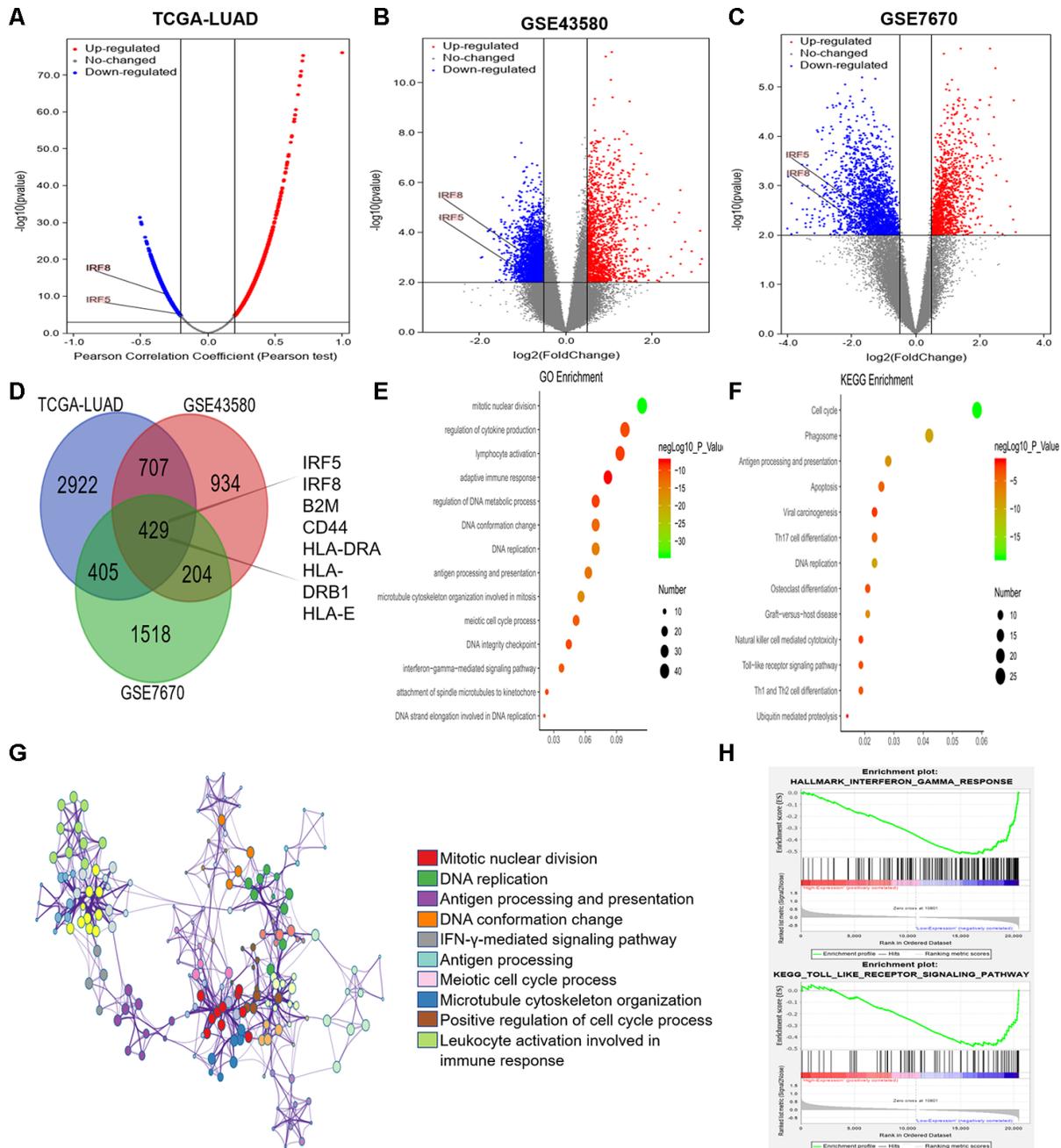
associated with poorer DMFS and OS in the PrognScan database (Figure 5D and 5E). In addition, *TRIM28* expression had little impact on breast cancer prognosis in the Kaplan-Meier plotter (Supplementary Figure 1) and no effect on breast cancer prognosis in GEPIA. These discrepancies may reflect differing data collection approaches or underlying mechanisms. However, in all the databases we examined, higher *TRIM28* expression was associated with a poorer prognosis of LUAD. These findings strongly suggested that *TRIM28* is a prognostic biomarker in LUAD.

The immune system is a critical regulator of tumor biology and has the capacity to support or inhibit tumor development, growth, invasion and metastasis. Tumor cells adopt a variety of mechanisms to avoid immune recognition and destruction, including: 1) downregulating human leukocyte antigen (HLA) class I molecules such as HLA-A, HLA-B, HLA-C and B2M on the cancer cell surface; 2) altering the antigen-presenting cell number or function; 3) lacking costimulation molecules such as B7-1, B7-2 and CD40; 4) promoting negative immune

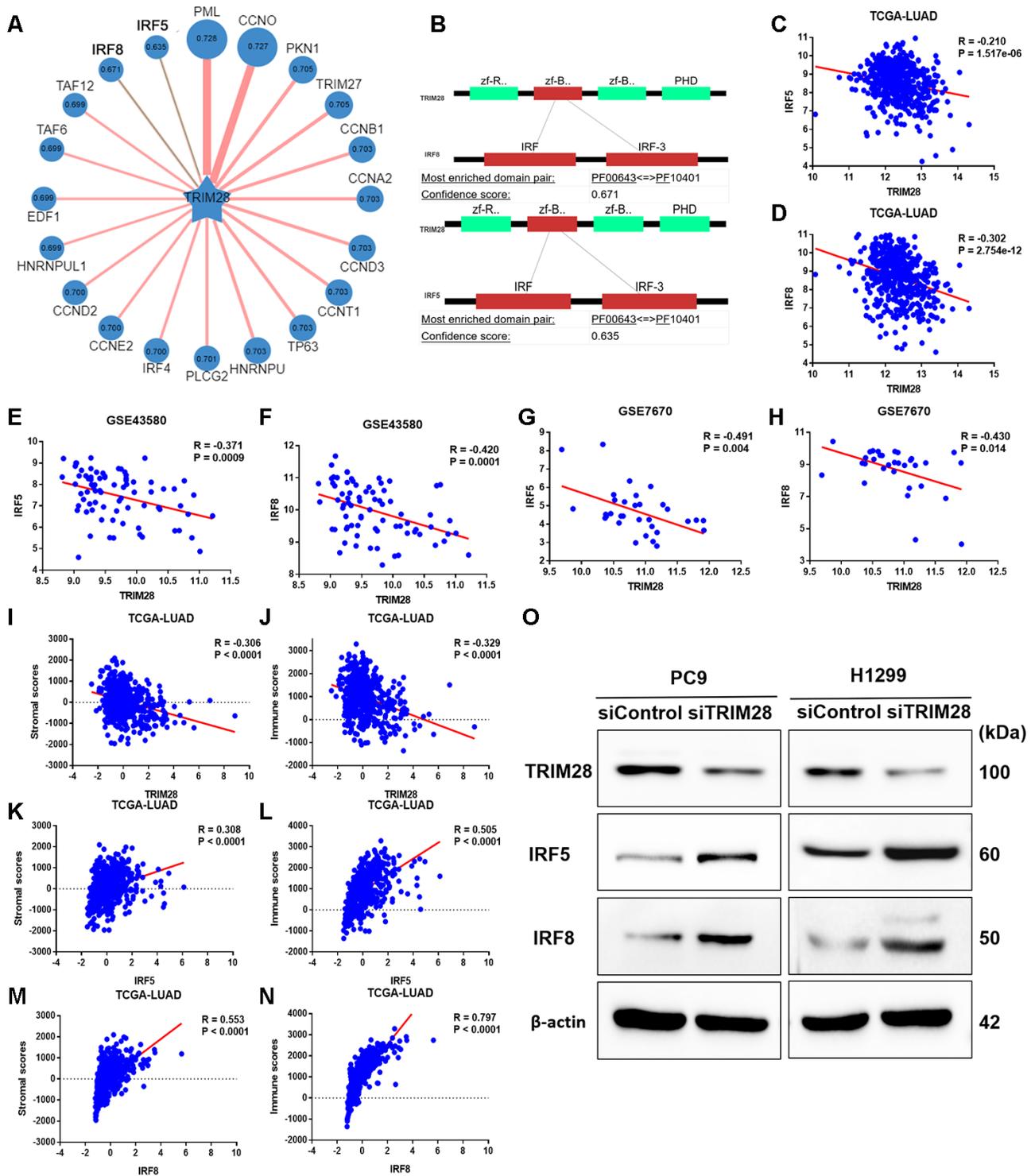
regulation by regulatory T cells and mesenchymal stem cells; 5) secreting immunosuppressive cytokines such as interleukin (IL)-10, transforming growth factor  $\beta$  and IL-6; 6) aberrantly expressing apoptosis-related molecules such as Fas, Fas ligand, tumor necrosis factor-related apoptosis inducing ligand and BAX; and 7) inhibiting effector cells via inhibitory ligands such as PD-L1, cytotoxic T-lymphocyte associated protein 4 and lymphocyte activation gene 3 [47–52]. The types and frequencies of these immune escape mechanisms vary among different cancer types. The proportions and activities of effector cells and antigen-presenting cells such as dendritic cells, B cells and macrophages are often reduced in the peripheral blood of cancer patients, while the numbers of immune-suppressive mesenchymal stem cells, natural killer T cells and regulatory T cells are generally elevated [49, 50]. Using different bioinformatics methods, we demonstrated that *TRIM28* expression correlated negatively with the immune infiltration of LUAD. Thus, we proposed that *TRIM28* may negatively regulate the TIME and thereby promote tumor development and progression.

There are several possible mechanisms by which TRIM28 could impair immune cell infiltration. 1) TRIM28 critically inhibits the induction of Foxp3, the number of regulatory T cells, Th17 cells differentiation [20, 53, 54], and macrophage activation [18, 24],

suggesting that TRIM28 may impair the TIME. 2) TRIM28 inhibits the effects of IRF5 on gene expression, and IRF5 has been reported to repress anti-inflammatory genes such as *IL-10* [18]. 3) TRIM28 negatively regulates IRF7, which is a potent transcription



**Figure 7. Analysis of DEGs according to TRIM28 expression in LUAD patients.** (A) Volcano map showing all the genes associated with TRIM28 in LUAD. (B, C) Volcano maps showing all the DEGs based on TRIM28 expression. (D) Venn diagram showing the number of common DEGs. (E, F) Top 14 GO enrichment terms (E) and top 13 KEGG enrichment terms (F) for all DEGs, analyzed in Metascape. (G) Associations among the top 10 cluster enrichment terms analyzed by Metascape, displayed as a network. An edge links terms with a similarity score > 0.3. (H) Gene Set Enrichment Analysis according to the expression of TRIM28 in TCGA. GO, gene ontology; DEGs, differentially expressed genes; LUAD, lung adenocarcinoma; GSEA, Gene Set Enrichment Analysis; TCGA, The Cancer Genome Atlas; GSE, Gene Expression Omnibus data series.



**Figure 8. The association of TRIM28 with IRF5 and IRF8.** (A) Network view of predicted E3-substrate interactions in UbiBrowser web services. In network view, the central node is the queried E3 ligase, and the surrounding nodes are the predicted substrates. The width of the edge reflects the confidence of the interaction. (B) The possible binding regions of IRF5 and IRF8 to TRIM28. (C–H) TRIM28 expression exhibited a significant negative correlation with IRF5 and IRF8 expression in TCGA (C, D), GSE43580 (E, F) and GSE7670 (G, H). (I–N) TRIM28 expression exhibited a strong negative correlation with stromal scores (I) and immune scores (J), while IRF5 and IRF8 levels exhibited strong positive relationships with stromal scores (K, M) and immune scores (L, N), respectively. (O) The expression of IRF5 and IRF8 after knocking down TRIM28 in two different LUAD cell lines (PC9 and H1299) through Western blotting. The gels have been run under the same experimental conditions. The blot of IRF5 in H1299 is obtained from the combined image merging the blot image and the ladder image. LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas; GSE, Gene Expression Omnibus data series.

factor of type I interferons and interferon-stimulated genes and is known as the master regulator of type I interferon-dependent immune responses [22]. 4) TRIM28 suppressed TNF- $\alpha$ -induced IL-6 production and transcriptional activation of NF- $\kappa$ B [24, 25]. 5) TRIM28 functions as an important negative regulator of the expression of IFN- $\beta$ , IFN- $\gamma$ , IL-6 and IL-8 during viral infection [55]. We found that TRIM28 may inhibit the interferon-gamma and Toll-like receptor signaling pathways by increasing the SUMOylation of IRF5 and IRF8, ultimately suppressing immune infiltration in LUAD. We confirmed these findings by performing a correlation analysis and validations in LUAD cell lines. Overall, TRIM28 appears to globally manage various immune-related cells and to reduce immune infiltration by altering the expression of diverse chemokines and molecular signaling pathways.

Using a multivariate Cox proportional hazards model, we confirmed that *TRIM28* expression and infiltrating B cell levels were independent predictors of survival in LUAD. Tumor-infiltrating B cell levels are strongly associated with the prognosis of various tumor types [56–59]. The mechanisms by which tumor-infiltrating B cells influence tumor immunity may include: 1) functioning as antigen-presenting cells to facilitate innate cellular immunity in the TIME; 2) activating CD8<sup>+</sup> T cells to promote antigen-specific antitumor immune responses [60]; and 3) promoting adaptive immunity by inducing the release of circulating cytokines to recruit immunosuppressive cells [61]. However, further study is needed to understand their functions and mechanisms.

The PD1/PD-L1 cascade is a highly effective therapeutic target in immunotherapy [4–7]. PD-L1 is expressed in a variety of cancer types in either a constitutive (or intrinsic) or interferon-induced manner. The results of our research and previous studies suggest that TRIM28 regulates interferons in multiple ways. Liang et al. [62] reported that verteporfin, a small-molecule inhibitor, inhibited PD-L1 by inducing autophagy and disrupting the STAT1-IRF1-TRIM28 signaling axis, thus exerting antitumor effects in immunotherapy. However, the efficacy of immunotherapy depends not only on immune infiltration and PD-L1 expression, but also on the tumor mutation burden, epidermal growth factor receptor mutation status and other unknown factors [63]. Thus, further research is needed to evaluate the potential of TRIM28 as a therapeutic target in immunotherapy.

There are several limitations to our research. First, this was a retrospective analysis based on public databases (TCGA and GEO). The number of included patients was limited, and it was difficult to account for variations in

race, age and geographic area. Thus, additional *in vivo* and *in vitro* experiments are required for functional and clinical verification. Second, considering the possible spatial and temporal heterogeneity of the TIME, immune and stromal assessments should ideally be performed at the core and infiltrating edges of the tumor, respectively. However, all the data in this study were from samples in the core area of the tumor.

In conclusion, the present study demonstrated that TRIM28 worsens the TIME and is highly expressed in LUAD. Increased *TRIM28* expression was associated with reduced levels of various infiltrating immune cells, and was an independent prognostic factor in LUAD. TRIM28 may negatively regulate the TIME by increasing the SUMOylation of IRF5 and IRF8. Thus, our research has provided new insights into the suppressive function of TRIM28 in the TIME and the potential of *TRIM28* as a prognostic biomarker in LUAD.

## MATERIALS AND METHODS

### Gene expression profile data

TCGA data containing RNA sequencing results and clinical information (level 3 data) were downloaded from the cBioPortal database (<http://www.cbioportal.org>). The *TRIM28* expression data from GSE32863 [33], GSE7670 [34], GSE19188 [35] and the Beer Lund dataset [36] were downloaded from Oncomine (<https://www.oncomine.org/resource/main.html>) [64]. Two mRNA expression datasets (GSE43580 and GSE7670) were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) [34, 45], and were based on the GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) and GPL96 (Affymetrix Human Genome U133A Array) platforms, respectively. The GSE43580 dataset included 77 LUAD patients, while the GSE7670 dataset included 28 LUAD patients.

### DEG identification

R software (version 3.6.1) was used for detailed analyses in this study. We used the LinkedOmics database (<http://www.linkedomics.org/login.php>) [44] to analyze the genes co-expressed with *TRIM28* in LUAD. All DEG analyses were performed using the "limma" R package. Fold-changes in gene expression were calculated with threshold criteria of a  $|\log_2\text{fold-change}| > 0.5$ , false discovery rate  $< 0.05$  and adjusted  $p < 0.001$  for DEG selection.

### Functional enrichment analysis of DEGs

To explore the functions of the overlapping DEGs, we performed GO and KEGG analyses in Metascape

(<http://metascape.org/gp>) [32]. The selected GO terms were from the "Biological Process" annotation datasets. The cutoff value for pathway screening was set to  $p < 0.01$ . The levels of significant DEGs were visualized on a heatmap based on hierarchical clustering analyses using the average linkage method.

### PPI network construction and analysis

STRING (Version 11.0, <http://string-db.org>) is a database of known and predicted PPI networks. We used this tool to construct PPI networks and predict potential interactions between candidate genes. Interactions were considered significant above a cutoff score of 0.4. In addition, Cytoscape software (Version 3.7.2, <http://www.cytoscape.org/>) [65] and the cytoHubba plugin [66] were used to explore the hub genes in the PPI network. Eleven methods can be used to explore essential nodes in PPI networks, but maximal clique centrality performs better than the others.

### RNA and protein expression analyses

Oncomine [64] was used to explore the *TRIM28* levels in different tumor types. The parameters were adjusted according to the following criteria:  $p$ -value of  $1e-4$ , fold change of 2 and gene ranking in the top 10%.

TIMER (<https://cistrome.shinyapps.io/timer/>) [67] is a comprehensive database that can be used to estimate the abundance of immune infiltrates and characterize the tumor-immune interactions across diverse tumor types. The levels of six tumor-infiltrating immune subsets, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells, are precalculated for 10,897 tumors from TCGA. We used TIMER to explore *TRIM28* expression in various tumor types.

In addition, the UALCAN database [68] can be used to analyze protein expression based on data from the Clinical Proteomic Tumor Analysis Consortium Confirmatory/Discovery dataset. We evaluated the protein levels of *TRIM28* in various cancers by performing a Clinical Proteomic Tumor Analysis Consortium analysis.

### Prognosis analysis

The PrognoScan database (<http://www.abren.net/PrognoScan/>) was used to determine the relationship between *TRIM28* levels and the prognoses of different tumor types [69]. PrognoScan analyzes the correlations between gene levels and prognostic indicators such as DFS and OS using a large number of public tumor

microarray datasets. The threshold was set to a Cox  $p$ -value of 0.05.

Kaplan-Meier plotter (<http://kmplot.com/analysis/>) [70], an online database of published microarray datasets, can be used to assess the impact of 54,675 genes on survival using 18,674 cancer samples, including samples from 5,143 breast, 1,816 ovarian, 2,437 lung and 1,065 gastric cancer patients. We used the Kaplan-Meier plotter to explore the relationship between *TRIM28* expression and prognosis in breast, lung, ovarian and gastric cancers. The HR, 95% CI and  $p$ -value were all calculated.

GEPIA (<http://gepia.cancer-pku.cn/index.html>) [71] is an online database that can be used for differential gene expression analysis, profile plotting, correlation analysis, patient survival analysis, similar gene detection and dimensionality reduction analysis based on TCGA and Genotype-Tissue Expression data. We used GEPIA to analyze the prognostic value of *TRIM28* expression based on the log-rank test in 33 cancer types.

### Stromal and immune score calculation and immune infiltration analysis

Using the "Estimate" R package, we calculated the stromal and immune scores of LUAD patients based on their gene expression profiles [11]. To verify the relationships between the target genes and the TIME, we used the TIMER database and another comprehensive database (TISIDB, <http://cis.hku.hk/TISIDB/index.php>) [72]. TISIDB integrates multiple heterogeneous data types. Spearman correlations between *TRIM28* levels and tumor-infiltrating lymphocyte levels across human cancers were analyzed. All hypothetical tests were two-sided, and  $p$ -values  $< 0.05$  were considered significant.

### Query for E3-TRIM28 interactions in UbiBrowser

UbiBrowser (<http://ubibrowser.ncpsb.org/>) [46] is an integrated bioinformatics platform that can be used to predict proteome-wide human E3-substrate networks based on naïve Bayesian networks. It currently contains 1,295 literature-reported E3-substrate interactions and 8,255 predicted E3-substrate interactions. We used it to predict the potential substrates of *TRIM28*.

### Cell culture and transfection

We obtained the PC9 cell line from the RIKEN BioResource Center (Tsukuba, Japan), and purchased the H1299 cell line from the American Type Culture Collection (Manassas, VA, USA). The cells were propagated in RPMI 1640 medium (Invitrogen,

Carlsbad, CA, USA) supplemented with 10 % fetal bovine serum and antibiotics.

For transfection, the cells ( $1 \times 10^5$ ) were seeded in six-well plates. The cells were transfected using jetPRIME reagent (Polyplus transfection) with 110 pmol of ON-TARGETplus SMARTpool-Human TRIM28 (L-005046-00-0020, Dharmacon) according to the manufacturers' protocols.

### Western blotting

The Western blot assay was conducted as previously described. Briefly, the cells were washed in cold  $1 \times$  phosphate-buffered saline and lysed in ice-cold radioimmunoprecipitation assay buffer supplemented with protease inhibitors on ice for 30 min. The protein concentrations were quantified using the bicinchoninic acid method according to the manufacturer's instructions. The protein samples were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and then transferred to nitrocellulose membranes (BioRad, USA). The membranes were blocked with 5% non-fat milk for one hour, and then incubated overnight with primary antibodies diluted in 2% bovine serum albumin. After being washed, the membranes were incubated with horseradish peroxidase-conjugated secondary antibodies for one hour. The primary antibodies used for immunodetection were rabbit monoclonal anti-TRIM28 (#4124, Cell Signaling Technology, USA), rabbit polyclonal anti-IRF5 (#20261, Cell Signaling Technology), rabbit monoclonal anti-IRF8 (#5628, Cell Signaling Technology) and rabbit polyclonal anti- $\beta$ -actin (#ab8227, Abcam, UK). The secondary antibody was obtained from Abcam (#ab205718). Blots were visualized with an enhanced chemiluminescence reagent (Supersignal; Pierce, Rockford, IL, USA).

### Statistical analysis

The results from the OncoPrint database are displayed with *p*-values, fold-changes and ranks. Interactive heatmaps were constructed using Next-Generation Clustered Heatmaps [73]. Prognoscan, GEPIA and Kaplan-Meier plots were used to create survival curves, and the results are displayed with HRs and *p*-values or Cox *p*-values. A multivariate Cox proportional hazards model was used to analyze the independent prognostic factors for lung cancer. *P*-values  $< 0.05$  were considered statistically significant.

### Abbreviations

ACC: adrenocortical carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma;

CI: confidence interval; DEGs: differentially expressed genes; DFS: disease-free survival; DMFS: distant metastasis-free survival; DSS: disease-specific survival; FP: first progression; HR: hazard ratio; KIRP: kidney renal papillary cell carcinoma; LGG: brain lower-grade glioma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; MESO: mesothelioma; NSCLC: non-small-cell lung cancer; OS: overall survival; PD1: programmed cell death-1; PD-L1: programmed cell death ligand-1; PPI: protein-protein interaction; RFS: recurrence-free survival; SKCM: skin cutaneous melanoma; TCGA: The Cancer Genome Atlas; TIME: tumor immune microenvironment; TRIM: Tripartite motif-containing.

### AUTHOR CONTRIBUTIONS

J.L. and X.H. conceived and designed the experiments; X.H., L.J.C. and J.L. performed the experiments; J.L., X.Q.M., H.J.W. and X.F.H. analyzed the data; H.B.W., W.J.L. and J.L. prepared the figures and tables; J.L., D.H. and Y.Q.Z. drafted the work or revised it critically for important content. All authors reviewed the manuscript.

### CONFLICTS OF INTEREST

The authors report no conflicts of interest.

### FUNDING

This study was supported by Grants-in-Aid from the Medical Innovation Talent Program, Health Department of Henan Province, China (Project No. 201702247).

### REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87–108. <https://doi.org/10.3322/caac.21262> PMID:[25651787](https://pubmed.ncbi.nlm.nih.gov/25651787/)
2. Jie-Liu, Li XY, Zhao YQ, Liu RQ, Zhang JB, Ma J, Chen LJ, Hu XF. Genotype-phenotype correlation in Chinese patients with pulmonary mixed type adenocarcinoma: relationship between histologic subtypes, TITF-1/SP-a expressions and EGFR mutations. *Pathol Res Pract.* 2014; 210:176–81. <https://doi.org/10.1016/j.prp.2013.11.013> PMID:[24370340](https://pubmed.ncbi.nlm.nih.gov/24370340/)
3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74. <https://doi.org/10.1016/j.cell.2011.02.013> PMID:[21376230](https://pubmed.ncbi.nlm.nih.gov/21376230/)

4. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, Gadgeel SM, Hida T, Kowalski DM, Dols MC, Cortinovis DL, Leach J, Polikoff J, et al, and OAK Study Group. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017; 389:255–65. [https://doi.org/10.1016/S0140-6736\(16\)32517-X](https://doi.org/10.1016/S0140-6736(16)32517-X) PMID:27979383
5. Herbst RS, Baas P, Kim DW, Felip E, Perez-Gracia JL, Han JY, Molina J, Kim JH, Arvis CD, Ahn MJ, Majem M, Fidler MJ, de Castro G Jr, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016; 387:1540–50. [https://doi.org/10.1016/S0140-6736\(15\)01281-7](https://doi.org/10.1016/S0140-6736(15)01281-7) PMID:26712084
6. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E, Waterhouse D, Ready N, Gainor J, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. 2015; 373:123–35. <https://doi.org/10.1056/NEJMoa1504627> PMID:26028407
7. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhäufel M, Arrieta O, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015; 373:1627–39. <https://doi.org/10.1056/NEJMoa1507643> PMID:26412456
8. Ali A, Goffin JR, Arnold A, Ellis PM. Survival of patients with non-small-cell lung cancer after a diagnosis of brain metastases. *Curr Oncol*. 2013; 20:e300–06. <https://doi.org/10.3747/co.20.1481> PMID:23904768
9. Waniczek D, Lorenc Z, Śnietura M, Wesecki M, Kopec A, Muc-Wierzoń M. Tumor-associated macrophages and regulatory T cells infiltration and the clinical outcome in colorectal cancer. *Arch Immunol Ther Exp (Warsz)*. 2017; 65:445–54. <https://doi.org/10.1007/s00005-017-0463-9> PMID:28343267
10. Zhang H, Liu H, Shen Z, Lin C, Wang X, Qin J, Qin X, Xu J, Sun Y. Tumor-infiltrating neutrophils is prognostic and predictive for postoperative adjuvant chemotherapy benefit in patients with gastric cancer. *Ann Surg*. 2018; 267:311–18. <https://doi.org/10.1097/SLA.0000000000002058> PMID:27763900
11. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013; 4:2612. <https://doi.org/10.1038/ncomms3612> PMID:24113773
12. Hao L, Leng J, Xiao R, Kingsley T, Li X, Tu Z, Yang X, Deng X, Xiong M, Xiong J, Zhang Q. Bioinformatics analysis of the prognostic value of Tripartite Motif 28 in breast cancer. *Oncol Lett*. 2017; 13:2670–78. <https://doi.org/10.3892/ol.2017.5764> PMID:28454449
13. Cheng B, Ren X, Kerppola TK. KAP1 represses differentiation-inducible genes in embryonic stem cells through cooperative binding with PRC1 and derepresses pluripotency-associated genes. *Mol Cell Biol*. 2014; 34:2075–91. <https://doi.org/10.1128/MCB.01729-13> PMID:24687849
14. Addison JB, Koontz C, Fugett JH, Creighton CJ, Chen D, Farrugia MK, Padon RR, Voronkova MA, McLaughlin SL, Livengood RH, Lin CC, Ruppert JM, Pugacheva EN, Ivanov AV. KAP1 promotes proliferation and metastatic progression of breast cancer cells. *Cancer Res*. 2015; 75:344–55. <https://doi.org/10.1158/0008-5472.CAN-14-1561> PMID:25421577
15. Jin X, Pan Y, Wang L, Zhang L, Ravichandran R, Potts PR, Jiang J, Wu H, Huang H. MAGE-TRIM28 complex promotes the warburg effect and hepatocellular carcinoma progression by targeting FBP1 for degradation. *Oncogenesis*. 2017; 6:e312. <https://doi.org/10.1038/oncsis.2017.21> PMID:28394358
16. Czerwińska P, Mazurek S, Wiznerowicz M. The complexity of TRIM28 contribution to cancer. *J Biomed Sci*. 2017; 24:63. <https://doi.org/10.1186/s12929-017-0374-4> PMID:28851455
17. Santoni de Sio FR, Barde I, Offner S, Kapopoulou A, Corsinotti A, Bojkowska K, Genolet R, Thomas JH, Luescher IF, Pinschewer D, Harris N, Trono D. KAP1 regulates gene networks controlling T-cell development and responsiveness. *FASEB J*. 2012; 26:4561–75. <https://doi.org/10.1096/fj.12-206177> PMID:22872677
18. Eames HL, Saliba DG, Krausgruber T, Lanfrancotti A, Ryzhakov G, Udalova IA. KAP1/TRIM28: an inhibitor of IRF5 function in inflammatory macrophages. *Immunobiology*. 2012; 217:1315–24. <https://doi.org/10.1016/j.imbio.2012.07.026> PMID:22995936
19. Zhou XF, Yu J, Chang M, Zhang M, Zhou D, Cammas F,

- Sun SC. TRIM28 mediates chromatin modifications at the TCR $\alpha$  enhancer and regulates the development of T and natural killer T cells. *Proc Natl Acad Sci USA*. 2012; 109:20083–88.  
<https://doi.org/10.1073/pnas.1214704109>  
PMID: [23169648](https://pubmed.ncbi.nlm.nih.gov/23169648/)
20. Chikuma S, Suita N, Okazaki IM, Shibayama S, Honjo T. TRIM28 prevents autoinflammatory T cell development in vivo. *Nat Immunol*. 2012; 13:596–603.  
<https://doi.org/10.1038/ni.2293> PMID: [22544392](https://pubmed.ncbi.nlm.nih.gov/22544392/)
21. Santos J, Gil J. TRIM28/KAP1 regulates senescence. *Immunol Lett*. 2014; 162:281–89.  
<https://doi.org/10.1016/j.imlet.2014.08.011>  
PMID: [25160591](https://pubmed.ncbi.nlm.nih.gov/25160591/)
22. Liang Q, Deng H, Li X, Wu X, Tang Q, Chang TH, Peng H, Rauscher FJ 3rd, Ozato K, Zhu F. Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J Immunol*. 2011; 187:4754–63.  
<https://doi.org/10.4049/jimmunol.1101704>  
PMID: [21940674](https://pubmed.ncbi.nlm.nih.gov/21940674/)
23. McNab FW, Rajsbaum R, Stoye JP, O'Garra A. Tripartite-motif proteins and innate immune regulation. *Curr Opin Immunol*. 2011; 23:46–56.  
<https://doi.org/10.1016/j.coi.2010.10.021>  
PMID: [21131187](https://pubmed.ncbi.nlm.nih.gov/21131187/)
24. Tsuruma R, Ohbayashi N, Kamitani S, Ikeda O, Sato N, Muromoto R, Sekine Y, Oritani K, Matsuda T. Physical and functional interactions between STAT3 and KAP1. *Oncogene*. 2008; 27:3054–59.  
<https://doi.org/10.1038/sj.onc.1210952>  
PMID: [18037959](https://pubmed.ncbi.nlm.nih.gov/18037959/)
25. Kamitani S, Togi S, Ikeda O, Nakasuji M, Sakauchi A, Sekine Y, Muromoto R, Oritani K, Matsuda T. Krüppel-associated box-associated protein 1 negatively regulates TNF- $\alpha$ -induced NF- $\kappa$ B transcriptional activity by influencing the interactions among STAT3, p300, and NF- $\kappa$ B/p65. *J Immunol*. 2011; 187:2476–83.  
<https://doi.org/10.4049/jimmunol.1003243>  
PMID: [21810609](https://pubmed.ncbi.nlm.nih.gov/21810609/)
26. Yokoe T, Toiyama Y, Okugawa Y, Tanaka K, Ohi M, Inoue Y, Mohri Y, Miki C, Kusunoki M. KAP1 is associated with peritoneal carcinomatosis in gastric cancer. *Ann Surg Oncol*. 2010; 17:821–28.  
<https://doi.org/10.1245/s10434-009-0795-8>  
PMID: [19898899](https://pubmed.ncbi.nlm.nih.gov/19898899/)
27. Cui Y, Yang S, Fu X, Feng J, Xu S, Ying G. High levels of KAP1 expression are associated with aggressive clinical features in ovarian cancer. *Int J Mol Sci*. 2014; 16:363–77.  
<https://doi.org/10.3390/ijms16010363>  
PMID: [25548895](https://pubmed.ncbi.nlm.nih.gov/25548895/)
28. Fitzgerald S, Sheehan KM, O'Grady A, Kenny D, O'Kennedy R, Kay EW, Kijanka GS. Relationship between epithelial and stromal TRIM28 expression predicts survival in colorectal cancer patients. *J Gastroenterol Hepatol*. 2013; 28:967–74.  
<https://doi.org/10.1111/jgh.12157>  
PMID: [23425061](https://pubmed.ncbi.nlm.nih.gov/23425061/)
29. Liu L, Zhao E, Li C, Huang L, Xiao L, Cheng L, Huang X, Song Y, Xu D. TRIM28, a new molecular marker predicting metastasis and survival in early-stage non-small cell lung cancer. *Cancer Epidemiol*. 2013; 37:71–78.  
<https://doi.org/10.1016/j.canep.2012.08.005>  
PMID: [22959342](https://pubmed.ncbi.nlm.nih.gov/22959342/)
30. Chen L, Chen DT, Kurtyka C, Rawal B, Fulp WJ, Haura EB, Cress WD. Tripartite motif containing 28 (Trim28) can regulate cell proliferation by bridging HDAC1/E2F interactions. *J Biol Chem*. 2012; 287:40106–18.  
<https://doi.org/10.1074/jbc.M112.380865>  
PMID: [23060449](https://pubmed.ncbi.nlm.nih.gov/23060449/)
31. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013; 6:p11.  
<https://doi.org/10.1126/scisignal.2004088>  
PMID: [23550210](https://pubmed.ncbi.nlm.nih.gov/23550210/)
32. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019; 10:1523.  
<https://doi.org/10.1038/s41467-019-09234-6>  
PMID: [30944313](https://pubmed.ncbi.nlm.nih.gov/30944313/)
33. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, Lam S, Gazdar AF, Laird-Offringa IA. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res*. 2012; 22:1197–211.  
<https://doi.org/10.1101/gr.132662.111>  
PMID: [22613842](https://pubmed.ncbi.nlm.nih.gov/22613842/)
34. Chen CH, Lai JM, Chou TY, Chen CY, Su LJ, Lee YC, Cheng TS, Hong YR, Chou CK, Whang-Peng J, Wu YC, Huang CY. VEGFA upregulates FLJ10540 and modulates migration and invasion of lung cancer via PI3K/AKT pathway. *PLoS One*. 2009; 4:e5052.  
<https://doi.org/10.1371/journal.pone.0005052>  
PMID: [19337377](https://pubmed.ncbi.nlm.nih.gov/19337377/)
35. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, van der Spek P, Foekens JA, Hoogsteden HC, Grosveld F, Philipsen S. Gene expression-based classification of non-small cell

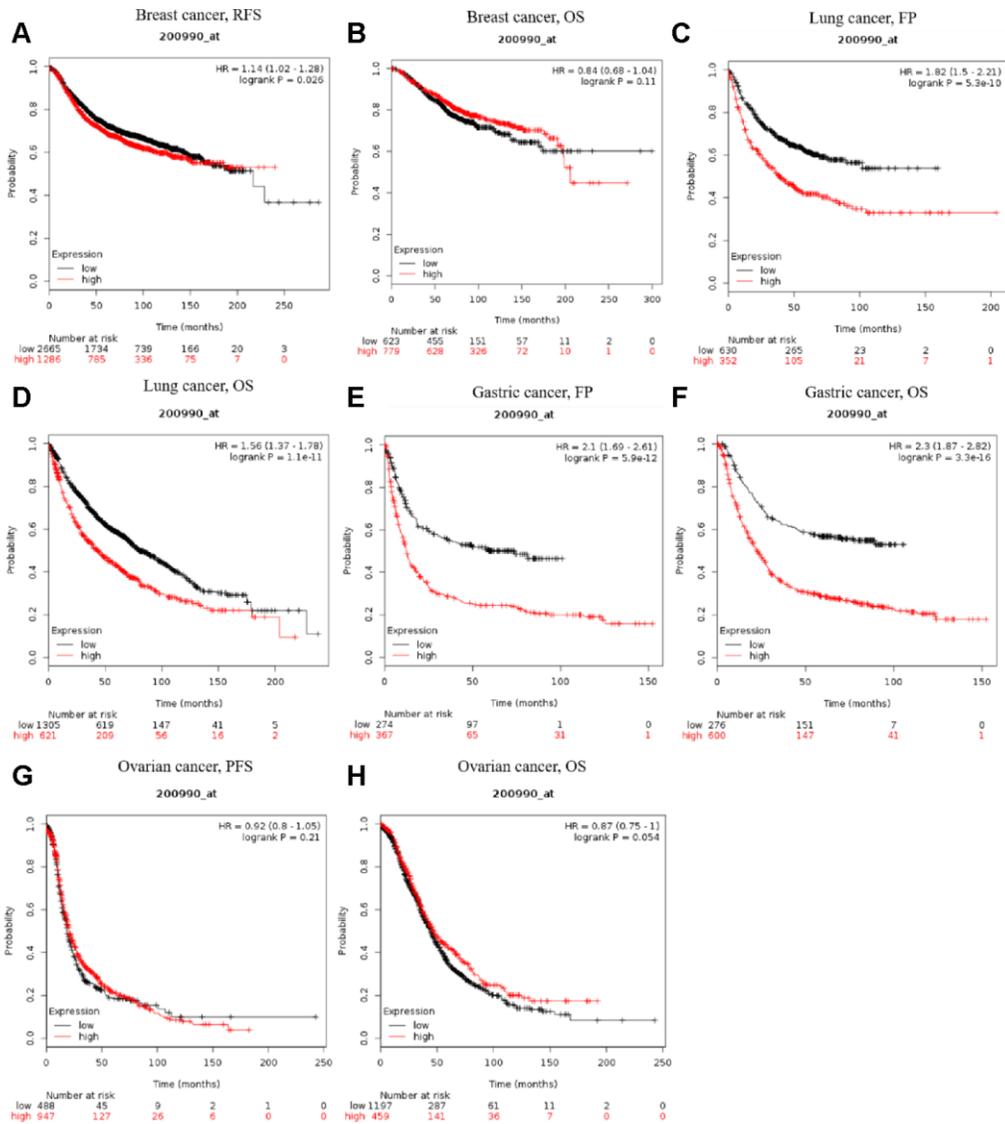
- lung carcinomas and survival prediction. *PLoS One*. 2010; 5:e10312.  
<https://doi.org/10.1371/journal.pone.0010312>  
PMID:[20421987](https://pubmed.ncbi.nlm.nih.gov/20421987/)
36. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*. 2002; 8:816–24.  
<https://doi.org/10.1038/nm733> PMID:[12118244](https://pubmed.ncbi.nlm.nih.gov/12118244/)
37. Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, Wong JE, Liu ET, Bergh J, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res*. 2006; 66:10292–301.  
<https://doi.org/10.1158/0008-5472.CAN-05-4414>  
PMID:[17079448](https://pubmed.ncbi.nlm.nih.gov/17079448/)
38. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, Bergh J. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA*. 2005; 102:13550–55.  
<https://doi.org/10.1073/pnas.0506230102>  
PMID:[16141321](https://pubmed.ncbi.nlm.nih.gov/16141321/)
39. Yau C, Benz CC. Genes responsive to both oxidant stress and loss of estrogen receptor function identify a poor prognosis group of estrogen receptor positive primary breast cancers. *Breast Cancer Res*. 2008; 10:R61.  
<https://doi.org/10.1186/bcr2120>  
PMID:[18631401](https://pubmed.ncbi.nlm.nih.gov/18631401/)
40. Chanrion M, Negre V, Fontaine H, Salvetat N, Bibeau F, Mac Grogan G, Mauriac L, Katsaros D, Molina F, Theillet C, Darbon JM. A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clin Cancer Res*. 2008; 14:1744–52.  
<https://doi.org/10.1158/1078-0432.CCR-07-1833>  
PMID:[18347175](https://pubmed.ncbi.nlm.nih.gov/18347175/)
41. Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kölbl H, Gehrman M. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res*. 2008; 68:5405–13.  
<https://doi.org/10.1158/0008-5472.CAN-07-5206>  
PMID:[18593943](https://pubmed.ncbi.nlm.nih.gov/18593943/)
42. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, Imoto S, Saito A, Ueno K, Hatanaka Y, Yoshida R, Higuchi T, Nomura M, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One*. 2012; 7:e43923.  
<https://doi.org/10.1371/journal.pone.0043923>  
PMID:[23028479](https://pubmed.ncbi.nlm.nih.gov/23028479/)
43. Senovilla L, Vacchelli E, Galon J, Adjemian S, Eggermont A, Fridman WH, Sautès-Fridman C, Ma Y, Tartour E, Zitvogel L, Kroemer G, Galluzzi L. Trial watch: prognostic and predictive value of the immune infiltrate in cancer. *Oncoimmunology*. 2012; 1:1323–43.  
<https://doi.org/10.4161/onci.22009>  
PMID:[23243596](https://pubmed.ncbi.nlm.nih.gov/23243596/)
44. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res*. 2018; 46:D956–63.  
<https://doi.org/10.1093/nar/gkx1090>  
PMID:[29136207](https://pubmed.ncbi.nlm.nih.gov/29136207/)
45. Tarca AL, Lauria M, Unger M, Bilal E, Boue S, Kumar Dey K, Hoeng J, Koepl H, Martin F, Meyer P, Nandy P, Norel R, Peitsch M, et al, and IMPROVER DSC Collaborators. Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER diagnostic signature challenge. *Bioinformatics*. 2013; 29:2892–99.  
<https://doi.org/10.1093/bioinformatics/btt492>  
PMID:[23966112](https://pubmed.ncbi.nlm.nih.gov/23966112/)
46. Li Y, Xie P, Lu L, Wang J, Diao L, Liu Z, Guo F, He Y, Liu Y, Huang Q, Liang H, Li D, He F. An integrated bioinformatics platform for investigating the human E3 ubiquitin ligase-substrate interaction network. *Nat Commun*. 2017; 8:347.  
<https://doi.org/10.1038/s41467-017-00299-9>  
PMID:[28839186](https://pubmed.ncbi.nlm.nih.gov/28839186/)
47. Spranger S. Mechanisms of tumor escape in the context of the t-cell-inflamed and the non-T-cell-inflamed tumor microenvironment. *Int Immunol*. 2016; 28:383–91.  
<https://doi.org/10.1093/intimm/dxw014>  
PMID:[26989092](https://pubmed.ncbi.nlm.nih.gov/26989092/)
48. Yaacoub K, Pedoux R, Tarte K, Guillaudeux T. Role of the tumor microenvironment in regulating apoptosis and cancer progression. *Cancer Lett*. 2016; 378:150–59.  
<https://doi.org/10.1016/j.canlet.2016.05.012>  
PMID:[27224890](https://pubmed.ncbi.nlm.nih.gov/27224890/)
49. Giraldo NA, Becht E, Vano Y, Sautès-Fridman C, Fridman WH. The immune response in cancer: from immunology to pathology to immunotherapy. *Virchows Arch*. 2015; 467:127–35.  
<https://doi.org/10.1007/s00428-015-1787-7>  
PMID:[26077464](https://pubmed.ncbi.nlm.nih.gov/26077464/)
50. Wu AA, Drake V, Huang HS, Chiu S, Zheng L. Reprogramming the tumor microenvironment: tumor-induced immunosuppressive factors paralyze T cells.

- Oncoimmunology. 2015; 4:e1016700.  
<https://doi.org/10.1080/2162402X.2015.1016700>  
PMID:[26140242](https://pubmed.ncbi.nlm.nih.gov/26140242/)
51. Otsuka A, Dreier J, Cheng PF, Nägeli M, Lehmann H, Felderer L, Frew IJ, Matsushita S, Levesque MP, Dummer R. Hedgehog pathway inhibitors promote adaptive immune responses in basal cell carcinoma. *Clin Cancer Res*. 2015; 21:1289–97.  
<https://doi.org/10.1158/1078-0432.CCR-14-2110>  
PMID:[25593302](https://pubmed.ncbi.nlm.nih.gov/25593302/)
52. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, Torrejon DY, Abril-Rodriguez G, Sandoval S, Barthly L, Saco J, Homet Moreno B, Mezzadra R, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med*. 2016; 375:819–29.  
<https://doi.org/10.1056/NEJMoa1604958>  
PMID:[27433843](https://pubmed.ncbi.nlm.nih.gov/27433843/)
53. Tanaka S, Pflieger C, Lai JF, Roan F, Sun SC, Ziegler SF. KAP1 regulates regulatory T cell function and proliferation in both Foxp3-dependent and -independent manners. *Cell Rep*. 2018; 23:796–807.  
<https://doi.org/10.1016/j.celrep.2018.03.099>  
PMID:[29669285](https://pubmed.ncbi.nlm.nih.gov/29669285/)
54. Gehrman U, Burbage M, Zueva E, Goudot C, Esnault C, Ye M, Carpièr JM, Burgdorf N, Hoyler T, Suarez G, Joannas L, Heurtebise-Chrétien S, Durand S, et al. Critical role for TRIM28 and HP1 $\beta/\gamma$  in the epigenetic control of T cell metabolic reprogramming and effector differentiation. *Proc Natl Acad Sci USA*. 2019; 116:25839–49.  
<https://doi.org/10.1073/pnas.1901639116>  
PMID:[31776254](https://pubmed.ncbi.nlm.nih.gov/31776254/)
55. Krischuns T, Günl F, Henschel L, Binder M, Willemsen J, Schloer S, Rescher U, Gerlt V, Zimmer G, Nordhoff C, Ludwig S, Brunotte L. Phosphorylation of TRIM28 enhances the expression of IFN- $\beta$  and proinflammatory cytokines during HPAIV infection of human lung epithelial cells. *Front Immunol*. 2018; 9:2229.  
<https://doi.org/10.3389/fimmu.2018.02229>  
PMID:[30323812](https://pubmed.ncbi.nlm.nih.gov/30323812/)
56. Tsou P, Katayama H, Ostrin EJ, Hanash SM. The emerging role of B cells in tumor immunity. *Cancer Res*. 2016; 76:5597–601.  
<https://doi.org/10.1158/0008-5472.CAN-16-0431>  
PMID:[27634765](https://pubmed.ncbi.nlm.nih.gov/27634765/)
57. Wang SS, Liu W, Ly D, Xu H, Qu L, Zhang L. Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer. *Cell Mol Immunol*. 2019; 16:6–18.  
<https://doi.org/10.1038/s41423-018-0027-x>  
PMID:[29628498](https://pubmed.ncbi.nlm.nih.gov/29628498/)
58. Gupta P, Chen C, Chaluvally-Raghavan P, Pradeep S. B cells as an immune-regulatory signature in ovarian cancer. *Cancers (Basel)*. 2019; 11:894.  
<https://doi.org/10.3390/cancers11070894>  
PMID:[31248034](https://pubmed.ncbi.nlm.nih.gov/31248034/)
59. Wouters MC, Nelson BH. Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer. *Clin Cancer Res*. 2018; 24:6125–35.  
<https://doi.org/10.1158/1078-0432.CCR-18-1481>  
PMID:[30049748](https://pubmed.ncbi.nlm.nih.gov/30049748/)
60. Ritchie DS, Yang J, Hermans IF, Ronchese F. B-lymphocytes activated by CD40 ligand induce an antigen-specific anti-tumour immune response by direct and indirect activation of CD8<sup>+</sup> T-cells. *Scand J Immunol*. 2004; 60:543–51.  
<https://doi.org/10.1111/j.0300-9475.2004.01517.x>  
PMID:[15584965](https://pubmed.ncbi.nlm.nih.gov/15584965/)
61. de Visser KE, Korets LV, Coussens LM. De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. *Cancer Cell*. 2005; 7:411–23.  
<https://doi.org/10.1016/j.ccr.2005.04.014>  
PMID:[15894262](https://pubmed.ncbi.nlm.nih.gov/15894262/)
62. Liang J, Wang L, Wang C, Shen J, Su B, Marisetty AL, Fang D, Kassab C, Jeong KJ, Zhao W, Lu Y, Jain AK, Zhou Z, et al. Verteporfin inhibits PD-L1 through autophagy and the STAT1-IRF1-TRIM28 signaling axis, exerting antitumor efficacy. *Cancer Immunol Res*. 2020; 8:952–65.  
<https://doi.org/10.1158/2326-6066.CIR-19-0159>  
PMID:[32265228](https://pubmed.ncbi.nlm.nih.gov/32265228/)
63. Shien K, Papadimitrakopoulou VA, Wistuba II. Predictive biomarkers of response to PD-1/PD-L1 immune checkpoint inhibitors in non-small cell lung cancer. *Lung Cancer*. 2016; 99:79–87.  
<https://doi.org/10.1016/j.lungcan.2016.06.016>  
PMID:[27565919](https://pubmed.ncbi.nlm.nih.gov/27565919/)
64. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincaid-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007; 9:166–80.  
<https://doi.org/10.1593/neo.07112> PMID:[17356713](https://pubmed.ncbi.nlm.nih.gov/17356713/)
65. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504.  
<https://doi.org/10.1101/gr.1239303>  
PMID:[14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
66. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014 (Suppl

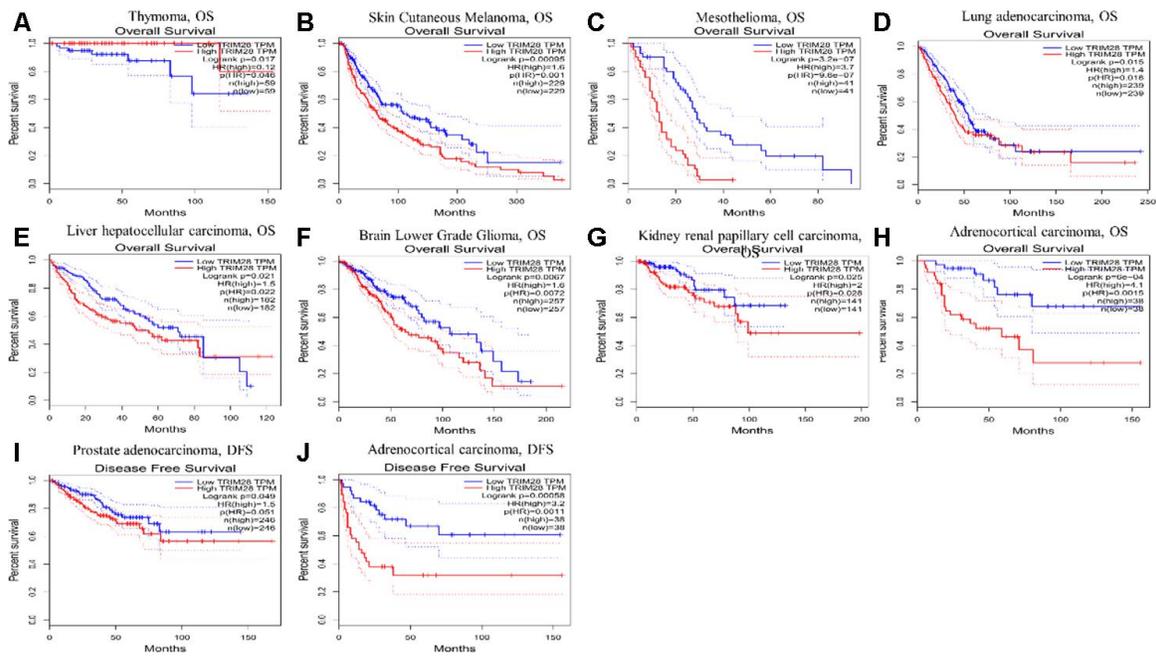
- 4); 8:S11.  
<https://doi.org/10.1186/1752-0509-8-S4-S11>  
PMID:[25521941](https://pubmed.ncbi.nlm.nih.gov/25521941/)
67. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
PMID:[29092952](https://pubmed.ncbi.nlm.nih.gov/29092952/)
68. Chandrashekar DS, Bashel B, Balasubramanya SA, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BV, Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia.* 2017; 19:649–58.  
<https://doi.org/10.1016/j.neo.2017.05.002>  
PMID:[28732212](https://pubmed.ncbi.nlm.nih.gov/28732212/)
69. Mizuno H, Kitada K, Nakai K, Sarai A. PrognosScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics.* 2009; 2:18.  
<https://doi.org/10.1186/1755-8794-2-18>  
PMID:[19393097](https://pubmed.ncbi.nlm.nih.gov/19393097/)
70. Lániczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, Gyórfy B. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat.* 2016; 160:439–46.  
<https://doi.org/10.1007/s10549-016-4013-7>  
PMID:[27744485](https://pubmed.ncbi.nlm.nih.gov/27744485/)
71. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45:W98–102.  
<https://doi.org/10.1093/nar/gkx247> PMID:[28407145](https://pubmed.ncbi.nlm.nih.gov/28407145/)
72. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SS, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics.* 2019; 35:4200–02.  
<https://doi.org/10.1093/bioinformatics/btz210>  
PMID:[30903160](https://pubmed.ncbi.nlm.nih.gov/30903160/)
73. Broom BM, Ryan MC, Brown RE, Ikeda F, Stucky M, Kane DW, Melott J, Wakefield C, Casasent TD, Akbani R, Weinstein JN. A galaxy implementation of next-generation clustered heatmaps for interactive exploration of molecular profiling data. *Cancer Res.* 2017; 77:e23–26.  
<https://doi.org/10.1158/0008-5472.CAN-17-0318>  
PMID:[29092932](https://pubmed.ncbi.nlm.nih.gov/29092932/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



**Supplementary Figure 1. Kaplan-Meier survival curves generated using Kaplan-Meier plotter for TRIM28 expression in different tumor types. (A, B) RFS and OS curves in breast cancer cohorts (n = 1,926, n = 641). (C, D) FP and OS curves in lung cancer cohorts (n = 982, n = 1,926). (E, F) FP and OS curves in gastric cancer cohorts (n = 641, n = 876). (G, H) PFS and OS curves in ovarian cancer cohorts (n = 1,435, n = 1,656). FP, first progression; PFS, Progression-free survival; OS, overall survival; RFS, recurrence-free survival.**



**Supplementary Figure 2.** Kaplan-Meier survival curves generated from the GEPIA database for TRIM28 expression in different tumor types (A–J). DFS, disease-free survival; OS, overall survival. 17. Clinical information data of 517 LUAD patients.

## **Supplementary Tables**

Please browse Full Text version to see the data of Supplementary Tables 1 to 4.

**Supplementary Table 1. Clinical information of 517 LUAD patients.**

**Supplementary Table 2. The immune and stromal scores of 517 LUAD patients.**

**Supplementary Table 3. The hub genes were screened by applying the cytoHubba plugin.**

**Supplementary Table 4. The substrates were predicted by using UbiBrowser.**

# Identification of a nomogram based on an 8-lncRNA signature as a novel diagnostic biomarker for head and neck squamous cell carcinoma

Rui Mao<sup>1,\*</sup>, Yuanyuan Chen<sup>2,\*</sup>, Lei Xiong<sup>3,\*</sup>, Yanjun Liu<sup>1,4</sup>, Tongtong Zhang<sup>5</sup>

<sup>1</sup>Affiliated Hospital of Southwest Jiaotong University, Chengdu, 610036, China

<sup>2</sup>Department of Pathology, The Third People's Hospital of Chengdu, Chengdu 610031, China

<sup>3</sup>Department of Otolaryngology, The Third People's Hospital of Chengdu, Chengdu 610031, China

<sup>4</sup>The Center of Gastrointestinal and Minimally Invasive Surgery, The Third People's Hospital of Chengdu, Chengdu 610031, China

<sup>5</sup>Medical Research Center, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, The Second Chengdu Hospital Affiliated to Chongqing Medical University, Chengdu 610031, Sichuan, China

\*Equal contribution

**Correspondence to:** Tongtong Zhang, Yanjun Liu; **email:** [mr1995@my.swjtu.edu.cn](mailto:mr1995@my.swjtu.edu.cn), [liujanjun@swjtu.edu.cn](mailto:liujanjun@swjtu.edu.cn)

**Keywords:** head and neck squamous carcinoma, long noncoding RNAs, prognosis, bioinformatics, qRT-PCR

**Received:** April 14, 2020

**Accepted:** August 17, 2020

**Published:** October 22, 2020

**Copyright:** © 2020 Mao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Long noncoding RNAs (lncRNAs) have been proposed as diagnostic or prognostic biomarkers of head and neck squamous carcinoma (HNSCC). The current study aimed to develop a lncRNA-based prognostic nomogram for HNSCC. lncRNA expression profiles were downloaded from The Cancer Genome Atlas (TCGA) database. After the reannotation of lncRNAs, the differential analysis identified 253 significantly differentially expressed lncRNAs in training set TCGA-HNSC (n = 300). The prognostic value of each lncRNA was first estimated in univariate Cox analysis, and 41 lncRNAs with P < 0.05 were selected as seed lncRNAs for Cox LASSO regression, which identified 11 lncRNAs. Multivariate Cox analysis was used to establish an 8-lncRNA signature with prognostic value. Patients in the high-signature score group exhibited a significantly worse overall survival (OS) than those in the low-signature score group, and the area under the receiver operating characteristic (ROC) curve for 3-year survival was 0.74. Multivariable Cox regression analysis among the clinical characteristics and signature scores suggested that the signature is an independent prognostic factor. The internal validation cohort, external validation cohort, and 102 HNSCC specimens quantified by qRT-PCR successfully validate the robustness of our nomogram.

## INTRODUCTION

The incidence and mortality of head and neck cancer have increased dramatically in recent decades. Most of the patients present advanced diseases with the characteristics of early invasion and metastasis [1, 2] [3]. Besides, despite advances in treatment, the 5-year

survival rate for head and neck cancer remains around 60%, which has improved only slightly over the past few decades [3, 4]. The current prognostic models for patients with HNSCC are based on clinicopathological parameters, but many cases with the same clinical stage show different results [2, 5]. Therefore, for patients with HNSCC, there is an urgent need for a

useful prognostic model that can predict the survival and prognosis of patients.

To identify lncRNAs associated with prognosis in HNSCC, we integrated gene matrix and clinical information from a TCGA dataset and the GSE65858 dataset to establish a nomogram with 8-lncRNA signature. Functional enrichment and WGCNA were performed to predict the potential functions of the gene modules, which are both related to the lncRNAs and clinical characteristics.

## RESULTS

### Preprocessing of the data sets

We downloaded the gene matrix of 546 samples from the TCGA-HNSC database, which included 502 tumour and 44 normal samples. We divided all HNSCC patients with complete information ( $n=499$ ) in TCGA-HNSC into training cohort and validation cohort, in a random manner according to a ratio of 3:2.

Moreover, From May 2017 to August 2018, a total of 102 frozen, surgically resected tumor tissues were obtained from patients with pathological diagnosis of HNSCC at Chengdu Third People's Hospital. The specimens were frozen with liquid nitrogen

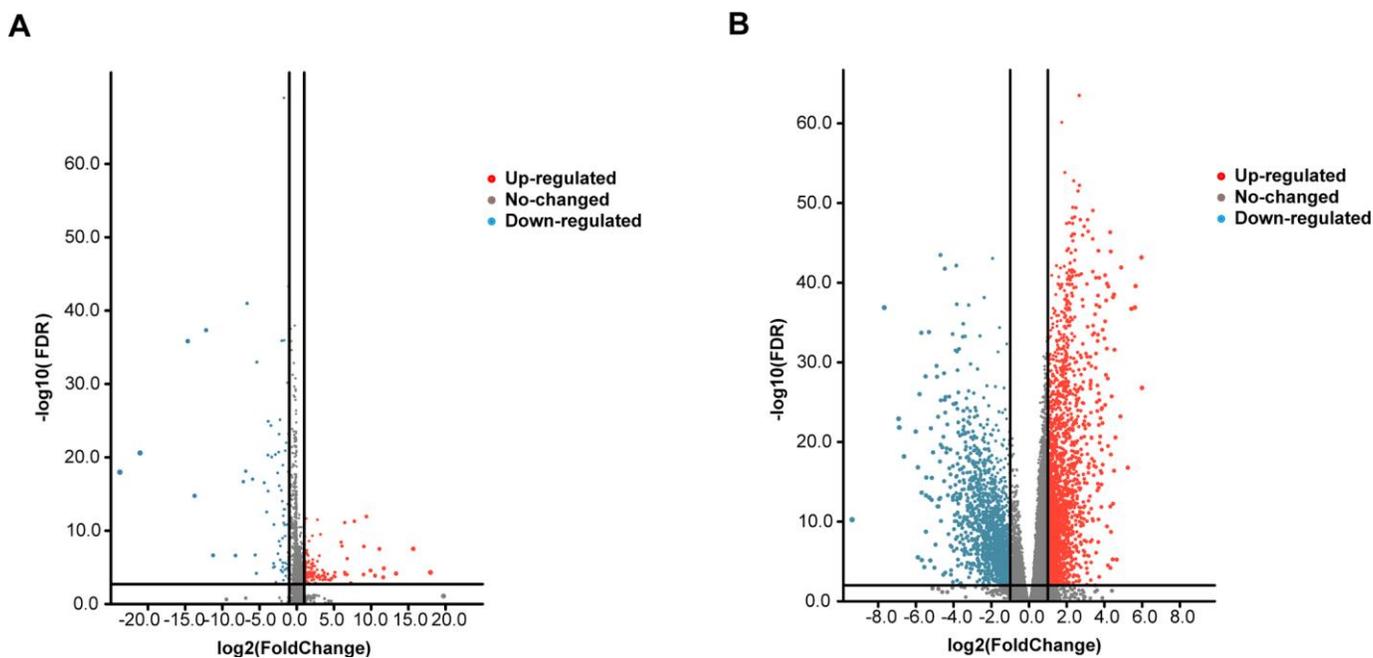
immediately after removal and transferred to the  $-80^{\circ}\text{C}$  refrigerator.

### Differential analysis

We conducted a differential analysis of the 300 tumor and 44 normal samples. Eventually, we obtained a total of 19754 mRNAs and 14847 lncRNAs. After obtaining the expression data, we identified differentially expressed genes using the software package EdgeR, selecting genes that had at least 2-fold higher expression levels in HNSCC samples (Poisson model  $\text{FDR} < 0.05$ ). Therefore, after screening, we obtained 4150 reliably expressed mRNAs and 253 lncRNAs (Figure 1A, 1B).

### Identification of 8-lncRNAs for predicting HNSCC patient survival

A total of 253 lncRNAs with significant differences were identified to have prognostic significance in univariate Cox survival analysis, and 41 with  $P < 0.05$  were screened out and applied in the following analysis (Figure 2A). As shown in Figure 2B, 2C, LASSO regression analysis identified 11 lncRNAs ( $\lambda$  value=11), which were then used in the multivariate Cox regression. Finally, 8 lncRNAs for predicting HNSCC patient survival were identified, including



**Figure 1. Volcano plot of the differentially expressed mRNAs and lncRNAs between HNSCC and para-carcinoma tissues.** Red indicates high expression, and blue indicates low expression ( $|\log_2\text{FC}| > 1$  and  $P$  value  $< 0.05$ ). The Y axis represents adjusted  $P$  values, and the X axis represents  $\log_2\text{FC}$  values. The RNAs studied in this article have been marked in the figure. (A) Volcano plot of the differentially expressed lncRNAs. (B) Volcano plot of the differentially expressed mRNAs.

MIR4435-2HG, LINC02541, MIR9-3HG, AC104083.1, AC099850.4, PTOV1-AS2, AC245041.2, and AL357033.4.

### **The role of the 8-lncRNA signature in HNSCC's prognosis**

The signature score of these 8 lncRNAs based on regression coefficients in multivariable Cox analysis was calculated as follows: signature score =  $(0.36314 \times \text{expression of MIR4435-2HG}) + (0.23003 \times \text{expression of LINC02541}) - (0.22031 \times \text{expression of MIR9-3HG}) - (0.23426 \times \text{expression of AC104083.1}) + (0.21344 \times \text{expression of AC099850.4}) - (0.27806 \times \text{expression of PTOV1-AS2}) + (0.25463 \times \text{expression of AC245041.2}) - (0.31513 \times \text{expression of AL357033.4})$ . Taking the median signature score as the dividing point, the patients were divided into high signature-score group and low-signature score group. (Figure 2D). Patients in the high-signature score group had a significantly worse OS than those in the low-signature score group (Figure 2E). Besides, the AUCs were assessed for 3years (AUC = 0.740) and 5years (AUC = 0.706) survival (Figure 2F), and the results suggest that the signature can effectively evaluate the prognosis of HNSCC patients.

### **Development of a prediction model integrating the 8-lncRNA signature and clinical characteristics**

We evaluated age, sex, lymph node (N) status, metastasis (M) status, tumor stage (stage), and new events (which include locoregional disease, locoregional recurrence, new primary tumor, and distant metastasis) using KM analysis. Next, we found that age, metastasis, and new event play an important role in the prognosis of HNSCC (Figure 3).

The signature was regarded as a predictor for HNSCC patients. We identified the significant variables through univariate Cox analysis. The multivariate model includes candidate variables with a P-value < 0.1 in univariate analysis. (Figure 4A). Finally, the results (Table 1) suggested that the independent risk factors for HNSCC, including: stage, M stage, new event, and signature score. Moreover, we compared the multivariate Cox regression results of the two groups with and without the signature score. Surprisingly, the C-index of the signature score-containing group (0.72) was higher than that of the signature score-free group (0.71) (Supplementary Figure 1). The nomogram model was built by using the coefficients of the multivariable Cox regression model (Figure 4B). The AUC for 3-year survival reached 0.788 (Figure 4C). What's more, the calibration curve shows that concerning the probabilities of 3-year OS and 5-year OS, the predicted

values are consistent with the observed values (Figure 4D). Finally, we calculated the total risk score based on each predictor in the nomogram model. Kaplan-Meier analysis showed that patients in the high-risk group had a significantly worse OS than those in the low-risk group (Figure 4E).

### **Validate the signature in the internal and external validation cohorts**

To determine the stability of this nomogram; we performed a similar analysis process in the validation cohort (n = 199). Taking the median signature score as the dividing point, the patients were divided into the high signature-score group (n = 100) and the low signature-score group (n = 99). with the median signature score as the cut-off point (Figure 5A). The Kaplan-Meier OS curves suggested that patients in the high-signature score group had a significantly worse OS than those in the low-signature score group (Figure 5B). The AUC value for 3-year survival exhibited by the 8-lncRNA signature reached 0.779 (Figure 5C). Besides, the calibration curve shows that concerning the probabilities of 3-year OS and 5-year OS, the predicted values are consistent with the observed values (Figure 5E). What's more, using the same total risk score formula in the internal validation cohort, the Kaplan-Meier OS curves showed that the OS of patients with the high-risk score was significantly worse than that of patients with the low-risk score (Figure 5F). The AUC exhibited by the total risk score for 3-year survival reached 0.796 (Figure 5E).

We also validated the robustness of the signature in GSE65858 (n = 270), which had an AUC of 0.785 for 3-year OS (Figure 6A, 6C). Moreover, the OS of patients with high-signature score was worse than those of patients with the low-signature score (Figure 6B). The Kaplan-Meier OS curves manifested that patients in the high total risk score group had a significantly worse OS than patients in the low total risk score group (Figure 6F). Similarly, the calibration curve showed good agreement between the predicted and observed values (Figure 6E), and the AUC exhibited by the total risk score for 3-year survival reached 0.811 (Figure 6D).

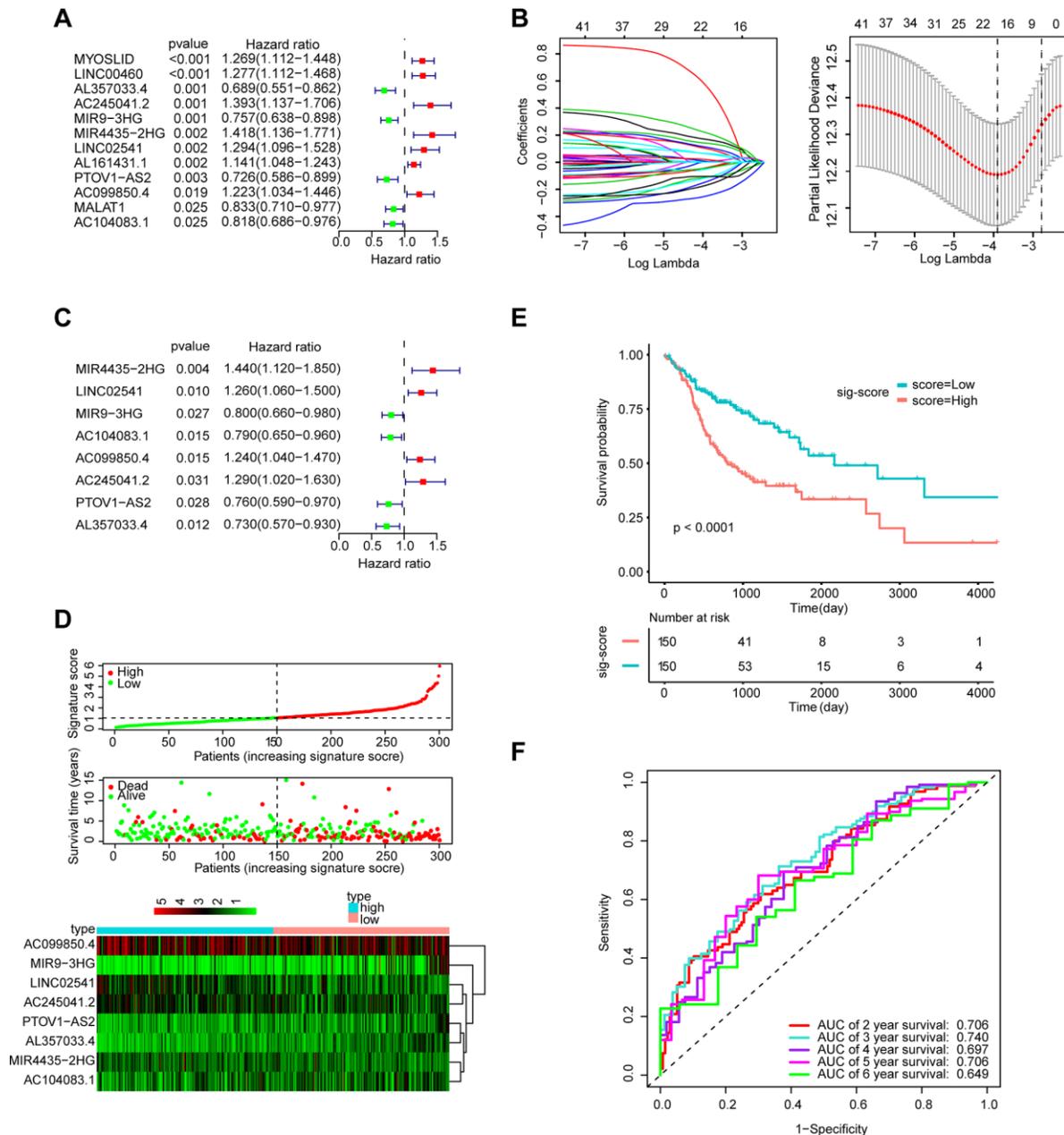
Furthermore, we measured the expression of these eight lncRNAs in 102 HNSCC samples by qRT-PCR (Figure 7A). The Kaplan-Meier curve showed that the OS of the patients with a high-signature score was significantly worse than that of the patients with a low-signature score (Figure 7B). The AUC for 3-year survival reached 0.942 (Figure 7C). The Kaplan-Meier curve showed that the OS of the patients with a high-risk score was significantly worse than that of the patients with a low-risk score

(Figure 7E). The calibration curve performs well (Figure 7D), and the AUC exhibited by the total risk score for 3-year survival reached 0.896 (Figure 7F).

Moreover, the above verification process was also performed for the entire TCGA-HNSC set (n=499) and revealed good results (Supplementary Figure 2).

## WGCNA

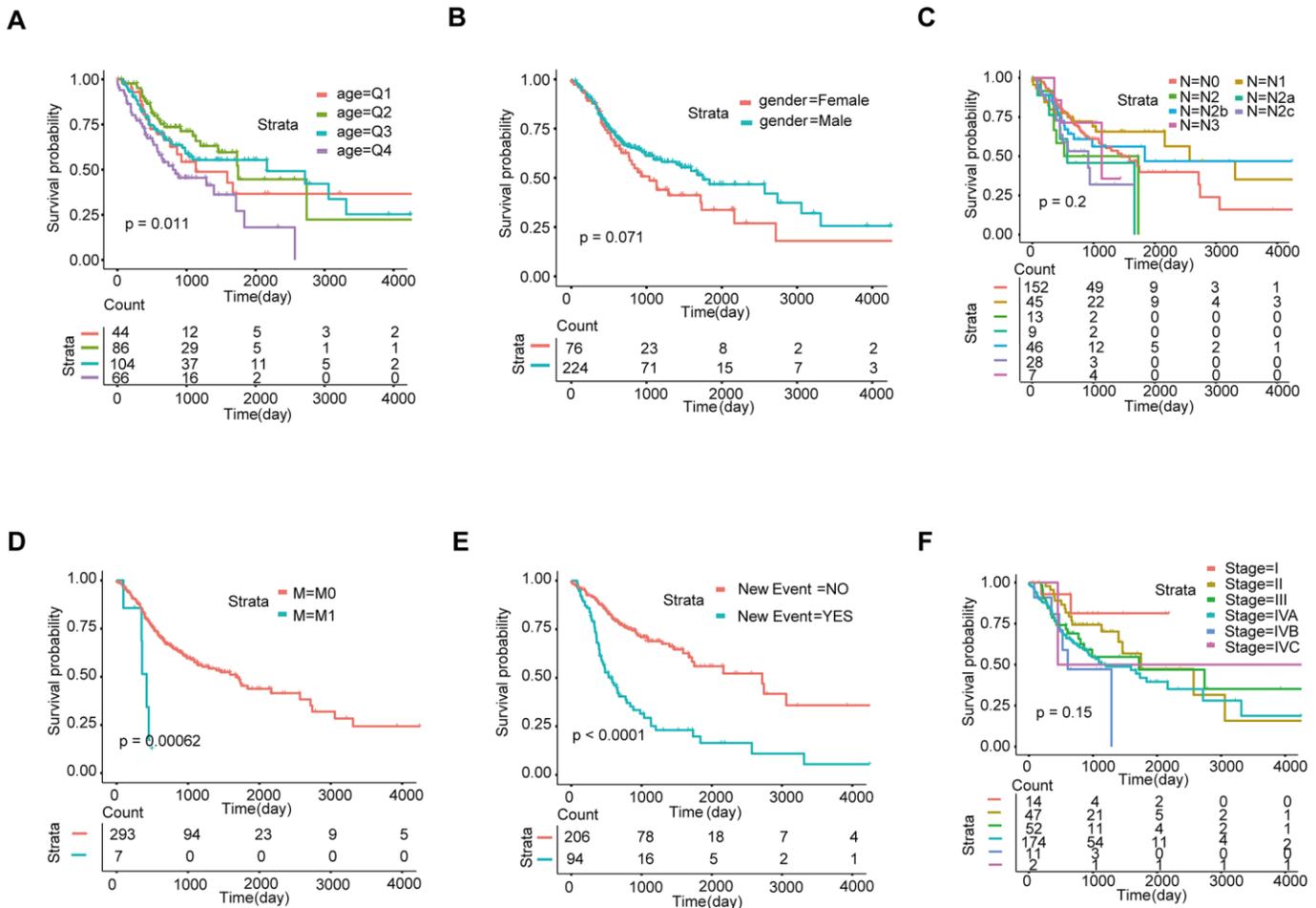
The gene co-expression system was established by WGCNA to screen the biologically significant gene modules related to the lncRNAs in the signature. To create a scale-free system, we set the soft threshold beta to 3 (Figure 8A). Besides, genes with similar patterns



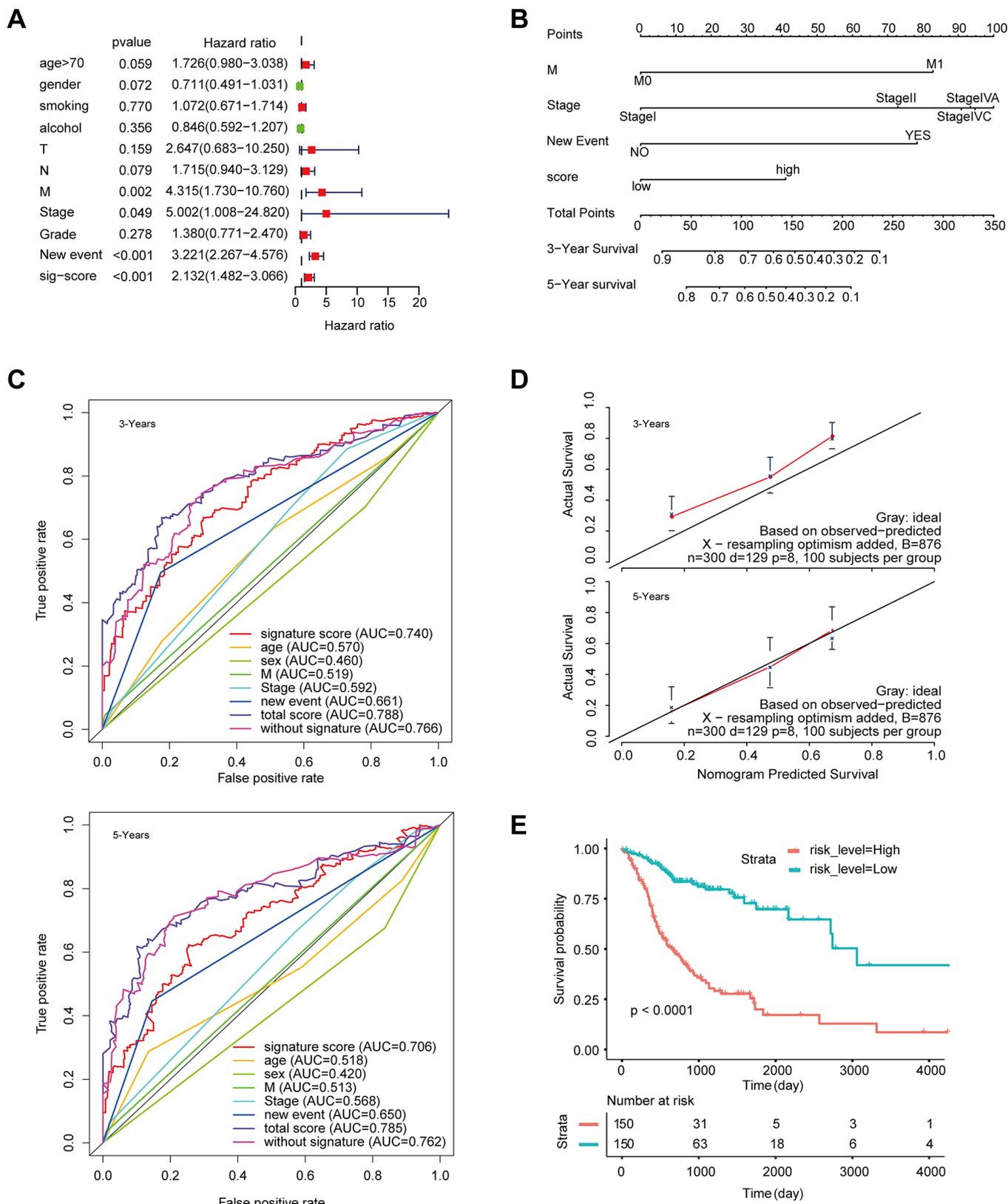
**Figure 2. Establishment and validation of the eight-lncRNA prognostic signature.** (A–C) The procedure of establishing the prognostic signature. (D) Correlation between the prognostic signature and the overall survival of patients in the TCGA cohort. The distribution of signature scores (top), survival time (middle) and lncRNA expression levels (bottom). The black dotted lines represent the median signature score cut-off dividing patients into the low- and high-signature groups. The red dots and lines represent the patients in the high-score group. The green dots and lines represent the patients in the low-score group. (E) Kaplan-Meier curves of OS based on the 8-lncRNA signature. (F) ROC curve analyses based on the 8-lncRNA signature.

were clustered in different modules (Figure 8B). The minimum cluster size was determined to be 30 per module. The gene modulus was determined by the dynamic shearing method. The module eigengene (ME) was calculated to explore the similarity of all modules (Figure 8C). Eigengenes were calculated to be correlated with clinical factors. Finally, a robust correlation between the gene significance and grade and signature score was identified (Figure 8D). The ten modules were clustered into two groups (Figure 8E). In order to evaluate the correlation between gene expression and survival time, we calculated the gene significance (Figure 9A). Then, we found that there was a strong correlation between the module members of the brown module and the genetic significance of OS. (cor-value = -0.47,  $P = 5.3e - 12$ ). The red module, whose hub gene contains MIR4435-2HG, was also negatively correlated with the OS (cor-value = -

0.2,  $P = 0.032$ ) (Figure 9B). Finally, we explore the GO term and KEGG pathway through functional enrichment analysis. (Figure 9C–9F). The results indicated that the biological processes (BP) of these genes mainly involved cell chemotaxis, leukocyte migration, immune response, cell-cell signaling, and so on. The results suggested that the molecular functions (MF) of these genes were related to actin binding, chemokine activity, chemokine receptor binding, ATPase binding, and so on. The results showed that the cellular components (CC) included collagen-containing extracellular matrix, plasma lipoprotein particle, growth cone and site of polarized growth. KEGG pathway functional enrichment showed that leukocyte transendothelial migration, cytokine-cytokine receptor interaction, cell adhesion molecules (CAMs), and the chemokine signaling pathway were mainly related to the genes in these modules.



**Figure 3. Screening of prognosis-related clinical characteristics by Kaplan-Meier analysis.** (A) Kaplan-Meier curves based on different age groups, where Q1, Q2, Q3, and Q4 represent quartiles. (B) Kaplan-Meier curves based on gender. (C) Kaplan-Meier curves based on different N stages. (D) Kaplan-Meier curves based on different M stages. (E) Kaplan-Meier curves based on new events. (F) Kaplan-Meier curves based on different tumor stages.



**Figure 4. Construction of a nomogram for overall survival prediction in HNSCC. (A)** Univariate and multivariate Cox regression analyses of clinical factors associated with overall survival. **(B)** The nomogram consists of M stage, new event, stage and the signature score based on the eight-lncRNA signature. **(C)** ROC curves according to the nomogram and lncRNA signature score. **(D)** Calibration curves of the nomogram for the estimation of survival rates at 3 and 5 years. **(E)** Kaplan-Meier curves of OS according to the total risk score.

**Table 1. The results of multivariate Cox analysis.**

	HR	Lower 95%CI	Upper 95%CI	P-value
<b>Age</b>				
<50y				
50-60y	0.825	0.450	1.511	0.533
60-70y	1.035	0.579	1.849	0.907
≥70	1.652	0.910	3.001	0.099
<b>sex</b>				
male vs female	0.696	0.462	1.049	0.083
<b>N</b>				
N0				
N1	0.706	0.406	1.228	0.218
N2	1.196	0.498	2.872	0.688
N2a	2.313	0.894	5.989	0.084
N2b	0.754	0.420	1.354	0.344
N2c	1.137	0.593	2.182	0.699
N3	0.388	0.114	1.319	0.129
<b>M</b>				
M1 vs M0	3.968	1.287	12.237	0.016*
<b>Stage</b>				
Stage I				
Stage II	2.535	0.573	11.216	0.220
Stage III	4.328	0.989	18.939	0.052
Stage IVA	4.547	1.078	19.178	0.039*
Stage IVB	5.015	0.956	26.299	0.057
Stage IVC	2.949	0.3231	26.922	0.338
<b>New event</b>				
yes vs no	3.032	2.081	4.418	<0.001***
<b>signature score</b>				
high vs low	1.904	1.304	2.780	<0.001***

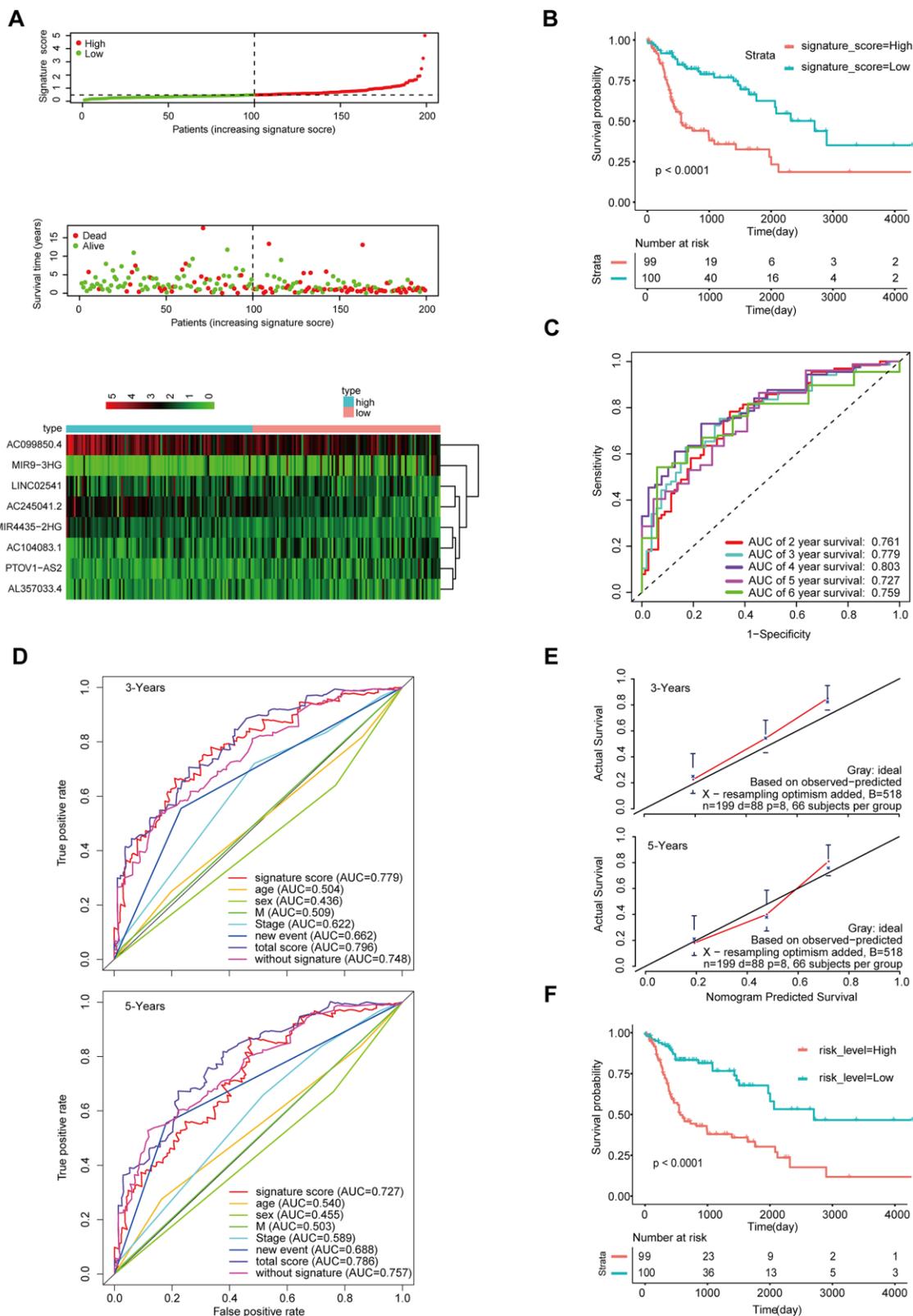
Abbreviations: HR, Hazard ratio; CI, Confidence interval; \*P<0.05; \*\* P<0.01; \*\*\* P<0.001.

We conducted a similar analysis process to estimate the correlation between gene expression and grade (Figure 10A). A strong correlation was found between the gene significance for grade and module membership in the turquoise module (which contains MIR9-3HG, AC099850.4 and PTOV1-AS2) (cor-value = 0.41, P = 7.2e - 23); the black module (which contains LINC02541) (cor-value = 0.35, P = 0.00047) and the red module (cor = 0.28, P = 0.0024) were both positively correlated with grade (Figure 10B). We constructed the lncRNA-mRNA network (weight>0.1) diagram of the hub lncRNAs in the turquoise module (Figure 10C). We also carried out functional enrichment analysis to explore the GO term and KEGG pathway (Figure 10D–10G). The results indicated that BP mainly involved cell proliferation, cell division, positive regulation of cell migration, and

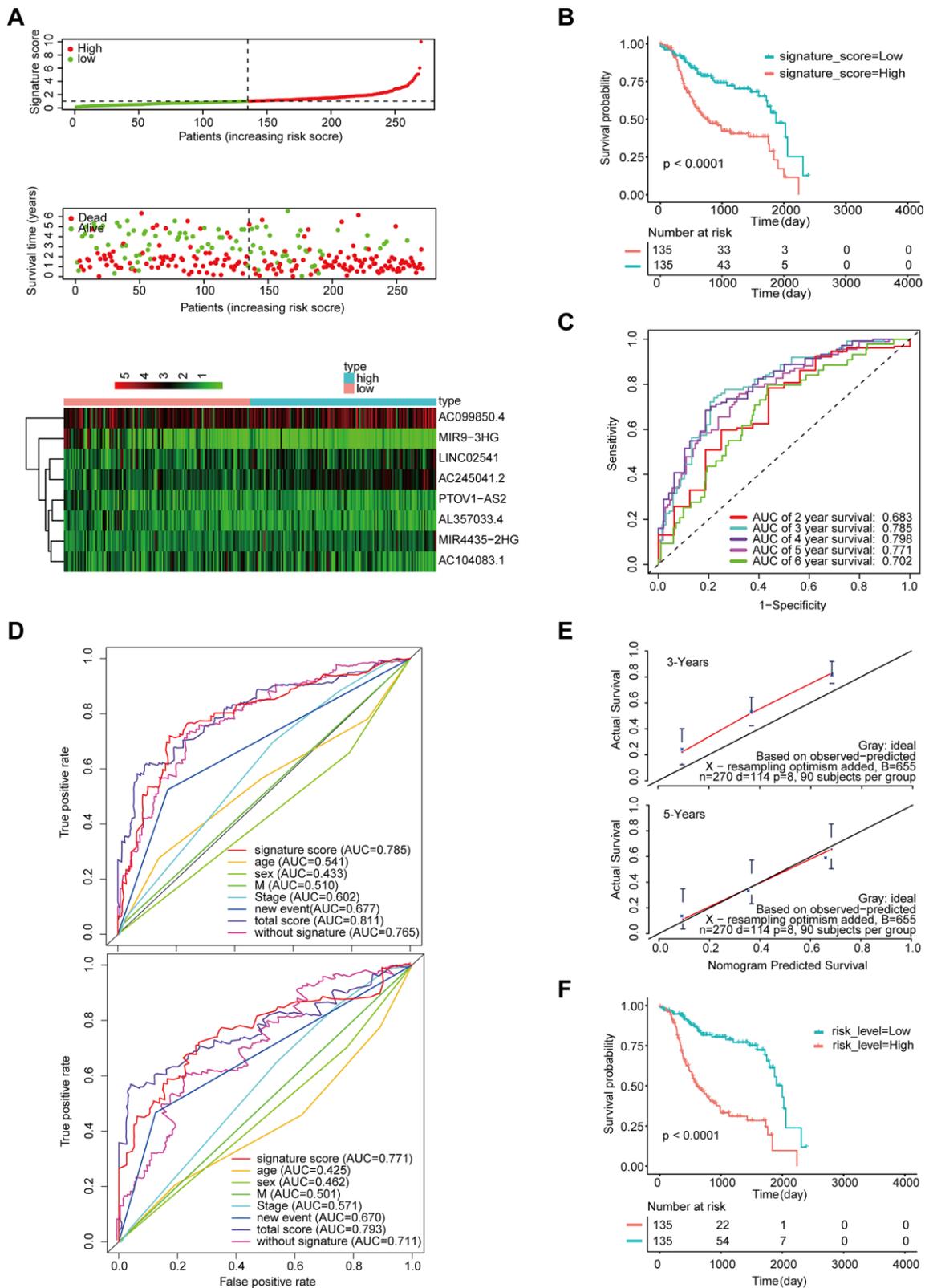
regulation of the cell cycle. The results showed that MF was related to catalytic activity, acting on DNA, protein binding, and DNA replication origin binding. The results showed that CC included proteinaceous extracellular matrix, chromosome, centromeric region, and extracellular matrix. Moreover, KEGG pathway functional enrichment showed that the cell cycle, the p53 signaling pathway, Cellular senescence, Mismatch repair, and DNA replication were mainly involved.

## DISCUSSION

Head and neck cancer ranks as the sixth leading malignancy worldwide, with almost 90% of cases classified as head and neck squamous cell carcinoma (HNSCC) [6]. Although the diagnosis and treatment



**Figure 5. Validation of the model by the internal validation set TCGA-HNSCC (n=199).** (A) Distribution of 8-lncRNA-based signature scores, lncRNA expression levels and patient survival durations in the internal validation set. (B) Kaplan-Meier curves of OS based on the 8-lncRNA signature. (C) ROC curve analyses based on the 8-lncRNA signature. (D) ROC curves according to the nomogram and lncRNA signature score. (E) Calibration curves of the nomogram for the estimation of survival rates at 3 and 5 years. (F) Kaplan-Meier curves of OS according to the total risk score.



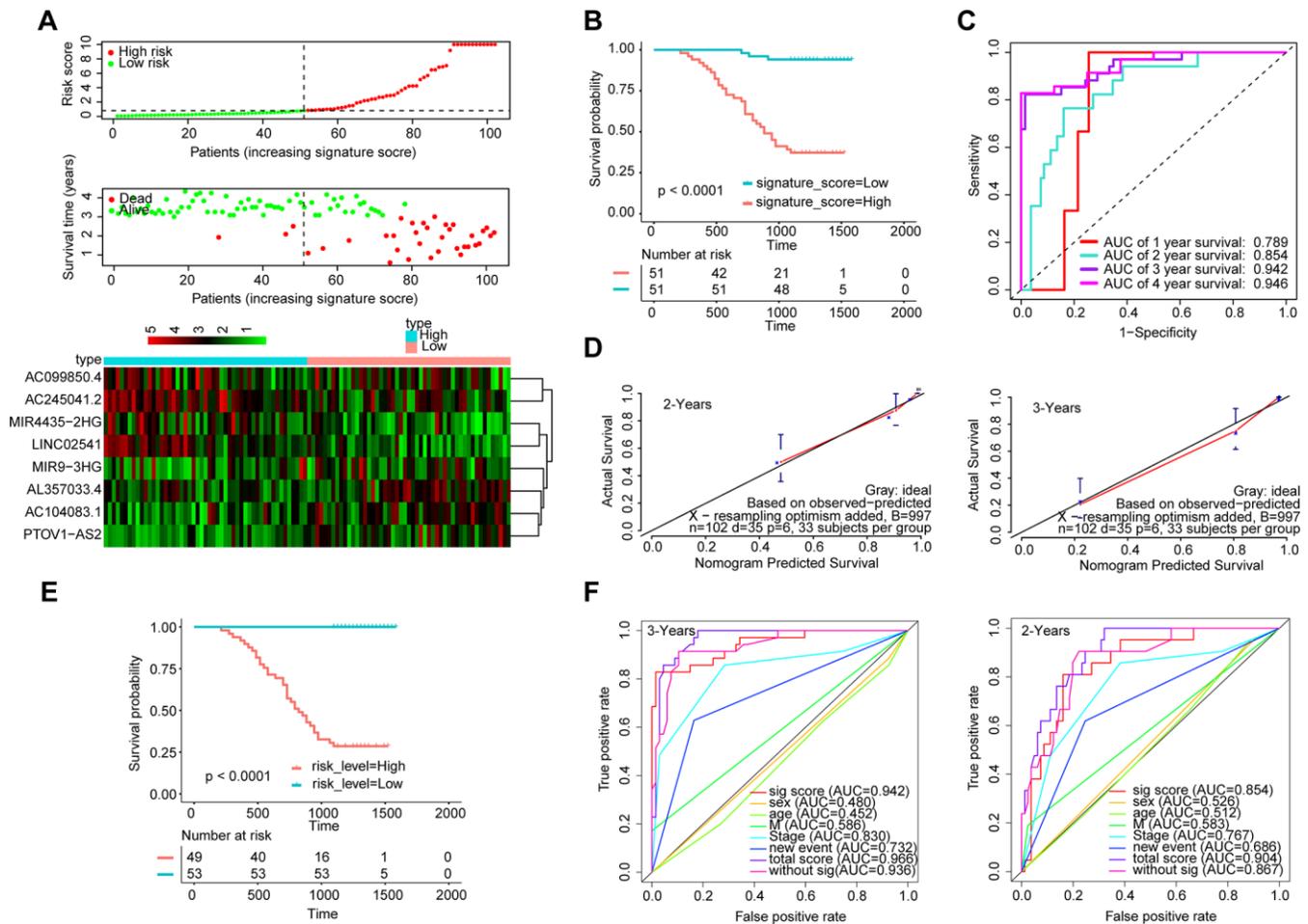
**Figure 6. Validation of the model by the external validation set GSE65858 (n=270).** (A) Distribution of 8-lncRNA-based signature scores, lncRNA expression levels and patient survival durations in the external validation set. (B) Kaplan-Meier curves of OS based on the 8-lncRNA signature. (C) ROC curve analyses based on the 8-lncRNA signature. (D) ROC curves according to the nomogram and lncRNA signature score. (E) Calibration curves of the nomogram for the estimation of survival rates at 3 and 5 years. (F) Kaplan-Meier curves of OS according to the total risk score.

have advanced in recent years, HNSCC still has a high incidence and mortality rate in developing countries [3]. Therefore, exploring diagnostic and prognostic biomarkers of HNSCC is urgent.

In the present study, we conducted a difference analysis between tumor and normal tissues in the TCGA-HNSC dataset. Through univariate Cox regression and LASSO analysis, we confirmed that lncRNAs were remarkably correlated with prognosis. Ultimately, eight lncRNAs (MIR4435-2HG, LINC02541, MIR9-3HG, AC104083.1, AC099850.4, PTOV1-AS2, AC245041.2, AL357033.4) were screened to compose a prognostic signature for HNSCC. A robust nomogram consisting of the signature, M, new event, and the stage was constructed for the prognostic prediction of HNSCC patients. Moreover, the AUC value of the signature-based nomogram was better than that of M, new event, and

the stage at 3 and 5 years. Besides, In this study, the AUC area analyzed by ROC curve is better than that of similar studies in most HNSCC [7, 8]. The results were verified in the internal validation set, the external validation set, and the qRT-PCR validation set of 102 HNSCC samples.

After a literature review, we found no research had been conducted about the mechanisms of the eight lncRNAs except MIR4435-2HG. MIR4435-2HG is the host gene of MIR4435-2, which is considered to be a biomarker in various cancers, such as oral squamous cell carcinoma [9], non-small-cell lung cancer cells [10], prostate carcinoma [11], gastric cancer [12], hepatocellular carcinoma [13] and lung cancer [14]. MIR4435-2HG promotes cancer cell migration and proliferation mainly by positively regulating TGF- $\beta$ 1 and activating the Wnt/ $\beta$ -catenin signaling pathway [9–14]. Interestingly, we found that



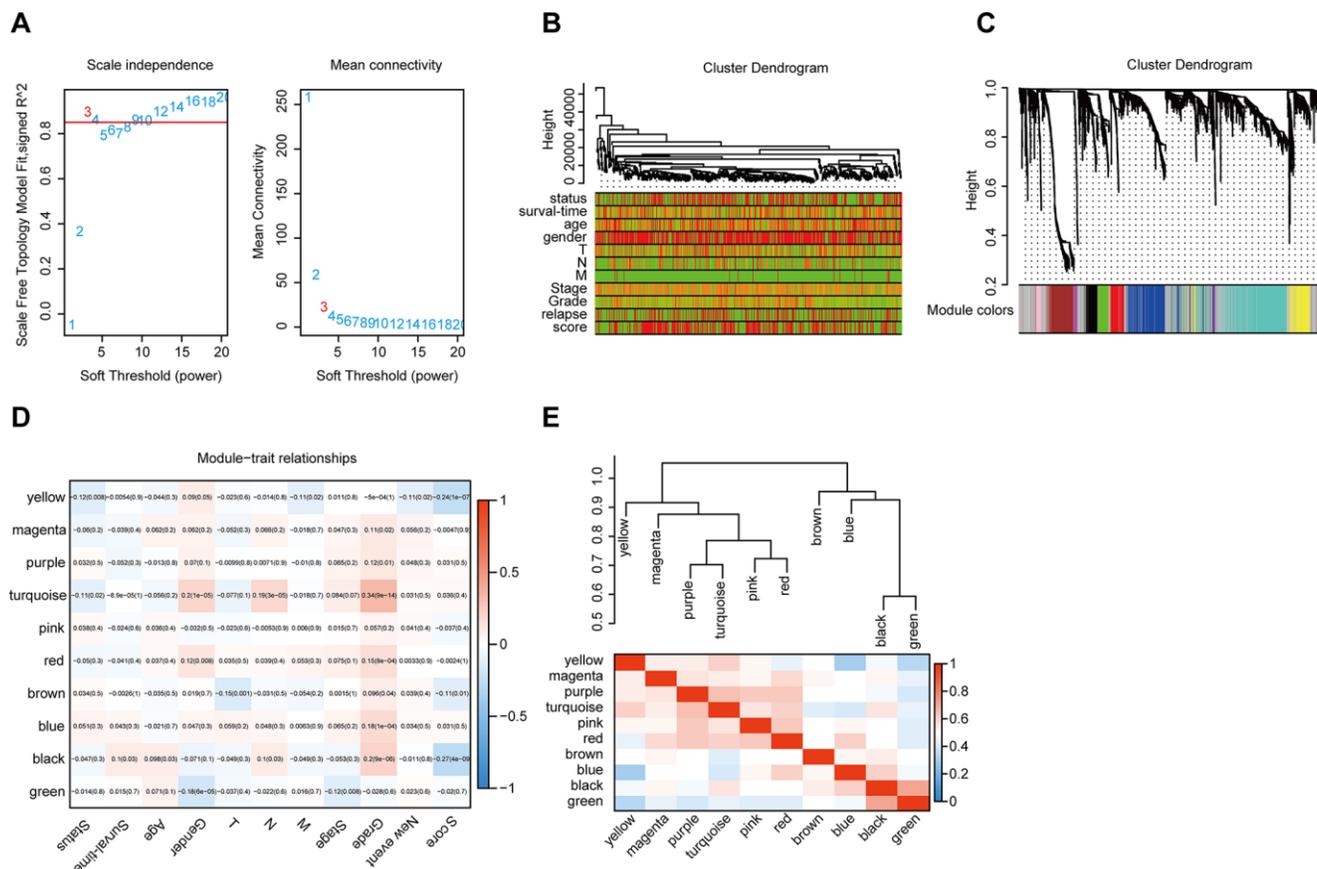
**Figure 7. Validation of the model by the qRT-PCR set (n=102).** (A) Distribution of 8-lncRNA-based signature scores, lncRNA expression levels and patient survival durations in the qRT-PCR validation set. (B) Kaplan-Meier curves of OS based on the 8-lncRNA signature. (C) ROC curve analyses based on the 8-lncRNA signature. (D) Calibration curves of the nomogram for the estimation of survival rates at 2 and 3 years. (E) Kaplan-Meier curves of OS according to the total risk score. (F) ROC curves according to the nomogram and lncRNA signature score.

the expression level of MIR4435-2HG was positively correlated with the risk score of patients with HNSCC in our study, which was consistent with the results of previously published literature. What is noteworthy is that HNSCC patients with high MIR4435-2HG expression appeared to have a poor prognosis.

To further clarify the mechanism of 8-lncRNAs affecting the survival of HNSCC patients, we selected AC099850.4 and AL357033.4, which showed the most differences in expression, for in vitro experiments. The results show that AL357033.4 overexpression could inhibit the proliferation of HNSCC cell FaDu and Hep-2. Moreover, knockdown of AC099850.4 could suppress the proliferation of FaDu and Hep-2 cells (Supplementary Figure 3). These results suggested that AL357033.4 and AC099850.4 may be involved in HNSCC proliferation and progression.

Nomograms have been developed in the majority of cancer types. For many cancers, the use of nomograms is more popular than traditional staging systems. [15–17], and thus, it has been proposed as an alternative or even a new standard [18–20]. In this study, a prognostic nomogram combining a lncRNA signature with clinical factors was established. Besides, our nomogram has better prediction accuracy than each factor alone.

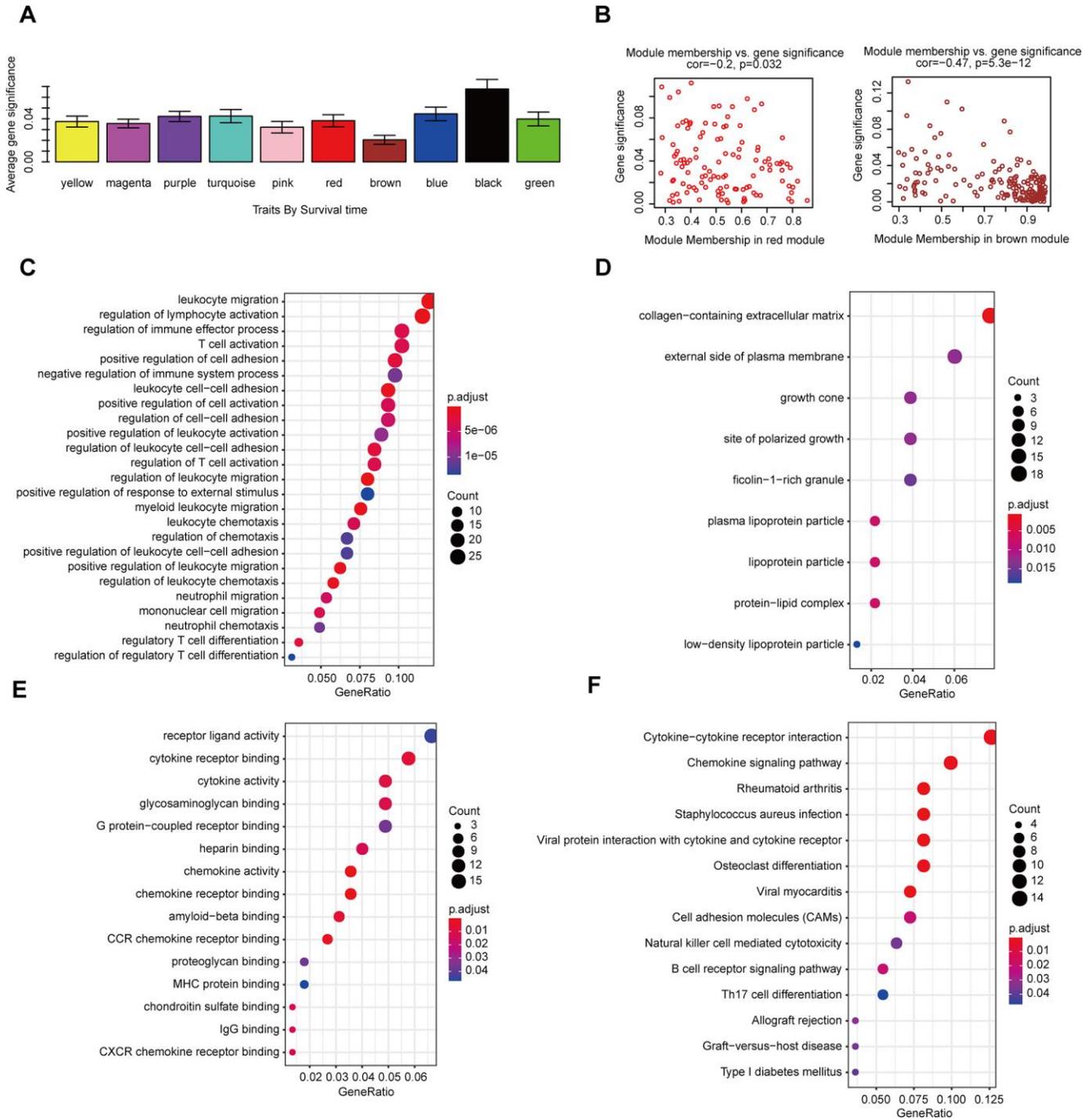
We used WGCNA and classified these genes into ten modules according to their expression profiles. Among these modules, we further pay attention to the gene modules that are highly related to various clinical features. Regarding survival time, the functional enrichment analysis indicated that the mRNAs associated with MIR4435-2HG were mainly associated with cellular signal transduction and the chemokine signaling pathway. Interestingly, aside from the feature of grade, the GO terms of the mRNAs that have a close connection with MIR4435-2HG,



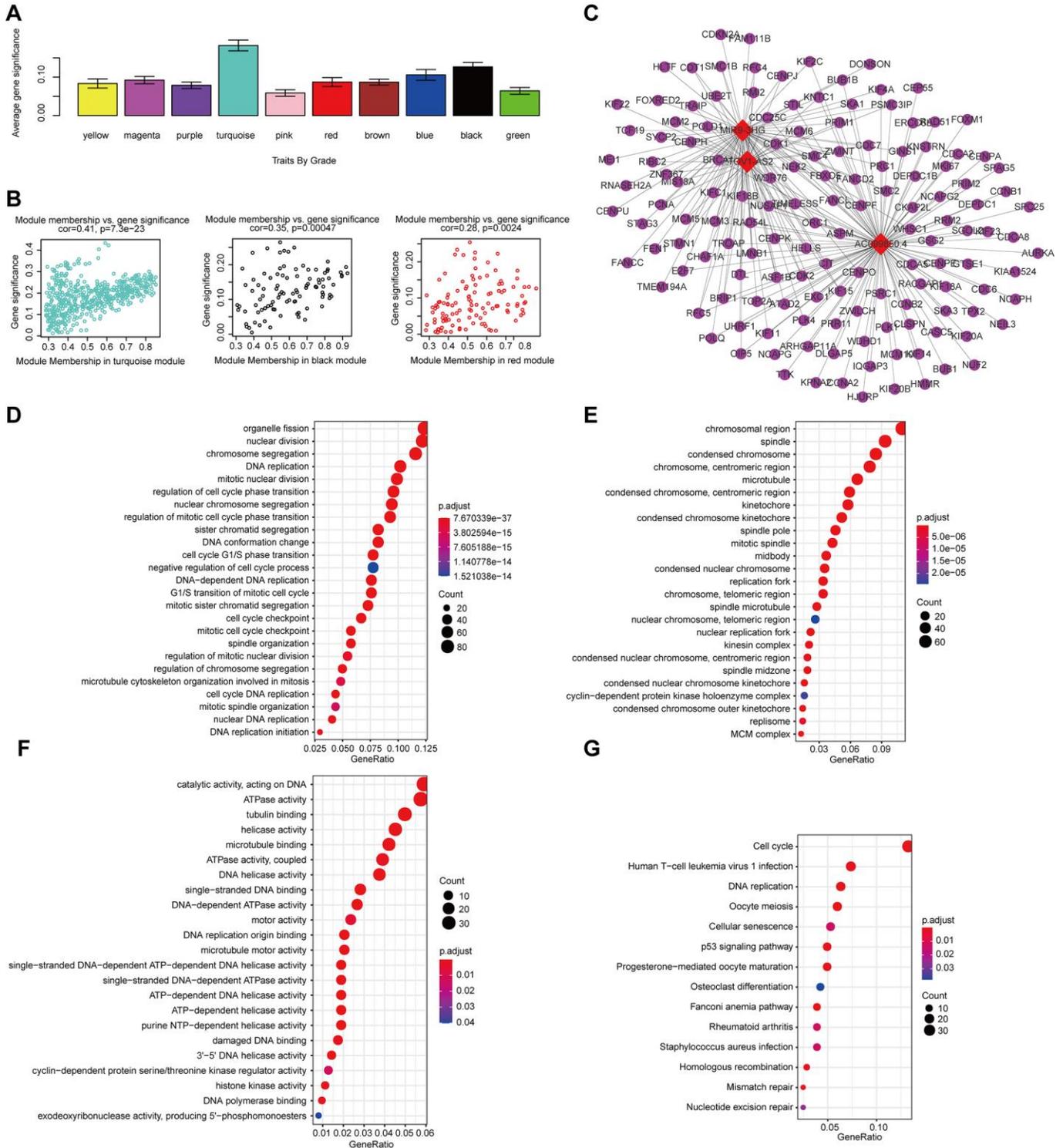
**Figure 8. WGCNA.** (A) Analysis of the scale-free topology model fit index for various soft-thresholding powers ( $\beta$ ) and the mean connectivity for various soft-thresholding powers. Overall, 3 was the most fitting power value. (B) Dendrogram of the genes and different clinical factors of HNSCC (survival time, survival status, sex, age, grade, stage, T stage, N stage, M stage, new event, signature score). (C) Dendrogram of the gene modules based on a dissimilarity measure. The branches of the cluster dendrogram correspond to the different gene modules. Each piece of the leaves on the cluster dendrogram corresponds to a gene. (D). Module-trait relationships. Heatmap of the correlation between module eigengenes and clinical characteristics of HNSCC. (E) Hierarchical clustering and heatmap of the hub gene network.

LINC02541, MIR9-3HG, AC099850.4, and PTOV1-AS2 were mainly focused on cell proliferation, cell division, and cell migration, while KEGG was mostly concentrated on tumor-related pathways such as the p53 signaling pathway, pathways in cancer, the cell cycle and ECM-receptor interaction.

In conclusion, we comprehensively evaluated the risk associated with clinical factors and lncRNAs and their contribution to prognosis and carried out risk stratification. The nomogram proposed in the present study objectively and accurately predicted the prognosis of patients with HNSCC.



**Figure 9. The correlation between the genes in the modules and survival time. (A)** Distribution of mean gene significance and standard deviation with survival time in the HNSCC modules. **(B)** Scatter plot of module eigengenes in red and brown modules. GO **(C–E)** and KEGG **(F)** pathway enrichment of eight modules. GO enrichment contains three categories: biological process **(C)**, cellular component **(D)** and molecular function **(E)**.



**Figure 10. The correlation between the genes in the modules and grade.** (A) Distribution of mean gene significance and standard deviation with grade in the HNSCC modules. (B) Scatter plot of the module eigengenes in the turquoise, black, and red modules. (C) The IncRNA-mRNA network (weight>0.1) of the hub IncRNAs in the turquoise module. Red and blue diamond shapes represent up- and downregulated IncRNAs, respectively. Purple circles represent mRNAs. GO (D–F) and KEGG (G) pathway enrichment of eight modules. GO enrichment contains three categories: biological process (D), cellular component (E) and molecular function (F).

## MATERIALS AND METHODS

### Data acquisition

The RNA-sequencing data of HNSCC patients were acquired from The Cancer Genome Atlas (TCGA) database (<http://cancergenome.nih.gov/>) and The Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) [21]. GSE65858 from GEO was conducted on the GPL10558 platform. Besides, we also followed 102 HNSCC patients in the Pathology Department and the Otolaryngology Department of Chengdu Third People's Hospital. The clinical features of patients with HNSCC are presented in Table 2.

### Differential analysis

The edgeR package in R software [22] were used to analyze the differentially expressed RNAs in HNSCC and adjacent normal tissues of the TCGA. Significantly expressed RNAs were identified by setting adjusted P values  $< 0.05$  and  $|\log_2FC$  (fold change)  $> 1$  ( $|\log_2FC > 1|$ ) and the adjusted FDR  $< .05$  [23, 24].

### The construction of the lncRNA-based prognostic signature

The prognostic value of 253 differentially expressed lncRNAs was first calculated in the univariate Cox analysis, and 41 lncRNAs with  $P < 0.05$  were identified as seed lncRNAs for LASSO regression analysis, which identified 11 lncRNAs (R 'glmnet', 'survival' packages). To determine the prognostic value of the lncRNAs, multivariate Cox regression was further performed using the R survival package based on each "significant" lncRNA identified in the above steps. A lncRNA with  $P < 0.05$  was defined as significant. The corresponding hazard ratios (HRs), 95% confidence intervals (CIs), and P-values were calculated.

### Prognostic evaluation using the 8-lncRNA signature

The signature score for each patient in the training group is calculated based on the formula (signature score =  $\exp\text{Gene1} \times \beta\text{Gene1} + \exp\text{Gene2} \times \beta\text{Gene2} + \exp\text{Gene} \times \beta\text{Gene}$  (where exp is the prognostic gene expression level and  $\beta$  represents the multivariate Cox regression model regression coefficient)). All samples are randomly divided into high- and low- signature score sets, with the median signature scores as the cut-off value [25]. The survival analysis of each group was evaluated through the Kaplan-Meier curve and the log-rank test. Receiver operating characteristic (ROC) curve analysis was employed to assess the specificity and sensitivity of the survival predictions according to the lncRNA signature scores (R package

"survivalROC"). A  $P$ -value  $< .05$  was considered significant.

### Development of a prediction model based on the 8-lncRNA signature and clinical characteristics

The gene signature score as a predictor for HNSCC patients was analyzed in the model. We determined the significant variables through univariate Cox regression analysis. The multivariate model includes candidate variables with a  $P$ -value  $< 0.1$  on univariate analysis. Finally, the multivariable Cox regression model began with the clinical candidate predictors as follows: stage, M stage, new event, and signature score. The nomogram model was built with the coefficients of the multivariable Cox regression model (using the R packages "rms", "Hmisc", "lattice", "Formula", and "foreign"). Then, we calculated the total risk score based on each predictor in the nomogram model and divided the HNSCC patients in the training and internal validation sets into two groups with the median risk score as the cut-off point. Kaplan-Meier curves and the log-rank test were used to compare the survival outcomes of the two groups. Receiver operating characteristic (ROC) curve analysis was employed to assess the accuracy and precision. of the survival predictions according to the total risk scores. Calibration curves were plotted to assess the calibration of the nomogram (R package "rms"). To quantify the discrimination performance of the nomogram, Harrell's C-index was measured. A  $P$ -value  $< .05$  was considered significant.

### Validation of the 8-lncRNA signature

The same risk formula was used to validate the internal validation set TCGA-HNSC ( $n = 199$ ), the entire set TCGA-HNSC ( $n = 499$ ), the external validation set GSE65858 ( $n = 270$ ) and the qRT-PCR set ( $n=102$ ).

### Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Total RNA was reverse-transcribed into cDNA with random primers using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Penzberg, Germany) following the manufacturer's instructions. The expression levels of the 8 lncRNAs were measured by qRT-PCR using FastStart Essential DNA Green Master mix (Roche, Penzberg, Germany) on a Roche LightCycler 480 (Roche, Penzberg, Germany). Relative expression was determined using inter-experiment normalization to GAPDH. All quantitative PCRs were conducted in triplicate. Divergent primers, rather than the more commonly used convergent primers, were designed for the lncRNAs. Primer specificity was verified using BLAST, with a single peak

**Table 2. The clinical features of patients with HNSCC.**

<b>Characteristics</b>	<b>Training dataset TCGA-HNSC (n=300)</b>	<b>Validation dataset TCGA-HNSC (n=199)</b>	<b>Validation dataset GSE65858 (n=270)</b>
<b>Age (y)</b>			
< 50	44	31	41
50-60	86	59	112
60-70	104	65	64
> 70	66	44	43
<b>Gender</b>			
Male	224	142	223
Female	76	57	47
<b>Survival status</b>			
Alive	171	111	94
Dead	129	88	176
<b>T</b>			
T1	20	14	35
T2	85	63	80
T3	83	49	58
T4	17	9	-
T4a	91	61	90
T4b	4	3	7
<b>N</b>			
N0	152	97	94
N1	45	38	32
N2	13	7	-
N2a	9	7	11
N2b	46	30	66
N2c	28	15	55
N3	7	5	12
<b>M</b>			
M0	293	192	263
M1	7	7	7
<b>Stage</b>			
I	14	11	18
II	47	33	37
III	52	38	37
IVA	174	114	155
IVB	11	2	16
IVC	2	1	7
<b>Grade</b>			
G1	31	32	-
G2	191	114	-
G3	71	52	-
G4	7	1	-
<b>New Event</b>			
Yes	94	74	133
No	206	125	137

in the melting curve indicating the generation of a specific product. Three experimental replicates were performed for each sample. Primers used in the study were presented in Supplementary Table 2.

### Construction of a weighted gene coexpression network

The procedure of WGCNA [26] included identifying the gene expression similarity matrix, adjacency matrix, and co-expression network. We set the cut-off as a Person correlation coefficient  $> 0.9$  and  $P < 0.001$  to screen gene coexpression with lncRNAs. Then, differentially expressed gene (DEG) analysis was performed among these genes, and we used the expression matrix composed of 4150 differential genes and the above 8 lncRNAs as input files. The power value of the adjacent matrix soft threshold is determined to be 9 to meet the scale-free topology standard. Hierarchical clustering analysis based on average linkage used the dynamic tree cut method for branch cutting (deep split = 2, cut height = 0.25, minimum cluster size = 30). If the similarity of the modules is  $> 0.9$ , they are merged. Based on the level of expression of each gene in each sample, we calculated the correlation between the genes in these modules and the individual phenotypes to measure the correlation between the gene and the phenotype (gene significance). The associations between the modules and variables were assessed to select the relevant modules. The lncRNA-mRNA network visualization was performed via Cytoscape software version 3.7.2 (<https://cytoscape.org/>) [16].

### Module function annotation

The enrichment analysis was conducted by DAVID [version 6.8] (<https://david.ncifcrf.gov>) [27] GO consists of three parts: biological processes (BP), molecular function (MF), and cellular composition (CC). Besides, all important GO or KEGG terms or genes are filtered into the meaning of  $P < .05$  and at least two mRNAs associated.

### Ethics statement

As the data (TCGA and GEO datasets) are publicly available, no ethical approval was required.

### Abbreviations

HNSCC: head and neck squamous carcinoma; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; lncRNAs: Long noncoding RNAs; ROC: receiver operating characteristic; OS: overall survival; AUC: the area under the curve; WGCNA: weighted

gene coexpression network; LASSO: least absolute shrinkage and selection operator; KEGG: Kyoto Encyclopedia of Genes and Genomes; qRT-PCR: Real-time quantitative reverse transcription polymerase chain reaction.

### AUTHOR CONTRIBUTIONS

M.-R, L.-Y.J and Z.-T.T conceived the project and designed the experiments. M.-R wrote the manuscript. C.-Y.Y and X.-L. carried out the statistical analysis and conducted the experiments. L.-Y.J. and Z.-T.T. contributed to manuscript revision. All authors provided suggestions during manuscript preparation and read the final version.

### ACKNOWLEDGMENTS

The authors would like to thank the efforts of staff in the Pathology Department and Otolaryngology Department of Chengdu Third People's Hospital.

### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

### FUNDING

This work was supported by grants from the National Natural Science Foundation of China (81502075) and the Foundation of Science and Technology of Sichuan Province (2019YJ0635). The funders had no role in the study design and implementation.

### REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136:E359–86. <https://doi.org/10.1002/ijc.29210> PMID:25220842
2. Marur S, Forastiere AA. Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment. *Mayo Clin Proc*. 2016; 91:386–96. <https://doi.org/10.1016/j.mayocp.2015.12.017> PMID:26944243
3. Mehanna H, Paleri V, West CM, Nutting C. Head and neck cancer—part 1: epidemiology, presentation, and prevention. *BMJ*. 2010; 341:c4684. <https://doi.org/10.1136/bmj.c4684> PMID:20855405
4. Payne K, Spruce R, Beggs A, Sharma N, Kong A, Martin T, Parmar S, Praveen P, Nankivell P, Mehanna H.

- Circulating tumor DNA as a biomarker and liquid biopsy in head and neck squamous cell carcinoma. *Head Neck*. 2018; 40:1598–604.  
<https://doi.org/10.1002/hed.25140>  
PMID:29542214
5. Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. *Mayo Clin Proc*. 2008; 83:489–501.  
<https://doi.org/10.4065/83.4.489> PMID:18380996
  6. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018; 68:7–30.  
<https://doi.org/10.3322/caac.21442>  
PMID:29313949
  7. Wang P, Jin M, Sun CH, Yang L, Li YS, Wang X, Sun YN, Tian LL, Liu M. A three-lncRNA expression signature predicts survival in head and neck squamous cell carcinoma (HNSCC). *Biosci Rep*. 2018; 38:BSR20181528.  
<https://doi.org/10.1042/BSR20181528>  
PMID:30355656
  8. Liu G, Zheng J, Zhuang L, Lv Y, Zhu G, Pi L, Wang J, Chen C, Li Z, Liu J, Chen L, Cai G, Zhang X. A prognostic 5-lncRNA expression signature for head and neck squamous cell carcinoma. *Sci Rep*. 2018; 8:15250.  
<https://doi.org/10.1038/s41598-018-33642-1>  
PMID:30323196
  9. Shen H, Sun B, Yang Y, Cai X, Bi L, Deng L, Zhang L. MIR4435-2HG regulates cancer cell behaviors in oral squamous cell carcinoma cell growth by upregulating TGF- $\beta$ 1. *Odontology*. 2020; 108:553–59.  
<https://doi.org/10.1007/s10266-020-00488-x>  
PMID:32016787
  10. Yang M, He X, Huang X, Wang J, He Y, Wei L. LncRNA MIR4435-2HG-mediated upregulation of TGF- $\beta$ 1 promotes migration and proliferation of nonsmall cell lung cancer cells. *Environ Toxicol*. 2020; 35:582–90.  
<https://doi.org/10.1002/tox.22893> PMID:31875359
  11. Zhang H, Meng H, Huang X, Tong W, Liang X, Li J, Zhang C, Chen M. lncRNA MIR4435-2HG promotes cancer cell migration and invasion in prostate carcinoma by upregulating TGF- $\beta$ 1. *Oncol Lett*. 2019; 18:4016–21.  
<https://doi.org/10.3892/ol.2019.10757>  
PMID:31516603
  12. Wang H, Wu M, Lu Y, He K, Cai X, Yu X, Lu J, Teng L. LncRNA MIR4435-2HG targets desmoplakin and promotes growth and metastasis of gastric cancer by activating Wnt/ $\beta$ -catenin signaling. *Aging (Albany NY)*. 2019; 11:6657–73.  
<https://doi.org/10.18632/aging.102164>  
PMID:31484163
  13. Kong Q, Liang C, Jin Y, Pan Y, Tong D, Kong Q, Zhou J. The lncRNA MIR4435-2HG is upregulated in hepatocellular carcinoma and promotes cancer cell proliferation by upregulating miRNA-487a. *Cell Mol Biol Lett*. 2019; 24:26.  
<https://doi.org/10.1186/s11658-019-0148-y>  
PMID:30988676
  14. Qian H, Chen L, Huang J, Wang X, Ma S, Cui F, Luo L, Ling L, Luo K, Zheng G. The lncRNA MIR4435-2HG promotes lung cancer progression by activating  $\beta$ -catenin signalling. *J Mol Med (Berl)*. 2018; 96:753–64.  
<https://doi.org/10.1007/s00109-018-1654-5>  
PMID:29872866
  15. Wang Y, Du L, Yang X, Li J, Li P, Zhao Y, Duan W, Chen Y, Wang Y, Mao H, Wang C. A nomogram combining long non-coding RNA expression profiles and clinical factors predicts survival in patients with bladder cancer. *Aging (Albany NY)*. 2020; 12:2857–79.  
<https://doi.org/10.18632/aging.102782>  
PMID:32047140
  16. Li W, Liu J, Zhao H. Identification of a nomogram based on long non-coding RNA to improve prognosis prediction of esophageal squamous cell carcinoma. *Aging (Albany NY)*. 2020; 12:1512–26.  
<https://doi.org/10.18632/aging.102697>  
PMID:31978896
  17. Gu J, Zhang X, Miao R, Ma X, Xiang X, Fu Y, Liu C, Niu W, Qu K. A three-long non-coding RNA-expression-based risk score system can better predict both overall and recurrence-free survival in patients with small hepatocellular carcinoma. *Aging (Albany NY)*. 2018; 10:1627–39.  
<https://doi.org/10.18632/aging.101497>  
PMID:30018179
  18. Sternberg CN. Are nomograms better than currently available stage groupings for bladder cancer? *J Clin Oncol*. 2006; 24:3819–20.  
<https://doi.org/10.1200/JCO.2006.07.1290>  
PMID:16864852
  19. Mariani L, Miceli R, Kattan MW, Brennan MF, Colecchia M, Fiore M, Casali PG, Gronchi A. Validation and adaptation of a nomogram for predicting the survival of patients with extremity soft tissue sarcoma using a three-grade system. *Cancer*. 2005; 103:402–08.  
<https://doi.org/10.1002/cncr.20778> PMID:15578681
  20. Wang L, Hricak H, Kattan MW, Chen HN, Scardino PT, Kuroiwa K. Prediction of organ-confined prostate cancer: incremental value of MR imaging and MR spectroscopic imaging to staging nomograms. *Radiology*. 2006; 238:597–603.  
<https://doi.org/10.1148/radiol.2382041905>  
PMID:16344335
  21. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30:207–10.

- <https://doi.org/10.1093/nar/30.1.207>  
PMID:[11752295](https://pubmed.ncbi.nlm.nih.gov/11752295/)
22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139–40.  
<https://doi.org/10.1093/bioinformatics/btp616>  
PMID:[19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/)
23. Fang SM, Hu BL, Zhou QZ, Yu QY, Zhang Z. Comparative analysis of the silk gland transcriptomes between the domestic and wild silkworms. *BMC Genomics*. 2015; 16:60.  
<https://doi.org/10.1186/s12864-015-1287-9>  
PMID:[25887670](https://pubmed.ncbi.nlm.nih.gov/25887670/)
24. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990; 9:811–18.  
<https://doi.org/10.1002/sim.4780090710>  
PMID:[2218183](https://pubmed.ncbi.nlm.nih.gov/2218183/)
25. Ma B, Li Y, Ren Y. Identification of a 6-lncRNA prognostic signature based on microarray re-annotation in gastric cancer. *Cancer Med*. 2020; 9:335–49.  
<https://doi.org/10.1002/cam4.2621>  
PMID:[31743579](https://pubmed.ncbi.nlm.nih.gov/31743579/)
26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9:559.  
<https://doi.org/10.1186/1471-2105-9-559>  
PMID:[19114008](https://pubmed.ncbi.nlm.nih.gov/19114008/)
27. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57.  
<https://doi.org/10.1038/nprot.2008.211>  
PMID:[19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)

## SUPPLEMENTARY MATERIALS

### Supplementary Methods

#### Cell line and Culture

FaDu and Hep-2 human laryngeal carcinoma cells were routinely cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco Corporation, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS; Invitrogen, Carlsbad, CA, USA), 100 U/mL penicillin G, and 100 U/mL streptomycin (Gibco, Carlsbad, CA, USA) in a humidified atmosphere of 95% air and 5% CO<sub>2</sub> at 37 °C. The medium was changed every 3 days.

#### Proliferation Assay

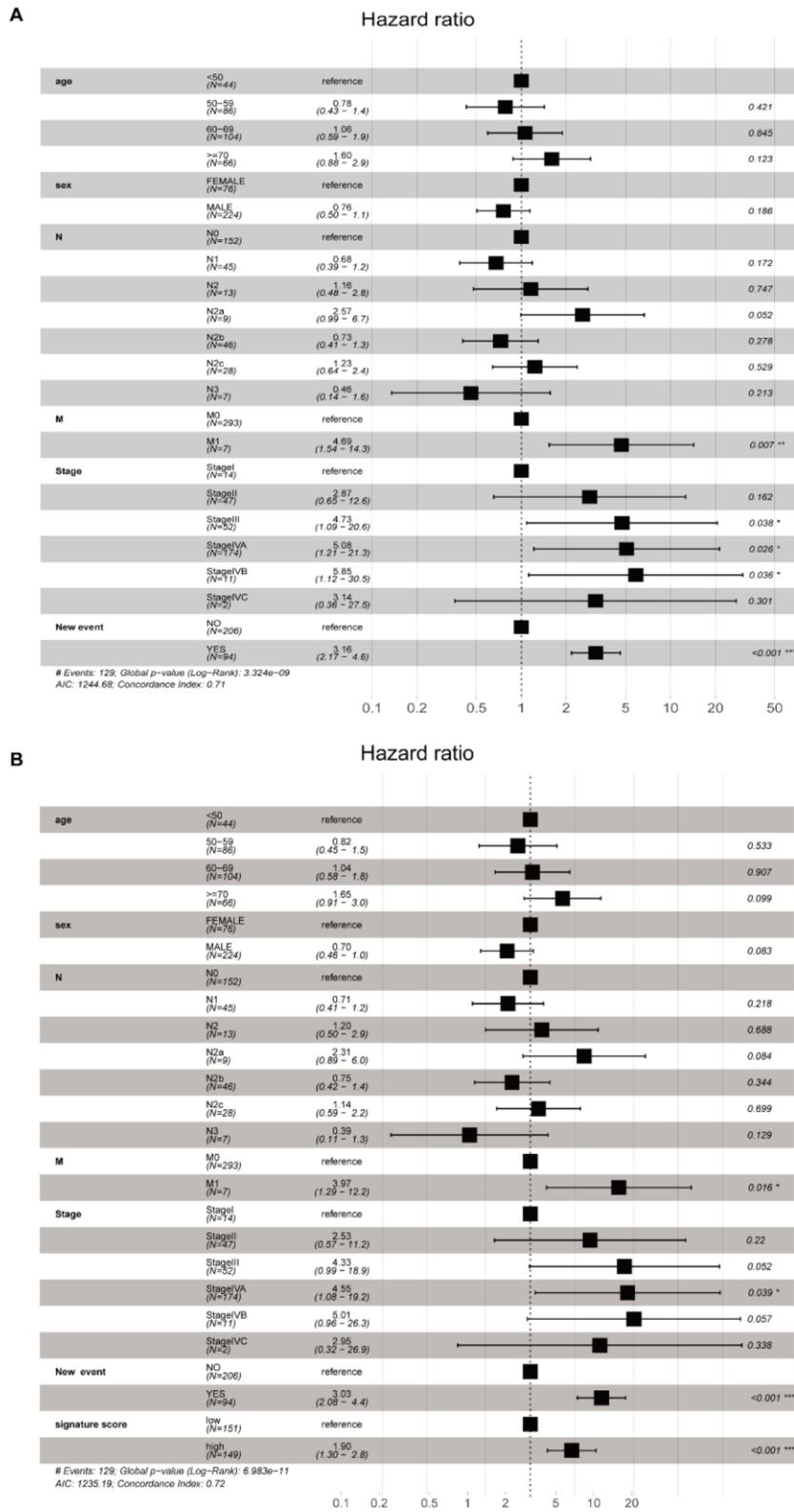
FaDu and Hep-2 cells in logarithmic growth phase were seeded in 96-well microplates with  $1 \times 10^4$  each well. The proliferation of FaDu and Hep-2 cells were assessed by using CCK-8 assay. After 24, 48, 72, and

96 h, cells were treated with 10  $\mu$ L of CCK-8 reagent (Dojindo Molecular Technologies, Kumamoto, Japan) and incubated at 37 °C for 1 h. An automatic microtiter plate reader was set to zero according to the control wells. The absorbance (A) of each well was measured at a wavelength of 450 nm.

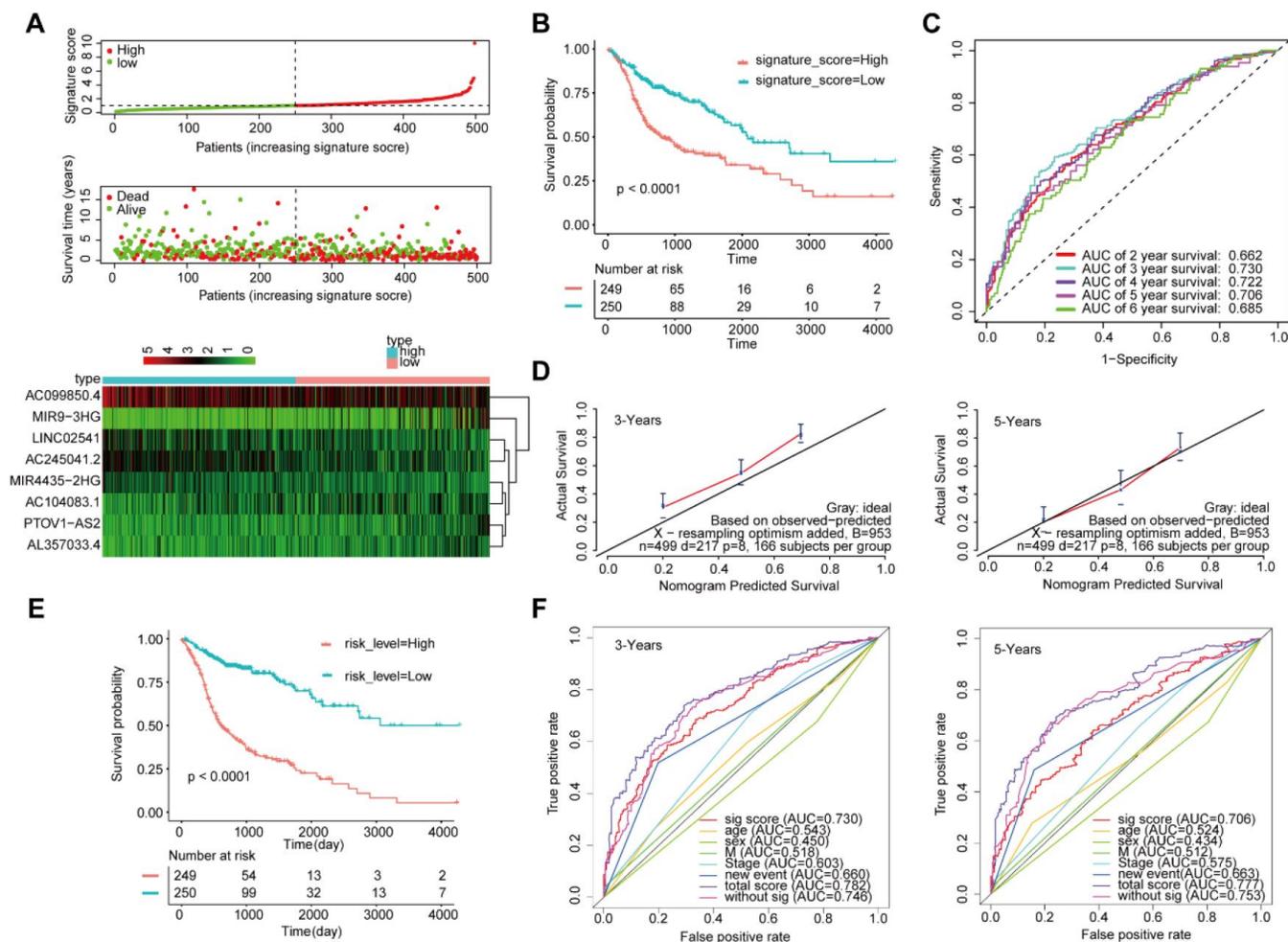
#### lncRNA knockdown and overexpression

Small interfering RNAs (siRNAs) of AC099850.4 and overexpression of AL357033.4 plasmid were constructed by GenePharma (Shanghai, China). Cells were transfected using Lipofectamine 2000 (Invitrogen, CA, USA). After 48 h of siRNA knockdown or plasmid transfection, lncRNA expression was measured using qRT-PCR.

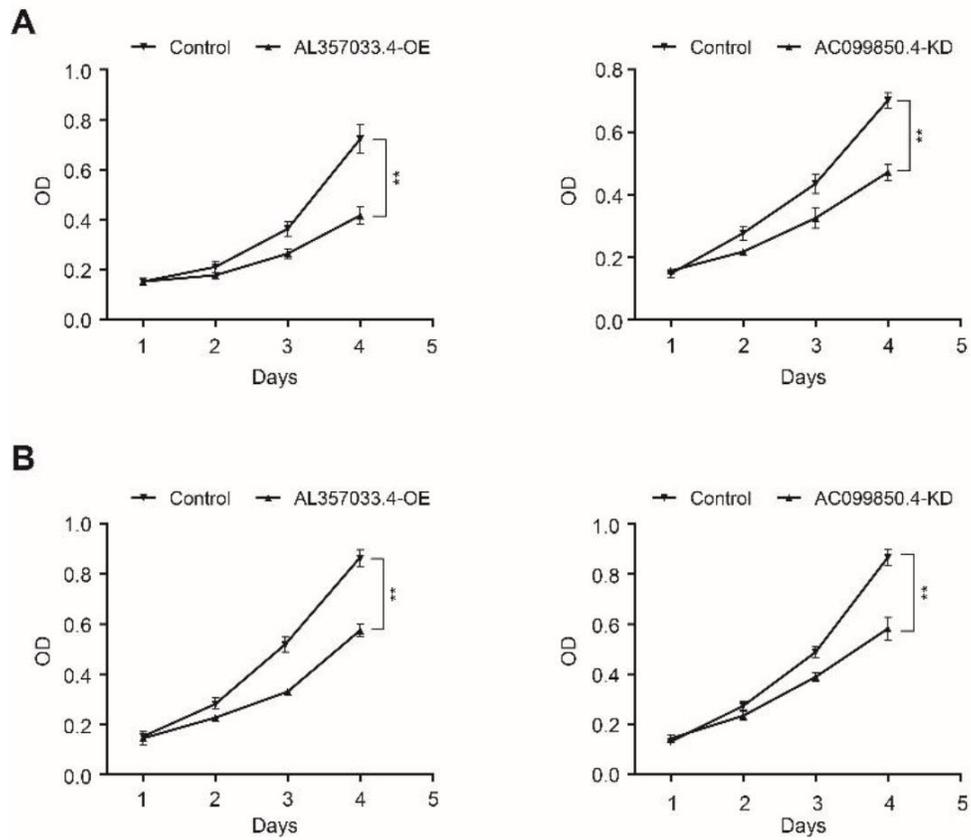
Supplementary Figures



Supplementary Figure 1. Comparison of the multivariate Cox regression results of the two groups with and without the signature score. (A) The group without the signature score. (B) The group with the signature score.



**Supplementary Figure 2. Validation of the model by the entire TCGA-HNSCC set (n=499).** (A) Distribution of 8-lncRNA-based signature scores, lncRNA expression levels and patient survival durations in the entire set. (B) Kaplan-Meier curves of OS based on the 8-lncRNA signature. (C) ROC curve analyses based on the 8-lncRNA signature. (D) Calibration curves of the nomogram for the estimation of survival rates at 3 and 5 years. (E) Kaplan-Meier curves of OS according to the total risk score. (F) ROC curves according to the nomogram and lncRNA signature score.



**Supplementary Figure 3.** Proliferation of FaDu (A) and Hep-2 (B) cells was analyzed using CCK-8 assay following lncRNAs overexpression or knockdown.

## Supplementary Table

Supplementary Table 1. Primers used in the study.

Gene	Sequence	Product length (bp)
MIR9-3HG	5'-GCCAGGCTTACTGTCTCTGG-3' 3'-GAGCCAAGACCTATCCACCA-5'	170
AC099850.4	5'-TCACCATGCCTGGGTAATTT-3' 3'-TGCCAAGGAATCTCTGAAGTC-5'	151
MIR4435-2HG	5'-CATTTTTCCCTGCTCTGCTC-3' 3'-ATCAAGGAGGGCTCATGTTG-5'	151
AC104083.1	5'-AGCCTTCCAGATTGTGAGGA-3' 3'-ACATTGGGCTGTCCA ACTTC-5'	209
AC245041.2	5'-AGGAGCCAGACATGTGGAGT-3' 3'-CAGAGCTTCTGCTGTGCAGT-5'	208
AL357033.4	5'-TCGGTCATCAGTTCATCAA-3' 3'-TAGCAGTGAACGCAGAGGTG-5'	166
LINC02541	5'-CCAGGCTGGAGTATGATGGT-3' 3'-TGGTGAAACCCCGTCTCTAC-5'	160
PTOV1-AS2	5'-CTGTCCCTGAGGAGTGGAGA-3' 3'-AGCGGTGGAGATGACGTTTC-5'	178
GAPDH	5'-GAAAGCCTGCCGGTGACTAA-3' 3'-GCCCAATACGACCAAATCAGAG-5'	150

# Immune-related biomarker risk score predicts prognosis in prostate cancer

Zezen Liu<sup>1,\*</sup>, Jiehui Zhong<sup>1,\*</sup>, Chao Cai<sup>1,\*</sup>, Jianming Lu<sup>2</sup>, Wenqi Wu<sup>1</sup>, Guohua Zeng<sup>1</sup>

<sup>1</sup>Department of Urology, Minimally Invasive Surgery Center, Guangdong Key Laboratory of Urology, Guangzhou Urology Research Institute, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China

<sup>2</sup>Guangdong Key Laboratory of Clinical Molecular Medicine and Diagnostics, Department of Urology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China

\*Equal contribution

**Correspondence to:** Guohua Zeng; email: [gzyzgh@vip.sina.com](mailto:gzyzgh@vip.sina.com)

**Keywords:** prostate cancer, PLK1, immune-related genes, immune microenvironment

**Received:** June 6, 2020

**Accepted:** July 25, 2020

**Published:** November 10, 2020

**Copyright:** © 2020 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

In this study, we constructed a model using a Cox proportional hazards model based on the expression of eight immune-related genes that were associated with prognosis in prostate cancer: EDNRB, ANGPTL2, TNFSF15, TNFRSF10D, EDN2, BMP2, NLRP14, and PLK1. We then identified associations between risk scores calculated with the model, tumor microenvironment characteristics, and immune cell infiltration. Prostate cancer patients in the high score group had poorer prognoses, and validation with the external GSE54460 dataset confirmed that the scoring model predicted biochemical recurrence with AUC values of 0.749 at 1 year, 0.804 at 3 years, and 0.774 at 5 years. Proportions of infiltrated M2 macrophages and regulatory T cells were increased in the high risk group, while CD8<sup>+</sup> T cells were increased in the low risk group. Network analysis revealed that PLK1 may be a key regulator of the immune-suppressive microenvironment in prostate cancer. Double immunofluorescence labeling of a prostate cancer tissue microarray indicated that PLK1 expression correlated positively with numbers of infiltrating macrophages. These results indicate that an immune-related, gene-based risk score effectively reflects immune microenvironment characteristics and predicts prognosis in prostate cancer.

## INTRODUCTION

Prostate cancer poses a serious threat to the health of men all over the world [1, 2]. Although radical surgical excision and radiation therapy are effective in treating prostate cancer, only lung cancer has a higher mortality rate. In China, where prostate cancer screening is relatively uncommon, most patients have locally advanced or metastatic prostate cancer upon diagnosis. In these cases, endocrine therapy is typically used as the primary treatment. After about 18 months of endocrine therapy, most prostate cancers become resistant to hormone treatments, and no other effective clinical treatments are currently available. Novel immunotherapies are an important development in tumor

treatment, and immune checkpoint inhibitors may be particularly effective. However, recent phase II clinical trials indicate that immune check point inhibitors are only effective against specific types of prostate cancer, and the disease control rate does not exceed 20% [3, 4].

The development of resistance to immunotherapies is the main reason for their poor efficacy in treating prostate cancer. PD-L1<sup>+</sup> or VISTA<sup>+</sup> M2 macrophages are the major drivers of prostate immunotherapy resistance [5]. M1 macrophages activated by the classical pathway can induce tissue inflammation, and T cells activated by inflammation can effectively kill or suppress prostate cancer cells. In contrast, M2 macrophages with anti-inflammatory characteristics can

promote tissue repair and immunosuppression. M2 macrophages exert protective effects in tumors by inhibiting tumor antigen presentation. This prevents T cells from recognizing tumor antigens and immune checkpoint inhibitors, which regulate T cell function, from effectively treating prostate cancer [6–8]. Additional causes of insensitivity to immunotherapy in prostate cancer remain poorly understood. Further explorations of the immune characteristics of prostate cancer are therefore needed.

Interactions between tumor cells and immune cells are complex, and tumor status can alter the tumor immune microenvironment. In this study, we examined relationships between gene expression and immune cell infiltration in prostate cancer. Immune genes associated with biochemical recurrence of prostate cancer were screened to construct a prognostic model. The effectiveness of the model was validated using an external dataset. Relationships between immune characteristic genes and immune cell infiltration in prostate cancer were analyzed. Finally, key genes that might induce crucial changes

in the prostate cancer immune microenvironment were identified (Figure 1). The relationship between expression of these potential key genes and M2 macrophage infiltration was verified in tissue samples.

## RESULTS

### Weighted correlation network analysis (WGCNA) identified prostate cancer-related genes associated with biochemical recurrence (BCR)

The TCGA database includes 5132 genes that are differentially expressed in prostate cancer compared to normal prostate tissue. Among them, 602 were immune-regulated genes (IRGs). Weighted correlation network analysis (WGCNA) was performed using these 602 candidate genes based on TCGA PRAD cancer sample expression profiles (Figure 2). Eight samples with abnormal clustering were removed during the screening process. The soft-thresholding power in the WGCNA ( $\beta=5$ ) was determined based on scale-free  $R^2 > 0.85$  (Figure 3A). Eight modules were identified by average

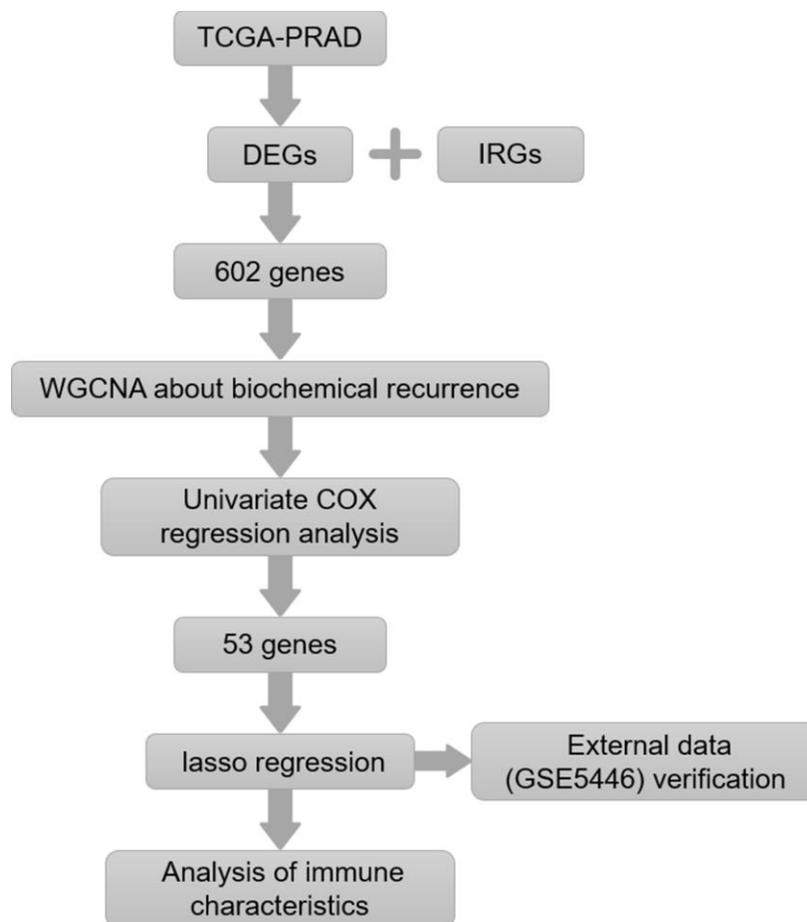


Figure 1. Flow chart of the experimental strategy.

linkage hierarchical clustering based the soft-thresholding power (Figure 3B). According to the hierarchical clustering of modules, there were significant correlations between most modules (Figure 3C). A network heatmap of all genes is shown in Figure 3D.

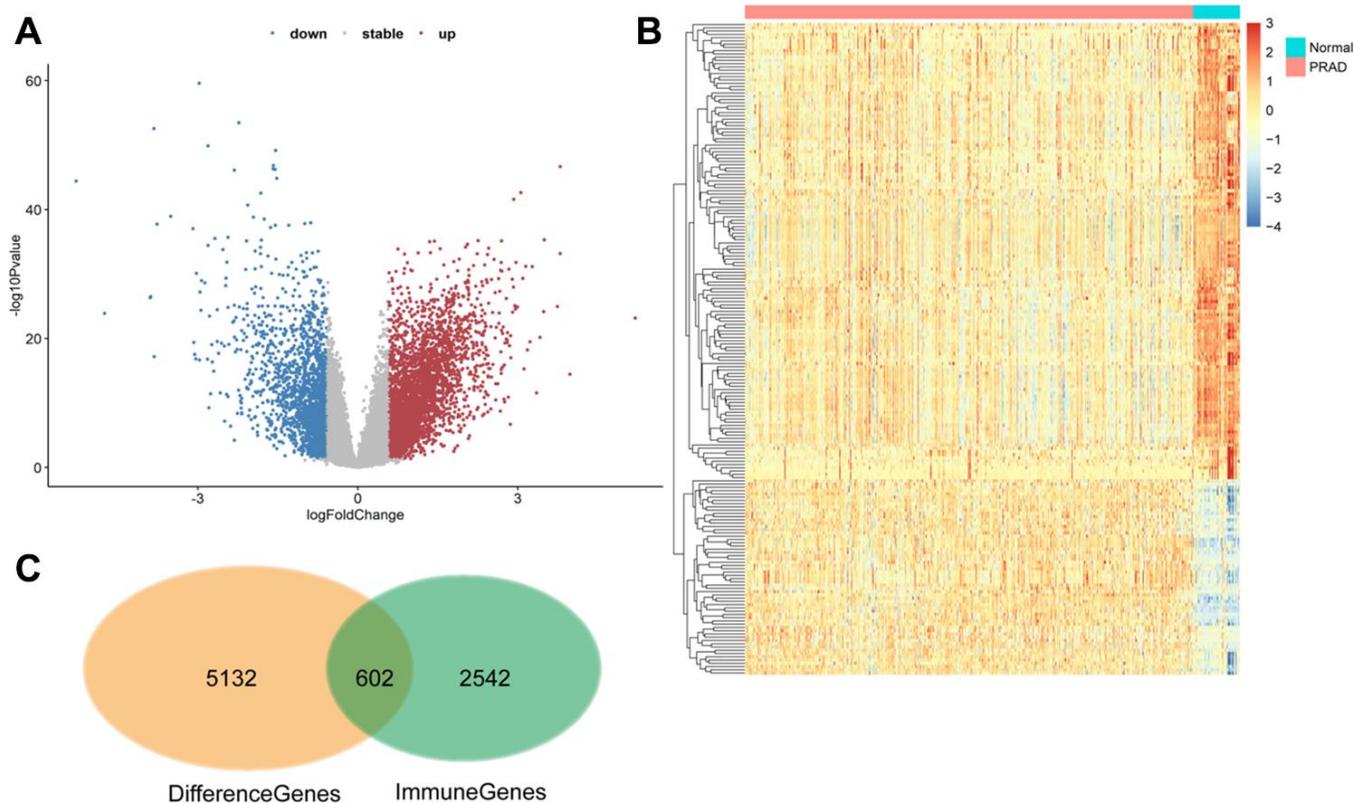
### Construction of an immunogenetic risk score associated with BCR in prostate cancer

We analyzed associations between all modules and clinical characteristics of prostate cancer patients. The brown, turquoise, pink, and yellow modules were highly associated with time to BCR, and the pink and red modules were also associated with BCR (Figure 4). Further screening of the genes included in these modules revealed a correlation coefficients of greater than 0.1 between their expression and BCR in prostate cancer and of greater than 0.5 for gene expression correlations within the modules. In total, 221 immune-related genes were associated with BCR in prostate cancer (Supplementary Table 3). We then performed Cox univariate regression analysis to identify correlations between the above genes and BCR based on clinical information for prostate cancer patients in

TCGA. A log-rank test revealed that 53 genes were associated with BCR in prostate cancer. Among these 53 genes, eight survival-associated IRGs (EDNRB, ANGPTL2, TNFSF15, TNFRSF10D, EDN2, BMP2, NLRP14, PLK1) were also identified using lasso regression analysis (Figure 5A, 5B). KM curves for PLK1, NLRP14, TNFRSF10D, and FGFR2 are shown in Figure 5C–5F; lasso regression coefficients for all eight IRGs are show in Table 1.

### Immunogenetic risk score is associated with BCR in prostate cancer

Risk scores for BCR in prostate cancer based on corresponding lasso coefficients were calculated for each sample in the training cohort (GSE54460). Risk score was significantly associated with BCR in prostate cancer patients; BCR occurred sooner in the high risk group (Figure 6A). An ROC curve was used to assess the effect of the Risk score. The AUC (Area Under Curve) was 0.749 at one year, 0.804 at 3 years, and 0.774 at 5 years in the testing cohort (Figure 6B). The ROC curve in the TCGA cohort was consistent with the curve in the training cohort (Figure 6C). The AUC was



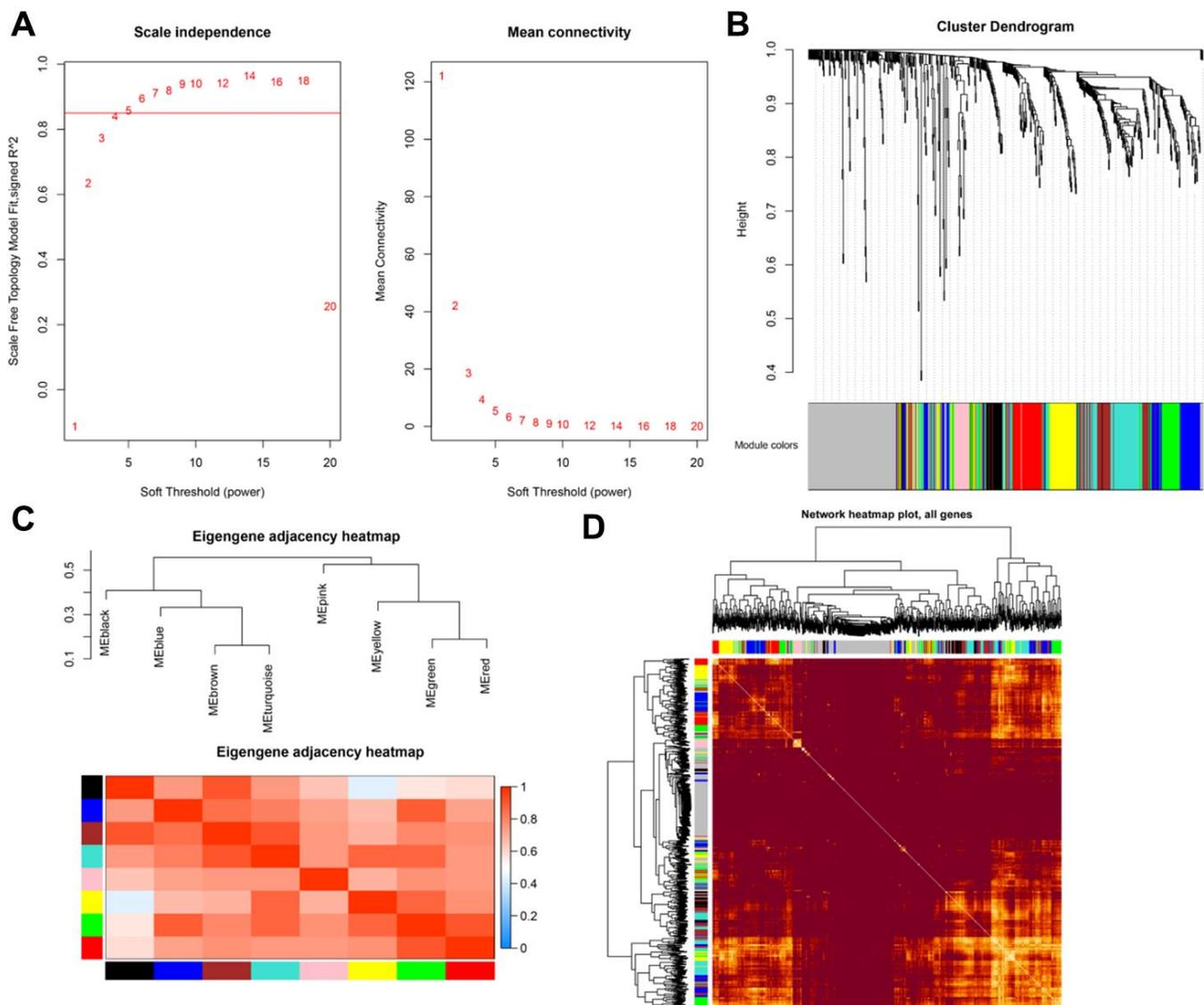
**Figure 2. 602 differentially expressed immune-related genes were identified.** (A and B) Volcano plot and heat maps showing differentially expressed genes in TCGA prostate cancer samples. (C) The 602 differentially expressed immune genes were considered candidate genes for the risk model.

0.644 at one year, 0.69 at 3 years, and 0.691 at 5 years in the TCGA cohort (Figure 6D). Risk score was therefore a useful predictor of BCR in prostate cancer. We also performed a multivariate correlation analysis to determine the impact of other clinical factors on the prognostic power of the risk score. Immune risk score was still associated with BCR in prostate cancer after multivariate adjustment ( $p=0.042$ ) (Table 2).

### Differences in immune characteristics between high and low risk groups

Next, we examined differences in the tumor microenvironment between low and high risk patients based on prostate cancer mRNA expression data from

the TCGA database. Genes that were differentially expressed between the high and low risk score groups are shown in the volcano plot in Supplementary Figure 1. No significant differences in immune and stroma genes were observed between the high and low risk score groups (Supplementary Figure 2). However, differences were observed in infiltration levels for 22 types immune cells in CIBERSORT data. Memory B cell, regulatory T cell, M2 macrophage, and dendritic cell infiltration were higher in the high risk group, while plasma cell, CD8<sup>+</sup>T cell, monocyte, and activated mast cell infiltration were higher in the low score group (Figure 7A). Overall immune cell infiltration data for the high and low risk score groups is shown in Figure 7B. Expression of immunoregulatory factors in the



**Figure 3.** (A) Scale-free fit index for soft-thresholding powers. (B) Dendrogram showing all differentially expressed genes clustered based on different metrics. (C) Heatmap of correlations between module eigengenes and clinical traits. (D) Visualization of gene networks.

tumor immune microenvironment also differed between the high and low risk groups. (Figure 8A); the 28 factors that differed significantly are shown in Supplementary Figure 3.

### Identification of pathways and gene ontology (GO) terms associated with high and low risk score groups

The ClusterProfiler R package was used to perform GSEA enrichment analysis on DEGs between the high and low score groups using gene sets from MsigDB as background genes. Log<sub>2</sub>(fold change) values from the differential expression analysis were used as the sorting criterion, and P value < 0.05 was used as the screening criterion (Supplementary Table 4). The results indicated that there were significant differences in “immune effector process,” “immune response,” “immune system process,” “innate immune response,” “positive regulation of immune system process,” “regulation of immune system process,” and “regulation of immune system process” between low and high risk groups (Figure 8B). GO enrichment analysis was also performed using p value < 0.05 and overlap > 0.75 as screening criteria (Supplementary Figure 4).

### PLK1 may be a key regulator of the immune microenvironment in prostate cancer

The LnCeVar database contains transcription factor regulation data validated in the literature. As shown in Figure 9, PLK1 is part of a rich regulation system. In a previous study, we found that infiltration of M2 macrophages promotes prostate cancer progression. Here, risk score was also associated with M2 macrophages. Among the genes included in the risk score, PLK1 had the largest coefficient, indicating that it may be the most critical factor in the model. We therefore performed double immunofluorescence labeling of CD163, a marker of M2 macrophages, and PLK1 in a prostate cancer tissue microarray (Figure 10A). PLK1 was positively correlated with CD163 in prostate cancer (Pca) samples (r=0.69, p<0.01), but not in benign prostatic hypertrophy (BPH) samples (r=0.12, p=0.63) (Figure 10B). Furthermore, PLK1 expression was significantly increased in prostate cancer compared to prostatic hyperplasia. (p=0.022) (Figure 10C).

### DISCUSSION

Current treatments for prostate cancer include surgery, radiation therapy, and endocrine therapy. Recent studies

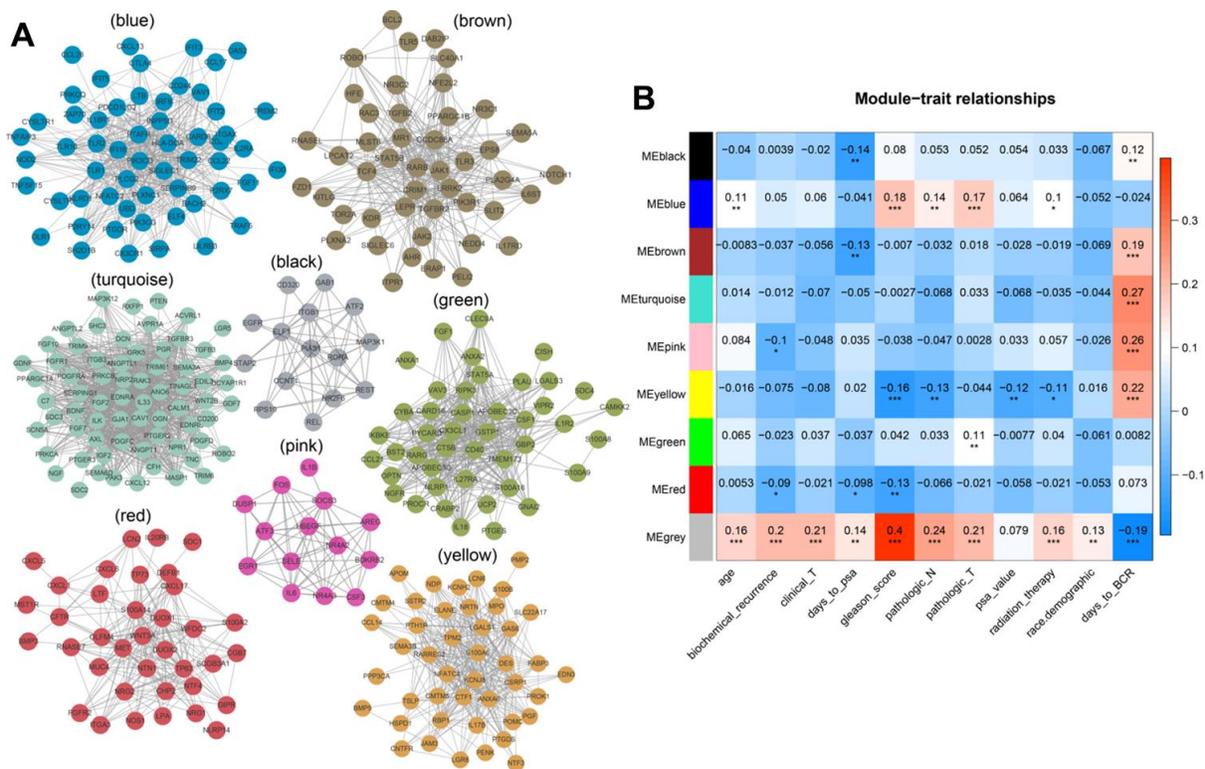
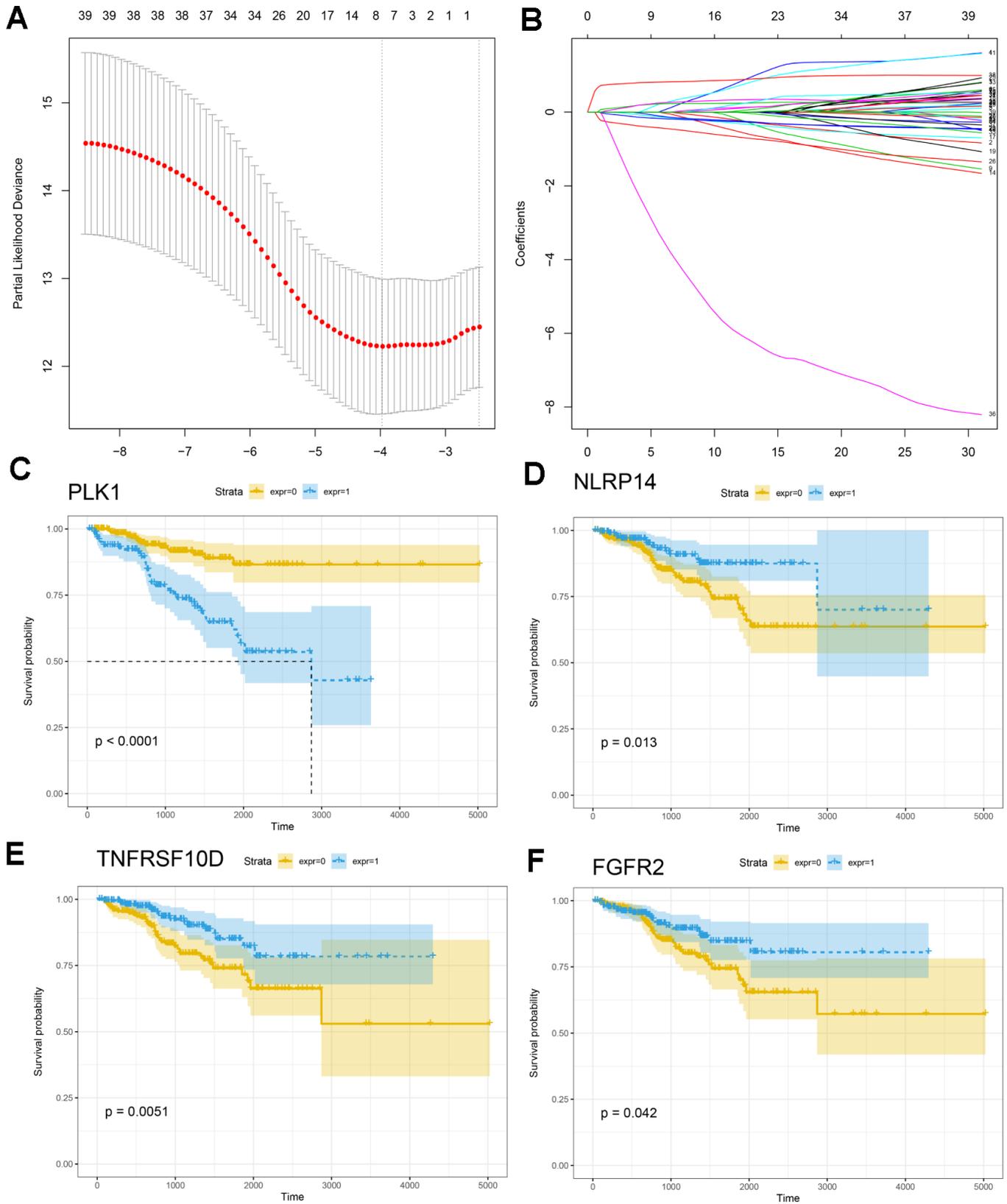


Figure 4. (A) Modular genetic correlation network map. Colors correspond to different modules. (B) Correlations between modules and clinical phenotypes. Red indicates positive correlations, blue indicates negative correlations.



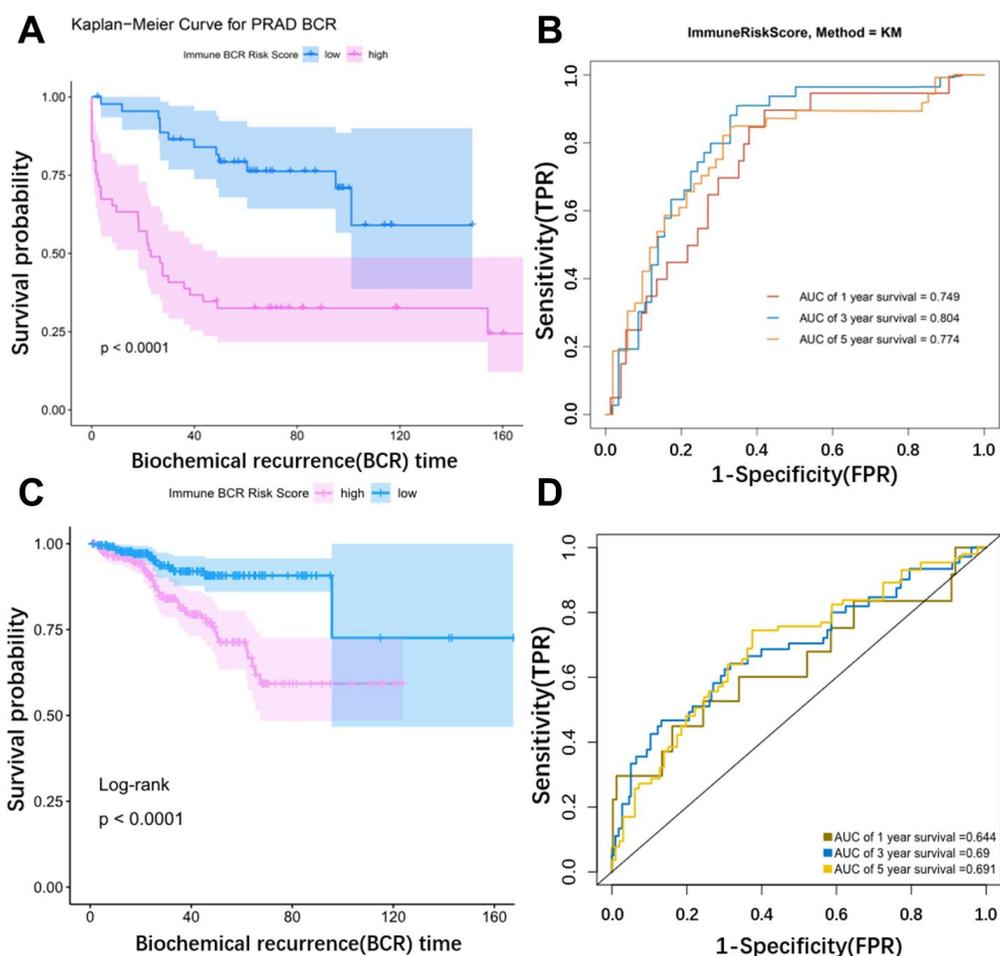
**Figure 5. Construction of the IRG-based prognostic model.** (A, B) The number of factors included in the model was determined through LASSO analysis. (C–F) KM curves for PLK1, NLRP14, TNFRSF10D, and FGFR2.

**Table 1. Lasso regression.**

Symbol	Lasso regression coefficient
DNRB	-0.057
ANGPTL2	0.168
TNFSF15	-0.007
TNFRSF10D	-0.357
EDN2	-0.123
BMP2	0.158
NLRP14	-1.952
PLK1	0.787

indicate that the tumor immune microenvironment plays an important role in the development and progression of prostate cancer, and immunotherapy can provide significant benefits to some prostate cancer patients. However, regulatory T cells, tolerogenic dendritic cells,

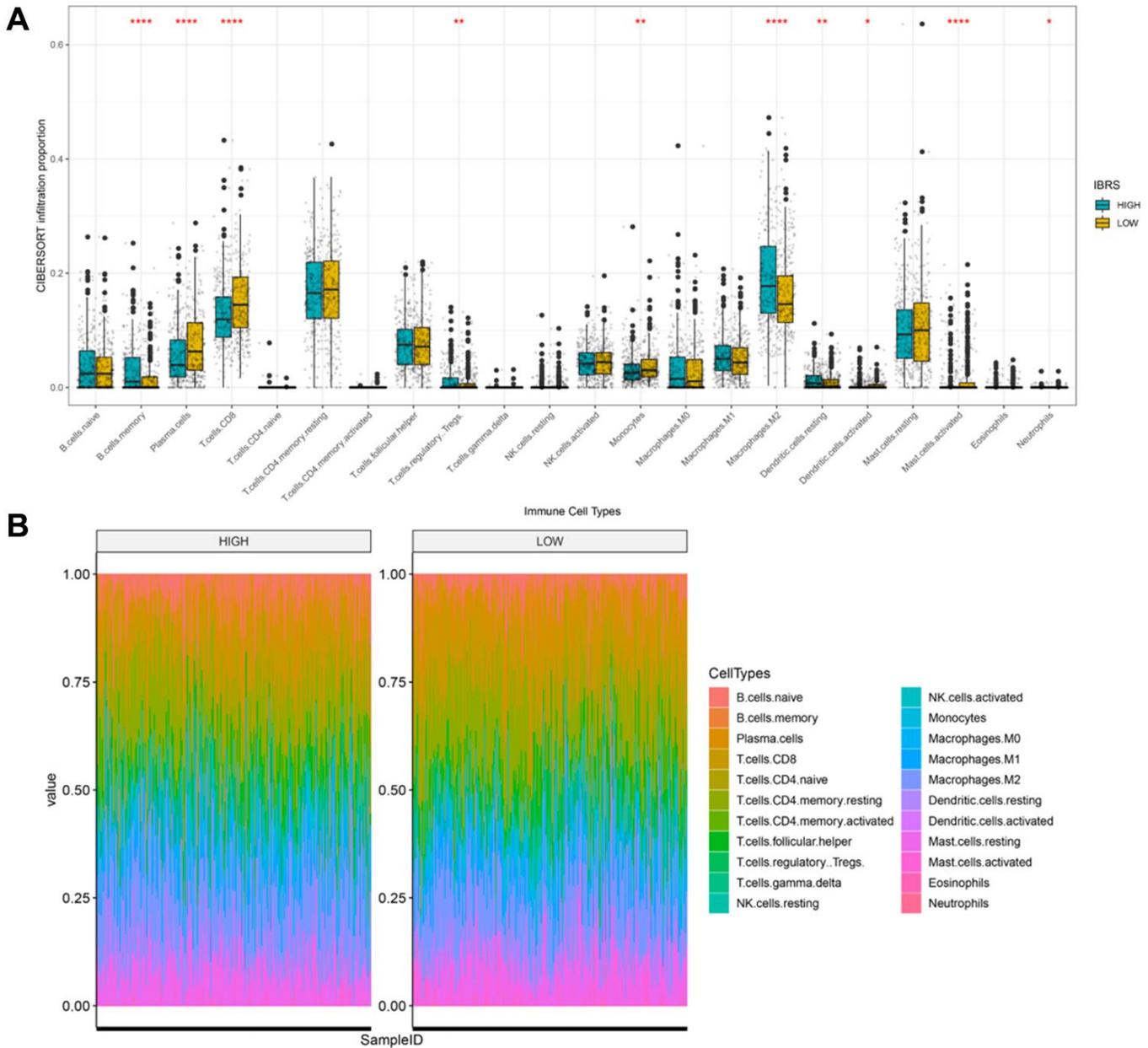
and non-functional T cells (CD4<sup>+</sup> and CD8<sup>+</sup>) can induce immunosuppression in the tumor microenvironment [9]. Such immunosuppression is the main obstacle to the efficacy of immunotherapy-induced anti-tumor immune responses. Multi-drug combination therapy, which can



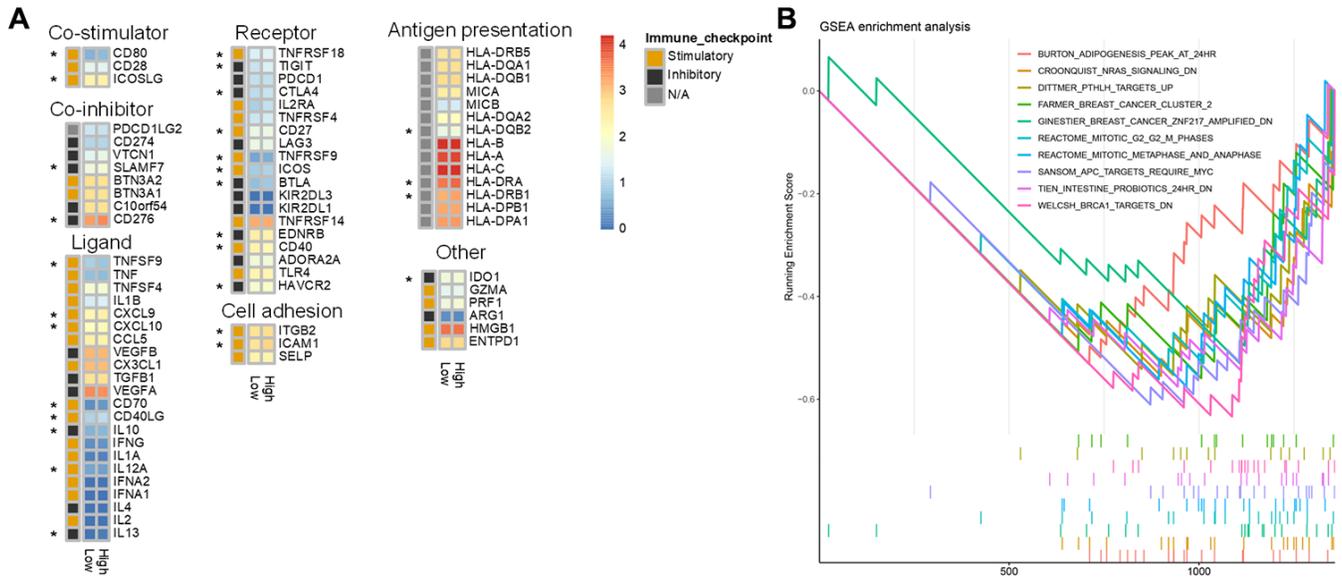
**Figure 6. Validation of the model using external data.** (A) KM curve for the external dataset (GSE54460). (B) Time dependent ROC curves. The AUC (Area Under Curve) was 0.749 at 1 year, 0.804 at 3 years, and 0.774 at 5 years in the GSE54460 cohort. (C) KM curve for TCGA. (D) Time dependent ROC curves. The AUC (Area Under Curve) was 0.644 at 1 year, 0.69 at 3 years, and 0.691 at 5 years in the TCGA cohort.

**Table 2. Multiple regression analysis was used to verify the model generated via lasso regression.**

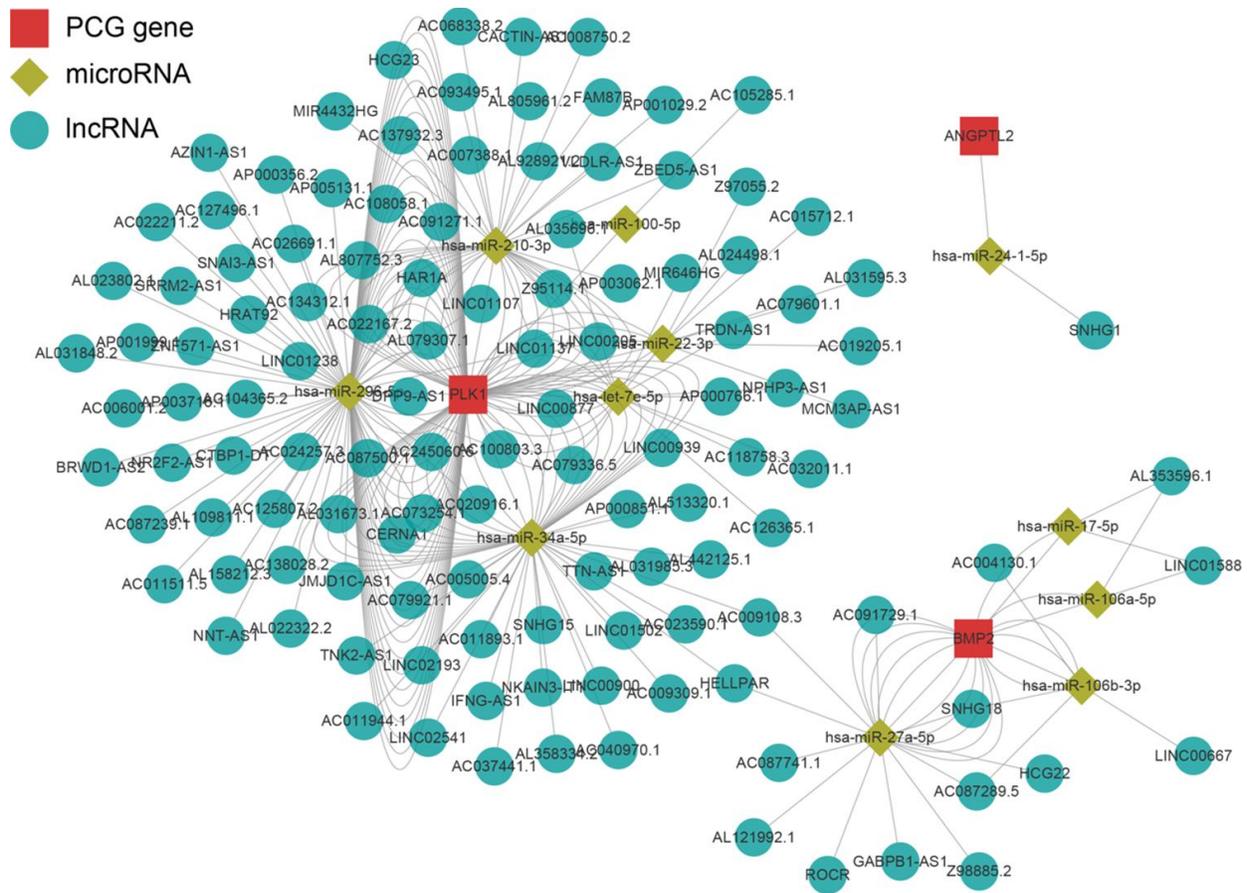
Variate	Univ HR (95% CI for HR)	Univ p value	Multiv HR (95% CI for HR)	Multiv p value
Age	1.0064 (0.9593-1.0559)	0.7938	0.9875 (0.9388-1.0388)	0.6265
Concentration (ng/μl)	1.0071 (1.0034-1.0107)	0.0001	0.9842 (0.9239-1.0484)	0.6207
Race (W VS B)	1.028 (0.4305-2.4546)	0.9505	1.2058 (0.4565-3.1853)	0.7057
Ratio 260/230	8.0448 (2.9966-21.5976)	<0.0001	4.9789 (0.7663-32.3485)	0.0927
Rpl13a Ct value	1.2157 (1.003-1.4734)	0.0465	1.1231 (0.8919-1.4142)	0.3234
Total yield (μg)	1.0909 (1.0402-1.1441)	0.0003	1.2078 (0.538-2.7116)	0.6472
Immune Risk score	7.2791 (2.3037-22.9999)	0.0007	3.7842 (1.0476-13.6691)	<b>0.0423</b>



**Figure 7. Cibersort was used to calculate infiltration scores for 22 immune cell types based on the TCGA prostate cancer expression profile. (A) Infiltration differences (ratio) in high and low risk groups. \*\*\*\*p<0.0001, \*\*\*p<0.001, \*\*p<0.01, \*p<0.05. (B) Infiltration profiles in the high and low risk groups.**



**Figure 8. Differences in immune characteristics between the high and low risk score groups. (A)** Expression of 75 immunomodulators in the high and low risk groups. “\*\*” indicates a difference in expression between the high and low risk groups. **(B)** Immune-related GSEA enrichment analysis.



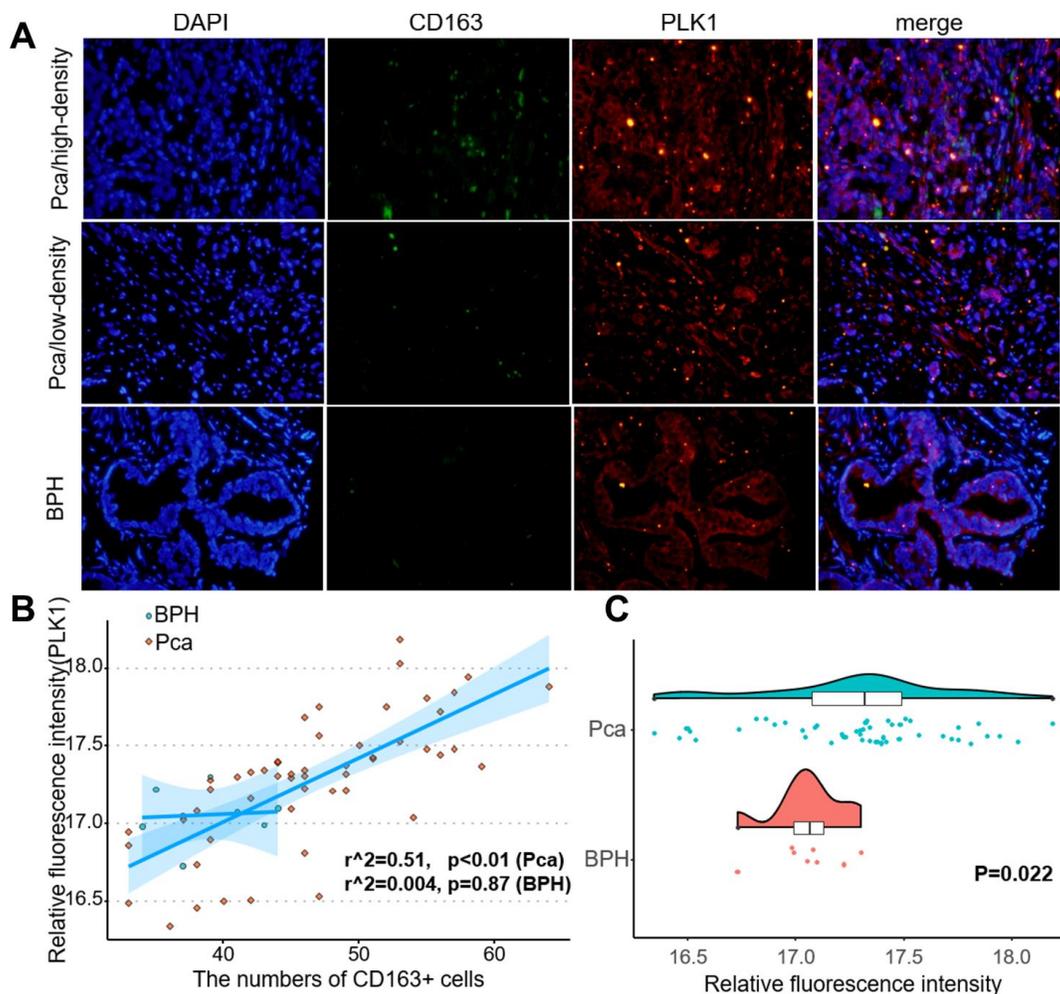
**Figure 9. Regulatory network of the key genes.** Red squares indicate key genes, green diamonds indicate microRNAs, and green circles indicate lncRNA. Key genes for which literature reporting validated regulatory networks was not available were omitted.

increase the impact of immune checkpoint inhibitors by regulating the tumor microenvironment, is an important method for improving the efficacy of cancer immunotherapy [10]. In this study, we examined the immune characteristics of prostate cancer to identify potential key genes that regulate the tumor immune microenvironment.

We found that 53 immune-related genes were associated with biochemical recurrence in prostate cancer. Among these, eight were identified as key genes by lasso regression. Validation with an external dataset indicated that the prediction model was highly accurate (5 year AUC=0.774). Moreover, biochemical recurrence occurred sooner and survival rates were lower in the high risk score group. Analysis of the LnCeVar database indicated that PLK1 had a particularly rich

regulatory system in prostate cancer among the eight key genes.

Decreased expression of EDNRB, TNFSF15, TNFRSF10D, EDN2, and NLRP14, as well as increased expression of ANGPTL2, BMP2, and PLK1, were associated with higher risk scores; EDNRB promoter methylation status was also associated with risk score. A meta-analysis of 11 similar studies indicated that the frequency of EDNRB methylation was substantially higher in prostate cancer compared with normal prostate tissues (OR = 5.42, 95 % CI = 1.98–14.88, P = 0.001), suggesting that EDNRB promoter methylation might increase the risk of prostate cancer [11]. Here, TNFSF15 expression was inversely associated with prostate cancer risk. This is consistent with the role of TNFSF15 as a downstream effector of AMPK that



**Figure 10. PLK1 expression correlated positively with M2 macrophage infiltration.** (A) Fluorescence imaging of human prostate cancer and adjacent noncancerous tissues with FITC-labeled CD163 and Cy3-labeled PLK1. Most green fluorescent signals were observed on the cytomembrane, while red fluorescent signals were primarily located in the cytoplasm in prostate tissue. (B) Numbers of green fluorescent cells and red fluorescence integral optical density were positively correlated in prostate cancer samples ( $r^2=0.51, p<0.01$ ). (C) PLK1 staining was more intense in prostate cancer tissues than in non-cancerous prostate tissues.

inhibits prostate cancer growth [12]. TNFRSF10D expression is also associated with prostate cancer and with the direct p53 effectors and ERK signaling pathways; here, TNFRSF10D was inversely related to prostate cancer biochemical recurrence risk score. Endothelins are involved in the regulation of various physiological processes, including plumage development in chickens, pigmentation, neural crest cell proliferation, differentiation, migration, cardiovascular development and functions, and pulmonary hypertension [13]. The endothelin EDN2 was inversely associated with prostate cancer immunological risk score in our model. As a member of a family of molecules that belong to a signal-induced multiprotein complex termed the inflammasome that activates proinflammatory caspase-1 and caspase-5, NLRP14 may play a regulatory role in the innate immune system. NLRP14 is considered an oncogene, and increased expression of NLRP14 is associated with increases in prostate cancer mortality [14]. This contradicts the inverse association found here between NLRP14 and risk score, and further study is warranted. ANGPTL2 is a secreted glycoprotein with homology to angiopoietins that may exert autocrine or paracrine effects on endothelial cells. ANGPTL2 also promotes M2 polarization of macrophages in non-small cell lung cancer [15]. In addition, ANGPTL2 may promote acquisition of androgen independence and tumor progression in prostate cancer by exerting autocrine and/or paracrine effects via the integrin  $\alpha 5 \beta 1$  receptor [16]. Here, we found a positive association between ANGPTL2 expression and immune-related risk score in prostate cancer. Finally, several studies suggest that BMP2 promotes progression and induces biochemical recurrence in prostate cancer [17–19], which is consistent with the positive association found here between BMP2 and risk score.

Our present results demonstrate that memory B cell, regulatory T cell, M2 macrophage, and dendritic cell infiltration were significantly increased in the high risk score group, while plasma cell, CD8+T cell, monocyte, and activated mast cell infiltration were higher in the low risk score group. M2 macrophages promote prostate cancer progression and help establish an immunosuppressive state in tumors [20–21]; this might help explain the increased M2 macrophage infiltration observed in the higher risk score group. In contrast, local increases in the density of infiltrating CD8+T cells in tumors is a marker of good prognosis; this might account for the increased CD8+T cell infiltration observed here in the low risk score group with better prognoses. Infiltration of these cell types reflects the immune microenvironment and can also predict prognosis in prostate cancer [22].

The role of PLK1 in prostate cancer is not clear at present [23]. However, numerous studies indicate that PLK1 can act as an oncogene, and PLK-1 inhibitors can effectively inhibit prostate cancer progression [24–26]. PLK1 can also inactivate other tumor suppressors [27, 28]. Another study suggests that PLK1 is a carcinogenic factor [29]. Here, we found that PLK1 expression was higher in the high risk group that experienced earlier biochemical recurrence. Moreover, downstream pathways regulated by PLK1 comprised the single largest pathway group in our biochemical recurrence prediction model for prostate cancer. PLK1 might therefore be one of the most important immune genes that contribute to biochemical recurrence in prostate cancer.

## MATERIALS AND METHODS

### Data acquisition

The training dataset was obtained from TCGA and included 498 cancer samples and 52 normal control samples; clinical information for those samples is shown in Supplementary Table 1. The GSE54460 testing dataset containing 106 samples was downloaded from the gene expression omnibus database (GEO: <https://www.ncbi.nlm.nih.gov/geo/>); clinical information for those samples is shown in Supplementary Table 2. Immune genes that were included in our analyses were obtained from InnateDB (<https://www.innatedb.ca/>) and ImmPort (<https://www.immport.org/home>).

### Identification of differentially expressed genes

Genes that were differentially expressed between normal and prostate cancer samples were identified based on the screening criteria of p value > 0.05 and log(fold-change) > 1.5.

### Weighted gene co-expression network analysis

TCGA expression data for 602 immune-related genes was used for WGCNA (weighted gene co-expression network analysis) to identify associations between gene expression modules and clinical characteristics [30]. During sample clustering, eight samples with abnormal clustering were identified and removed from the analysis (Supplementary Figure 1). The co-expression network was then constructed and divided into modules. A scale-free network coefficient of greater than 0.85 was used to ensure that the co-expression network conformed to the scale-free network standards. Gene significance (GS) indicates the strength of linear correlations between the expression of different gene modules and clinical features. Modules with  $P \leq 0.01$  and higher GS values

were identified as survival-related modules and included in subsequent analysis.

### Univariate Cox regression and lasso regression

The Survival package for R was used to perform univariate Cox regression and the KM test. The Glmnet package was used for lasso regression. Prognosis related genes were screen and correlation regression coefficients were obtained. The resulting risk score was validated in the GEO dataset (GSE54460). “TimeROC” was used to draw receiver operating characteristic curves (ROC), and the area under the curve (AUC) was calculated. Samples were divided between the high and low risk groups using the median risk score (RS) as a cutoff value. KM curves were used to evaluate the survival of prostate cancer patients.

### Comparison of immune cell infiltration between high and low risk groups

The “ESTIMATE” package was used to calculate the microenvironment score, and the “CIBERSORT” package was used to assess the proportions of 22 leukocyte subtypes based on differences in mRNA expression between the high and low risk groups [31]. The Wilcox test was used to evaluate differences between the high low risk score groups.

### Double immunofluorescence

Tissue microarrays (DC-Pro11018, Xian, China) including 74 prostate cancer tissues and 6 non-cancer prostate tissues, along with associated detailed clinical information, were purchased from Alenabio Biotech (Xian, China). Clinical information for the prostate cancer tissue microarray is shown in Supplementary Table 5. The sections were incubated overnight at 4°C with antibodies against anti-CD163 (rabbit, 1:100, A8383, Elabscience, Wuhan, China) and anti-PLK1 (mouse, 1:50, TA500393S, ORIGENE, Rockville, USA). The sections were then washed three times with cold PBS and stained with Cy3 Goat Anti-Mouse IgG (H+L) (1:100, AS008, ABclonal, Wuhan, China) or FITC Goat Anti-Rabbit IgG (H+L) (1:100, AS011, AS008, ABclonal, Wuhan, China) secondary antibodies. Nuclei were stained with DAPI. Stained tissues were visualized using an Olympus IX73 microscope (Waltham, MA). IOD (integral optical density) was calculated using ImageJ (1.46r, National Institutes of Health, USA).

### Statistical analysis

All statistical analyses were performed using R software (version 3.6.3, <http://www.R-project.org>). A two-sided  $P < 0.05$  indicated a statistically significant difference.

## AUTHOR CONTRIBUTIONS

Zezen Liu: conceptualization, writing of original draft; Jiehui Zhong: conceptualization, formal analysis; Chao Cai: formal analysis; Jianming Lu: methodology; Guohua Zeng: funding acquisition, investigation.

## CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## FUNDING

This work was financed by grants from the China Postdoctoral Science Foundation(No. 2020M672590), the National Natural Science Foundation of China (No. 81670643 and 81870483), the Collaborative Innovation Project of Guangzhou Education Bureau (No. 1201620011), the Guangzhou Science Technology and Innovation Commission (No. 201704020193), and the Science and Technology Planning Project of Guangdong Province (No. 2017B030314108).

## REFERENCES

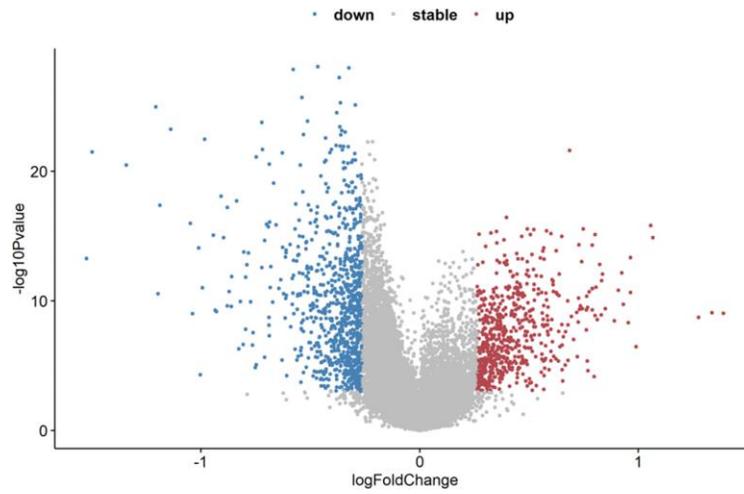
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:31912902
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424. <https://doi.org/10.3322/caac.21492> PMID:30207593
3. Jindal V. Immunotherapy: a glimmer of hope for metastatic prostate cancer. *Chin Clin Oncol.* 2018; 7:61. <https://doi.org/10.21037/cco.2018.02.01> PMID:29860848
4. Antonarakis ES, Piulats JM, Gross-Goupil M, Goh J, Ojamaa K, Hoimes CJ, Vaishampayan U, Berger R, Sezer A, Alanko T, de Wit R, Li C, Omlin A, et al. Pembrolizumab for treatment-refractory metastatic castration-resistant prostate cancer: multicohort, open-label phase II KEYNOTE-199 study. *J Clin Oncol.* 2020; 38:395–405. <https://doi.org/10.1200/JCO.19.01638> PMID:31774688
5. Gao J, Ward JF, Pettaway CA, Shi LZ, Subudhi SK, Vence LM, Zhao H, Chen J, Chen H, Efstathiou E, Troncoso P, Allison JP, Logothetis CJ, et al. VISTA is an inhibitory immune checkpoint that is increased after ipilimumab therapy in patients with prostate cancer. *Nat Med.* 2017; 23:551–55.

- <https://doi.org/10.1038/nm.4308> PMID:[28346412](https://pubmed.ncbi.nlm.nih.gov/28346412/)
6. Uderhardt S, Martins AJ, Tsang JS, Lämmermann T, Germain RN. Resident macrophages cloak tissue microlesions to prevent neutrophil-driven inflammatory damage. *Cell*. 2019; 177:541–55.e17. <https://doi.org/10.1016/j.cell.2019.02.028> PMID:[30955887](https://pubmed.ncbi.nlm.nih.gov/30955887/)
  7. El-Kenawi A, Gatenbee C, Robertson-Tessi M, Bravo R, Dhillion J, Balagurunathan Y, Berglund A, Vishvakarma N, Ibrahim-Hashim A, Choi J, Luddy K, Gatenby R, Pilon-Thomas S, et al. Acidity promotes tumour progression by altering macrophage phenotype in prostate cancer. *Br J Cancer*. 2019; 121:556–566. <https://doi.org/10.1038/s41416-019-0542-2> PMID:[31417189](https://pubmed.ncbi.nlm.nih.gov/31417189/)
  8. Kowal J, Kornete M, Joyce JA. Re-education of macrophages as a therapeutic strategy in cancer. *Immunotherapy*. 2019; 11:677–89. <https://doi.org/10.2217/imt-2018-0156> PMID:[31088236](https://pubmed.ncbi.nlm.nih.gov/31088236/)
  9. Jafari S, Molavi O, Kahroba H, Hejazi MS, Maleki-Dizaji N, Barghi S, Kiaie SH, Jadidi-Niaragh F. Clinical application of immune checkpoints in targeted immunotherapy of prostate cancer. *Cell Mol Life Sci*. 2020; 77:3693–710. <https://doi.org/10.1007/s00018-020-03459-1> PMID:[32006051](https://pubmed.ncbi.nlm.nih.gov/32006051/)
  10. Boettcher AN, Usman A, Morgans A, VanderWeele DJ, Sosman J, Wu JD. Past, current, and future of immunotherapies for prostate cancer. *Front Oncol*. 2019; 9:884. <https://doi.org/10.3389/fonc.2019.00884> PMID:[31572678](https://pubmed.ncbi.nlm.nih.gov/31572678/)
  11. Yuan Y, Du Y, Wang L, Liu X. The value of endothelin receptor type B promoter methylation as a biomarker for the risk assessment and diagnosis of prostate cancer: a meta-analysis. *Pathol Res Pract*. 2020; 216:152796. <https://doi.org/10.1016/j.prp.2019.152796> PMID:[31926772](https://pubmed.ncbi.nlm.nih.gov/31926772/)
  12. Zhou J, Yang Z, Tsuji T, Gong J, Xie J, Chen C, Li W, Amar S, Luo Z. LITAF and TNFSF15, two downstream targets of AMPK, exert inhibitory effects on tumor growth. *Oncogene*. 2011; 30:1892–900. <https://doi.org/10.1038/onc.2010.575> PMID:[21217782](https://pubmed.ncbi.nlm.nih.gov/21217782/)
  13. Ling L, Maguire JJ, Davenport AP. Endothelin-2, the forgotten isoform: emerging role in the cardiovascular system, ovarian development, immunology and cancer. *Br J Pharmacol*. 2013; 168:283–95. <https://doi.org/10.1111/j.1476-5381.2011.01786.x> PMID:[22118774](https://pubmed.ncbi.nlm.nih.gov/22118774/)
  14. Wang H, Shen L, Li Y, Lv J. Integrated characterisation of cancer genes identifies key molecular biomarkers in stomach adenocarcinoma. *J Clin Pathol*. 2020; 73:579–86. <https://doi.org/10.1136/jclinpath-2019-206400> PMID:[32034058](https://pubmed.ncbi.nlm.nih.gov/32034058/)
  15. Wei X, Nie S, Liu H, Sun J, Liu J, Li J, Li S, Wang S, Han S, Wang J, Sun Y. Angiopoietin-like protein 2 facilitates non-small cell lung cancer progression by promoting the polarization of M2 tumor-associated macrophages. *Am J Cancer Res*. 2017; 7:2220–33. PMID:[29218246](https://pubmed.ncbi.nlm.nih.gov/29218246/)
  16. Sato R, Yamasaki M, Hirai K, Matsubara T, Nomura T, Sato F, Mimata H. Angiopoietin-like protein 2 induces androgen-independent and Malignant behavior in human prostate cancer cells. *Oncol Rep*. 2015; 33:58–66. <https://doi.org/10.3892/or.2014.3586> PMID:[25370833](https://pubmed.ncbi.nlm.nih.gov/25370833/)
  17. Lai TH, Fong YC, Fu WM, Yang RS, Tang CH. Osteoblasts-derived BMP-2 enhances the motility of prostate cancer cells via activation of integrins. *Prostate*. 2008; 68:1341–53. <https://doi.org/10.1002/pros.20799> PMID:[18512729](https://pubmed.ncbi.nlm.nih.gov/18512729/)
  18. Horvath LG, Henshall SM, Kench JG, Turner JJ, Golovsky D, Brenner PC, O'Neill GF, Kooner R, Stricker PD, Grygiel JJ, Sutherland RL. Loss of BMP2, Smad8, and Smad4 expression in prostate cancer progression. *Prostate*. 2004; 59:234–42. <https://doi.org/10.1002/pros.10361> PMID:[15042598](https://pubmed.ncbi.nlm.nih.gov/15042598/)
  19. Tae BS, Cho S, Kim HC, Kim CH, Kang SH, Lee JG, Kim JJ, Park HS, Cheon J, Oh MM, Kang SG. Decreased expression of bone morphogenetic protein-2 is correlated with biochemical recurrence in prostate cancer: Immunohistochemical analysis. *Sci Rep*. 2018; 8:10748. <https://doi.org/10.1038/s41598-018-28566-9> PMID:[30013089](https://pubmed.ncbi.nlm.nih.gov/30013089/)
  20. Zheng T, Ma G, Tang M, Li Z, Xu R. IL-8 secreted from M2 macrophages promoted prostate tumorigenesis via STAT3/MALAT1 pathway. *Int J Mol Sci*. 2018; 20:98. <https://doi.org/10.3390/ijms20010098> PMID:[30591689](https://pubmed.ncbi.nlm.nih.gov/30591689/)
  21. Liu ZZ, Han ZD, Liang YK, Chen JX, Wan S, Zhuo YJ, Cai ZD, Deng YL, Lin ZY, Mo RJ, He HC, Zhong WD. TRIB1 induces macrophages to M2 phenotype by inhibiting IKB-zeta in prostate cancer. *Cell Signal*. 2019; 59:152–62. <https://doi.org/10.1016/j.cellsig.2019.03.017> PMID:[30926388](https://pubmed.ncbi.nlm.nih.gov/30926388/)
  22. Evans JD, Morris LK, Zhang H, Cao S, Liu X, Mara KC, Stish BJ, Davis BJ, Mansfield AS, Dronca RS, Iott MJ, Kwon ED, Foote RL, et al. Prospective

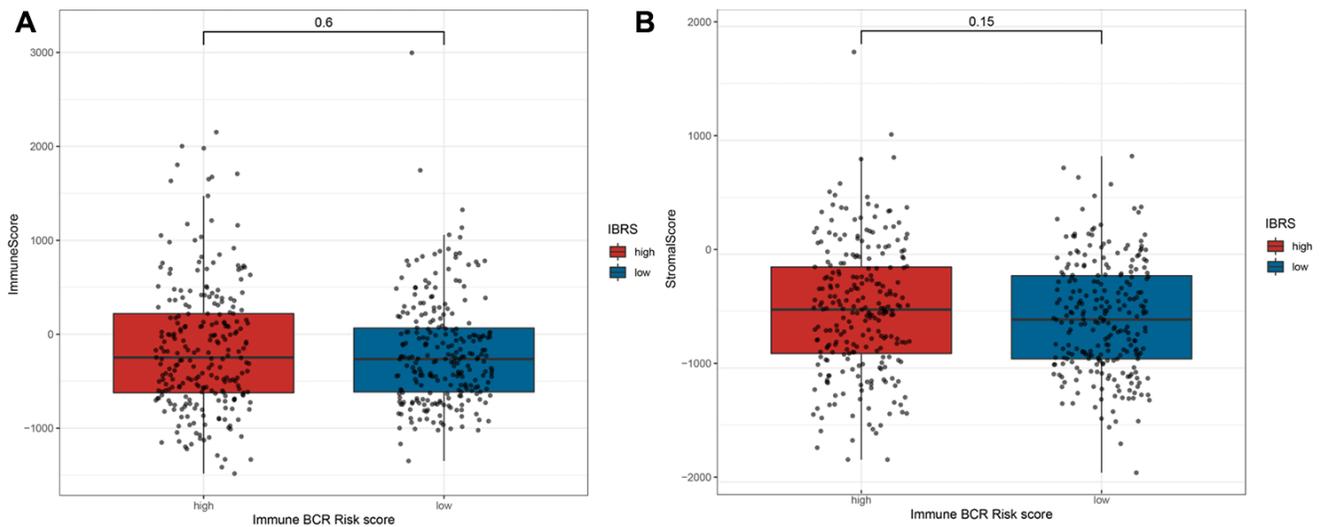
- immunophenotyping of CD8<sup>+</sup> T cells and associated clinical outcomes of patients with oligometastatic prostate cancer treated with metastasis-directed SBRT. *Int J Radiat Oncol Biol Phys.* 2019; 103:229–40.  
<https://doi.org/10.1016/j.ijrobp.2018.09.001>  
PMID:30205124
23. de Cárcer G. The Mitotic Cancer Target Polo-Like Kinase 1: Oncogene or Tumor Suppressor? *Genes (Basel).* 2019; 10:208.  
<https://doi.org/10.3390/genes10030208>  
PMID:30862113
24. Li J, Wang R, Kong Y, Broman MM, Carlock C, Chen L, Li Z, Farah E, Ratliff TL, Liu X. Targeting Plk1 to enhance efficacy of olaparib in castration-resistant prostate cancer. *Mol Cancer Ther.* 2017; 16:469–79.  
<https://doi.org/10.1158/1535-7163.MCT-16-0361>  
PMID:28069876
25. Shao C, Ahmad N, Hodges K, Kuang S, Ratliff T, Liu X. Inhibition of polo-like kinase 1 (Plk1) enhances the antineoplastic activity of metformin in prostate cancer. *J Biol Chem.* 2015; 290:2024–33.  
<https://doi.org/10.1074/jbc.M114.596817>  
PMID:25505174
26. Shin SB, Woo SU, Yim H. Cotargeting Plk1 and androgen receptor enhances the therapeutic sensitivity of paclitaxel-resistant prostate cancer. *Ther Adv Med Oncol.* 2019; 11:1758835919846375.  
<https://doi.org/10.1177/1758835919846375>  
PMID:31156720
27. Eckerdt F, Strebhardt K. Polo-like kinase 1: target and regulator of anaphase-promoting complex/cyclosome-dependent proteolysis. *Cancer Res.* 2006; 66:6895–98.  
<https://doi.org/10.1158/0008-5472.CAN-06-0358>  
PMID:16849530
28. Seki A, Coppinger JA, Du H, Jang CY, Yates JR 3rd, Fang G. Plk1- and beta-TrCP-dependent degradation of bora controls mitotic progression. *J Cell Biol.* 2008; 181:65–78.  
<https://doi.org/10.1083/jcb.200712027>  
PMID:18378770
29. Simizu S, Osada H. Mutations in the plk gene lead to instability of plk protein in human tumour cell lines. *Nat Cell Biol.* 2000; 2:852–54.  
<https://doi.org/10.1038/35041102> PMID:11056542
30. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559.  
<https://doi.org/10.1186/1471-2105-9-559>  
PMID:19114008
31. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019; 37:773–82.  
<https://doi.org/10.1038/s41587-019-0114-2>  
PMID:31061481

## SUPPLEMENTARY MATERIALS

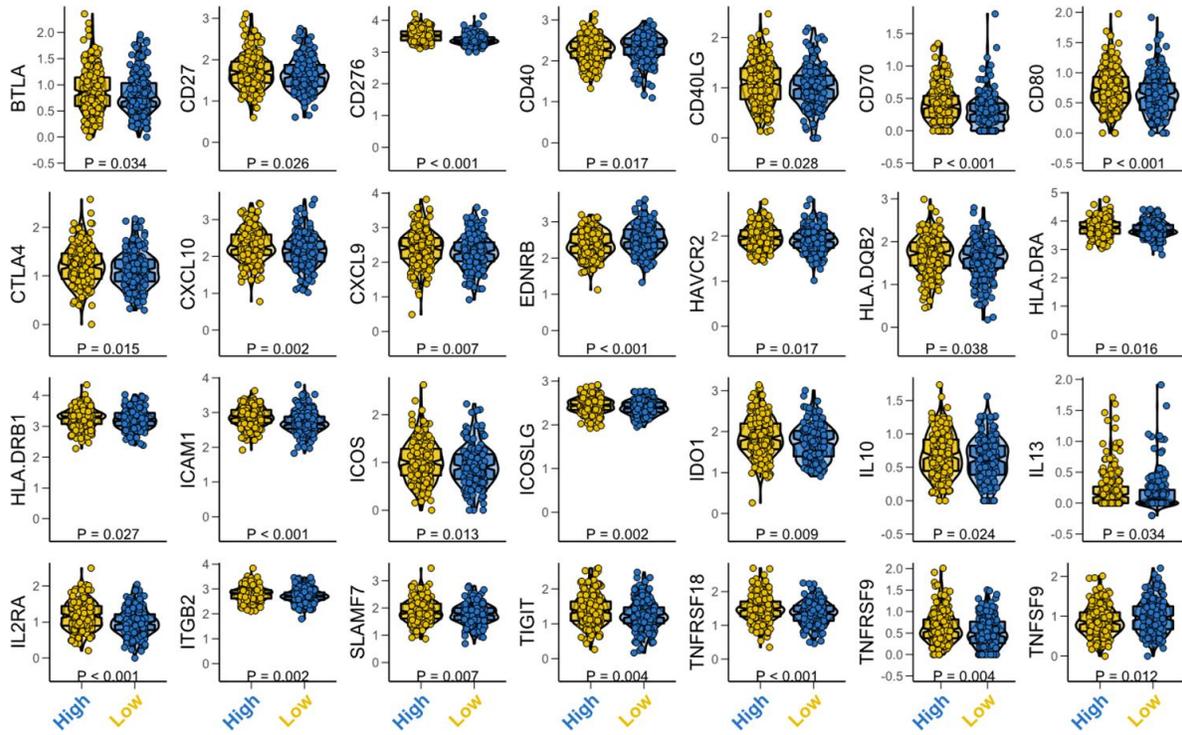
### Supplementary Figures



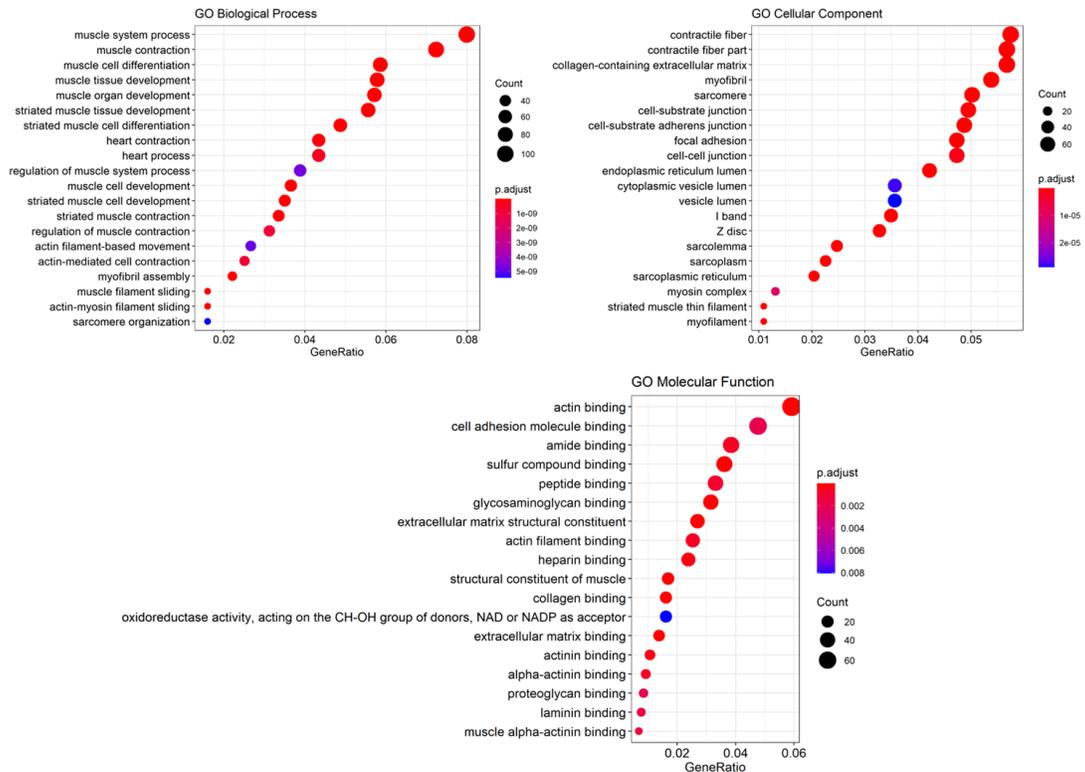
**Supplementary Figure 1.** Volcano map showing genes that were differential expressed between the high and low risk score groups. Red dots represent genes that are up-regulated in the high risk group, and blue dots represent genes that are down-regulated in the high risk group.



**Supplementary Figure 2.** (A) Immune scores and (B) matrix scores in the high and low risk groups.



Supplementary Figure 3. 28 immunomodulatory factors differentially expressed between the high and low risk score groups.



Supplementary Figure 4. GO enrich analysis was performed for differentially expressed genes.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 3 and 4.

**Supplementary Table 1. Clinical information of TCGA prostate cancer samples.**

<b>Characteristics</b>		<b>Counts or percent</b>
Age	median(range)	61(41-78)
Biochemical_recurrence(BCR)	Yes	68(12.8%)
	No	464(87.2%)
Days_to_BCR	median(range)	708.0(58.0-2459.0)
Days_to_last_follow_up	median(range)	958(23-5024)
Clinical_T	T1a	1(0.2%)
	T1b	2(0.4%)
	T1c	212(42.1%)
	T2	13(2.6%)
	T2a	72(14.3%)
	T2b	72(14.3%)
	T2c	62(12.3%)
	T3a	48(9.5%)
	T3b	19(3.8%)
	T4	2(0.4%)
Days_to_psa	median(range)	518.5(-164.0 - 3447.0)
Gleason_score	median(range)	7(6-10)
Pathologic_N	N0	442(82.2%)
	N1	96(17.8%)
Pathologic_T	T2a	14(2.3%)
	T2b	12(1.9%)
	T2c	211(34.3%)
	T3a	203(33.0%)
	T3b	162(26.3%)
	T4	14(2.3%)
Psa_value	median(range)	0.1(0-323)
Radiation_therapy	NO	448(86.8%)
	YES	68(13.2%)
Race.demographic	american indian or alaska native	1(0.16%)
	asian	13(2.09%)
	black or african american	81(13.00%)
	not reported	18(2.89%)
	white	510(81.9%)

**Supplementary Table 2. Clinical information of GSE54460 prostate cancer samples.**

<b>Characteristics</b>		<b>Counts or percent</b>
Age	median(range)	61.7(43-78)
Concentration(ng.µl)	median(range)	95.39(17.04-757.53)
race		
	Black	22(31.4%)
	White	48(68.6%)
Ratio(260/230)	median(range)	0.87(0.16-1.86)
Rpl13a.Ct.value	median(range)	28.61(24.50-30.90)
Total yield(µg)	median(range)	6.55(1.50-25.36)
BCR		
	0	51(48.1%)
	1	55(51.8%)
months to BCR	median(range)	21.40(0-154.22)
months to lastFollow	median(range)	68.50(0.70-180.56)

**Supplementary Table 3. 221 immune-related genes included in the candidate gene set.**

**Supplementary Table 4. Results of GSEA enrichment analysis for the high and low risk score groups.**

**Supplementary Table 5. Clinical information of tissue microarray of prostate cancer.**

<b>Characteristics</b>	<b>Counts or mean</b>
Age (years)	68.29±10.20
<b>T classification (n)</b>	
T1	2
T2	49
T3	20
T4	2
<b>N classification</b>	
N0	66
N1	7
<b>M classification</b>	
M0	69
M1	4
<b>Grade</b>	
1	4
2	18
2-3	2
3	38
unknown	11
<b>Stage</b>	
I	3
II	43
IIB	1
III	15
IIIB	1
IV	10

## Serum KIAA1199 is an advanced-stage prognostic biomarker and metastatic oncogene in cholangiocarcinoma

Xiangyu Zhai<sup>1,2,\*</sup>, Wei Wang<sup>1,\*</sup>, Yunlong Ma<sup>1</sup>, Yijia Zeng<sup>3</sup>, Dandan Dou<sup>4</sup>, Haoning Fan<sup>5</sup>, Jianping Song<sup>1</sup>, Xin Yu<sup>4</sup>, Danqing Xin<sup>4</sup>, Gang Du<sup>1</sup>, Zhengchen Jiang<sup>1</sup>, Hao Zhang<sup>1</sup>, Xinlu Zhang<sup>6</sup>, Bin Jin<sup>1,2</sup>

<sup>1</sup>Department of Surgery, School of Medicine, Cheeloo College of Medicine, Shandong University, Jinan, China

<sup>2</sup>Department of General Surgery, Qilu Hospital of Shandong University, Jinan, China

<sup>3</sup>Radiology Department, Qilu Hospital of Shandong University, Jinan, China

<sup>4</sup>School of Basic Medical Sciences, Shandong University, Jinan, China

<sup>5</sup>College of Traditional Chinese Medicine, Southern Medical University, Guangzhou, China

<sup>6</sup>Ji Nan Central Hospital, Jinan, China

\*Equal contribution

**Correspondence to:** Bin Jin; email: [jinbin9449@126.com](mailto:jinbin9449@126.com), <https://orcid.org/0000-0002-1411-2631>

**Keywords:** cholangiocarcinoma, KIAA1199, serum, bile, biomarker

**Received:** August 31, 2019

**Accepted:** July 20, 2020

**Published:** November 10, 2020

**Copyright:** © 2020 Zhai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**Background:** Cell proliferation and migration are the determinants of malignant tumor progression, and a better understanding of related genes will lead to the identification of new targets aimed at preventing the spread of cancer. Some studies have shown that KIAA1199 (CEMIP) is a transmembrane protein expressed in many types of noncancerous cells and cancer cells. However, the potential role of KIAA1199 in the progression of cholangiocarcinoma (CCA) remains unclear.

**Results:** Analysis of cancer-related databases showed that KIAA1199 is overexpressed in CCA. ELISA, immunohistochemistry, Western blotting and qPCR indicated high expression levels of KIAA1199 in serum, CCA tissues and CCA cell lines. In the serum (n = 41) and large sample validation (n = 177) cohorts, higher KIAA1199 expression was associated with shorter overall survival and disease-free survival times. At the cellular level, KIAA1199 overexpression (OE) promoted CCA growth and metastasis. Subcutaneous tumor xenograft experiments showed that KIAA1199 enhances CCA cell proliferation. Additionally, the expression levels of components in the EMT-related TGF- $\beta$  pathway changed significantly after KIAA1199 upregulation and silencing.

**Conclusion:** KIAA1199 is a promising new diagnostic molecule and therapeutic target in CCA. The serum KIAA1199 level can be used as a promising clinical tool for predicting the overall postoperative outcomes of patients with CCA.

**Methods:** CCA-related KIAA1199 data were downloaded from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. To assess the prognostic impact of KIAA1199, an enzyme-linked immunosorbent assay (ELISA) was used to measure the serum level of KIAA1199 in 41 patients who underwent surgical resection. Immunohistochemical staining, Western blotting and qPCR were used to verify and retrospectively review the expression levels of KIAA1199 in cancer tissue specimens from 177 CCA patients. The effect of KIAA1199 on CCA was evaluated by cell-based functional assays and subcutaneous tumor xenograft experiments. The expression levels of proteins associated with epithelial-mesenchymal transition (EMT) and activation of relevant signaling pathways were measured via Western blotting.

## INTRODUCTION

Cholangiocarcinoma (CCA) is an epithelial malignancy that occurs in different parts of the biliary tree and constitutes the second most common primary hepatic malignancy. CCA can be classified into three main types according to its anatomical location: perihepatic, distal and intrahepatic CCA (pCCA, dCCA and iCCA, respectively) [1]. The incidence of CCA is lower in Western countries (0.1-0.2/100,000 population) than in Southeast Asia (50-113/100,000 population), and CCA affects more males than females [2]. Currently, surgical lesion removal is a common curative treatment for CCA. Because of the asymptomatic nature of early-stage CCA, most patients are diagnosed at an advanced stage [3]. Surgical resection does not significantly improve long-term survival, and the rates of recurrence and metastasis are high [4–8]. Therefore, identification of a biomarker associated with CCA metastasis and prognosis is critical for the treatment of CCA. These biomarkers can serve as potential diagnostic molecules and therapeutic targets in CCA.

KIAA1199 (CEMIP, <https://www.ncbi.nlm.nih.gov/gene/57214>) is a protein indicator of cell migration that is localized to the perinuclear space (likely the ER) and cell membrane and is encoded by a gene located on chromosome 15q25.1 [9–11]. A related study showed that patients developed hearing loss due to mutations in KIAA1199 [12]. Moreover, severe tumor invasion and uncontrolled proliferation are associated with KIAA1199 overexpression (OE). Several studies have supported these observations in various cancers, including gastric carcinoma [13], breast carcinoma [14], colorectal tumors [15], hepatocellular carcinoma [16], prostate carcinoma [17] and oral squamous cell carcinoma [18]. In addition, these studies identified the clinical relevance of KIAA1199 to cancer (e.g., disease stage/5-year survival rate), but in most of the studies, the cohort was too small to allow any meaningful conclusions to be drawn. Moreover, KIAA1199 has been found in cell supernatants [19]. Therefore, KIAA1199 may be a secretory factor that promotes tumor development. In this study, we examined the expression level of KIAA1199 in CCA and sought to determine whether serum KIAA1199 levels can be used as a postoperative prognostic indicator.

## RESULTS

### Public database forecast results

The TCGA and GEO databases were used to compare KIAA1199 mRNA expression levels in patients with

CCA. KIAA1199 expression was significantly higher in CCA patients than in normal subjects (Figure 1A, 1B).

### KIAA1199 expression in human CCA tissues and cell lines

We evaluated KIAA1199 expression in CCA and adjacent tissues from 177 patients by immunohistochemistry. KIAA1199 expression was higher in carcinoma tissues (143/177 (80.8%); Table 1, Figure 1C) than in paracancer tissues (85/177 (48%); Table 1, Figure 1C). In addition, KIAA1199 expression levels were measured in twenty pairs of clinical CCA and paracancer tissues by qPCR and Western blotting. KIAA1199 protein and mRNA levels were higher in CCA tissues than in normal tissues ( $P < 0.001$  for KIAA1199; Figure 1D, 1E).

Among the CCA cell lines, HucCCT1, RBE and HCCC9810 cells had the highest and QBC939 cells had the lowest KIAA1199 expression levels (Figure 1F, 1G). Hucct-1 were thus used as the cell model for RNA knockdown and QBC939 were used as the cell model for RNA overexpression.

### Localization of KIAA1199

To unravel the mechanism underlying KIAA1199-mediated cell migration, we determined the subcellular localization of KIAA1199 by employing immunostaining and fluorescent tagging approaches. The Hucct-1 cell line and CCA tissue exhibited high mRNA and protein expression levels of KIAA1199 and were thus used to examine the subcellular localization of endogenous KIAA1199 protein. Immunofluorescence analysis showed that KIAA1199 is localized primarily in the perinuclear space of CCA cells (presumably the ER, including the outer nuclear membrane and ER tubules) and the plasma membrane (Figure 2A, 2B). KIAA1199 was fused with GFP at the C-terminus and transfected into QBC939 cells. We found that KIAA1199 was expressed on the cell membrane. GFP (lack of signal peptide; spread throughout cells) was used as a control (Figure 2C).

The SignalP v5.0 results indicated that KIAA1199 contains signal peptides (Figure 2D). According to TMHMM Server v2.0, KIAA1199 comprises 602 amino acids, all of which are extracellular, and no transmembrane domain (Figure 2E). Thus, the SignalP v5.0 and TMHMM Server v2.0 results also indicated that KIAA1199 is a classically secreted protein.

In addition, large amounts of exogenous KIAA1199 protein were secreted by CCA cells. We detected the KIAA1199 protein in the culture medium of Hucct-1

cells. However, after the expression of KIAA1199 was silenced, the KIAA1199 protein was not detected in the culture medium (Figure 2F).

### KIAA1199 is a secreted molecule and is highly expressed in bile and serum

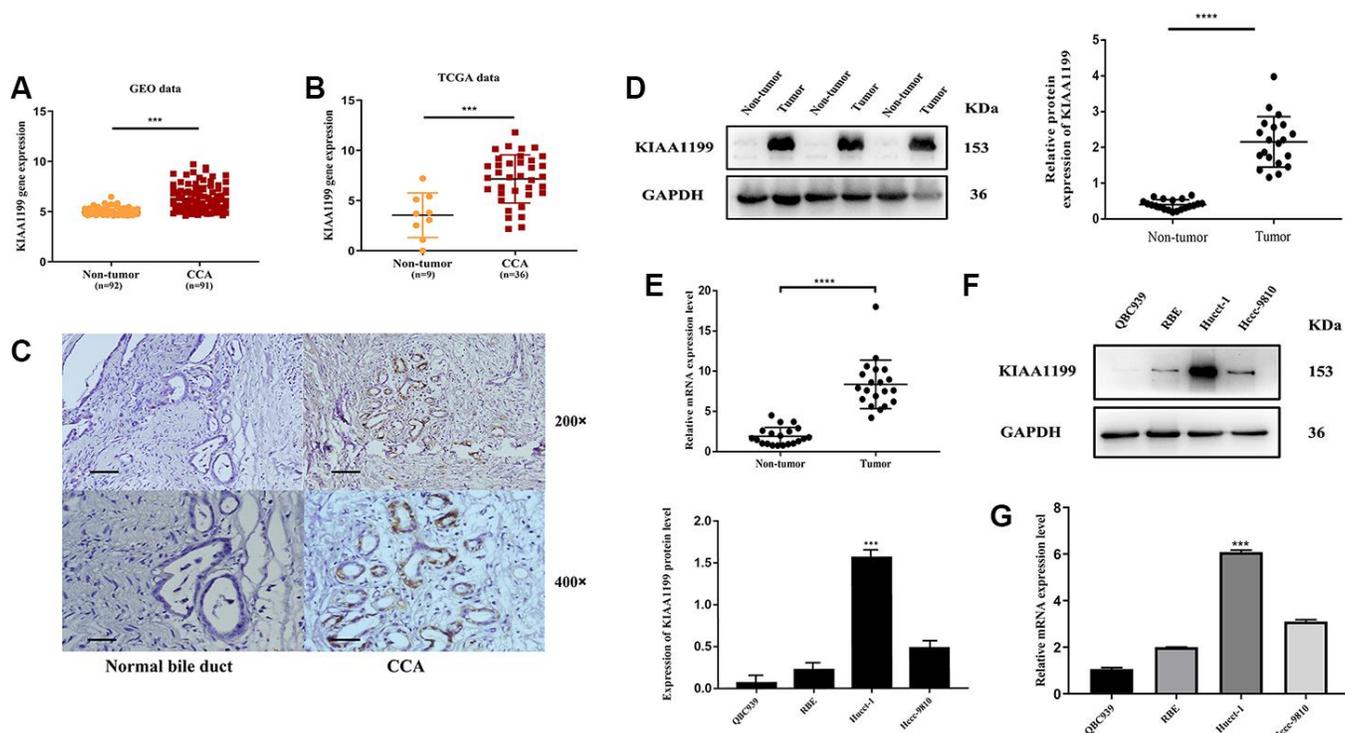
Then, we used ELISA to detect the expression of KIAA1199 in bile and serum. KIAA1199 levels were high in bile from CCA patients with cholestasis (median: 13.54 ng/ml (range: 3.059-26.39 ng/ml), Figure 3A) and no expression of KIAA1199 was detected in bile of healthy individuals (patients with right hepatic hemangioma or who had undergone cholecystectomy). In addition, we measured KIAA1199 levels in 41 paired serum samples from CCA patients and healthy subjects. As shown in Figure 3B, serum KIAA1199 levels were significantly higher in the CCA patients than in the normal subjects (median: 0.1842 (range: 0-1.75) ng/ml vs. 4.78 (range: 0.365-11.698) ng/ml,  $P < 0.001$ ).

A ROC curve was constructed (Figure 3C). Compared with the traditional diagnostic marker CA199, serum KIAA1199 has obvious diagnostic advantages. The area

under the curve (AUC) for serum KIAA1199 is 0.995, and that for CA199 is 0.894.

### Relationship between serum KIAA1199 levels and prognosis

We investigated the relationship between serum KIAA1199 levels and prognosis in patients with CCA. In the serum cohort, we evaluated the preoperative serum KIAA1199 levels in 41 patients with CCA. Most patients had advanced-stage CCA (stage III/IV,  $n = 29$  [70.7%]). We stratified the CCA patients according to their serum expression level of KIAA1199 with a cutoff value of 4.7841 ng/ml, which was the median value in CCA patients. The patients were divided into two groups according to this cutoff value: a high serum KIAA1199 group ( $n=21$ ) and a low serum KIAA1199 group ( $n=20$ ). Survival curves were plotted using the Kaplan-Meier method, and survival times were assessed by the log-rank test. The overall survival (OS) times of patients with high serum KIAA1199 levels were significantly shorter than those of patients with low serum KIAA1199 levels ( $P < 0.05$ , Figure 3D). The median OS times and 36- and 60-month survival rates in the high and low serum KIAA1199 groups were 28



**Figure 1. Bioinformatics prediction.** (A) GEO: (GSE76297), Normal:  $n=92$ , CCA:  $n=91$ , ( $p < 0.001$ ). (B) TCGA database, Normal:  $n=9$ , CCA:  $n=36$ , ( $p < 0.001$ ). Protein and mRNA expression of KIAA1199 in CCA tissues and CCA cell lines: (C) The relative expression levels of KIAA1199 in Pathological sections. (D) The relative protein expression levels of KIAA1199 in 20 pairs of CCA and adjacent normal tissues ( $p < 0.001$ ). (E) The relative mRNA expression levels of KIAA1199 in 20 pairs of CCA and adjacent normal tissues ( $p < 0.001$ ). (F, G) The relative protein and mRNA expression of KIAA1199 in four CCA cell lines.

**Table 1. Differential expression of KIAA1199 in CCA tissues and corresponding paracarcinoma tissues (n=177).**

	KIAA1199 expression		P-value
	Low (%)	High (%)	
<b>Carcinoma tissues</b>	34 (19.2%)	143 (80.8)	<0.001
<b>Paracarcinoma tissues</b>	92 (52%)	85 (48%)	

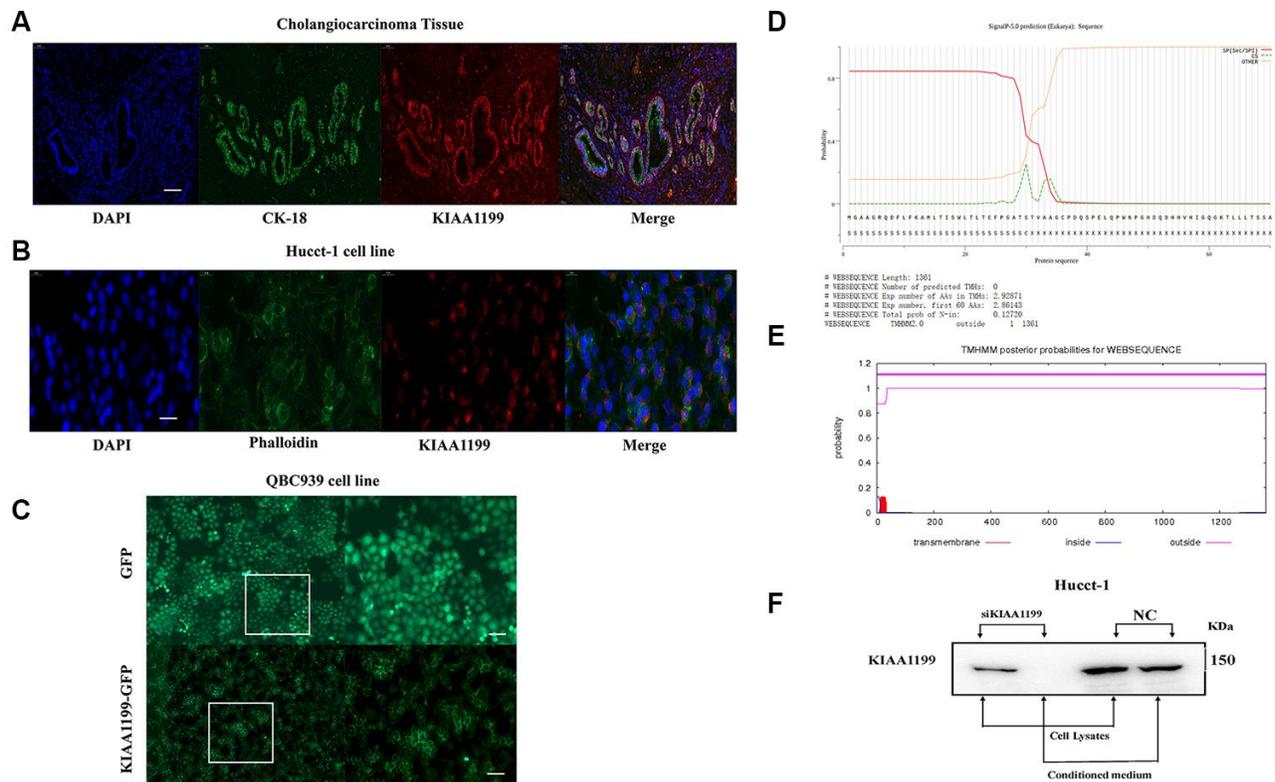
and 39 months, 37.5% and 51.2%, and 6.25% and 18.8%, respectively. Moreover, the disease-free survival (DFS) times of patients with high serum KIAA1199 levels were significantly shorter than those of patients with low serum KIAA1199 levels ( $P < 0.05$ , Figure 3E). The median DFS times and 12- and 36-month DFS rates in the high and low serum KIAA1199 groups were 27.0 and 34.0 months, 31.5% and 38.5%, and 6.25% and 20.9%, respectively.

Then, we evaluated the relationship between serum KIAA1199 expression and related clinical parameters (Table 2). Serum KIAA1199 levels in CCA patients were markedly related to LN metastasis ( $P < 0.05$ ) and

TNM stage ( $P < 0.05$ ). However, the KIAA1199 expression level was not significantly associated with other parameters, for example, age, sex, histological grade, CA19-9 level or tumor location.

### Relationship between KIAA1199 and epithelial-mesenchymal transition (EMT)

We examined the relationship between KIAA1199 and EMT-related proteins in CCA. The expression of KIAA1199 was higher in carcinoma tissues (Table 1, Figure 4A A1, A2) than in paracancer tissues (Table 1, Figure 4A E1, E2). The expression of E-cadherin was high in 49 of the 177 CCA specimens (27.7%; Table 3,



**Figure 2. (A)** Immunofluorescence localization of KIAA1199 and CK-18 in cholangiocarcinoma tissue (red, KIAA1199; green, CK-18; blue, DAPI). **(B)** Immunofluorescence localization of KIAA1199 and Phalloidin in Hucct-1 cell line (red, KIAA1199; green, Phalloidin; blue, DAPI). **(C)** Microscopic determination of KIAA1199 cellular localization using QBC939 cells transfected with green fluorescent protein (GFP), KIAA1199-GFP chimeric cDNAs. **(D)** KIAA1199 had signal peptides. **(E)** KIAA1199 has 602 amino acids, all of which are extracellular, and there is no transmembrane domain (TMD). **(F)** Western blot analysis was performed to detect KIAA1199 in cell lysates and conditioned culturing medium from Hucct-1 and Hucct-1 knockdown KIAA1199 (siKIAA1199).

Figure 4A B1, B2). According to Spearman's rank correlation analysis, KIAA1199 and E-cadherin exhibited a profound negative relationship in the CCA samples ( $r=-0.371$ ,  $P<0.001$ ; Table 3). N-cadherin expression was high in 127 of the 177 CCA tissue samples (71.8%; Table 3, Figure 4A C1, C2), and the KIAA1199 and N-cadherin expression levels were significantly positively correlated in CCA specimens according to Spearman's rank correlation analysis ( $r=0.311$ ,  $P<0.001$ ; Table 3). Additionally, vimentin expression was high in 136 of the 177 CCA tissues (76.8%; Table 3, Figure 4A D1, D2), and Spearman's rank correlation analysis identified a significant positive correlation between the KIAA1199 and vimentin expression levels in CCA tissues ( $r=0.351$ ,  $P<0.001$ ; Table 3).

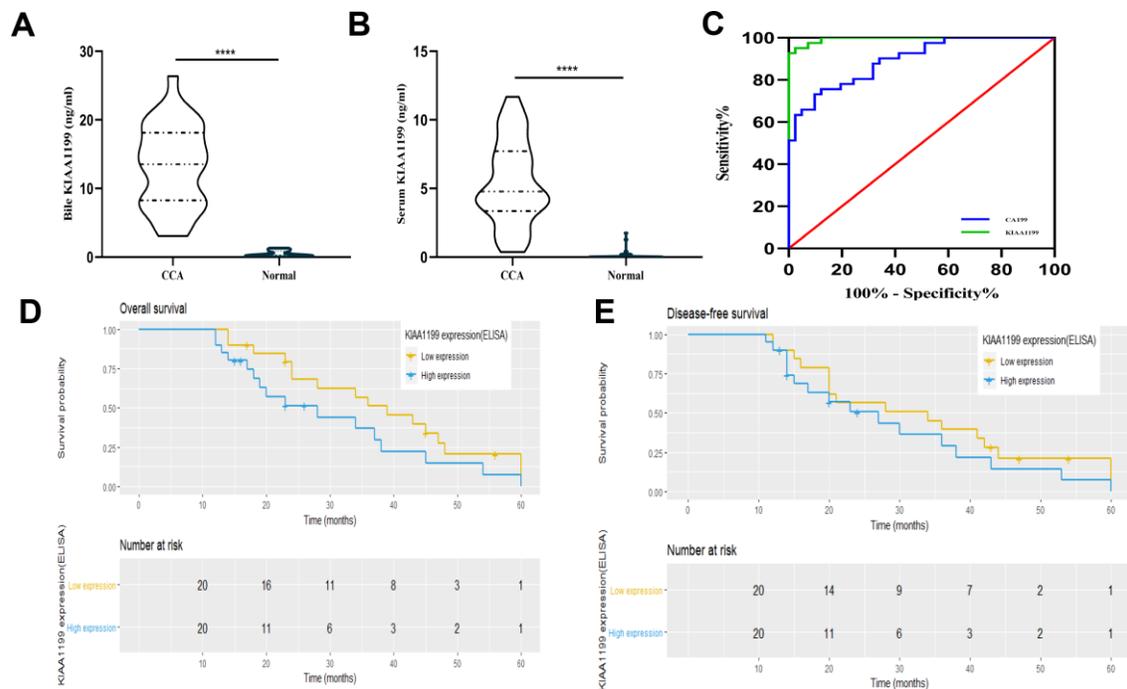
The potential effects of KIAA1199 on the expression levels of EMT-related proteins were further investigated in CCA cell lines. When KIAA1199 was knocked down in the HuCCT1 cell line, the protein expression level of E-cadherin was increased, while those of N-cadherin and vimentin were decreased (Figure 4B). In contrast, when KIAA1199 was overexpressed in the QBC939 cell line, the protein expression level of E-cadherin was decreased, while those of N-cadherin and vimentin were increased (Figure 4C).

## Validation in a large sample retrospective cohort

We used a large sample cohort to further validate the effect of KIAA1199 on the prognosis of CCA and to assess the relationship between KIAA1199 expression and clinical parameters in CCA patients. As shown in Table 4, high KIAA1199 expression was associated with the histological grade ( $P<0.05$ ), LN metastasis ( $P<0.05$ ), the TNM stage ( $P<0.05$ ), and the CA19-9 level ( $P<0.05$ ).

Survival curves were generated using the Kaplan-Meier method, and differences in the survival times were assessed by the log-rank test (Figure 5A). CCA patients with high KIAA1199 expression had longer DFS and OS times than those with low KIAA1199 expression ( $P<0.05$  for both). The median DFS times in the high and low KIAA1199 expression groups were 12 and 23 months, respectively. The median OS times in the high and low KIAA1199 expression groups were 15 and 56 months, respectively.

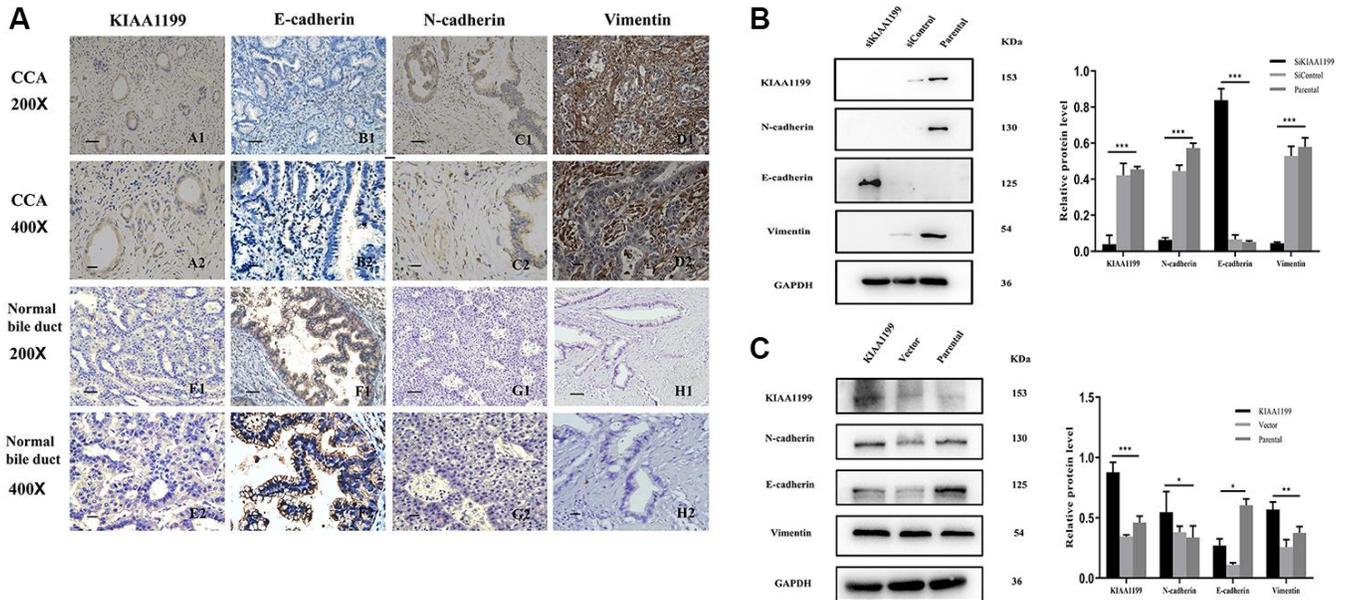
The univariate analysis results (Figure 5B) showed that age, sex, tumor location, histological grade, TNM stage, LN metastasis and CA19-9 level were associated with both OS and DFS. Multivariate survival analysis with correction (Figure 5C) showed that high KIAA1199 expression, high histological grade, advanced TNM



**Figure 3.** (A) Comparison of bile KIAA1199 levels in CCA patients (n=33) and healthy controls (n=15),  $p<0.001$ . (B) Comparison of serum KIAA1199 levels in CCA patients (n=41) and healthy controls (n=41),  $p<0.001$ . (C) ROC curve was used to evaluate serum KIAA1199 and ca199 diagnostic performances. (D, E) Kaplan-Meier analyses for overall survival (OS) and disease-free survival (DFS) in 41 CCA patients according to serum KIAA1199 levels. The quartile method is used to divide the data into four parts, and the dotted line in the figure is the dividing line of each part.

**Table 2. Correlations between serum KIAA1199 expression and clinicopathological characteristics in CCA.**

Parameters	KIAA1199 expression(ELISA)		P-value
	High(n=21)	Low(n=20)	
<b>Age (years)</b>			
≤60	8	6	0.585
>60	13	14	
<b>Gender</b>			
Male	12	13	0.606
Female	9	7	
<b>Tumor location</b>			
Intrahepatic	6	5	0.796
Perihilar/ Distal	15	15	
<b>Histological grade</b>			
Well	17	16	0.939
Moderate-poor	4	4	
<b>TNM stage</b>			
I-II	3	9	0.031
III-IV	18	11	
<b>Ca19-9(U/ml)</b>			
≤129	8	9	0.654
>129	13	11	
<b>Lymph Node Metastasis</b>			
No	3	11	0.006
Yes	18	9	



**Table 3. Correlations of KIAA1199 expression with N-cadherin, E-cadherin and vimentin in CCA.**

Immunoreactivity	KIAA1199 expression		r-value	P-value
	Low	High		
<b>N-cadherin expression</b>				
Low	20	30	0.311	<0.001
High	14	113		
<b>E-cadherin expression</b>				
Low	13	115	-0.371	<0.001
High	21	28		
<b>Vimentin expression</b>				
Low	19	22	0.351	<0.001
High	15	121		

**Table 4. Correlations between KIAA1199 expression and clinicopathological characteristics in CCA.**

Parameters	KIAA1199 expression		P-value
	Negative	Positive	
<b>Age (years)</b>			0.446
≤60	13	65	
>60	21	78	
<b>Gender</b>			0.786
Male	18	72	
Female	16	71	
<b>Tumor location</b>			0.698
Intrahepatic	5	30	
Perihilar/Distal	29	113	
<b>Histological grade</b>			0.012
Well	23	60	
Moderate-poor	11	83	
<b>TNM stage</b>			0.013
I-II	20	51	
III-IV	14	92	
<b>Ca19-9(U/ml)</b>			0.050
≤129	13	81	
>129	21	62	
<b>Lymph Node Metastasis</b>			0.012
No	18	43	
Yes	16	100	

stage and LN metastasis were associated with reduced OS and DFS times. These results indicate that KIAA1199 expression is an independent predictive factor for both DFS (HR=1.62, 95% CI: 1.03-2.55, P<0.05) and OS (HR=1.89, 95% CI: 1.21-2.97; P=0.005).

**KIAA1199 promotes CCA cell proliferation, cell migration and invasion**

To investigate the role of KIAA1199 in vitro, KIAA1199 was knocked down with siRNA. siRNA-1 was selected for further experiments because

it yielded the largest reduction in KIAA1199 expression in the transient transfection experiment according to both Western blotting and qPCR (Figure 6A, 6B).

The results of retrospective clinical studies suggest that KIAA1199 is involved in tumor migration and invasion. Therefore, we selected cells with low (QBC939) and high (HuCCT1) KIAA1199 expression for these in vitro experiments. After silencing KIAA1199 with siRNA-1 in Hucct-1 and overexpressing KIAA1199 in QBC939 with lentivirus carrying LV-KIAA1199 (Figure 6F, 6G).

The CCK-8 proliferation assay results showed that KIAA1199 knockdown suppressed Hucct-1 cell proliferation (Figure 6C) and KIAA1199 overexpression promoted QBC939 cell proliferation (Figure 6H). The results of wound healing assays showed that KIAA1199 overexpression enhanced but KIAA1199 knockdown suppressed CCA cell migration (Figure 6D, 6I). Moreover, the results of transwell assays revealed that KIAA1199 upregulation drastically increased but KIAA1199 knockdown dramatically decreased invasiveness (Figure 6E, 6J).

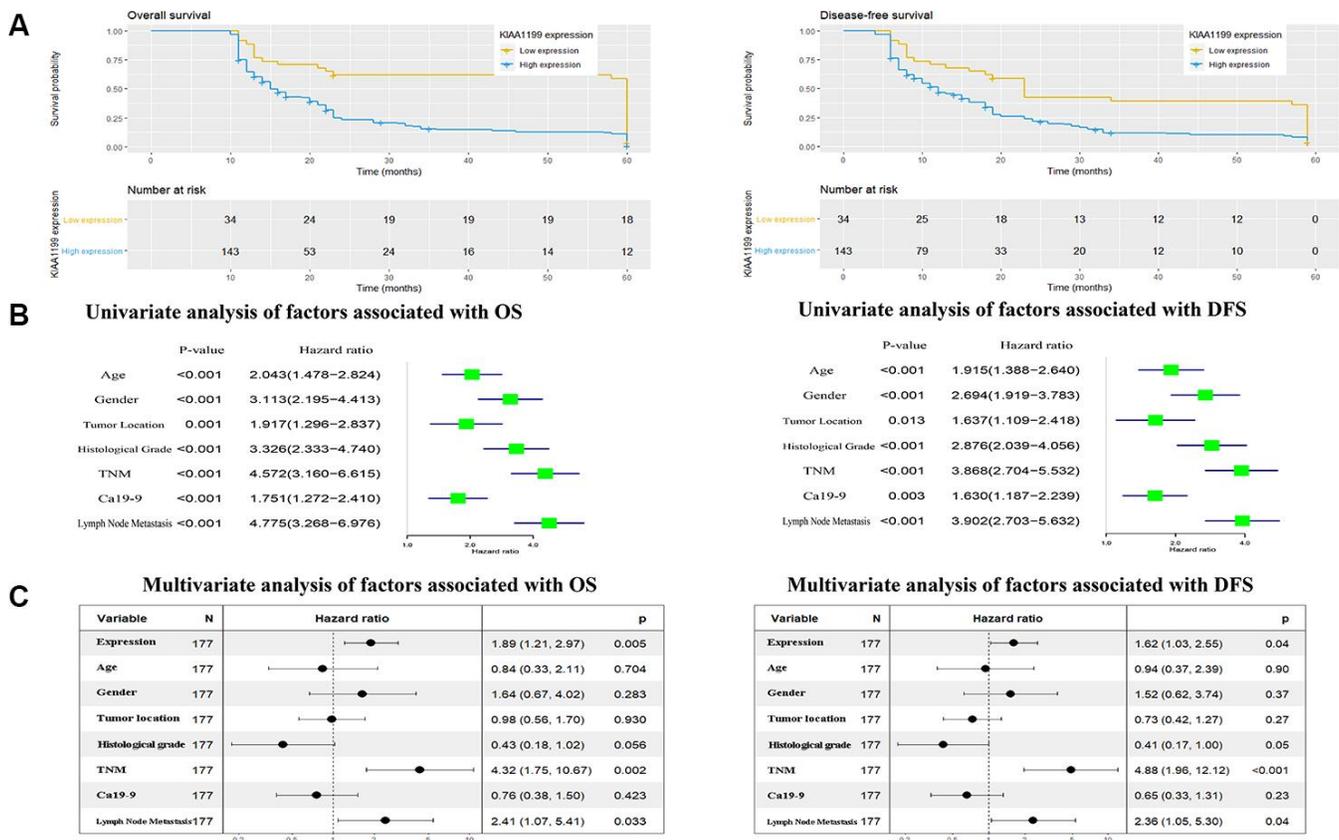
### KIAA1199 promotes CCA growth *in vivo*

Our studies of clinical data and cell lines indicated that KIAA1199 plays an important role in promoting the proliferation and development of CCA. We then examined the effect of KIAA1199 on tumor growth in a nude mouse subcutaneous xenograft model (Figure 7A). The size and weight of tumors from mice in the shKIAA1199 group were significantly reduced compared to those from mice in the control group ( $P < 0.001$ ). The size and weight of tumors from mice in the KIAA1199 group were significantly increased compared to those from mice in the vector group

( $P < 0.001$ ) (Figure 7B, 7C). KI67 results are consistent with the above results, KIAA1199 overexpression accelerates the proliferation of transplanted tumors (Figure 7D). Immunohistochemistry and WB were then performed to detect KIAA1199 expression in tumor tissues (Figure 7E, 7F).

### KIAA1199 upregulates the TGF- $\beta$ -PI3K-AKT-mediated EMT signaling pathway

Alterations in KIAA1199 expression can influence the behaviors, including proliferation and invasion, of CCA tumor cells. The immunostaining and Western blot results indicated that KIAA1199 expression was strongly positively correlated with the expression of EMT-related proteins, such as N-cadherin and vimentin, and was negatively correlated with the expression of E-cadherin. Thus, we measured the expression of proteins in the classical EMT pathway—the SMAD-independent pathway. TGF- $\beta$  was dramatically downregulated after KIAA1199 silencing, and the TGF- $\beta$ -regulating proteins PI3k, AKT and mTOR were significantly downregulated in response to KIAA1199 silencing (Figure 8A). In contrast, when KIAA1199 was overexpressed in the QBC939 cell line, the expression



**Figure 5.** (A) Kaplan-Meier analysis of overall survival (OS) and disease-free survival (DFS) in 177 patients with CCA according to KIAA1199 staining. (B, C) Univariate and multivariate analyses of factors associated with survival and recurrence.

levels of TGF- $\beta$ , PI3k, AKT and mTOR were higher in the OE cells than in the negative control cells (Figure 8B).

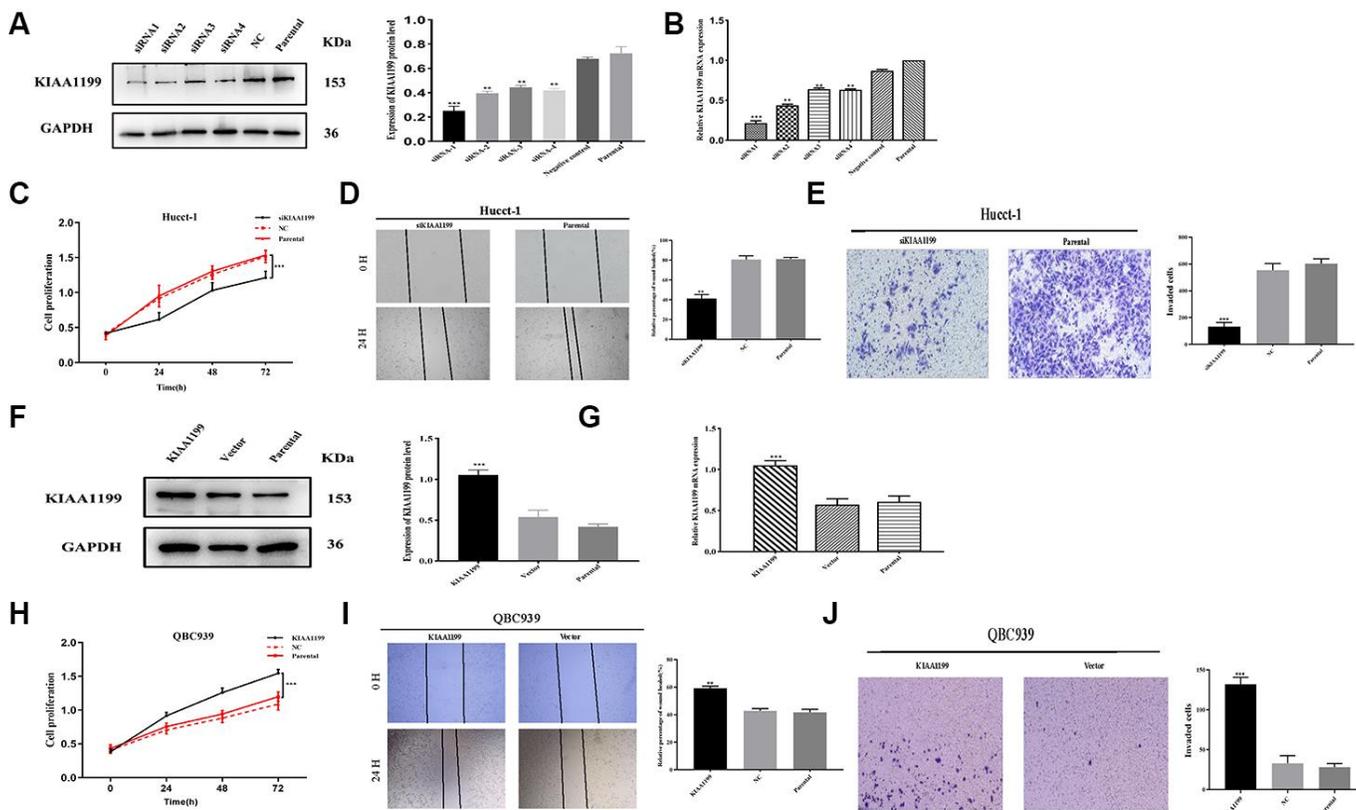
To further confirm that KIAA1199 enhances the proliferation and migration of CCA by upregulating the TGF- $\beta$ -PI3K-AKT-mediated EMT signaling pathway, we treated QBC939 cells with the small molecule TGF- $\beta$  inhibitor SB431542 and small molecule PI3K inhibitor LY294002. Both SB431542 and LY294002 inhibited KIAA1199-induced migration of cells cultured in conditioned medium (CM) containing KIAA1199 (Figure 8C, 8D).

## DISCUSSION

Jiang [16], Birkenkamp-Demtroder [11] and Matsuzaki [13] studied the role of KIAA1199 in hepatocellular cancer, colorectal cancer and gastric cancer. However,

in most studies, the research cohort was too small to allow any meaningful conclusions to be drawn. Moreover, the role of KIAA1199 in CCA is not well understood, and the available research is limited. For the first time, we identified the function and specific mechanism of KIAA1199 in human CCA progression by examining a large patient cohort and performing long-term follow-up evaluations. The present study constitutes a first step toward filling this research gap.

In previous studies [15], KIAA1199 was considered to be a glycosylated protein, localized primarily in the perinuclear space (likely the ER, including the outer nuclear membrane and ER tubules) and plasma membrane. Recently, researchers [19] also found that KIAA1199 has a parallel  $\beta$ -helix repeat (PbH1) domain, which comprises a  $\beta$ -helix repeat containing eight conserved glycine residues, five repeating  $\beta$ -strands, one  $\alpha$ -helix and two GG domains. Many G8-containing



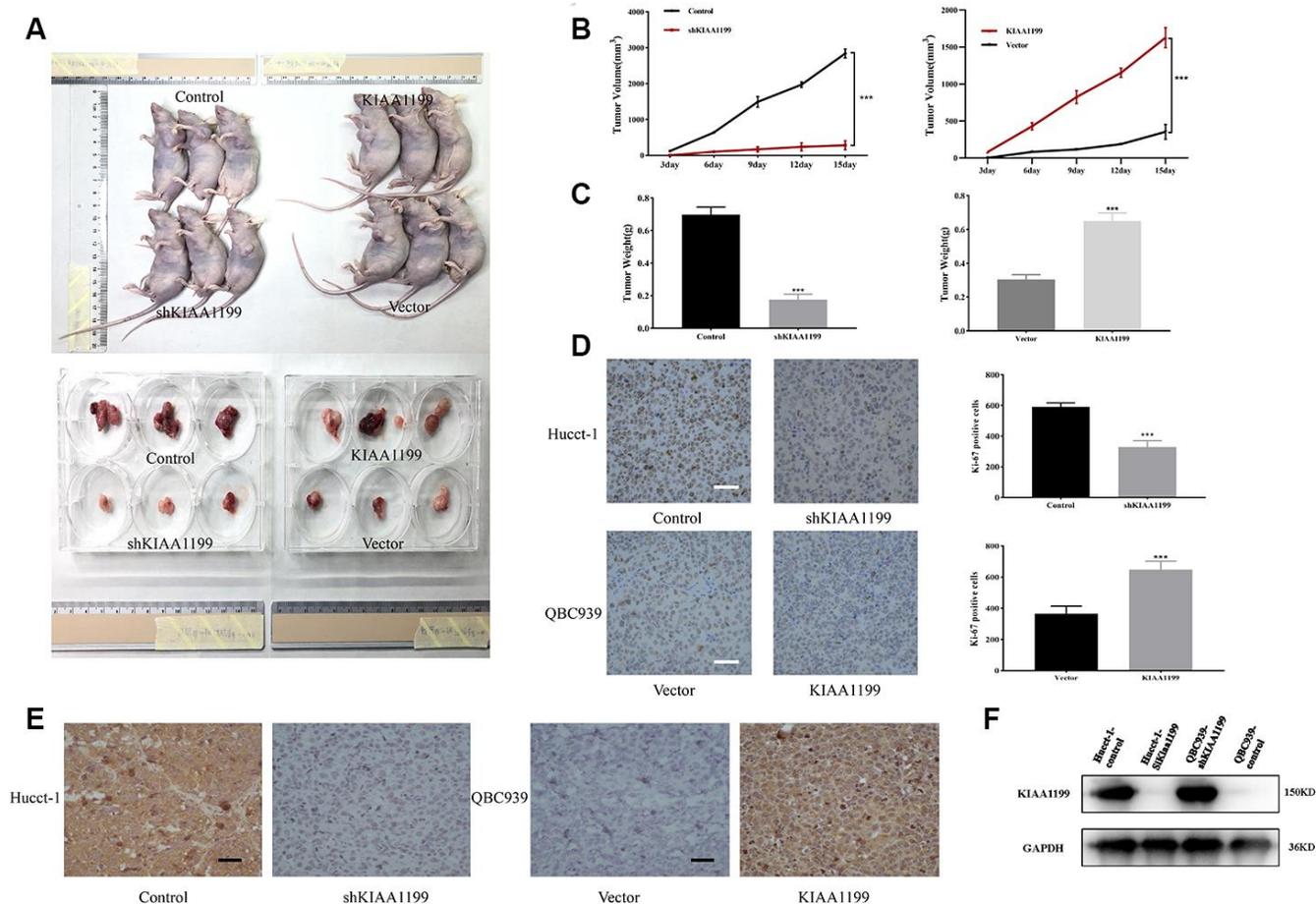
**Figure 6. KIAA1199 regulates proliferation and invasion in CCA cell lines.** (A, B) The relative protein and mRNA expression of four small interfering RNA in siKIAA1199-transfected cells compared with control and parental cells. (F, G) Overexpression of KIAA1199 in QBC939 cells with lentivirus infection was verified by western blotting and qPCR. Proliferation of Hucct-1 cells was detected with CCK-8 after silencing KIAA1199 in Hucct-1 cells (C) and overexpressing KIAA1199 in QBC939 cells (H) in normal medium with 10% FBS. Wound healing assay was applied to evaluate migration of Hucct-1 (D) and QBC939 (I). 24 h after a scratch in the cell monolayer, the wound size was measured again. Migration of Hucct-1 and QBC939 cells was assessed with transwell assay (E, J). After KIAA1199 knockdown and overexpression, cells were seeded in the upper transwell chamber and incubated for 24 h, with FBS in the lower chamber. (original magnification:  $\times 200$ ; scale bar, 20  $\mu$ m). Data, mean  $\pm$  S.D., and representative of three independent experiments.

proteins are often considered integral membrane proteins with signal peptides and/or transmembrane segments [12, 20, 21]. Thus, KIAA1199 may be a secretory factor that participates in extracellular ligand binding and processing. In this study, we provided evidence supporting this hypothesis via predictions with SignalP v5.0 and TMHMM Server v2.0 and detection of KIAA1199 in cell culture supernatant, bile and serum.

Bile is formed by transmembrane molecules expressed in the bile duct, including water channels (e.g., aquaporins), transporters (e.g., SGLT1, a Na<sup>+</sup>-glucose transporter) and transforming proteins (e.g., SLC4A2, a Cl<sup>-</sup>/HCO<sub>3</sub><sup>-</sup> exchanger) [22]. These molecules are expressed at the apical and basolateral membranes of cholangiocytes and promote the movement of water, electrolytes and solutes, thus altering bile volume and composition [23]. Cholestasis refers to the retention of

normal bile constituents, including toxic bile acids, within the liver. Patients with CCA often develop cholestasis because the tumor blocks the biliary tract. The high level of KIAA1199 detected in bile from CCA patients with cholestasis confirmed that KIAA1199 is a secreted protein. The high level of KIAA1199 detected in preoperative serum from patients with CCA further confirms this conclusion.

According to the findings of our studies, we conclude the following: (1) KIAA1199 is a secreted protein. (2) Serum KIAA1199 levels are higher in CCA patients than in normal individuals. (3) KIAA1199 expression is high in serum and resected tumor tissue. (4) High KIAA1199 levels can be detected in the bile of patients with CCA cholestasis. (5) Cells secrete KIAA1199 into the supernatant, as shown by cell-based experiments, and the amount of secreted KIAA1199 is related to the



**Figure 7.** (A) Xenografts were established in nude mice with stable KIAA1199 knockdown or overexpression cell. KIAA1199 knockdown decreased the volume (B) and weight (C) of xenograft tumors. KIAA1199 overexpression increased the volume and weight of xenograft tumors. Tumor diameter was measured every 3 day. (D) Ki67 staining showed that KIAA1199 can improve the proliferation ability of xenograft tumors. (E, F) Immunohistochemistry and WB detected the expression of KIAA1199 in xenograft tumors. \*, \*\* and \*\*\* represented  $P < 0.05$ ,  $P < 0.01$  and  $P < 0.001$  by Student's t-test, between the indicated groups.

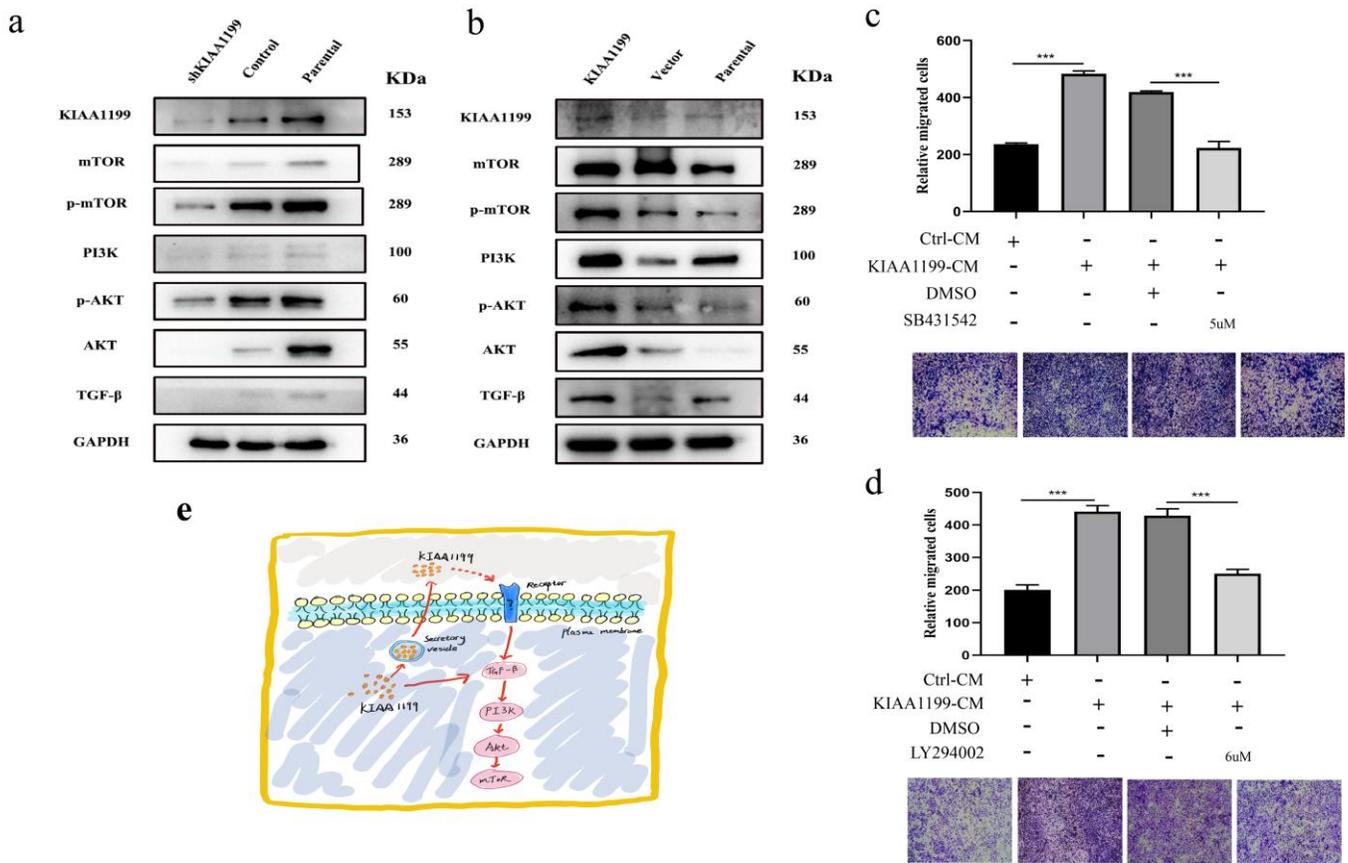
cellular expression level of KIAA1199 [19]. We speculate that KIAA1199 secreted by CCA tumors is a potential source of KIAA1199 in bile and serum.

Then, we explored the clinical implications of the serum KIAA1199 level in CCA patients prior to curative surgery, because the serum KIAA1199 level may be a marker for predicting poor prognosis in patients with CCA. Analysis of clinical medical records indicated that the diagnostic accuracy of serum KIAA1199 is higher than that of the traditional diagnostic marker CA199. In addition, OS and DFS were shown to differ significantly between the low and high serum KIAA1199 groups.

Patients in the high serum KIAA1199 expression group had worse OS and DFS than those in the low serum

KIAA1199 expression group. In a retrospectively reviewed large sample validation cohort, statistical analysis of the clinical information associated with the pathological sections confirmed that KIAA1199 upregulation is associated with poor prognosis. This result was consistent with the serum analysis results and indicates that high KIAA1199 expression is an independent predictor of poor prognosis.

The relationship between the KIAA1199 expression level and the prognosis of patients with CCA prompted us to further investigate the role of KIAA1199 at the cellular level. Our data showed that CCA cell migration and invasion were significantly inhibited after KIAA1199 silencing and confirmed that KIAA1199 overexpression promoted CCA metastasis and invasion. Thus, the expression level of KIAA1199 in CCA cell



**Figure 8. The expression of KIAA1199 and TGF-β-PI3K-Akt pathway-associated proteins by western blot analyses. (A)** Western blot analysis of KIAA1199 and TGF-β-PI3K-Akt pathway-associated proteins in KIAA1199 silenced Hucct1 cell line. **(B)** Western blot analysis of KIAA1199 and TGF-β-PI3K-Akt pathway-associated proteins in KIAA1199 overexpressed QBC939 cell line. **(C)** QBC939 were pretreated with TGF-β inhibitor (SB431542, 5 μM) for 2 h and transwell migration assay was performed in the absence or presence of KIAA1199 conditioned medium (CM). **(D)** QBC939 were pretreated with PI3K inhibitor (LY294002, 6 μM) for 2 h and transwell migration assay was performed in the absence or presence of KIAA1199 conditioned medium (CM). **(E)** KIAA1199-mediated EMT may occur through a non-Smad pathway. At least three independent experiments were preformed, data presented as mean ± SD, \*, \*\* and \*\*\* represented P<0.05, P < 0.01 and P < 0.001 by Student's t-test, between the indicated group.

lines influences their proliferation and migration, indicating that KIAA1199 plays an important role in CCA progression.

Patients with high KIAA1199 expression have poor prognoses, mainly due to the highly metastatic nature of this disease. Cell-based functional experiments again demonstrated the effects of KIAA1199 on cell proliferation and invasion. However, the mechanism by which KIAA1199 promotes tumor metastasis is not clear. During tumor metastasis, shed tumor cells enter the circulatory system by intravasation. Tumor cells in the circulatory system migrate out of the circulatory system by extravasation, enter distant tissues and form tiny metastatic clones; metastatic foci are then formed through proliferation of these cells [24]. Currently, the most studied mechanism is EMT. The role of EMT in CCA metastasis has received increasing attention [24]. EMT has been shown to significantly accelerate the metastasis of epithelial-derived carcinomas such as CCA [25, 26]. During EMT, epithelial tumor cells lose their polarity, their adhesion phenotype is altered to the mesenchymal phenotype, and their invasion and migration capacities are enhanced. Moreover, EMT is an early event in the distant metastasis of tumor cells [27, 28]. EMT can accelerate the growth of tumor cells [29] and promote tumor invasion and metastasis. E-cadherin [30], N-cadherin [31] and vimentin [28] are important protein markers of EMT. Decreased levels of E-cadherin can lead to decreased cell adhesion, thus promoting cell invasion and metastasis, and loss of E-cadherin expression has historically been considered the most prominent feature of EMT [32]. In addition, the expression of vimentin, N-cadherin and other proteins is increased in cells exhibiting the interstitial phenotype [31].

Here, we found that low E-cadherin expression correlates with high KIAA1199 expression in CCA. In contrast, the expression levels of vimentin and N-cadherin were positively correlated with that of KIAA1199 in CCA. Therefore, the expression levels of EMT-related proteins are closely related to high KIAA1199 expression levels in CCA, indicating that KIAA1199 may promote CCA by enhancing EMT.

Numerous signaling pathways mediate EMT. The TGF- $\beta$  signaling pathway is thought to be the most important signaling pathway for EMT induction during development and in cancer, and other pathological conditions [33]. In some in vitro cultured epithelial cell lines, TGF- $\beta$  stimulation alone can induce EMT [34]. TGF- $\beta$  signaling-mediated EMT can be activated by either the classical SMAD-dependent pathway or the SMAD-independent pathway. In the TGF- $\beta$  signaling-mediated SMAD-independent signaling pathway, PI3K-AKT-mTOR signaling is activated, leading to

transcriptional regulation. In addition, activated AKT can trigger EMT by inhibiting the transcriptional regulation of ribonucleoprotein E1 (hnRNPE1) [33]. In this study, results clarify that KIAA1199-mediated EMT may occur via the SMAD-independent pathway (Figure 8E).

Although our research links exogenous KIAA1199 with the proliferation and migration of CCA cells, the specific mechanisms by which KIAA1199 affects CCA cells remain unclear (e.g., whether the effects are autonomous or nonautonomous and whether a KIAA1199 receptor or binding partner is involved).

Although various models have been used to verify the role of KIAA1199 in different cell lines, and the role of KIAA1199 has been examined in xenograft models, the lack of a gene knockout mouse model prohibits further evaluation of the physiological effects of KIAA1199. This inability is another limitation of this study. Moreover, since this study was conducted primarily through retrospective analysis, the results may be biased. Prospective cohorts are needed to further validate our conclusion.

Despite these limitations, our study is the first to demonstrate the potential of KIAA1199 as a novel prognostic biomarker for CCA. In addition, KIAA1199 is a promising new diagnostic molecule and therapeutic target in CCA. Serum KIAA1199 levels are also a potential clinical tool for predicting tumor recurrence and overall prognosis in patients after curative surgery.

## MATERIALS AND METHODS

### Bioinformatics prediction

KIAA1199 expression in CCA was predicted using the cancer-related databases The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). The GSE76297 (normal: n=92, CCA: n=91) and TCGA (normal: n=9, CCA: n=36) datasets were downloaded from these databases for differential gene expression analysis in CCA. We used the TCGA database-based tumor data analysis website (UALCAN, <http://ualcan.path.uab.edu/>) to obtain clinical data analysis results.

### CCA cell lines

The QBC939, HCCC9810, HuCCT1 and RBE cell lines were purchased from Procell Life Science & Technology Corporation (Wuhan, China). CCA cell lines were cultured in RPMI 1640 medium (HyClone, Canada) or DMEM (HyClone, Canada) supplemented with 10% fetal bovine serum (Gibco, NY, USA) and 100 U/ml antibiotic

solution (Gibco, NY, USA). The cell incubator was maintained at 37°C with 5% CO<sub>2</sub> and 95% humidity.

### Patient enrollment and follow-up

From January 2007 to December 2017, 177 patients undergoing curative resection for pathologically confirmed CCA were enrolled in this study. The patients' clinical information, including age, sex, tumor location, tumor histological grade, tumor-node-metastasis (TNM) stage, lymph node (LN) metastasis status, and serum carbohydrate antigen 19-9 (CA19-9) level, was recorded. Table 4 lists the detailed clinical information of the 177 CCA patients in this study. Twenty fresh frozen cancer tissues, 41 preoperative serum samples from CCA patients (Table 3), 15 bile samples from healthy individuals (patients with right hepatic hemangioma or who had undergone cholecystectomy) and 33 bile samples from CCA patients were collected between 2010 and 2020. In accordance with the Declaration of Helsinki, the patients provided signed written consent for the collection of bile, serum and tissue samples used in this study. This study was approved by the Medical Ethics Committee of Shandong University and Qilu Hospital of Shandong University.

### Secreted protein prediction

The presence of N-terminal signal peptides and the lack of transmembrane domains (TMDs) are the two major characteristics of secreted proteins. Proteins that meet these two criteria are classified as the computational secretome [35].

We used SignalP v5.0 (<http://www.cbs.dtu.dk/services/SignalP/index.php>) to identify N-terminal signal peptides and TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) to predict transmembrane domains with the default parameters. The KIAA1199 [Homo sapiens] proteome was obtained from the NCBI database (<https://www.ncbi.nlm.nih.gov/protein/>).

### Measurement of KIAA1199 levels in bile and serum

KIAA1199 levels in human bile and serum samples were measured. Whole blood samples were incubated in a serum separation tube (SST) at 4°C overnight and were then centrifuged at 1000 × g for 15 min. The serum was immediately removed and tested, or the sample was aliquoted and stored at -80°C. Repeated freeze-thaw cycles were avoided. Bile was collected from CCA patients with cholestasis, and the samples were centrifuged for 20 min at 12000 × g and 2-8°C. The levels of KIAA1199 in serum and bile were measured by enzyme-linked immunosorbent assay

(ELISA). The KIAA1199 levels were quantified with Human KIAA1199 ELISA kits (SER965Hu, Cloud-Clone, USA and CSB-E13092h, CUSABIO, China).

### Immunohistochemistry

A two-step protocol was used for the immunohistochemical analysis according to the manufacturer's instructions (SP-9001, ZSGB-BIO; Beijing, China). The pathological sections were dewaxed, subjected to antigen retrieval in a microwave oven, and incubated with the appropriate primary antibodies (diluted 1:100) overnight. Subsequently, the pathological sections were incubated for 30 min with the secondary antibody (SA00004-2; Proteintech, Wuhan, China; diluted 1:400) at room temperature and were stained with diaminobenzidine (DAB Substrate Kit, DA1010; Solarbio, China). Finally, the pathological sections were dehydrated and fixed after counterstaining with hematoxylin. Images were acquired at 200× and 400× magnification with an Eclipse 80i microscope (Nikon, Tokyo, Japan).

The following primary antibodies were used: polyclonal rabbit anti-KIAA1199 (21129-1-AP; Proteintech; Wuhan, China); anti-E-cadherin (20874-1-AP; Proteintech); anti-N-cadherin (22018-1-AP; Proteintech); and anti-vimentin (10366-1-AP; Proteintech), anti-KI67 (27309-1-AP; Proteintech).

Semiquantitative methods were used to quantify the immunohistochemical results, namely, the staining intensity (absent, weak, moderate, strong=0, 1, 2, 3) and percentage of stained cells (<10%, 10%–50%, >50%=1, 2, 3). A total score of <3 for the sum of the scores for the two parameters was considered to indicate low expression. A total score of ≥3 was considered to indicate high expression. All results were scored independently by two experienced pathologists.

### Immunofluorescence

Frozen sections (4-μm thick) of liver tissue were prepared for immunofluorescence staining. In brief, sections were stained by overnight incubation with primary antibodies, such as anti-KIAA1199 (Proteintech, 1:200), anti-CK18 (Abcam, 1:200), and anti-phalloidin (Abcam, 1:200). Alexa 488-conjugated goat anti-rabbit (A11008, Thermo Fisher; 1:1000) or Alexa 568-conjugated goat anti-mouse (A-21043, Invitrogen; 1:1000) secondary antibodies were used. After three washes with PBST, the slides were incubated with DAPI (Beyotime, 1:2000) for 5 min. In addition, LC3 lentivirus was transfected into CCA cells according to the manufacturer's recommendations. Images were acquired with a fluorescence microscope (LSM 780, Carl Zeiss).

**Table 5. Related sequences of KIAA1199.**

	Sequence
<b>KIAA1199 F</b>	5' -AGGCGTGACACTGTCTCGGCTACAG-3'
<b>KIAA1199 P</b>	5' -CCACTCCACGTCTTGAACCCAC-3'
<b>GAPDH F</b>	5' -TTGGTATCGTGGAAGGACTCA-3'
<b>GAPDH R</b>	5' -TGTCATCATATTTGGCAGGTT-3'
<b>siRNA-1</b>	5'-GCAATCGTCCCATTGATATAC-3'
<b>siRNA-2</b>	5'-GCTGCAGGATCTGAGGAAACT -3'
<b>siRNA-3</b>	5'-GGTTATGACCCACCCACATAC -3'
<b>siRNA-4</b>	5'-GGGATAAGACATCTGTGTTCC-3'
<b>NC(negative control)</b>	5'-TTCTCCGAACGTGTACAGT-3'

### Western blot analysis

When cells reached ~80% confluence, the complete medium was removed, and the cells were washed with PBS and cultured in serum-free medium for 8 h. The medium was then collected and supplemented with 1 mM PMSF and 1× EDTA-free complete Protease Inhibitor. The mixture was centrifuged at 100 × g for 5 min to remove suspended cells and was then centrifuged at 3000 × g with an Amicon Ultra centrifuges (10 kDa NMWL) to concentrate the clarified medium.

Proteins were extracted from CCA cell lines or tissues with radioimmunoprecipitation assay (RIPA) buffer (89900, Thermo Scientific, California, USA) containing protease and phosphatase inhibitors (36978, Thermo Scientific, California, USA). Protein concentrations were measured according to the kit manufacturer's instructions (Beyotime BCA Protein Assay Kit, P0009, China). Then, proteins were separated by 10% SDS-PAGE and transferred to polyvinylidene fluoride (PVDF) membranes (Millipore, NY, USA). Membranes were blocked with 5% BSA blocking buffer (SW3015, Solarbio, Beijing, China) and incubated with primary antibodies (diluted 1:1000) overnight at 4°C. Membranes were then incubated for 1 h with the secondary antibody (diluted 1:5000), and the protein bands were visualized with a Western Fluorescence Kit (BeyoECL Plus, Beyotime, P0018S, China) and a Chemiluminescence Imaging System (T-4600, Tanon, Shanghai, China). Band densities were quantified with ImageJ, and the target protein levels were normalized to those of GAPDH.

### Quantitative RT-PCR (qPCR)

Total RNA was extracted from cell lines and tissues using a FastPure Cell/Tissue Total RNA Isolation Mini Kit (RC101, Vazyme, Nanjing, China) according to the manufacturer's protocol. The RNA concentration was determined, and total RNA was reverse transcribed to

cDNA (P113-01, Vazyme, Nanjing, China). qPCR was conducted using SYBR Green reagent (FP303, TIANGEN, Beijing, China) in a sequence detection system (ABI 7900HT, Applied Biosystems, MA, USA). The  $2^{-\Delta\Delta C_t}$  method was used to analyze the experimental results. All primer sequence information is shown in Table 5.

### KIAA1199 silencing and OE

For silencing, a small interfering RNA (siRNA) against KIAA1199 was designed. siRNA was transfected into CCA cell lines with Xfect™ RNA Transfection Reagent (631450, Takara, USA) according to the manufacturer's instructions. The target siRNA sequences are listed in Table 5. The lentiviral vectors overexpressing KIAA1199 (LV-KIAA1199) and the corresponding knockdown vectors (LV-shKIAA1199) were synthesized by Jinan Boshang Corporation (Jinan, China). Empty vectors were used as the corresponding controls.

### In vitro

We used CCK-8 (Selleck, USA) assays to measure the proliferation of related carcinoma cell lines. The absorbance at 490 nm was measured at 0 h, 24 h, 48 h, and 72 h. Wound healing assays were used to evaluate cell migration. The migrated cells at the edge of the wound were imaged at 0 and 24 h after wounding, and migration was assessed as follows: (average wound area at 24 h - average wound area at 0 h)/average wound area at 0 h. Transwell assays were used to evaluate cell invasion. Transwell inserts (Corning, USA) with an 8-µm pore size were placed in 24-well plates. After incubation for 24 h at 37°C in a 5% CO<sub>2</sub> atmosphere, the noninvasive cells remained on the upper surface of the filter, and the highly invasive cells that invaded through the filter were fixed with methanol and stained with crystal violet. Images were acquired at 200× magnification with an Eclipse TS100 microscope (Nikon, Tokyo, Japan).

## ***In vivo***

Twelve nude mice were randomly divided into 4 groups of 3 mice each. Hucct1-control, Hucct1-siKIAA1199, QBC939-control, and QBC939-shKIAA1199 CCA cells were suspended separately in sterile PBS, and these cell suspensions were injected into the left axilla of the nude mice. Each nude mouse was injected with 200  $\mu$ l of suspension containing approximately  $5 \times 10^6$  cells. The length (L) and width (W) of the tumors were measured using a Vernier caliper every three days, and the tumor volume (TV) was calculated according to the following formula:  $TV = (L \times W^2) / 2$ . The nude mice were sacrificed 15 days later, and the tumor tissues were completely resected and weighed. KIAA1199 expression in tumor tissues was detected via Western blotting.

## **Statistical analysis**

Statistical analysis was performed using SPSS 16.0 (Chicago, USA) and GraphPad Prism v. 6.01 (San Diego, CA, USA). The values are expressed as the mean  $\pm$  standard error of the mean. Receiver operating characteristic (ROC) curve analysis was used to evaluate the sensitivity and specificity of serum KIAA1199 for diagnosing CCA. An unpaired t-test was used to compare quantitative variables. Pearson's  $\chi^2$  test or Fisher's exact test was used to compare qualitative variables. The patient survival curves were plotted using the Kaplan-Meier method, and the log-rank test was used to identify significant differences among groups. Multivariate analysis was performed with a Cox regression model.  $P < 0.05$  was considered to indicate a statistically significant difference.

## **AUTHOR CONTRIBUTIONS**

Bin Jin designed experiments; Xiangyu Zhai and Wei Wang carried out experiments; Yunlong Ma and Yijia Zeng helped data curation; Xiangyu Zhai wrote the manuscript; Dandan Dou helped with Elisa and Western blotting experiments; Jianping Song, Xin Yu and Danqing Xin analyzed experimental results; Xinlu Zhang, Gang Du, Zhengchen Jiang and Hao Zhang made clinical investigation; Haoning Fan edited this manuscript.

## **CONFLICTS OF INTEREST**

The author reports no conflicts of interest in this work.

## **FUNDING**

This study is supported by the National Natural Science Foundation of China (NO. 81571367), Key R & D

project of Shandong Province (NO.2017GSF218021). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## **REFERENCES**

1. Razumilava N, Gores GJ. Cholangiocarcinoma. *Lancet*. 2014; 383:2168–79.  
[https://doi.org/10.1016/S0140-6736\(13\)61903-0](https://doi.org/10.1016/S0140-6736(13)61903-0)  
PMID:[24581682](https://pubmed.ncbi.nlm.nih.gov/24581682/)
2. Tyson GL, El-Serag HB. Risk factors for cholangiocarcinoma. *Hepatology*. 2011; 54:173–84.  
<https://doi.org/10.1002/hep.24351>  
PMID:[21488076](https://pubmed.ncbi.nlm.nih.gov/21488076/)
3. Jarnagin WR, Fong Y, DeMatteo RP, Gonen M, Burke EC, Bodniewicz BS J, Youssef BA M, Klimstra D, Blumgart LH. Staging, resectability, and outcome in 225 patients with hilar cholangiocarcinoma. *Ann Surg*. 2001; 234:507–17.  
<https://doi.org/10.1097/00000658-200110000-00010>  
PMID:[11573044](https://pubmed.ncbi.nlm.nih.gov/11573044/)
4. Puhalla H, Schuell B, Pokorny H, Kornek GV, Scheithauer W, Gruenberger T. Treatment and outcome of intrahepatic cholangiocellular carcinoma. *Am J Surg*. 2005; 189:173–77.  
<https://doi.org/10.1016/j.amjsurg.2004.11.009>  
PMID:[15720985](https://pubmed.ncbi.nlm.nih.gov/15720985/)
5. Shinohara ET, Mitra N, Guo M, Metz JM. Radiation therapy is associated with improved survival in the adjuvant and definitive treatment of intrahepatic cholangiocarcinoma. *Int J Radiat Oncol Biol Phys*. 2008; 72:1495–501.  
<https://doi.org/10.1016/j.ijrobp.2008.03.018>  
PMID:[18472359](https://pubmed.ncbi.nlm.nih.gov/18472359/)
6. Tamandl D, Herberger B, Gruenberger B, Puhalla H, Klingler M, Gruenberger T. Influence of hepatic resection margin on recurrence and survival in intrahepatic cholangiocarcinoma. *Ann Surg Oncol*. 2008; 15:2787–94.  
<https://doi.org/10.1245/s10434-008-0081-1>  
PMID:[18685896](https://pubmed.ncbi.nlm.nih.gov/18685896/)
7. Yedibela S, Demir R, Zhang W, Meyer T, Hohenberger W, Schönleben F. Surgical treatment of mass-forming intrahepatic cholangiocarcinoma: an 11-year western single-center experience in 107 patients. *Ann Surg Oncol*. 2009; 16:404–12.  
<https://doi.org/10.1245/s10434-008-0227-1>  
PMID:[19037702](https://pubmed.ncbi.nlm.nih.gov/19037702/)
8. Zografos GN, Farfaras A, Zagouri F, Chrysikos D,

- Karaliotas K. Cholangiocarcinoma: principles and current trends. *Hepatobiliary Pancreat Dis Int.* 2011; 10:10–20.  
[https://doi.org/10.1016/s1499-3872\(11\)60001-5](https://doi.org/10.1016/s1499-3872(11)60001-5)  
PMID:[21269929](https://pubmed.ncbi.nlm.nih.gov/21269929/)
9. Raish M, Khurshid M, Ansari MA, Chaturvedi PK, Bae SM, Kim JH, Park EK, Park DC, Ahn WS. Analysis of molecular cytogenetic alterations in uterine leiomyosarcoma by array-based comparative genomic hybridization. *J Cancer Res Clin Oncol.* 2012; 138:1173–86.  
<https://doi.org/10.1007/s00432-012-1182-6>  
PMID:[22419440](https://pubmed.ncbi.nlm.nih.gov/22419440/)
  10. Dash DP, Silvestri G, Hughes AE. Fine mapping of the keratoconus with cataract locus on chromosome 15q and candidate gene analysis. *Mol Vis.* 2006; 12:499–505.  
PMID:[16735990](https://pubmed.ncbi.nlm.nih.gov/16735990/)
  11. Birkenkamp-Demtroder K, Maghnouj A, Mansilla F, Thorsen K, Andersen CL, Øster B, Hahn S, Ørntoft TF. Repression of KIAA1199 attenuates Wnt-signalling and decreases the proliferation of colon cancer cells. *Br J Cancer.* 2011; 105:552–61.  
<https://doi.org/10.1038/bjc.2011.268> PMID:[21772334](https://pubmed.ncbi.nlm.nih.gov/21772334/)
  12. Abe S, Usami S, Nakamura Y. Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. *J Hum Genet.* 2003; 48:564–70.  
<https://doi.org/10.1007/s10038-003-0079-2>  
PMID:[14577002](https://pubmed.ncbi.nlm.nih.gov/14577002/)
  13. Matsuzaki S, Tanaka F, Mimori K, Tahara K, Inoue H, Mori M. Clinicopathologic significance of KIAA1199 overexpression in human gastric cancer. *Ann Surg Oncol.* 2009; 16:2042–51.  
<https://doi.org/10.1245/s10434-009-0469-6>  
PMID:[19434458](https://pubmed.ncbi.nlm.nih.gov/19434458/)
  14. Evensen NA, Kuscu C, Nguyen HL, Zarrabi K, Dufour A, Kadam P, Hu YJ, Pulkoski-Gross A, Bahou WF, Zucker S, Cao J. Unraveling the role of KIAA1199, a novel endoplasmic reticulum protein, in cancer cell migration. *J Natl Cancer Inst.* 2013; 105:1402–16.  
<https://doi.org/10.1093/jnci/djt224> PMID:[23990668](https://pubmed.ncbi.nlm.nih.gov/23990668/)
  15. Tiwari A, Schneider M, Fiorino A, Haider R, Okoniewski MJ, Roschitzki B, Uozio A, Menigatti M, Jiricny J, Marra G. Early insights into the function of KIAA1199, a markedly overexpressed protein in human colorectal tumors. *PLoS One.* 2013; 8:e69473.  
<https://doi.org/10.1371/journal.pone.0069473>  
PMID:[23936024](https://pubmed.ncbi.nlm.nih.gov/23936024/)
  16. Jiang Z, Zhai X, Shi B, Luo D, Jin B. KIAA1199 overexpression is associated with abnormal expression of EMT markers and is a novel independent prognostic biomarker for hepatocellular carcinoma. *Onco Targets Ther.* 2018; 11:8341–48.  
<https://doi.org/10.2147/OTT.S187389> PMID:[30538502](https://pubmed.ncbi.nlm.nih.gov/30538502/)
  17. Michishita E, Garcés G, Barrett JC, Horikawa I. Upregulation of the KIAA1199 gene is associated with cellular mortality. *Cancer Lett.* 2006; 239:71–77.  
<https://doi.org/10.1016/j.canlet.2005.07.028>  
PMID:[16157444](https://pubmed.ncbi.nlm.nih.gov/16157444/)
  18. Chanthammachat P, Promwikorn W, Pruegsanusak K, Roytrakul S, Srisomsap C, Chokchaichamnankit D, Svasti J, Boonyaphiphat P, K S, Thongsuksai P. Comparative proteomic analysis of oral squamous cell carcinoma and adjacent non-tumour tissue from Thailand. *Arch Oral Biol.* 2013; 58:1677–85.  
<https://doi.org/10.1016/j.archoralbio.2013.08.002>  
PMID:[24112734](https://pubmed.ncbi.nlm.nih.gov/24112734/)
  19. Chen L, Shi K, Andersen TL, Qiu W, Kassem M. KIAA1199 is a secreted molecule that enhances osteoblastic stem cell migration and recruitment. *Cell Death Dis.* 2019; 10:126.  
<https://doi.org/10.1038/s41419-018-1202-9>  
PMID:[30755597](https://pubmed.ncbi.nlm.nih.gov/30755597/)
  20. Jenkins J, Mayans O, Pickersgill R. Structure and evolution of parallel beta-helix proteins. *J Struct Biol.* 1998; 122:236–46.  
<https://doi.org/10.1006/jsbi.1998.3985> PMID:[9724625](https://pubmed.ncbi.nlm.nih.gov/9724625/)
  21. He QY, Liu XH, Li Q, Studholme DJ, Li XW, Liang SP. G8: a novel domain associated with polycystic kidney disease and non-syndromic hearing loss. *Bioinformatics.* 2006; 22:2189–91.  
<https://doi.org/10.1093/bioinformatics/btl123>  
PMID:[16632497](https://pubmed.ncbi.nlm.nih.gov/16632497/)
  22. Lazaridis KN, LaRusso NF. The cholangiopathies. *Mayo Clin Proc.* 2015; 90:791–800.  
<https://doi.org/10.1016/j.mayocp.2015.03.017>  
PMID:[25957621](https://pubmed.ncbi.nlm.nih.gov/25957621/)
  23. Banales JM, Sáez E, Uriz M, Sarvide S, Urribarri AD, Splinter P, Tietz Bogert PS, Bujanda L, Prieto J, Medina JF, LaRusso NF. Up-regulation of microRNA 506 leads to decreased Cl-/HCO3- anion exchanger 2 expression in biliary epithelium of patients with primary biliary cirrhosis. *Hepatology.* 2012; 56:687–97.  
<https://doi.org/10.1002/hep.25691> PMID:[22383162](https://pubmed.ncbi.nlm.nih.gov/22383162/)
  24. Tsai JH, Yang J. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* 2013; 27:2192–206.  
<https://doi.org/10.1101/gad.225334.113>  
PMID:[24142872](https://pubmed.ncbi.nlm.nih.gov/24142872/)
  25. Nitta T, Mitsuhashi T, Hatanaka Y, Miyamoto M, Oba K, Tsuchikawa T, Suzuki Y, Hatanaka KC, Hirano S, Matsuno Y. Prognostic significance of epithelial-mesenchymal transition-related markers in

- extrahepatic cholangiocarcinoma: comprehensive immunohistochemical study using a tissue microarray. *Br J Cancer*. 2014; 111:1363–72.  
<https://doi.org/10.1038/bjc.2014.415> PMID:[25077440](https://pubmed.ncbi.nlm.nih.gov/25077440/)
26. Baulida J. Epithelial-to-mesenchymal transition transcription factors in cancer-associated fibroblasts. *Mol Oncol*. 2017; 11:847–59.  
<https://doi.org/10.1002/1878-0261.12080> PMID:[28544627](https://pubmed.ncbi.nlm.nih.gov/28544627/)
27. Puisieux A, Brabletz T, Caramel J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol*. 2014; 16:488–94.  
<https://doi.org/10.1038/ncb2976> PMID:[24875735](https://pubmed.ncbi.nlm.nih.gov/24875735/)
28. Maier HJ, Wirth T, Beug H. Epithelial-mesenchymal transition in pancreatic carcinoma. *Cancers (Basel)*. 2010; 2:2058–83.  
<https://doi.org/10.3390/cancers2042058> PMID:[24281218](https://pubmed.ncbi.nlm.nih.gov/24281218/)
29. Solanas G, Porta-de-la-Riva M, Agustí C, Casagolda D, Sánchez-Aguilera F, Larriba MJ, Pons F, Peiró S, Escrivà M, Muñoz A, Duñach M, de Herreros AG, Baulida J. E-cadherin controls beta-catenin and NF-kappaB transcriptional activity in mesenchymal gene expression. *J Cell Sci*. 2008; 121:2224–34.  
<https://doi.org/10.1242/jcs.021667> PMID:[18565826](https://pubmed.ncbi.nlm.nih.gov/18565826/)
30. Handra-Luca A, Hong SM, Walter K, Wolfgang C, Hruban R, Goggins M. Tumour epithelial vimentin expression and outcome of pancreatic ductal adenocarcinomas. *Br J Cancer*. 2011; 104:1296–302.  
<https://doi.org/10.1038/bjc.2011.93> PMID:[21448168](https://pubmed.ncbi.nlm.nih.gov/21448168/)
31. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2014; 15:178–96.  
<https://doi.org/10.1038/nrm3758> PMID:[24556840](https://pubmed.ncbi.nlm.nih.gov/24556840/)
32. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal*. 2014; 7:re8.  
<https://doi.org/10.1126/scisignal.2005189> PMID:[25249658](https://pubmed.ncbi.nlm.nih.gov/25249658/)
33. Wendt MK, Tian M, Schiemann WP. Deconstructing the mechanisms and consequences of TGF- $\beta$ -induced EMT during cancer progression. *Cell Tissue Res*. 2012; 347:85–101.  
<https://doi.org/10.1007/s00441-011-1199-1> PMID:[21691718](https://pubmed.ncbi.nlm.nih.gov/21691718/)
34. Willis BC, Borok Z. TGF-beta-induced EMT: mechanisms and implications for fibrotic lung disease. *Am J Physiol Lung Cell Mol Physiol*. 2007; 293:L525–34.  
<https://doi.org/10.1152/ajplung.00163.2007> PMID:[17631612](https://pubmed.ncbi.nlm.nih.gov/17631612/)
35. Zeng R, Gao S, Xu L, Liu X, Dai F. Prediction of pathogenesis-related secreted proteins from stemphylium lycopersici. *BMC Microbiol*. 2018; 18:191.  
<https://doi.org/10.1186/s12866-018-1329-y> PMID:[30458731](https://pubmed.ncbi.nlm.nih.gov/30458731/)

## Could microtubule inhibitors be the best choice of therapy in gastric cancer with high immune activity: mutant DYNC1H1 as a biomarker

Jin Bai<sup>1,2,3,4</sup>, BoWen Yang<sup>1,2,3,4</sup>, Ruichuan Shi<sup>1,2,3,4</sup>, Xinye Shao<sup>1,2,3,4</sup>, Yujing Yang<sup>1,2,3,4</sup>, Fang Wang<sup>1,2,3,4</sup>, Jiawen Xiao<sup>5</sup>, Xiujuan Qu<sup>1,2,3,4</sup>, Yunpeng Liu<sup>1,2,3,4</sup>, Ye Zhang<sup>6</sup>, Zhi Li<sup>1,2,3,4</sup>

<sup>1</sup>Department of Medical Oncology, The First Hospital of China Medical University, Shenyang 110001, China

<sup>2</sup>Key Laboratory of Anticancer Drugs and Biotherapy of Liaoning Province, The First Hospital of China Medical University, Shenyang 110001, China

<sup>3</sup>Liaoning Province Clinical Research Center for Cancer, Shenyang 110001, China

<sup>4</sup>Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education, Shenyang 110001, China

<sup>5</sup>Department of Medical Oncology, Shenyang Fifth People Hospital, Tiexi District, Shenyang 110001, China

<sup>6</sup>Laboratory I of Cancer Institute, The First Hospital of China Medical University, Shenyang 110001, China

**Correspondence to:** Zhi Li, Ye Zhang; **email:** [zli@cmu.edu.cn](mailto:zli@cmu.edu.cn), [y Zhang21@cmu.edu.cn](mailto:y Zhang21@cmu.edu.cn)

**Keywords:** immune checkpoint blockade, microtubule inhibitors, mutation, DYNC1H1, gastric cancer

**Received:** January 13, 2020

**Accepted:** August 27, 2020

**Published:** November 20, 2020

**Copyright:** © 2020 Bai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

Immune checkpoint blockade (ICB) has achieved unprecedented breakthroughs in various cancers, including gastric cancer (GC) with high immune activity (MSI-H or TMB-H), yet clinical benefits from ICB were moderate. Here we aimed to identify the most appropriate drugs which can improve outcomes in GC. We firstly compared MSI-H and TMB-H GC samples with normal samples in TCGA-STAD cohort, respectively. After that, Connectivity Map database repurposed nine candidate drugs (CMap score < -90). Then, microtubule inhibitors (MTIs) were screened as the significant candidate drugs with their representative gene sets strongly enriched ( $p < 0.05$ ) via GSEA. GDSC database validated higher activities of some MTIs in GC cells with MSI-H and TMB-H ( $p < 0.05$ ). Furthermore, some MTIs activities were positively associated with mutant Dynein Cytoplasmic 1 Heavy Chain 1 (DYNC1H1) ( $p < 0.05$ ) based on NCI-60 cancer cell line panel. DYNC1H1 was high frequently alteration in GC and was positively associated with TMB-H and MSI-H. Mutant DYNC1H1 may be accompanied with down-regulation of MTIs-related genes in GC or change the binding pocket to sensitize MTIs. Overall, this study suggested that some MTIs may be the best candidate drugs to treat GC with high immune activity, especially patients with DYNC1H1 mutated.

### INTRODUCTION

High immune activity is one of the essential characteristics of malignant tumors [1]. Currently, as the advent of immune checkpoint blockade (ICB), GC patients with microsatellite instability-high (MSI-H) exhibited sensitivity to ICB in multiple clinical trials [2–4]. Tumors with high tumor mutational burden (TMB-H) also showed better responses to ICB and more prolonged overall survival than those with low

TMB in gastric cancer [5]. Therefore, MSI-H and TMB-H can be considered as the high immune activity biomarkers in GC to repurpose better drugs that can improve clinical outcomes.

ICB has achieved overwhelming breakthroughs in the treatment of different malignant tumors such as GC [6, 7]. ICB significantly prolonged the overall survival of GC patients [2, 4, 8, 9]. However, the therapeutic benefit was limited to specific subgroups such as MSI-

H and TMB-H and so on. Therefore, ongoing studies are underway to improve the efficacy of ICB through the following two main ways: 1) selecting patients with high immune activity subgroups for ICB treatment [5]; 2) in combination with conventional chemotherapy regimens or other drugs [10, 11]. However, there were still no standard recommendation for the combination drugs. Hence, we considered it valuable to study that the drugs which target the high immune activity subgroups may be a more appropriate choice.

Currently, it is difficult to obtain population data in the short term. High-throughput databases of pharmacogenomics are practical approaches for preliminary drug repurposing. For instance, the Connectivity Map (CMap) database provides drug-induced gene expression profiles from 72 cell lines by 27,927 compounds [12]. This database has been used to identify rapamycin for improving the prognosis of GC [13]. Another study reported that ERCC1 and DPD might be the oxaliplatin-resistant genes, based on the NCI-60 cancer cell line panel [14, 15]. Another study constructed a methodology of determining the optimal combination of chemotherapy drugs for gastric cancer patients based on chemotherapeutic drug responses to DCF from the genomics of drug sensitivity in cancer (GDSC) database [16, 17].

In the current study, we selected TCGA-STAD (Stomach Adenocarcinoma) patients with high immune activity (MSI-H, and TMB-H) as the research objects. We then used online omics tools such as CMap, NCI-60, GDSC databases to identify candidate drugs that could more potential for the treatment of GC with high immune activity as well as the predictive biomarker for the drug sensitivity of candidate drugs and possible mechanisms.

## RESULTS

### Identification of differentially expressed genes (DEGs) in gastric cancer with high immune activity

To identify DEGs in GC with high immune activity, we used the GC (N = 375) and adjacent normal tissues (N = 32) of the TCGA-STAD cohort. 12.5% (47/375) GC samples displayed MSI-H, and 15.7% (118/375) of GC patients were TMB-H. Indeed, DEGs were screened from each group compared with normal tissues. Overall, a total of 539 up-regulated and 1353 down-regulated genes were found in the MSI-H group, 632 up-regulated and 979 down-regulated genes were screened from TMB-H group. Together, there were 1375 common DEGs in the two groups: 436 common up-regulated and 939 common down-regulated genes (common DEGs were shown in Supplementary Table 1).

### Identification of high immune activity targeting candidate drugs for gastric cancer

We firstly used DEGs of MSI-H and TMB-H groups to query in the CMap database respectively to repurpose candidate drugs. CMap results revealed nine drugs both from the two groups were significantly negatively correlated with the DEGs (connectivity score <-90). These nine drugs can be classified into seven types (Figure 1A) according to the annotation information of the CMap database and previous studies [18–23].

We next sought to the significant candidate drugs. We screened the representative functional gene sets of the seven drug types from the Molecular Signatures Database (MSigDB) (Supplementary Table 2). Based on the common DEGs, GSEA was performed with representative functional gene sets for each drug type. GSEA results showed that some microtubule inhibitors-related pathways: "GO\_MICROTUBULE\_BINDING", "GO\_MICROTUBULE\_DEPOLYMERIZATION" were significantly enriched ( $p$ -value < 0.05) in DEGs of high immune activity, while other types of drug showed no statistical significance ( $p$ -value > 0.05) (Figure 1B).

Furthermore, we utilized the genomics of drug sensitivity in cancer (GDSC) database to validate the anticancer sensitivities of MTIs in GC with high immune activity. A total of 13 tubulin related drugs (such as MTIs, AURK inhibitors, and KIF inhibitors) and 26 gastric cancer cells were available in the GDSC database. For the TMB levels and MSI status in GC cell lines, some MTIs showed significant associations with TMB levels (Figure 2A, docetaxel:  $R = 0.44$ ,  $p$ -value = 0.004) and MSI-H (Figure 2D, docetaxel:  $p$ -value < 0.001). However, no significance was found in other tubulin related drugs, including AURK inhibitors (Figure 2B, 2E, alisertib:  $p$ -values > 0.05) and KIF inhibitors (Figure 2C, 2F, ARRY-520:  $p$ -values > 0.05) (complete results were shown in Supplementary Table 2).

Moreover, survival analysis revealed three MTIs related genes (like BUB1B) in TMB group and 12 MTIs related genes (like ABCG2) in MSI-H group had significant prognostic values (Supplementary Figure 1). Together, we referred that some MTIs were more likely to treat GC with high immune activity.

### Prediction of the mutant gene for candidate drugs activity and its association with high immune activity

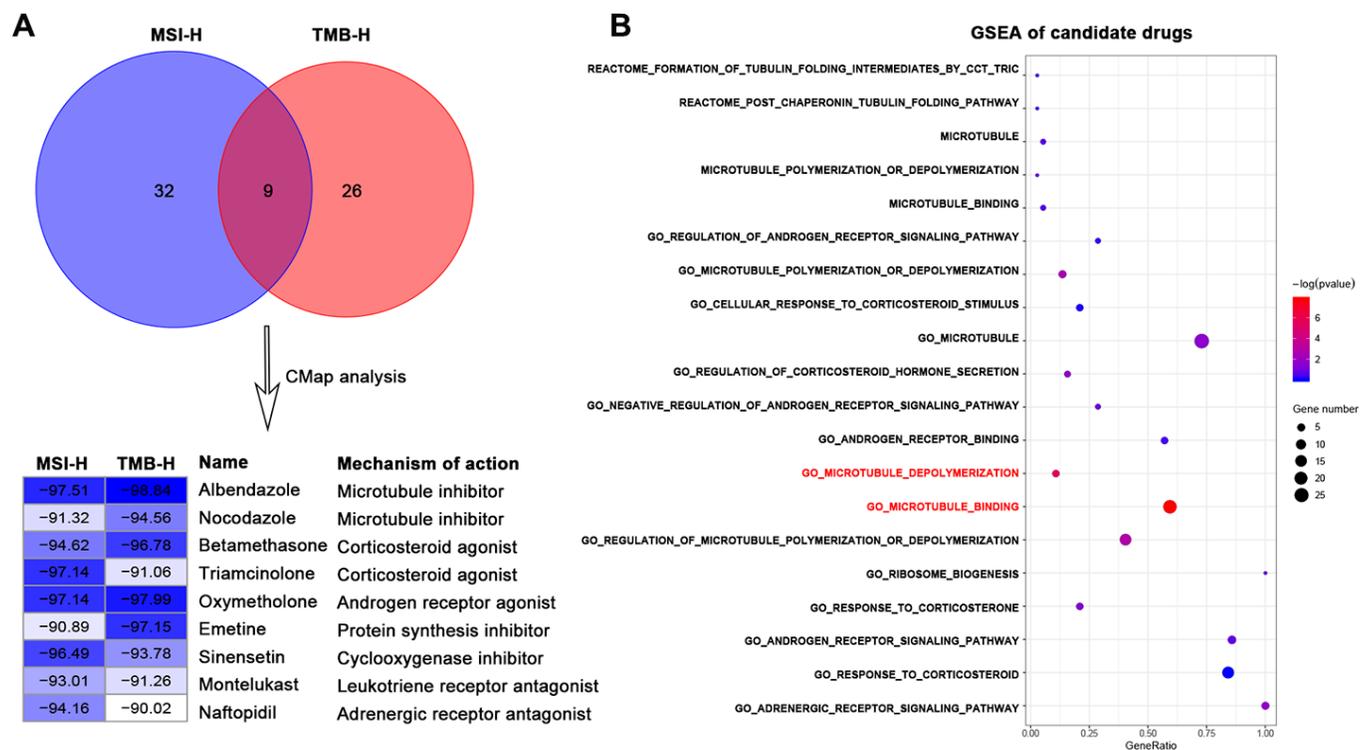
To find the characteristic gene which could suggest the activities of MTIs in GC with high immune activity, we started with a PPI network analysis. Based on the 17 common DEGs directly enriched in the MTIs representative pathways, we constructed a PPI network

with 30 nodes and 289 edges (Figure 3A). After that we carried out a mutational landscape analysis on these 30 node genes and revealed that three genes (DYNC1H1, KIF26B, and CENPF) were mutated at a high frequency (mutation frequency  $\geq 7\%$ ) in the TCGA-STAD cohort (Figure 3B).

To characterize the roles of these three genes in high immune activity, we separately compared the levels of TMB, MSI score, and mRNA expression between these three genes mutated and wild type tumors based on the TCGA-STAD, NCI-60 cell lines and immunotherapy datasets. As shown in Figure 4 and Supplementary Figure 2, mutations in DYNC1H1 were significantly correlated with TMB-H and MSI-H in these above datasets. Mutations in CENPF were considerably correlated with TMB-H and MSI-H in the TCGA-STAD cohort, but not in the NCI-60 cell lines or immunotherapy dataset. Mutations in KIF26B were significantly correlated with TMB-H and MSI-H in the TCGA-STAD cohort, but not immunotherapy datasets (KIF26B was not detected in NCI-60 cell lines dataset).

Besides, there was no significant difference of mRNA expression levels among three genes mutated and wild type tumors in TCGA-STAD or NCI-60 cell lines dataset (Figure 4C, 4F). Together, these results indicated that mutant DYNC1H1 was strongly associated with enhanced tumor immune activity.

Furthermore, we investigated the importance of mutant DYNC1H1 in TCGA pan-cancer, showing that the mutation frequency of DYNC1H1 was 0.4%-19.7%, average mutation frequency was 5.17% in 15 cancers. DYNC1H1 mutated tumors were associated with higher TMB levels than its wild type tumors (Supplementary Figure 3A). Accordingly, the mutation frequency of DYNC1H1 was positively associated with TMB levels in 9 cancers (Supplementary Figure 3B). Moreover, DYNC1H1 was highly mutated in Uterine Corpus Endometrial Carcinoma (UCEC) and GC. There were more than 26 missense mutation spots of DYNC1H1 in GC (Supplementary Figure 3C, 3D) These results showed that the mutant DYNC1H1 was consistent with high immune activity of TMB levels in multiple cancers.



**Figure 1. Identification of candidate drugs that may treat gastric cancer with high immune activity in TCGA-STAD cohort. (A)** Venn diagram (top) showing the number of common candidate drugs (CMap score  $< -90$ ) from MSI-H (blue) and TMB-H (red) groups via the connectivity map database. Connectivity score table (down) displaying nine common candidate drugs, each row responsible for a drug and columns corresponding to MSI-H and TMB-H groups. The score labels representing the connectivity score of each drug in each group, and right sides of the table indicating the name and mechanism of each drug. These nine drugs can be classified into seven types. **(B)** GSEA results for candidate drugs based on the functional gene sets of these seven drug types from the Molecular Signatures Database by the function of enricher of clusterprofiler package. Pathways in red font were significantly enriched ( $p$  value  $< 0.05$ ).

## Association of mutant DYNC1H1 with enhanced microtubule inhibitors activities

To explore the interaction between mutant DYNC1H1 and MTIs, we firstly selected 41 MTIs included in NCI-60 cell lines dataset. 13 of these MTIs have significantly increased activities in the DYNC1H1 mutated cell lines (Figure 5,  $p$ -value  $< 0.05$ ). Furthermore, a clear trend of higher MTIs activities was observed in DYNC1H1-mutated GC cells (Supplementary Figure 4). These results suggested that mutations in DYNC1H1 may be an essential biomarker for the antitumor effects of MTIs.

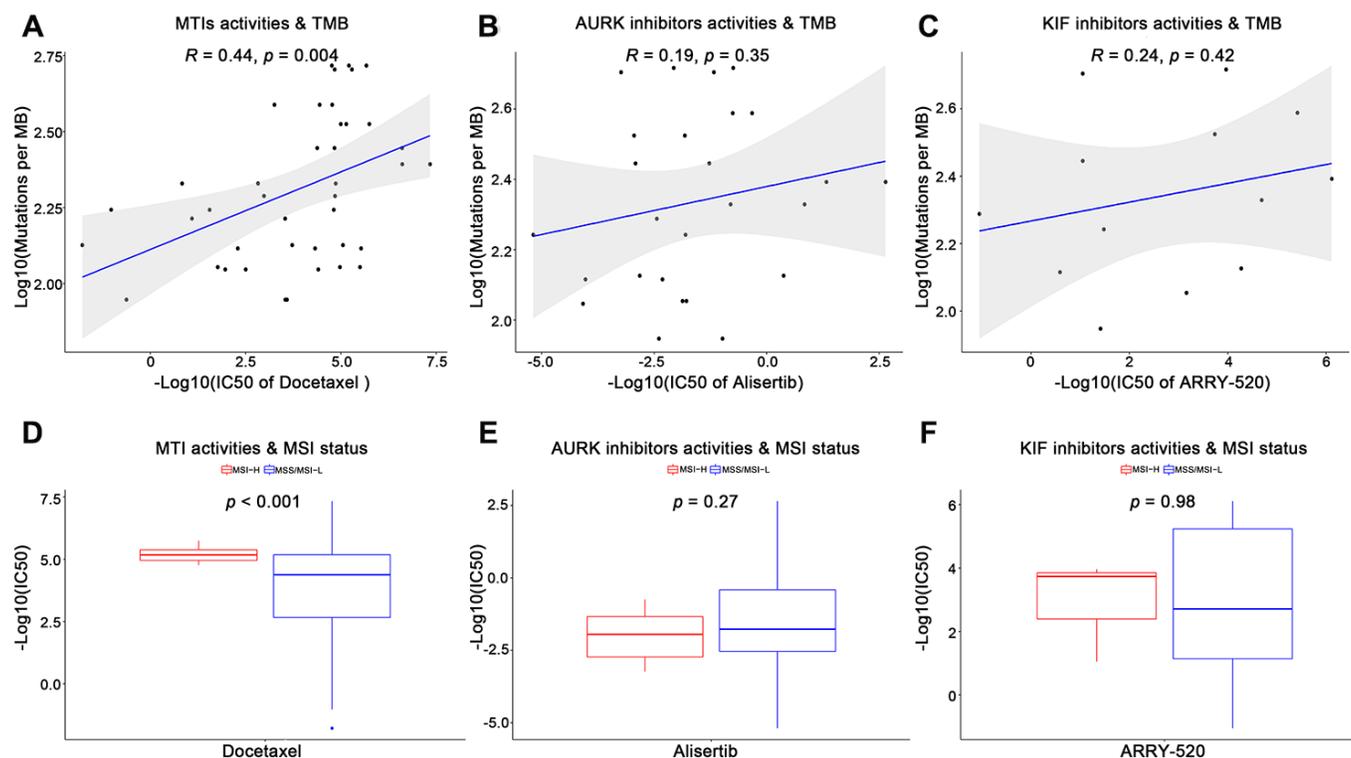
## Effects of mutant DYNC1H1 on enhanced microtubule inhibitors activities

To further explore how mutant DYNC1H1 increase drug sensitivities of MTIs, we used a differential analysis method to identify mutant DYNC1H1 related genes by comparing DYNC1H1 mutated tumor with wild type tumors (37/338) in TCGA-STAD. A total of 199 genes were significantly correlated with mutant DYNC1H1 (23 up-regulated and 176 down-regulated

genes) (Figure 6A). Based on NCI-60 cell lines dataset, the expressions of four genes (IGF2, MAL, KRT13, CALCA) were negatively associated with MTIs sensitivities (Figure 6B–6E,  $p$ -value  $< 0.05$ ). Accordingly, IGF2 expression levels were lower in paclitaxel sensitive breast cancer cells than in paclitaxel resistant breast cancer cells in GSE90564 (Figure 6F,  $p$ -value  $< 0.05$ ). Together, these results suggested that the process of mutant DYNC1H1 sensitizing MTIs may be followed by decreased expressions of MTIs related genes such as IGF2, KRT13, CALCA and MAL.

## DISCUSSION

The advent of cancer immunotherapy, such as the approval of anti-PD1 monoclonal antibodies, has altered the treatment paradigm of many malignancies including GC [6, 7]. However, the clinical benefits were quite low and limited to high immune activity subtypes patients [2–5]. In this study, MTIs were repurposed as the more appropriate drugs to treat GC with high immune activity (TMB-H and MSI-H) based on CMap database and GSEA (Figure 7). Then, MTIs like docetaxel showed increased activities when the mutations in DYNC1H1.



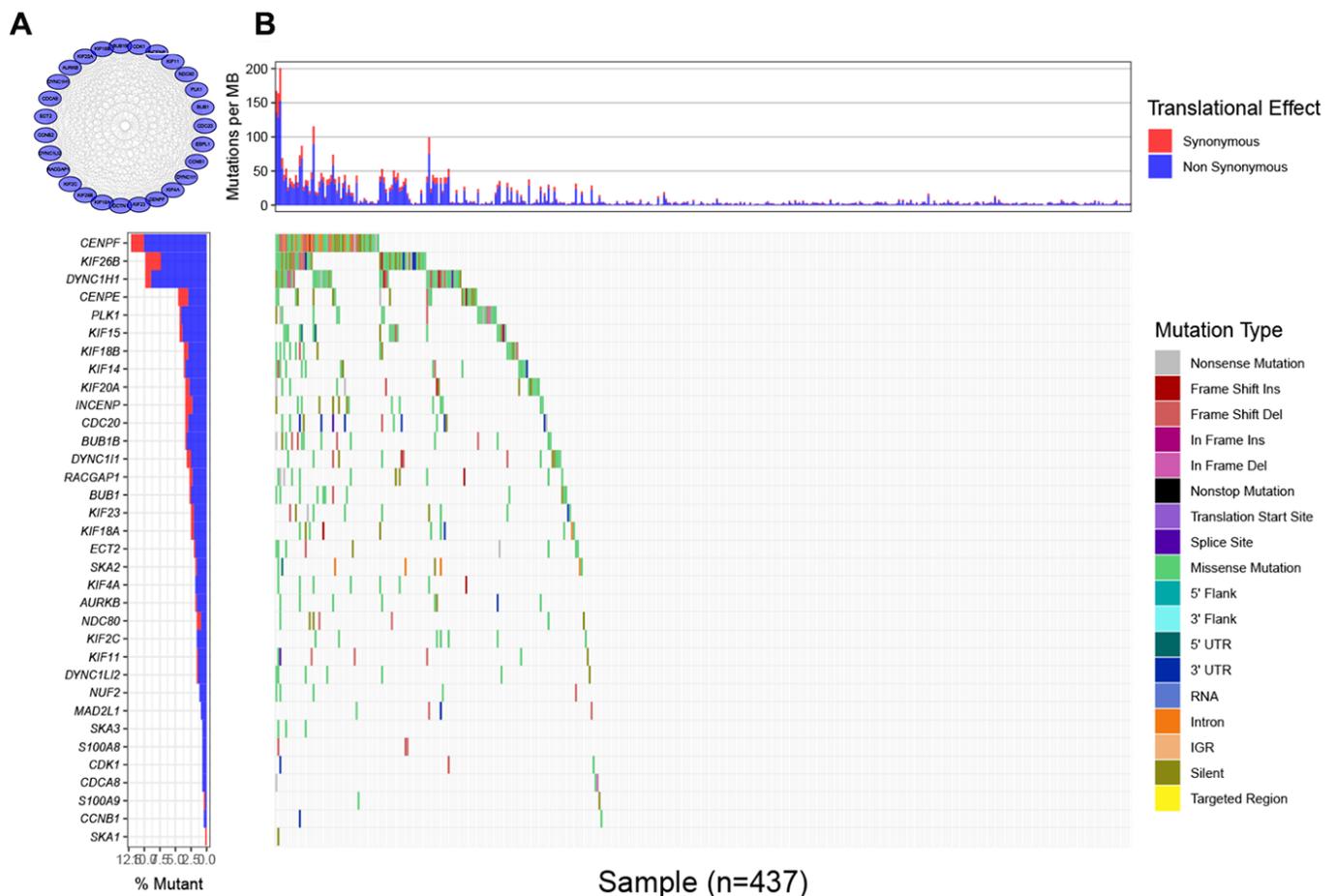
**Figure 2. Validation of microtubule inhibitors (MTIs) in gastric cancer cell lines with high immune activity based on GDSC database.** (A–C) Plot depicting the correlations of IC<sub>50</sub> values of some tubulin related drugs ((A) MTIs, (B) AURK inhibitors, (C) KIF inhibitors) with TMB levels in GC cell lines based on GDSC database using Spearman's correlation. (D–F) Histograms showing the different levels of IC<sub>50</sub> values of some tubulin related drugs ((D) MTIs, (E) AURK inhibitors, (F) KIF inhibitors) between MSI-H and MSS/MSI-L GC cell lines based on GDSC database using Student's *t* test.  $P$ -value  $< 0.05$  was considered significant.

Furthermore, mutant DYNC1H1 can act as a biomarker for MTIs activities possibly because it may be followed by lower levels of MTIs-related genes, or some mutations in DYNC1H1 may change the binding confirmation of MTIs and microtubules. In summary, our research indicated that GC patients with high immune activity may benefit from some MTIs that can be sensitized by mutant DYNC1H1.

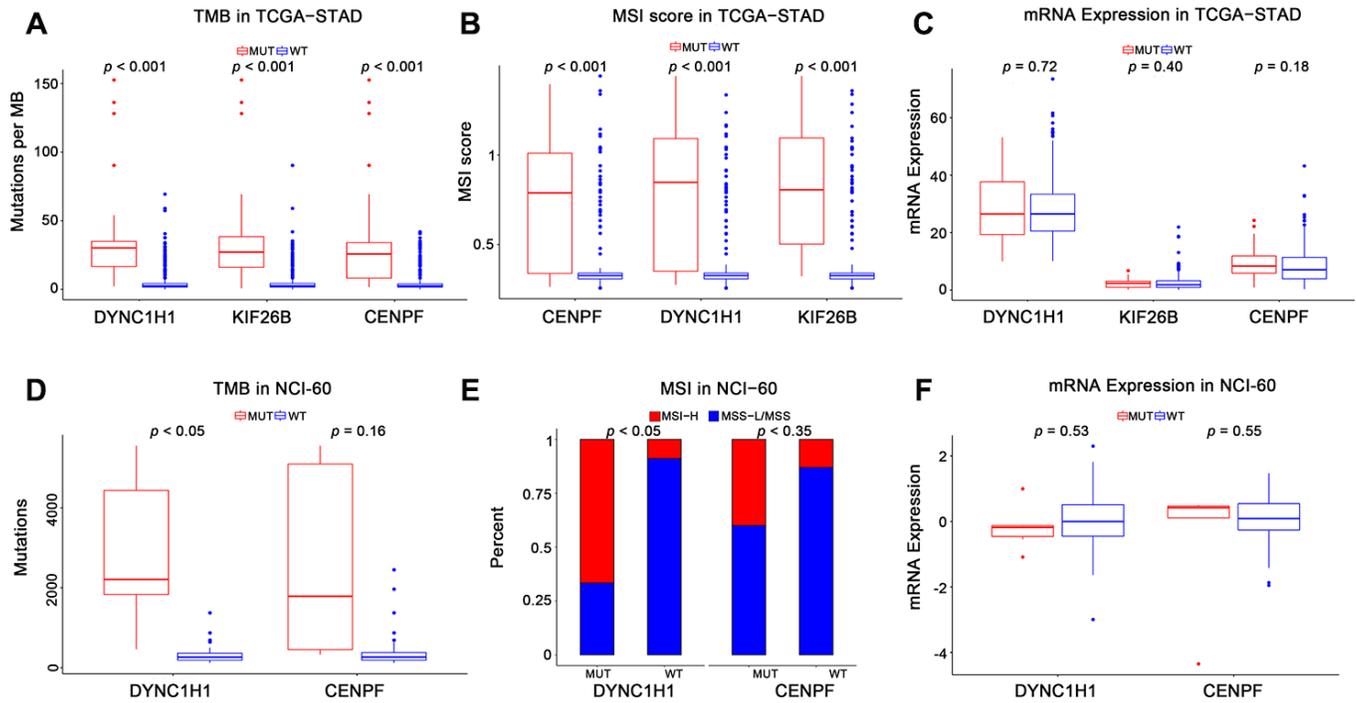
This study presented that some MTIs may treat high immune activity subgroup GC patients by several lines of evidence. Firstly, nocodazole and alendazole were recognized as MTIs from CMap database (Figure 1A). Nocodazole can affect the dynamics of microtubules by specifically binding to the cell motility apparatus [18]. Apart from the indicator of an anthelmintic benzimidazole carbamate [19, 20], alendazole also can inhibit microtubule polymerization for the treatment of gastric cancer and other cancers [21–23]. Then, some

MTIs representative pathways can be significantly enriched in high immune activity of GC: "GO\_MICROTUBULE\_BINDING", "GO\_MICROTUBULE\_DEPOLYMERIZATION", whilst other drug types repurposed by CMap showed no significance (Figure 1B). Moreover, the activities of some MTIs (such as docetaxel) were strongly higher in MSI-H and TMB-H gastric cancer cells in the GDSC database. However, activities of other tubulin related drugs (such as AURK inhibitors, KIF inhibitors) revealed no significance in different MSI status and TMB levels (Figure 2).

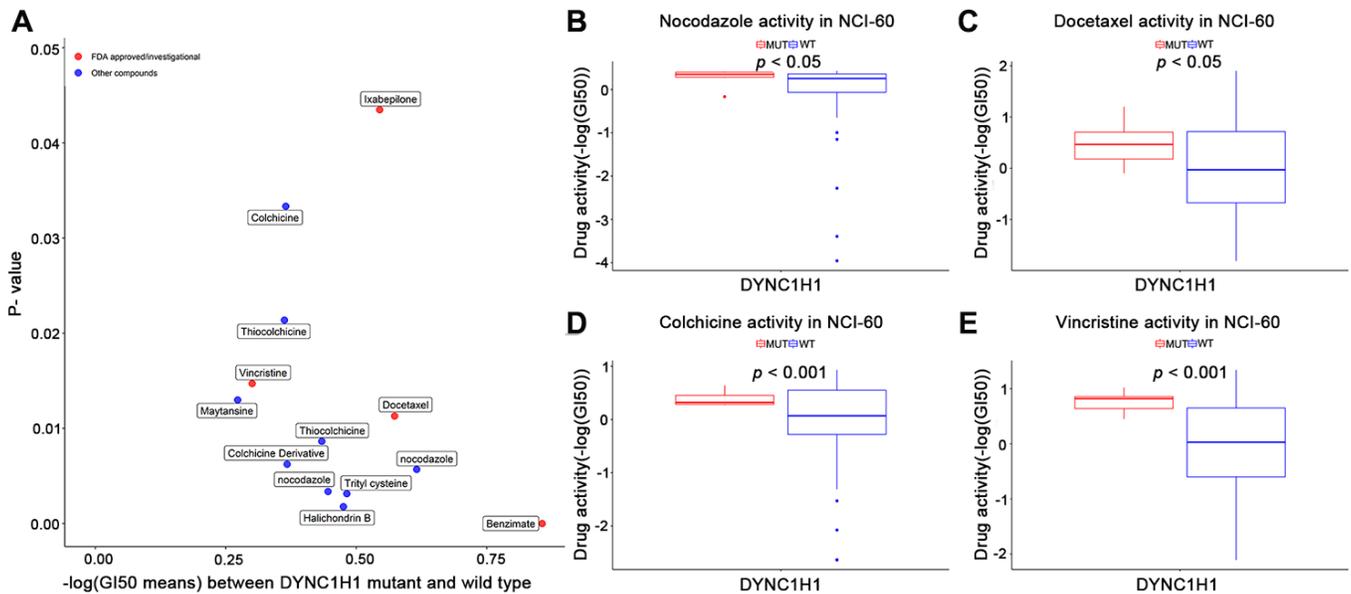
MTIs are widely known for binding to the microtubules [24, 25]. Among them, taxanes (paclitaxel and docetaxel) have shown antitumor activity in the treatment of GC according to the National Comprehensive Cancer Network (NCCN) guidelines. In addition to antiproliferative effects, growing studies have reported that some MTIs can be



**Figure 3. Mutational landscape of genes involved in pathways of microtubule inhibitors (MTIs) in TCGA-STAD cohort. (A)** Protein-protein interaction (PPI) network of genes that significantly enriched in microtubule inhibitors (MTIs) representative gene sets. **(B)** Mutational landscape of node genes from PPI network showing that CENPF, KIF26B, and DYNC1H1 were highly mutated in TCGA-STAD cohort (alteration frequency  $\geq 7\%$ ) by the GenVisR package. Top for somatic mutation rate of each sample, bottom left for the total mutation frequency of each gene, and bottom right for specific mutation type of each gene in each sample.



**Figure 4. Association of mutant genes with high immune activity in TCGA-STAD and NCI-60 cell lines datasets.** Levels of TMB, MSI score, and mRNA expression in TCGA-STAD (A–C) and NCI-60 cell lines (D–F), stratified by genes (DYNC1H1, CENPF, and KIF26B) mutation status. MUT: mutated, WT: wild type. KIF26B was not detected in NCI-60 cell lines dataset. All  $p$ -values were obtained by Student’s t-test, in addition to  $\chi^2$  test for (E).  $P$ -value  $< 0.05$  was considered significant.



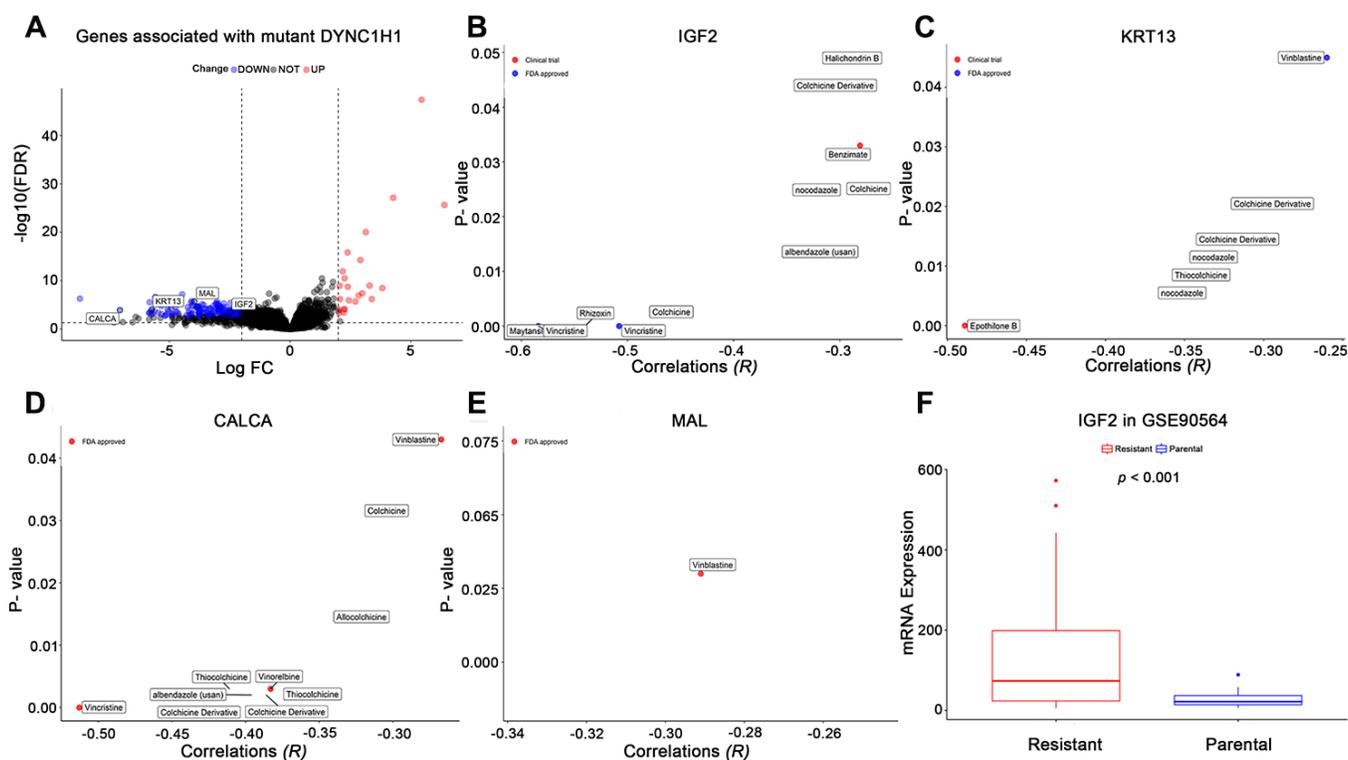
**Figure 5. Association of mutant DYNC1H1 with enhanced microtubule inhibitors (MTIs) activities in NCI-60 cell lines dataset.** (A) Volcano plot for the different activities of MTIs between DYNC1H1 mutated and wild type NCI-60 cell lines dataset. The x-axis represented the different levels of mean  $-\log\text{GI50}$ , and the y-axis showed  $p$ -values obtained by Student’s t-test. (B–E) Histograms depicting different MTIs activities of nocodazole (B), docetaxel (C), colchicine (D), vincristine (E), stratified by DYNC1H1 mutation status in NCI-60 cell lines dataset ( $p$ -value  $< 0.05$  by Student’s t-test). MUT: mutated, WT: wild type.  $P$ -value  $< 0.05$  was considered significant. DYNC1H1 was mutated in six NCI-60 cell lines: HCC\_2998, HCT\_116, HCT\_15, KM12, MOLT\_4, and UACC\_62.

immunostimulatory properties, including their abilities to reprogram the immune-suppressive M2 profile of tumor-associated macrophages to immune-stimulating M1 profile [26], stimulate maturation and activities of dendritic cells [27, 28], and decrease the accumulation and immunosuppressive activities of tumor-infiltrating MDSCs [29]. Clinically, single-agent paclitaxel even showed a better median progression-free survival than the anti-PD1 pembrolizumab (4.1 months vs 1.5 months) in the KEYNOTE-061 clinical trial (PDL1 CPS>1) [4]. These evidences supported that some MTIs may achieve good outcomes in gastric cancer patients with high immune activity.

Our results clearly demonstrated that drug activities of some MTIs were strongly associated with mutant DYNC1H1 (Figure 6). DYNC1H1 is the heavy chain of cytoplasmic dynein which acts as a motor protein using ATP to travel along the microtubule (MTs) toward minus end [30, 31]. This complex participates in multiple cell processes, such as spindle formation in

mitosis and transportation of various cellular cargoes. Mutant DYNC1H1 was reported to impede the ATP hydrolysis cycle which help bind to MTs in the neurological diseases [32]. Most mutations of DYNC1H1 were tied up with the occurrence and development of pancreatic cancer [33–35], suggesting that mutations in DYNC1H1 may play a vital role in the complex biological process of malignant tumors. However, the effect of its mutation on gastric cancer remains unknown.

In this study, mutant DYNC1H1 was closely associated with MTIs-related genes (IGF2, KRT13, MAL, and CALCA) (Figure 5). IGF2 and anti-apoptotic gene KRT1 were significantly overexpressed in MTIs-resistant cancer cells, and exhaustion of IGF2 can restore paclitaxel sensitivity [36, 37]. Also, CALCA and MAL were reported to be highly expressed in the allergic reactions of MTIs [38, 39]. On the other hand, DYNC1H1 directly binds to MTs via the structure called MTBD [40, 30]. Besides, MTIs (like paclitaxel



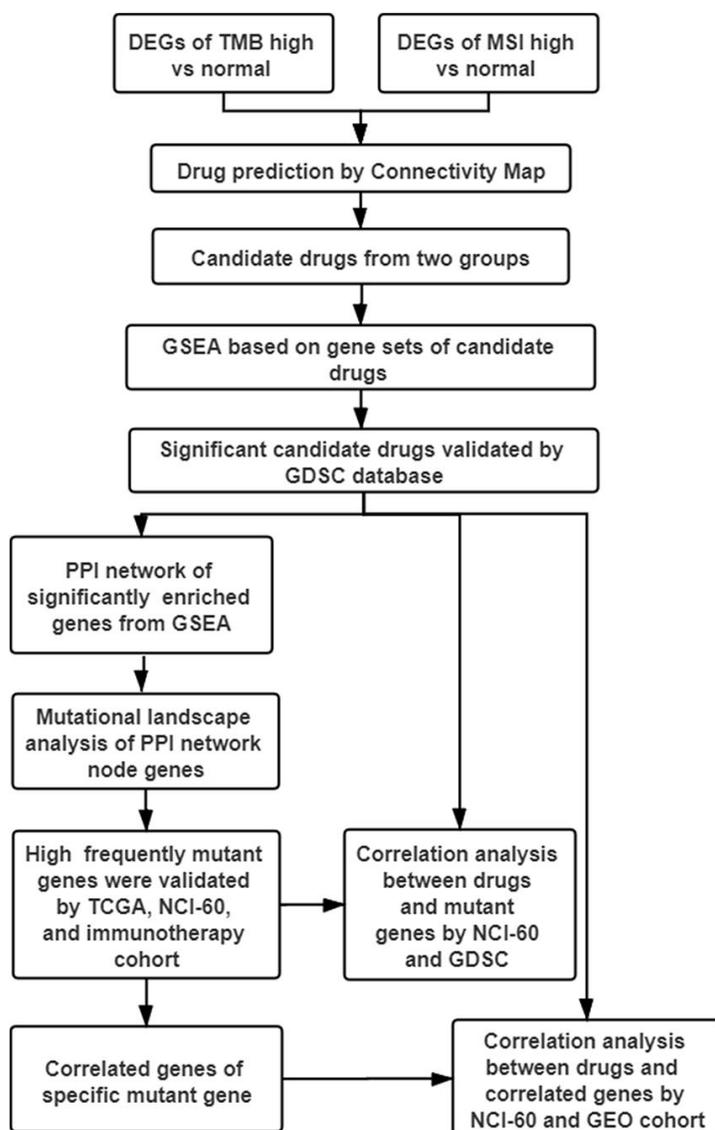
**Figure 6. Effects of mutant DYNC1H1 on enhanced microtubule inhibitors (MTIs) activities.** (A) Volcano plot for the DEGs between DYNC1H1 mutated and wild type gastric cancer patients in TCGA-STAD cohort. The x-axis represented log2 (fold change): mutant DYNC1H1 compared with wild type patients, and the y-axis represented significant difference as  $-\log_{10}(\text{FDR})$ . The criteria of  $\text{FDR} < 0.05$  and  $|\log_2\text{FC}| \geq 2$  were considered significant by the function TCGAbiolinks\_DEA of TCGAbiolinks. (B–E) Correlations between MTIs activities and mRNA expressions of IGF2, KRT13, CALCA, and MAL in NCI-60 cell lines dataset, respectively. *P*-value estimated using Pearson’s correlation. The x-axis represented the correlation coefficient, and the y-axis showed the significance ( $-\log_{10}$  *p*-value). (F) Histograms depicting different mRNA expressions of IGF2 between resistant and parental cell lines in GSE90564 dataset. *P*-value estimated using Student’s *t*-test. *P*-value < 0.05 was considered significant.

and docetaxel) exert antitumor effects through binding to  $\beta$ -tubulin [41, 42]. Hence we speculated that mutant DYNC1H1 may narrow the binding pocket of MTBD and  $\beta$ -tubulin and then increase binding pocket of MTIs and  $\beta$ -tubulin, thereby enhancing the anti-microtubule effect of MTIs. We identified 26 missense mutation sites of DYNC1H1 in the TCGA-STAD mutation profile (Supplementary Figure 3). However, there were no studies clearly illustrated the correlation of these mutation sites of DYNC1H1 with gastric cancer.

Our study has some limitations such as few normal samples included in this study but no other suitable datasets available for validation. We cannot also rule out the possibility that other drugs or combination

therapy may show more benefits for GC patients with high immune activity. Therefore, we plan to use our own clinical samples for verification analysis in the future. Further investigation of how mutant DYNC1H1 sensitizes MTIs will be necessary in future studies.

To conclude, our study identified some MTIs such as docetaxel that could potentially be the best drugs for GC with high immune activity (TMB-H and MSI-H). Mutant DYNC1H1 significantly positively correlated with TMB-H and MSI-H in GC or various cancers. We found that mutant DYNC1H1 can sensitize MTIs, possibly because it was accompanied with down-regulation of some MTIs resistant or side effect genes. It may also change the binding pocket of MTIs and microtubules.



**Figure 7. Work flow of the study.** DEGs: differentially expressed genes; TMB-H: high tumor mutational burden; MSI-H: high microsatellite instability; GSEA: gene sets enrichment analysis; GDSC: genomics of drug sensitivity in cancer; PPI: protein-protein interaction; TCGA: The Cancer Genome Atlas; NCI-60: National Cancer Institute 60.

## MATERIALS AND METHODS

### Gastric cancer dataset and differentially expressed genes (DEGs) associated with high immune activity

GC dataset was acquired from GDC Data Portal TCGA-STAD (n = 407). "Level 3" RNA sequencing data (raw count) and clinical information were downloaded by using R package TCGAbiolinks [43]. Then, patients were classified into two groups: TMB-H group (median as cut-point of tumor mutational rate), and MSI-H group. Finally, differential analysis of the two groups were both used by the function TCGAbiolinks\_DEA. DEGs were determined with the criteria of  $|\log_2(\text{FC})| > 2$  and  $\text{FDR} < 0.05$ .

### Drug prediction

The CMap database (<https://clue.io/cmap>) was used to identify drugs that can effectively treat high immune activity subgroups GC. Input data require a range of 10 ~ 150 genes that are up-regulated and/or down-regulated. Because the number of DEGs in the two groups of TMB-H and MSI-H were more than 150, we selected top150 up-regulated DEGs following the FDR in ascending order. A so-called Connectivity Score (-100 ~100) is an indicator for evaluating the correlation between a drug and input genes. The score less than -90 can be considered that the drug is significantly negatively related to the input genes, that is, the drug can dramatically reverse the role of these genes to treat specific diseases. Drugs correlated with the input genes of TMB-H and MSI-H group were selected as candidate drugs for subsequent analysis.

### Gene sets enrichment analysis (GSEA)

GSEA was used to further screen the significant candidate drugs. Candidate drugs-related gene sets were selected from MSigDB [44], and common DEGs were used as input data. GSEA was performed separately with gene sets of each drug type using the function enricher of clusterprofiler [45] package. The drug corresponding to the significantly enriched gene sets ( $p\text{-value} \leq 0.05$ ) was regarded as significant candidate drug.

### Gastric cancer data of drug sensitivity and high immune activity

To validate the role of significant candidate drugs for treating gastric cancer with high immune activity, we extracted the half maximal inhibitory concentration (IC50) values of them in GC cell lines from GDSC (<https://www.cancerxgene.org/downloads/anova>) database. MSI status of each GC cell line was also provided by GDSC. TMB levels of each cell line were

retrieved from the CCLE (<https://portals.broadinstitute.org/ccle/data>) database.

### Protein-protein interaction (PPI) network

Search Tool for the Retrieval of Interacting Genes (STRING; [string-db.org](http://string-db.org)) was used to build a PPI network for genes directly enriched in gene sets of significant candidate drugs. The parameter of interaction was set as the interaction score  $> 0.9$  and no more than 20 interactors. Cytoscape software 3.7.0 was used to visualize the PPI network.

### Mutational landscape analysis

Somatic mutational landscape analysis was used to explore characteristic genes of significant candidate drugs. Mutational profile of TCGA-STAD was obtained from TCGA data portal by the GDCquery\_Maf function of the TCGAbiolinks package in R. Using "GenVisR" package to process and visualize the mutational burden of PPI network node genes. The specific mutant genes were identified with mutation frequency  $\geq 7\%$  [46].

### Data of gene mutation and high immune activity

To compare different levels of high immune activity (TMB, MSI) and mRNA expression between specific mutant genes mutated and wild type tumors. TCGA-STAD, NCI-60, immunotherapy cohort (Allen cohort), and TCGA pan-cancer datasets were selected.

For TCGA-STAD dataset, gene mutational profile and mRNA expression data, TMB levels (tumor mutational rate), MSI score [47] data were used. For NCI-60 cancer cell lines dataset, binary gene mutation and mRNA expression data were available from R package rcellminer [14], TMB levels of each cell line were downloaded from cbiportal (<https://www.cbiportal.org>) database, and microsatellite status data were downloaded from cosmic (<https://cancer.sanger.ac.uk/cosmic/>) database. For Allen cohort, consisting of 110 advanced-stage melanoma patients treated with anti-CTLA-4 therapy, provided complete somatic mutation profile [48]. For TCGA pan-cancer, mutational profiles were obtained from the TCGA data portal by the GDCquery\_Maf function of the TCGAbiolinks package in R. TMB levels were calculated based on non-synonymous data. Mutation frequencies of specific mutant genes in each cancer type were estimated respectively.

### Data of drug sensitivity and gene mutation

To verify the relationship between specific mutant genes and significant candidate drugs, we used

genomics data and drug sensitivity of NCI-60 cell lines dataset and GDSC database. For NCI-60 cell lines, the concentration of drug was presented as cause 50% growth inhibition (GI50) value and was available from R package rcellminer. Gene mutation information was described above Gastric cancer data of drug sensitivity and high immune activity. For GDSC database, gene mutation information was downloaded from the GDSC website (<https://www.cancerxgene.org/downloads>). IC50 values of candidate drugs were described above. Data of gene mutation and high immune activity.

### The effect of specific mutant genes

To investigate the impact of specific mutant genes on significant candidate drugs. Firstly, we used differential analysis method to find genes significantly associated with the specific mutant genes. According to significant mutant genes status, DEGs were identified by the mutated group compared to the wild type group in the TCGA-STAD (the criteria of  $|\log_2(\text{FC})| > 2$  and  $\text{FDR} < 0.05$ ). Secondly, to analyze the relationships between these DEGs and significant candidate drugs activities, the gene expression data and drug activity data of NCI-60 cell lines and GSE90564 (paclitaxel sensitivity and resistance dataset,  $n = 38$ ) were used.

### Statistical analysis

Differential analysis was mainly conducted by the function TCGAbiolinks\_DEA of TCGAbiolinks package. Survival curves were performed by the Kaplan-Meier method and compared by the log-rank test via the R package survival. Group comparisons were performed by Student's *t* test for continuous variables, while  $\chi^2$  test for categorical variables. The correlation of IC50 values of candidate drugs with TMB levels in GC cell lines of GDSC database, GI50 values of candidate drugs with gene expression or gene mutation in NCI-60 cancer cell lines dataset were identified by Spearman's correlation analysis. All statistical tests were two-sided, and *p*-value  $< 0.05$  was considered significant. Statistical analysis was performed by R software version 3.6.1(v. 3.5.2 (<http://www.r-project.org>)).

### Abbreviations

ICB: immune checkpoint blockade; GC: gastric cancer; STAD: stomach adenocarcinoma; TCGA: The Cancer Genome Atlas; MSI-H: high microsatellite instability; TMB-H: high tumor mutational burden; CMap: Connectivity Map; MTIs: microtubule inhibitors; DYNC1H1: Dynein Cytoplasmic 1 Heavy Chain 1; DEGs: differentially expressed genes; GSEA: gene sets enrichment analysis; PPI: protein-protein interaction; NCI-60: National Cancer Institute 60; GDSC: genomics

of drug sensitivity in cancer; IC50: inhibitory concentration; GI50: growth inhibition; UCEC: Uterine Corpus Endometrial Carcinoma.

### AUTHOR CONTRIBUTIONS

Zhi Li conceived the study. Jin Bai performed all bioinformatic analyses and edited the manuscript. BoWen Yang, Ruichuan Shi, Xinye Shao provided assistance in mutational landscape analysis, data collection, and statistical analysis, respectively. Yujing Yang and Fang Wang revised the manuscript. Xiujuan Qu, Yunpeng Liu, Jiawen Xiao, and Ye Zhang discussed the results and provided guidance to the study. All authors read and approved the final manuscript.

### ACKNOWLEDGMENTS

The authors would like to acknowledge the Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education (China Medical University, Shenyang, China) for providing the space and equipment for conducting the experiments.

### CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

### FUNDING

The National Key Research and Development Program of China (NO.2017YFC1308900); National Natural Science Foundation of China (NO.81972331, No. 81302023); The General Projects of Liaoning Province Colleges and Universities (LFWK201706).

### REFERENCES

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646–74. <https://doi.org/10.1016/j.cell.2011.02.013> PMID:[21376230](https://pubmed.ncbi.nlm.nih.gov/21376230/)
2. Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, Sun W, Jalal SI, Shah MA, Metges JP, Garrido M, Golan T, Mandala M, et al. Safety and efficacy of pembrolizumab monotherapy in patients with previously treated advanced gastric and gastroesophageal junction cancer: phase 2 clinical KEYNOTE-059 trial. *JAMA Oncol*. 2018; 4:e180013. <https://doi.org/10.1001/jamaoncol.2018.0013> PMID:[29543932](https://pubmed.ncbi.nlm.nih.gov/29543932/)
3. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M, Lee S, Park SH, Park JO, et al. Comprehensive molecular

- characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med*. 2018; 24:1449–58.  
<https://doi.org/10.1038/s41591-018-0101-z>  
PMID:30013197
4. Shitara K, Özgüroğlu M, Bang YJ, Di Bartolomeo M, Mandalà M, Ryu MH, Fornaro L, Olesiński T, Caglevic C, Chung HC, Muro K, Goekkurt E, Mansoor W, et al, and KEYNOTE-061 investigators. Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2018; 392:123–33.  
[https://doi.org/10.1016/S0140-6736\(18\)31257-1](https://doi.org/10.1016/S0140-6736(18)31257-1)  
PMID:29880231
  5. Wang F, Wei XL, Wang FH, Xu N, Shen L, Dai GH, Yuan XL, Chen Y, Yang SJ, Shi JH, Hu XC, Lin XY, Zhang QY, et al. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase Ib/II clinical trial NCT02915432. *Ann Oncol*. 2019; 30:1479–86.  
<https://doi.org/10.1093/annonc/mdz197>  
PMID:31236579
  6. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB, Leming PD, Spigel DR, Antonia SJ, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*. 2012; 366:2443–54.  
<https://doi.org/10.1056/NEJMoa1200690>  
PMID:22658127
  7. Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, Drake CG, Camacho LH, Kauh J, Odunsi K, Pitot HC, Hamid O, Bhatia S, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*. 2012; 366:2455–65.  
<https://doi.org/10.1056/NEJMoa1200694>  
PMID:22658128
  8. Kang YK, Boku N, Satoh T, Ryu MH, Chao Y, Kato K, Chung HC, Chen JS, Muro K, Kang WK, Yeh KH, Yoshikawa T, Oh SC, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017; 390:2461–2471.  
[https://doi.org/10.1016/S0140-6736\(17\)31827-5](https://doi.org/10.1016/S0140-6736(17)31827-5)  
PMID:28993052
  9. Muro K, Chung HC, Shankaran V, Geva R, Catenacci D, Gupta S, Eder JP, Golan T, Le DT, Burtneess B, McRee AJ, Lin CC, Pathiraja K, et al. Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncol*. 2016; 17:717–26.  
[https://doi.org/10.1016/S1470-2045\(16\)00175-3](https://doi.org/10.1016/S1470-2045(16)00175-3)  
PMID:27157491
  10. Salati M, Orsi G, Smyth E, Aprile G, Beretta G, De Vita F, Di Bartolomeo M, Fanotto V, Lonardi S, Morano F, Pietrantonio F, Pinto C, Rimassa L, et al. Gastric cancer: translating novel concepts into clinical practice. *Cancer Treat Rev*. 2019; 79:101889.  
<https://doi.org/10.1016/j.ctrv.2019.101889>  
PMID:31445415
  11. Boku N, Ryu MH, Kato K, Chung HC, Minashi K, Lee KW, Cho H, Kang WK, Komatsu Y, Tsuda M, Yamaguchi K, Hara H, Fumita S, et al. Safety and efficacy of nivolumab in combination with S-1/capecitabine plus oxaliplatin in patients with previously untreated, unresectable, advanced, or recurrent gastric/gastroesophageal junction cancer: interim results of a randomized, phase II trial (ATTRACTION-4). *Ann Oncol*. 2019; 30:250–58.  
<https://doi.org/10.1093/annonc/mdy540>  
PMID:30566590
  12. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, Lahr DL, Hirschman JE, Liu Z, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017; 171:1437–52.e17.  
<https://doi.org/10.1016/j.cell.2017.10.049>  
PMID:29195078
  13. Gong C, Hu Y, Zhou M, Yao M, Ning Z, Wang Z, Ren J. Identification of specific modules and hub genes associated with the progression of gastric cancer. *Carcinogenesis*. 2019; 40:1269–77.  
<https://doi.org/10.1093/carcin/bgz040>  
PMID:30805585
  14. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, Doroshow J, Pommier Y. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res*. 2012; 72:3499–511.  
<https://doi.org/10.1158/0008-5472.CAN-12-1370>  
PMID:22802077
  15. Baba H, Watanabe M, Okabe H, Miyamoto Y, Sakamoto Y, Baba Y, Iwatsuki M, Chikamoto A, Beppu T. Upregulation of ERCC1 and DPD expressions after oxaliplatin-based first-line chemotherapy for metastatic colorectal cancer. *Br J Cancer*. 2012; 107:1950–55.  
<https://doi.org/10.1038/bjc.2012.502>  
PMID:23169295
  16. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR,

- Ramaswamy S, Futreal PA, Haber DA, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 2013; 41:D955–61.  
<https://doi.org/10.1093/nar/gks111>  
PMID:23180760
17. Lee JH, Park YR, Jung M, Lim SG. Gene regulatory network analysis with drug sensitivity reveals synergistic effects of combinatory chemotherapy in gastric cancer. *Sci Rep.* 2020; 10:3932.  
<https://doi.org/10.1038/s41598-020-61016-z>  
PMID:32127608
18. Florian S, Mitchison TJ. Anti-microtubule drugs. *Methods Mol Biol.* 2016; 1413:403–21.  
[https://doi.org/10.1007/978-1-4939-3542-0\\_25](https://doi.org/10.1007/978-1-4939-3542-0_25)  
PMID:27193863
19. Köhler P. The biochemical basis of anthelmintic action and resistance. *Int J Parasitol.* 2001; 31:336–45.  
[https://doi.org/10.1016/s0020-7519\(01\)00131-x](https://doi.org/10.1016/s0020-7519(01)00131-x)  
PMID:11400692
20. Gil-Grande LA, Rodriguez-Caabeiro F, Prieto JG, Sánchez-Ruano JJ, Brasa C, Aguilar L, García-Hoz F, Casado N, Bárcena R, Alvarez AI, Dal-Ré R. Randomised controlled trial of efficacy of albendazole in intra-abdominal hydatid disease. *Lancet.* 1993; 342:1269–72.  
[https://doi.org/10.1016/0140-6736\(93\)92361-v](https://doi.org/10.1016/0140-6736(93)92361-v)  
PMID:7901585
21. Zhang X, Zhao J, Gao X, Pei D, Gao C. Anthelmintic drug albendazole arrests human gastric cancer cells at the mitotic phase and induces apoptosis. *Exp Ther Med.* 2017; 13:595–603.  
<https://doi.org/10.3892/etm.2016.3992>  
PMID:28352336
22. Pourgholami MH, Akhter J, Wang L, Lu Y, Morris DL. Antitumor activity of albendazole against the human colorectal cancer cell line HT-29: in vitro and in a xenograft model of peritoneal carcinomatosis. *Cancer Chemother Pharmacol.* 2005; 55:425–32.  
<https://doi.org/10.1007/s00280-004-0927-6>  
PMID:15565325
23. Patel K, Doudican NA, Schiff PB, Orlow SJ. Albendazole sensitizes cancer cells to ionizing radiation. *Radiat Oncol.* 2011; 6:160.  
<https://doi.org/10.1186/1748-717X-6-160>  
PMID:22094106
24. Field JJ, Kanakkanthara A, Miller JH. Microtubule-targeting agents are clinically successful due to both mitotic and interphase impairment of microtubule function. *Bioorg Med Chem.* 2014; 22:5050–59.  
<https://doi.org/10.1016/j.bmc.2014.02.035>  
PMID:24650703
25. Dumontet C, Jordan MA. Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat Rev Drug Discov.* 2010; 9:790–803.  
<https://doi.org/10.1038/nrd3253>  
PMID:20885410
26. Wanderley CW, Colón DF, Luiz JP, Oliveira FF, Viacava PR, Leite CA, Pereira JA, Silva CM, Silva CR, Silva RL, Speck-Hernandez CA, Mota JM, Alves-Filho JC, et al. Paclitaxel reduces tumor growth by reprogramming tumor-associated macrophages to an M1 profile in a TLR4-dependent manner. *Cancer Res.* 2018; 78:5891–900.  
<https://doi.org/10.1158/0008-5472.CAN-17-3480>  
PMID:30104241
27. Shurin GV, Tourkova IL, Kaneno R, Shurin MR. Chemotherapeutic agents in noncytotoxic concentrations increase antigen presentation by dendritic cells via an IL-12-dependent mechanism. *J Immunol.* 2009; 183:137–44.  
<https://doi.org/10.4049/jimmunol.0900734>  
PMID:19535620
28. Kaneno R, Shurin GV, Tourkova IL, Shurin MR. Chemomodulation of human dendritic cell function by antineoplastic agents in low noncytotoxic concentrations. *J Transl Med.* 2009; 7:58.  
<https://doi.org/10.1186/1479-5876-7-58>  
PMID:19591684
29. Sevko A, Michels T, Vrohligs M, Umansky L, Beckhove P, Kato M, Shurin GV, Shurin MR, Umansky V. Antitumor effect of paclitaxel is mediated by inhibition of myeloid-derived suppressor cells and chronic inflammation in the spontaneous melanoma model. *J Immunol.* 2013; 190:2464–71.  
<https://doi.org/10.4049/jimmunol.1202781>  
PMID:23359505
30. Nishida N, Komori Y, Takarada O, Watanabe A, Tamura S, Kubo S, Shimada I, Kikkawa M. Structural basis for two-way communication between dynein and microtubules. *Nat Commun.* 2020; 11:1038.  
<https://doi.org/10.1038/s41467-020-14842-8>  
PMID:32098965
31. Cianfrocco MA, DeSantis ME, Leschziner AE, Reck-Peterson SL. Mechanism and regulation of cytoplasmic dynein. *Annu Rev Cell Dev Biol.* 2015; 31:83–108.  
<https://doi.org/10.1146/annurev-cellbio-100814-125438> PMID:26436706
32. Hoang HT, Schlager MA, Carter AP, Bullock SL. DYNC1H1 mutations associated with neurological diseases compromise processivity of dynein-dynactin-cargo adaptor complexes. *Proc Natl Acad Sci USA.* 2017; 114:E1597–606.  
<https://doi.org/10.1073/pnas.1620141114>  
PMID:28196890

33. Gong LB, Wen T, Li Z, Xin X, Che XF, Wang J, Liu YP, Qu XJ. DYNC1I1 promotes the proliferation and migration of gastric cancer by up-regulating IL-6 expression. *Front Oncol.* 2019; 9:491.  
<https://doi.org/10.3389/fonc.2019.00491>  
PMID:[31249807](https://pubmed.ncbi.nlm.nih.gov/31249807/)
34. Palaniappan A, Ramar K, Ramalingam S. Computational identification of novel stage-specific biomarkers in colorectal cancer progression. *PLoS One.* 2016; 11:e0156665.  
<https://doi.org/10.1371/journal.pone.0156665>  
PMID:[27243824](https://pubmed.ncbi.nlm.nih.gov/27243824/)
35. Sucularli C, Arslantas M. Computational prediction and analysis of deleterious cancer associated missense mutations in DYNC1H1. *Mol Cell Probes.* 2017; 34:21–29.  
<https://doi.org/10.1016/j.mcp.2017.04.004>  
PMID:[28455235](https://pubmed.ncbi.nlm.nih.gov/28455235/)
36. Huang GS, Brouwer-Visser J, Ramirez MJ, Kim CH, Hebert TM, Lin J, Arias-Pulido H, Qualls CR, Prossnitz ER, Goldberg GL, Smith HO, Horwitz SB. Insulin-like growth factor 2 expression modulates taxol resistance and is a candidate biomarker for reduced disease-free survival in ovarian cancer. *Clin Cancer Res.* 2010; 16:2999–3010.  
<https://doi.org/10.1158/1078-0432.CCR-09-3233>  
PMID:[20404007](https://pubmed.ncbi.nlm.nih.gov/20404007/)
37. Lian F, Xing X, Yuan G, Schäfer C, Rauser S, Walch A, Röcken C, Ebeling M, Wright MB, Schmid RM, Ebert MP, Burgermeister E. Farnesoid X receptor protects human and murine gastric epithelial cells against inflammation-induced damage. *Biochem J.* 2011; 438:315–23.  
<https://doi.org/10.1042/BJ20102096> PMID:[21619550](https://pubmed.ncbi.nlm.nih.gov/21619550/)
38. Itoh Y, Sendo T, Hirakawa T, Takasaki S, Goromaru T, Nakano H, Oishi R. Pemirolast potently attenuates paclitaxel hypersensitivity reactions through inhibition of the release of sensory neuropeptides in rats. *Neuropharmacology.* 2004; 46:888–94.  
<https://doi.org/10.1016/j.neuropharm.2003.11.018>  
PMID:[15033348](https://pubmed.ncbi.nlm.nih.gov/15033348/)
39. Roglio I, Bianchi R, Camozzi F, Carozzi V, Cervellini I, Crippa D, Lauria G, Cavaletti G, Melcangi RC. Docetaxel-induced peripheral neuropathy: protective effects of dihydroprogesterone and progesterone in an experimental model. *J Peripher Nerv Syst.* 2009; 14:36–44.  
<https://doi.org/10.1111/j.1529-8027.2009.00204.x>  
PMID:[19335538](https://pubmed.ncbi.nlm.nih.gov/19335538/)
40. Kon T, Sutoh K, Kurisu G. X-ray structure of a functional full-length dynein motor domain. *Nat Struct Mol Biol.* 2011; 18:638–42.  
<https://doi.org/10.1038/nsmb.2074>  
PMID:[21602819](https://pubmed.ncbi.nlm.nih.gov/21602819/)
41. Schiff PB, Horwitz SB. Taxol assembles tubulin in the absence of exogenous guanosine 5'-triphosphate or microtubule-associated proteins. *Biochemistry.* 1981; 20:3247–52.  
<https://doi.org/10.1021/bi00514a041>  
PMID:[6113842](https://pubmed.ncbi.nlm.nih.gov/6113842/)
42. Schiff PB, Fant J, Horwitz SB. Promotion of microtubule assembly in vitro by taxol. *Nature.* 1979; 277:665–67.  
<https://doi.org/10.1038/277665a0>  
PMID:[423966](https://pubmed.ncbi.nlm.nih.gov/423966/)
43. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016; 44:e71.  
<https://doi.org/10.1093/nar/gkv1507>  
PMID:[26704973](https://pubmed.ncbi.nlm.nih.gov/26704973/)
44. Li Z, Li AD, Xu L, Bai DW, Hou KZ, Zheng HC, Qu XJ, Liu YP. SPARC expression in gastric cancer predicts poor prognosis: results from a clinical cohort, pooled analysis and GSEA assay. *Oncotarget.* 2016; 7:70211–22.  
<https://doi.org/10.18632/oncotarget.12191>  
PMID:[28053291](https://pubmed.ncbi.nlm.nih.gov/28053291/)
45. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–87.  
<https://doi.org/10.1089/omi.2011.0118>  
PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
46. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, Castro MA, Gibb EA, Kanchi RS, et al, and TCGA Research Network. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell.* 2017; 171:540–56.e25.  
<https://doi.org/10.1016/j.cell.2017.09.007>  
PMID:[28988769](https://pubmed.ncbi.nlm.nih.gov/28988769/)
47. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, Reeser JW, Yu L, Roychowdhury S. Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol.* 2017; 2017:PO.17.00073.  
<https://doi.org/10.1200/PO.17.00073>  
PMID:[29850653](https://pubmed.ncbi.nlm.nih.gov/29850653/)
48. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MH, Goldinger SM, Utikal J, Hassel JC, Weide B, et al. Genomic

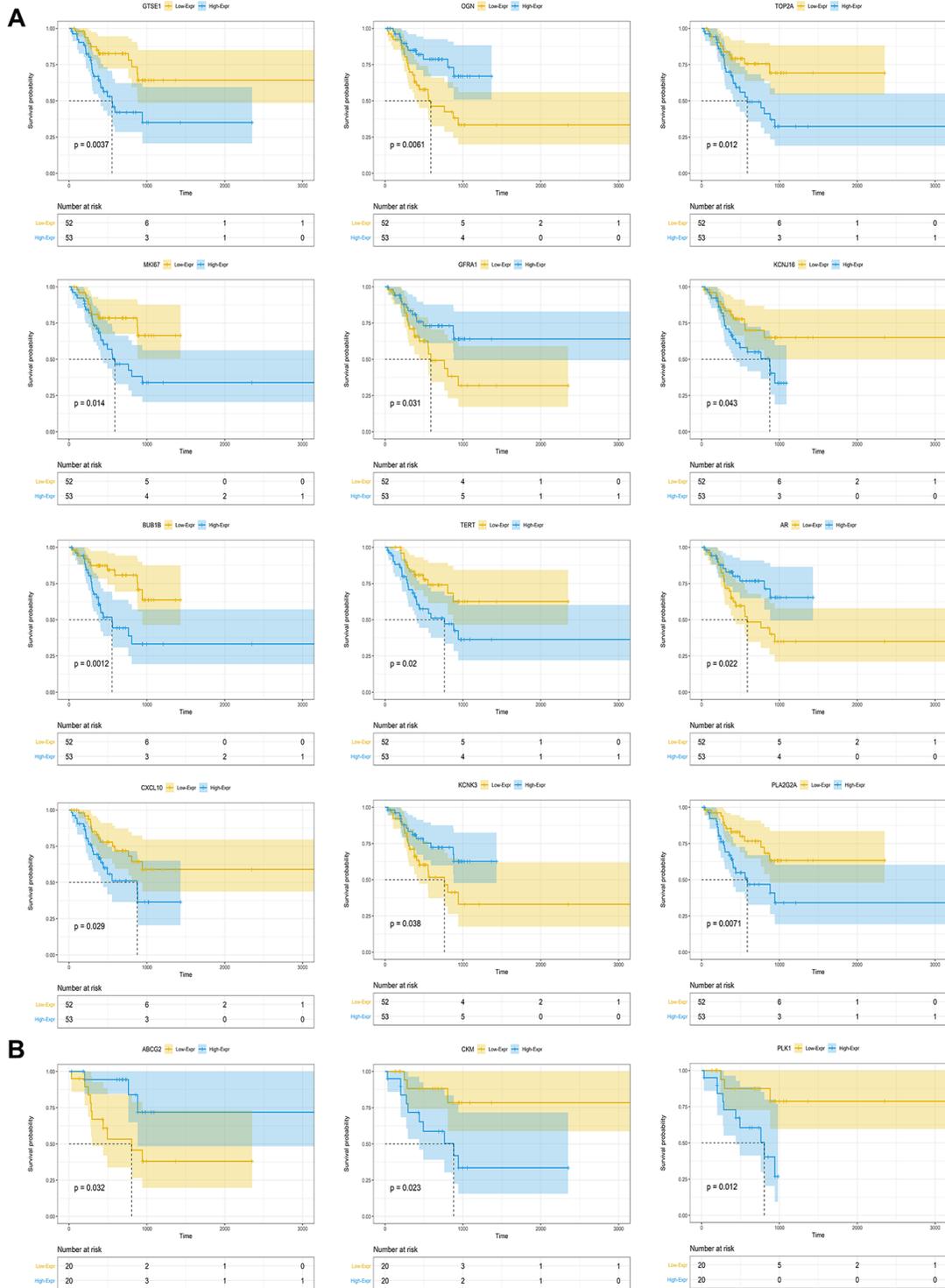
correlates of response to CTLA-4 blockade in metastatic melanoma. Science. 2015; 350:207–11.

<https://doi.org/10.1126/science.aad0095>

PMID:[26359337](https://pubmed.ncbi.nlm.nih.gov/26359337/)

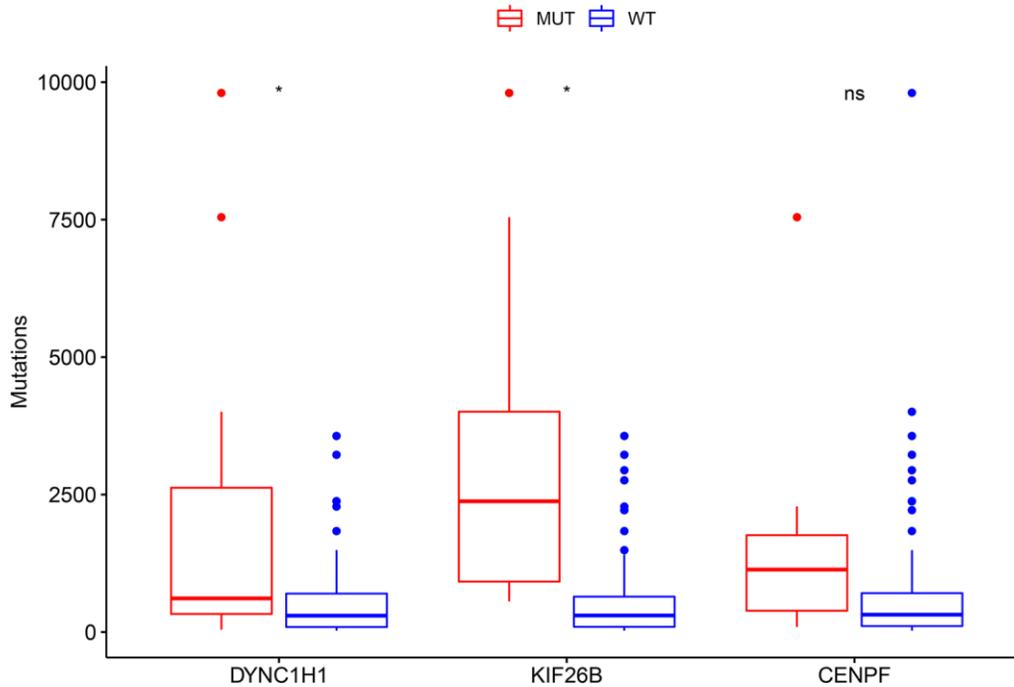
# SUPPLEMENTARY MATERIALS

## Supplementary Figures

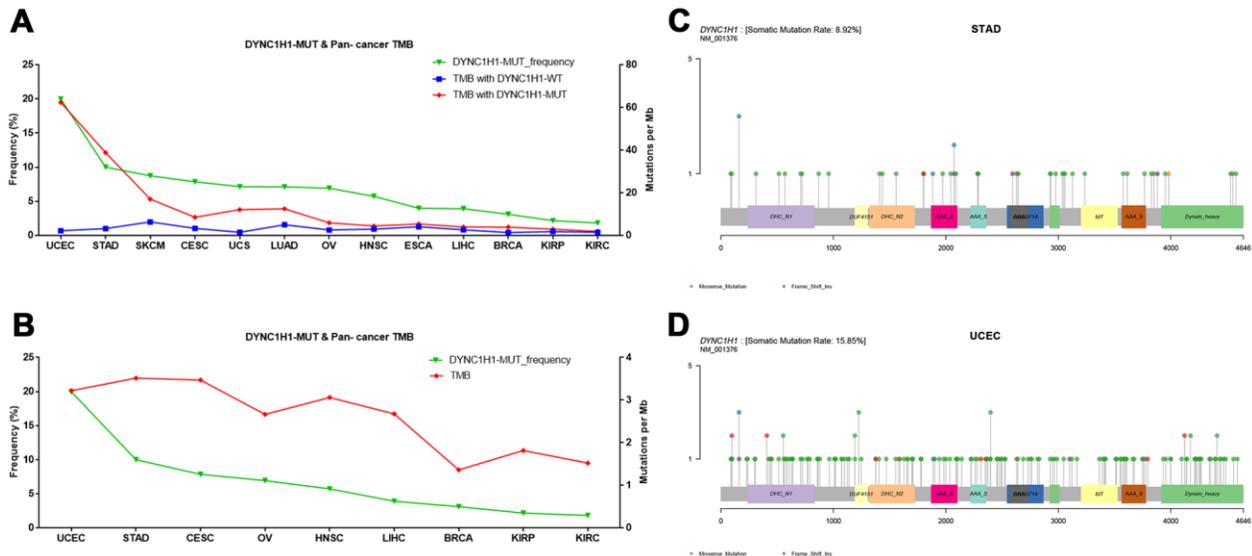


**Supplementary Figure 1. The microtubule inhibitors (MTIs) related DEGs in the survival analysis of TCGA-STAD cohort.** Kaplan-Meier plots for overall survival of DEGs related to MTIs in TMB-H group (A) and MSI-H group (B) by log-rank test.  $P$ -value  $<0.05$  was considered significant.

### TMB in immunotherapy cohort

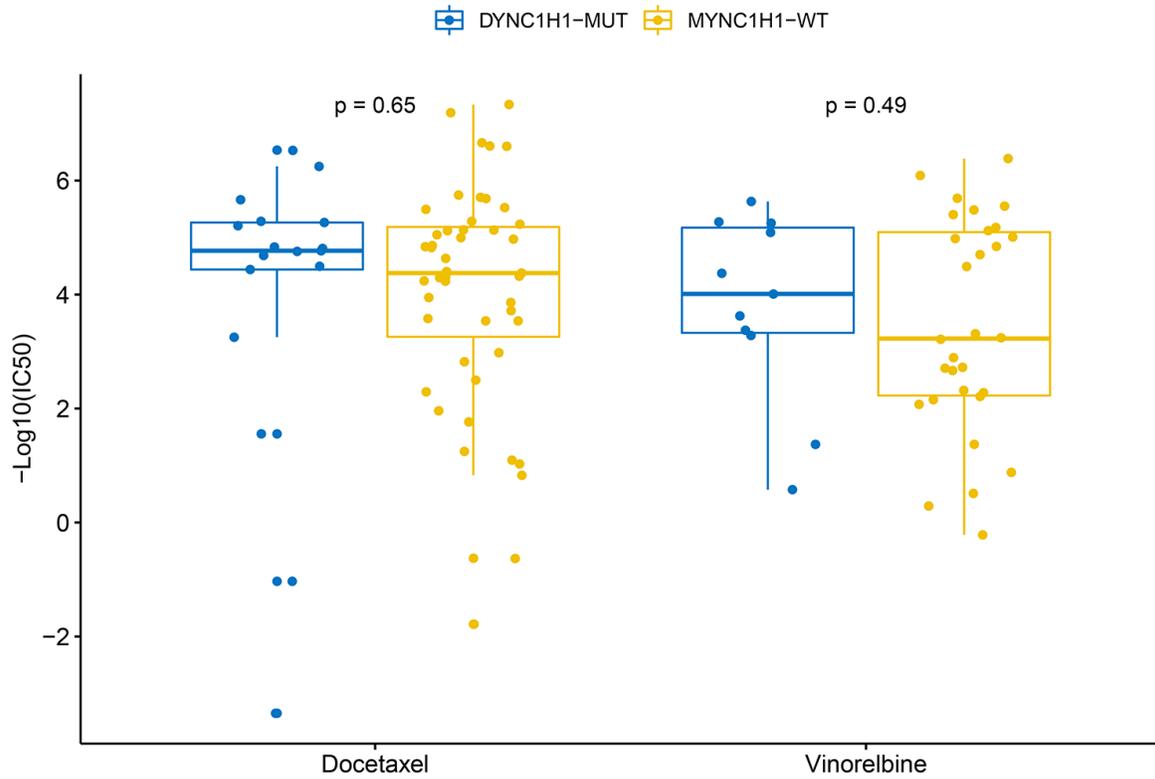


**Supplementary Figure 2. Association of mutant DYNC1H1 with TMB in immunotherapy cohort.** Levels of TMB in immunotherapy cohort (Allen cohort), stratified by genes (DYNC1H1, KIF26B, and CENPF) mutation status. MUT: mutated, WT: wild type. Student's t-test obtained  $p$ -values. (\* $P < 0.05$ ; ns, not significant).



**Supplementary Figure 3. Association of mutant DYNC1H1 with TMB in various cancers.** (A) Line chart showing DYNC1H1 mutation frequency (green) and levels of TMB with DYNC1H1 mutated (red) and wild type (blue) in 15 cancers. (B) Line chart presenting DYNC1H1 mutation frequency (green) and levels of TMB (red) in 9 cancers. (C, D) Lollipopplots presenting the mutation spots of DYNC1H1 in gastric cancer (GC)(C) and Uterine Corpus Endometrial Carcinoma (UCEC)(D). Abbreviation: UCEC, Uterine Corpus Endometrial Carcinoma; STAD, Stomach Adenocarcinoma; SKCM, Skin Cutaneous Melanoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; UCS, Uterine Carcinosarcoma; LUAD, Lung Adenocarcinoma; OV, Ovarian serous cystadenocarcinoma; HNCS, Head and Neck Cancer; ESCA, Esophageal Cancer; LIHC, Liver hepatocellular carcinoma; BRCA, Breast invasive carcinoma; KIRP, Kidney renal papillary cell carcinoma; KIRC, Kidney renal clear cell carcinoma.

### DYNC1H1 status in STAD cells of GDSC



**Supplementary Figure 4. Association of mutant DYNC1H1 with microtubule inhibitors (MTIs) in gastric cancer cell lines with high immune activity based on GDSC database.** Histograms depicting different MTIs activities of docetaxel and vinorelbine, stratified by DYNC1H1 status in gastric cancer cell lines based on GDSC database ( $p$ -value  $< 0.05$  by Student's  $t$ -test). MUT: mutated, WT: wild type.

## SUPPLEMENTARY TABLES

Please browse Full Text version to see the data of Supplementary Tables 1 and 3.

**Supplementary Table 1. The common DEGs of MSI-H group and TMB-H group.**

**Supplementary Table 2. The functional gene sets of each drug type from molecular signatures database.**

Candidate drug types	Functional gene sets
1. Microtubule inhibitor	BIOCARTA_G2_PATHWAY
	BIOCARTA_MTOR_PATHWAY
	BIOCARTA_NFKB_PATHWAY
	GO_MICROTUBULE
	GO_MICROTUBULE_BINDING
	GO_MICROTUBULE_DEPOLYMERIZATION
	GO_MICROTUBULE_NUCLEATION
	GO_MICROTUBULE_POLYMERIZATION
	GO_MICROTUBULE_POLYMERIZATION_OR_DEPOLYMERIZATION
	GO_REGULATION_OF_MICROTUBULE_POLYMERIZATION_OR_DEPOLYMERIZATION
	GO_S_ADENOSYLHOMOCYSTEINE_METABOLIC_PROCESS
	MICROTUBULE
	MICROTUBULE_BINDING
	MICROTUBULE_POLYMERIZATION_OR_DEPOLYMERIZATION
	REACTOME_POST_CHAPERONIN_TUBULIN_FOLDING_PATHWAY
	REACTOME_FORMATION_OF_TUBULIN_FOLDING_INTERMEDIATES_BY_CCT_TRIC
2. Corticosteroid agonist	GO_CELLULAR_RESPONSE_TO_CORTICOSTEROID_STIMULUS
	GO_REGULATION_OF_CORTICOSTEROID_HORMONE_SECRETION
	GO_RESPONSE_TO_CORTICOSTEROID
	GO_RESPONSE_TO_CORTICOSTERONE
3. Androgen receptor agonist	GO_ANDROGEN_RECEPTOR_BINDING
	GO_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY
	GO_NEGATIVE_REGULATION_OF_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY
	GO_REGULATION_OF_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY
4. Protein synthesis inhibitor	GO_TESTOSTERONE_BIOSYNTHETIC_PROCESS
	GO_POLYSOME
	GO_RIBOSOMAL_SMALL_SUBUNIT_BIOGENESIS
	GO_RIBOSOMAL_LARGE_SUBUNIT_BIOGENESIS
	GO_RIBOSOME_BIOGENESIS
	GO_SIGNAL_RECOGNITION_PARTICLE_ENDOPLASMIC_RETICULUM_TARGETING
	GO_TRNA_CATABOLIC_PROCESS
GO_TRNA_METABOLIC_PROCESS	
5. Cyclooxygenase inhibitor	GO_EUKARYOTIC_TRANSLATION_INITIATION_FACTOR_3_COMPLEX
	0
6. Leukotriene receptor antagonist	GO_LEUKOTRIENE_D4_BIOSYNTHETIC_PROCESS

7. Adrenergic receptor antagonist

- GO\_ALPHA\_2A\_ADRENERGIC\_RECEPTOR\_BINDING
- GO\_ADRENERGIC\_RECEPTOR\_BINDING
- GO\_ADRENERGIC\_RECEPTOR\_SIGNALING\_PATHWAY
- GO\_BETA\_2\_ADRENERGIC\_RECEPTOR\_BINDING
- GO\_NEGATIVE\_REGULATION\_OF\_ADENYLATE\_CYCLASE\_ACTIVATING\_ADRENERGIC\_RECEPTOR\_SIGNALING\_PATHWAY

---

**Supplementary Table 3. Association of tubulin related drugs in gastric cancer cell lines with high immune activity based on GDSC database.**

## Identification of “regulation of RhoA activity panel” as a prognostic and predictive biomarker for gastric cancer

Wenwen Huang<sup>1,\*</sup>, Songhui Zhao<sup>2,\*</sup>, Cheng Zhang<sup>1</sup>, Zhongwu Li<sup>3</sup>, Sai Ge<sup>1</sup>, Baofeng Lian<sup>2</sup>, Hui Feng<sup>4</sup>, Kai Wang<sup>2</sup>, Ruihua Xu<sup>5</sup>, Jiafu Ji<sup>6</sup>, Jing Gao<sup>7,#</sup>, Weiwei Shi<sup>2,8,#</sup>, Lin Shen<sup>1,#</sup>

<sup>1</sup>Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education, Beijing), Peking University Cancer Hospital and Institute, Beijing 100142, China

<sup>2</sup>OrigiMed Inc., Shanghai 201112, China

<sup>3</sup>Department of Pathology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education, Beijing), Peking University Cancer Hospital and Institute, Beijing 100142, China

<sup>4</sup>Shanghai Junshi Biosciences Co., Ltd, Shanghai 201203, China

<sup>5</sup>State Key Laboratory of Oncology in South China, Department of Medical Oncology, Sun Yat-sen University Cancer Center, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China

<sup>6</sup>Department of Gastrointestinal Surgery, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education, Beijing), Peking University Cancer Hospital and Institute, Beijing 100142, China

<sup>7</sup>National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, China

<sup>8</sup>Department of Thoracic Surgery, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai 200433, China

\*Equal contribution

#Senior co-authors

**Correspondence to:** Lin Shen, Weiwei Shi, Jing Gao; **email:** [shenlin@bjmu.edu.cn](mailto:shenlin@bjmu.edu.cn), [shiw@origimed.com](mailto:shiw@origimed.com), [gaojing\\_pumc@163.com](mailto:gaojing_pumc@163.com), <https://orcid.org/0000-0002-3613-9413>

**Keywords:** regulation of RhoA activity, gastric cancer, prognosis and predictive biomarker, tumor migration, tumor microenvironment

**Received:** February 25, 2020

**Accepted:** September 14, 2020

**Published:** December 03, 2020

**Copyright:** © 2020 Huang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

RhoA is a member of the RHO family GTPases and is associated with essential functions in gastric cancer. In this study, we identified a gastric cancer biomarker, termed the “regulation of RhoA activity panel” (RRAP). Patients with gastric cancer from The Cancer Genome Atlas database were divided into training (N=160) and validation (N=155) cohorts. A cohort of 109 Chinese gastric cancer patients was utilized as an independent validation. Patients with mutated RRAP showed significantly better overall survival than patients with wild type RRAP. We also analyzed the association between RRAP and the migration capacity, immune-related signatures, and the tumor microenvironment. RRAP-mutant tumors had a significantly lower degree of lymph node metastasis and lower activities of migration-related pathways. These tumors also showed significantly increased immune cell infiltration and cytotoxic activity. Furthermore, two independent patient cohorts who received immune checkpoint blockade therapy were assessed for RRAP mutant status. As expected, for both immunotherapy cohorts, higher response rates to immune checkpoint blockade therapy were observed in patients with RRAP-mutant tumors than in patients with wild type RRAP tumors. Overall, this study indicates that the RRAP gene set is a potential biomarker for gastric cancer prognosis and therapeutic selection.

## INTRODUCTION

Gastric cancer is one of the most frequently occurring and lethal malignancies [1]. The Lauren classification divides gastric cancer into 2 main subtypes, intestinal and diffuse [2]. However, this classification provides limited guidance for disease prognosis and treatment decisions. Global efforts to characterize gastric cancer at the molecular level from the perspective of cancer genomics and transcriptomics have been made, including those from The Cancer Genome Atlas (TCGA) [3] and the Asian Cancer Research Group [4]. Gastric cancer patients are classified into 4 subtypes based on gene expression profiling; each subtype exhibits distinct patterns of molecular alterations, disease progression, and prognosis [4]. Although these large-scale efforts have provided comprehensive insights into gastric cancer, they have not translated into a clinical benefit. A genomic-based molecular biomarker with prognostic and/or therapy predictive value is still needed for gastric cancer.

Among the various genomic alternations that occur in gastric cancer, *RHOA* mutation plays a critical role in the development and progression of cancer by regulating actin organization [5], cell migration [6], cytokinesis and the cell cycle [7]. Recent studies have also suggested its potential role in modulating the tumor microenvironment (TME) of cancers [8, 9]. Although the overexpression of RhoA has been frequently recognized in various cancers and was found to be significantly associated with poor prognosis in gastric cancer [10], similar overall survival (OS) rates were nevertheless observed between patients with *RHOA* mutant and wild-type gastric cancers [11, 12]; therefore, the prognostic value of this gene mutation is poor. In 2017, Shi et al. [13] established that mutations within the gene set designated as the “regulation of RhoA activity pathway” were associated with better progression-free survival (PFS) and overall survival in HER2+ breast cancer patients. This gene set includes RhoA, as well as guanine nucleotide exchange factors and GTPase-activating proteins [14], both of which are involved in regulating RhoA activity. Furthermore, altered RhoA signaling has been reported in gastric cancers, especially in the diffuse type gastric cancer [3, 15–21]. However, the clinical significance of this pathway in overall gastric cancer remains unresolved. Mutations in “regulation of RhoA activity pathway” gene set may be involved in the prognosis and therapeutic prediction of gastric cancer through their effect on RhoA and its effector molecule activity.

Based on the pathway changes of RhoA activity, we developed a statistically optimized gene subset as a biomarker by applying a genetic algorithm to a training

gastric cancer cohort obtained from the gastric cancer dataset of TCGA. We validated this biomarker in a nonoverlapped TCGA-validation cohort and in an independent Chinese gastric cancer (CGC) cohort. The association of this biomarker with lymph node metastasis, migration-related pathways, immune-related signatures, and the TME was assessed to glean insights into possible related mechanisms. Inspired by its effect on the TME, we collected 2 independent cohorts of gastric cancer patients (denoted as IM1 and IM2) who received immune checkpoint blockade (ICB) treatment and revealed the potential predictive capability of this biomarker.

## RESULTS

### Prognostic biomarker “regulation of RhoA activity panel” for gastric cancer

As shown in Figure 1, the potential biomarker “regulation of RhoA activity panel” (RRAP) was calculated from the TCGA training cohort with a genetic algorithm on the “regulation of RhoA activity pathway” gene set (see Methods). The resulting optimal solution, containing 20 genes, is hereafter denoted as RRAP (Figure 1B). These gene mutations were significantly associated with gastric cancer (Supplementary Table 1). According to whether or not mutations occurred in the coding region of any of the RRAP genes, patients were classified as RRAP-wild type or RRAP-mutant. In the TCGA training cohort, RRAP-mutant patients displayed better OS compared to RRAP-wild type patients, with a hazard ratio of 0.4 (95% confidence interval: 0.2-0.79, p-value = 0.006, Figure 2A). Importantly, this was also validated in the TCGA validation cohort, with a hazard ratio of 0.48 (95% confidence interval: 0.26-0.91, p-value = 0.021, Figure 2B). Furthermore, we incorporated age, gender, Lauren classification, pathologic stage, and RRAP into a multivariate Cox analysis for overall survival. The results showed that RRAP was an independent prognostic factor after adjusting for these clinicopathological parameters in both the TCGA training (hazard ratio = 0.35, 95% confidence interval: 0.17-0.7, p-value = 0.003; Supplementary Figure 1A) and TCGA validation cohorts (hazard ratio = 0.51, 95% confidence interval: 0.26-1, p-value = 0.05; Supplementary Figure 1B).

The reliability of RRAP was assessed using an independent patient cohort that was a subset of the 109 CGC patients. These patients were whole-exome sequenced, and 26 (23.9%) were identified as RRAP-mutant. This cohort exhibited a mutational proportion similar to the TCGA training (26.3%) and TCGA validation (28.3%) cohorts. For the reliability

assessment cohort, Kaplan-Meier curves showed that RRAP-mutant was significantly associated with longer OS (hazard ratio = 0.13, 95% confidence interval: 0.03-0.52, p-value = 6.66e-4, Figure 2C). Multivariate Cox regression analysis was performed as described above, and the results again showed that RRAP was an independent prognostic factor (Supplementary Figure 1C). We also merged data of the TCGA training, TCGA validation, and CGC cohorts into overall cohort. RRAP-mutant patients displayed better OS compared to RRAP-wild type patients, with a hazard ratio of 0.38 (95% confidence interval: 0.25-0.59, p-value = 7.37e-6) in the overall cohort (Figure 2D and Supplementary Figure 1D).

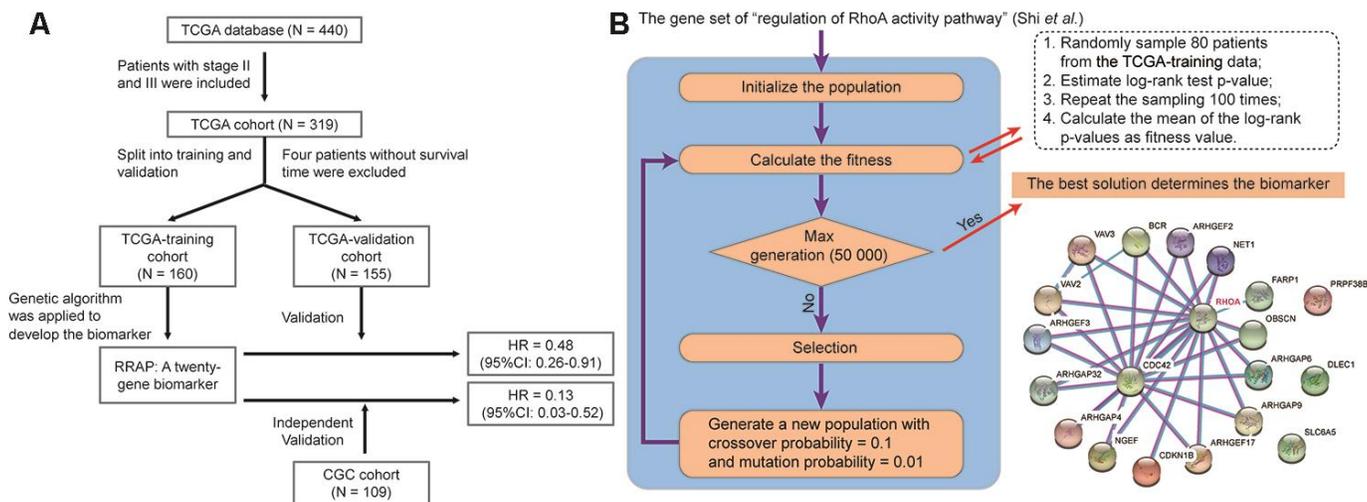
### RRAP was associated with cell migration activity

Given the role of RhoA in cell migration, we assessed the effect of RRAP on tumor metastasis. Strikingly, patients with RRAP-mutant tumors had a significantly lower rate of lymph node metastasis (pN3 of regional lymph node) compared with RRAP-wild type patients (11.5% vs 36.1%, respectively, Fisher's exact test p-value = 0.026) in the CGC cohort, which was also confirmed in the overall TCGA cohort (12.8% vs 24.7%, respectively, p-value = 0.03). We also found that patients with RRAP-mutant tumors had a moderately lower distant metastasis recurrence risk compared to patients with RRAP-wild type tumors in the CGC cohort (15.4% vs 31.3%, respectively, p-value = 0.136); the same trend was observed in the overall TCGA cohort (3.6% vs. 7.9%, respectively, p-value = 0.26). This suggested that RRAP-mutant tumors may have an impaired migration capacity. At the molecular level, we explored the RNA-seq data collected from

TCGA and estimated the enrichment score of 4 migration-related functions—adherens junctions, cell adhesion molecules, focal adhesion, and regulation of the actin cytoskeleton. All of these functions showed significantly lower activity in RRAP-mutant tumors compared to RRAP-wild type tumors (Figure 3A–3D). Further analysis revealed that 39 genes related to the 4 functions exhibited significantly lower expression in RRAP-mutant tumors (log2 fold change > 1 and adjusted p-value < 0.01, Figure 3E). We next performed univariate Cox and Kaplan-Meier analysis on the 39 genes and found that 5 genes (*CLDN11*, *CLDN6*, *CLDN9*, *VTN*, and *F2*; Supplementary Table 2) were significantly associated with poor OS (Figure 3F, log-rank p-value < 0.01). Kaplan-Meier analysis revealed that the low expression of these 5 genes, along with *CNTN1*, *CNTN2*, *CADM1*, *NCAM1*, *FGF19*, and *FGF20*, were significantly associated with better OS (Figure 3G, Supplementary Figure 2). Taken together, these results indicated an association between RRAP mutation and tumor cell migration, thus resulting in metastasis.

### Association of RRAP with immune-related signatures

We next explored the correlation between RRAP and the immune-related signatures of tumor mutational burden (TMB), neoantigen burden (NAB), and deficient mismatch repair (dMMR). Higher TMB was observed in RRAP-mutant patients in the TCGA training, TCGA validation, and CGC cohorts, with p-values of 3.7e-8, 9e-11, and 6.3e-5, respectively (Figure 4A). A significantly higher NAB (available only in the CGC cohort) was observed in RRAP-mutant patients (p-value

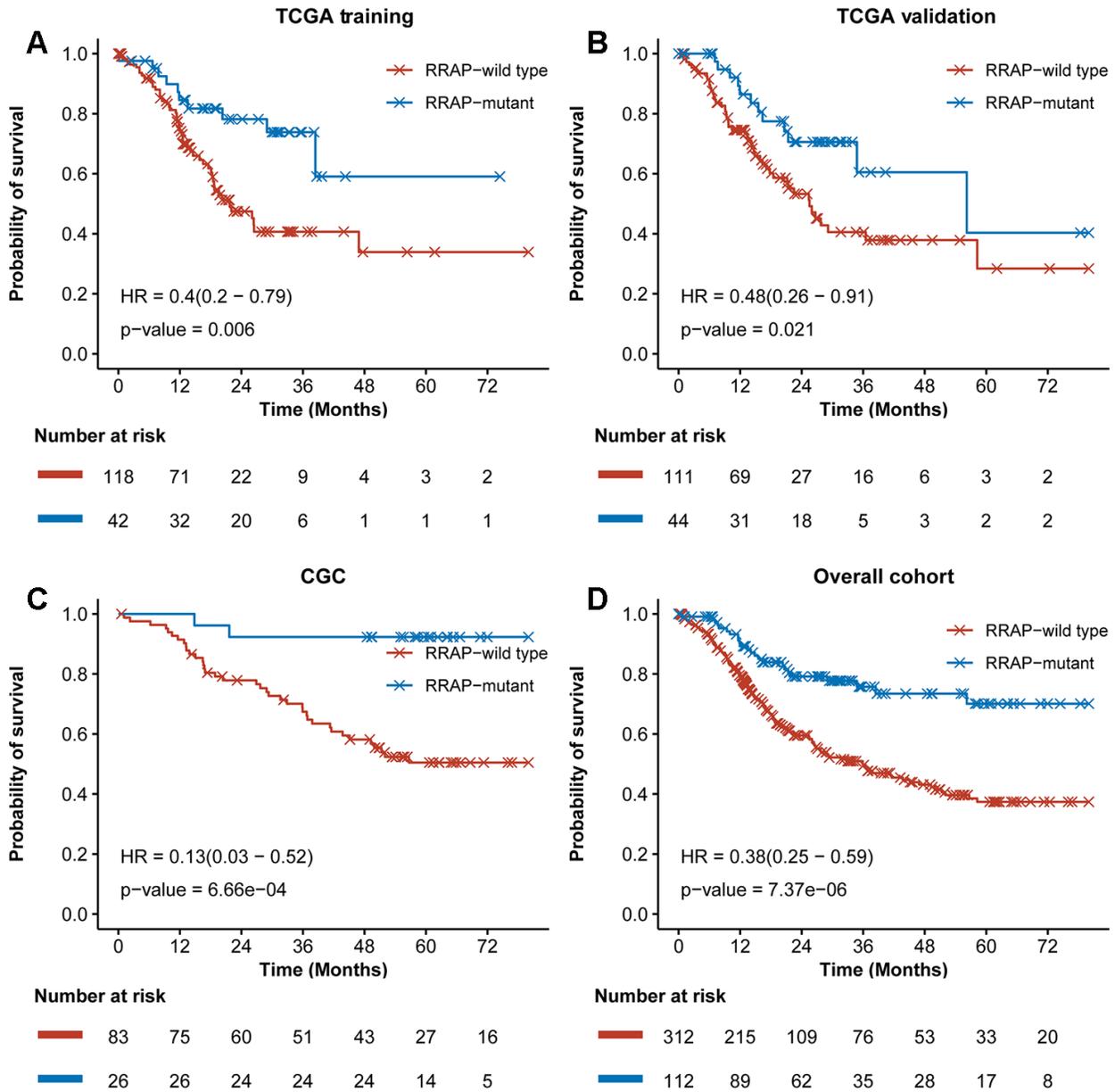


**Figure 1. Identification of RRAP biomarker.** (A) Outline of the cohort construction and analysis workflow. (B) RRAP selection. The right panel shows the interaction of genes among RRAP.

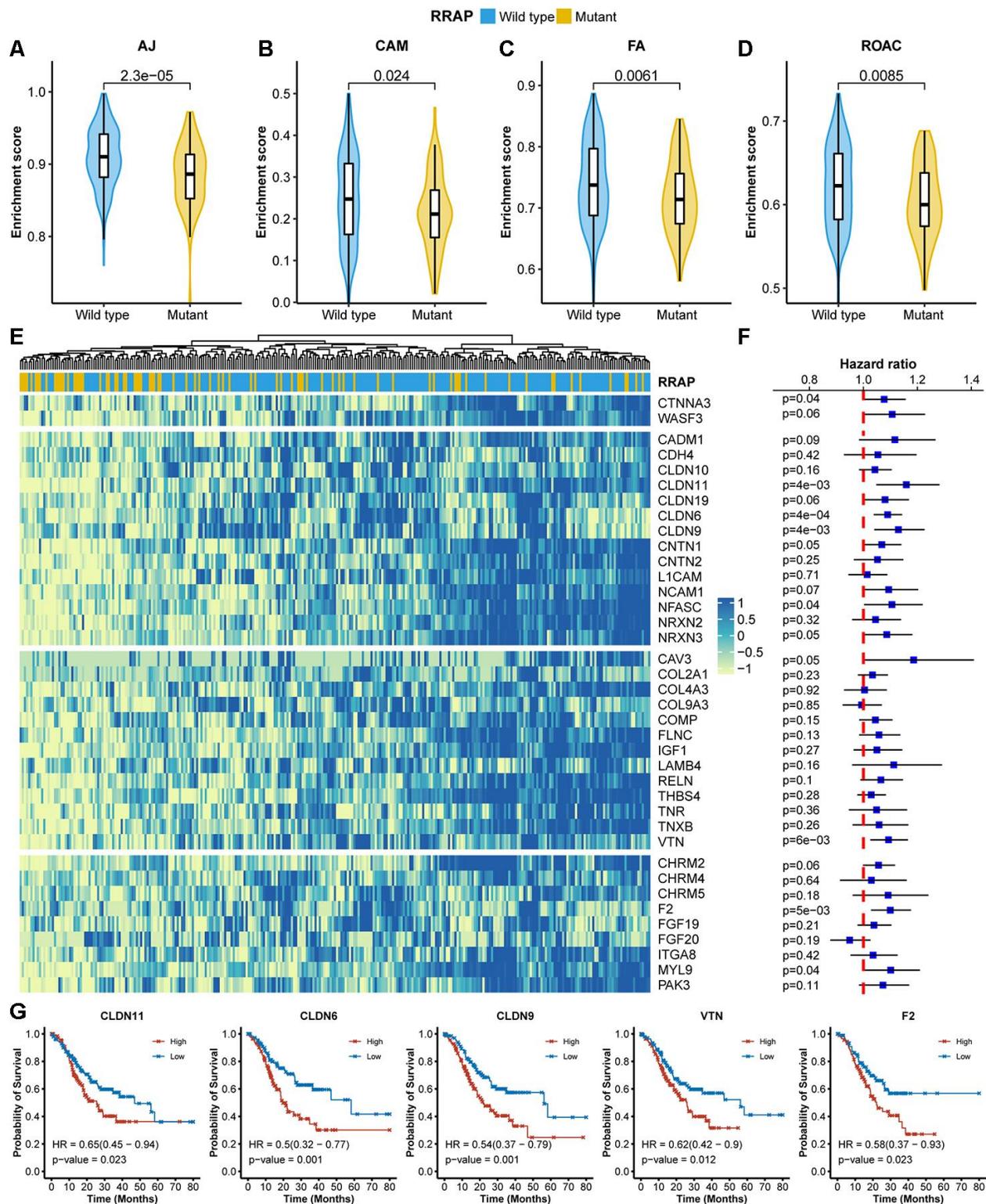
= 0.002, Figure 4B). Additionally, there was a significantly higher fraction of mutational signatures associated with dMMR in the RRAP-mutant tumors compared to RRAP-wild type tumors in all 3 cohorts (Figure 4C). Although the 3 immune-related signatures in this study could predict the response to immunotherapy in multiple cancers [22], TMB and NAB were not associated with the overall survival of gastric cancer patients, while dMMR was related to overall survival in the TCGA validation cohort (Figure 4D).

### Effect of RRAP mutation on the tumor microenvironment

In addition to the 3 signatures above, we further explored the association between RRAP and the TME using TCGA expression data. For this purpose, we collected tumor immune infiltrate data published by TCGA and compared immune cellular fraction between RRAP-mutant and RRAP-wild type tumors [23]. RRAP-mutant tumors showed significantly higher levels of infiltration of activated CD4<sup>+</sup> memory T cells



**Figure 2. Training and validation of RRAP.** Kaplan-Meier survival curves for RRAP-wild type and RRAP-mutant patients in the (A) TCGA training cohort, (B) TCGA validation cohort, (C) an independent CGC cohort, (D) and the overall cohort containing merged data from the three cohorts. P-values were estimated with the log-rank test, and hazard ratios (HRs) were estimated with the Cox model.

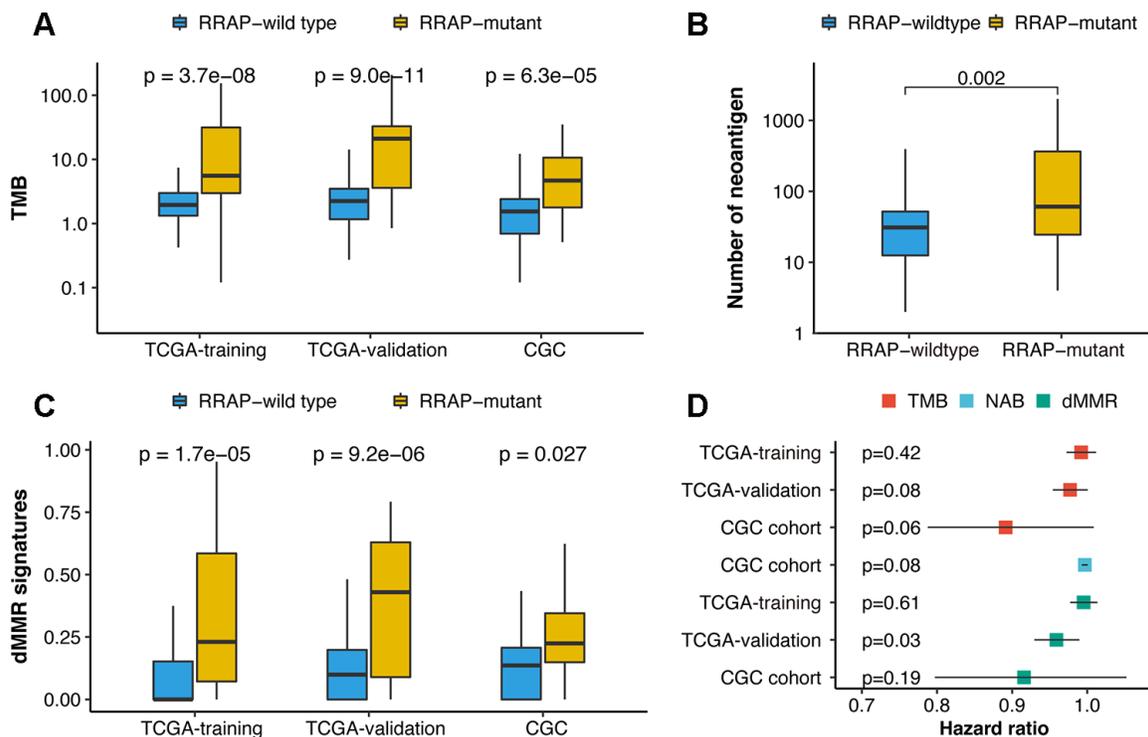


**Figure 3. Association of RRAP with lower cell migration activity.** (A–D) Pathway enrichment scores between RRAP-wild type and RRAP-mutant were compared among 4 functions: adherens junction (AJ, A), cell adhesion molecules (CAM, B), focal adhesion (FA, C), and regulation of the actin cytoskeleton (ROAC, D). The Wilcoxon rank-sum test was applied to estimate differences. (E) Heatmap of differentially expressed genes among the 4 functions; the column was clustered and annotated with RRAP status. (F) Forest plot of association of between gene expression and overall survival. Squares indicate the hazard ratios and error bars represent the 95% confidence interval; the log-rank test was performed to estimate p-values. (G) Kaplan-Meier overall survival curves for patients with high and low expression levels of CLDN11, CLDN6, CLDN9, VTN, and F2. P-values were estimated with the log-rank test, and HRs were estimated with the Cox model.

(p-value =  $8.6 \times 10^{-7}$ ) and M1 macrophages (p-value = 0.00077) (Figure 5A), both of which are reportedly associated with longer survival in multiple cancers [24, 25]. In addition, we observed higher infiltration of CD8+ T cells—the primary effector in anti-tumor immunity—in RRAP-mutant tumors compared to RRAP-wild type tumors, although it did not reach a statistical significance (p-value = 0.07385, Figure 5A). There was a significant difference in the levels of CD8+ infiltration in the independent CGC cohort. IHC staining was used to evaluate the infiltration of CD8+ tumor infiltrating lymphocytes (TILs) in 52 tumors from the CGC cohort; 40 of these were RRAP-wild type tumors and 12 were RRAP-mutant tumors. Consistent with the TCGA RNAseq results, RRAP-mutant tumors exhibited increased infiltration of CD8+ TILs compared to RRAP-wild type tumors (p-value = 0.026) (Figure 5B–5F). Given the functional dependence of CD4+ T cells and M1 macrophages on human leukocyte antigen (HLA) class II molecules, further analyses of several HLA class II molecules were also performed. The results showed significantly higher expression of these molecules in RRAP-mutant tumors compared with RRAP-wild type tumors (Figure 5G), including HLA-

DMA (p-value = 0.0106), HLA-DQA1 (p-value = 0.0235), and HLA-DRA (p-value = 0.0091). Taken together, an increased anti-tumor immune response by CD4+ T cells and macrophages could be seen for RRAP-mutant tumors.

Granule exocytosis-related molecules (e.g., PRF1/GZMA/GZMB) and death ligand pathways (e.g., the Fas/FasL apoptotic killing pathway) are involved in the cytotoxic effects of CD8+ T cells [26] and possibly of CD4+ T cells [27, 28]. We found that PRF1, GZMA, and GZMB were significantly increased in RRAP-mutant tumors compared to RRAP-wild type tumors, with p-values of 0.0056, 0.002, and  $4.5 \times 10^{-5}$ , respectively (Figure 5H). Additionally, RRAP was associated with significantly increased FAS (p-value = 0.0039) and FASLG expression (p-value = 0.0207) (Figure 5I). The apoptosis pathway was also evaluated and exhibited higher activity in RRAP-mutant tumors than in RRAP-wild type tumors (p = 0.0014, Figure 5J). These results strongly suggested that RRAP mutation affected the TME, providing evidence for the association of RRAP-mutant with better OS on the basis of increasing anti-tumor activity in the TME (Supplementary Figure 3).

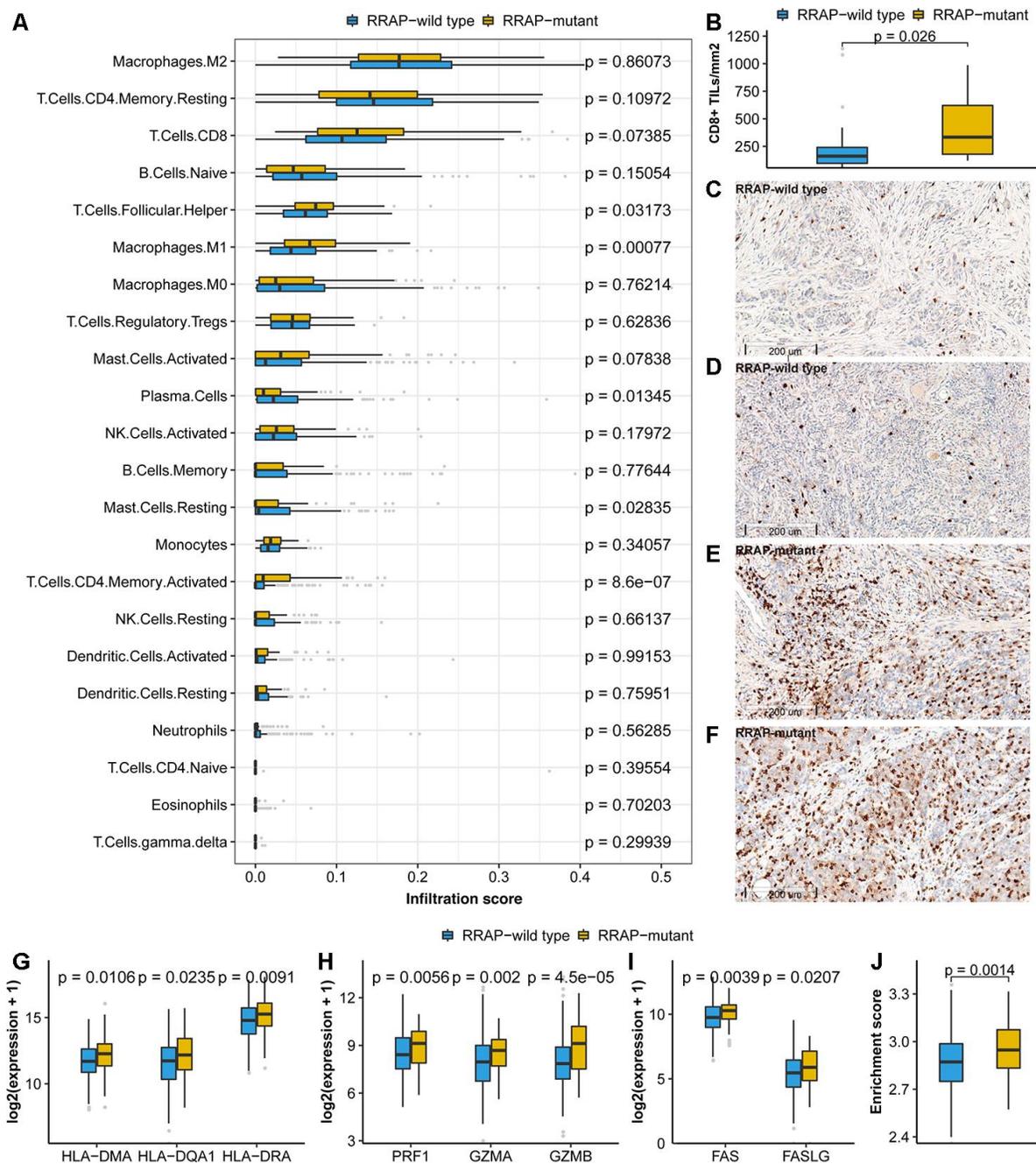


**Figure 4. Association of RRAP with immune-related biomarkers.** (A) Boxplot for differences in the tumor mutational burden (TMB) between RRAP-wild type and RRAP-mutant tumors in the TCGA training, TCGA validation, and CGC cohorts. (B) Boxplot for differences in the neoantigens between RRAP-wild type and RRAP-mutant tumors in the CGC cohort. (C) Boxplot for differences in the dMMR percentage between RRAP-wild type and RRAP-mutant tumors in TCGA training, TCGA validation, and CGC cohort. The Wilcoxon rank-sum test was applied to compare the differences. (D) Association of overall survival and TMB, neoantigen burden (NAB), and dMMR in the 3 cohorts. The hazard ratio was estimated with univariate Cox analysis, and the log-rank test was applied to calculate the p-value.

## RRAP as a potential predictive biomarker for checkpoint inhibitor-based immunotherapy

Inspired by the significant effect of RRAP on prognosis, immune-related signatures and the TME, we further

investigated the association between RRAP and the response to ICB therapy. Of the IM1 cohort who received ICB therapy (N = 37), 7 patients were identified as RRAP-mutant and 30 were RRAP-wild type. After treatment, 15 (50%) patients in the RRAP-



**Figure 5. Association of RRAP with the tumor microenvironment.** (A) Immune infiltrations estimated by CIBERSORT were compared between RRAP-wild type tumors and RRAP-mutant tumors. (B) The density of CD8+ TILs was compared between RRAP-wild type and RRAP-mutant tumors. (C–F) The representative immunohistochemistry for CD8 images of RRAP-wild type tumor tissue (C, D) and RRAP-mutant tumor tissue (E, F). (G–I) Expression values were compared for HLA class II genes (G), cytotoxic effector molecules (H), and apoptosis-related genes (I). (J) The boxplot shows the difference in enrichment score of the apoptosis pathway between RRAP-wild type and RRAP-mutant tumors. The Wilcoxon rank-sum test was applied to estimate the p-values.

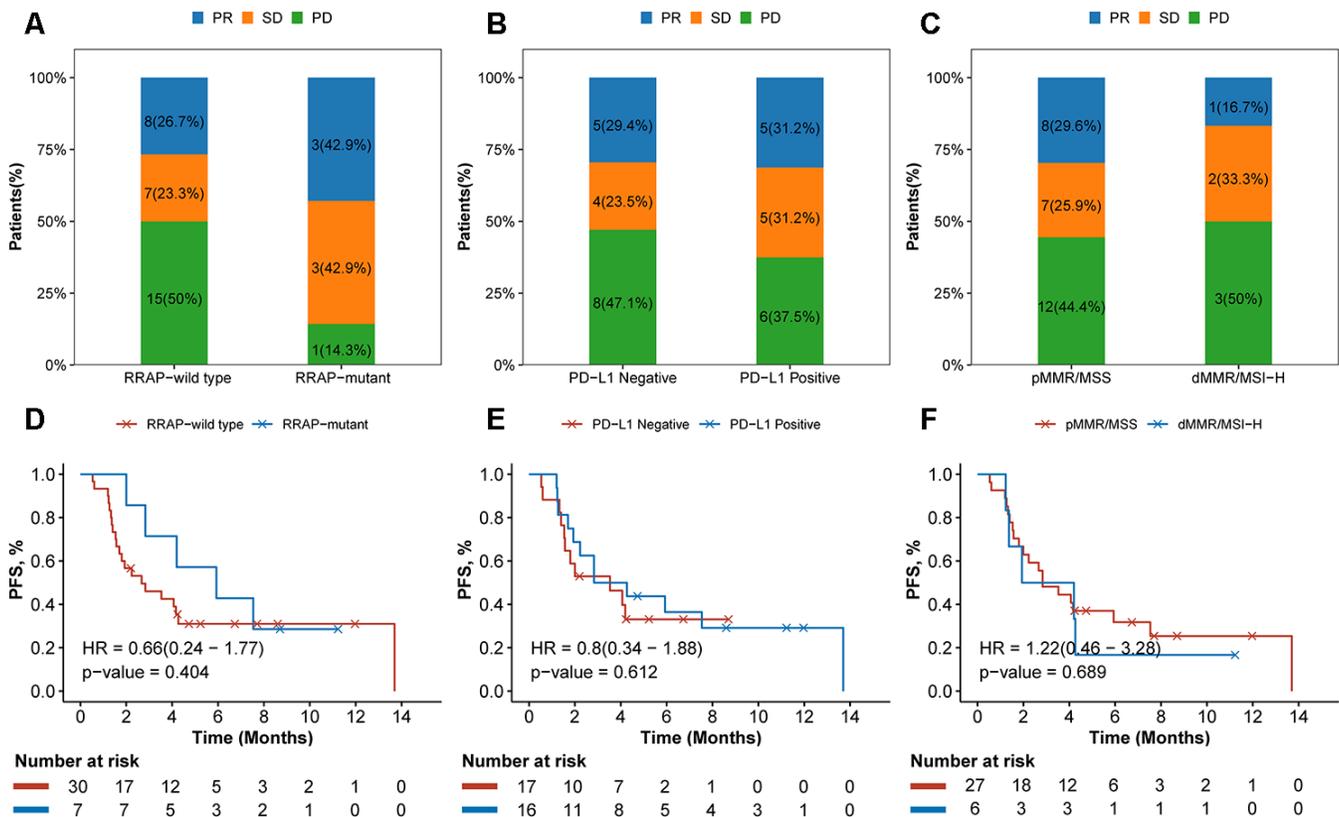
wild type group had progressive disease. In contrast, only 1 (14.3%) patient in the RRAP-mutant group had progressive disease (Figure 6A). We also evaluated the efficacy of ICB therapy between groups stratified by PD-L1 expression and microsatellite instability (MSI) and/or mismatch repair (MMR) status, and no notable differences were found (Figure 6B, 6C). The Kaplan-Meier curves suggested that RRAP-mutant favored PFS compared with RRAP-wild type (Figure 6D), with respective median PFS times of 5.93 months (95% confidence interval: 2.83-not available) and 2.67 months (95% confidence interval: 1.70-not available). However, there was no difference in PFS between PD-L1-positive and PD-L1-negative patients (median PFS: 3.55, 95% confidence interval: 1.93-not available and 3.53, 95% confidence interval: 1.57-not available, respectively) (Figure 6E). The same observation was made when comparing dMMR/MSI-H and pMMR/MSS patients (median PFS: 3.07 95% confidence interval: 1.37-not available and 2.83, 95% confidence interval: 2-not available, respectively) (Figure 6F). In the IM2 cohort (N = 47), 12 patients were classified as RRAP-mutant; these patients had a better overall response rate

than RRAP-wild type patients (33.3% vs 8.6%, respectively, Fisher's exact test p-value = 0.06). These results suggested that RRAP may play a role as a potential predictive biomarker for checkpoint inhibitor-based immunotherapy in gastric cancer.

## DISCUSSION

In this study, we identified RRAP as a biomarker, validated its prognostic effect in TCGA and CGC gastric cancer data sets, and investigated its association with tumor metastasis, the TME, and its potential prediction value for ICB therapies.

For the TCGA training, TCGA validation, and CGC cohorts, the frequencies of RRAP-mutant were elevated (26.3%, 28.3%, and 23.9%, respectively) compared to *RHOA* mutations alone (7.5%, 4.5%, and 3.7%, respectively), suggesting that RRAP can be more broadly utilized as a predictive biomarker for gastric cancer. RRAP-mutant proved to be an independent prognostic factor and was significantly associated with favorable overall survival, regardless of histological



**Figure 6. Association of RRAP and immunotherapy efficacy.** (A–C) Stacked barplot showing the proportion of patients with progressive disease (PD), stable disease (SD), or partial response (PR) for each group divided by RRAP (A), PD-L1 (B) and MSI/MMR status (C). (D–F) The text indicates the number and percentage of patients in each group. Kaplan-Meier survival curves for patients grouped by RRAP (D), PD-L1 (E), and MSI/MMR status (F). P-values were estimated with the log-rank test, and HRs were estimated with the Cox model.

classification (i.e., not only in diffuse gastric cancer). This clinical significance is likely attributed to altered RhoA activity disrupting the process of tumor invasion and metastasis when the process is impaired by RRAP mutation. Clinically, our results showed that patients with RRAP-mutant tumors had a significantly lower rate of lymph node metastasis (pN3) and a lower risk of distant metastasis recurrence. At the molecular level, our analysis suggested that RRAP-mutant tumors exhibited a low activity of migration-related functions (adhesion junctions, cell adhesion molecules, focal adhesion, and regulation of the actin cytoskeleton). In these functions, 11 genes (*CLDN11*, *CLDN6*, *CLDN9*, *VTN*, *F2*, *CNTN1*, *CNTN2*, *CADM1*, *NCAM1*, *FGF19*, and *FGF20*) that were downregulated in RRAP-mutant tumors were significantly associated with improved OS. Our findings are supported by previous results, not only regarding the role of RhoA in tumor cell invasion and metastasis [29] but also regarding the association of RhoA activity with gastric cancer prognosis [30, 31].

Recent efforts have shown that RhoA can modulate the TME [8, 9]. Therefore, we further assessed the significant association between RRAP and clinical outcomes in terms of the TME. Our results indicated that RRAP-mutant patients had higher fractions of activated CD4+ memory T cells, CD8+ T cells, and M1 macrophages, all of which have been reported as biomarkers that are positively associated with overall survival in multiple cancers [24, 25]. We further found that HLA-DMA, HLA-DQA1, and HLA-DRA, on which CD4+ T cells and M1 macrophages functionally depend [32, 33], also showed significantly increased expression in RRAP-mutant tumors. Our experimental CD8+ IHC results were highly consistent with our bioinformatics analysis. Moreover, we also showed that mutated RRAP may affect the TME by regulating the expression levels of granule exocytosis-related molecules (PRF1/GZMA/GZMB) and the death ligand pathway (Fas/FasL apoptotic killing pathway), which are involved in the cytotoxic effects of CD8+ T cells and possibly of CD4+ T cells. Taken together, these results strongly indicated an association between RRAP-mutant and better OS on the basis of increasing antitumor activity in the TME (Supplementary Figure 3). The association between well-known immunotherapy-related biomarkers (TMB, NAB, and dMMR) and the RRAP status was also assessed. Interestingly, all three biomarkers were significantly higher in RRAP-mutant tumors; however, only RRAP had prognostic significance. Moreover, RhoA signaling plays an important role in inducing activating innate immune and adaptive T cell responses [34]. For example, the downregulation of CDC42 reduces NK cell-mediated killing, allowing cancer cells to escape from the human immune response [35]. VAV3, a Rho

family GTPase, activates multiple cell signaling pathways, including NK cell activation [36]. And the role of RhoA for phagocytosis has been studied in macrophages [37, 38]. Besides, RhoA signaling in T cells and B cells is pivotal for activation and migration [34]. The association of RRAP-mutant with immune activation and anti-tumor activity in the tumor immune microenvironment may provide clues to the predictive effect of RRAP in immunotherapy response.

The US Food and Drug Administration approved MSI-H/dMMR as a biomarker for immunotherapy; MSI-H/dMMR occurs in only 4-22% of gastric cancer cases [3, 39, 40] with an approximately 40-57% response [40, 41]. Many studies have shown that TMB, CD8+ TILs, and PD-L1 expression correlate with the efficacy of immunotherapy [42-44]; however, the association between these markers and the clinical benefit of gastric cancer immunotherapy is uncertain [45]. Accordingly, in the current study, 2 retrospective analyses on the IM1 and IM2 cohorts were performed to predict the efficacy of ICB therapy based on RRAP status. The response rate of RRAP-mutant patients was higher than that of RRAP-wild type patients within both the IM1 and IM2 cohorts. The differences are expected to be significant in a larger sized cohort. A clinical trial based on the RRAP biomarker is being designed. Our findings indicate that RRAP plays an important role in the regulation of the TME. Its status as a predictive biomarker will be further verified in larger clinical cohorts receiving immunotherapy.

In summary, this is one of the few efforts of biomarker identification for disease prognosis and therapeutic response based on pathway genomic characteristics in gastric cancer. In gastric cancer patients, RRAP-mutant tumors were correlated with a better prognosis, regardless of the histological classification and clinicopathological parameters. We investigated this correlation mainly from the perspective of RRAP regulating tumor invasion, metastasis, and the TME. Patients with RRAP-mutant tumors showed a better response to checkpoint inhibitor-based immunotherapy. These findings shed light on the clinical implications of the RRAP-mutant biomarker and may be used to guide personalized therapy for gastric cancer patients.

## MATERIALS AND METHODS

### Gastric cancer cohorts

The 4 gastric cancer patient cohorts included in this study were the TCGA cohort (N=315, Supplementary Table 3), the CGC cohort (N = 109, Supplementary Table 3), and 2 immunotherapy cohorts, IM1 (N = 37) and IM2 (N = 47). Clinical and mutation data from the

TCGA database (N = 440) were downloaded from the cBioportal database [46], and survival data were collected from a previous study [47]. Only patients with American Joint Committee on Cancer (AJCC) version 8 stage II and III were included (N = 319). Four patients without a survival time were excluded; the final number of patients in the TCGA cohort was 315. The TCGA training (N = 160) and validation (N = 155) cohorts were generated by random sampling to be approximately equal in sample size. Formalin-fixed, paraffin-embedded tumor tissue samples from the CGC cohort were collected from 109 stage II–III treatment-naive gastric cancer patients who had primary gastric cancer resection at Peking University Cancer Hospital and Institute between 2008 and 2015. The tumor tissue samples of the IM1 cohort were obtained from metastatic gastric cancer patients before they started ICB treatment. After the samples were collected, these 37 patients received at least 1 cycle of any ICB therapy regardless of the agent's target (i.e., PD-1/PD-L1). The IM2 cohort containing 47 patients who received PD-1 inhibitor therapy (toripalimab) was collected from multiple centers, details of which were previously published [48].

This study was approved by the medical ethics committee of the Peking University Cancer Hospital, and participants provided informed consent. The design and implementation of the study complied with the local regulations and guidelines and with the basic principles of the Declaration of Helsinki.

### **Whole-exome sequencing analysis and variant filtering**

Tumor tissue and adjacent nontumor tissue samples from CGC cohort patients were subjected to whole-exome sequencing. All whole-genome sequencing (including DNA extraction and quality control) was performed in the Origimed laboratory. This laboratory is College of American Pathologists-accredited and is a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory. All tumor tissue slides were reviewed by two independent pathologists, and samples with estimated tumor purity greater than 20% were included in the study. In detail, DNA was extracted from the formalin-fixed, paraffin-embedded samples according to the manufacturer's instructions. Next, ~500 ng of genomic DNA was sheared to a mean fragment length of 200 bp and labeled with a 6–8 base barcode during polymerase chain reaction (PCR) amplification. Exomes were captured using a SureSelectXT Human All Exon V6 (Agilent Technologies). Sequencing was performed with an Illumina HiSeqX instrument using 150 base paired-end reads. Reads were trimmed with AdapterRemoval v2

[49], aligned to the human reference genome (GRCh37) by Burrows-Wheeler Aligner v0.7.5a [50], and PCR duplicates were removed by Picard (version 1.47). Somatic variants were identified by GATK4 Mutect2 (version 4.0.6) [51] and then annotated with SnpEff (version 4.3b) [52]. Variants in the common dbSNP database (version 147) or those having a frequency above 1.5% in the Exome Sequencing Project 6500 or 1000 genome project were excluded from further consideration. The variant files have been uploaded to the European Variation Archive (PRJEB31906). Low-confidence variants in both the TCGA and CGC cohorts were removed by applying the following filters: (1) total coverage < 30, (2) variant allele depth < 7, and (3) variant allele frequency < 0.05.

### **Feature selection and biomarker validation**

The prognostic biomarker was calculated by a genetic algorithm that was implemented in the Python package *pyeasyga* [50, 53]. In brief, each solution was a binary vector with a 0 or 1 value that had the same length as the original full set of the “regulation of RhoA activity pathway” biomarker (length = 48); 1 indicated that the corresponding gene harbored a nonsynonymous substitution or indel in the coding region, and 0 indicated neither of these changes. The goal was to find the best solution containing the genes with their values equal to 1, which was considered as the optimal gene subset (i.e., biomarker). The parameters of the genetic algorithm were empirically set as follows: a total population size of 40 solutions, crossover probability of 0.1, mutation probability of 0.01, and max generation of 50 000. The fitness value was calculated as follows: a cohort that was randomly selected (size = 80, 50%) from the 160 patients of the TCGA training cohort was further divided into a mutation and wild-type group according to the mutation status of the genes that corresponded to value 1 of each solution; a log-rank p-value was calculated for this division, the whole process was performed 100 times, and the mean p-values were calculated as the fitness value. Finally, the best solution was generated, and the resulting genes with a value of 1 were regarded as the final subset (i.e., the biomarker).

### **Tumor mutational burden, neoantigens, and mutational signatures**

The tumor mutational burden (TMB) was estimated by dividing the total number of somatic variants by the coding region size. All somatic variants in the coding and splicing regions were counted, and the coding region size was estimated as 33 Mb based on RefSeq Genome Annotation (GRCh37). HLA typing was performed with OptiType (version 1.2.1) [54]. All nonsynonymous mutations that were identified were

translated into peptides of 17 amino acids with an in-house pipeline. The sliding window (approximately 9-11 amino acids in size) method was used to identify substrings within the 17mer that had a predicted HLA class I binding affinity of less than 500 nM to any patient-specific HLA allele. The binding affinity for amino acids and alleles was analyzed using NetMHCpan v3.0 [55]. The neoantigen burden was estimated as the total number of substrings with a binding affinity less than 500 nM. Mutational signature contributions were identified using the R package *deconstrucSigs* (version 1.8.0) [56] with 30 signatures documented by the Catalogue of Somatic Mutations in Cancer as a reference [57]; samples with an error rate less than 0.15 and mutation counts greater than 30 were considered for mutational signature analysis. Deficient mismatch repair was estimated as the sum of signature 6, 15, 20, and 26 [57].

### RNA-seq data analysis

RNA-seq data of the TCGA cohort were extracted with TCGAAbiolinks [58]. Raw read counts were normalized with DESeq2 [59] and then log<sub>2</sub> transformed. The ssGSEA algorithm in the R package GSVA [60] was used to evaluate the pathway enrichment score. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway gene sets were retrieved from the KEGG database by the R package gage [61].

### Immunohistochemistry staining and evaluation

Immunohistochemistry (IHC) analysis for CD8 (clone SP16, ZSGB-BIO) was evaluated within intratumor areas. Aperio Scanscope (Aperio Technologies Vista, CA, USA) was used to quantify CD8<sup>+</sup> density by the rare event tissue test method. We counted the total number of CD8<sup>+</sup> cells of each area based on 6 randomly captured visual fields (400×400 μm<sup>2</sup>) and defined the density of CD8<sup>+</sup> TILs as the total cell number per square millimeter. IHC staining of anti-PD-L1 (clone SP142, Spring Bioscience) was also annotated within intratumor areas. The percentages of cancer cells and immune cells with anti-PD-L1 staining were measured in each area based on 3 visual fields in darkly stained areas (400×400 μm<sup>2</sup>). The expression of PD-L1 was defined as positive when ≥ 1% of the tumor/stromal cells were positive. IHC-stained sections were scored independently by 2 gastrointestinal pathologists blinded to the clinicopathological parameters and biomarker results.

### Mismatch repair/microsatellite instability testing

To evaluate the mismatch repair (MMR) and/or microsatellite instability (MSI) status, MLH1 (clone

ES05, Gene Tech), MSH2 (clone 25D12, Gene Tech), MSH6 (clone EP49, Gene Tech) and PMS2 (clone EP51, Gene Tech) were stained. The complete loss of expression of one or more proteins was considered as dMMR. In some cases, the microsatellite stability status was calculated by using a single multiplex PCR that assessed five microsatellite loci (BAT-25, BAT-26, D2S123, D5S346, and D17S250) [62]. For interpretation, instability at more than one locus referred to high microsatellite instability (MSI-H), instability at a single locus referred to low microsatellite instability (MSI-L), and no instability at any locus referred to stable microsatellites (MSS) [63].

### Statistical analysis

Survival analysis was carried out by the Kaplan-Meier method, and the difference between the groups was compared using the log-rank test. The hazard ratio and the 95% confidence interval were estimated by the Cox model. The multivariate Cox model was performed to adjust for confounding factors. Fisher's exact test was used to compare proportions. The Wilcoxon rank-sum test was used to compare continuous values. All statistical analyses were performed with R software 3.5.3.

### Abbreviations

CGC: Chinese gastric cancer; HLA: human leukocyte antigen; ICB: immune checkpoint blockade; IHC: immunohistochemistry; MMR: mismatch repair; MSI: microsatellite instability; MSS: microsatellite stable; NAB: neoantigen burden; OS: overall survival; PFS: progression free survival; RRAP: regulation of RhoA activity panel; TCGA: The Cancer Genome Atlas; TILs: infiltrating lymphocytes; TMB: tumor mutational burden; TME: tumor microenvironment.

### AUTHOR CONTRIBUTIONS

LS, WS and JG conceived and designed the study. WH and SZ performed the experiments and analyzed the data. CZ, ZL and SG performed the sample preparation and the pathological evaluation. BL, HF, and KW contributed the reagents, materials, and analysis tools. RX and JJ involved in the sample collection and provided the diagnostic data. WH, SZ and WS wrote the manuscript. All authors read and approved the final manuscript.

### ACKNOWLEDGMENTS

We thank all the patients who participated in this study, their families; the clinicians and the staff of Peking University Cancer Hospital and Institute, for data management and logistic support.

## CONFLICTS OF INTEREST

The authors have declared that no conflicts of interest exists.

## FUNDING

This research was supported by funding from the National Natural Science Foundation of China (No. 91959205, and No. 81872341), and the National Key Research and Development Program of China (No. 2017YFC1308900).

## REFERENCES

1. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, Hancock J, et al, and Global Burden of Disease Cancer Collaboration. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA Oncol.* 2017; 3:524–48.  
<https://doi.org/10.1001/jamaoncol.2016.5688>  
PMID:[27918777](https://pubmed.ncbi.nlm.nih.gov/27918777/)
2. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand.* 1965; 64:31–49.  
<https://doi.org/10.1111/apm.1965.64.1.31>  
PMID:[14320675](https://pubmed.ncbi.nlm.nih.gov/14320675/)
3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014; 513:202–09.  
<https://doi.org/10.1038/nature13480>  
PMID:[25079317](https://pubmed.ncbi.nlm.nih.gov/25079317/)
4. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med.* 2015; 21:449–56.  
<https://doi.org/10.1038/nm.3850>  
PMID:[25894828](https://pubmed.ncbi.nlm.nih.gov/25894828/)
5. Mitchison TJ, Cramer LP. Actin-based cell motility and cell locomotion. *Cell.* 1996; 84:371–79.  
[https://doi.org/10.1016/s0092-8674\(00\)81281-7](https://doi.org/10.1016/s0092-8674(00)81281-7)  
PMID:[8608590](https://pubmed.ncbi.nlm.nih.gov/8608590/)
6. Lauffenburger DA, Horwitz AF. Cell migration: a physically integrated molecular process. *Cell.* 1996; 84:359–69.  
[https://doi.org/10.1016/s0092-8674\(00\)81280-5](https://doi.org/10.1016/s0092-8674(00)81280-5)  
PMID:[8608589](https://pubmed.ncbi.nlm.nih.gov/8608589/)
7. Olson MF, Paterson HF, Marshall CJ. Signals from ras and rho GTPases interact to regulate expression of p21Waf1/Cip1. *Nature.* 1998; 394:295–99.  
<https://doi.org/10.1038/28425> PMID:[9685162](https://pubmed.ncbi.nlm.nih.gov/9685162/)
8. Kalpana G, Figy C, Yeung M, Yeung KC. Reduced RhoA expression enhances breast cancer metastasis with a concomitant increase in CCR5 and CXCR4 chemokines signaling. *Sci Rep.* 2019; 9:16351.  
<https://doi.org/10.1038/s41598-019-52746-w>  
PMID:[31705019](https://pubmed.ncbi.nlm.nih.gov/31705019/)
9. Nam GH, Lee EJ, Kim YK, Hong Y, Choi Y, Ryu MJ, Woo J, Cho Y, Ahn DJ, Yang Y, Kwon IC, Park SY, Kim IS. Combined rho-kinase inhibition and immunogenic cell death triggers and propagates immunity against cancer. *Nat Commun.* 2018; 9:2165.  
<https://doi.org/10.1038/s41467-018-04607-9>  
PMID:[29867097](https://pubmed.ncbi.nlm.nih.gov/29867097/)
10. Huang KH, Lan YT, Chen MH, Chao Y, Lo SS, Li AF, Wu CW, Chiou SH, Yang MH, Shyr YM, Fang WL. The correlation between RhoA expression and clinicopathological characteristics in gastric cancer patients after curative surgery. *World J Surg.* 2015; 39:2289–99.  
<https://doi.org/10.1007/s00268-015-3095-4>  
PMID:[26013205](https://pubmed.ncbi.nlm.nih.gov/26013205/)
11. Ushiku T, Ishikawa S, Kakiuchi M, Tanaka A, Katoh H, Aburatani H, Lauwers GY, Fukayama M. RHOA mutation in diffuse-type gastric cancer: a comparative clinicopathology analysis of 87 cases. *Gastric Cancer.* 2016; 19:403–11.  
<https://doi.org/10.1007/s10120-015-0493-0>  
PMID:[25823974](https://pubmed.ncbi.nlm.nih.gov/25823974/)
12. Röcken C, Behrens HM, Böger C, Krüger S. Clinicopathological characteristics of RHOA mutations in a Central European gastric cancer cohort. *J Clin Pathol.* 2016; 69:70–75.  
<https://doi.org/10.1136/jclinpath-2015-202980>  
PMID:[26251521](https://pubmed.ncbi.nlm.nih.gov/26251521/)
13. Shi W, Jiang T, Nuciforo P, Hatzis C, Holmes E, Harbeck N, Sotiriou C, Peña L, Loi S, Rosa DD, Chia S, Wardley A, Ueno T, et al. Pathway level alterations rather than mutations in single genes predict response to HER2-targeted therapies in the neo-ALTTO trial. *Ann Oncol.* 2017; 28:128–35.  
<https://doi.org/10.1093/annonc/mdw434>  
PMID:[28177460](https://pubmed.ncbi.nlm.nih.gov/28177460/)
14. Sahai E, Marshall CJ. RHO-GTPases and cancer. *Nat Rev Cancer.* 2002; 2:133–42.  
<https://doi.org/10.1038/nrc725> PMID:[12635176](https://pubmed.ncbi.nlm.nih.gov/12635176/)
15. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, Siu HC, Deng S, Chu KM, Law S, Chan KH, Chan AS, Tsui WY, et al. Whole-genome sequencing and comprehensive

- molecular profiling identify new driver mutations in gastric cancer. *Nat Genet.* 2014; 46:573–82.  
<https://doi.org/10.1038/ng.2983>  
PMID:24816253
16. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, Watanabe Y, Ichimura T, Ushiku T, Funahashi S, et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet.* 2014; 46:583–87.  
<https://doi.org/10.1038/ng.2984> PMID:24816255
  17. Xu XT, Song QB, Yao Y, Ruan P, Tao ZZ. Inhibition of RhoA/ROCK signaling pathway promotes the apoptosis of gastric cancer cells. *Hepatogastroenterology.* 2012; 59:2523–26.  
<https://doi.org/10.5754/hge12147> PMID:22584506
  18. Murray D, Horgan G, Macmathuna P, Doran P. NET1-mediated RhoA activation facilitates lysophosphatidic acid-induced cell migration and invasion in gastric cancer. *Br J Cancer.* 2008; 99:1322–29.  
<https://doi.org/10.1038/sj.bjc.6604688>  
PMID:18827818
  19. Leyden J, Murray D, Moss A, Arumuguma M, Doyle E, McEntee G, O’Keane C, Doran P, MacMathuna P. Net1 and myeov: computationally identified mediators of gastric cancer. *Br J Cancer.* 2006; 94:1204–12.  
<https://doi.org/10.1038/sj.bjc.6603054>  
PMID:16552434
  20. Lin KY, Wang LH, Hseu YC, Fang CL, Yang HL, Kumar KJ, Tai C, Uen YH. Clinical significance of increased guanine nucleotide exchange factor Vav3 expression in human gastric cancer. *Mol Cancer Res.* 2012; 10:750–59.  
<https://doi.org/10.1158/1541-7786.MCR-11-0598-T>  
PMID:22544459
  21. Zhang H, Schaefer A, Wang Y, Hodge RG, Blake DR, Diehl JN, Papageorge AG, Stachler MD, Liao J, Zhou J, Wu Z, Akarca FG, de Klerk LK, et al. Gain-of-function RHOA mutations promote focal adhesion kinase activation and dependency in diffuse gastric cancer. *Cancer Discov.* 2020; 10:288–305.  
<https://doi.org/10.1158/2159-8290.CD-19-0811>  
PMID:31771969
  22. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer.* 2019; 19:133–50.  
<https://doi.org/10.1038/s41568-019-0116-x>  
PMID:30755690
  23. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, et al. The Immune Landscape of Cancer. *Immunity.* 2018; 48:812–830.e14.  
<https://doi.org/10.1016/j.immuni.2018.03.023>  
PMID:29628290
  24. Pagès F, Berger A, Camus M, Sanchez-Cabo F, Costes A, Molitor R, Mlecnik B, Kirilovsky A, Nilsson M, Damotte D, Meatchi T, Bruneval P, Cugnenc PH, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med.* 2005; 353:2654–66.  
<https://doi.org/10.1056/NEJMoa051424>  
PMID:16371631
  25. Ohri CM, Shikotra A, Green RH, Waller DA, Bradding P. Macrophages within NSCLC tumour islets are predominantly of a cytotoxic M1 phenotype associated with extended survival. *Eur Respir J.* 2009; 33:118–26.  
<https://doi.org/10.1183/09031936.00065708>  
PMID:19118225
  26. Martínez-Lostao L, Anel A, Pardo J. How do cytotoxic lymphocytes kill cancer cells? *Clin Cancer Res.* 2015; 21:5047–56.  
<https://doi.org/10.1158/1078-0432.CCR-15-0685>  
PMID:26567364
  27. Trapani JA, Smyth MJ. Functional significance of the perforin/granzyme cell death pathway. *Nat Rev Immunol.* 2002; 2:735–47.  
<https://doi.org/10.1038/nri911> PMID:12360212
  28. Green DR, Ferguson TA. The role of fas ligand in immune privilege. *Nat Rev Mol Cell Biol.* 2001; 2:917–24.  
<https://doi.org/10.1038/35103104>  
PMID:11733771
  29. Narumiya S, Tanji M, Ishizaki T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. *Cancer Metastasis Rev.* 2009; 28:65–76.  
<https://doi.org/10.1007/s10555-008-9170-7>  
PMID:19160018
  30. Yoon C, Cho SJ, Aksoy BA, Park DJ, Schultz N, Ryeom SW, Yoon SS. Chemotherapy resistance in diffuse-type gastric adenocarcinoma is mediated by RhoA activation in cancer stem-like cells. *Clin Cancer Res.* 2016; 22:971–83.  
<https://doi.org/10.1158/1078-0432.CCR-15-1356>  
PMID:26482039
  31. Chang KK, Cho SJ, Yoon C, Lee JH, Park DJ, Yoon SS. Increased RhoA activity predicts worse overall survival in patients undergoing surgical resection for lauren diffuse-type gastric adenocarcinoma. *Ann Surg Oncol.* 2016; 23:4238–46.  
<https://doi.org/10.1245/s10434-016-5357-2>  
PMID:27364501
  32. Harding CV, Geuze HJ. Class II MHC molecules are present in macrophage lysosomes and phagolysosomes that function in the phagocytic processing of listeria monocytogenes for presentation to T cells. *J Cell Biol.* 1992; 119:531–42.  
<https://doi.org/10.1083/jcb.119.3.531>  
PMID:1400590

33. König R, Huang LY, Germain RN. MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. *Nature*. 1992; 356:796–98.  
<https://doi.org/10.1038/356796a0> PMID:1574118
34. Bros M, Haas K, Moll L, Grabbe S. RhoA as a key regulator of innate and adaptive immunity. *Cells*. 2019; 8:733.  
<https://doi.org/10.3390/cells8070733> PMID:31319592
35. Al Absi A, Wurzer H, Guerin C, Hoffmann C, Moreau F, Mao X, Brown-Clay J, Petrolli R, Casellas CP, Dieterle M, Thiery JP, Chouaib S, Berchem G, et al. Actin cytoskeleton remodeling drives breast cancer cell escape from natural killer-mediated cytotoxicity. *Cancer Res*. 2018; 78:5631–43.  
<https://doi.org/10.1158/0008-5472.CAN-18-0441> PMID:30104240
36. Rodriguez RM, Suarez-Alvarez B, Lavín JL, Mosén-Ansorena D, Baragaño Raneros A, Márquez-Kisinousky L, Aransay AM, Lopez-Larrea C. Epigenetic networks regulate the transcriptional program in memory and terminally differentiated CD8<sup>+</sup> T cells. *J Immunol*. 2017; 198:937–49.  
<https://doi.org/10.4049/jimmunol.1601102> PMID:27974453
37. Nakaya M, Tanaka M, Okabe Y, Hanayama R, Nagata S. Opposite effects of rho family GTPases on engulfment of apoptotic cells by macrophages. *J Biol Chem*. 2006; 281:8836–42.  
<https://doi.org/10.1074/jbc.M510972200> PMID:16439364
38. Kim SY, Kim S, Bae DJ, Park SY, Lee GY, Park GM, Kim IS. Coordinated balance of Rac1 and RhoA plays key roles in determining phagocytic appetite. *PLoS One*. 2017; 12:e0174603.  
<https://doi.org/10.1371/journal.pone.0174603> PMID:28376111
39. Salem ME, Puccini A, Grothey A, Raghavan D, Goldberg RM, Xiu J, Korn WM, Weinberg BA, Hwang JJ, Shields AF, Marshall JL, Philip PA, Lenz HJ. Landscape of tumor mutation load, mismatch repair deficiency, and PD-L1 expression in a large patient cohort of gastrointestinal cancers. *Mol Cancer Res*. 2018; 16:805–12.  
<https://doi.org/10.1158/1541-7786.MCR-17-0735> PMID:29523759
40. Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, Sun W, Jalal SI, Shah MA, Metges JP, Garrido M, Golan T, Mandala M, et al. Safety and efficacy of pembrolizumab monotherapy in patients with previously treated advanced gastric and gastroesophageal junction cancer: phase 2 clinical KEYNOTE-059 trial. *JAMA Oncol*. 2018; 4:e180013.  
<https://doi.org/10.1001/jamaoncol.2018.0013> PMID:29543932
41. Marabelle A, Le DT, Ascierto PA, Di Giacomo AM, De Jesus-Acosta A, Delord JP, Geva R, Gottfried M, Penel N, Hansen AR, Piha-Paul SA, Doi T, Gao B, et al. Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J Clin Oncol*. 2020; 38:1–10.  
<https://doi.org/10.1200/JCO.19.02105> PMID:31682550
42. Zhu J, Powis de Tenbossche CG, Cané S, Colau D, van Baren N, Lurquin C, Schmitt-Verhulst AM, Liljeström P, Uyttenhove C, Van den Eynde BJ. Resistance to cancer immunotherapy mediated by apoptosis of tumor-infiltrating lymphocytes. *Nat Commun*. 2017; 8:1404.  
<https://doi.org/10.1038/s41467-017-00784-1> PMID:29123081
43. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, et al, and KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016; 375:1823–33.  
<https://doi.org/10.1056/NEJMoa1606774> PMID:27718847
44. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, Moreira AL, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015; 348:124–28.  
<https://doi.org/10.1126/science.aaa1348> PMID:25765070
45. Shitara K, Özgüroğlu M, Bang YJ, Di Bartolomeo M, Mandalà M, Ryu MH, Fornaro L, Olesiński T, Caglevic C, Chung HC, Muro K, Goekkurt E, Mansoor W, et al, and KEYNOTE-061 investigators. Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2018; 392:123–33.  
[https://doi.org/10.1016/S0140-6736\(18\)31257-1](https://doi.org/10.1016/S0140-6736(18)31257-1) PMID:29880231
46. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401–04.  
<https://doi.org/10.1158/2159-8290.CD-12-0095> PMID:22588877
47. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV,

- Omberg L, Wolf DM, Shriver CD, et al, and Cancer Genome Atlas Research Network. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018; 173:400–16.e11.  
<https://doi.org/10.1016/j.cell.2018.02.052>  
PMID:[29625055](https://pubmed.ncbi.nlm.nih.gov/29625055/)
48. Wang F, Wei XL, Wang FH, Xu N, Shen L, Dai GH, Yuan XL, Chen Y, Yang SJ, Shi JH, Hu XC, Lin XY, Zhang QY, et al. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase Ib/II clinical trial NCT02915432. *Ann Oncol*. 2019; 30:1479–86.  
<https://doi.org/10.1093/annonc/mdz197>  
PMID:[31236579](https://pubmed.ncbi.nlm.nih.gov/31236579/)
  49. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016; 9:88.  
<https://doi.org/10.1186/s13104-016-1900-2>  
PMID:[26868221](https://pubmed.ncbi.nlm.nih.gov/26868221/)
  50. Gould Rothberg BE, Berger AJ, Molinaro AM, Subtil A, Krauthammer MO, Camp RL, Bradley WR, Ariyan S, Kluger HM, Rimm DL. Melanoma prognostic model using tissue microarrays and genetic algorithms. *J Clin Oncol*. 2009; 27:5772–80.  
<https://doi.org/10.1200/JCO.2009.22.8239>  
PMID:[19884546](https://pubmed.ncbi.nlm.nih.gov/19884546/)
  51. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013; 31:213–19.  
<https://doi.org/10.1038/nbt.2514> PMID:[23396013](https://pubmed.ncbi.nlm.nih.gov/23396013/)
  52. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012; 6:80–92.  
<https://doi.org/10.4161/fly.19695> PMID:[22728672](https://pubmed.ncbi.nlm.nih.gov/22728672/)
  53. Dolled-Filhart M, Rydén L, Cregger M, Jirström K, Harigopal M, Camp RL, Rimm DL. Classification of breast cancer using genetic algorithms and tissue microarrays. *Clin Cancer Res*. 2006; 12:6459–68.  
<https://doi.org/10.1158/1078-0432.CCR-06-1383>  
PMID:[17085660](https://pubmed.ncbi.nlm.nih.gov/17085660/)
  54. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics*. 2014; 30:3310–16.  
<https://doi.org/10.1093/bioinformatics/btu548>  
PMID:[25143287](https://pubmed.ncbi.nlm.nih.gov/25143287/)
  55. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med*. 2016; 8:33.  
<https://doi.org/10.1186/s13073-016-0288-x>  
PMID:[27029192](https://pubmed.ncbi.nlm.nih.gov/27029192/)
  56. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016; 17:31.  
<https://doi.org/10.1186/s13059-016-0893-4>  
PMID:[26899170](https://pubmed.ncbi.nlm.nih.gov/26899170/)
  57. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, Boyault S, Burkhardt B, Butler AP, et al, and Australian Pancreatic Cancer Genome Initiative, and ICGC Breast Cancer Consortium, and ICGC MMLL-Seq Consortium, and ICGC PedBrain. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415–21.  
<https://doi.org/10.1038/nature12477>  
PMID:[23945592](https://pubmed.ncbi.nlm.nih.gov/23945592/)
  58. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44:e71.  
<https://doi.org/10.1093/nar/gkv1507>  
PMID:[26704973](https://pubmed.ncbi.nlm.nih.gov/26704973/)
  59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15:550.  
<https://doi.org/10.1186/s13059-014-0550-8>  
PMID:[25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
  60. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013; 14:7.  
<https://doi.org/10.1186/1471-2105-14-7>  
PMID:[23323831](https://pubmed.ncbi.nlm.nih.gov/23323831/)
  61. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009; 10:161.  
<https://doi.org/10.1186/1471-2105-10-161>  
PMID:[19473525](https://pubmed.ncbi.nlm.nih.gov/19473525/)
  62. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the

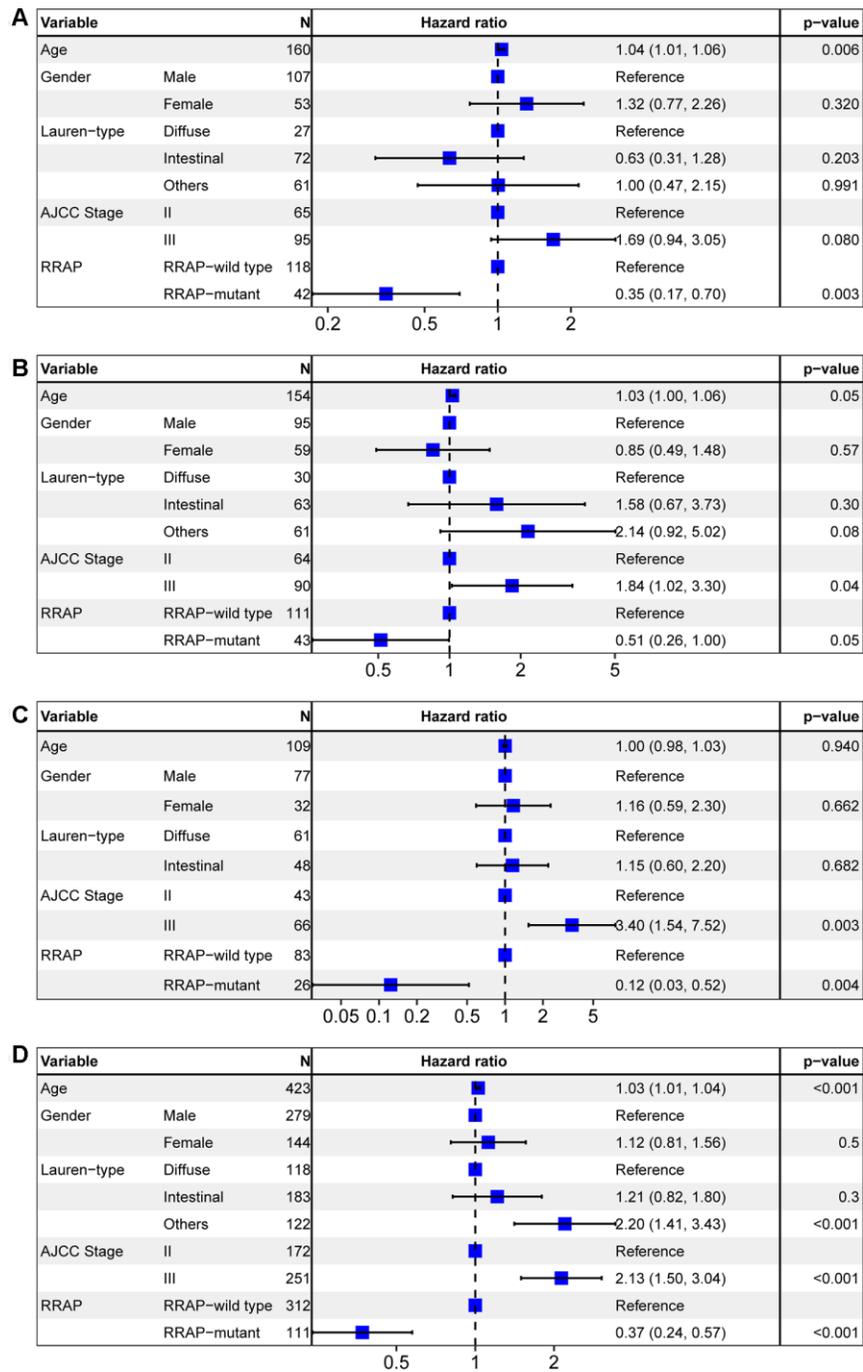
determination of microsatellite instability in colorectal cancer. *Cancer Res.* 1998; 58:5248–57.  
PMID:[9823339](https://pubmed.ncbi.nlm.nih.gov/9823339/)

63. Jang M, Kwon Y, Kim H, Kim H, Min BS, Park Y, Kim TI, Hong SP, Kim WK. Microsatellite instability test using

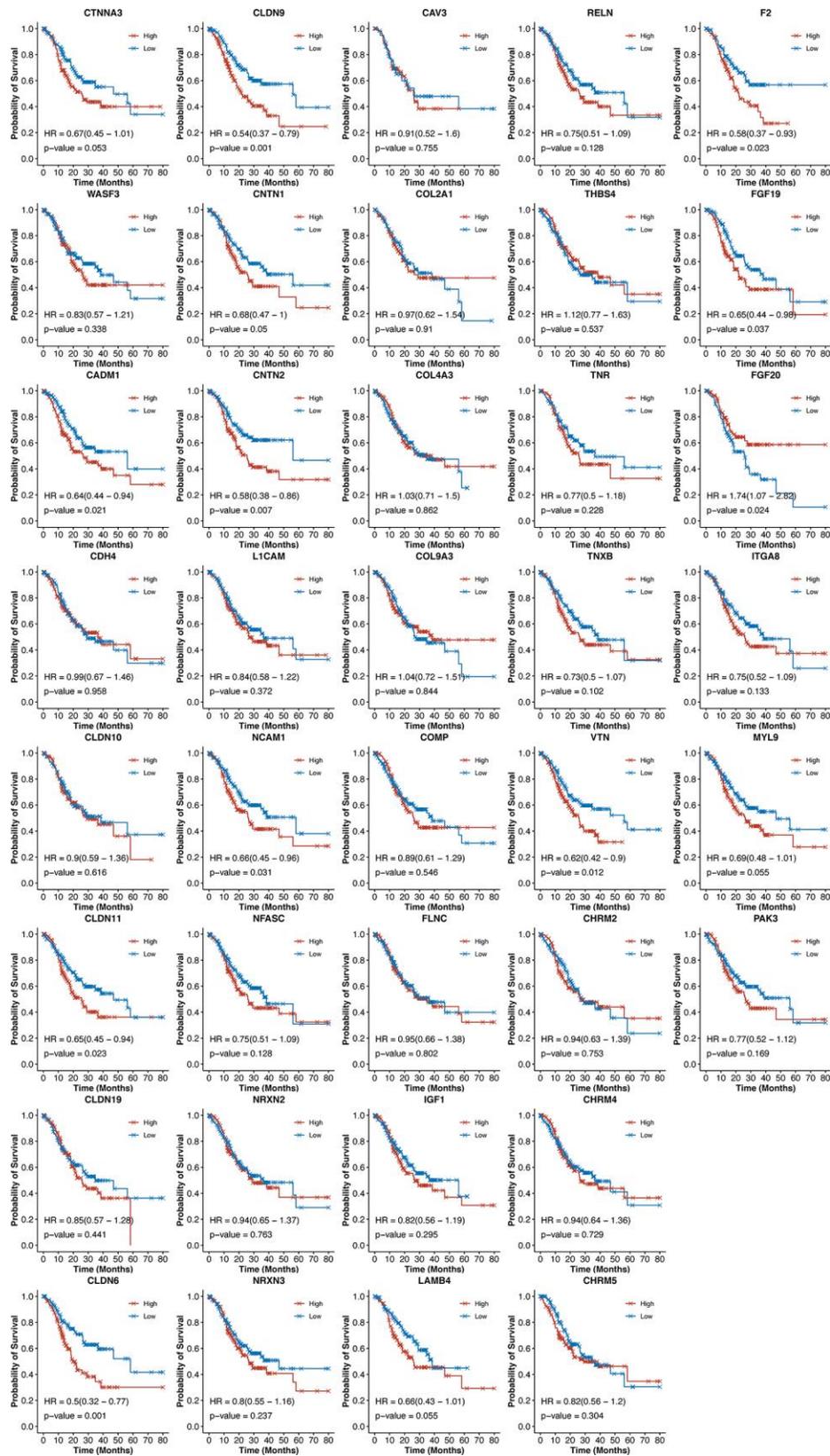
peptide nucleic acid probe-mediated melting point analysis: a comparison study. *BMC Cancer.* 2018; 18:1218.  
<https://doi.org/10.1186/s12885-018-5127-6>  
PMID:[30514259](https://pubmed.ncbi.nlm.nih.gov/30514259/)

SUPPLEMENTARY MATERIALS

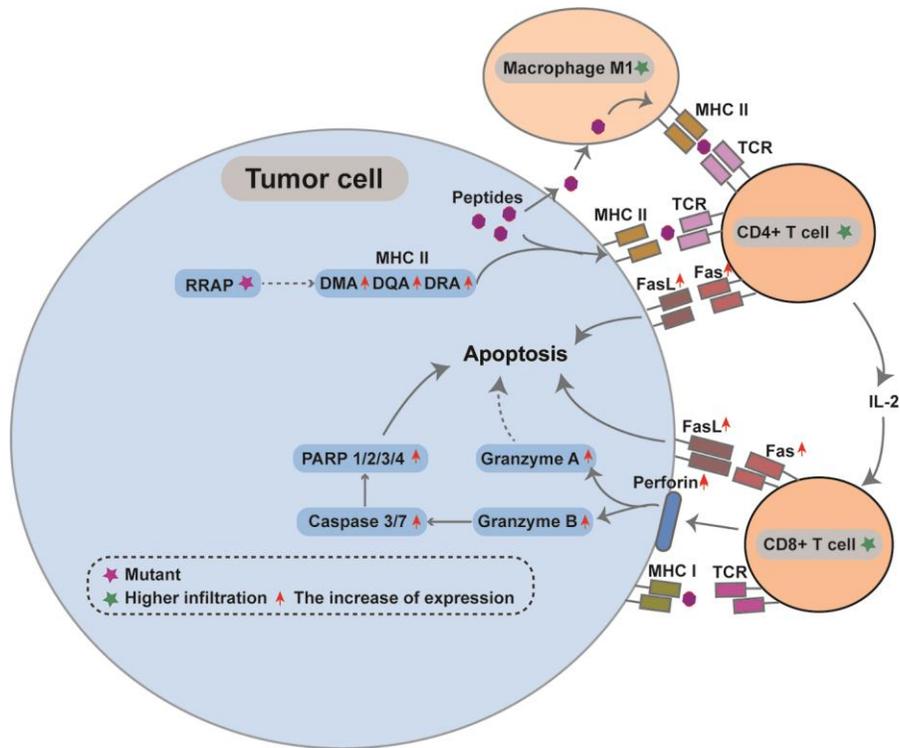
Supplementary Figures



**Supplementary Figure 1.** Multivariate Cox model analysis, with age, gender, Lauren type, AJCC stage, and RRAP as confounding factors, performed on the TCGA training cohort (A), TCGA validation cohort (B), CGC cohort (C), and the overall cohort containing merged data from the three cohorts (D).



Supplementary Figure 2. Kaplan-Meier overall survival curves for patients with high and low expression of 39 differentially expressed genes of the 4 migration-related functions. P-values were estimated with the log-rank test, and hazard ratios were estimated with the Cox model.



Supplementary Figure 3. Diagram of the interaction of immune cells and tumor cells in the RRAP-mutant tumor microenvironment.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1.

**Supplementary Table 1. Genes contained in the RRAP biomarker gene set and their mutation frequency in the TCGA training, TCGA validation, and CGC cohorts, along with migration activity.**

**Supplementary Table 2. The CLDN11, CLDN6, CLDN9, VTN, and F2 with cell migration activity.**

Symbol	Conclusion	Experiment model	PMID
CLDN6	Higher <b>CLDN6</b> expression promotes the cell proliferation and migration ability of gastric cancer. Increased expression of <b>CLDN6</b> predicts poor prognosis in gastric cancer patients.	MKN28, AGS, MKN7, NUGC-3, xenograft subcutaneous tumor model	31827075, 20874001, 27914788, 31654186
CLDN9	<b>CLDN9</b> overexpression in gastric adenocarcinoma cells has been reported to increase cell migration and proliferation. The expression of <b>CLDN9</b> in gastric cancer correlates to poor prognosis. Higher <b>CLDN9</b> expression promotes invasiveness of several solid tumors in vitro and metastasis in vivo.	AGS, HL7702, a spontaneous metastasis model	20874001, 29031421, 31418417, 26669782
CLDN11	Silencing of <b>CLDN11</b> is associated with increased invasiveness, proliferation, and migration of gastric cancer cells. DNA hypermethylation is associated with the downregulation of <b>CLDN11</b> in gastric cancer cells.	MKN28, MKN74, AGS, HFE145	32119960, 28962204, 19956721
F2	<b>F2</b> is generated in the TME, promoting the migration and metastasis of tumor cells.	B16, HT29-D4, SUM149, MC38 cells	30654498, 15539922, 12707033, 23280128
VTN	Decreased expression of <b>VTN</b> promotes gastric cancer cell growth and metastasis. <b>VTN</b> is related to the migration and invasion of other solid tumors, such as ovarian adenocarcinoma, fibrosarcoma and nasopharyngeal carcinoma.	AGS, IGROV1, HT-1080,5-8F, HNE2, HONE1 cells	25789040, 9935212, 16052409, 29123267

**Supplementary Table 3. Clinical characteristics in TCGA-training, TCGA-validation, and CGC cohorts.<sup>a</sup>**

<b>Characteristic</b>	<b>TCGA- training (N=160)</b>	<b>TCGA-validation (N=155)</b>	<b>CGC (N=109)</b>
Age, years			
Mean (SD)	64.3 (10.8)	66.2 (9.9)	57.9 (13.0)
Median (Q1 – Q3)	66 (57-72)	68 (58-73)	58 (51-67)
Min – Max	30-84	43-90	25-83
Gender - no. (%)			
Male	107 (66.9%)	96 (61.9%)	77 (70.6%)
Female	53 (33.1%)	59 (38.1%)	32 (29.4%)
AJCC Stage - no. (%)			
II	65 (40.6%)	65 (41.9%)	43 (39.4%)
III	95 (59.4%)	90 (58.1%)	66 (60.6%)
Lauren classification - no. (%)			
Diffuse	27 (16.9%)	30 (19.4%)	61 (56.0%)
Intestinal	72 (45.0%)	64 (41.3%)	48 (44.0%)
Unknown	61 (38.1%)	61 (39.3%)	—
Tumor location - no. (%)			
Gastroesophageal junction	33 (21.0%)	31 (20.0%)	34 (31.2%)
Non-gastroesophageal junction	123 (76.9%)	121 (78.1%)	75 (68.8%)
Unknown	4 (3.0%)	3 (2.0%)	—
Adjuvant chemotherapy - no. (%)			
Yes	81 (50.6%)	70 (45.2%)	98 (90.0%)
No	—	—	8 (7.3%)
Unknown	79 (49.4%)	85 (54.8%)	3 (2.7%)

<sup>a</sup>Because of rounding, percentages may not total 100.

# KRAS mutations are negatively correlated with immunity in colon cancer

Xiaorui Fu<sup>1,2,\*</sup>, Xinyi Wang<sup>1,2,\*</sup>, Jinzhong Duanmu<sup>1</sup>, Taiyuan Li<sup>1</sup>, Qunguang Jiang<sup>1</sup>

<sup>1</sup>Department of Gastrointestinal Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, People's Republic of China

<sup>2</sup>Queen Mary College, Medical Department, Nanchang University, Nanchang, Jiangxi, People's Republic of China

\*Equal contribution

**Correspondence to:** Qunguang Jiang; **email:** [ndyfy03699@ncu.edu.cn](mailto:ndyfy03699@ncu.edu.cn)

**Keywords:** KRAS mutations, immunity, colon cancer, tumor-infiltrating immune cells, inflammation

**Received:** March 27, 2020

**Accepted:** October 8, 2020

**Published:** November 26, 2020

**Copyright:** © 2020 Fu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

The heterogeneity of colon cancer tumors suggests that therapeutics targeting specific molecules may be effective in only a few patients. It is therefore necessary to explore gene mutations in colon cancer. In this study, we obtained colon cancer samples from The Cancer Genome Atlas, and the International Cancer Genome Consortium. We evaluated the landscape of somatic mutations in colon cancer and found that KRAS mutations, particularly rs121913529, were frequent and had prognostic value. Using ESTIMATE analysis, we observed that the KRAS-mutated group had higher tumor purity, lower immune score, and lower stromal score than the wild-type group. Through single-sample Gene Set Enrichment Analysis and Gene Set Enrichment Analysis, we found that KRAS mutations negatively correlated with enrichment levels of tumor infiltrating lymphocytes, inflammation, and cytolytic activities. HLA gene expression and checkpoint-related genes were also lower in the KRAS-mutated group. Finally, we found 24 immune-related genes that differed in expression between the KRAS-mutated and wild-type samples, which may provide clues to the mechanism of KRAS-related immune alteration. Our findings are indicative of the prognostic and predictive value of KRAS and illustrate the relationship between KRAS mutations and immune activity in colon cancer.

## INTRODUCTION

Colon cancer is the third leading cause of cancer deaths, with more than 1 million new cases diagnosed every year [1]. Heterogeneity is a characteristic of colon cancer whereby the pattern of mutations differ significantly among patients [2]. Mutations in essential genes can affect the proliferation, differentiation, apoptosis, survival capacity, and distant metastasis of tumor cells [3]. Thus, therapeutic methods that target specific biomolecules or genes are effective in a small fraction of patients. It is necessary to explore gene mutations and more potential therapeutic targets for colon cancer. *APC*, a tumor suppressor gene, is the most frequently mutated gene in patients with colon cancer and influences the Wnt/ $\beta$ -catenin pathway [4]. Mutated

*APC* has been observed in early stage colon cancer and is correlated with clinical outcomes [5]. However, it was seldom detected in patients with late stage colon cancer and metastasis. In contrast, the *TP53* inactivating mutation is usually observed in more advanced tumors [6, 7]. Presently, RAS is the only predictive biomarker in the application of anti-EGFR agents to treat wild-type colon cancer [8, 9]. *KRAS* encodes a p21 protein, which couples with GTPase to transform GTP into GDP and regulates signaling pathways related to cellular growth and survival. When *KRAS* is mutated, the downstream signaling pathway (mitogen-activated protein kinase, MAPK) is activated, leading to cellular proliferation and tumor progression. In addition, *KRAS* mutations are predictive markers for breast, lung, ovarian, head/neck, and pancreatic cancers [4, 8, 9, 11]. For example, Jung

et al. found that *KRAS* mutations were correlated with poor prognosis in patients with breast cancer, together with AKT signaling pathway activation, estrogen negative, and basal-like gene expression patterns. As BRAF is downstream of RAS in the MAPK/ERK signaling pathway, mutated BRAF is assumed to have the same resistance to the anti-EGFR agent as to the RAS-mutated colon tumor [10]. Furthermore, microsatellite instability (MSI) is another pathogenesis factor, and, if detected at an early stage, improves patient outcome [11, 12]. However, the mechanism of these gene mutations is unclear and personalized treatment requires further research on clinical biomarkers.

The minority of human colon cancers are genetically driven, including Lynch syndrome, familial adenomatous, and hamartomata's polyposis [13]. The majority of colon cancer cases correlate with environmental and nonhereditary events, such as chronic inflammatory disease [14, 15]. Previous studies using mouse colon cancer models highlighted the importance of chronic inflammation in the development of colon cancer. In addition, these studies illustrated the mechanisms of inflammation-driven carcinogenesis in the intestine [16]. The infiltrating immune cells and their cytokines play a role in the inflammatory response. A higher neoantigen mutational load was positively correlated with T-lymphocyte infiltration and survival outcomes in patients with colon cancer [17]. Cytokines can be pro-inflammatory (IL-1, IL-6, IL-8, tumor-necrosis factor, transforming growth factor- $\beta$  (TGF- $\beta$ )) or anti-inflammatory (IL-1ra, IL-4, IL-10, IL-13) [18, 19]. Based on the successful utilization of immune checkpoint inhibitors, immunotherapy has gained grounds in clinical oncology practice in the last decade. Although patients with colon cancer have not benefited from immunotherapy, several studies have shown that colon tumors with high mutational burden may be potential targets of immune checkpoint inhibitors [20, 21]. Inhibition of MEK upregulates IFN-gamma-mediated human leukocyte antigen (HLA) and programmed death-1 receptor (PD-L1) expression in melanoma, colorectal, and breast cancers [22, 23]. The product of HLA genes—MHC protein—can also regulate the immune system [24]. There is an increasing role for PD-1 inhibition in MSI colon cancer, while the generalized activity of PD-1 inhibitors has not been seen in microsatellite stable (MSS) colon cancer [25, 26]. Thus, there is a need to study the relationship between specific genetic variants and immune events as well as alternative approaches to treat patients with different genetic characteristics.

In this study, we performed a comprehensive evaluation of somatic mutations in colon cancer. We found that *KRAS* mutations had a strong negative correlation with

immunity and was of great prognostic value. We used single-sample Gene Set Enrichment Analysis (ssGSEA) and Gene Set Enrichment Analysis (GSEA) to identify the corresponding immune signatures of *KRAS* mutations and evaluated the relationship between *KRAS*-related pathways and immune cell infiltration. We compared the infiltration of immune cells, tumor mutational burden (TMB), HLA gene expression, and checkpoint-related genes between the *KRAS*-mutated and wild-type samples. Finally, in order to provide clues for the mechanism of *KRAS*-related immune alteration, we screened immune-related genes that differed in expression between the *KRAS*-mutated and wild-type samples.

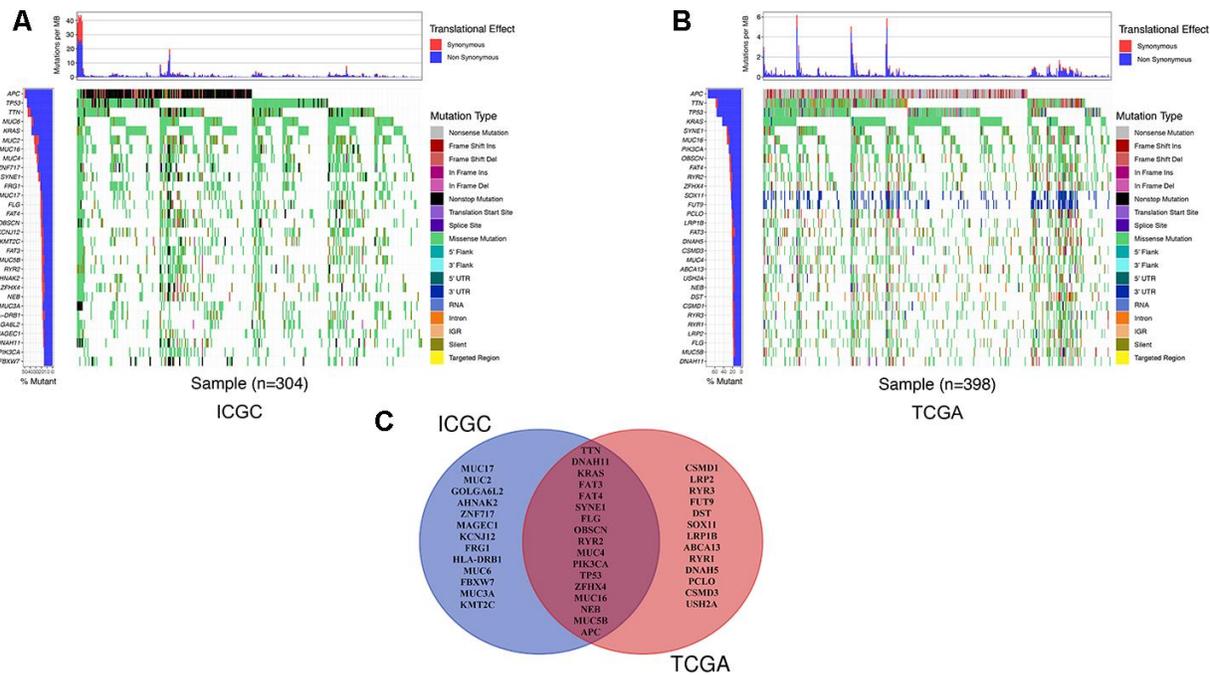
## RESULTS

### The landscape of genetic mutations in colon cancer

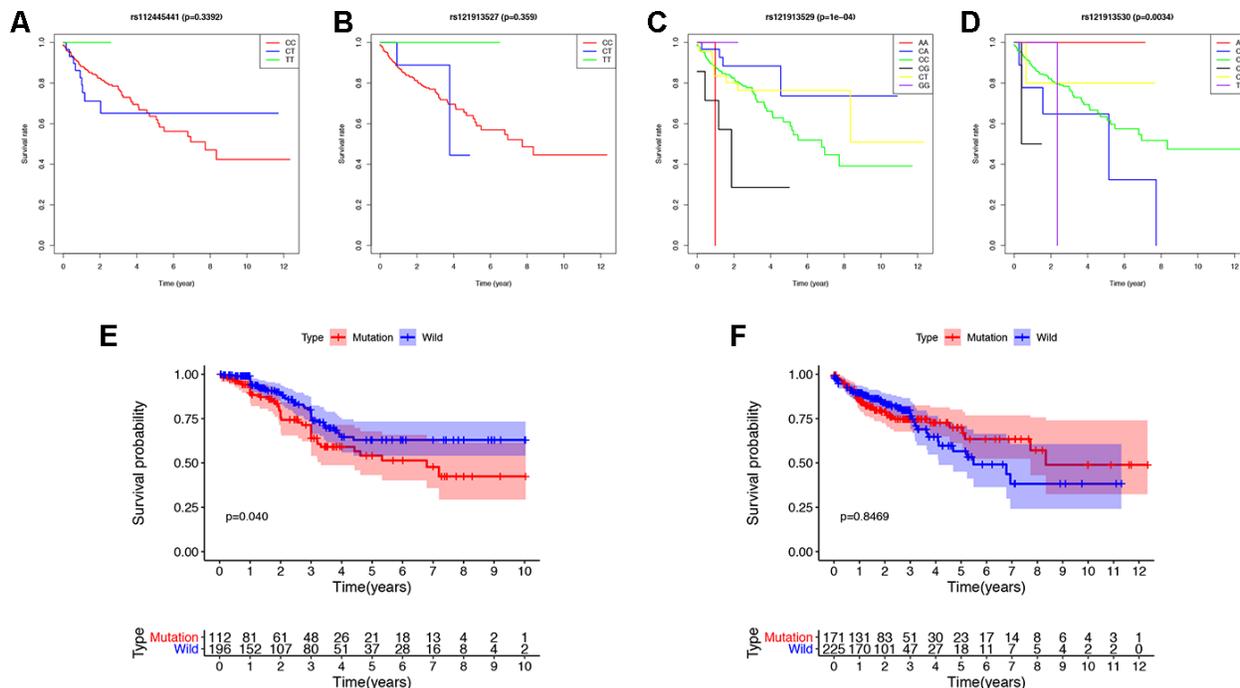
We detected the top thirty mutated genes in colon cancer samples from the ICGC database, and the top five of mutated genes were *APC*, *TP53*, *TTN*, *MUC6*, and *KRAS* (Figure 1A). We also detected the top thirty mutated genes in colon cancer samples from the The Cancer Genome Atlas (TCGA) database, and the top five mutated genes were *APC*, *TTN*, *TP53*, *KRAS*, and *SYNE1* (Figure 1B). Among the detected genes, 17 were members of the 30 most frequently mutated genes in the ICGC and TCGA databases (Figure 1C). The expression of some genes were significantly different in the mutated group than the wild-type group and included *APC* ( $p = 0.003$ ), *DNAH11* ( $p = 0.021$ ), *FAT3* ( $p = 0.031$ ), *FAT4* ( $p = 0.002$ ), *KRAS* ( $p = 0.039$ ), *MUC5B* ( $p < 0.001$ ), *PIK3CA* ( $p = 0.022$ ), and *TP53* ( $p < 0.001$ ) (Supplementary Figure 1). Next, we analyzed the mutational frequency of specific loci in the TCGA cohort. We found that rs121913529 in *KRAS* had the highest mutational frequency with 90 out of 399 patients having a mutation in this locus (Supplementary Table 2). Therefore, we predicted that *KRAS* mutations play an important role in colon cancer. We also performed survival analysis of four types of *KRAS* mutations whose mutation frequency were higher than 10/399: rs112445441 ( $p = 0.339$ ), rs121913527 ( $p = 0.359$ ), rs121913529 ( $p < 0.001$ ), and rs121913530 ( $p = 0.003$ ) (Figure 2A–2D). *KRAS*-mutated groups also showed worse survival outcomes compared to the wild-type groups in the ICGC cohort ( $p = 0.040$ ) (Figure 2E–2F). These results indicate that *KRAS* mutations, particularly rs121913529, have prognostic value in colon cancer.

### *KRAS* mutations are negatively correlated with immune activities in colon cancer

In order to explore the underlying mechanism of *KRAS* mutations in colon cancer, we performed GSEA to identify correlated pathways. We noticed that *KRAS*



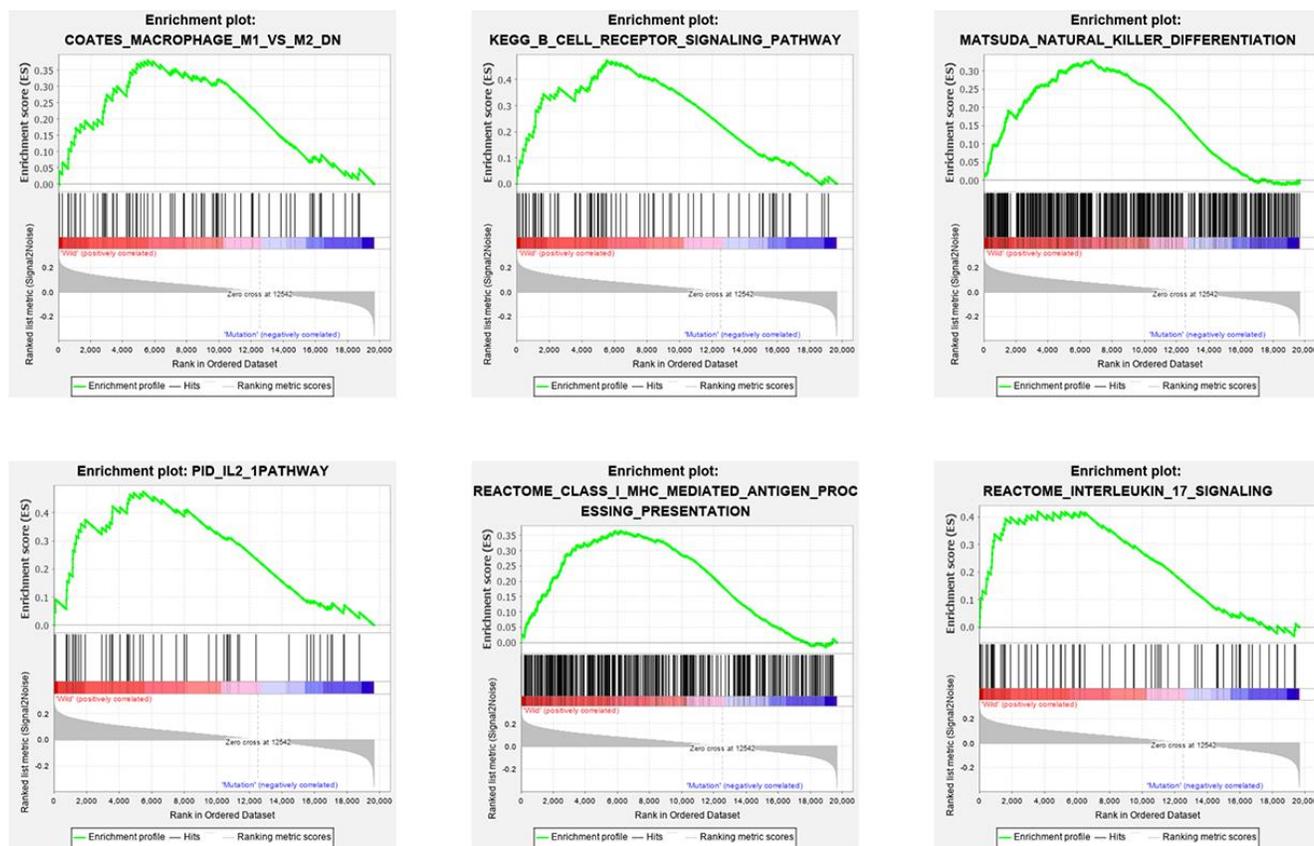
**Figure 1. The landscape of genetic mutations in colon cancer.** (A) The 30 most frequent mutations of samples in the ICGC database. The percentage of patients with mutations, translation effect (synonymous or non-synonymous), and mutation types were given. (B) The 30 most frequent mutations of samples in the TCGA database. (C) A Venn diagram of mutated genes. There were 17 genes that were members of the 30 most frequently mutated genes in the ICGC and TCGA databases.



**Figure 2. Survival analysis of KRAS mutations in colon cancer.** (A–D) The survival rate of patients with different base-pairs in rs112445441, rs121913527, rs121913529, and rs121913530. (E, F) Comparison of the survival rates between KRAS-mutated and wild-type patients from the ICGC and TCGA databases. The difference in survival rate was statistically significant in ICGC (P=0.04), but not in TCGA (P=0.8469). In TCGA analysis, there were 171 and 225 patients with and without KRAS mutations, respectively. In ICGC analysis, there were 112 and 196 patients with and without KRAS mutations, respectively.

mutations had high enrichment in some immune signatures: macrophage M1 and M2, natural killer cell (NK cell) differentiation, class I MHC-mediated antigen processing, B-cell receptor signaling, IL-2, and IL-17 pathways (Figure 3). We assessed the immunity of tumor samples by applying the ssGSEA approach to the transcriptomes of TCGA colon cancer samples (Figure 4A). We incorporated 30 immune-related pathways and infiltrating cells to estimate the immune capacity of colon cancer tissues. We found significantly lower enrichment levels in 13 pathways within the *KRAS*-mutated group: pDCs, Treg, inflammation-promoting, Th1 cells, HLA, T cell co-stimulation, cytolytic activity, tumor infiltration lymphocyte (TIL), T cell co-inhibiting, T helper cells, neutrophils, macrophages, and checkpoint (Supplementary Table 3). Furthermore, we compared the tumor purity, immune score, and stromal score between the *KRAS*-mutated and wild-type groups. The *KRAS*-mutated group had lower immune and stromal scores while its tumor purity was higher than that in the wild-type group. This revealed that the *KRAS* mutation negatively correlated with immune activities (Figure 4B–4D). Moreover, we explored the correlation

between *KRAS* mutation and specific immune signatures by analyzing expression levels of signature-related genes and immune cell infiltration. We compared the ssGSEA scores of 16 immune cell infiltration signatures between the *KRAS*-mutated and wild-type groups and found that the infiltration of macrophages ( $p = 0.033$ ), neutrophils ( $p = 0.026$ ), pDCs ( $p < 0.001$ ), T-helper cells ( $p = 0.024$ ), Th1 cells ( $p = 0.011$ ), and Tregs ( $p = 0.001$ ) were lower in *KRAS*-mutated group (Figure 5A). The TIL signature—composed of 117 genes—showed significantly higher enrichment in the wild-type than the *KRAS*-mutated groups ( $p = 0.015$ ), and 64 out of 117 genes in this signature showed lower expression levels in the *KRAS*-mutated group (Figure 5B, Supplementary Table 4). As for the inflammation-promoting signature, the *KRAS*-mutated group showed lower enrichment and 9 out of 15 genes in this signature had decreased expression ( $p = 0.002$ , Figure 5C, Supplementary Table 5). Granzyme A (GZMA) and perforin 1 (PRF1) secreted by cytotoxic T-cells and NK cells are able to kill tumor cells [27]. GZMA is a tryptase that leads to caspase-independent apoptosis, while PRF1 is a pore-forming



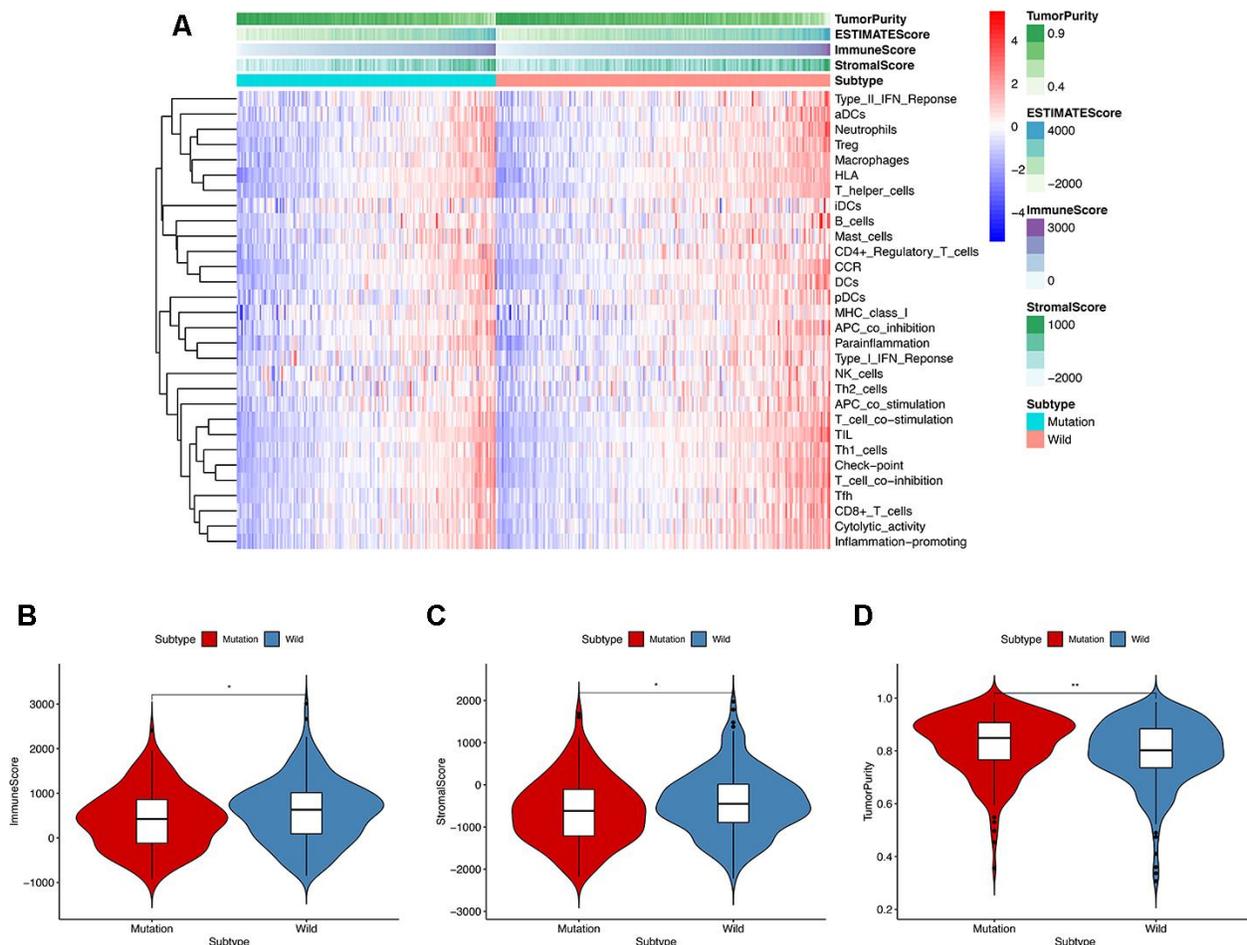
**Figure 3. Six immune pathways were enriched with *KRAS* mutations.** As shown in the enrichment plots, *KRAS* mutations were positively correlated with the immune pathways: macrophage M1 and M2, NK cell differentiation, class I MHC-mediated antigen processing, B cell receptor signaling, IL-2 and IL-17 pathways.

enzyme that facilitates the entry of granzymes into the target cells. Both effector molecules were considerably overexpressed upon CD8+ T cell activation [28]. The cytolytic activity was calculated as the mean of GZMA and PRF1 expression [29, 30]. The *KRAS*-mutated group had lower GZMA and PRF1 expression (Figure 5D–5F). These observations demonstrate that *KRAS* mutations are negatively correlated with immune cell infiltration, cytotoxic cell activity, and inflammatory response in colon cancer.

### Exploring the mechanism and function of *KRAS* mutation in immune activities

We analyzed TMB, HLA gene expression, and checkpoint-related genes in the *KRAS*-mutated and wild-type groups. Among the 19 HLA genes, 12 showed significantly lower expression levels in the

*KRAS*-mutated group compared to the wild-type group (Figure 6A). For the checkpoint-related genes, we observed that *BTLA*, *CD80*, *CD86*, *CTLA4*, *IDO1*, *PDCD1LG2*, and *TIGIT* had decreased expression in the *KRAS*-mutated group (Figure 6B). Gene mutations can generate neoantigens that mediate anti-tumor immune activities, and TMB has also been shown to have a strong correlation with tumor immunity. However, there were no significant differences in TMB between the *KRAS*-mutated and wild-type groups. This suggests that TMB cannot explain their difference in immunity (Figure 6C). RAS-related pathways obtained from KEGG included RAS, PI3K-AKT, mTOR, MAPK, FOXO, and ERBB signaling pathways. With the exception of mTOR and ERBB signaling pathways, most of the RAS-related pathways positively correlated with the immune signature. RAS ( $r = 0.61$ ) and FOXO ( $r = 0.5$ ) signaling pathways had strong positive



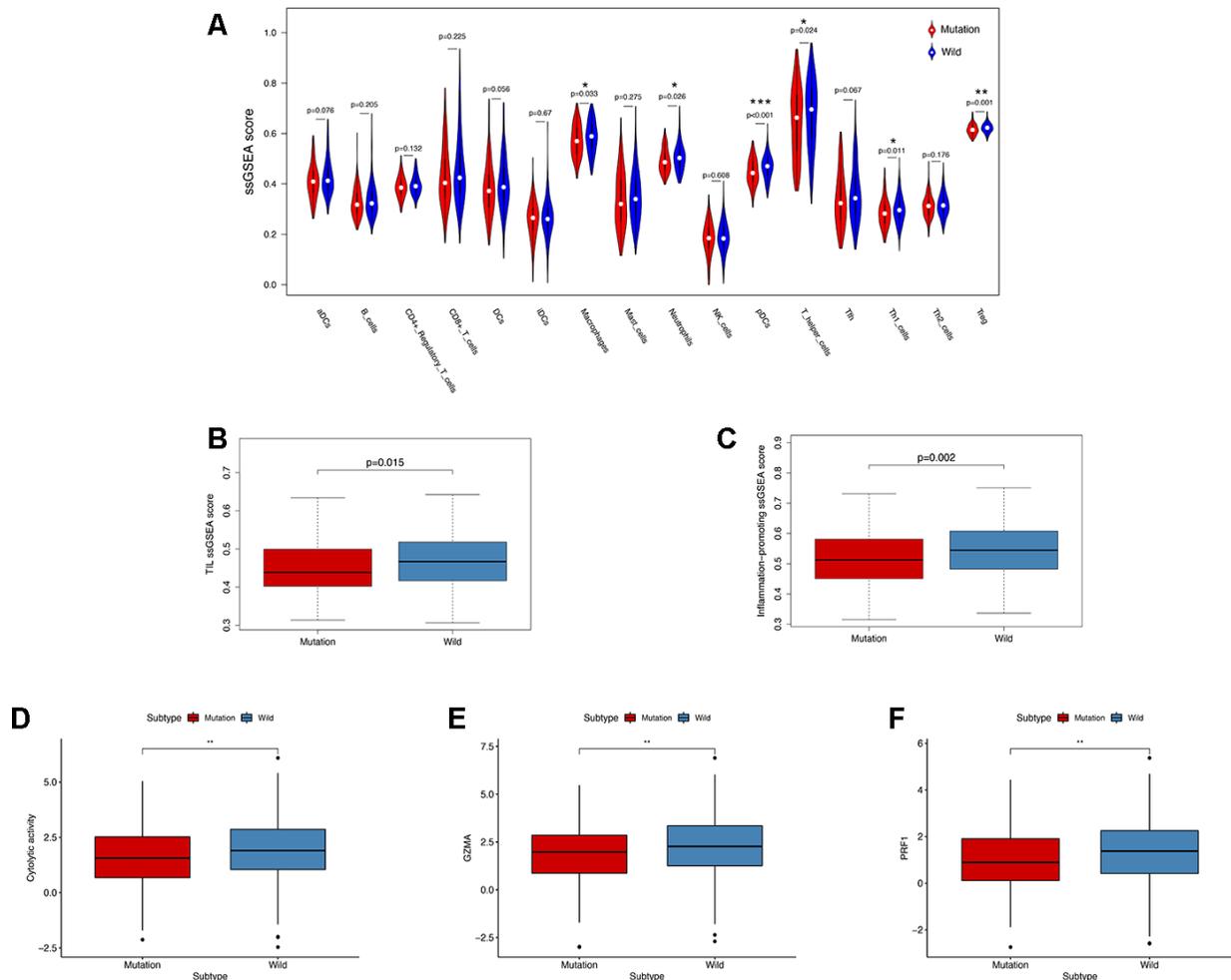
**Figure 4. ssGSEA and ESTIMATE analysis of the relationship between *KRAS* mutations and immune activities.** (A) In ssGSEA, 30 immune-related pathways were incorporated to estimate the immune capacity of individual colon cancer samples. These gene sets were composed of immune cells and processes. The tumor purity, immune score, and stromal score are also shown in the heatmap. (B–D) Using the Mann-Whitney test, we found that the *KRAS*-mutated group was of lower immune and stromal score while its tumor purity was higher than the wild-type group.

correlations with neutrophils. RAS ( $r = 0.53$ ) and PI3K-ATK signaling pathways ( $r = 0.56$ ) exhibited a positive correlation with macrophages. There was also a positive correlation between the RAS signaling pathway and T-helper cells. Interestingly, neutrophils, macrophages, and T-helpers also showed differences in ssGSEA between the *KRAS*-mutated and wild-type groups. Finally, we used the Wilcoxon test to screen for differently expressed genes between the *KRAS*-mutated and wild-type groups, with  $FDR < 0.05$ . We calculated the Pearson correlation coefficients between these differentially expressed genes (Supplementary Table 6) and the immune score of every colon cancer sample from the TCGA datasets. We identified 24 genes, which had  $cor > 0.8$  and  $p < 0.05$  (Figure 7, Supplementary Table 7). The annotations of these genes are shown in Supplementary Table 8. *KRAS* mutations may affect the

expression of these genes to further adjust the immune microenvironment.

## DISCUSSION

Based on the gene mutational landscape in colon cancer, we found that *KRAS* mutations ranked in the top five of mutated genes in the TCGA and ICGC databases. *KRAS* is a member of the RAS family, which are G-proteins involved in intracellular signaling [31]. The contribution of RAS to anti-apoptosis, proliferation, and metastasis ability of cancer cells has been well validated [32, 33]. These activities were achieved via activation of several downstream effector pathways such as canonical PI3K-AKT-mTOR and RAF-MEK-ERK kinase cascades [34]. Numerous studies indicated that *KRAS* mutations serve as prognostic and predictive



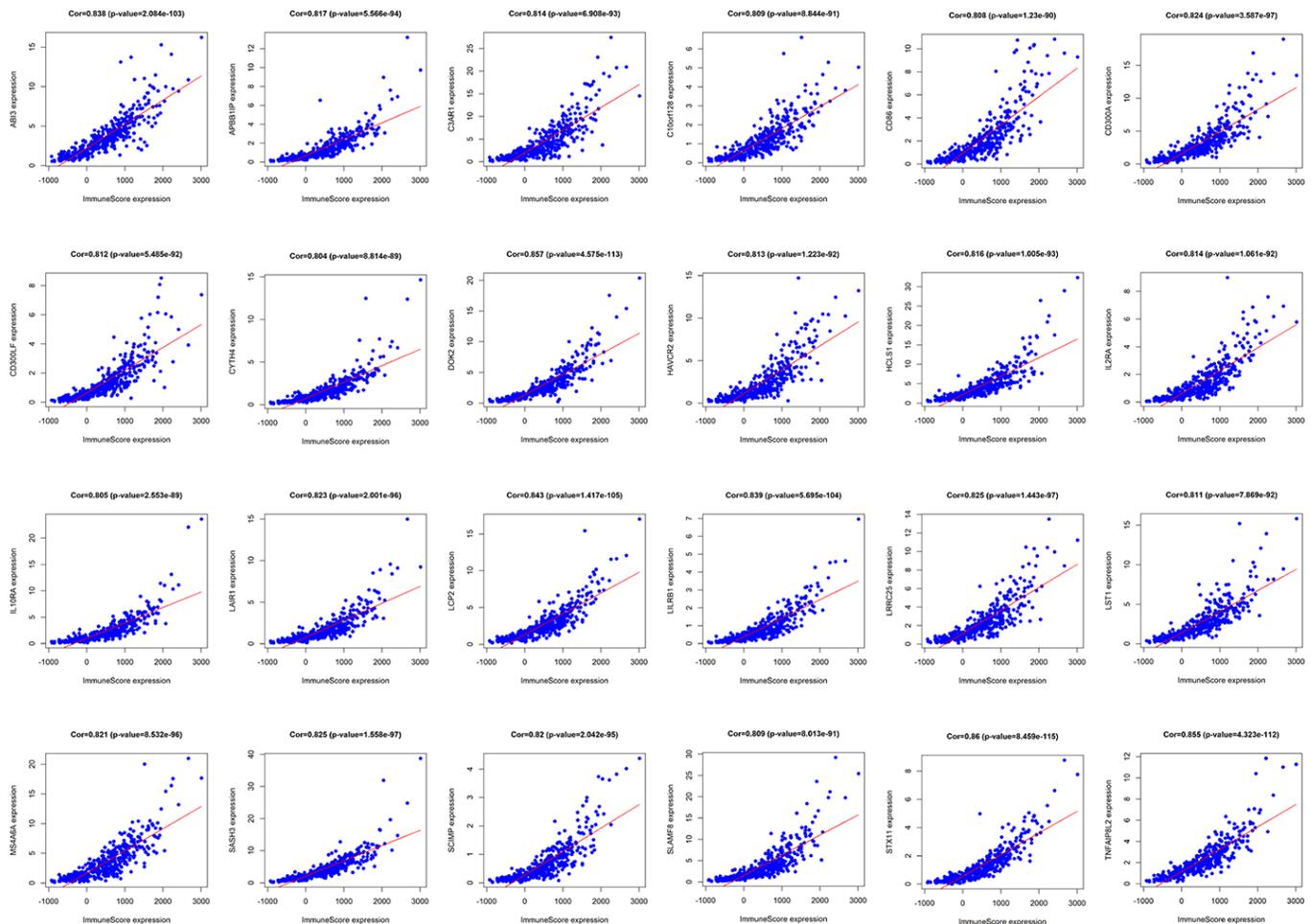
**Figure 5. *KRAS*-mutated and wild-type groups differ in immune cell infiltration, inflammation, and cytolytic activities.** (A) Relative infiltration levels of 17 immune cells in the *KRAS*-mutated and wild-type groups. P values were calculated using a Mann-Whitney test. (B, C) By comparing the ssGSEA score of two immune signatures, the *KRAS*-mutated group showed lower enrichment levels of TIL and inflammation-promoting. (D–F) The *KRAS*-mutated group showed a lower level of cytolytic activities and GZMA and PRF1 expression (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ).



and anti-CTLA-4 or anti-PD-L1 treatment to induce productive immune responses in the body. Some checkpoint-related genes (*BTLA*, *CD80*, *CD86*, *CTLA4*, *IDO1*, *PDCD1LG2*, and *TIGIT*) had decreased expression in the *KRAS*-mutated group, providing potential opportunities for immunotherapy in colon cancer.

Recent studies demonstrated that TMB could serve as a predictive biomarker for immunotherapy. Neoantigens presented by tumor cells can activate the infiltration of CD8<sup>+</sup> T-cells to recognize antigens and release cytolytic enzyme into tumor cells [42]. However, the correlation among gene mutations, TMB, and immune activities in colon cancer remains unknown. Our results showed no significant difference in TMB between the *KRAS*-mutated and wild-type groups. This indicated that TMB could not explain the relationship between *KRAS* mutations and the immune response. HLA gene expression was different between the *KRAS*-mutated

and wild-type groups, supporting the role of HLA genes in regulating *KRAS*-related immune activities. To explore the mechanism of immune differences between the two groups, we identified 24 differentially expressed genes, which strongly correlated with the immune score. These genes may participate in the upstream or downstream pathways of *KRAS*-related immune alteration. For example, overexpression of CD86 is one of the most recognized characteristics of M1 macrophages and a predictive biomarker for immunotherapy [43]. Redente et al. reported that the number of macrophages was increased in the background of a mutated oncogenic *KRAS*, providing support for mutated *KRAS* directing macrophage infiltration in tumor tissue [44]. There is a need to further investigate the predictive value of *KRAS* for immune activity in colon cancer. Future research should detail the mechanism of how *KRAS* mutation and its downstream signaling pathways alter the immune activities and clinical phenotypes of colon cancer.



**Figure 7. Pearson correlation analysis of immune score and differentially expressed genes.** There were 24 differentially expressed genes that had a strong correlation with immune score, with cor>0.8 and P<0.05.

## CONCLUSIONS

In colon cancer, the *KRAS* gene was of high mutational frequency and rs121913529 was the most frequently mutated locus. Two loci of *KRAS* (rs121913529 and rs121913530) had prognostic value in patients with colon cancer. *KRAS* mutations had a strong negative correlation with TIL, inflammation, cytolytic activities, and HLA genes. Seven checkpoint-related genes (*BTLA*, *CD80*, *CD86*, *CTLA4*, *IDO1*, *PDCD1LG2*, and *TIGIT*) had decreased expression in the *KRAS*-mutated group, providing potential opportunities for immunotherapy in colon cancer. The *KRAS*-mutated group showed lower infiltration of macrophages, neutrophils, T-helper and T-regulatory cells. In order to explore the underlying mechanism, we also detected 24 immune-related genes that differed in expression in the *KRAS*-mutated and wild-type groups.

## MATERIALS AND METHODS

### Downloaded data

We obtained somatic mutation data and clinical information of colon cancer samples from the TCGA database (n=399) via the GDC data portal (<https://portal.gdc.cancer.gov/repository>) and ICGC database (n=321, <http://dcc.icgc.org/releases/current/Projects>). We downloaded the RNA-seq data (level 3, HTSeq-FPKM) of 473 colon cancer patients with clinical information from the TCGA database. The mutation data was paired with the RNA-seq data according to patient ID. The annotations of genes were obtained from the Uniprot database (<https://www.uniprot.org/>).

### Analysis of somatic gene mutations in colon cancer

For TCGA, we downloaded the “Masked Somatic Mutation” subtype of somatic mutation data and used the VarScan software for processing. We used an R package called “maftools” [45] to analyze and visualize the Mutation Annotation Format of somatic variants. We annotated TSV files containing somatic variant information from ICGC according to the hg19 reference genome. Both cohorts were visualized by the GenVisR package. The definition of TMB is the total number of coding errors of somatic genes per million bases, including base-pair substitutions, insertions, and deletions [20]. We counted all base-pair substitutions in the coding region of specific genes, except silent mutations that failed to alter amino acids. To calculate the TMB score of each sample, we divided the total number of mutations by the exome size (38 Mb). We analyzed the difference in overall survival rates between the mutated and wide-type groups using an R package called “survival.”

## Implementation of ssGSEA and GSEA

We performed ssGSEA to acquire the enrichment score for each immune-related pair [46] and sample using an R package called “GSVA” [47] (Supplementary Table 1). We obtained 30 immune gene sets from several literature sources, including immune cell types and functions [48], tumor-infiltrating lymphocytes (TILs) [49], proinflammatory [50], para-inflammation (PI) [51], cytokine and cytokine receptor (CCR) [52], human leukocyte antigen (HLA) [53], regulatory T (Treg) cells [54], and immune checkpoints [55]. The ssGSEA applied gene signatures expressed by immune cells and pathways to the colon cancer samples. The approach used in our study involved immune cells and pathways in innate and adaptive immunity. We used an R package called “ESTIMATE” to calculate the immune score, tumor purity, and stromal score of every tumor sample [56]. The stromal score is defined as the presence of stroma in tumor tissue. The immune score is defined as the infiltration of immune cells in tumor tissue. The tumor purity score is defined as tumor purity. We performed GSEA to explore the correlated pathways of *KRAS* mutation. Gene ontology gene sets “c2.cp.v7.0.symbols.gmt” obtained from the Molecular Signatures Database (MSigDB, <http://software.broadinstitute.org/gsea/downloads.jsp>) were used for the enrichment analysis. The enriched gene set was considered statistically significant when the p value was less than 0.05. We demonstrated the correlation between *KRAS* mutations and immune cell infiltration by comparing the immune cell infiltration in *KRAS*-mutated and wild-type groups.

### Correlation between the RAS-related pathway and immune activity in colon cancer

We obtained RAS-related pathways from KEGG. We analyzed the activities of RAS-related pathways by ssGSEA scores. We used a first order partial correlation to assess the correlation between immune signatures and RAS pathways [57]. We used the Spearman correlation test to evaluate the correlation with a significance threshold of  $p < 0.05$ .

### Statistical analysis

We used the Wilcoxon test to screen for gene expression differences between the *KRAS*-mutated and wild-type groups. We analyzed the relationship between differentially expressed genes and the immune score by calculating the Pearson correlation coefficients. The expression levels of checkpoint-related and HLA genes in the *KRAS*-mutated and wild-type groups were analyzed by the Mann–Whitney U test.

## Abbreviations

The Cancer Genome Atlas: TCGA; International Cancer Genome Consortium: ICGC; Single-Sample Gene Set Enrichment Analysis: ssGSEA; Gene set enrichment analysis: GSEA; microsatellite instability: MSI; tumor-necrosis factor: TNF; transforming growth factor- $\beta$ : TGF- $\beta$ ; human leukocyte antigen: HLA; programmed death-1 receptor: PD-L1; microsatellite stable: MSS; tumor mutational burden: TMB; natural killer cell: NK cell; tumor infiltration lymphocyte: TIL; granzyme A: GZMA; perforin 1: PRF1; Mutation Annotation Format: MAF.

## AUTHOR CONTRIBUTIONS

Qunguang Jiang, Xiaorui Fu and Xinyi Wang designed the study. Xinyi Wang and Taiyuan Li collected the mRNA transcriptome data and clinical information from TCGA and ICGC. Xiaorui Fu and Xinyi Wang performed analyses on the TCGA and ICGC data. Qunguang Jiang and Jinzhong Duanmu performed statistical analyses. Xiaorui Fu wrote the manuscript. Qunguang Jiang and Xinyi Wang reviewed and revised the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge contributions from the ICGC and TCGA databases.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This work was supported by a grant from National Natural Science Foundation of China (No:81560397 and 81660403).

## REFERENCES

1. Labianca R, Beretta GD, Kildani B, Milesi L, Merlin F, Mosconi S, Pessi MA, Prochilo T, Quadri A, Gatta G, de Braud F, Wils J. Colon cancer. *Crit Rev Oncol Hematol*. 2010; 74:106–33.  
<https://doi.org/10.1016/j.critrevonc.2010.01.010>  
PMID:[20138539](https://pubmed.ncbi.nlm.nih.gov/20138539/)
2. Punt CJ, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat Rev Clin Oncol*. 2017; 14:235–46.  
<https://doi.org/10.1038/nrclinonc.2016.171>  
PMID:[27922044](https://pubmed.ncbi.nlm.nih.gov/27922044/)
3. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, Reid JG, Santibanez J, Shinbrot E, et al. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487:330–37.  
<https://doi.org/10.1038/nature11252>  
PMID:[22810696](https://pubmed.ncbi.nlm.nih.gov/22810696/)
4. Kwong LN, Dove WF. APC and its modifiers in colon cancer. *Adv Exp Med Biol*. 2009; 656:85–106.  
[https://doi.org/10.1007/978-1-4419-1145-2\\_8](https://doi.org/10.1007/978-1-4419-1145-2_8)  
PMID:[19928355](https://pubmed.ncbi.nlm.nih.gov/19928355/)
5. Narayan S, Roy D. Role of APC and DNA mismatch repair genes in the development of colorectal cancers. *Mol Cancer*. 2003; 2:41.  
<https://doi.org/10.1186/1476-4598-2-41>  
PMID:[14672538](https://pubmed.ncbi.nlm.nih.gov/14672538/)
6. Soussi T, Wiman KG. TP53: an oncogene in disguise. *Cell Death Differ*. 2015; 22:1239–49.  
<https://doi.org/10.1038/cdd.2015.53> PMID:[26024390](https://pubmed.ncbi.nlm.nih.gov/26024390/)
7. Iacopetta B. TP53 mutation in colorectal cancer. *Hum Mutat*. 2003; 21:271–76.  
<https://doi.org/10.1002/humu.10175> PMID:[12619112](https://pubmed.ncbi.nlm.nih.gov/12619112/)
8. Arrington AK, Heinrich EL, Lee W, Duldulao M, Patel S, Sanchez J, Garcia-Aguilar J, Kim J. Prognostic and predictive roles of KRAS mutation in colorectal cancer. *Int J Mol Sci*. 2012; 13:12153–68.  
<https://doi.org/10.3390/ijms131012153>  
PMID:[23202889](https://pubmed.ncbi.nlm.nih.gov/23202889/)
9. Siddiqui AD, Piperdi B. KRAS mutation in colon cancer: a marker of resistance to EGFR-I therapy. *Ann Surg Oncol*. 2010; 17:1168–76.  
<https://doi.org/10.1245/s10434-009-0811-z>  
PMID:[19936839](https://pubmed.ncbi.nlm.nih.gov/19936839/)
10. Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm Sin B*. 2015; 5:390–401.  
<https://doi.org/10.1016/j.apsb.2015.07.001>  
PMID:[26579470](https://pubmed.ncbi.nlm.nih.gov/26579470/)
11. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010; 138:2073–87.e3.  
<https://doi.org/10.1053/j.gastro.2009.12.064>  
PMID:[20420947](https://pubmed.ncbi.nlm.nih.gov/20420947/)
12. Nojadedh JN, Behrouz Sharif S, Sakhinia E. Microsatellite instability in colorectal cancer. *EXCLI J*. 2018; 17:159–68.  
<https://doi.org/10.17179/excli2017-948>  
PMID:[29743854](https://pubmed.ncbi.nlm.nih.gov/29743854/)
13. Terzić J, Grivennikov S, Karin E, Karin M. Inflammation and colon cancer. *Gastroenterology*. 2010; 138:2101–14.e5.  
<https://doi.org/10.1053/j.gastro.2010.01.058>  
PMID:[20420949](https://pubmed.ncbi.nlm.nih.gov/20420949/)

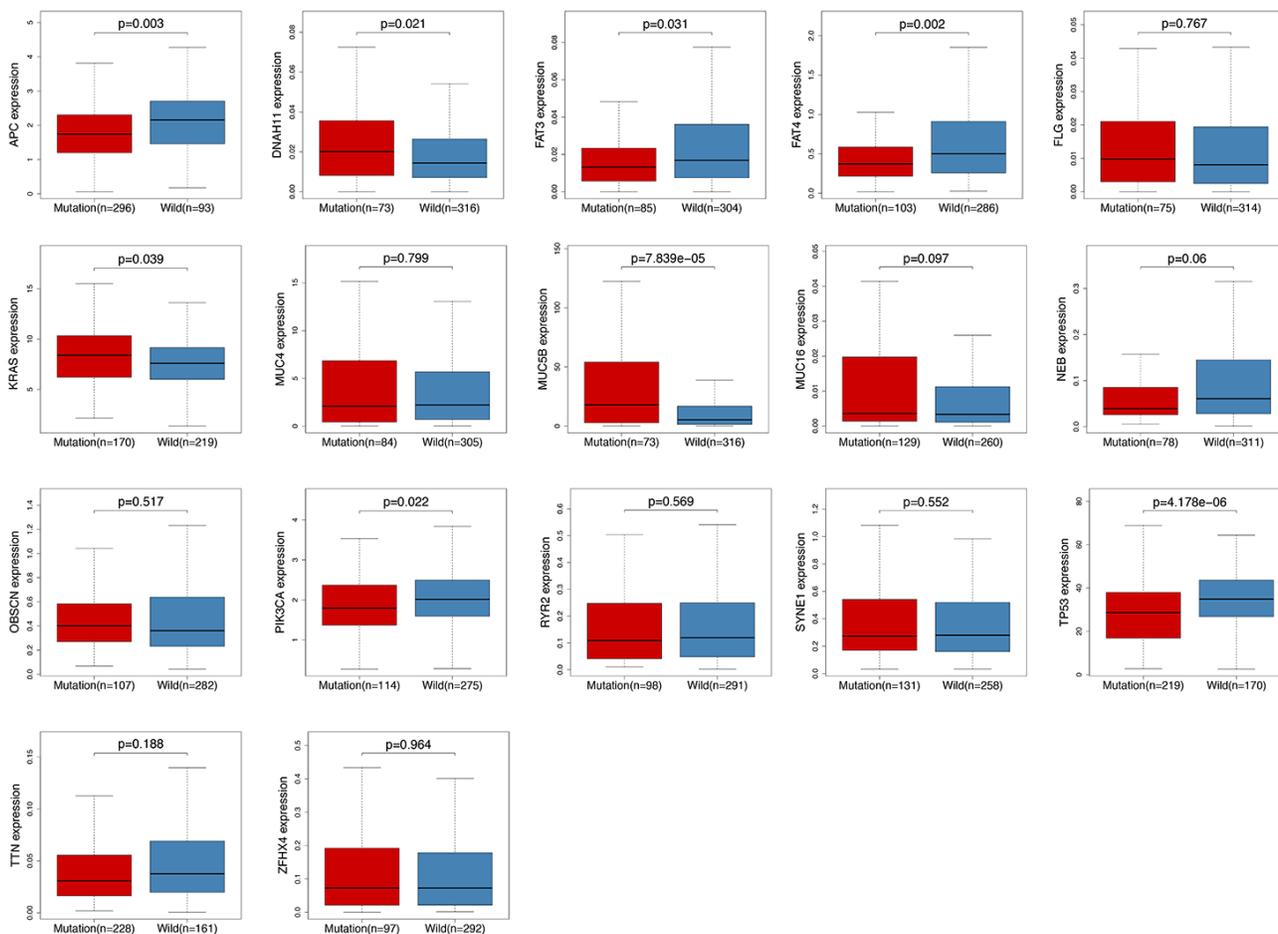
14. Church J. Molecular genetics of colorectal cancer. *Semin Colon Rectal Surg.* 2016; 27:172–75.  
<https://linkinghub.elsevier.com/retrieve/pii/S1043148916300203>
15. Rustgi AK. The genetics of hereditary colon cancer. *Genes Dev.* 2007; 21:2525–38.  
<https://doi.org/10.1101/gad.1593107> PMID:[17938238](https://pubmed.ncbi.nlm.nih.gov/17938238/)
16. Kang CW, Dutta A, Chang LY, Mahalingam J, Lin YC, Chiang JM, Hsu CY, Huang CT, Su WT, Chu YY, Lin CY. Apoptosis of tumor infiltrating effector TIM-3+CD8+ T cells in colon cancer. *Sci Rep.* 2015; 5:15659.  
<https://doi.org/10.1038/srep15659> PMID:[26493689](https://pubmed.ncbi.nlm.nih.gov/26493689/)
17. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet.* 2019; 51:202–06.  
<https://doi.org/10.1038/s41588-018-0312-8> PMID:[30643254](https://pubmed.ncbi.nlm.nih.gov/30643254/)
18. Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. *Gastroenterology.* 2011; 140:1807–16.  
<https://doi.org/10.1053/j.gastro.2011.01.057> PMID:[21530747](https://pubmed.ncbi.nlm.nih.gov/21530747/)
19. Cullis J, Das S, Bar-Sagi D. Kras and tumor immunity: friend or foe? *Cold Spring Harb Perspect Med.* 2018; 8:a031849.  
<https://doi.org/10.1101/cshperspect.a031849> PMID:[29229670](https://pubmed.ncbi.nlm.nih.gov/29229670/)
20. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther.* 2017; 16:2598–608.  
<https://doi.org/10.1158/1535-7163.MCT-17-0386> PMID:[28835386](https://pubmed.ncbi.nlm.nih.gov/28835386/)
21. Miao D, Margolis CA, Vokes NI, Liu D, Taylor-Weiner A, Wankowicz SM, Adeegbe D, Keliher D, Schilling B, Tracy A, Manos M, Chau NG, Hanna GJ, et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet.* 2018; 50:1271–81.  
<https://doi.org/10.1038/s41588-018-0200-2> PMID:[30150660](https://pubmed.ncbi.nlm.nih.gov/30150660/)
22. Chang CH, Qiu J, O’Sullivan D, Buck MD, Noguchi T, Curtis JD, Chen Q, Gindin M, Gubin MM, van der Windt GJ, Tonc E, Schreiber RD, Pearce EJ, Pearce EL. Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell.* 2015; 162:1229–41.  
<https://doi.org/10.1016/j.cell.2015.08.016> PMID:[26321679](https://pubmed.ncbi.nlm.nih.gov/26321679/)
23. Brandt R, Sell T, Lüthen M, Uhlitz F, Klinger B, Riemer P, Giesecke-Thiel C, Schulze S, El-Shimy IA, Kunkel D, Fauler B, Mielke T, Mages N, et al. Cell type-dependent differential activation of ERK by oncogenic KRAS in colon cancer and intestinal epithelium. *Nat Commun.* 2019; 10:2919.  
<https://doi.org/10.1038/s41467-019-10954-y> PMID:[31266962](https://pubmed.ncbi.nlm.nih.gov/31266962/)
24. Parkhurst MR, Joo J, Riley JP, Yu Z, Li Y, Robbins PF, Rosenberg SA. Characterization of genetically modified t-cell receptors that recognize the CEA:691-699 peptide in the context of HLA-A2.1 on human colorectal cancer cells. *Clin Cancer Res.* 2009; 15:169–80.  
<https://doi.org/10.1158/1078-0432.CCR-08-1638> PMID:[19118044](https://pubmed.ncbi.nlm.nih.gov/19118044/)
25. Yaghoubi N, Soltani A, Ghazvini K, Hassanian SM, Hashemy SI. PD-1/ PD-L1 blockade as a novel treatment for colorectal cancer. *Biomed Pharmacother.* 2019; 110:312–18.  
<https://doi.org/10.1016/j.biopha.2018.11.105> PMID:[30522017](https://pubmed.ncbi.nlm.nih.gov/30522017/)
26. Nguyen LT, Ohashi PS. Clinical blockade of PD1 and LAG3—potential mechanisms of action. *Nat Rev Immunol.* 2015; 15:45–56.  
<https://doi.org/10.1038/nri3790> PMID:[25534622](https://pubmed.ncbi.nlm.nih.gov/25534622/)
27. Woodsworth DJ, Dreolini L, Abraham L, Holt RA. Targeted cell-to-cell delivery of protein payloads via the granzyme-perforin pathway. *Mol Ther Methods Clin Dev.* 2017; 7:132–45.  
<https://doi.org/10.1016/j.omtm.2017.10.003> PMID:[29201936](https://pubmed.ncbi.nlm.nih.gov/29201936/)
28. Johnson BJ, Costelloe EO, Fitzpatrick DR, Haanen JB, Schumacher TN, Brown LE, Kelso A. Single-cell perforin and granzyme expression reveals the anatomical localization of effector CD8+ T cells in influenza virus-infected mice. *Proc Natl Acad Sci USA.* 2003; 100:2657–62.  
<https://doi.org/10.1073/pnas.0538056100> PMID:[12601154](https://pubmed.ncbi.nlm.nih.gov/12601154/)
29. Zaravinos A, Roufas C, Nagara M, de Lucas Moreno B, Oblovatskaya M, Efstathiades C, Dimopoulos C, Ayiomamitis GD. Cytolytic activity correlates with the mutational burden and deregulated expression of immune checkpoints in colorectal cancer. *J Exp Clin Cancer Res.* 2019; 38:364.  
<https://doi.org/10.1186/s13046-019-1372-z> PMID:[31429779](https://pubmed.ncbi.nlm.nih.gov/31429779/)
30. Narayanan S, Kawaguchi T, Yan L, Peng X, Qi Q, Takabe K. Cytolytic activity score to assess anticancer immunity in colorectal cancer. *Ann Surg Oncol.* 2018; 25:2323–31.

- <https://doi.org/10.1245/s10434-018-6506-6>  
PMID:29770915
31. Simanshu DK, Nissley DV, McCormick F. RAS proteins and their regulators in human disease. *Cell*. 2017; 170:17–33.  
<https://doi.org/10.1016/j.cell.2017.06.009>  
PMID:28666118
32. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. *Nat Rev Cancer*. 2011; 11:761–74.  
<https://doi.org/10.1038/nrc3106> PMID:21993244
33. Fernández-Medarde A, Santos E. Ras in cancer and developmental diseases. *Genes Cancer*. 2011; 2:344–58.  
<https://doi.org/10.1177/1947601911411084>  
PMID:21779504
34. Zhou B, Der CJ, Cox AD. The role of wild type RAS isoforms in cancer. *Semin Cell Dev Biol*. 2016; 58:60–69.  
<https://doi.org/10.1016/j.semcdb.2016.07.012>  
PMID:27422332
35. Demory Beckler M, Higginbotham JN, Franklin JL, Ham AJ, Halvey PJ, Imasuen IE, Whitwell C, Li M, Liebler DC, Coffey RJ. Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. *Mol Cell Proteomics*. 2013; 12:343–55.  
<https://doi.org/10.1074/mcp.M112.022806>  
PMID:23161513
36. Calcagno SR, Li S, Colon M, Kreinest PA, Thompson EA, Fields AP, Murray NR. Oncogenic k-ras promotes early carcinogenesis in the mouse proximal colon. *Int J Cancer*. 2008; 122:2462–70.  
<https://doi.org/10.1002/ijc.23383>  
PMID:18271008
37. Tong L, Yang XX, Liu MF, Yao GY, Dong JY, Ye CS, Li M. Mutational analysis of key EGFR pathway genes in Chinese breast cancer patients. *Asian Pac J Cancer Prev*. 2012; 13:5599–603.  
<https://doi.org/10.7314/apjcp.2012.13.11.5599>  
PMID:23317280
38. Liao W, Overman MJ, Boutin AT, Shang X, Zhao D, Dey P, Li J, Wang G, Lan Z, Li J, Tang M, Jiang S, Ma X, et al. KRAS-IRF2 axis drives immune suppression and immune therapy resistance in colorectal cancer. *Cancer Cell*. 2019; 35:559–72.e7.  
<https://doi.org/10.1016/j.ccell.2019.02.008>  
PMID:30905761
39. Skoulidis F, Byers LA, Diao L, Papadimitrakopoulou VA, Tong P, Izzo J, Behrens C, Kadara H, Parra ER, Canales JR, Zhang J, Giri U, Gudikote J, et al. Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. *Cancer Discov*. 2015; 5:860–77.  
<https://doi.org/10.1158/2159-8290.CD-14-1236>  
PMID:26069186
40. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, Moreira AL, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015; 348:124–28.  
<https://doi.org/10.1126/science.aaa1348>  
PMID:25765070
41. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhäufel M, Arrieta O, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015; 373:1627–39.  
<https://doi.org/10.1056/NEJMoa1507643>  
PMID:26412456
42. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. 2013; 14:1014–22.  
<https://doi.org/10.1038/ni.2703> PMID:24048123
43. Edwards JP, Zhang X, Frauwirth KA, Mosser DM. Biochemical and functional characterization of three activated macrophage populations. *J Leukoc Biol*. 2006; 80:1298–307.  
<https://doi.org/10.1189/jlb.0406249>  
PMID:16905575
44. Redente EF, Dwyer-Nield LD, Merrick DT, Raina K, Agarwal R, Pao W, Rice PL, Shroyer KR, Malkinson AM. Tumor progression stage and anatomical site regulate tumor-associated macrophage and bone marrow-derived monocyte polarization. *Am J Pathol*. 2010; 176:2972–85.  
<https://doi.org/10.2353/ajpath.2010.090879>  
PMID:20431028
45. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018; 28:1747–56.  
<https://doi.org/10.1101/gr.239244.118>  
PMID:30341162
46. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009; 462:108–12.  
<https://doi.org/10.1038/nature08460> PMID:19847166
47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub

- TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50.  
<https://doi.org/10.1073/pnas.0506580102>  
 PMID:16199517
48. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015; 160:48–61.  
<https://doi.org/10.1016/j.cell.2014.12.033>  
 PMID:25594174
49. Massink MP, Kooi IE, Martens JW, Waisfisz Q, Meijers-Heijboer H. Genomic profiling of CHEK2\*1100delC-mutated breast carcinomas. *BMC Cancer*. 2015; 15:877.  
<https://doi.org/10.1186/s12885-015-1880-y>  
 PMID:26553136
50. Bedognetti D, Hendrickx W, Marincola FM, Miller LD. Prognostic and predictive immune gene signatures in breast cancer. *Curr Opin Oncol*. 2015; 27:433–44.  
<https://doi.org/10.1097/CCO.0000000000000234>  
 PMID:26418235
51. Aran D, Lasry A, Zinger A, Biton M, Pikarsky E, Hellman A, Butte AJ, Ben-Neriah Y. Widespread parainflammation in human cancer. *Genome Biol*. 2016; 17:145.  
<https://doi.org/10.1186/s13059-016-0995-z>  
 PMID:27386949
52. Wong HS, Chang CM, Liu X, Huang WC, Chang WC. Characterization of cytokinome landscape for clinical responses in human cancers. *Oncoimmunology*. 2016; 5:e1214789.  
<https://doi.org/10.1080/2162402X.2016.1214789>  
 PMID:27999736
53. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017; 18:248–62.  
<https://doi.org/10.1016/j.celrep.2016.12.019>  
 PMID:28052254
54. Liu Z, Li M, Jiang Z, Wang X. A comprehensive immunologic portrait of triple-negative breast cancer. *Transl Oncol*. 2018; 11:311–29.  
<https://doi.org/10.1016/j.tranon.2018.01.011>  
 PMID:29413765
55. De Simone M, Arrigoni A, Rossetti G, Guarini P, Ranzani V, Politano C, Bonnal RJ, Provasi E, Sarnicola ML, Panzeri I, Moro M, Crosti M, Mazzara S, et al. Transcriptional landscape of human tissue lymphocytes unveils uniqueness of tumor-infiltrating T regulatory cells. *Immunity*. 2016; 45:1135–47.  
<https://doi.org/10.1016/j.immuni.2016.10.021>  
 PMID:27851914
56. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013; 4:2612.  
<https://doi.org/10.1038/ncomms3612> PMID:24113773
57. Kim S. Ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun Stat Appl Methods*. 2015; 22:665–74.  
<https://doi.org/10.5351/CSAM.2015.22.6.665>  
 PMID:26688802

## SUPPLEMENTARY MATERIALS

### Supplementary Figure



**Supplementary Figure 1. Correlation between somatic mutations with gene expression.** Among the 17 genes detected in Figure 1C, the mutation of 8 genes were correlated with their expression levels: *APC*, *DNAH11*, *FAT3*, *FAT4*, *KRAS*, *MUC5B*, *PIK3CA*, and *TP53*. The correlation analysis was performed based on TCGA.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 4, 6, 7.

**Supplementary Table 1. The genetic composition of 30 immune signatures.**

aDCs	CD83;LAMP3;CCL1
APC_co_inhibition	C10orf54;CD274;LGALS9;PDCD1LG2;PVRL3
APC_co_stimulation	CD40;CD58;CD70;ICOSLG;SLAMF1;TNFSF14;TNFSF15;TNFSF18;TNFSF4;TNFSF8;TNFSF9
B_cells	BACH2;BANK1;BLK;BTLA;CD79A;CD79B;FCRL1;FCRL3;HVCN1;RALGPS2
CCR	CCL16;TPO;TGFB2;CXCL2;CCL14;TGFB3;IL11RA;CCL11;IL4I1;IL33;CXCL12;CXCL10;BMPER;BMP8A;CXCL11;IL21R;IL17B;TNFRSF9;ILF2;CX3CR1;CCR8;TNFSF12;CSF3;TNFSF4;BMP3;CX3CL1;BMP5;CXCR2;TNFRSF10D;BMP2;CXCL14;CCL28;CXCL3;BMP6;CCL21;CXCL9;CCL23;IL6;TNFRSF18;IL17RD;IL17D;IL27;CCL7;IL1R1;CXCR4;CXCR2P1;TGFB11;IFNGR1;IL9R;IL1RAPL1;IL11;CSF1;IL20RA;IL25;TNFRSF4;IL18;ILF3;CCL20;TNFRSF12A;IL6ST;CXCL13;IL12B;TNFRSF8;IL6R;BMP2;IFNE;IL1RAPL2;IL3RA;BMP4;CCL24;TNFSF13B;CCR4;IL2RA;IL32;TNFRSF10C;IL22RA1;BMP1A;CXCR5;CXCR3;IFNA8;IL17REL;IFNB1;IFNAR1;TNFRSF1B;CCL17;IFNL1;IL16;IL1RL1;ILK;CCL25;ILDR2;CXCR1;IL36RN;IL34;TGFB1;IFNG;IL19;ILKAP;BMP2K;CCR10;ILDR1;EPO;CCR7;IL17C;IL23A;CCR5;IL7;EPOR;CCL13;IL2RG;IL31RA;TNFAIP6;IFNL2;BMP1;IL12RB1;TNFAIP8;IL4R;TNFRSF6B;TNFAIP8L1;TNFRSF10B;IFNL3;CCL5;CXCL6;CXCL1;CCR3;TNFSF11;CSF1R;IL21;IL1RAP;IL12RB2;CCL1;IL17RA;CCR1;IL1RN;TNFRSF11B;TNFRSF14;IL13;IL2RB;BMP8B;CCL2;IL24;IL18RAP;TGFB1;TNFSF10;TNFRSF11A;CXCL5;IL5RA;TNFSF9;IL1RL2;TNFRSF13C;IL36G;IL15RA;TNFRSF21;CXCL8;IL22RA2;TNFAIP8L2;IL18R1;IFNLR1;CXCR6;CCL3L3;TNFRSF1A;IL17RE;IFNGR2;IL17RC;TNFAIP8L3;ILVBL;TGFBRAPI;CCL4L1;CSF2RA;CCRN4L;CCL26;TNFAIP1;CCRL2;IFNA10;TNFRSF17;IFNA13;IL20;IL18BP;CCL3L1;TNFSF12-TNFSF13;IL5;IL23R;IL26;TNF;TGFA;CSF2;IL1F10;CXCL17;TNFSF13;IFNA4;IL37;IL12A;IL7R;IFNA1;IL1A;IL4;IL2;CCL22;CSF3R;IL10;IFNK;TGFB2;IL1R2;IL1B;IL17F;IL27RA;IL15;TNFSF8;IL36B;XCL1;CXCL16;TNFRSF19;IL3;CCL3;IFNA2;BMP1B;IFNA21;TNFSF18;CCL8;IL17RB;TNFRSF25;IL22;IL10RB;IFNAR2;CCL18;IFNA16;CSF2RB;IL36A;TNFAIP3;IL13RA2;IL13RA1;CCR9;TNFRSF10A;IFNA7;IFNW1;XCL2;TNFSF14;CCR2;BMP15;BMP10;CCL15;CCL14;TGFB1;IFNA5;BMP7;IFNA14;IL20RB;IL10RA;IFNA17;CCR6;TGFB3;CCL15;CCL4;CCL27;TNFRSF13B;TNFAIP2;IL31;IL17A;TNFSF15;CCL19;IFNA6;IL9
CD8+_T_cells	CD8A
CD4+ Regulatory T cells	C15orf53;CTLA4;FOXP3;GPR15;IL32;IL4;IL5
Check-point	IDO1;LAG3;CTLA4;TNFRSF9;ICOS;CD80;PDCD1LG2;TIGIT;CD70;TNFSF9;ICOSLG;KIR3DL1;CD86;PDCD1;LAIR1;TNFRSF8;TNFSF15;TNFRSF14;IDO2;CD276;CD40;TNFRSF4;TNFSF14;HHLA2;CD244;CD274;HAVCR2;CD27;BTLA;LGALS9;TMIGD2;CD28;CD48;TNFRSF25;CD40LG;ADORA2A;VTCN1;CD160;CD44;TNFSF18;TNFRSF18;BTNL2;C10orf54;CD200R1;TNFSF4;CD200;NRP1
Cytolytic_activity	PRF1;GZMA
DCs	CCL17;CCL22;CD209;CCL13
HLA	HLA-E;HLA-DPB2;HLA-C HLA-J;HLA-DQB1;HLA-DQB2;HLA-DQA2;HLA-DQA1;HLA-A;HLA-DMA;HLA-DOB;HLA-DRB1;HLA-H;HLA-B;HLA-DRB5;HLA-DOA;HLA-DPB1;HLA-DRA;HLA-DRB6;HLA-L;HLA-F;HLA-G;HLA-DMB;HLA-DPA1
iDCs	CD1A;CD1E
Inflammation-promoting	CCL5;CD19;CD8B;CXCL10;CXCL13;CXCL9;GNLY;GZMB;IFNG;IL12A;IL12B;IRF1;PRF1;STAT1;TBX21
Macrophages	C11orf45;CD68;CLEC5A;CYBB;FUCA1;GPNMB;HS3ST2;LGGMN;MMP9;TM4SF19
Mast_cells	CMA1;MS4A2;TPSAB1
MHC_class_I	B2M;HLA-A;TAP1
Neutrophils	EVI2B;HSD17B11;KDM6B;MEGF9;MNDA;NLRP12;PADI4;SELL;TRANK1;VNN3
NK_cells	KLRC1;KLRF1
Parainflammation	CXCL10;PLAT;CCND1;LGGMN;PLAUR;AIM2;MMP7;ICAM1;MX2;CXCL9;ANXA1;TLR2;PLA2G2D;ITGA2;MX1;HMOX1;CD276;TIRAP;IL33;PTGES;TNFRSF12A;SCARB1;CD14;BLNK;IFIT3;RETNLB;IFIT2;ISG15;OAS2;REL;OAS3;CD44;PPARG;BST2;OAS1;NOX1;PLA2G2A;IFIT1;IFITM3;IL1RN
pDCs	CLEC4C;CXCR3;GZMB;IL3RA;IRF7;IRF8;LILRA4;PHEX;PLD4;PTCRA
T_cell_co-inhibition	BTLA;C10orf54;CD160;CD244;CD274;CTLA4;HAVCR2;LAG3;LAIR1;TIGIT
T_cell_co-stimulation	CD2;CD226;CD27;CD28;CD40LG;ICOS;SLAMF1;TNFRSF18;TNFRSF25;TNFRSF4;TNFRSF8;TNFRSF9;TNFSF14
T_helper_cells	CD4
Tfh	PDCD1;CXCL13;CXCR5
Th1_cells	IFNG;TBX21;CTLA4;STAT4;CD38;IL12RB2;LTA;CSF2

Th2_cells	PMCH;LAIR2;SMAD2;CXCR6;GATA3;IL26
TIL	ITM2C;CD38;THEMIS2;GLYR1;ICOS;F5;TIGIT;KLRD1;IRF4;PRKCQ;FCRL5;SIRPG;LPXN;IL2RG;CCL5;LCK;TRAF3 IP3;CD86;MAL;LILRB1;DOK2;CD6;PAG1;LAX1;PLEK;PIK3CD;SLAMF1;XCL1;GPR171;XCL2;TBX21;CD2;CD53;KL HL6;SLAMF6;CD40;SIT1;TNFRSF4;CD79A;CD247;LCP2;CD3D;CD27;SH2D1A;FYB;ARHGAP30;ACAP1;CST7;CD3G ;IL2RB;CD3E;FCRL3;CORO1A;ITK;TCL1A;CYBB;CSF2RB;IKZF1;NCF4;DOCK2;CCR2;PTPRC;PLAC8;NCKAP1L;IL 7R;6- Sep;CD28;STAT4;CD8A;LY9;CD48;HCST;PTPRCAP;SASH3;ARHGAP25;LAT;TRAT1;IL10RA;PAX5;CCR7;DOCK11; PARVG;SPNS1;CD52;HCLS1;ARHGAP9;GIMAP6;PRKCB;MS4A1;GPR18;TBC1D10C;GVINP1;P2RY8;EVI2B;VAMP5 ;KLRK1;SELL;MPEG1;MS4A6A;ARHGAP15;MFNG;GZMK;SELPLG;TARP;GIMAP7;FAM65B;INPP5D;ITGA4;MZB1; GPSM3;STK10;CLEC2D;IL16;NLRC3;GIMAP5;GIMAP4;IFFO1;CFH;PVRIG;CFHR1
Treg	IL12RB2;TMPRSS6;CTSC;LAPTM4B;TFRC;RNF145;NETO2;ADAT2;CHST2;CTLA4;NFE2L3;LIMA1;IL1R2;ICOS;HS DL2;HTATIP2;FKBP1A;TIGIT;CCR8;LTA;SLC35F2;IL21R;AHCYL1;SOCS2;ETV7;BCL2L1;RRAGB;ACSL4;CHRNA6; BATF;LAX1;ADPRH;TNFRSF4;ANKRD10;CD274;CASP1;LY75;NPTN;SSTR3;GRSF1;CSF2RB;TMEM184C;NDFIP2;Z BTB38;ERL1;TRAF3;NAB1;HS3ST3B1;LAYN;JAK1;VDR;LEPROT;GCNT1;PTPRJ;IKZF2;CSF1;ENTPD1;TNFRSF18;M ETTL7A;KSR1;SSH1;CADM1;IL1R1;ACP5;CHST7;THADA;CD177;NFAT5;ZNF282;MAGEH1
Type_I_IFN_Reponse	DDX4;IFIT1;IFIT2;IFIT3;IRF7;ISG20;MX1;MX2;RSAD2;TNFSF10
Type_II_IFN_Reponse	GPR146;SELP;AHR

Supplementary Table 2. The mutational frequency of genes in the whole genome.

	Gene	Number
rs121913529	KRAS	90
rs113488022	BRAF	49
rs112445441	KRAS	31
rs104886003	PIK3CA	30
rs28934578	TP53	26
rs121913332	APC	25
rs781215815	RNF43	22
rs121913530	KRAS	19
rs782665429	DOCK3	18
rs121913279	PIK3CA	17
rs764719749	ACVR2A	15
rs121913333	APC	14
rs11540652	TP53	13
rs772920507	BMPR2	13
rs121913527	KRAS	13
rs28934574	TP53	12
rs28934576	TP53	12
rs587781392	APC	12
rs121913273	PIK3CA	12
rs770033147	SVIL	11
rs121913331	APC	11
rs121913287	PIK3CA	11
rs397516436	TP53	10
rs62619935	APC	10
rs549924573	RCVRN	10
rs759765382	RPL22	9
rs763321097	TEAD2	9
rs372217972	QKI	9
rs767148651	HNRNPL	8
rs137854574	APC	8
rs759448855	GLI1	8
rs121912651	TP53	8
rs137854580	APC	8
rs121913343	TP53	8
rs199566425	C9orf47	7
rs377767347	SMAD4	7
rs778658185	BCORL1	7
rs756774416	LMNTD2	7
rs781677398	ZDHHC8	7
rs766438669	CSMD3	7
rs397515734	APC	7
rs775633847	RAB28	7
rs762648935	MBD6	7
rs143104828	CSF2RA	6
rs746563015	CNTLN	6
rs762448666	ELMSAN1	6
rs121913237	NRAS	6
rs121913329	APC	6
rs777980924	FHOD3	6
rs762805003	CD93	6
rs761047150	C9orf131	6
rs563361433	ONECUT1	6
rs748072217	KIAA1024	6
rs778388564	PCDH19	6

rs151073460	FUS	6
rs748349403	SLC8A2	6
rs765803753	MICAL3	6
rs748850271	PHGR1	6

The table shows mutational frequency of locus more than 5.

**Supplementary Table 3. The enrichment level of 30 immune signature in KRAS-mutated and wild-type groups were compared by Mann-Whitney test.**

<b>Immune Term</b>	<b>P Value</b>
pDCs	<b>0.000284</b>
Treg	<b>0.000637</b>
Inflammation_promoting	<b>0.002</b>
Th1_cells	<b>0.011</b>
HLA	<b>0.012</b>
T_cell_co_stimulation	<b>0.014</b>
Cytolytic_activity	<b>0.015</b>
TIL	<b>0.015</b>
T_cell_co_inhibition	<b>0.023</b>
T_helper_cells	<b>0.024</b>
Neutrophils	<b>0.026</b>
Macrophages	<b>0.033</b>
Check_point	<b>0.037</b>
DCs	0.056
Tfh	0.067
CCR	0.069
aDCs	0.076
Type_II_IFN_Response	0.083
CD4+_Regulatory_T_cells	0.132
Type_I_IFN_Response	0.158
Th2_cells	0.176
B_cells	0.205
CD8+_T_Cells	0.225
Mast_cells	0.275
APC_co_stimulation	0.289
MHC_class_I	0.468
NK_cells	0.608
iDcs	0.67
Parainflammation	0.831
APC_co_inhibition	0.998

**Supplementary Table 4. Comparing the expression of genes in TIL signature between KRAS-mutated and wild-type groups.**

**Supplementary Table 5. Comparing the expression of genes in inflammation-activities signature between KRAS-mutated and wild-type groups.**

Gene	WildMean	MutationMean	logFC	pValue
IL12A	0.282451	0.430636	0.608472	<b>3.36E-06</b>
GZMB	7.518245	13.26506	0.819164	<b>4.63E-06</b>
CXCL9	12.71954	17.27174	0.441366	<b>0.001911</b>
CXCL10	24.05325	29.22686	0.281063	<b>0.006641</b>
PRF1	3.191711	4.06926	0.350436	<b>0.009838</b>
CXCL13	4.056069	6.041906	0.574921	<b>0.020125</b>
GNLY	2.000802	2.894216	0.532595	<b>0.021077</b>
IFNG	0.27372	0.389159	0.507657	<b>0.028949</b>
STAT1	37.46403	44.23563	0.239703	<b>0.049644</b>
TBX21	0.247669	0.293948	0.247147	0.061036
IRF1	15.09839	17.20898	0.188767	0.064267
CCL5	12.90831	15.46207	0.260433	0.186282
CD8B	1.457315	1.674802	0.200678	0.279428
CD19	0.511267	0.742265	0.537859	0.283726
IL12B	0.033961	0.038595	0.184569	0.507014

**Supplementary Table 6. Detection of differentially expressed genes in KRAS-mutated and wild-type groups.**

**Supplementary Table 7. The correlation coefficients of differential expressed genes and immune score.**

**Supplementary Table 8. The annotations of differentially expressed genes with strong correlation with immune score.**

Gene names	Entry	Protein names	Length
STX11	O75558	Syntaxin-11	287
DOK2	O60496	Docking protein 2 (Downstream of tyrosine kinase 2)	412
TNFAIP8L2	Q6P589	Tumor necrosis factor alpha-induced protein 8-like protein 2	184
LCP2	Q13094	Lymphocyte cytosolic protein 2	533
LILRB1	Q8NHL6	Leukocyte immunoglobulin-like receptor subfamily B member 1	650
ABI3	Q9P2A4	ABI gene family member 3	366
LRRC25	Q8N386	Leucine-rich repeat-containing protein 25	305
SASH3	O75995	SAM and SH3 domain-containing protein 3	380
CD300A	Q9UGN4	CMRF35-like molecule 8	299
LAIR1	Q6GTX8	Leukocyte-associated immunoglobulin-like receptor 1	287
MS4A6A	Q9H2W1	Membrane-spanning 4-domains subfamily A member 6A	248
SCIMP	Q6UWF3	SLP adapter and CSK-interacting membrane protein	145
APBB1IP	Q7Z5R6	Amyloid beta A4 precursor protein-binding family B member 1-interacting protein	666
HCLS1	P14317	Hematopoietic lineage cell-specific protein	486
C3AR1	Q16581	C3a anaphylatoxin chemotactic receptor	482
IL2RA	P01589	Interleukin-2 receptor subunit alpha	272
HAVCR2	Q8TDQ0	Hepatitis A virus cellular receptor 2	301
CD300LF	Q8TDQ1	CMRF35-like molecule 1 (Immune receptor expressed on myeloid cells 1)	290
LST1	O00453	Leukocyte-specific transcript 1 protein	97
LST1	Q9Y6L6	Solute carrier organic anion transporter family member 1B1	691
SLAMF8	Q9P0V8	SLAM family member 8 (B-lymphocyte activator macrophage expressed)	285
C10orf128	Q5T292	Transmembrane protein 273	105
CD86	P42081	T-lymphocyte activation antigen CD86	329
IL10RA	Q13651	Interleukin-10 receptor subunit alpha	578
CYTH4	Q9UIA0	Cytohesin-4	394

# Heparanase is a novel biomarker for immune infiltration and prognosis in breast cancer

Wen-Jing Yang<sup>1</sup>, Lin Shi<sup>1</sup>, Xiao-Min Wang<sup>1</sup>, Guo-Wang Yang<sup>1</sup>

<sup>1</sup>Department of Oncology, Beijing Hospital of Traditional Chinese Medicine, Capital Medical University, Beijing 100010, China

**Correspondence to:** Guo-Wang Yang; **email:** [yangguowang\\_bhtcm@126.com](mailto:yangguowang_bhtcm@126.com), <https://orcid.org/0000-0002-0799-3864>

**Keywords:** heparanase, immune infiltration, prognosis, breast cancer

**Received:** October 19, 2020

**Accepted:** April 29, 2021

**Published:** August 30, 2021

**Copyright:** © 2021 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Heparanase (HPSE), an endoglycosidase that cleaves heparan sulfate, regulates a variety of biological processes that promote tumor progression. In this study, we analyzed the correlation between HPSE expression and prognosis in cancer patients, using multiple databases (Oncomine, TIMER, PrognoScan, GEPIA, Kaplan–Meier plotter, miner v4.1, DAVID). HPSE expression was significantly increased in bladder, breast, lung, and stomach cancer compared to matched normal tissues. The increased HPSE expression correlated with poor prognosis and increased immune infiltration levels of B cells, CD8+ and CD4+ T cells, macrophages, neutrophils and dendritic cells in bladder and breast cancer. In breast cancer, the high HPSE expression was associated with basal-like subtypes, younger age (0–40), advanced Scarff-Bloom-Richardson grade, Nottingham Prognostic Index and p53 mutation status. In addition, using a mouse model of breast cancer, our data showed that HPSE upregulated IL-10 expression and promoted macrophage M2 polarization and T cell exhaustion. Together, our data provide a novel immunological perspective on the mechanisms underlying breast cancer progression, and indicate that HPSE may serve as a biomarker for immune infiltration and prognosis in breast cancer.

## INTRODUCTION

Breast cancer is one of the most common tumors and the leading cause of cancer-related deaths in women within the globe [1]. Although improved the strategies for early diagnosis and treatment, the prognosis is still poor, mainly due to inherent aggressive behavior and lack of recognized treatment targets [2]. Therefore, there is an urgent need to develop more sensitive and specific biomarkers for the prognosis of breast cancer patients. In the last 2 decades, immunotherapy including programmed death-1 (PD-1), cytotoxic T lymphocyte associated antigen 4 (CTLA4), and programmed death ligand-1 (PD-L1) inhibitors, demonstrated major breakthroughs and became the major therapeutic approach in solid tumors, such as non-small-cell lung carcinoma (NSCLC) and malignant melanoma [3, 4]. Breast cancer, harboring lots of activated tumor-

infiltrating lymphocytes (TILs), is one of the most promising targets of immunotherapy among solid tumors [5–10]. Unfortunately, only a fraction of breast cancer patients responds well to immunotherapy. Since TILs serve as an independent favorable prognostic factor, and a predictive marker of chemotherapy, neoadjuvant therapy, and immunotherapy responses in breast cancer [11–17], identification of specific TILs-associated biomarkers may contribute to development of specific targeted immunotherapies in breast cancer.

The tumor microenvironment (TME), containing tumor cells and non-tumor cells, such as endothelial cells, immune cells, and fibroblasts [18], makes an important impact in tumor metastasis and progression [19–23]. Heparanase (HPSE) is the only mammalian endoglycosidase which can cleaves heparan sulfate (HS), regulates remodeling of the basement membranes

and extracellular matrix, as well as promotes the release of many HS-related molecules including cytokines, growth factors, and enzymes. HPSE is upregulated in many types of human tumors [24–27], and this elevation contributes to tumor angiogenesis, growth, metastasis, chemoresistance, and poor prognosis [28–32]. Inhibitors targeting HPSE, such as PI-88 (muparfostat), SST0001 (roneparstat), PG545 (pixatimod), and M-402 (necuparanib) have entered clinical trials. Although the role of HPSE in tumor cells has been well documented, its interaction with non-tumor cells in the TME has not been sufficiently explored. Recent studies have suggested that HPSE interaction with immune cells which contains T cells, B cells, NK cells, macrophages, neutrophils, and dendritic cells, can have both pro- and anti-tumorigenic roles, depending on the setting [33]. In addition, one research has indicated that by increasing HPSE expression in the ILs, tumors can regulate gene expression of many other tumor and non-tumor cells [34]. Thus, analyzing the interaction between HPSE, breast cancer cells, and TILs might show a novel immunological perspective to understand the mechanisms of tumor progression and further improve the clinical practice in breast cancer therapy.

In this research, we analyzed HPSE expression and the role played in the prognosis of cancer patients. In addition, we investigated HPSE association with tumor-infiltrating immune cells and related immune markers in bladder and breast cancer, and analyzed the HPSE correlation with clinicopathological parameters in breast cancer.

## RESULTS

### HPSE mRNA expression in different kinds of human cancer

To analyze HPSE expression in different kinds of cancer, HPSE mRNA expression in different tumors and matched control tissues were performed using the Oncomine database. The analysis revealed a statistically increased HPSE expression in bladder, brain, CNS, breast, gastric, leukemia, lung, lymphoma, and sarcoma tumors compared to matched normal tissues. However, a decreased HPSE expression was found in colorectal, head and neck, and esophageal cancers (Figure 1A).

To validate the Oncomine results, we analyzed HPSE expression in The Cancer Genome Atlas (TCGA) using the Timer database. As shown in Figure 1B, the HPSE expression was significantly increased in bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), esophageal carcinoma (ESCA), lung adenocarcinoma (LUAD), lung

squamous carcinoma (LUSC), stomach adenocarcinoma (STAD), and thyroid carcinoma (THCA). In contrast, the HPSE expression was decreased in colon adenocarcinoma (COAD), head and neck cancers (HNSC), kidney renal clear cell carcinoma (KIRC), kidney chromophobe (KICH), liver hepatocellular carcinoma (LIHC), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), and uterine corpus endometrial carcinoma (UCEC).

Comparison of Oncomine and Timer results indicated that the HPSE expression was significantly increased in bladder, breast, lung, and stomach cancer, while it was decreased in colon, head, and neck cancer. Thus, we next analyzed the association of HPSE expression and prognosis in the above cancers.

### High HPSE expression impacts prognosis in bladder and breast cancer

In order to investigate whether the HPSE expression correlates with prognosis in bladder, breast, gastric, lung, colorectal, head, and neck cancer patients, we used the PrognoScan, GEPIA, and Kaplan–Meier plotter databases to evaluate the impact of HPSE expression on survival. The relationships between HPSE expression and prognosis in different cancers using the PrognoScan database is shown in Table 1. Notably, HPSE expression significantly impacted prognosis in bladder and breast cancers. Analysis of HPSE expression and prognosis in different cancers using the GEPIA database showed that high HPSE expression was related to poor DFS and OS rates in bladder cancer (Figure 2B). Analysis of HPSE expression and cancer prognosis using the Kaplan–Meier plotter database showed that high HPSE expression was related to poor RFS, DMFS, PPS, and OS rates in breast cancer; it was also associated with poor PPS and OS rates in stomach cancer (Figure 2B). Therefore, it is conceivable (confirmed by at least 2 databases) that a high HPSE expression is an independent risk factor, and is associated with poor prognosis in bladder and breast cancer.

### HPSE expression correlates with infiltrating immune cells in bladder and breast cancer

TILs are an independent predict factors of survival in cancer. Therefore, we analyzed whether the HPSE expression was associated with immune infiltration levels in bladder and breast cancer. We assessed the correlation of HPSE expression with immune infiltration levels in bladder and breast cancers by the TIMER database. The results showed that in bladder cancer, the HPSE expression is not related to infiltrating levels of B cells, medium correlation with infiltrating

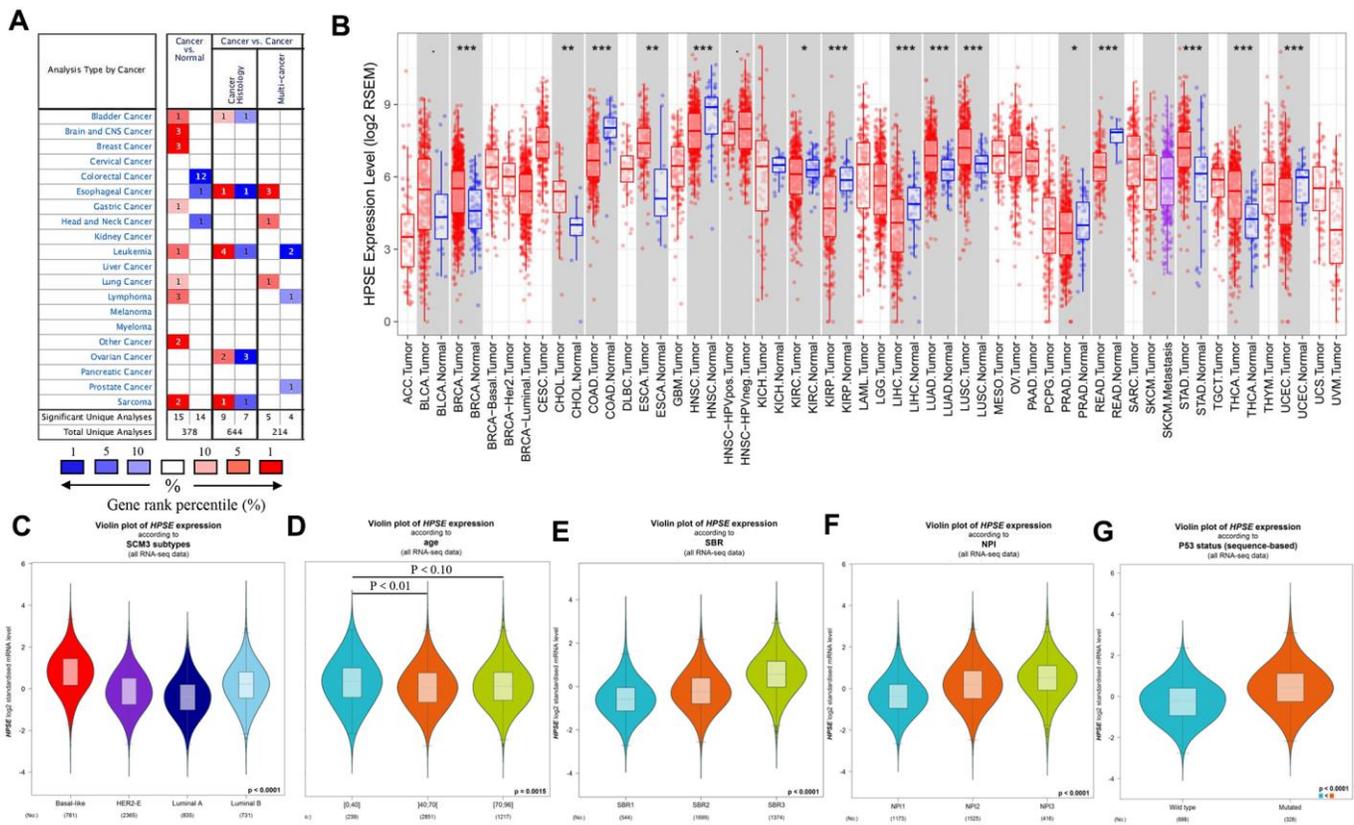
levels of CD8+ T cells, weak correlation with infiltrating levels of CD4+ T cells and macrophages, and a strong correlation with infiltrating levels of neutrophils and dendritic cells (DCs). In breast cancer, the HPSE expression showed a medium correlation with infiltrating levels of B cells, CD8+ T cells, CD4+ T cells, and macrophages, and a strong correlation with infiltrating levels of neutrophils and DCs (Figure 3).

To further explore the relationship between HPSE and the infiltrating immune cells, we investigated the association between HPSE and immune cell markers for tumor-associated macrophages (TAM), M1 macrophages, M2 macrophages, monocytes, NK cells, exhausted T cells, Tfh cells, Th1 cells, Th2 cells, Th17 cells, and Treg cells in bladder and breast cancer by the GEPIA database (Table 2). The results showed that in bladder cancer, the HPSE expression had no correlation

with Tfh cells, a weak positive correlation with TAM and Treg cells, and a weak negative correlation with M1 macrophages and Th2 cells. The HPSE expression also showed a medium positive correlation with M2 macrophages, monocytes, NK cells, exhausted T cells, and Th1 and Th17 cells in bladder cancer. In breast cancer, the HPSE expression had no correlation with Th2 cells, a weak positive correlation with M1 macrophages, NK cells, and Th17 cells, a medium positive correlation with TAM, exhausted T cells, Tfh cells, Th1 cells, and Treg cells, and a strong positive correlation with M2 macrophages and monocytes.

### HPSE expression correlates with clinicopathological parameters in breast cancer

The correlation between HPSE expression and clinicopathological parameters was explored by the bc-



**Figure 1. HPSE expression in different types of human cancers, and relationship between HPSE expression and clinicopathological parameters in breast cancer patients.** (A) Increased or decreased HPSE in data sets of different cancers compared with normal tissues in the Oncomine database. The graphic demonstrates the numbers of datasets with statistically upregulated (red) or downregulated (blue) expression of the target mRNA. The grid color is determined by the best gene rank percentile for the analyses within the grid. The Arabic number in each grid represents the number of analyses that met the criteria Gene HPSE. The gene rank was analyzed by the percentile of target genes of HPSE in the top of all genes measured. (B) Human HPSE expression in different tumor types in TCGA database determined by TIMER (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). The results are shown for the relationship between HPSE expression and SCM3 intrinsic molecular subtype (C), age (D), SBR (E), NPI (F), and P53 status (G).

**Table 1. The relationships between HPSE expression and prognosis of different cancers in PrognScan.**

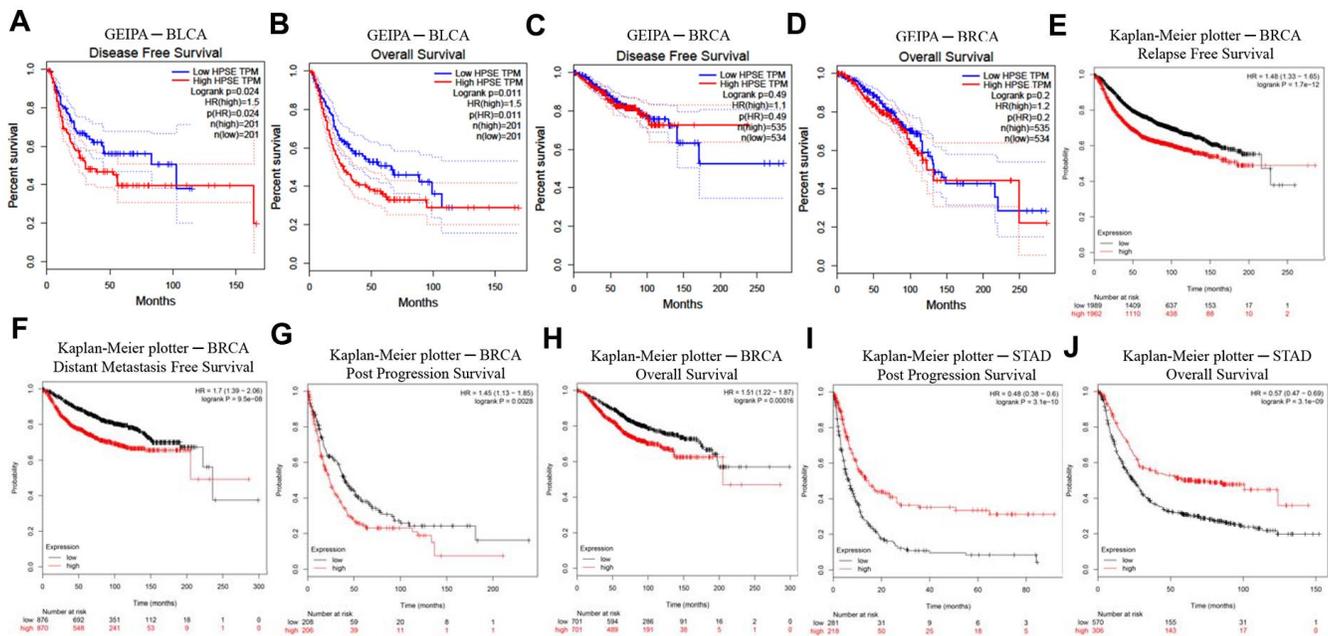
Dataset	Cancer type	Endpoint	N	P-value	HR [95% CI]
GSE13507	Bladder cancer	Disease Specific Survival	165	0.0259141	2.49 [1.12 - 5.56]
GSE1456-GPL96	Breast cancer	Disease Specific Survival	159	0.00755681	1.93 [1.19 - 3.11]
GSE1456-GPL96	Breast cancer	Overall Survival	159	0.0184731	1.60 [1.08 - 2.35]
GSE3494-GPL96	Breast cancer	Disease Specific Survival	236	0.0195956	1.50 [1.07 - 2.11]
GSE4922-GPL96	Breast cancer	Disease Free Survival	249	0.0150854	1.38 [1.06 - 1.79]

GenExMiner online tool. HPSE expression was compared with different clinicopathological parameters including intrinsic molecular subtype, age, Scarff-Bloom-Richardson (SBR) grade, Nottingham Prognostic Index (NPI), and P53 status. For intrinsic molecular subtypes, the HPSE expression in Basal-like subtype was significantly higher than that in Her2+, Luminal A, and Luminal B subtypes; the HPSE expression in Luminal B subtype was significantly higher than in Her2+ and Luminal A subtypes; and the HPSE expression in Her2+ subtype was significantly higher than in Luminal A subtype (Figure 1C). Regarding age, the HPSE expression in 0-40 years group was significantly higher than in 40-70 and 70-96 years groups; no significant difference was found between 40-70 years group and 70-96 years group

(Figure 1D). The SBR histological grade evaluates the degree of duct formation, nucleus pleomorphism, and nuclear division count, while the NPI index stratifies patients into prognostic groups according to lymph node stage, tumor size, and tumor grade. Breast cancer patients with higher SBR grade and NPI tended to express higher levels of HPSE (Figure 1E, 1F). Regarding P53 status, the HPSE expression in mutated group was significantly higher than in wild type group (Figure 1G).

**HPSE expression correlates with M2 macrophage polarization and IL-10 in breast cancer**

HPSE target genes in human breast cancer tissues are listed in Table 3; they are discriminated by R. Using the



**Figure 2.** Kaplan–Meier survival curves comparing high and low HPSE expression in different cancer types in GEIPA (A–D) and Kaplan–Meier plotter (E–J). (A, B) Survival curves of DFS and OS in BLCA. (C, D) Survival curves of DFS and OS in BRCA. (E–H) Survival curves of RFS, DMFS, PPS and OS in BRCA. (I, J) Survival curves of PPS and OS in STAD. Bladder urothelial carcinoma (BLCA); breast invasive carcinoma (BRCA); disease-free survival (DFS); distant metastasis-free survival (DMFS); disease-specific survival (DSS); heparanase (HPSE); Gene Expression Profiling Interactive Analysis (GEPIA); overall survival (OS); post progression survival (PPS); relapse-free survival (RFS); stomach adenocarcinoma (STAD).

DAVID software, we found that a total of 17 GO functions (Biological Processes) were enriched (Figure 4); this was supported by analysis of immune markers (Table 4).

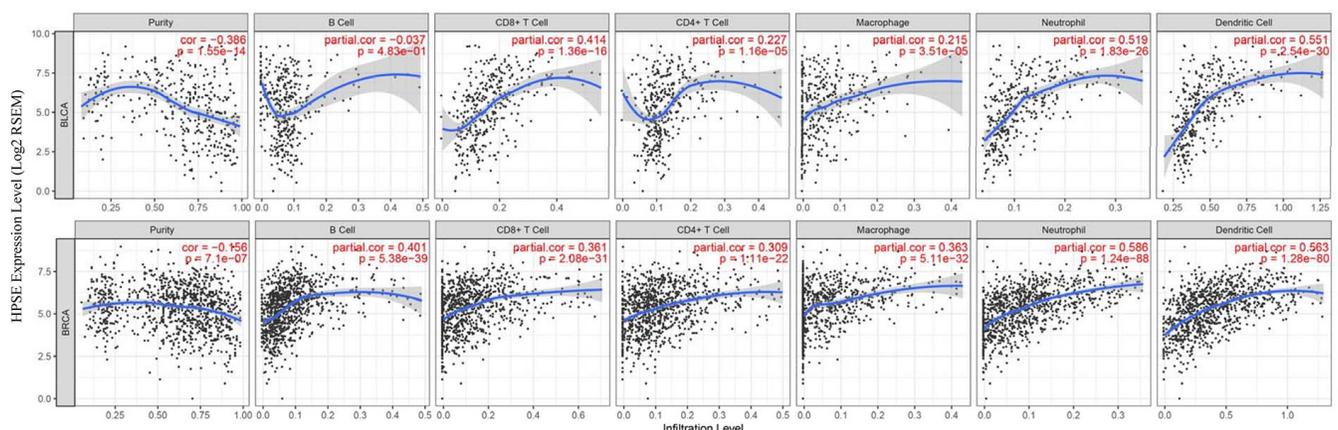
Next, using Human MDA-MB-231-HPSE and MDA-MB-231-mock cells which with high and low expression of HPSE (Figure 5A), we validated our data in a mouse model of breast cancer, using mice overexpressing HPSE and a control mock group. Because macrophages usually infiltrate at the edges of tumor tissues, we analyzed the infiltration of macrophages by immunohistochemical staining (IHC) at the tumor tissue edges. IHC staining showed that the HPSE expression correlated with an increased expression of CD163 (Figure 5B) and VSIG4 (Figure 5C), which are markers of M2 macrophages. Compared with MOCK group, M2 macrophages in HPSE group tended to infiltrate into tumor tissues. In addition, the HPSE expression related to an increased expression of IL-10 (Figure 5D), which is known to induce macrophage polarization into the M2 phenotype.

## DISCUSSION

HPSE, the only mammalian endoglycosidase that cleaves HS, contributes to tumor angiogenesis, growth, metastasis, chemoresistance, and poor prognosis in multiple tumors [24–32]. In this research, we demonstrate that the HPSE expression is obviously increased in bladder, breast, lung, and stomach cancer, but decreased in colon, head, and neck cancer. In addition, our data show that the high HPSE expression is an independent risk factor in bladder and breast cancer, indicating that

HPSE can be used as a prognostic biomarker for bladder and breast cancer. These findings are supported by previous studies demonstrating that high HPSE gene and protein levels are associated with bladder cancer invasion and metastasis [35]. Analysis of HPSE in urine from 282 individuals showed that the urine HPSE levels were elevated during bladder cancer progression [36]. HPSE mRNA levels in bladder cancer tissues related to tumor stage, histological grade, size, number, recurrence and lymph node metastasis [37]. Inhibition of HPSE expression suppressed invasion, migration and adhesion of bladder cancer cells [38]. In addition, HPSE overexpression accelerated the obesity-associated breast cancer progression [39]. However, the HPSE expression in metastatic lesions does not always reflect the expression in primary tumors. In stage I breast cancer patients, a strong HPSE staining was associated with shorter overall survival rates [40]. Tumor growth, vascularization and recurrence were significantly reduced by inhibition the procoagulant activity of HPSE [41]. Furthermore, inhibition of HPSE expression reduced tumor metastasis by reducing extracellular regulated protein kinase (ERK) and phosphorylation of protein kinase B (Akt) [42].

The immune cells in TME play a critical role in tumor progression, and are recognized as an independent predict factor of cancer survival. Moreover, the immune cells- associated HPSE has important pro-tumorigenic and anti-tumorigenic functions [33]. A previous study has indicated that the increased HPSE expression in TILs can regulate gene expression in other tumor and non-tumor cells [34], suggesting that there might be an HPSE-TME crosstalk that can affect occurrence,



**Figure 3. Correlation of HPSE expression with immune infiltration level in BLCA and BRCA.** In BLCA, HPSE expression has no correlation with infiltrating levels of B cells, medium correlation with infiltrating levels of CD8+ T cells, weak correlation with infiltrating levels of CD4+ T cells and macrophages, and a strong correlation with infiltrating levels of neutrophils and dendritic cells. In BRCA, HPSE expression has medium correlation with infiltrating levels of B cells, CD8+ T cells, CD4+ T cells, and macrophages, and a strong correlation with infiltrating levels of neutrophils and dendritic cells.

**Table 2. Correlation analysis between HPSE and relate genes and markers of immune cells in GEIPA.**

Description	Gene makers	BLCA			BRCA		
		R	P-value	Mean	R	P-value	Mean
B cell	CD19	-0.012	0.8		0.036	0.23	
	CD79A	-0.0034	0.95		0.055	0.071	
T cell	CD2	0.33	***		0.31	***	
	CD3D	0.22	***		0.18	***	
CD8+T	CD3E	0.29	***	<b>0.28(***)</b>	0.2	***	<b>0.23(***)</b>
	CD8A	0.34	***		0.25	***	
	CD8B	0.042	0.4	<b>0.34(***)</b>	0.25	***	<b>0.25(***)</b>
Dendritic cell	CD1C	-0.012	0.82		0.024	0.43	
	HLA-DPA1	0.31	***		0.39	***	
	HLA-DPB1	0.29	***		0.26	***	
	HLA-DQB1	0.23	***		0.16	***	
	HLA-DRA	0.34	***		0.41	***	
	ITGAX	0.21	***		0.34	***	
	NRP1	0.31	***	<b>0.2817(***)</b>	0.22	***	<b>0.2967(***)</b>
TAM	CCL2	0.21	***		0.28	***	
	CD68	0.46	***		0.52	***	
	IL-10	0.2	***	<b>0.29(***)</b>	0.53	***	<b>0.4433(***)</b>
M1	IRF5	-0.12	0.014		0.34	***	
	NOS2	0.013	0.8		0.082	***	
	PTGS2	0.017	0.73	<b>-0.12(0.014)</b>	0.044	0.15	<b>0.211(***)</b>
M2	CD163	0.37	***		0.55	***	
	MS4A4A	0.4	***		0.54	***	
	VSIG4	0.34	***	<b>0.37(***)</b>	0.42	***	<b>0.5033(***)</b>
Monocyte	CD86	0.44	***		0.65	***	
	CSF1R	0.38	***	<b>0.41(***)</b>	0.38	***	<b>0.515(***)</b>
Natural killer cell	KIR2DL1	0.03	0.54		-0.0055	0.86	
	KIR2DL3	0.047	0.34		0.18	***	
	KIR2DL4	0.35	***		0.33	***	
	KIR2DS4	0.03	0.55		0.08	*	
	KIR3DL1	0.071	0.16		0.19	***	
	KIR3DL2	0.088	0.077		0.21	***	
	KIR3DL3	0.035	0.48	<b>0.35(***)</b>	-0.019	0.53	<b>0.198(***)</b>
Neutrophils	CCR7	-0.21	***		0.041	0.180	
	CEACAM8	0.11	0.034		-0.0035	0.91	
	ITGAM	0.2	***	<b>0.0333(***)</b>	0.19	***	<b>0.19(***)</b>
T cell exhaustion	CTLA4	0.35	***		0.38	***	
	GZMB	0.066	0.19		0.35	***	
	HAVCR2	0.47	***		0.6	***	
	LAG3	0.4	***		0.47	***	
	PDCD1	0.27	***	<b>0.3725(***)</b>	0.3	***	<b>0.42(***)</b>
Tfh	BCL6	-0.14	*		0.036	0.240	
	IL21	0.23	***	<b>0.045(***)</b>	0.34	***	<b>0.34(***)</b>
Th1	IFNG	0.32	***		0.38	***	
	STAT1	0.46	***		0.51	***	
	STAT4	0.37	***		0.29	***	
	TBX21	0.28	***		0.31	***	
	TNF	0.26	***	<b>0.338(***)</b>	0.25	***	<b>0.348(***)</b>
Th2	GATA3	-0.33	***		-0.23	***	
	IL13	0.086	0.085		0.092	*	

	STAT5A	0.16	*		0.18	***	
Th17	STAT6	-0.091	0.068	<b>-0.017(***)</b>	-0.0018	*	<b>0.01005(***)</b>
	IL17A	-0.047	0.34		0.079	*	
Treg	STAT3	0.34	***	<b>0.34(***)</b>	0.23	***	<b>0.1545(***)</b>
	CCR8	0.28	***		0.32	***	
	FOXP3	0.36	***		0.33	***	
	STAT5B	0.059	0.23		-0.0053	0.860	
	TGFB1	0.13	0.0084	<b>0.2567(***)</b>	0.061	*	<b>0.237(***)</b>

TAM, Tumor-associated macrophage; Th, T helper cell; Tfh, Follicular helper T cell; Treg, Regulatory T cell; R value of Spearman's correlation, the absolute value of R, 0-0.09 means no correlation, 0.1-0.3 means weak correlation, 0.3-0.5 means medium correlation, 0.5-1.0 means strong correlation, + means positive correlation, - means negative correlation; MEAN means the average value of relate genes and markers of one specific immune cells with statistical significance. P-value < 0.01 was considered statistical difference, \* P-value < 0.01; \*\* P-value < 0.001; \*\*\* P-value < 0.0001.

**Table 3. Correlation analysis between HPSE and relate genes and markers of immune cells in breast cancer gene-expression miner v4.1.**

Description	Gene makers	Basal-Like		HER2+		Luminal-A		Luminal-B		TNBC	
		R	P value	R	P value	R	P value	R	P value	R	P value
TAM	CCL2	0.42	***	0.33	*	0.43	***	0.54	***	0.42	***
	CD68	0.48	***	0.4	***	0.38	***	0.39	*	0.51	***
	IL-10	0.52	***	0.44	***	0.61	***	0.54	***	0.51	***
M1	IRF5	0.31	***	0.26	0.0277	0.44	***	0.48	***		***
	NOS2	0.08	0.0553	0.32	*	0.1	0.1526	-0.04	0.7538		*
	PTGS2	-0.1	0.0259	0.04	0.7597	0.26	**	0.26	0.0361		0.2842
M2	CD163	0.61	***	0.52	***	0.64	***	0.65	***	0.67	***
	MS4A4A	0.53	***	0.5	***	0.66	***	0.66	***	0.56	***
	VSIG4	0.4	***	0.4	**	0.56	***	0.56	***	0.46	***
Monocyte	CD86	0.62	***	0.64	***	0.73	***	0.82	***	0.64	***
	CSF1R	0.4	***	0.55	***	0.59	***	0.6	***	0.42	***
T cell exhaustion	CTLA4	0.41	***	0.23	0.0492	0.33	***	0.41	**	0.38	***
	GZMB	0.42	***	0.2	0.0994	0.27	**	0.23	0.0643	0.38	***
	HAVCR2	0.6	***	0.7	***	0.75	***	0.75	***	0.64	***
	LAG3	0.56	***	0.35	0.0024	0.41	***	0.47	***	0.54	***
	PDCD1	0.41	***	0.22	0.0643	0.26	**	0.36	*	0.41	***
Th1	IFNG	0.41	***	0.26	0.0282	0.34	***	0.32	*	0.39	***
	STAT1	0.57	***	0.45	***	0.46	***	0.54	***	0.57	***
	STAT4	0.42	***	0.21	0.0731	0.33	***	0.32	0.0101	0.38	***
	TBX21	0.45	***	0.29	0.0155	0.32	***	0.29	0.0218	0.42	***
	TNF	0.24	***	0.15	0.2001	0.16	0.023	0.35	*	0.34	***

TAM, Tumor-associated macrophage; Th, T helper cell. R value of Spearman's correlation, the absolute value of R, 0-0.09 means no correlation, 0.1-0.3 means weak correlation, 0.3-0.5 means medium correlation, 0.5-1.0 means strong correlation, + means positive correlation, - means negative correlation. P-value < 0.01 was considered statistical difference, \* P-value < 0.01; \*\* P-value < 0.001; \*\*\* P-value < 0.0001.

development, and fate of tumors. Our study proved the correlation between HPSE expression and tumor-infiltrating immune cells, and supported the intimate correlation between HPSE expression and immune cell infiltration in bladder and breast cancer.

TAMs are the most plentiful immune cells within the TME [43, 44]. According to their function, macrophages can be divided into M1 and M2

phenotypes. The M1 phenotype is characterized by pro-inflammatory cytokines (IL-1 $\beta$ , TNF $\alpha$ , INF $\alpha$ , IL-6, IL-12, IL-16) and tumoricidal activity. In contrast, the M2 phenotype is characterized by anti-inflammatory cytokines (IL-4, IL-10, IL-13, TGF- $\beta$ ) that promote angiogenesis, tissue remodeling, and repair [45, 46]. Therefore, macrophages play a dual role in tumor growth: on one hand, initiating immune responses against transformed cells; and on the other hand,

promoting tumor angiogenesis and growth [47–50]. Our data showed that HPSE expression had no or weak correlation with the markers of M1 macrophage, such as NOS2, PTGS2, and IRF5, while HPSE expression had medium or strong correlation with M2 macrophage markers, such as CD163, VSIG4, and MS4A4A. In addition, IL10, a crucial anti-inflammatory cytokine that regulates M2 macrophage polarization, had a strong positive correlation with HPSE expression, indicating that HPSE overexpression might regulate the IL10-mediated M2 macrophage polarization. Our results indicated that HPSE might activate Tregs (CCR8, FOXP3, and TGFB1). In addition, HAVCR2 (TIM-3), a crucial gene that regulates T cell exhaustion, positively correlated with HPSE expression, suggesting that HPSE expression might regulate the TIM-3 mediated T cell exhaustion. These results demonstrated the important role of HPSE in regulating recruitment and activity of immune infiltrating cells in breast cancer, and suggested that HPSE might regulate immune escape in the breast cancer microenvironment.

Macrophages are often found both in primary and metastatic tumors, and contribute to tumor progression [51, 52]. In HPSE knockout mice, macrophages decreased the levels of pro-inflammatory cytokines (TNF- $\alpha$ , IL-1 $\beta$ , CXCL2, and IL6), and impaired infiltration into lung tumors, which were smaller than tumors in wild type animals [53]. In a mouse model of

pancreatic cancer, HPSE overexpression was associated with increased macrophage expression of M2 cytokines IL-6, IL-10, CCL-2, VEGF, and macrophage scavenger receptor-2 (MSR-2), increased tumor size, and increased levels of tumor-infiltrating macrophages [54]. In addition, increased HPSE expression in epithelium has been associated with inflammation and inflammation-associated tumorigenesis, such as inflammatory bowel disease (IBD) [55], pancreatitis [56], and esophageal carcinoma [57]. HPSE has been also suggested as a key regulator of the tumor microenvironment [53] and tumor progression [58]. These findings show that HPSE, regardless of its cellular source, promotes tumorigenesis.

Inflammation-induced HPSE is involved in coupling inflammation and cancer [53]. Toll-like receptors (TLRs) lie upstream of the signaling cascade that leads to cytokine induction by HPSE [59, 60]. HPSE is required for macrophage activation, crosstalk with the tumor microenvironment, and tumorigenesis; the mechanism involves HPSE-mediated TLR activation at the cell membrane, followed by Erk/p38/JNK activation and AP1-mediated transcription [61]. However, the exact mechanism of how HPSE regulates the macrophage phenotype is not understood. Our data indicate that HPSE participates in TLR signaling and JAK-STAT cascade by promoting cytokine production. Specifically, HPSE seems to promote macrophage

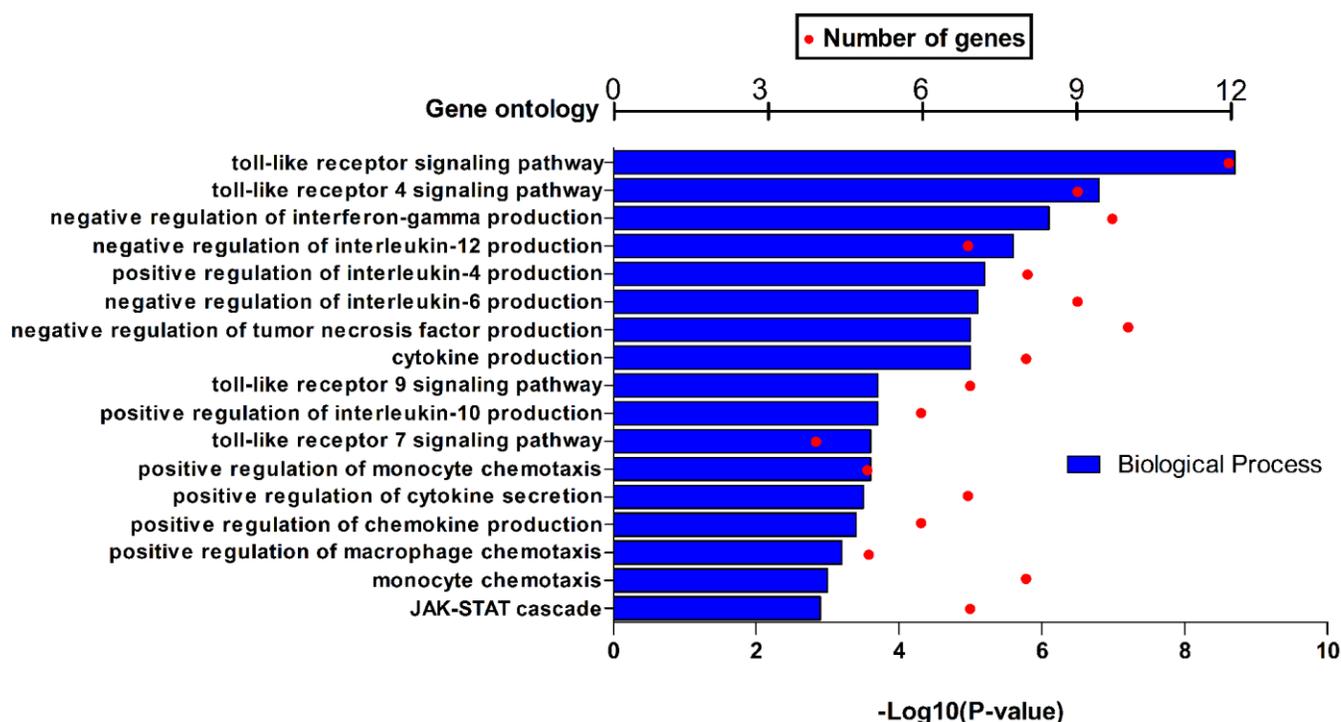
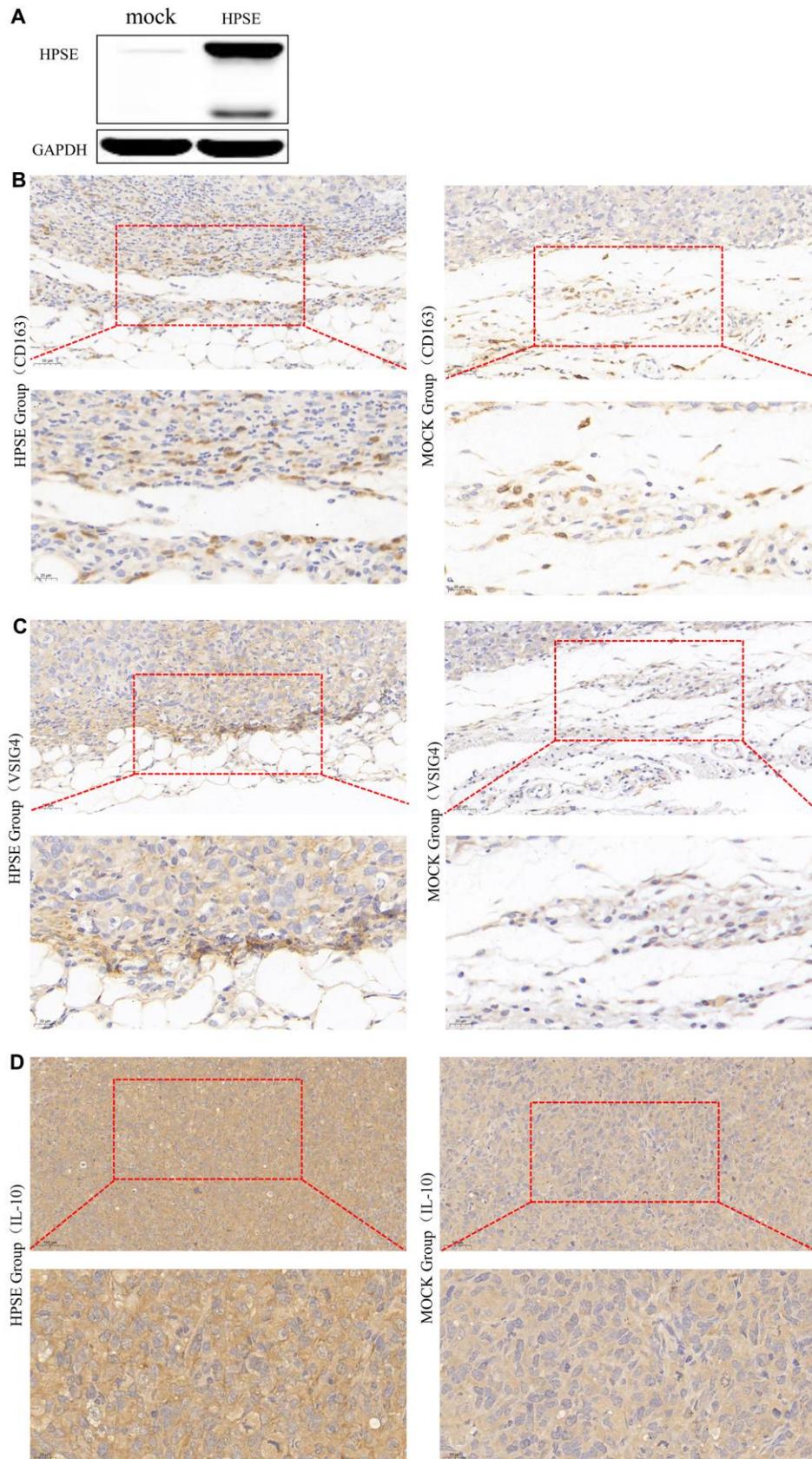


Figure 4. GO function annotations (biological process) for HPSE and target genes in BRCA in DAVID.

**Table 4. HPSE GO function annotations (biological process) for all and intrinsic molecular subtype (RIMSPC) in bc-GenExMiner v4.4.**

<b>Id</b>	<b>Description</b>	<b>Genes</b>	<b>P-value</b>
<b>Basal-like</b>			
GO:0051607	defense response to virus	DTX3L, HERC5, IFIH1, IFIT2, IFIT3, ISG15, OAS1, OAS2, OASL, PLSCR1, RSAD2, RTP4	1.92E-14
GO:0009615	response to virus	CCL8, IFIH1, IFIT2, IFIT3, OAS1, OAS2, OASL, RSAD2	1.34E-10
GO:0060337	type I interferon signaling pathway	IFIT2, IFIT3, ISG15, OAS1, OAS2, OASL, RSAD2	2.27E-10
GO:0045071	negative regulation of viral genome replication	ISG15, OAS1, OASL, PLSCR1, RSAD2	3.82E-08
GO:0060700	regulation of ribonuclease activity	OAS1, OAS2, OASL	5.53E-08
<b>HER2+</b>			
GO:0010629	negative regulation of gene expression	CCL3, CCR1, HAVCR2, LGALS9B, LGMN, MSR1	8.44E-06
GO:0030502	negative regulation of bone mineralization	CCL3, CCR1, SRGN	8.71E-06
GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	FCER1G, IFI30	1.94E-05
GO:0060333	interferon-gamma-mediated signaling pathway	FCGR1B, HLA-DQA2, HLA-H, IFI30	3.46E-05
GO:0006954	inflammatory response	ADORA3, CCL3, CCL8, CCR1, FPR3, HAVCR2, TLR1	3.67E-05
<b>Luminal A</b>			
GO:0006955	immune response	C1QC, CCR1, CD86, CTSS, CXCL10, FCGR1B, FCGR2B, FCGR2C, GPR65, HLA-DRA, IGHA1, IGHV3-33, IGHV3-53, IGKV1D-16, IGKV1D-17, IGKV1D-33, IGKV1D-42, IGKV1D-43, IGKV2D-30, IGKV3D-15, IGKV3OR2-268, IGKV5-2, IGKV6D-21, IGSF6, LST1, MARCH1, NCF4, TLR1, TLR4, TLR7, TNFSF13B, TRGC2, TRGV3	6.44E-27
GO:0045087	innate immune response	C1QA, C1QB, C1QC, CD14, CLEC7A, CORO1A, CYBB, FCER1G, HAVCR2, IFI16, IGHA1, IGHV3-33, IGHV3-35, IGHV3-53, IGHV3-74, IGHV3OR15-7, IGHV4OR15-8, LY86, LY96, LYN, MPEG1, NCF1, NCF2, RNASE6, TLR1, TLR4, TLR7, TREM2, TRGV3, TRGV9, TYROBP	3.20E-23
GO:0006911	phagocytosis	AIF1, FCER1G, FCGR2B, IGHA1, IGHV3-33, IGHV3-35, IGHV3-53, IGHV3-74, IGHV3OR15-7, IGHV4OR15-8, ITGB2, MSR1, TREM2	3.49E-14
GO:0006954	inflammatory response	AIF1, C3AR1, CCL8, CCR1, CD14, CLEC7A, CXCL10, CYBB, FCGR2B, FPR3, HAVCR2, HCK, IFI16, ITGB2, LY86, LY96, SIGLEC1, THEMIS2, TLR1, TLR4	1.10E-13
GO:0006958	complement activation	C1QA, C1QB, C1QC, IGHA1, IGHV3-33, IGHV3-35, IGHV3-53, IGHV3-74, IGHV3OR15-7, IGHV4OR15-8, IGKV1D-16, IGKV1D-33, IGKV2D-30, IGKV5-2	2.94E-13
<b>Luminal B</b>			
GO:0051607	defense response to virus	CXCL10, GBP1, HERC5, IFI16, IFI44L, IFIH1, IFIT2, IFIT3, MX1, MX2, OAS1, OAS2, OAS3, OASL, PLSCR1, RNASE6, RSAD2, TLR3, TLR7, TRIM22	4.02E-21
GO:0009615	response to virus	CCL8, IFI44, IFIH1, IFIT2, IFIT3, MX1, MX2, OAS1, OAS2, OAS3, OASL, RSAD2, TRIM22	5.31E-15
GO:0006954	inflammatory response	ADORA3, C3AR1, CCL2, CCL8, CCR1, CLEC7A, CXCL10, CXCL11, FPR3, HAVCR2, IFI16, LGALS9, LIPA, NMI, SIGLEC1, THEMIS2, TLR1, TLR4	3.42E-13
GO:0006955	immune response	CCR1, CD80, CD86, CTSC, CTSS, CXCL10, CXCL11, FCGR1B, FCGR2C, GPR65, IFI44, IFI44L, IGLV6-57, MARCH1, TLR1, TLR4, TLR7, TNFSF13B, TRIM22	5.83E-13
GO:0060337	type I interferon signaling pathway	IFIT2, IFIT3, MX1, MX2, OAS1, OAS2, OAS3, OASL, RSAD2	4.89E-11



**Figure 5.** HPSE expression between MDA-MB-231-HPSE and MDA-MB-231-mock cells (A). HPSE expression correlates with IL-10 induced M2 macrophage polarization in a mouse breast tumor model: IHC staining of (B) CD163, (C) VSIG4, and (D) IL-10.

polarization towards the M2 phenotype by suppressing interferon-gamma, IL-12, and IL-6 production, and upregulating IL-4 and IL-10 production [62–64]. In addition, our clinicopathological parameter analysis indicated that HPSE overexpression is associated with basal-like subtypes, younger age (0–40), advanced SBR grade, advanced NPI grade, and P53 mutated status.

Our previous bioinformatics and experimental studies demonstrated that HPSE promotes malignant progression, angiogenesis and metastasis in breast cancer by enhancing the crosstalk between tumor cells and platelet [65]. In this study, multi-database integration analysis indicated that high HPSE expression contributes to macrophage M2 polarization and T cell exhaustion, thus promoting tumor growth. These findings were corroborated also by IHC staining of breast tumor tissues; however, since the nude mouse model is not suitable for T cell analysis, we focused on M2 macrophage polarization. The results indicated that the protein levels of IL-10, CD163, and VSIG4 were significantly increased in breast tumor tissues in mice overexpressing HPSE, indicating that HPSE might promote macrophage M2 polarization (CD163, VSIG4) by upregulating IL-10. Together, our results show that HPSE may serve as a novel biomarker for immune infiltration and prognosis in breast cancer.

## MATERIALS AND METHODS

### Analysis of the expression of HPSE

#### *Oncomine database analysis*

The Oncomine database (<http://www.oncomine.org>) [66, 67] collects transcriptomic cancer data for biomedical study. Using the Oncomine database, the HPSE expression was compared between cancer tissues and their matched normal tissues. The threshold was: p-value  $\leq 1E-4$ , fold change  $\geq 2$ , and gene rank  $\geq$  top 10%.

#### *TIMER database analysis*

TIMER is a database incorporating 10009 samples with 23 cancer types based on TCGA (<https://cistrome.shinyapps.io/timer/>); HPSE expression in various cancers was compared between cancer tissues and their matched normal tissues.

### HPSE and clinical prognosis

#### *PrognScan database analysis*

The PrognScan (<http://www.abren.net/PrognScan/>) is an online database used to evaluate the biological relationship between gene expression and prognostic

contains overall survival (OS), relapse-free survival (RFS), distant metastasis-free survival (DMFS), disease-specific survival (DSS), and disease-free survival (DFS) in various types of cancers [68], and provide corresponding p-value, hazard ratio (HR), and 95% confidence intervals. Therefore, it has been used to analyze the correlation between the expression of HPSE and survival in different cancers with the adjusted cox p-value  $< 0.05$ .

#### *GEPIA database analysis*

Gene Expression Profiling Interactive Analysis (GEPIA) (<http://gepia.cancer-pku.cn/index.html>) is used to perform survival analysis (OS and DFS) depended on RNA sequencing data from TCGA database [69]. And in our study, was used to analyze correlation between HPSE mRNA expression and survival in various types of cancers; HR and log-rank p-values were provided, and the threshold was p-value  $< 0.05$ .

#### *Kaplan–Meier plotter database analysis*

Kaplan–Meier plotter database (<http://kmplot.com/analysis/>) is used for analyzing gene association with OS, RFS, DMFS, and post progression survival (PPS) in breast, ovarian, lung and gastric cancer [70]. And in our study, was used to identify the correlation between HPSE mRNA expression and survival in the above four cancer types. The HR and log-rank p-values were provided, and the threshold was p-value  $< 0.05$ .

### HPSE and infiltrating immune cells and markers

#### *TIMER database analysis*

TIMER (<https://cistrome.shinyapps.io/timer/>) using deconvolution statistical method to analyze the infiltration levels of immune cells including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells (DCs) based on gene expression profiles [71]. And in our study, was used to analyze the correlation between HPSE expression and the above immune infiltrating cells in bladder and breast cancers, and provided partial correlation coefficients; the threshold was p-value  $< 0.01$ . The absolute value of R, 0–0.09 meant no correlation, 0.1–0.3 meant weak correlation, 0.3–0.5 meant medium correlation, and 0.5–1.0 meant strong correlation.

#### *GEPIA database analysis*

We used GEPIA database to identify the correlation between HPSE and related genes and markers in immune cells. the threshold was p-value  $< 0.01$ ; R, 0–0.09 meant no correlation, 0.1–0.3 meant weak correlation, 0.3–0.5 meant medium correlation, and 0.5–1.0 meant strong correlation.

## HPSE in breast cancer

### *Breast cancer gene-expression miner v4.1 database analysis*

Breast Cancer Gene-Expression Miner v4.1 (bcGenExMiner v4.1) [72, 73], was utilized to evaluate the correlation between HPSE expression and clinicopathological parameters in breast cancer.

### *GO functional annotation analysis*

The database for annotation, visualization, and integrated discovery (DAVID) v.6.8 (<https://david.ncifcrf.gov>) [74] was used to perform GO [75] functional annotation analyses (positive target genes of HPSE). The background list parameter was human genome, and the threshold was p-value < 0.05.

### *In vivo studies*

Human MDA-MB-231-HPSE and MDA-MB-231-mock cells, which represented breast cancer cell lines with high and low expression of HPSE, were handled by Dr. Israel Vlodavsky, the GFP inserts were performed in Glyconovo Technologies Co., Ltd, and the HPSE expression have been detected between MDA-MB-231-HPSE and MDA-MB-231-mock cells by Western blot analysis, and the specific operation method were described previously [65].

Cell culture, animal care, and establishment of a nude mouse model were described previously [65].

The breast tissues were fixed by immersing them in 10% neutral buffered formalin at room temperature for 24 h, and the paraffin embedding process was performed. Cut the paraffin block of tumor tissue into 5µm sections for IHC staining: The tissue slides were dewaxed in xylene and rehydrated in a graded alcohol bath. Slides were immersed into EDTA antigen extraction buffer and microwaved, and then treated with 3% hydrogen peroxide in methanol to quench the activity of endogenous peroxidase, and combined with 3% bovine serum white, the proteins are incubated together to block non-specific binding. Mouse anti-CD163 (1:100; 93498S, CST), anti-IL-10 (1:100; ab34843, Abcam), and anti-VSIG4 (1:100; PA5-52018, Thermo) antibodies were incubated overnight at 4° C, and then incubation with horseradish-peroxidase (HRP)-conjugated secondary antibody. Expression of CD163, IL-10, and VSIG4 in tissues was assessed by two pathologists. The staining outcomes were assessed as the intensity on a scale of 0 to 3; 0 (no staining), 1 (weak staining), 2 (moderate staining), and 3 (strong staining). Positive tumor cell percentage was semi-quantitatively assessed on a scale of 0 to 4; 0 (none), 1 (1–25%), 2 (26–50%), 3 (51–75%), and 4 (>75%).

Histochemical score (H-score) of staining was calculated by multiplying these two variables.

## Abbreviations

HPSE: heparanase; CTLA4: cytotoxic T lymphocyte associated antigen 4; PD-1: programmed death-1; PD-L1: programmed death ligand-1; NSCLC: non-small-cell lung carcinoma; TILs: tumor-infiltrating lymphocytes; TME: Tumor microenvironment; HS: heparan sulfate; GO: gene ontology; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; CHOL: cholangiocarcinoma; ESCA: esophageal carcinoma; LUAD: lung adenocarcinoma; LUSC: Lung squamous carcinoma; STAD: stomach adenocarcinoma; THCA: thyroid carcinoma; COAD: colon adenocarcinoma; HNSC: head and neck cancer; KIRC: kidney renal clear cell carcinoma; KICH: kidney chromophobe; LIHC: liver hepatocellular carcinoma; PRAD: prostate adenocarcinoma; READ: rectum adenocarcinoma; UCEC: uterine corpus endometrial carcinoma; SBR: Scarff–Bloom–Richardson; NPI: Nottingham Prognostic Index; Akt: phosphorylation of protein kinase B; ERK: extracellular regulated protein kinase; CXCL2: C-X-C motif chemokine ligand 2; VEGF: vascular endothelial growth factor; MSR-2: macrophage scavenger receptor-2; IBD: inflammatory bowel disease; TLRs: Toll-like receptors; TCGA: The Cancer Genome Atlas; OS: overall survival; RFS: relapse-free survival; DMFS: distant metastasis-free survival; DSS: disease-specific survival; DFS: disease-free survival; HR: hazard ratio; GEPIA: Gene Expression Profiling Interactive Analysis; DCs: dendritic cells; H-score: Histochemical score.

## AUTHOR CONTRIBUTIONS

Conceptualization: Guo-Wang Yang; Methodology: Wen-Jing Yang, Lin Shi; Validation: Wen-Jing Yang; Analysis: Wen-Jing Yang, Lin Shi; Investigation: Wen-Jing Yang, Xiao-Min Wang; Data curation: Wen-Jing Yang, Lin Shi; Writing-original draft preparation: Wen-Jing Yang; Writing-review and editing: Guo-Wang Yang, Xiao-Min Wang; Supervision: Guo-Wang Yang; Project administration: Guo-Wang Yang. Funding acquisition: Guo-Wang Yang and Xiao-Min Wang.

## ACKNOWLEDGMENTS

The authors would like to thank Hai-Bo Wang (Department of Biochemistry and Molecular Biology, Capital Medical University, Beijing, China.) and Xin-Liu (Beijing Hospital of Traditional Chinese Medicine, Capital Medical University, Beijing, China) for technical support.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers 81774039, 81673924).

## REFERENCES

1. Fahad Ullah M. Breast Cancer: Current Perspectives on the Disease Status. *Adv Exp Med Biol.* 2019; 1152: 51–64.  
[https://doi.org/10.1007/978-3-030-20301-6\\_4](https://doi.org/10.1007/978-3-030-20301-6_4)  
PMID:[31456179](https://pubmed.ncbi.nlm.nih.gov/31456179/)
2. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol.* 2016; 13:674–90.  
<https://doi.org/10.1038/nrclinonc.2016.66>  
PMID:[27184417](https://pubmed.ncbi.nlm.nih.gov/27184417/)
3. Barbee MS, Ogunniyi A, Horvat TZ, Dang TO. Current status and future directions of the immune checkpoint inhibitors ipilimumab, pembrolizumab, and nivolumab in oncology. *Ann Pharmacother.* 2015; 49:907–37.  
<https://doi.org/10.1177/1060028015586218>  
PMID:[25991832](https://pubmed.ncbi.nlm.nih.gov/25991832/)
4. Ninomiya K, Hotta K. Pembrolizumab for the first-line treatment of non-small cell lung cancer. *Expert Opin Biol Ther.* 2018; 18:1015–21.  
<https://doi.org/10.1080/14712598.2018.1522300>  
PMID:[30207786](https://pubmed.ncbi.nlm.nih.gov/30207786/)
5. Stanton SE, Adams S, Disis ML. Variation in the Incidence and Magnitude of Tumor-Infiltrating Lymphocytes in Breast Cancer Subtypes: A Systematic Review. *JAMA Oncol.* 2016; 2:1354–60.  
<https://doi.org/10.1001/jamaoncol.2016.1061>  
PMID:[27355489](https://pubmed.ncbi.nlm.nih.gov/27355489/)
6. García-Teijido P, Cabal ML, Fernández IP, Pérez YF. Tumor-Infiltrating Lymphocytes in Triple Negative Breast Cancer: The Future of Immune Targeting. *Clin Med Insights Oncol.* 2016 (Suppl 1); 10:31–39.  
<https://doi.org/10.4137/CMO.S34540> PMID:[27081325](https://pubmed.ncbi.nlm.nih.gov/27081325/)
7. Nanda R, Chow LQ, Dees EC, Berger R, Gupta S, Geva R, Pusztai L, Pathiraja K, Aktan G, Cheng JD, Karantza V, Buisseret L. Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. *J Clin Oncol.* 2016; 34:2460–67.  
<https://doi.org/10.1200/JCO.2015.64.8931>  
PMID:[27138582](https://pubmed.ncbi.nlm.nih.gov/27138582/)
8. Adams S, Schmid P, Rugo HS, Winer EP, Loirat D, Awada A, Cescon DW, Iwata H, Campone M, Nanda R, Hui R, Curigiano G, Toppmeyer D, et al. Pembrolizumab monotherapy for previously treated metastatic triple-negative breast cancer: cohort A of the phase II KEYNOTE-086 study. *Ann Oncol.* 2019; 30:397–404.  
<https://doi.org/10.1093/annonc/mdy517>  
PMID:[30475950](https://pubmed.ncbi.nlm.nih.gov/30475950/)
9. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, Denkert C, Park YH, Hui R, Harbeck N, Takahashi M, Foukakis T, Fasching PA, et al, and KEYNOTE-522 Investigators. Pembrolizumab for Early Triple-Negative Breast Cancer. *N Engl J Med.* 2020; 382:810–21.  
<https://doi.org/10.1056/NEJMoa1910549>  
PMID:[32101663](https://pubmed.ncbi.nlm.nih.gov/32101663/)
10. Adams S, Diéras V, Barrios CH, Winer EP, Schneeweiss A, Iwata H, Loi S, Patel S, Henschel V, Chui SY, Rugo HS, Emens LA, Schmid P. Patient-reported outcomes from the phase III IMpassion130 trial of atezolizumab plus nab-paclitaxel in metastatic triple-negative breast cancer. *Ann Oncol.* 2020; 31:582–89.  
<https://doi.org/10.1016/j.annonc.2020.02.003>  
PMID:[32178964](https://pubmed.ncbi.nlm.nih.gov/32178964/)
11. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, Pfitzner BM, Salat C, Loi S, Schmitt WD, Schem C, Fisch K, Darb-Esfahani S, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J Clin Oncol.* 2015; 33:983–91.  
<https://doi.org/10.1200/JCO.2014.58.1967>  
PMID:[25534375](https://pubmed.ncbi.nlm.nih.gov/25534375/)
12. Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ, Wolff AC, Wood WC, Davidson NE, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol.* 2014; 32:2959–66.  
<https://doi.org/10.1200/JCO.2013.55.0491>  
PMID:[25071121](https://pubmed.ncbi.nlm.nih.gov/25071121/)
13. Dieci MV, Mathieu MC, Guarneri V, Conte P, Delalogue S, Andre F, Goubar A. Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. *Ann Oncol.* 2015; 26:1698–704.  
<https://doi.org/10.1093/annonc/mdv239>  
PMID:[25995301](https://pubmed.ncbi.nlm.nih.gov/25995301/)
14. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, Budczies J, Huober J,

- Klauschen F, Furlanetto J, Schmitt WD, Blohmer JU, Karn T, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* 2018; 19:40–50. [https://doi.org/10.1016/S1470-2045\(17\)30904-X](https://doi.org/10.1016/S1470-2045(17)30904-X) PMID:29233559
15. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, Loi S. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat Rev Clin Oncol.* 2016; 13:228–41. <https://doi.org/10.1038/nrclinonc.2015.215> PMID:26667975
  16. Pruneri G, Vingiani A, Denkert C. Tumor infiltrating lymphocytes in early breast cancer. *Breast.* 2018; 37:207–14. <https://doi.org/10.1016/j.breast.2017.03.010> PMID:28363679
  17. Borcherdig N, Kolb R, Gullicksrud J, Vikas P, Zhu Y, Zhang W. Keeping Tumors in Check: A Mechanistic Review of Clinical Response and Resistance to Immune Checkpoint Blockade in Cancer. *J Mol Biol.* 2018; 430:2014–29. <https://doi.org/10.1016/j.jmb.2018.05.030> PMID:29800567
  18. Lim YZ, South AP. Tumour-stroma crosstalk in the development of squamous cell carcinoma. *Int J Biochem Cell Biol.* 2014; 53:450–58. <https://doi.org/10.1016/j.biocel.2014.06.012> PMID:24955488
  19. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* 2013; 19:1423–37. <https://doi.org/10.1038/nm.3394> PMID:24202395
  20. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nat Rev Cancer.* 2009; 9:239–52. <https://doi.org/10.1038/nrc2618> PMID:19279573
  21. Yin Y, Yao S, Hu Y, Feng Y, Li M, Bian Z, Zhang J, Qin Y, Qi X, Zhou L, Fei B, Zou J, Hua D, Huang Z. The Immune-microenvironment Confers Chemoresistance of Colorectal Cancer through Macrophage-Derived IL6. *Clin Cancer Res.* 2017; 23:7375–87. <https://doi.org/10.1158/1078-0432.CCR-17-1283> PMID:28928161
  22. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74. <https://doi.org/10.1016/j.cell.2011.02.013> PMID:21376230
  23. Cammarota R, Bertolini V, Pennesi G, Bucci EO, Gottardi O, Garlanda C, Laghi L, Barberis MC, Sessa F, Noonan DM, Albini A. The tumor microenvironment of colorectal cancer: stromal TLR-4 expression as a potential prognostic marker. *J Transl Med.* 2010; 8:112. <https://doi.org/10.1186/1479-5876-8-112> PMID:21059221
  24. Hammond E, Khurana A, Shridhar V, Dredge K. The Role of Heparanase and Sulfatases in the Modification of Heparan Sulfate Proteoglycans within the Tumor Microenvironment and Opportunities for Novel Cancer Therapeutics. *Front Oncol.* 2014; 4:195. <https://doi.org/10.3389/fonc.2014.00195> PMID:25105093
  25. Zheng LD, Tong QS, Tang ST, Du ZY, Liu Y, Jiang GS, Cai JB. Expression and clinical significance of heparanase in neuroblastoma. *World J Pediatr.* 2009; 5:206–10. <https://doi.org/10.1007/s12519-009-0039-9> PMID:19693465
  26. Rivara S, Milazzo FM, Giannini G. Heparanase: a rainbow pharmacological target associated to multiple pathologies including rare diseases. *Future Med Chem.* 2016; 8:647–80. <https://doi.org/10.4155/fmc-2016-0012> PMID:27057774
  27. Vlodaysky I, Beckhove P, Lerner I, Pisano C, Meirovitz A, Ilan N, Elkin M. Significance of heparanase in cancer and inflammation. *Cancer Microenviron.* 2012; 5:115–32. <https://doi.org/10.1007/s12307-011-0082-7> PMID:21811836
  28. Gomes AM, Stelling MP, Pavão MS. Heparan sulfate and heparanase as modulators of breast cancer progression. *Biomed Res Int.* 2013; 2013:852093. <https://doi.org/10.1155/2013/852093> PMID:23984412
  29. Edovitsky E, Elkin M, Zcharia E, Peretz T, Vlodaysky I. Heparanase gene silencing, tumor invasiveness, angiogenesis, and metastasis. *J Natl Cancer Inst.* 2004; 96:1219–30. <https://doi.org/10.1093/jnci/djh230> PMID:15316057
  30. Yang Y, Macleod V, Bendre M, Huang Y, Theus AM, Miao HQ, Kussie P, Yaccoby S, Epstein J, Suva LJ, Kelly T, Sanderson RD. Heparanase promotes the spontaneous metastasis of myeloma cells to bone. *Blood.* 2005; 105:1303–09. <https://doi.org/10.1182/blood-2004-06-2141> PMID:15471949
  31. Shteingauz A, Boyango I, Naroditsky I, Hammond E, Gruber M, Doweck I, Ilan N, Vlodaysky I. Heparanase Enhances Tumor Growth and Chemoresistance by Promoting Autophagy. *Cancer Res.* 2015; 75:3946–57. <https://doi.org/10.1158/0008-5472.CAN-15-0037> PMID:26249176
  32. Ramani VC, Zhan F, He J, Barbieri P, Nosedà A, Tricot G, Sanderson RD. Targeting heparanase overcomes

- chemoresistance and diminishes relapse in myeloma. *Oncotarget*. 2016; 7:1598–607.  
<https://doi.org/10.18632/oncotarget.6408>  
PMID:26624982
33. Mayfosh AJ, Baschuk N, Hulett MD. Leukocyte Heparanase: A Double-Edged Sword in Tumor Progression. *Front Oncol*. 2019; 9:331.  
<https://doi.org/10.3389/fonc.2019.00331>  
PMID:31110966
34. Theodoro TR, de Matos LL, Sant Anna AV, Fonseca FL, Semedo P, Martins LC, Nader HB, Del Giglio A, da Silva Pinhal MA. Heparanase expression in circulating lymphocytes of breast cancer patients depends on the presence of the primary tumor and/or systemic metastasis. *Neoplasia*. 2007; 9:504–10.  
<https://doi.org/10.1593/neo.07241> PMID:17603633
35. Gohji K, Okamoto M, Kitazawa S, Toyoshima M, Dong J, Katsuoaka Y, Nakajima M. Heparanase protein and gene expression in bladder cancer. *J Urol*. 2001; 166: 1286–90.  
PMID:11547059
36. Shafat I, Pode D, Peretz T, Ilan N, Vlodayvsky I, Nisman B. Clinical significance of urine heparanase in bladder cancer progression. *Neoplasia*. 2008; 10:125–30.  
<https://doi.org/10.1593/neo.07875> PMID:18283334
37. Zhao W, Wang XS, Niu HT, Wang LL, Han BM, Xia SJ. Clinical relevance of heparanase mRNA expression in bladder cancer and its usefulness as a detection marker in voided urine. *Mol Med Rep*. 2009; 2:327–31.  
<https://doi.org/10.3892/mmr.00000104>  
PMID:21475833
38. Tatsumi Y, Miyake M, Shimada K, Fujii T, Hori S, Morizawa Y, Nakai Y, Anai S, Tanaka N, Konishi N, Fujimoto K. Inhibition of Heparanase Expression Results in Suppression of Invasion, Migration and Adhesion Abilities of Bladder Cancer Cells. *Int J Mol Sci*. 2020; 21:3789.  
<https://doi.org/10.3390/ijms21113789>  
PMID:32471161
39. Hermano E, Goldberg R, Rubinstein AM, Sonnenblick A, Maly B, Nahmias D, Li JP, Bakker MA, van der Vlag J, Vlodayvsky I, Peretz T, Elkin M. Heparanase Accelerates Obesity-Associated Breast Cancer Progression. *Cancer Res*. 2019; 79:5342–54.  
<https://doi.org/10.1158/0008-5472.CAN-18-4058>  
PMID:31481501
40. Vornicova O, Naroditsky I, Boyango I, Shachar SS, Mashlach T, Ilan N, Vlodayvsky I, Bar-Sela G. Prognostic significance of heparanase expression in primary and metastatic breast carcinoma. *Oncotarget*. 2017; 9:6238–44.  
<https://doi.org/10.18632/oncotarget.23560>  
PMID:29464068
41. Crispel Y, Axelman E, Tatour M, Kogan I, Nevo N, Brenner B, Nadir Y. Peptides inhibiting heparanase procoagulant activity significantly reduce tumour growth and vascularisation in a mouse model. *Thromb Haemost*. 2016; 116:669–78.  
<https://doi.org/10.1160/TH16-02-0119>  
PMID:27535490
42. Wu YJ, Hsu WJ, Wu LH, Liou HP, Pangilinan CR, Tyan YC, Lee CH. Hinokitiol reduces tumor metastasis by inhibiting heparanase via extracellular signal-regulated kinase and protein kinase B pathway. *Int J Med Sci*. 2020; 17:403–13.  
<https://doi.org/10.7150/ijms.41177>  
PMID:32132875
43. Singh R, Mishra MK, Aggarwal H. Inflammation, Immunity, and Cancer. *Mediators Inflamm*. 2017; 2017:6027305.  
<https://doi.org/10.1155/2017/6027305>  
PMID:29234189
44. Mantovani A, Sica A. Macrophages, innate immunity and cancer: balance, tolerance, and diversity. *Curr Opin Immunol*. 2010; 22:231–37.  
<https://doi.org/10.1016/j.coi.2010.01.009>  
PMID:20144856
45. Biswas SK, Sica A, Lewis CE. Plasticity of macrophage function during tumor progression: regulation by distinct molecular mechanisms. *J Immunol*. 2008; 180:2011–17.  
<https://doi.org/10.4049/jimmunol.180.4.2011>  
PMID:18250403
46. De Palma M, Lewis CE. Macrophage regulation of tumor responses to anticancer therapies. *Cancer Cell*. 2013; 23:277–86.  
<https://doi.org/10.1016/j.ccr.2013.02.013>  
PMID:23518347
47. Biswas SK, Mantovani A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nat Immunol*. 2010; 11:889–96.  
<https://doi.org/10.1038/ni.1937>  
PMID:20856220
48. Zebedin E, Freissmuth M, Sexl V. Signal interception-based therapies—a double-edged sword in Bcr/abl-induced malignancies? *Leuk Lymphoma*. 2008; 49: 620–24.  
<https://doi.org/10.1080/10428190701858872>  
PMID:18398721
49. Soloviev DA, Hazen SL, Szpak D, Bledzka KM, Ballantyne CM, Plow EF, Pluskota E. Dual role of the leukocyte integrin  $\alpha$ M $\beta$ 2 in angiogenesis. *J Immunol*. 2014; 193:4712–21.  
<https://doi.org/10.4049/jimmunol.1400202>  
PMID:25261488

50. Mills CD, Lenz LL, Harris RA. A Breakthrough: Macrophage-Directed Cancer Immunotherapy. *Cancer Res.* 2016; 76:513–16.  
<https://doi.org/10.1158/0008-5472.CAN-15-1737>  
PMID:[26772756](https://pubmed.ncbi.nlm.nih.gov/26772756/)
51. Nielsen SR, Schmid MC. Macrophages as Key Drivers of Cancer Progression and Metastasis. *Mediators Inflamm.* 2017; 2017:9624760.  
<https://doi.org/10.1155/2017/9624760>  
PMID:[28210073](https://pubmed.ncbi.nlm.nih.gov/28210073/)
52. Cersosimo F, Lonardi S, Bernardini G, Telfer B, Mandelli GE, Santucci A, Vermi W, Giurisato E. Tumor-Associated Macrophages in Osteosarcoma: From Mechanisms to Therapy. *Int J Mol Sci.* 2020; 21:5207.  
<https://doi.org/10.3390/ijms21155207>  
PMID:[32717819](https://pubmed.ncbi.nlm.nih.gov/32717819/)
53. Gutter-Kapon L, Alishekevitz D, Shaked Y, Li JP, Aronheim A, Ilan N, Vlodavsky I. Heparanase is required for activation and function of macrophages. *Proc Natl Acad Sci USA.* 2016; 113:E7808–17.  
<https://doi.org/10.1073/pnas.1611380113>  
PMID:[27849593](https://pubmed.ncbi.nlm.nih.gov/27849593/)
54. Hermano E, Meirovitz A, Meir K, Nussbaum G, Appelbaum L, Peretz T, Elkin M. Macrophage polarization in pancreatic carcinoma: role of heparanase enzyme. *J Natl Cancer Inst.* 2014; 106:dju332.  
<https://doi.org/10.1093/jnci/dju332>  
PMID:[25326645](https://pubmed.ncbi.nlm.nih.gov/25326645/)
55. Lerner I, Hermano E, Zcharia E, Rodkin D, Bulvik R, Doviner V, Rubinstein AM, Ishai-Michaeli R, Atzmon R, Sherman Y, Meirovitz A, Peretz T, Vlodavsky I, Elkin M. Heparanase powers a chronic inflammatory circuit that promotes colitis-associated tumorigenesis in mice. *J Clin Invest.* 2011; 121:1709–21.  
<https://doi.org/10.1172/JCI43792> PMID:[21490396](https://pubmed.ncbi.nlm.nih.gov/21490396/)
56. Khamaysi I, Singh P, Nasser S, Awad H, Chowers Y, Sabo E, Hammond E, Gralnek I, Minkov I, Nosedá A, Ilan N, Vlodavsky I, Abassi Z. The Role of Heparanase in the Pathogenesis of Acute Pancreatitis: A Potential Therapeutic Target. *Sci Rep.* 2017; 7:715.  
<https://doi.org/10.1038/s41598-017-00715-6>  
PMID:[28386074](https://pubmed.ncbi.nlm.nih.gov/28386074/)
57. Brun R, Naroditsky I, Waterman M, Ben-Izhak O, Groisman G, Ilan N, Vlodavsky I. Heparanase expression by Barrett's epithelium and during esophageal carcinoma progression. *Mod Pathol.* 2009; 22:1548–54.  
<https://doi.org/10.1038/modpathol.2009.115>  
PMID:[19749739](https://pubmed.ncbi.nlm.nih.gov/19749739/)
58. Bhattacharya U, Gutter-Kapon L, Kan T, Boyango I, Barash U, Yang SM, Liu J, Gross-Cohen M, Sanderson RD, Shaked Y, Ilan N, Vlodavsky I. Heparanase and Chemotherapy Synergize to Drive Macrophage Activation and Enhance Tumor Growth. *Cancer Res.* 2020; 80:57–68.  
<https://doi.org/10.1158/0008-5472.CAN-19-1676>  
PMID:[31690669](https://pubmed.ncbi.nlm.nih.gov/31690669/)
59. Lesina M, Kurkowski MU, Ludes K, Rose-John S, Treiber M, Klöppel G, Yoshimura A, Reindl W, Sipos B, Akira S, Schmid RM, Algül H. Stat3/Socs3 activation by IL-6 transsignaling promotes progression of pancreatic intraepithelial neoplasia and development of pancreatic cancer. *Cancer Cell.* 2011; 19:456–69.  
<https://doi.org/10.1016/j.ccr.2011.03.009>  
PMID:[21481788](https://pubmed.ncbi.nlm.nih.gov/21481788/)
60. Blich M, Golan A, Arvatz G, Sebbag A, Shafat I, Sabo E, Cohen-Kaplan V, Petcherski S, Avniel-Polak S, Eitan A, Hammerman H, Aronson D, Axelman E, et al. Macrophage activation by heparanase is mediated by TLR-2 and TLR-4 and associates with plaque progression. *Arterioscler Thromb Vasc Biol.* 2013; 33:e56–65.  
<https://doi.org/10.1161/ATVBAHA.112.254961>  
PMID:[23162016](https://pubmed.ncbi.nlm.nih.gov/23162016/)
61. Goodall KJ, Poon IK, Phipps S, Hulett MD. Soluble heparan sulfate fragments generated by heparanase trigger the release of pro-inflammatory cytokines through TLR-4. *PLoS One.* 2014; 9:e109596.  
<https://doi.org/10.1371/journal.pone.0109596>  
PMID:[25295599](https://pubmed.ncbi.nlm.nih.gov/25295599/)
62. Campbell MJ, Tonlaar NY, Garwood ER, Huo D, Moore DH, Khramtsov AI, Au A, Baehner F, Chen Y, Malaka DO, Lin A, Adeyanju OO, Li S, et al. Proliferating macrophages associated with high grade, hormone receptor negative breast cancer and poor clinical outcome. *Breast Cancer Res Treat.* 2011; 128:703–11.  
<https://doi.org/10.1007/s10549-010-1154-y>  
PMID:[20842526](https://pubmed.ncbi.nlm.nih.gov/20842526/)
63. Kratochvill F, Neale G, Haverkamp JM, Van de Velde LA, Smith AM, Kawauchi D, McEvoy J, Roussel MF, Dyer MA, Qualls JE, Murray PJ. TNF Counterbalances the Emergence of M2 Tumor Macrophages. *Cell Rep.* 2015; 12:1902–14.  
<https://doi.org/10.1016/j.celrep.2015.08.033>  
PMID:[26365184](https://pubmed.ncbi.nlm.nih.gov/26365184/)
64. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. *Nat Rev Clin Oncol.* 2017; 14: 399–416.  
<https://doi.org/10.1038/nrclinonc.2016.217>  
PMID:[28117416](https://pubmed.ncbi.nlm.nih.gov/28117416/)
65. Yang WJ, Zhang GL, Cao KX, Liu XN, Wang XM, Yu MW, Li JP, Yang GW. Heparanase from triple-negative breast

- cancer and platelets acts as an enhancer of metastasis. *Int J Oncol*. 2020; 57:890–904.  
<https://doi.org/10.3892/ijo.2020.5115> PMID:32945393
66. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6:1–6.  
[https://doi.org/10.1016/s1476-5586\(04\)80047-2](https://doi.org/10.1016/s1476-5586(04)80047-2)  
PMID:15068665
67. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007; 9:166–80.  
<https://doi.org/10.1593/neo.07112> PMID:17356713
68. Mizuno H, Kitada K, Nakai K, Sarai A. PrognScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics*. 2009; 2:18.  
<https://doi.org/10.1186/1755-8794-2-18>  
PMID:19393097
69. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017; 45:W98–102.  
<https://doi.org/10.1093/nar/gkx247>  
PMID:28407145
70. Lániczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, Gyórfy B. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat*. 2016; 160:439–46.  
<https://doi.org/10.1007/s10549-016-4013-7>  
PMID:27744485
71. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res*. 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
PMID:29092952
72. Jézéquel P, Campone M, Gouraud W, Guérin-Charbonnel C, Leux C, Ricolleau G, Campion L. bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Res Treat*. 2012; 131:765–75.  
<https://doi.org/10.1007/s10549-011-1457-7>  
PMID:21452023
73. Jézéquel P, Frénel JS, Campion L, Guérin-Charbonnel C, Gouraud W, Ricolleau G, Campone M. bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses. *Database (Oxford)*. 2013; 2013:bas060.  
<https://doi.org/10.1093/database/bas060>  
PMID:23325629
74. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003; 4:P3.  
PMID:12734009
75. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–29.  
<https://doi.org/10.1038/75556>  
PMID:10802651

## PPM1D is a potential prognostic biomarker and correlates with immune cell infiltration in hepatocellular carcinoma

Zhangtao Yu<sup>1,2,3,\*</sup>, Yinghui Song<sup>1,2,3,4,\*</sup>, Mengting Cai<sup>5</sup>, Bo Jiang<sup>1,2,3</sup>, Zhihua Zhang<sup>1,2,3</sup>, Le Wang<sup>1,2,3</sup>, Yu Jiang<sup>1,2,3</sup>, Lianhong Zou<sup>6</sup>, Xiehong Liu<sup>6</sup>, Nanhui Yu<sup>6</sup>, Xianhai Mao<sup>1,2,3</sup>, Chuang Peng<sup>1,2,3</sup>, Sulai Liu<sup>1,2,3,4</sup>

<sup>1</sup>Department of Hepatobiliary Surgery, Hunan Research Center of Biliary Disease, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, , Changsha 410005, Hunan Province, China

<sup>2</sup>Clinical Medical Technology Research Center of Hunan Provincial for Biliary Disease Prevention and Treatment, Changsha 410005, Hunan Province, China

<sup>3</sup>Biliary Disease Research Laboratory of Hunan Provincial People's Hospital, Key Laboratory of Hunan Normal University, Changsha 410005, Hunan Province, China

<sup>4</sup>Central Laboratory of Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha 410005, Hunan Province, China

<sup>5</sup>Department of Nuclear Medicine, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha 410005, Hunan Province, China

<sup>6</sup>Hunan Provincial Institute of Emergency Medicine, The First Affiliated Hospital of Hunan Normal University, Hunan Provincial People's Hospital, Changsha 410005, Hunan Province, China

\*Equal contribution

**Correspondence to:** Sulai Liu; email: [liusulai@hunnu.edu.cn](mailto:liusulai@hunnu.edu.cn)

**Keywords:** PPM1D, Wip1, prognostic biomarker, immune infiltrates, hepatocellular carcinoma

**Received:** November 5, 2020

**Accepted:** August 10, 2021

**Published:** September 1, 2021

**Copyright:** © 2021 Yu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**Background:** Protein phosphatase magnesium-dependent 1 delta (PPM1D), also referred to as wild-type p53-induced phosphatase 1 (Wip1) or protein phosphatase 2C delta (PP2Cδ), is an oncogenic nuclear serine/threonine phosphatase belonging to the PP2C family. However, the knowledge regarding PPM1D mRNA expression, tumor immunity, and the prognosis in hepatocellular carcinoma (HCC) is scanty.

**Methods:** We analyzed *PPM1D*, including its expression in both the normal and tumor tissue using the Sangerbox database and Tumor Immune Estimation Resource (TIMER). We evaluated its correlation with prognosis in different tumor types by the Kaplan-Meier plotter and Gene Expression Profiling Interactive Analysis (GEPIA). The correlations between PPM1D and the cancer immune infiltrates were determined using TIMER. The correlations between PPM1D expression and gene marker sets of the immune infiltrates were established by both the TIMER and GEPIA. Immunohistochemistry was performed to detect the expression of Wip1 protein encoded by PPM1D in HCC, and the relationship between Wip1 expression and the prognosis of HCC were analyzed.

**Results:** We found out that PPM1D mRNA expression was significantly higher in several human cancers, including HCC, than in the corresponding normal human tissues. The PPM1D mRNA high expression in HCC was significantly correlated with poor prognosis. The expression was associated with progression-free survival (PFS) in multiple HCC patients' cohorts (PFS HR = 1.5,  $P = 0.0066$ ). This was especially in early stage (stage 1) and AJCC\_T 1 of HCC. Besides, PPM1D mRNA expression indicated a positive correlation with tumor-infiltrating

Monocytes, tumor-associated macrophages (TAMs), M1 Macrophage, M2 Macrophage, dendritic cells (DCs), T-helper (Th) and Treg. Wip1 was higher in HCC than paracancerous tissue. High expression of Wip1 was associated with poor prognosis of HCC.

**Conclusion:** Our findings suggested that PPM1D mRNA is critical in activating tumor immunity. Besides, they implied that PPM1D could be a potential prognostic biomarker for cancer progression. Moreover, it correlated with tumor immune cell infiltration in HCC.

## INTRODUCTION

Globally, hepatocellular carcinoma (HCC) ranks fifth in the incidence and ranks second in leading cause of cancer-related deaths of all malignant tumors [1]. There are many factors that can cause HCC, hepatitis B is well known as the main risk factor of HCC, and there are other risk factors such as alcoholic liver disease, hepatitis C and carcinogens for instance aflatoxin [2–5]. The sole radical treatment for HCC is surgical resection [6, 7]. However, most HCC patients are normally diagnosed at intermediate to advanced stages. As such they become ineligible for radical therapies. Besides, the rate of recurrence is high even with surgical resection [7]. At present, it is thus important to determine original prognostic markers and therapeutic targets for HCC.

*PPM1D* gene, encoding Wip1 phosphatase, is expressed in neutrophils, stem cells, macrophages, hematopoietic progenitors, B and T lymphocytes in peripheral blood and bone marrow [8]. Previous studies indicate that PPM1D is amplified and overexpressed in various tumors and it is hence considered an oncogene [9, 10]. PPM1D is a phosphatase which promote growth and its numerous downstream targets are important tumor promoting factors. In our previous study, we found PPM1D was overexpressed in renal clear cell carcinoma and intrahepatic cholangiocarcinoma. Its overexpression is related to poor prognosis [11–14]. It is also reported that PPM1D is one of underlying prognostic biomarkers and treatment targets for HCC [15]. However, its functions and mechanisms of HCC progression is unknown.

Based on the expression of specific markers, we studied the expression of PPM1D, its association with prognosis of HCC, while level of the various tumor-infiltrating immune cells (TIICs) in this research. Tumor Immune Estimation Resource (TIMER), Gene Expression Profiling Interactive Analysis (GEPIA) and the Kaplan-Meier plotter databases were applied in the analyses above. Our results revealed the vital contribution of PPM1D in HCC prognosis. Besides, they implied that expression of PPM1D might regulate tumor immunity through modulating the infiltration of the immune cells of HCC.

## MATERIALS AND METHODS

### PPM1D gene expression database analysis

PPM1D mRNA status of various tumors including HCC and the corresponding normal tissues were identified from the sangerbox database (<http://sangerbox.com/Tool>). The threshold was established with respect to the following values: P-value of 0.001, fold change of 1.5.

### Kaplan-meier plotter database analysis

Our researchers used Kaplan-Meier plotter to establish the correlation between the expression of genes and the survival rates of 21 different tumors from >10,000 tumor samples. We used the Kaplan-Meier plotter to analyse the correlation between PPM1D gene expression and the survival rates of the lung, ovarian, breast, gastric, pancreatic and liver tumors. We select the required conditions according to the default settings to get the required results; and this was foundation of hazard ratios (HR), with log-rank P-values and 95% confidence intervals.

### TIMER database analysis

TIMER (<https://cistrome.shinyapps.io/timer/>) is a systematic database, and we used it to analyse the TIICs in 32 tumor types from >10,000 tumor samples in The Cancer Genome Atlas (TCGA) database systematically (<https://cistrome.shinyapps.io/timer/>). Foundation on the investigation of the gene expression profiles, TIMER inferred the abundance of the TIICs. PPM1D expression in various tumor types was studied. We also determined the relationship between the status of PPM1D gene and the plenty of the infiltrating immune cells (macrophages, CD8+ T cells, CD4+ T cells, dendritic cells (DCs), B cells and neutrophils), according to the expression of the specific marker gene in various tumors inclusive of HCC. The gene markers for the TIICs included the markers for monocytes, neutrophils, natural killer (NK) cells, tumor-associated macrophages (TAMs), DCs, CD8+ T cells, T cells (general), B cells, T-helper 2 (Th2) cells, T-helper 1 (Th1) cells, M1 macrophages, M2 macrophages, exhausted T cells, T-helper 17 (Th17) cells, follicular helper T (Tfh) cells

and regulatory T cells (Tregs). These gene markers were consulted in the prior researches [16–19].

### Gene correlation analysis in GEPIA

GEPIA database (<http://gepia.cancer-pku.cn/index.html>) confirmed the significantly correlated genes in TIMER. It aided in analyzing the RNA expression data of the GTEx and TCGA projects inclusive of 8,587 normal and 9,736 cancer tissue samples. Besides, GEPIA helped in generating the survival curves, which includes disease-free survival (DFS) and overall survival (OS) rates.

### Immunohistochemistry

The tissues were fixed by formalin, embedded by paraffin, and sliced into thin sections. Then, we immunostained the tissue sections with a polyclonal anti-Wip1 antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA), by using standard techniques to detect the expression of Wip1 protein [11]. Afterwards, all tissue sections were read by optical microscopies and evaluated for  $\geq 5$  fields at a  $\times 400$  magnification independently by more than two pathological professors who were not conscious of any outcome or clinical data. While the cell membrane and cytoplasm stained brown, the tissue sections were considered as wip1 positive. All images were captured by microscopic imaging system (Nikon E1000, Tokyo, Japan).

### Statistical analysis

Based on the Sangerbox and TIMER database, the gene expression analysis entailed determination of the ranks, fold changes and *P*-values. We used Kaplan-Meier plots to generate the survival curves. Gene expression corrections were performed in the GEPIA and TIMER databases by Spearman's correlation analysis. The statically significant difference was considered when  $p < 0.05$ .

### Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the ethical committee of Hunan Normal University.

### Availability of data and materials

Data and materials are included in the manuscript.

## RESULTS

### The levels of PPM1D mRNA in various human tumors

To determine the differences between PPM1D mRNA expression in normal and tumor tissues, the status of PPM1D mRNA were analyzed through Sangerbox and TIMER database. The results of the analyses demonstrated that the PPM1D expression was higher in various tumors, such as esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), colon adenocarcinoma (COAD), cholangiocarcinoma (CHOL), adrenocortical carcinoma (ACC), stomach adenocarcinoma (STAD), lung adenocarcinoma (LUAD), pancreatic adenocarcinoma (PAAD), brain lower grade glioma (LGG), breast invasive carcinoma (BRCA), hepatocellular carcinoma (HCC), acute myeloid leukemia (LAML), skin cutaneous melanoma (SKCM), testicular germ cell tumors (TGCT), prostate adenocarcinoma (PRAD), kidney renal papillary cell carcinoma (KIRP), kidney renal clear cell carcinoma (KIRC), thyroid carcinoma (THCA) and uterine carcinosarcoma (UCS) tissue; and they were significantly lower in rectum adenocarcinoma (READ), uterine corpus endometrial carcinoma (UCEC), lung squamous cell carcinoma (LUSC) and ovarian serous cystadenocarcinoma (OV) than in the normal tissues, which were shown in Figure 1A. The analysis of TCGA RNA-seq data demonstrated that mRNA expression of PPM1D was significantly lower in BRCA, kidney chromophobe (KICH), lung squamous cell carcinoma (LUSC), COAD, LUAD, THCA, and UCEC comparing with the corresponding normal tissue. Though, PPM1D expression was remarkably higher in HCC in comparison to the corresponding normal tissue (Figure 1B). However, there was difference in the expression of PPM1D in some tumors, such as BRCA, COAD, LUAD and THCA, between the two databases. The results of the two databases consistently show that PPM1D was significantly extremely expressed in CHOL and HCC.

### Prognostic significance of PPM1D expression in cancers

High PPM1D expressions were not associated with OS in HCC (Figure 2A), however the mRNA expressions of PPM1D was significant related with poor prognosis in HCC (PFS: HR = 1.5, 95% CI = 1.12 to 2.02,  $P = 0.0066$ ; Figure 2B). However, PFS in PAAD (Figure 2F) have no associated with mRNA expression of PPM1D. In gastric cancer, the high PPM1D mRNA expression was significant correlated with better prognosis (Figure 2C, 2D and 2E). These findings imply that the prognostic significance of the status of PPM1D expressions in different tumors is inconsistent (Figure 2).

In some tumors, high expression of PPM1D indicates a poor prognosis such as MESO, high expression of PPM1D in some cancers indicates a good prognosis such as KIRC, and the expression level of PPM1D in some cancers has no correlation with the prognosis such as HNSC (Supplementary Figure 1). The above results suggest that the prognostic significance of PPM1D mRNA expression in various tumors depends on tumor types.

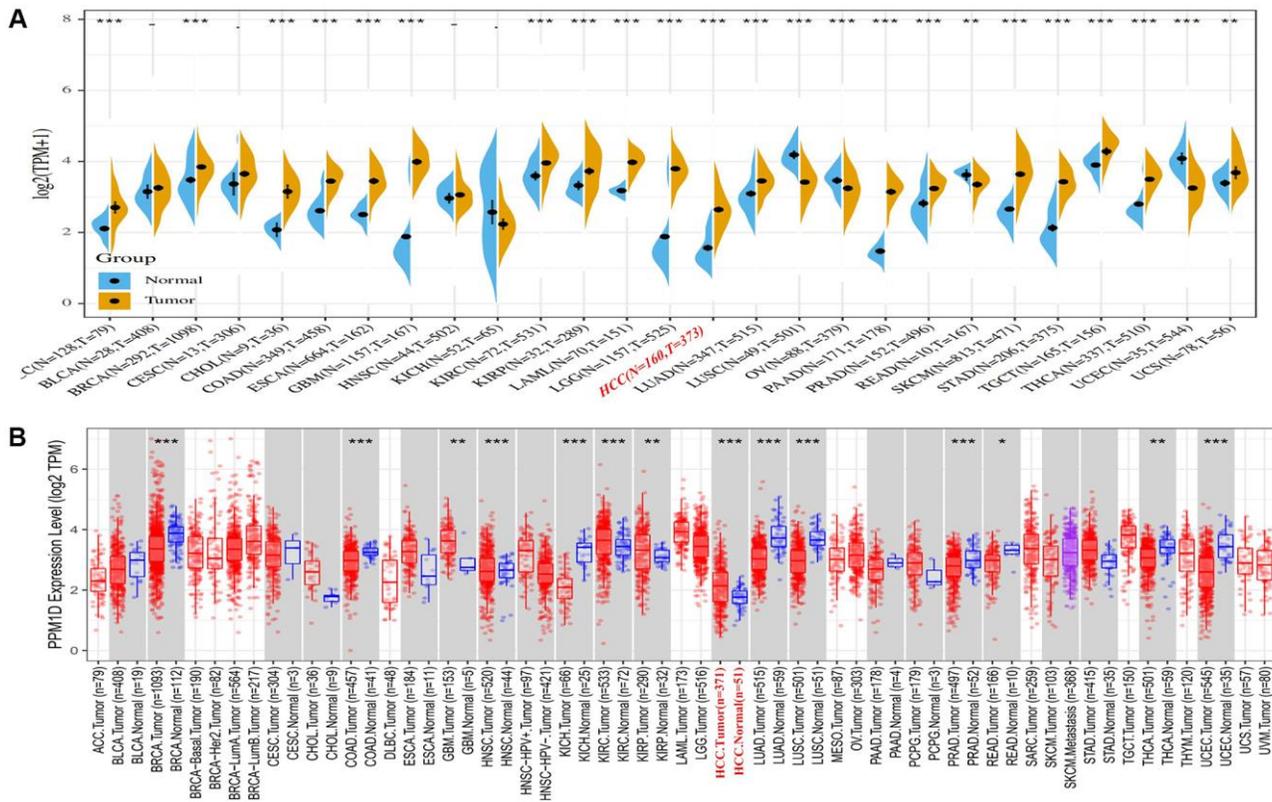
### Association of PPM1D expressions with different clinical HCC patients through Kaplan-Meier plotter

Higher expression of PPM1D associated with poor OS, PFS, RFS and DSS (Table 1). Combined with univariate and multivariate analyses of clinicopathological factors affecting PFS of HCC patients (Table 2), we found PPM1D expression was associated with male, non-vascular infiltrating, Asian, and hepatitis virus-infected

of HCC patients. In particular, the high PPM1D mRNA expression correlated with worse disease-specific (DSS), recurrence-free (RFS) and progression-free survival (PFS) in stage 1 and AJCC-T 1 of HCC patients. However, the expression of PPM1D does not seem to be related to the prognosis for the middle and late stages of HCC (Table 1).

### The association of the levels of PPM1D with the infiltration level of immune cells in HCC

Based on TIMER database, we analyzed the association of PPM1D mRNA expression with the infiltrating immune cells in 8 types of digestive tumors inclusive of HCC. These results showed the expression of PPM1D mRNA remarkably associated with tumor purity, CD8+ T cells, macrophages, CD4+ T cells, DCs and B cells of them. (Figure 3 and Supplementary Figure 2).



**Figure 1. PPM1D expression in different types of human cancers.** (A) High or low expression of PPM1D in different human cancer tissues compared with normal tissues using the Sangerbox database. (B) The level of PPM1D expression in different tumor types from the TCGA database in TIMER. Note: \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Abbreviations: ACC: Adrenocortical carcinoma; BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; CHOL: Cervical and endocervical cancers(CESC), Cholangiocarcinoma; COAD: Colon adenocarcinoma; DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LAML: Acute Myeloid Leukemia; LGG: Brain Lower Grade Glioma; HCC: Hepatocellular Carcinoma; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; MESO: Mesothelioma; OV: Ovarian serous cystadenocarcinoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach adenocarcinoma; TGCT: Testicular Germ Cell Tumors; THCA: Thyroid carcinoma; UCEC: Uterine Corpus Endometrial Carcinoma; UCS: Uterine Carcinosarcoma; UVM: Uveal Melanoma.

We also detected that higher PPM1D mRNA expression related with higher immune cells infiltration and worse prognosis in HCC (Figure 3). The PPM1D mRNA expression positively associated with the infiltration rates of the CD4+ T cells ( $r = 0.329$ ,  $P = 4.06e-10$ ), CD8+ T cells ( $r = 0.237$ ,  $P = 9.15e-06$ ), B cells ( $r = 0.276$ ,  $P = 2.00e-07$ ), macrophages ( $r = 0.402$ ,  $P = 1.16e-14$ ), and DCs ( $r = 0.385$ ,  $p = 1.88e-13$ ) in HCC tissues. These findings indicate that PPM1D is vital in immune cells infiltration in HCC, particularly in macrophages. Although these findings indicate differences in the level of tumor infiltration by the immune cell, the mRNA expression of PPM1D and prognosis in various tumors, the results generally imply that mRNA expression of PPM1D modulates the infiltration of the immune cells to cancer tissues.

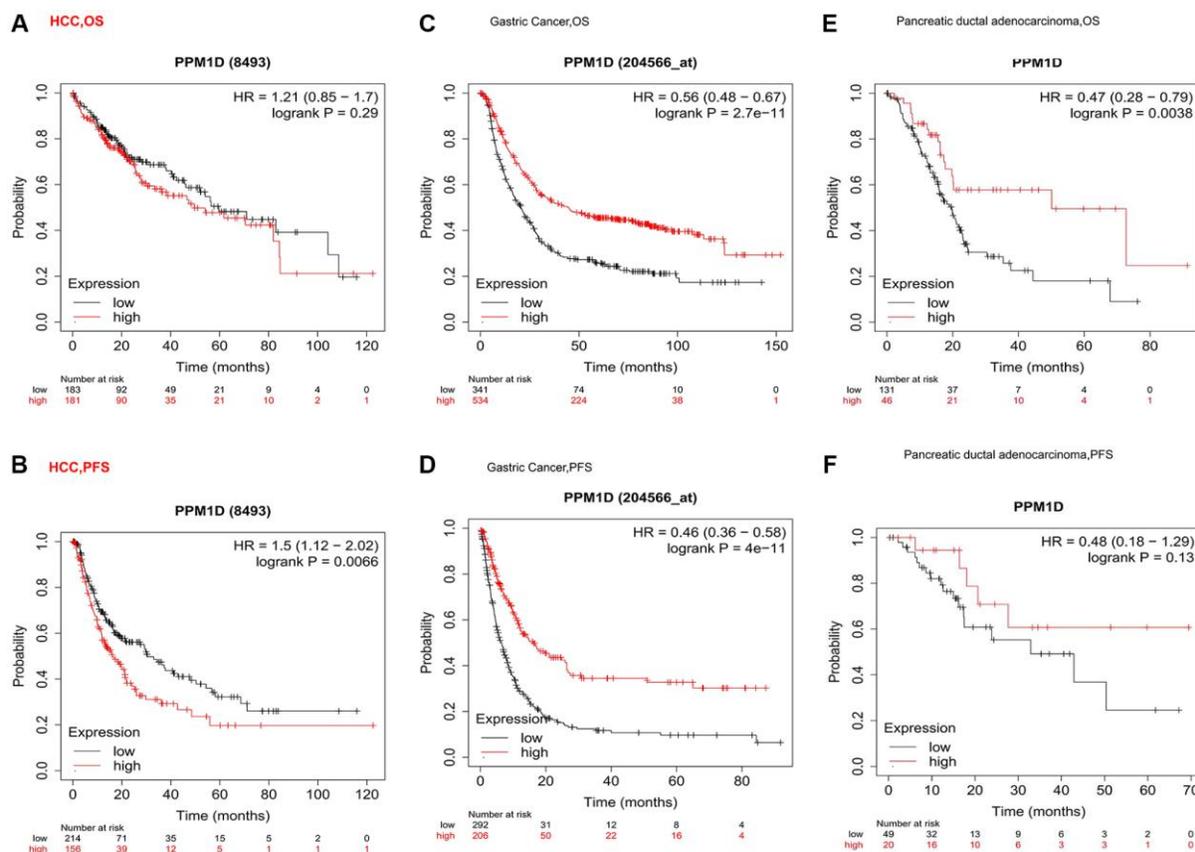
### Association analysis of PPM1D mRNA expression with the marker of distinct subgroups of immune cell

The relation between PPM1D mRNA expression and the level of TIICs in HCC tissues was analysed based on the GEPIA and TIMER databases. Considering that

the analysis of immune infiltration is affected by tumor purity of clinical samples, so we used purity to adjust the analysis (Table 3). The immune cells in HCC tissue including T cells (general), B cells, CD8+ T cells, TAMs, DCs, NK cells, M1 and M2 macrophages, monocytes, Tregs, Th17, Th1, Th2, exhausted T and Tfh cells were analyzed, using CHOL as the control.

According to the GEPIA and TIMER databases, PPM1D mRNA expression in HCC tissue remarkably associated with the expression of marker gene due to TAMs, tumor-infiltrating Monocytes, M1 Macrophage, M2 Macrophage, DCs, Th and Treg (Figure 4 and Table 4), while the association was not remarkable in CHOL (Supplementary Figure 3 and Table 4).

The mRNA expression of PPM1D revealed that the association with the expression of the markers of particular immune cell including the monocyte marker, such as CD86 ( $r = 0.392$ ;  $P = 3.77e-14$ ), IRF5 ( $r = 0.381$ ;  $P = 2.26e-13$ ), COX2 ( $r = 0.404$ ;  $P = 5.96e-15$ ), CD163 ( $r = 0.393$ ;  $P = 3.21e-14$ ), BDCA-4 ( $r = 0.541$ ;  $P = 1.43e-27$ ), STAT1 ( $r = 0.446$ ;  $P = 3.10e-18$ ), STAT6 ( $r = 0.366$ ;  $P = 2.41e-12$ ), STAT5A ( $r = 0.404$ ;



**Figure 2.** Kaplan-Meier survival curve analysis of the prognostic significance of high and low expression of PPM1D in different types of human cancers using the Kaplan-Meier plotter database (A-F). (A, B) OS and High PPM1D expression was correlated with poor PFS in HCC ( $n = 364$ ,  $n = 370$ ). (C, D) Survival curves of OS and PFS survival curves in the gastric cancer ( $n = 875$ ,  $n = 498$ ). (E, F) OS and PFS survival curves in the Pancreatic ductal adenocarcinoma ( $n = 177$ ,  $n = 69$ ). Abbreviations: OS: overall survival; PFS: progression-free survival.

**Table 1. Correlation of PPM1D mRNA expression and prognosis in HCC with different characteristics by Kaplan-Meier plotter.**

Charac teristics	OS (n = 364)			PFS (n = 366)			RFS (n = 313)			DSS (n = 357)		
	N	Hazard Ratio	P-value	N	Hazard Ratio	p-value	N	Hazard Ratio	p-value	N	Hazard Ratio	p-value
<b>SEX</b>												
Male	246	1.87 (0.91–3.87)	0.085	246	1.52 (1.05–2.2)	<b>0.027</b>	208	1.62 (1.07–2.46)	<b>0.023</b>	241	1.87 (0.91–3.87)	0.085
Female	118	1.69 (0.87–3.27)	0.120	120	1.66 (0.98–2.8)	0.057	105	1.55 (0.84–2.84)	0.160	116	2.12 (1–4.45)	<b>0.044</b>
<b>STAGE</b>												
1	170	1.56 (0.85–2.88)	0.150	170	2.31 (1.4–3.8)	<b>0.000</b>	153	2.19 (1.27–3.76)	<b>0.004</b>	167	2.72 (1.12–6.61)	<b>0.021</b>
2	83	1.71 (0.68–4.28)	0.250	84	0.8 (0.42–1.51)	0.490	74	0.67 (0.34–1.29)	0.230	82	2.55 (0.56–9.52)	0.21
3	83	0.65 (0.36–1.18)	0.150	83	0.75 (0.44–1.29)	0.300	68	0.78 (0.42–1.47)	0.440	81	0.65 (0.32–1.34)	0.24
4	5	–	–	–	–	–	–	–	–	–	–	–
<b>AJCC_T</b>												
1	180	1.54 (0.86–2.77)	0.140	180	2.18 (1.34–3.54)	0.001	160	2.06 (1.22–3.49)	<b>0.006</b>	177	2.25 (1.01–5.04)	<b>0.042</b>
2	90	1.8 (0.73–4.42)	0.190	92	0.77 (0.42–1.41)	0.400	79	0.68 (0.36–1.27)	0.220	89	2.95 (0.67–12.9)	0.13
3	78	0.58 (0.31–1.09)	0.086	78	1.57 (0.76–3.25)	0.220	65	0.76 (0.4–1.44)	0.390	75	0.52 (0.23–1.18)	0.11
4	13	–	–	13	–	–	6	–	–	13	–	–
<b>Vascular invasion</b>												
None	203	1.72 (1.03–2.87)	0.036	204	2.11 (1.33–3.35)	<b>0.001</b>	175	1.95 (1.21–3.15)	0.006	200	1.81 (0.88–3.71)	0.099
Micro	90	2.14 (0.93–4.93)	0.066	91	1.41 (0.79–2.49)	0.240	81	0.52 (0.26–1.01)	0.050	88	2.45 (0.76–7.97)	0.12
<b>Race</b>												
white	181	0.73 (0.45–1.18)	0.190	183	1.45 (0.98–2.16)	0.060	147	1.41 (0.9–2.21)	0.130	177	1.34 (0.71–2.52)	0.37
Asian	155	1.9 (1.03–3.49)	<b>0.035</b>	155	1.8 (1.12–2.88)	<b>0.014</b>	143	2.01 (1.2–3.36)	<b>0.007</b>	152	5.57 (1.31–9.65)	0.0086
<b>Hepatitis virus</b>												
Yes	150	2.17 (1.11–4.28)	<b>0.021</b>	152	1.97 (1.24–3.12)	<b>0.003</b>	138	2.07 (1.26–3.4)	<b>0.003</b>	149	2.75 (1.13–6.69)	<b>0.02</b>
None	167	1.54 (0.97–2.44)	0.067	167	1.54 (0.97–2.44)	0.067	142	0.72 (0.43–1.18)	0.190	163	2 (0.99–4.05)	<b>0.049</b>

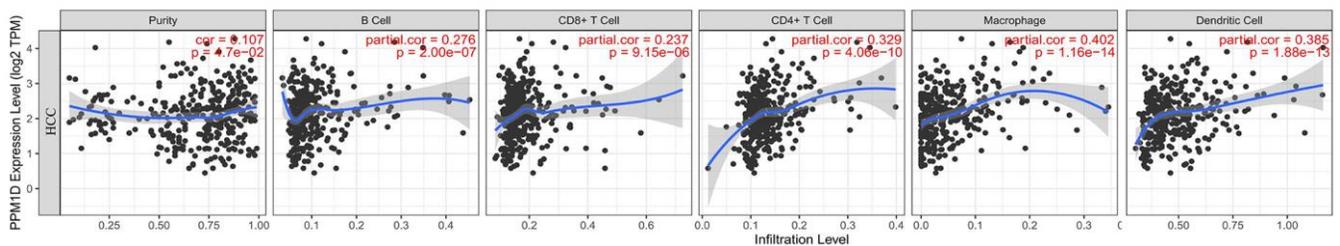
$P = 5.25e-15$ ), CCR8 ( $r = 0.513$ ;  $P = 1.41e-24$ ), and STAT5B ( $r = 0.731$ ;  $P = 5.82e-59$ ) (Table 1). These findings imply that PPM1D mRNA expression correlated with infiltration of the immune cell in HCC.

### Immunohistochemical examination about the expression of Wip1 in HCC samples

We collected 40 patient specimens for immunohistochemical examination and completed the follow-up work.

All the patients were divided into two groups with good prognosis and poor prognosis by the median survival time. Then the expression levels of Wip1 were analyzed in the two groups. Wip1 was highly expressed in hepatocellular carcinoma compared with precancerous tissue (Figure 5). There were 6 patients with high wip1 expression in good prognosis group, and 15 patients with high wip1 expression in poor prognosis group. In summary, higher expressions of Wip1 were intently correlated to worse prognosis, as shown in Table 5.

## HCC



**Figure 3. Correlation analysis of PPM1D expression and infiltration levels of immune cells in HCC tissues based on the TIMER database.** PPM1D expression in HCC tissues positively correlates with tumor purity and infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, and DCs.

**Table 2. Univariate and multivariate analyses of clinicopathological factors affecting PFS of HCC patients.**

Characteristics	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
SEX				
Male	1.52 (1.05–2.2)	0.027	1.42 (0.95–1.8)	0.069
Female	1.66 (0.98–2.8)	0.057	-	-
STAGE				
1	2.31(1.4–3.8)	0.00071	2.02 (1.6–3.2)	0.005
2	0.8(0.42–1.51)	0.49	-	-
3	0.75(0.44–1.29)	0.3	-	-
AJCC_T				
1	2.18 (1.34–3.54)	0.0013	1.62 (1.42–2.85)	0.02
2	0.77 (0.42–1.41)	0.4	-	-
3	1.57 (0.76–3.25)	0.22	-	-
Vascular invasion				
None	2.11 (1.33–3.35)	0.0012	1.85 (1.45–3.20)	0.034
Micro	1.41 (0.79–2.49)	0.24	-	-
Race				
White	1.45 (0.98–2.16)	0.06	-	-
Asian	1.8 (1.12–2.88)	0.014	1.62 (0.92–2.45)	0.32
Hepatitis virus				
Yes	1.97 (1.24–3.12)	0.0033	1.52 (0.67–2.86)	0.47
None	1.54 (0.97–2.44)	0.067	-	-

**Table 3. Correlation analysis between PPM1D and related genes and markers of immune cells in TIMER.**

Description	Gene markers	HCC				CHOL			
		None		Purity		None		Purity	
		Core	P	Core	P	Core	P	Core	P
CD8+ T cell	CD8A	0.17	**	0.254	***	1.107	0.533	0.019	0.914
	CD8B	0.07	0.18	0.15	**	0.073	0.673	0.003	0.987
T cell (general)	CD3D	0.03	0.59	0.1	0.06	0.232	0.173	0.152	0.385
	CD3E	0.1	0.06	0.206	***	0.172	0.315	0.066	0.705
	CD2	0.08	0.12	0.182	***	0.151	0.377	0.046	0.791
B cell	CD19	0.18	***	0.222	***	0.307	0.068	0.243	0.160
	CD79A	0.14	**	0.238	***	0.287	0.090	0.221	0.203
Monocyte	CD86	0.27	***	0.392	***	0.127	0.460	0.023	0.897
	CSF1R	0.21	***	0.326	***	0.172	0.315	0.100	0.568
TAM	CCL2	0.19	***	0.295	***	0.281	0.096	0.246	0.155
	CD68	0.18	***	0.255	***	0.181	0.289	0.132	0.449
	IL10	0.23	***	0.324	***	0.251	0.140	0.158	0.364
M1 Macro-phage	NOS2	0.21	***	0.223	***	0.307	0.069	0.312	0.068
	IRF5	0.39	***	0.381	***	0.219	0.199	0.179	0.304
	PTGS2	0.28	***	0.404	***	0.383	0.021	0.338	0.046
M2 Macro-phage	CD163	0.28	***	0.393	***	0.392	0.018	0.343	0.044
	VSIG4	0.18	***	0.285	***	0.28	0.099	0.216	0.213
	MS4A4A	0.22	***	0.33	***	0.286	0.091	0.21	0.227
Natural killer cell	KIR2DL1	0.06	0.23	0.058	0.28	-0.202	0.237	-0.239	0.166

	KIR2DL3	0.18	***	0.217	***	-0.015	0.929	-0.04	0.818
	KIR2DL4	0.12	0.02	0.152	**	-0.214	0.211	-0.272	0.114
	KIR3DL1	0.13	0.01	0.157	**	-0.124	0.471	-0.158	0.364
	KIR3DL2	0.08	0.12	0.130	0.02	-0.105	0.542	-0.113	0.516
	KIR3DL3	0.04	0.5	0.017	0.75	-0.191	0.265	-0.225	0.193
	KIR2DS4	0.09	0.08	0.099	0.07	-0.114	0.507	-0.149	0.392
Dendritic cell	HLA-DPB1	0.16	**	0.253	***	0.088	0.608	-0.004	0.982
	HLA-DQB1	0.07	0.21	0.147	**	0.311	0.066	0.269	0.118
	HLA-DRA	0.22	***	0.321	***	0.071	0.681	-0.033	0.850
	HLA-DPA1	0.22	***	0.322	***	0.056	0.743	-0.048	0.783
	BDCA-1	0.19	***	0.263	***	0.136	0.428	0.052	0.767
	BDCA-4	0.51	***	0.541	***	0.515	**	0.484	**
	CD11c	0.27	***	0.356	***	0.288	0.089	0.219	0.206
Th1	TBX21	0.10	0.05	0.176	**	0.147	0.390	0.033	0.852
	STAT4	0.18	***	0.232	***	0.099	0.564	0.04	0.820
	STAT1	0.41	***	0.446	***	0.4	*	0.379	*
	TNF	0.12	0.02	0.187	***	0.037	0.83	-0.067	0.703
	INF- $\alpha$	0.22	***	0.309	***	0.319	0.059	0.294	0.087
Th2	GATA3	0.16	**	0.280	***	0.046	0.79	-0.082	0.638
	STAT6	0.39	***	0.366	***	0.529	**	0.548	***
	STAT5A	0.35	***	0.404	***	0.228	0.181	0.196	0.258
	IL13	0.09	0.08	0.097	0.07	0.04	0.818	-0.011	0.95
Tfh	BCL6	0.41	***	0.401	***	0.407	*	0.4	*
	IL21	0.12	0.02	0.162	**	0.326	0.053	0.286	0.095
Th17	STAT3	0.49	***	0.528	***	0.547	***	0.557	***
	IL17A	0.1	0.06	0.107	0.05	0.053	0.759	0.005	0.979
Treg	FOXP3	0.3	***	0.335	***	0.264	0.12	0.189	0.277
	CCR8	0.43	***	0.513	***	0.344	*	0.293	0.088
	STAT5B	0.74	***	0.731	***	0.642	***	0.637	***
	TGFB1	0.21	***	0.297	***	0.307	0.069	0.265	0.124
T cell exhaustion	PD-1	0.14	**	0.198	***	0.203	0.234	0.157	0.369
	CTLA4	0.11	0.03	0.186	***	0.225	0.186	0.171	0.325
	LAG3	0.11	0.04	0.149	**	0.08	0.64	0.005	0.977
	TIM-3	0.23	***	0.351	***	0.126	0.464	0.036	0.839
	GZMB	0.04	0.43	0.075	0.17	0.039	0.823	-0.058	0.742

## DISCUSSION

This study illustrated that mRNA expression of PPM1D were correlated with several human cancer prognosis. High PPM1D mRNA levels correlated with poor prognosis in HCC. Furthermore, PPM1D mRNA levels associated with the level of TIICs, with respect to the status of the markers for the various immune cell types in HCC. These findings suggested that PPM1D is an underlying prognostic marker for HCC and other tumors.

The Sangerbox, TIMER, GEPIA, and Kaplan-Meier Plotter databases were used to study the status of

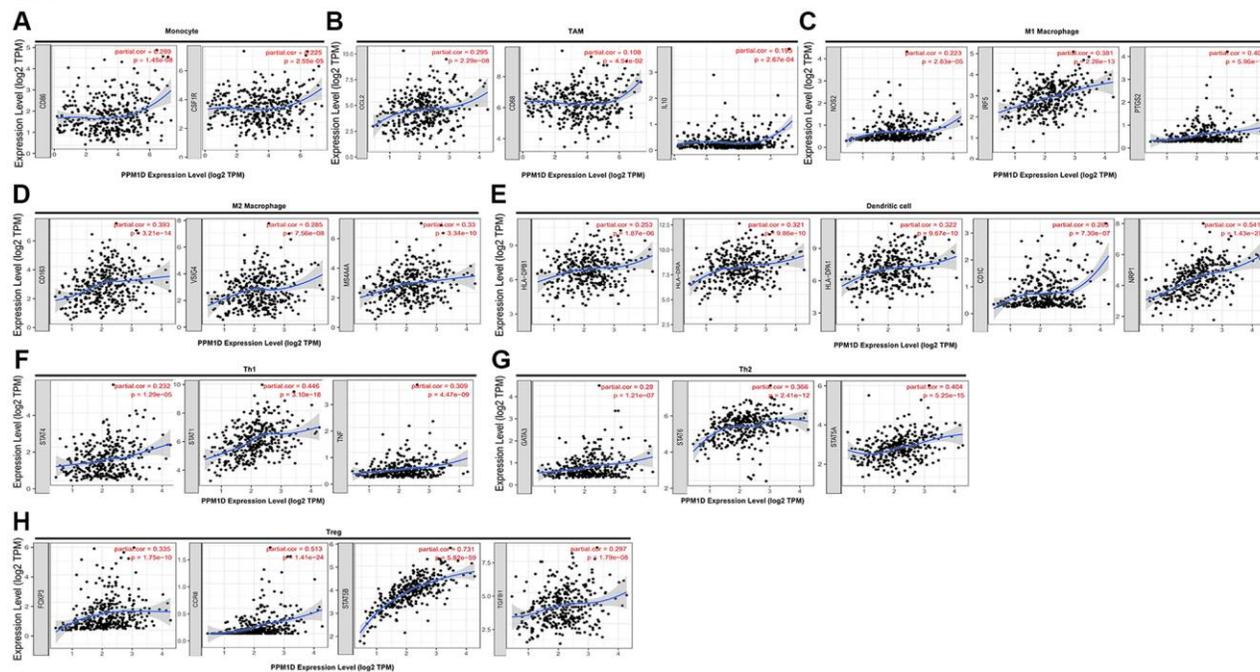
PPM1D mRNA in tumor tissue. The analysis demonstrated that PPM1D expression was significantly upregulated in most cancers. However, the expression of PPM1D differed in various cancers. This reflects differences in different methods of data collection as well as hidden causative mechanism. Furthermore, our results of PPM1D expression was consistent in HCC tissues with other databases. Moreover, we also found PPM1D protein was upregulated in HCC tissue than the paracancerous tissue. The GEPIA database gene expression analysis revealed that higher PPM1D expression related with worse prognosis in cancers including HCC, BLCA, CESC, MESO and UVM. Besides, Kaplan-Meier Plotter analysis disclosed that

PPM1D higher expression related with worse prognosis in HCC. Through immunohistochemistry, we also verified the high expression of PPM1D protein. The high PPM1D expression related with poor prognosis of the patients in early stage (stage 1 and AJCC\_1), and long PFS and OS in HCC patients with low PPM1D expression (Table 1). The findings imply that PPM1D is

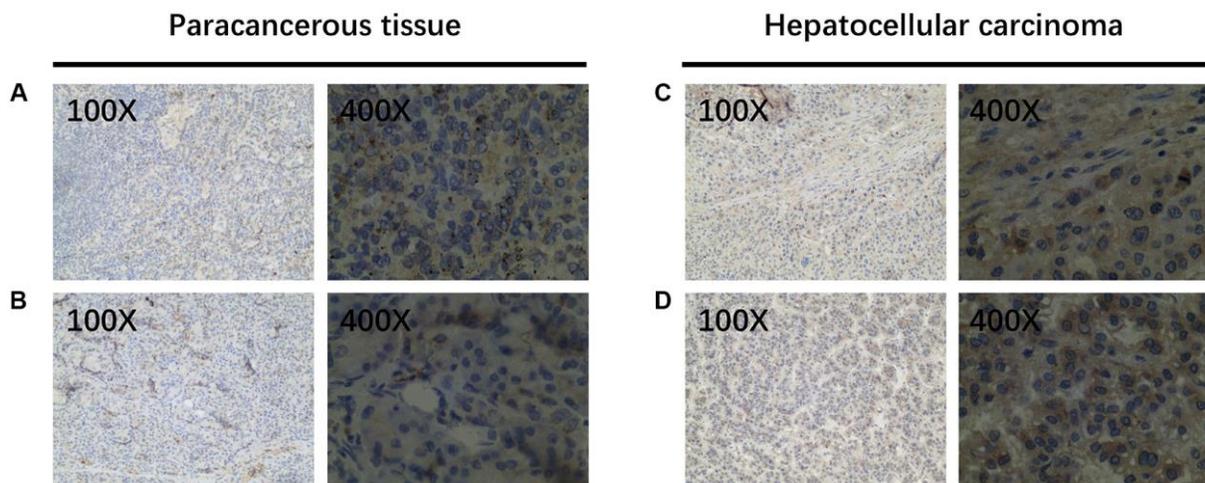
one of underlying prognostic markers for HCC, especially for patients in early stage.

Dissimilar to traditional tumor treatments such as chemotherapy and radiotherapy, immunotherapy is an innovative treatment method that dynamically regulates the immune system to attack tumor cells in diverse

## HCC



**Figure 4.** Correlation analysis of PPM1D mRNA expression and the expression of marker genes of infiltrating immune cells in HCC (A–H) using the TIMER database. (A–G) The scatter plots show correlation between PPM1D expression and the gene markers of (A) Monocytes (CD86 and CSF1R); (B) TAMs (CCL2, IL-10 and CD68); (C) M1 Macrophage (NOS2, IRF5 and PTGS2); (D) M2 Macrophage (CD163, VSIG4 and MS4A4A); (E) DCs (HLA-DPB1, HLA-DRA, HLA-DPA1, CD1C and NRP1); (F) Th1 cells (STAT4, STAT1 and TNF); (G) Th2 cells (GATA3, STAT6 and STAT5A) and (H) Tregs (FOXP3, CCR8, STAT5B and TGFB1) in HCC samples ( $n = 371$ ). PPM1D gene was on the x-axis and the related marker genes were on the y-axis.



**Figure 5.** WIP1 expression in paracancerous and hepatocellular carcinoma tissues. (A, B) Wip1 is low expressed in paracancerous tissues; (C, D) Wip1 is highly expressed in tumor tissues.

**Table 4. Correlation analysis between PPM1D and related genes and markers in GEPIA.**

Description	Gene markers	HCC				CHOL			
		Tumor		Normal		Tumor		Normal	
		R	P	R	P	R	P	R	P
Monocyte	CD86	0.28	***	0.55	***	0.13	0.44	0.60	0.097
	CD115	0.25	***	0.52	***	0.20	0.25	0.23	0.55
TAM	CCL2	0.18	***	0.23	0.11	0.25	0.14	0.55	0.13
	CD68	0.21	***	0.57	***	0.21	0.21	0.43	0.25
	IL10	0.25	***	0.39	**	0.29	0.08	0.38	0.32
M1 Macrophage	iNOS	0.25	***	0.14	0.33	0.37	*	-0.10	0.80
	IRF5	0.38	***	0.26	0.07	0.21	0.22	0.25	0.52
M2 Macrophage	VSIG4	0.17	**	0.43	**	0.30	0.08	0.05	0.91
	MS4A4A	0.22	***	0.50	***	0.30	0.08	0.13	0.74
Dendritic cell	HLA-DPB1	0.18	***	0.45	***	0.10	0.57	0.13	0.74
	HLA-DQB1	-0.05	0.38	0.20	0.17	0.24	0.16	0.20	0.61
	HLA-DRA	0.22	***	0.47	***	0.042	*	0.18	0.64
	HLA-DPA1	0.22	***	0.46	***	0.026	0.88	0.12	0.78
	BDCA-1(CD11C)	0.18	***	0.31	*	0.073	0.67	0.05	0.91
	CD11c(ITGAX)	0.30	***	0.40	**	0.30	0.07	0.62	0.09

“R” is a correlation coefficient, the higher the value, the stronger the correlation.

**Table 5. The relationship between Wip1 expression and prognosis in HCC.**

Prognosis	Wip1 expression		$\chi^2$	P value
	Low	High		
Good prognosis	14	6	8.120	0.004
Poor prognosis	5	15		

targets and directions [20]. In various basic experiments and clinical researches, it has been found that immunotherapy does have an incomparable advantage over traditional cancer treatments, and can prolong OS and PFS [21]. Current research shows that immunotherapy also plays a significant role in HCC [22, 23]. In recent years, immune checkpoint inhibitors (anti-CTLA-4, anti-PD-1 and anti-PD-L1 antibodies) have shown therapeutic potentiality on advanced HCC [24, 25]. Therefore, an in-depth understanding of tumor immune infiltration can help provide prognostic predictors for immunotherapy, and provide more precise and individualized treatment for HCC.

A previous study showed that a PPM1D-deficient mouse has defective immune response [26]. There was cumulative evidence suggested that a PPM1D-deficient mouse has immune response deficiency [27], neutrophil migration and development via p38-MAPK-STAT1 pathway [28]; macrophage migration and phagocytosis

[29]; dendritic spine memory and morphology via p53MAPK signaling. PPM1D mutations assembled in H3F3A-mutated malignant brainstem gliomas, and high intratumoral CD8+ T cell density was less common in the H3F3A-mutated tumors. Moreover, patients with H3F3A-mutated tumors experienced worse prognoses in comparison to other patients. It is speculated that PPM1D may be related to poor prognosis resulting from low intracranial CD8+ T cells [30].

Our study demonstrates that higher PPM1D expression correlated with worse prognosis in HCC and infiltration of different types of immune cell inclusive of CD4+, CD8+ T cells, macrophages, B cells and DCs. As an important component of the tumor microenvironment, TAM play a vital role in immune regulation. The increase of TAM and PD-L1 in liver cancer shows the characteristics of immune escape [31]. We established the relationship between expression of PPM1D mRNA and TAM markers, IL-10 and CD68, and iNOS, M1

macrophage marker, IRF5 and COX2. This implies that PPM1D activates the infiltration and TAM's activity. PPM1D expression was also associated with the expression of the markers various subsets of Th cells, including Tregs (FOXP3, STAT5B and TGF- $\beta$ ), Tfh (BCL6), and Th1 (STAT-1, STAT-4 and TNF- $\alpha$ ), Th2 (GATA3, STAT6, STAT35A). This suggests the vital contribution of the PPM1D in adjusting tumor-infiltration of the Th cells. Besides, the expression of the deficient T cells marker, TIM-3 and PD-1, which were critical inhibitory immune checkpoint proteins definitely associated with the PPM1D expression. Interestingly, PPM1D is highly expressed in HCC and CHOL from Sangerbox and TIMER database., and our previous researches have illustrated that PPM1D is associated with poor prognosis of cholangiocarcinoma [11]. However, the analysis found that PPM1D is not correlated to the immune cell infiltration of CHOL, suggesting that the association of PPM1D on immune cell infiltration has a special role in HCC, which is worthy of further study.

Currently, researchers are concerned about the contribution of PPM1D in HCC. It is reported that mRNA of PPM1D was remarkably higher in HCC than corresponding normal tissue from 86 HCC. The high PPM1D was related to TNM stage, alpha-fetoprotein (AFP) level, tumor size, recurrence incidence and the family history for HCC. Nevertheless, PPM1D expression and age, gender, alcohol intake, lymph node metastasis, hepatitis B virus infection or portal vein invasion did not show any correlation [32]. In this study, we realized that the upregulation of PPM1D correlated with poor prognosis in specific patients who had some specific clinical characteristics inclusive of alcohol consumers, Asians, males, those in early stages (stage 1 and AJCC\_1), and the suffering from hepatitis viral infections. The findings in this study consistent with our previous study [11, 14]. However, the credibility of this study could be higher than that of a single-center study since it emanates from multiple databases. In general, our findings imply that a high expression of PPM1D could be related with poor prognosis of HCC.

Final note: Our research has some limitations. The first point is that many of our data sources are previously reported data, not all of our own clinical data. The second point is that some tumors in the database have too few samples.

## CONCLUSIONS

Our findings suggest that PPM1D might be an underlying prognostic marker for HCC, that could be taken to determine the rate of immune cells infiltration in cancer tissue. The comparatively high levels of

PPM1D in HCC and other cancers imply a high risk for tumor relapse after treatment. As such, close medical monitoring is vital for these patients.

## AUTHOR CONTRIBUTIONS

Zhangtao Yu, Zhihua Zhang, Yu Jiang, Lianhong Zou and Xiehong Liu performed the analysis data; Yinghui Song, Nanhui Yu and Le Wang contributed significantly to analysis and manuscript preparation; Chuang Peng and Mengting Cai performed the data analyses and wrote the manuscript; Bo Jiang, Chuang Peng and Sulai Liu contributed to the conception of the study; Zhangtao Yu, Yinghui Song, Xianhai Mao helped perform the analysis with constructive discussions.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

## FUNDING

This work was financially supported by following funds: Youth Talent of Hunan (Grant No. 2020RC3066); Postdoctoral Innovation Talents Project (NO. 2020RC2064); Hunan Provincial Natural Science Foundation of China (Grant No. 2020JJ5610); Hunan Provincial Development and Reform Commission Project (NO. 2019FGW26); The Project of Improving the Diagnosis and Treatment Capacity of Hepatobiliary, Pancreas and Intestine in Hunan Province (Xiangwei [2019] No. 118); China Postdoctoral Science Foundation (NO.2020M68115/ 2021T140197); Hunan Natural Science Fund for Excellent Young Scholars (NO. 2021JJ20003).

## REFERENCES

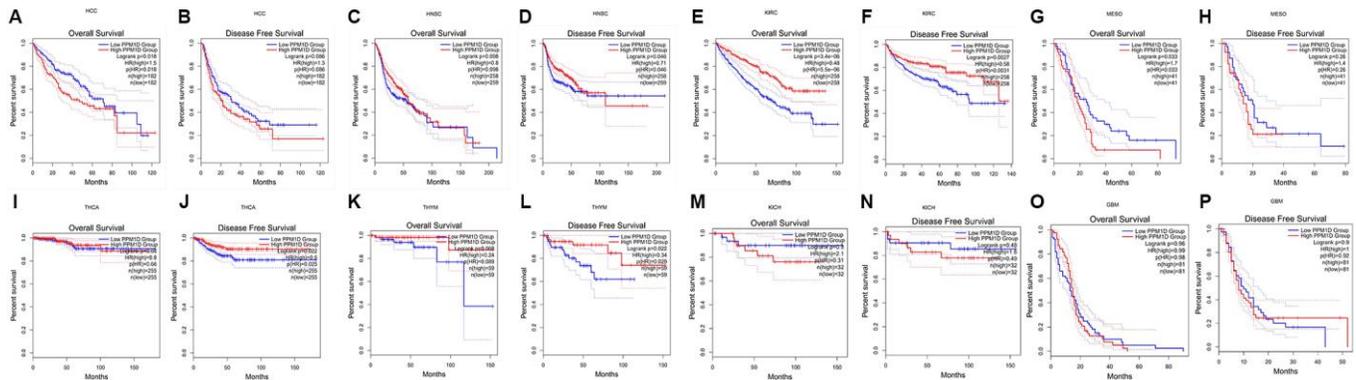
1. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, Allen C, Hansen G, Woodbrook R, Wolfe C, Hamadeh RR, Moore A, Werdecker A, et al, and Global Burden of Disease Cancer Collaboration. The Global Burden of Cancer 2013. *JAMA Oncol.* 2015; 1:505–27.  
<https://doi.org/10.1001/jamaoncol.2015.0735>  
PMID:[26181261](https://pubmed.ncbi.nlm.nih.gov/26181261/)
2. Makarova-Rusher OV, Altekruze SF, McNeel TS, Ulahannan S, Duffy AG, Graubard BI, Greten TF, McGlynn KA. Population attributable fractions of risk factors for hepatocellular carcinoma in the United States. *Cancer.* 2016; 122:1757–65.  
<https://doi.org/10.1002/cncr.29971>  
PMID:[26998818](https://pubmed.ncbi.nlm.nih.gov/26998818/)

3. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol*. 2004; 130:417–22.  
<https://doi.org/10.1007/s00432-004-0552-0>  
PMID:15042359
4. Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, Al-Raddadi R, Alvis-Guzman N, Amoako Y, Artaman A, Ayele TA, Barac A, Bensenor I, et al, and Global Burden of Disease Liver Cancer Collaboration. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA Oncol*. 2017; 3:1683–91.  
<https://doi.org/10.1001/jamaoncol.2017.3055>  
PMID:28983565
5. Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology*. 2004 (Suppl 1); 127:S35–50.  
<https://doi.org/10.1053/j.gastro.2004.09.014>  
PMID:15508101
6. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol*. 2018; 69:182–236.  
<https://doi.org/10.1016/j.jhep.2018.03.019>  
PMID:29628281
7. Kishi Y, Hasegawa K, Sugawara Y, Kokudo N. Hepatocellular carcinoma: current management and future development-improved outcomes with surgical resection. *Int J Hepatol*. 2011; 2011:728103.  
<https://doi.org/10.4061/2011/728103>  
PMID:21994868
8. Uyanik B, Grigorash BB, Goloudina AR, Demidov ON. DNA damage-induced phosphatase Wip1 in regulation of hematopoiesis, immune system and inflammation. *Cell Death Discov*. 2017; 3:17018.  
<https://doi.org/10.1038/cddiscovery.2017.18>  
PMID:28417018
9. Fiscella M, Zhang H, Fan S, Sakaguchi K, Shen S, Mercer WE, Vande Woude GF, O'Connor PM, Appella E. Wip1, a novel human protein phosphatase that is induced in response to ionizing radiation in a p53-dependent manner. *Proc Natl Acad Sci U S A*. 1997; 94:6048–53.  
<https://doi.org/10.1073/pnas.94.12.6048>  
PMID:9177166
10. Lu X, Nguyen TA, Moon SH, Darlington Y, Sommer M, Donehower LA. The type 2C phosphatase Wip1: an oncogenic regulator of tumor suppressor and DNA damage response pathways. *Cancer Metastasis Rev*. 2008; 27:123–35.  
<https://doi.org/10.1007/s10555-008-9127-x>  
PMID:18265945
11. Liu S, Jiang B, Li H, He Z, Lv P, Peng C, Wang Y, Cheng W, Xu Z, Chen W, Liu Z, Zhang B, Shen S, Xiang S. Wip1 is associated with tumorigenesis and metastasis through MMP-2 in human intrahepatic cholangiocarcinoma. *Oncotarget*. 2017; 8:56672–83.  
<https://doi.org/10.18632/oncotarget.18074>  
PMID:28915621
12. Liu S, Jiang J, Huang L, Jiang Y, Yu N, Liu X, Lv Y, Li H, Zou L, Peng C, Yu X, Jiang B. iNOS is associated with tumorigenicity as an independent prognosticator in human intrahepatic cholangiocarcinoma. *Cancer Manag Res*. 2019; 11:8005–22.  
<https://doi.org/10.2147/CMAR.S208773>  
PMID:31692584
13. Chen C, Jiang J, Fang M, Zhou L, Chen Y, Zhou J, Song Y, Kong G, Zhang B, Jiang B, Li H, Peng C, Liu S. MicroRNA-129-2-3p directly targets Wip1 to suppress the proliferation and invasion of intrahepatic cholangiocarcinoma. *J Cancer*. 2020; 11:3216–24.  
<https://doi.org/10.7150/jca.41492>  
PMID:32231727
14. Liu S, Qi L, Han W, Wan X, Jiang S, Li Y, Xie Y, Liu L, Zeng F, Liu Z, Zu X. Overexpression of wip1 is associated with biologic behavior in human clear cell renal cell carcinoma. *PLoS One*. 2014; 9:e110218.  
<https://doi.org/10.1371/journal.pone.0110218>  
PMID:25334029
15. Xu Z, Cao C, Xia H, Shi S, Hong L, Wei X, Gu D, Bian J, Liu Z, Huang W, Zhang Y, He S, Lee NP, Chen J. Protein phosphatase magnesium-dependent 1δ is a novel tumor marker and target in hepatocellular carcinoma. *Front Med*. 2016; 10:52–60.  
<https://doi.org/10.1007/s11684-016-0433-3>  
PMID:26809466
16. Henriksson J, Chen X, Gomes T, Ullah U, Meyer KB, Miragaia R, Duddy G, Pramanik J, Yusa K, Lahesmaa R, Teichmann SA. Genome-wide CRISPR Screens in T Helper Cells Reveal Pervasive Crosstalk between Activation and Differentiation. *Cell*. 2019; 176:882–96.e18.  
<https://doi.org/10.1016/j.cell.2018.11.044>  
PMID:30639098
17. Cózar B, Greppi M, Carpentier S, Narni-Mancinelli E, Chiossone L, Vivier E. Tumor-Infiltrating Natural Killer Cells. *Cancer Discov*. 2021; 11:34–44.  
<https://doi.org/10.1158/2159-8290.CD-20-0655>  
PMID:33277307
18. Ngambenjawong C, Gustafson HH, Pun SH. Progress in tumor-associated macrophage (TAM)-targeted therapeutics. *Adv Drug Deliv Rev*. 2017; 114:206–21.  
<https://doi.org/10.1016/j.addr.2017.04.010>  
PMID:28449873

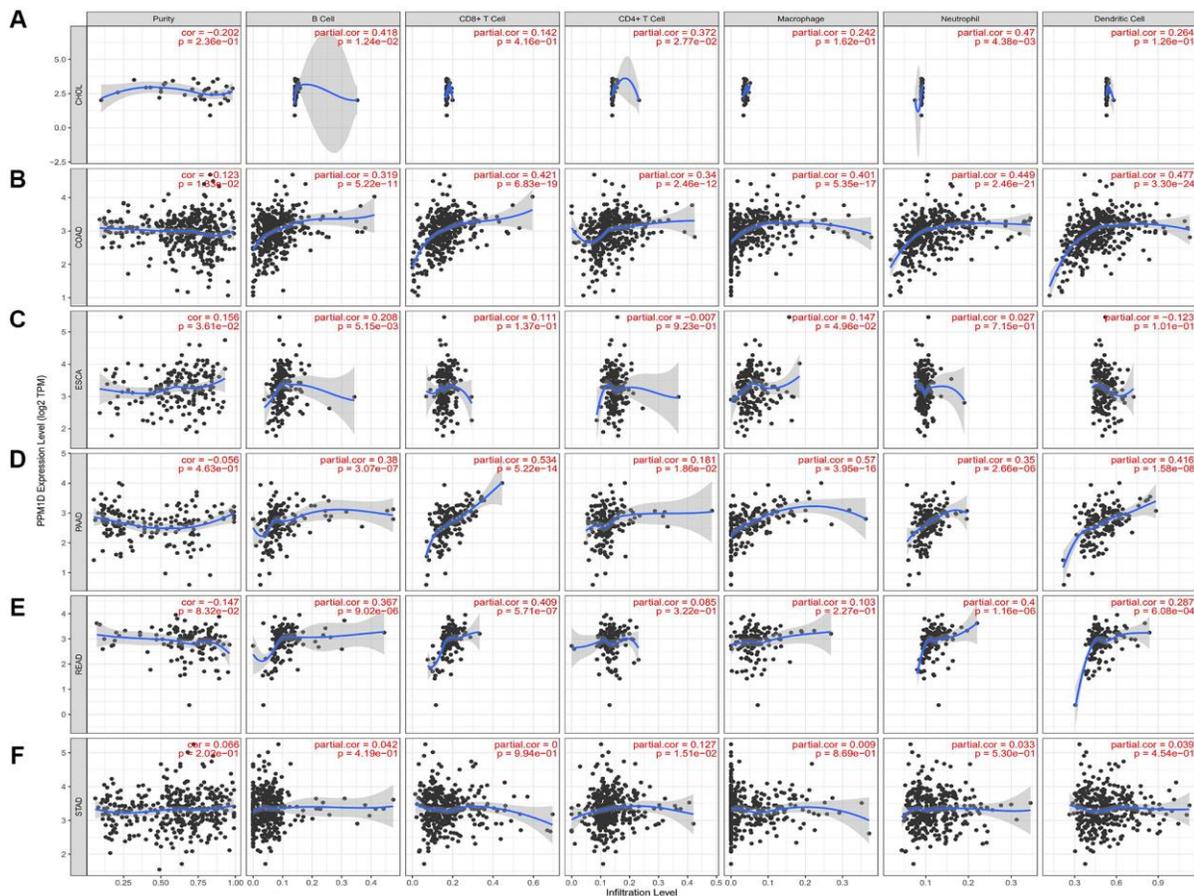
19. Zhang JQ, Zeng S, Vitiello GA, Seifert AM, Medina BD, Beckman MJ, Loo JK, Santamaria-Barria J, Maltbaek JH, Param NJ, Moral JA, Zhao JN, Balachandran V, et al. Macrophages and CD8<sup>+</sup> T Cells Mediate the Antitumor Efficacy of Combined CD40 Ligation and Imatinib Therapy in Gastrointestinal Stromal Tumors. *Cancer Immunol Res.* 2018; 6:434–47. <https://doi.org/10.1158/2326-6066.CIR-17-0345> PMID:29467128
20. Couzin-Frankel J. Breakthrough of the year 2013. *Cancer immunotherapy. Science.* 2013; 342:1432–33. <https://doi.org/10.1126/science.342.6165.1432> PMID:24357284
21. Tan S, Li D, Zhu X. Cancer immunotherapy: Pros, cons and beyond. *Biomed Pharmacother.* 2020; 124:109821. <https://doi.org/10.1016/j.biopha.2020.109821> PMID:31962285
22. El Dika I, Khalil DN, Abou-Alfa GK. Immune checkpoint inhibitors for hepatocellular carcinoma. *Cancer.* 2019; 125:3312–19. <https://doi.org/10.1002/cncr.32076> PMID:31290997
23. Wu X, Gu Z, Chen Y, Chen B, Chen W, Weng L, Liu X. Application of PD-1 Blockade in Cancer Immunotherapy. *Comput Struct Biotechnol J.* 2019; 17:661–74. <https://doi.org/10.1016/j.csbj.2019.03.006> PMID:31205619
24. Zongyi Y, Xiaowu L. Immunotherapy for hepatocellular carcinoma. *Cancer Lett.* 2020; 470:8–17. <https://doi.org/10.1016/j.canlet.2019.12.002> PMID:31811905
25. Prieto J, Melero I, Sangro B. Immunological landscape and immunotherapy of hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol.* 2015; 12:681–700. <https://doi.org/10.1038/nrgastro.2015.173> PMID:26484443
26. Choi J, Appella E, Donehower LA. The structure and expression of the murine wildtype p53-induced phosphatase 1 (Wip1) gene. *Genomics.* 2000; 64:298–306. <https://doi.org/10.1006/geno.2000.6134> PMID:10756097
27. Schito ML, Demidov ON, Saito S, Ashwell JD, Appella E. Wip1 phosphatase-deficient mice exhibit defective T cell maturation due to sustained p53 activation. *J Immunol.* 2006; 176:4818–25. <https://doi.org/10.4049/jimmunol.176.8.4818> PMID:16585576
28. Kamada R, Kudoh F, Yoshimura F, Tanino K, Sakaguchi K. Inhibition of Ser/Thr phosphatase PPM1D induces neutrophil differentiation in HL-60 cells. *J Biochem.* 2017; 162:303–08. <https://doi.org/10.1093/jb/mvx032> PMID:28486685
29. Tang Y, Liu L, Sheng M, Xiong K, Huang L, Gao Q, Wei J, Wu T, Yang S, Liu H, Mu Y, Li K. Wip1 knockout inhibits the proliferation and enhances the migration of bone marrow mesenchymal stem cells. *Exp Cell Res.* 2015; 334:310–22. <https://doi.org/10.1016/j.yexcr.2015.03.018> PMID:25839408
30. Zhang Y, Pan C, Wang J, Cao J, Liu Y, Wang Y, Zhang L. Genetic and immune features of resectable malignant brainstem gliomas. *Oncotarget.* 2017; 8:82571–82. <https://doi.org/10.18632/oncotarget.19653> PMID:29137285
31. Zhu Y, Yang J, Xu D, Gao XM, Zhang Z, Hsu JL, Li CW, Lim SO, Sheng YY, Zhang Y, Li JH, Luo Q, Zheng Y, et al. Disruption of tumour-associated macrophage trafficking by the osteopontin-induced colony-stimulating factor-1 signalling sensitises hepatocellular carcinoma to anti-PD-L1 blockade. *Gut.* 2019; 68:1653–66. <https://doi.org/10.1136/gutjnl-2019-318419> PMID:30902885
32. Li GB, Zhang XL, Yuan L, Jiao QQ, Liu DJ, Liu J. Protein phosphatase magnesium-dependent 1δ (PPM1D) mRNA expression is a prognosis marker for hepatocellular carcinoma. *PLoS One.* 2013; 8:e60775. <https://doi.org/10.1371/journal.pone.0060775> PMID:23556002

SUPPLEMENTARY MATERIALS

Supplementary Figures

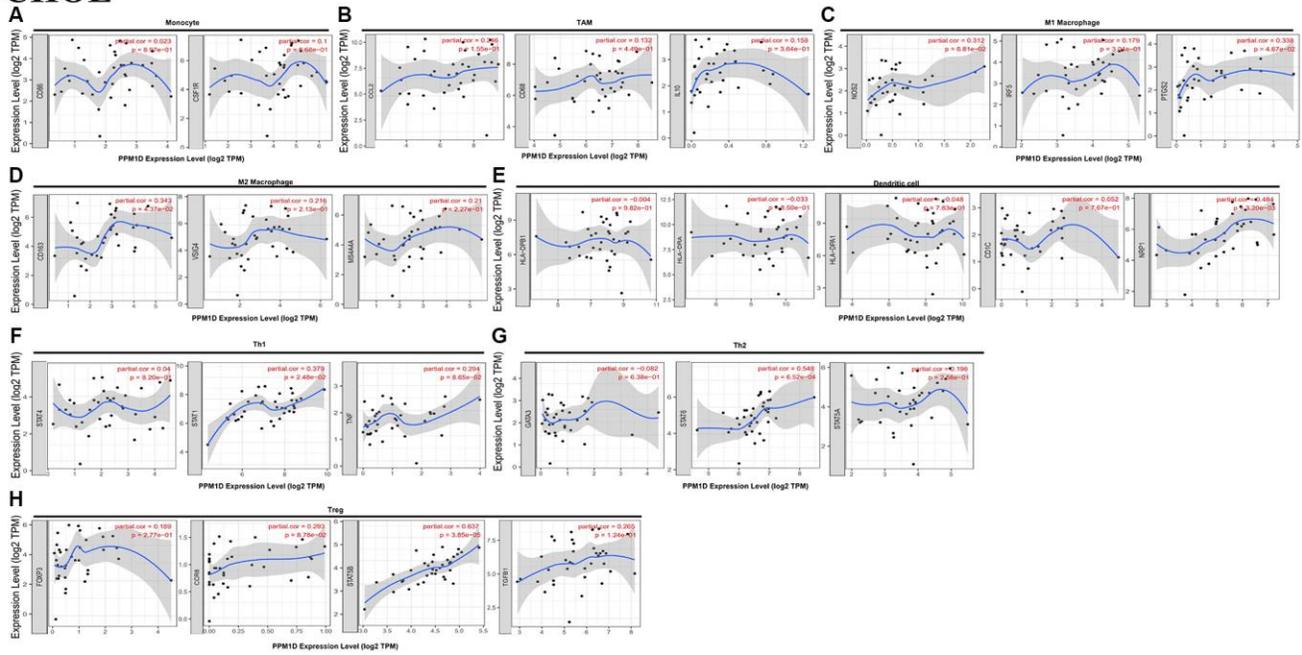


**Supplementary Figure 1. Correlation of PPM1D mRNA expression with prognostic values in diverse types of cancer in the GEPIA databases.** Overall survival and disease free curves comparing the high and low expression of PPM1D in Hepatocellular carcinoma (HCC) (A, B), Head and Neck squamous cell carcinoma (HNSC) (C, D), Kidney renal clear cell carcinoma (KIRC) (E, F), Mesothelioma (MESO) (G, H), Thyroid carcinoma (THCA) (I, J), Thymoma (THYM) (K, L), Kidney Chromophobe (KICH) (M, N), Glioblastoma multiforme (GBM) (O, P).



**Supplementary Figure 2. Correlation analysis of PPM1D expression and infiltration levels of immune cells in digestive system tumors tissues based on the TIMER database.** PPM1D expression in HCC tissues positively correlates with tumor purity and infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, and DCs. (A) Cholangiocarcinoma (CHOL), (B) Colon adenocarcinoma (COAD), (C) Esophageal carcinoma (ESCA), (D) Pancreatic adenocarcinoma (PAAD), (E) Rectum adenocarcinoma (READ), (F) stomach adenocarcinoma (STAD).

# CHOL



**Supplementary Figure 3.** Correlation analysis of PPM1D mRNA expression and the expression of marker genes of infiltrating immune cells in CHOL (A–H) using the TIMER database. (A–G) The scatter plots show correlation between PPM1D expression and the gene markers of (A) Monocytes (CD86 and CSF1R); (B) TAMs (CCL2, IL-10 and CD68); (C) M1 Macrophage (NOS2, IRF5 and PTGS2); (D) M2 Macrophage (CD163, VSIG4 and MS4A4A); (E) DCs (HLA-DPB1, HLA-DRA, HLA-DPA1, CD1C and NRP1); (F) Th1 cells (STAT4, STAT1 and TNF); (G) Th2 cells (GATA3, STAT6 and STAT5A) and (H) Tregs (FOXP3, CCR8, STAT5B and TGFB1) in HCC samples ( $n = 36$ ). PPM1D gene was on the x-axis and the related marker genes were on the y-axis.

# CCR5 is a prognostic biomarker and an immune regulator for triple negative breast cancer

Xin Wang<sup>1,\*</sup>, Yong Han<sup>2,3,\*</sup>, Jiamin Peng<sup>4</sup>, Jie He<sup>1</sup>

<sup>1</sup>Thoracic Surgery Department, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>2</sup>Department of Thoracic Surgery, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China

<sup>3</sup>Key Laboratory of Tumor Molecular Diagnosis and Individualized Medicine of Zhejiang Province, Zhejiang, China

<sup>4</sup>Department of Clinical Laboratory, Tongde Hospital of Zhejiang Province, Hangzhou, Zhejiang 310012, China

\*Equal contribution

**Correspondence to:** Jie He; email: [prof.jiehe@gmail.com](mailto:prof.jiehe@gmail.com), <https://orcid.org/0000-0002-0285-5403>

**Keywords:** CCR5, TNBC, multi-omics landscape, immune therapy, prognosis

**Received:** March 9, 2021

**Accepted:** September 18, 2021

**Published:** October 30, 2021

**Copyright:** © 2021 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

This study aims to explore the clinical implications and potential mechanistic functions of CCR5 in triple negative breast cancer. Briefly, we demonstrated that CCR5 is overexpressed in TNBC and is associated with better prognosis of TNBC. CCR5 expression is positively correlated with tumor immune cell infiltration and tumor immune response related pathways. Multi-omics data analyses identified CCR5 associated genomic and proteomic changes. CCR5 overexpression was associated with better overall survival in TNBC patients with TP53 mutation. We also summarized the latest findings on ICB efficacy related genes and explored the association between CCR5 and those genes. These results indicated that CCR5 is a potential tumor suppressor gene and individualized therapeutic strategy could be established based on multi-omics background and expression pattern of ICB related genes. In conclusion, CCR5 is associated with better survival of TNBC patients with TP53 mutation, which may exert its roles through tumor immune environment.

## INTRODUCTION

Cancer caused approximately 10 million of deaths in the year 2020 worldwide [1]. With the development of new technologies and experimental methods and devices, numerous research advancement have been achieved in both basic and clinical fields in the last decade [2, 3]. Triple negative breast cancer (TNBC) is the most aggressive subtype of breast cancer with limited therapeutic options [4]. Recent studies revealed several underlying molecular mechanisms and potential drugs for TNBC [5, 6]. However, there is still a long way to go before clinical application. It is crucial to clarify the mechanisms underly TNBC's aggressive phenotype and explore novel prognostic biomarkers and therapeutic targets.

It was reported that CCR5-Δ32/Δ32 genotype was associated with reduced life expectancy despite the protective effect of the mutation against HIV by analyzing genotyping and death registry information of 409,693 individuals of British ancestry [7]. This report indicated that CCR5 might have some positive roles in human fitness and thus caught the attention of the whole world immediately.

Actually, the function of CCR5 remain elusive despite intensive study in recent years. For instance, some researchers claimed that CCR5 could promote breast and gastric cancer progression, while others demonstrated that the expression of CCR5 in both CD4+ and CD8+ T cells was critical in boosting anti-tumor immune response [8–11]. We and our

collaborators showed that CCR5 overexpression was associated with better prognosis of breast cancer, lung cancer, liver cancer, rectal cancer and cervical cancer which might owing to its association with immune cell infiltration [12]. In this study, we will analysis the expression and prognostic value of CCR5 in TNBC and explore the underlying molecular mechanisms using multi-omics data. We will also discuss the association between CCR5 and ICB efficacy related genes and offer potential therapeutic options for TNBC.

## RESULTS

### CCR5 expression and its association with survival and TIL in basal-like or triple negative breast cancer

CCR5 is overexpressed in breast cancer in comparison to matched normal control (Supplementary Figure 1,  $P < 0.0001$ ). Its expression level in basal-like breast cancer or TNBC are significantly higher than in non-basal-like or non-TNBC samples (Figure 1A, expression data is obtained from bc-GenExMiner database,  $P < 0.0001$ ). Methylation level of CCR5 promoter region was significantly associated with CCR5 expression (Spearman  $r = 0.5912$ ,  $P < 0.0001$ , Figure 1B), which meant that methylation might be one of the main regulators of CCR5 expression. As is shown in Figure 1C, Kaplan–Meier analyses indicated that CCR5 expression was positively correlated with overall survival (OS), recurrence free survival (RFS) and distant metastasis-free survival (DMFS) (merged TNBC gene expression data and corresponding clinical information were from KM plot, HR = 0.41, 0.39 and 0.32;  $P = 0.0007$ ,  $<0.0001$  and  $<0.0001$ , respectively), which was further confirmed by analyzing TCGA

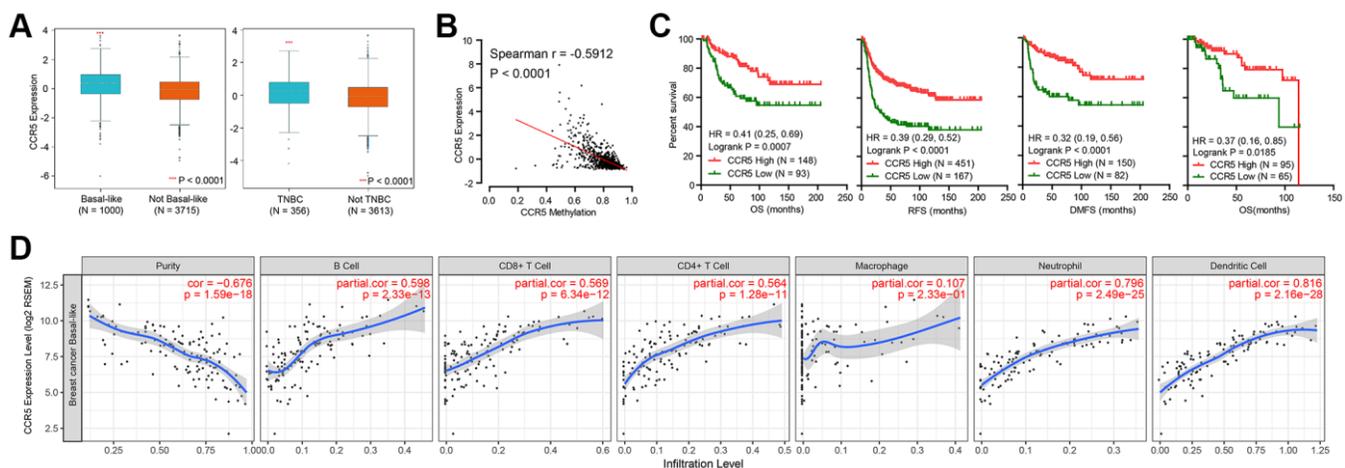
TNBC data (OS, HR = 0.37,  $P = 0.0185$ ). Tumor immune cell infiltration analyses using TIMER in basal-like breast cancer samples showed that adjusted CCR5 expression was significantly correlated with tumor infiltration of B cell, CD8+ T cell, CD4+ T cell, Neutrophil and Dendritic cell ( $P < 0.0001$ ), while there was no correlation between CCR5 expression and Macrophage infiltration ( $P = 0.233$ ) (Figure 1D).

### GSEA analysis of CCR5 using 1570 breast cancer samples

Since CCR5 is overexpressed in TNBC/basal-like breast cancer and is associated with better survival and tumor immune cell infiltration, we wish to elucidate the potential roles of CCR5 in breast cancer using an expression profiling dataset containing 1570 breast cancer samples. GSEA analyses results (Figure 2) indicated that CCR5 expression was positively correlated with innate inflammation pathways such as NF- $\kappa$ B pathway (Figure 2A, NES = 2.5,  $P < 0.0001$ ). It was also associated with TCR pathway, Th1Th2 pathway, T cell to Natural killer pathway and immune checkpoint related signatures (Figure 2B–E, NES = 2.35, 2.17, 2.27, 2.13, respectively, all  $P < 0.0001$ ). Moreover, CCR5 was associated with apoptosis (Figure 2F, NES = 2.04,  $P < 0.0001$ ). These results indicated that CCR5 might repress breast cancer progression through NF- $\kappa$ B pathway, immune cell activation and pro-apoptosis.

### The genomic landscape of CCR5 high and low expressed TNBC patients

Next, the genomic landscape of TNBC patients with CCR5 high and low expression were compared and



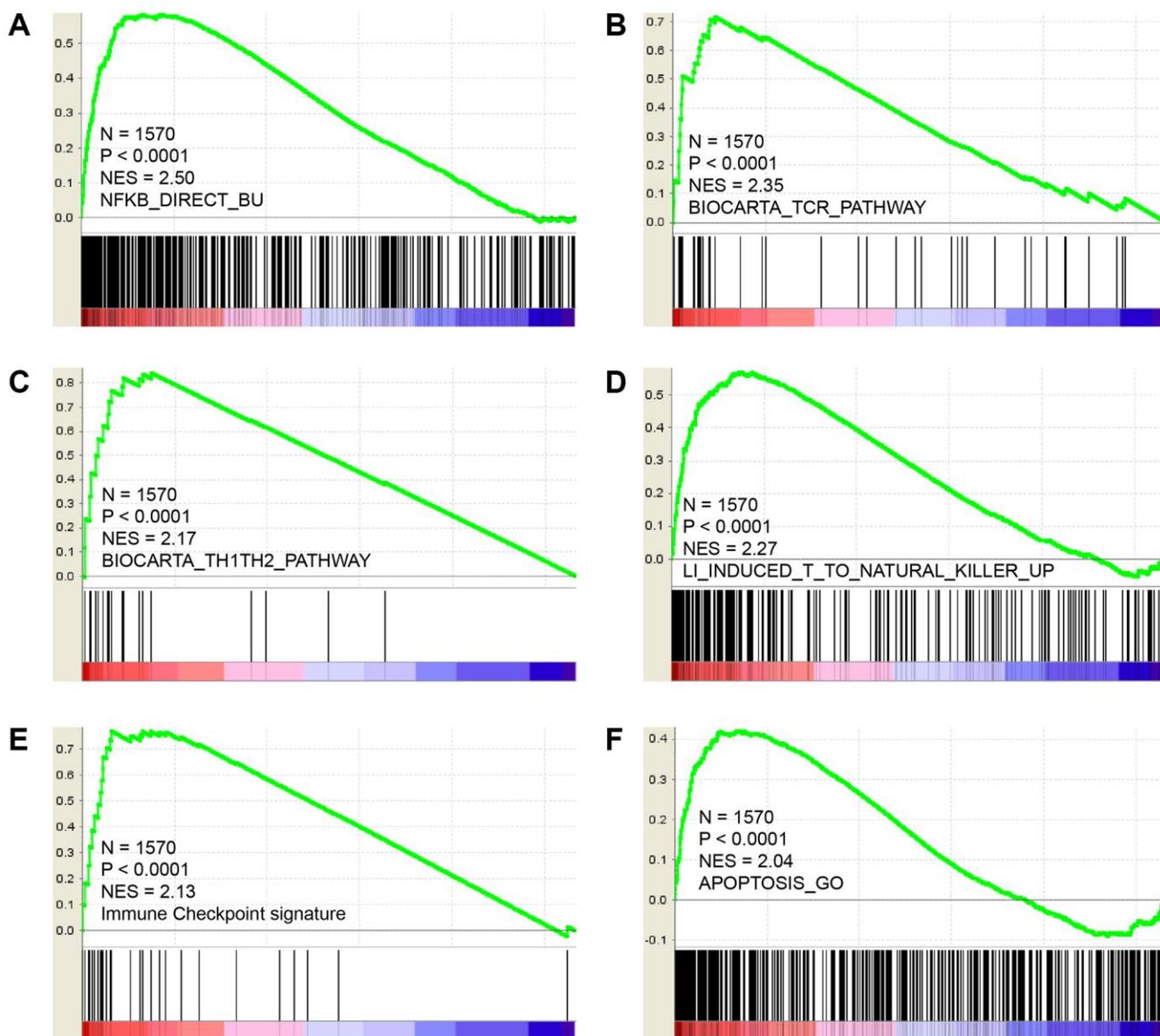
**Figure 1.** (A) CCR5 is overexpressed in basal-like or TNBC subtypes of breast cancer compared to not basal-like or non-TNBC subtypes. (B) The expression of CCR5 is negatively correlated CCR5 promoter methylation levels. (C) Patients with CCR5 high expression have better survivals compared to CCR5 low expression group. (D) CCR5 expression is positively correlated with tumor infiltration immune cells such as B cell, CD8+ T cell, CD4+ T cell, Neutrophil and Dendritic cell.

presented in Figure 3. Figure 3A shows the comparison of mutation profiles between CCR5 high and low groups. We can see that TP53 has the highest mutation rate both in CCR5 high and CCR5 low TNBC patients (78.08% vs. 67.12%). Since the mutation rate of TP53 in breast cancer is around 30%, the high mutation rate of TP53 in TNBC indicates its potential role in TNBC progression. The following mutated genes are TTN, USH2A and PIK3CA etc. (19.18% vs. 23.29%, 6.85% vs. 12.33% and 10.96% vs. 9.59%, respectively). Figure 3B presents the comparison of CNA profiles between CCR5 high and low groups. The top three gene with copy number variation are MYC, EXT1 and RAD21 (Amplification, 37.97% vs. 35.06%, 31.65% vs. 27.27% and 30.38% vs. 27.27%, respectively). Detailed

mutation and CNA profiling data comparison were presented in Supplementary Table 1.

### Prognostic value of CCR5 in TP53 wildtype/mutation TNBC patients

Results in the previous section showed that TP53 was highly mutated in TNBC, the association between CCR5 expression and TP53 mutation was explored in this part. Using expression and mutation data of TCGA, we demonstrated that there was no statistical difference of CCR5 expression in TNBC patient with or without TP53 mutation (Figure 4A). Kaplan-Meier analysis results showed that CCR5 expression is not associated with overall survival of TNBC patients in TP53



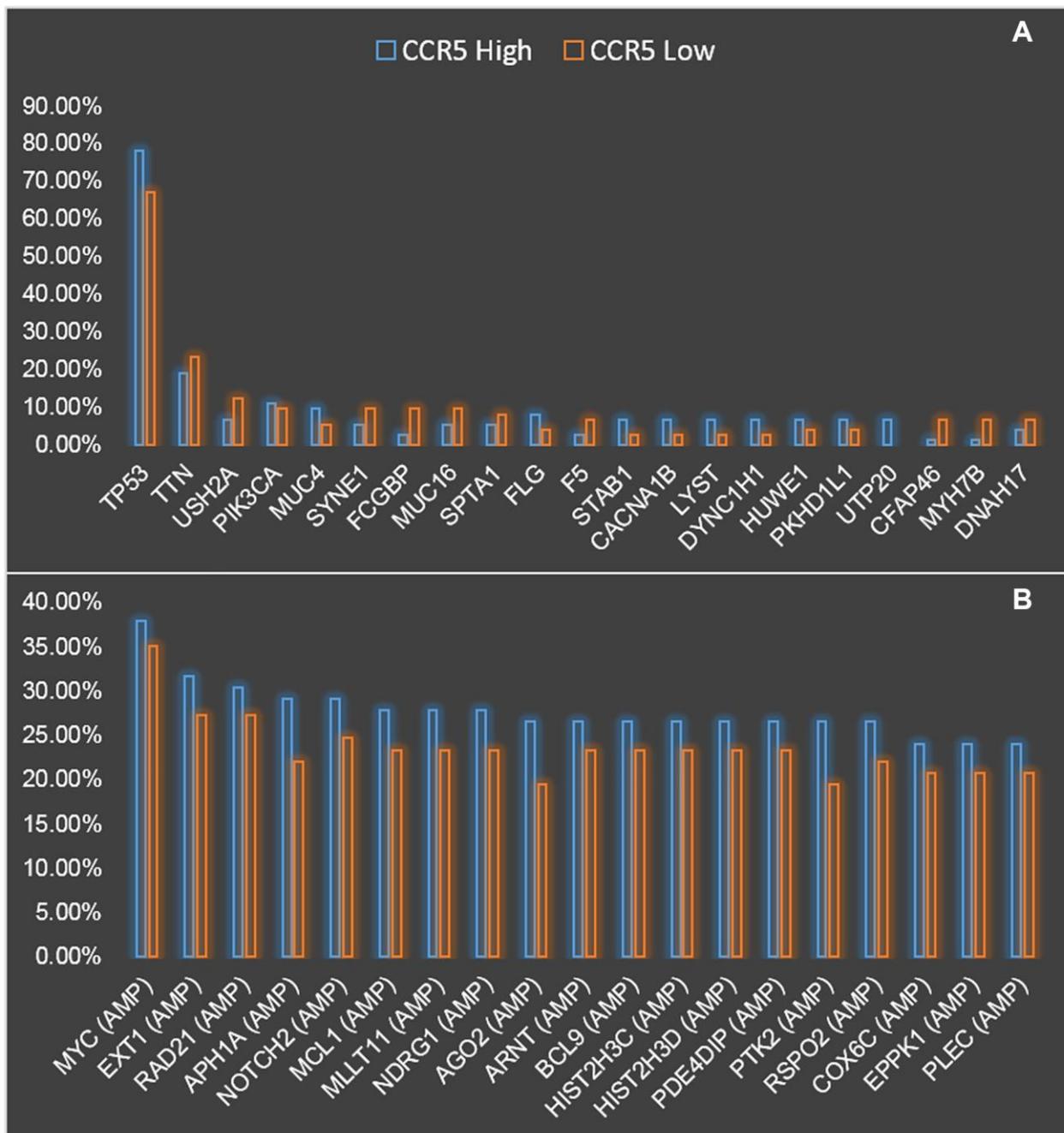
**Figure 2.** GSEA results show that CCR5 is positively correlated with NF- $\kappa$ B pathway (A), TCR pathway (B), TH1/TH2 pathway (C), Natural killer cell up (D), immune checkpoint signature (E) and Apoptosis (F).

wildtype group (HR = 0.93,  $P = 0.93$ , Figure 4B), while high expression of CCR5 was associated with better overall survival in TNBC patients with TP53 mutation (HR = 0.27, Logrank  $p = 0.0083$ , Figure 4C). The above results indicate that CCR5 expression is not correlated with TP53 mutation status, however, the function of CCR5 in mitigating TNBC progression may rely on TP53 mutation. We tried to explain this interesting phenomenon by analysis the expression pattern of CCR5 and immune checkpoint markers in TNBC patients with or without P53 mutation (Supplementary

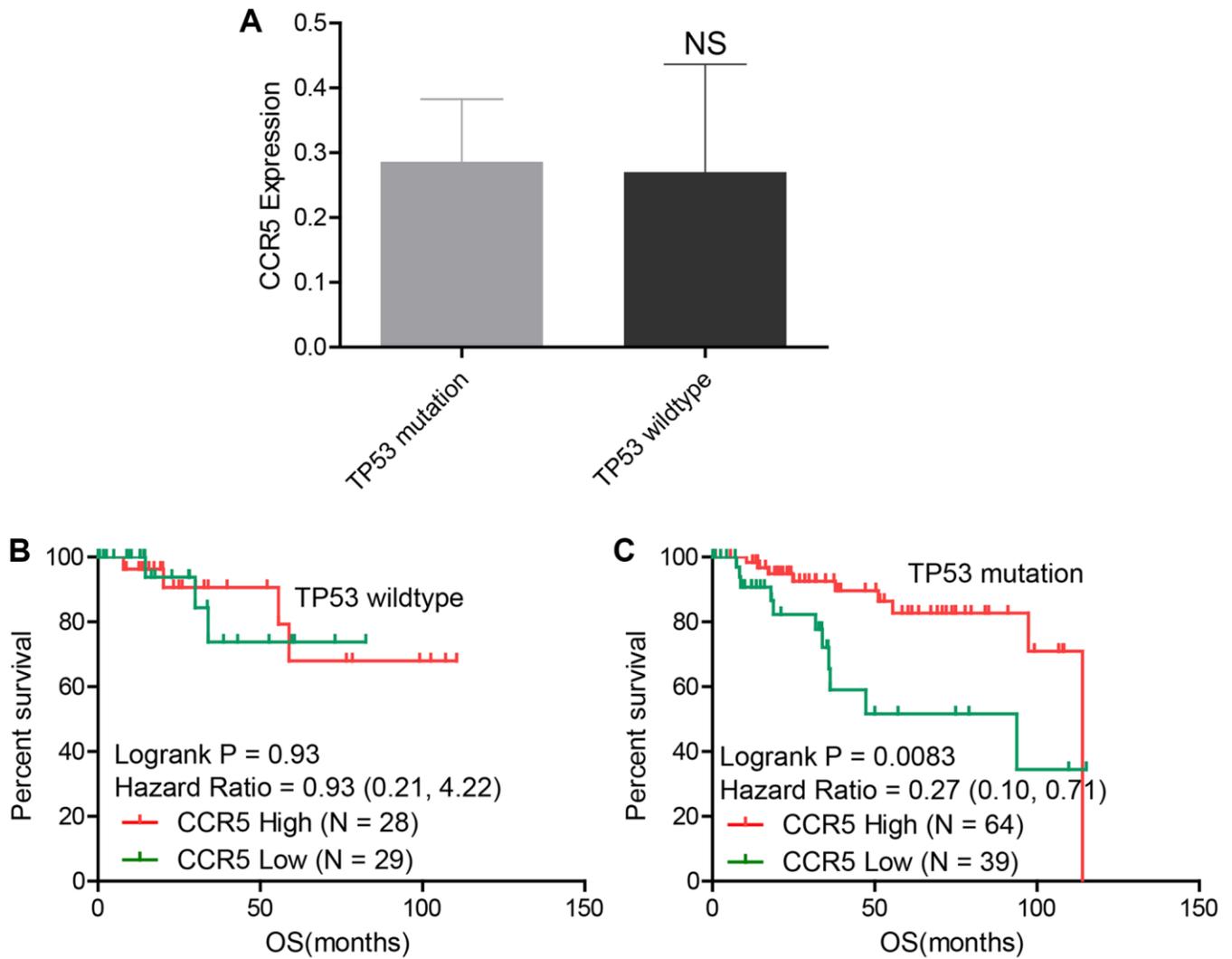
Figures 2–3). The correlation between CCR5 expression and infiltrated immune cells were also visualized in Supplementary Figures 4–5.

### Proteomic analysis of CCR5 high and low expressed TNBC patients

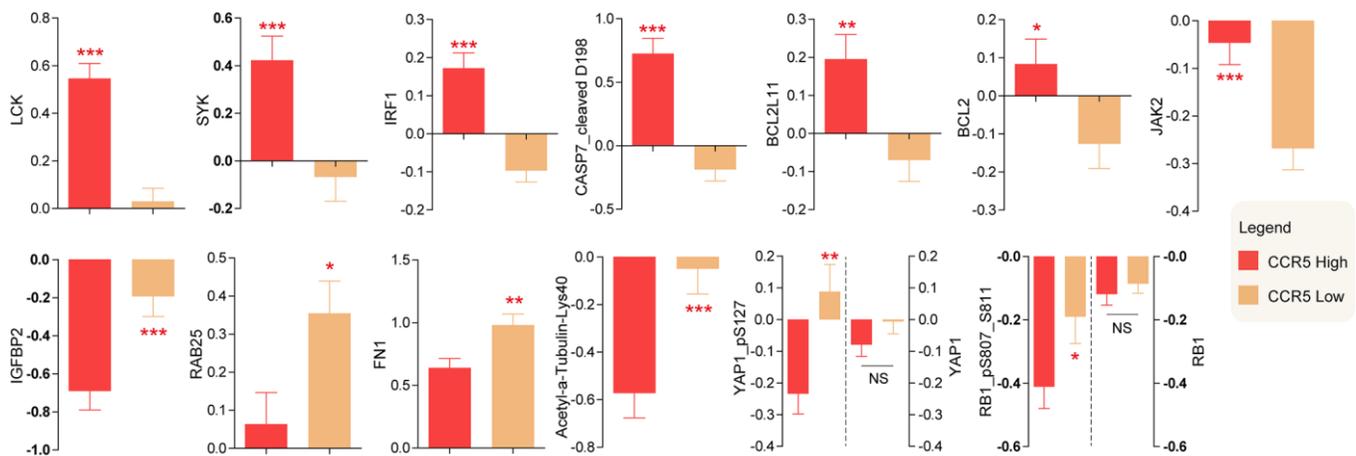
Reverse phase protein array (RPPA) data of TNBC, which containing 226 antibodies, was downloaded from TCGA. Differentially changed proteins were computed and top changed proteins were presented in Figure 5



**Figure 3.** Comparison of mutation (A) and CNV (B) landscapes between CCR5 high expression and low expression groups.



**Figure 4.** (A) There is no statistical difference of CCR5 expression between TP53 wildtype and TP53 mutation groups in TNBC samples. (B) CCR5 expression is not associated with OS in TNBC patients with wildtype TP53. (C) Patients with high CCR5 expression have better survival in TP53 mutation TNBC samples.



**Figure 5.** Differentially expressed proteins between CCR5 high expression and low expression groups.

(For detailed information, please see Supplementary Table 1). LCK, SYK, IRF1, CASP7\_cleaved D198, BCL2L11, BCL2 and JAK2 were higher in TNBC patients with high CCR5 expression, while IGFBP2, RAB25, FN1, Acetyl- $\alpha$ -tubulin-Lys40, YAP1\_pS127, RB1\_pS807\_S811 were lower in CCR5 high TNBC patients. There are no statistical differences of YAP1 and RB1 between CCR5 high and CCR5 low TNBC patients. These proteomic changes might partly account for CCR5 associated good prognosis.

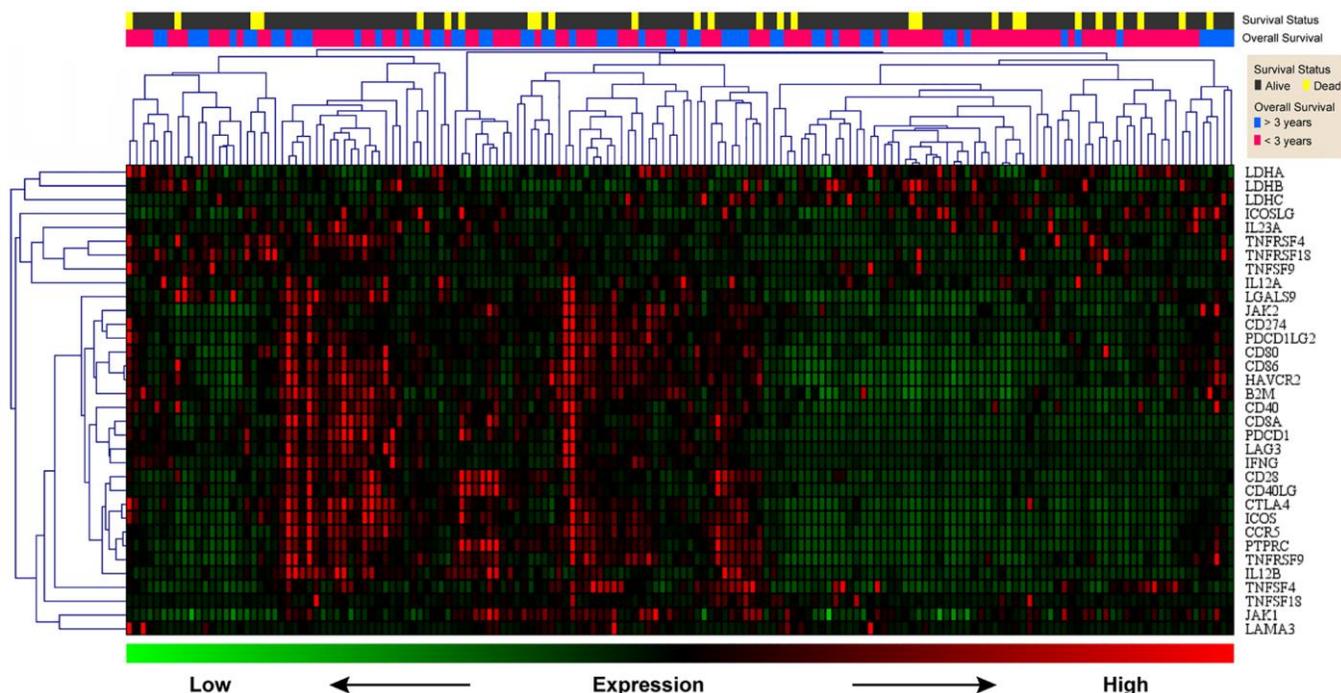
### Clustering analysis of CCR5 and immune checkpoint related signatures in TNBC samples

Previously, we demonstrated that CCR5 expression was correlated with tumor immune cell infiltration. Since TILs were often associated with the efficacy of immunotherapy, the prognostic of CCR5 in immunotherapy would be very interesting. Here, we performed ROC curve analyses using gene expression data of melanoma treated with immune checkpoint inhibitor (GSE91061, pre-treatment and on-treatment) to show the predictive power of CCR5 in predicting disease control rate (DCR). The AUC of pre-treatment CCR5 value (pre), on-treatment CCR5 value (on) and 'on-treatment minus pre-treatment CCR5 value (on-pre)' were 0.5509, 0.7338 and 0.7546, respectively (Supplementary Figure 6). The prognostic power of on-pre was the highest (sensitivity = 75%, specificity = 77.78%). These results showed that CCR5 elevation

after immune checkpoint blockade (ICB) was a significant prognostic marker, which might also have therapeutic implications. Since CCR5 is associated with ICB efficacy, heat map and cluster analysis were performed using TNBC data from TCGA through MeV to show the expression pattern of CCR5 and immune checkpoint related genes. Figure 6 shows that these genes are clustered into three main groups: Lactate dehydrogenase (LDH); ICOSLG, IL23A, TNFRSF4, TNFRSF18, TNFSF9 and IL12A; CCR5 and other ICB efficacy related genes including PDCD1 (PD1), CD274 (PDL1), JAK2 and LAG3. Specifically, CCR5 expression pattern is more similar to PTPRC, TNFRSF9, ICOS and CTLA4. The above results indicate that CCR5 is a prognostic marker for ICB treatment and could serve as a potential therapeutic target. Cox regression for survival analysis using CCR5 and immune checkpoint related genes was performed. Risk groups were divided using best cutoff point of risk score. Figure 7 indicated that patients in low risk group had better OS than high risk group.

### Association analyses of CCR5 and ICB efficacy related genes in TNBC samples

Next, we discussed the latest findings on molecular basis of ICB resistance and analyzed the association between CCR5 and those genes. We demonstrate that in TNBC patients CCR5 is positively correlated with IFN- $\gamma$ , IL12B and key non-canonical NF- $\kappa$ B pathway genes,



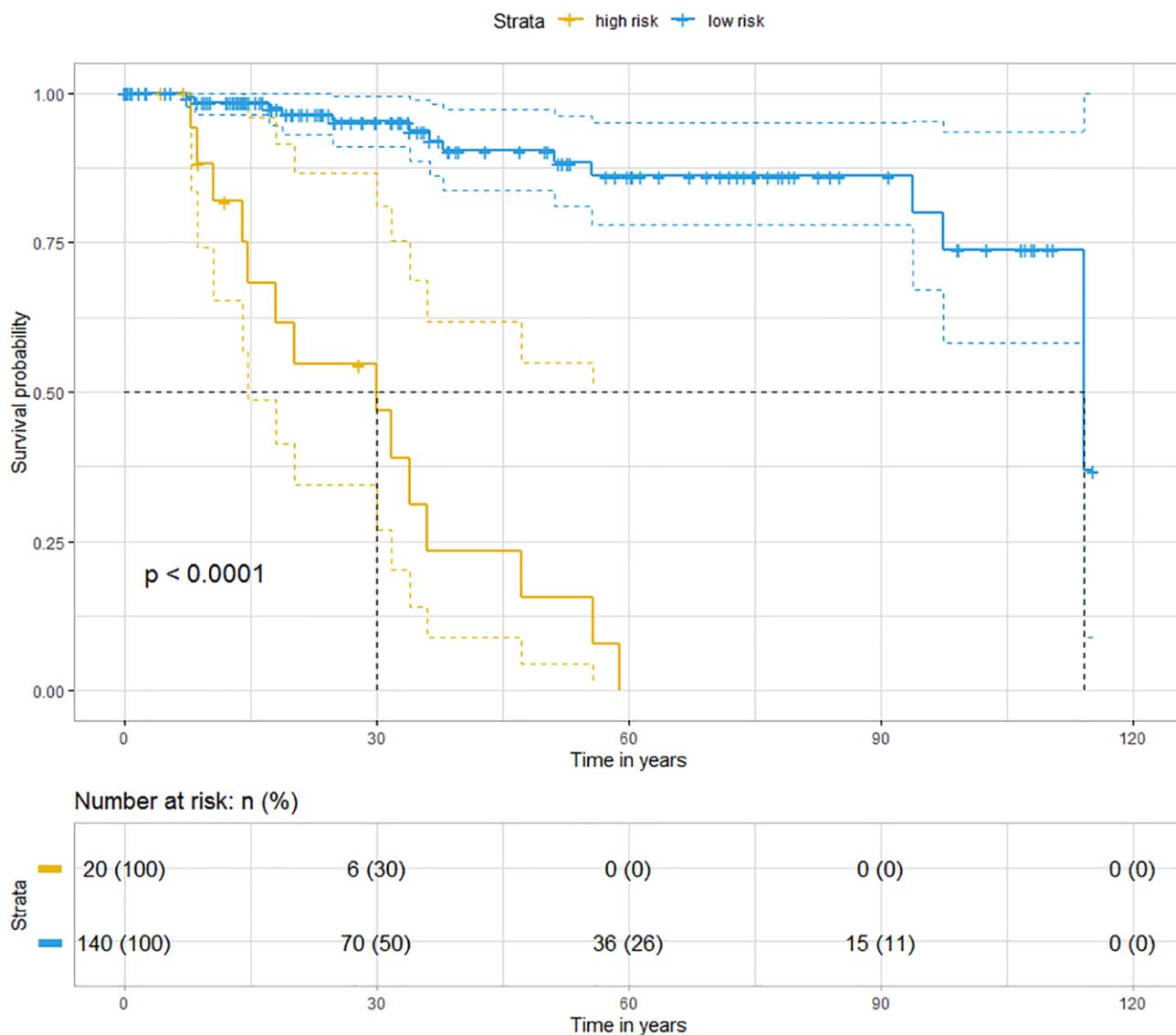
**Figure 6. Heat map of CCR5 and immune checkpoint related genes.** Green represents low expression level and red represents high expression level. The two upper rows represent survival and survival status, respectively.

namely CD40, ABCB11, NFKB2, RELB and MAP3K14 (Figure 8A and Figure 8C–8H). Besides, CCR5 is also positively correlated with T cell costimulator CD28 and immune checkpoint receptor LAG3 (Figure 8I–8J). While the correlation between CCR5 and IL12A does not have statistical significance (Figure 8B). FGL1 is overexpressed in breast cancer in comparison to matched normal control (Supplementary Figure 7A,  $P = 0.0039$ ), but there is no statistical difference between basal-like and non-basal-like group of breast cancer (Supplementary Figure 7B). The correlation between CCR5 and FGL1 does not have statistical significance (Supplementary Figure 7C). SIGLEC15 and YTHDF1 are overexpressed in breast cancer and especially non-basal-like subtype (Supplementary Figures 8A–8B and 9A–9B). There is no correlation between the expression of CCR5 and

SIGLEC15, YTHDF1 (Supplementary Figures 8C and 9C,  $P = 0.347$  and  $0.422$ , respectively). The above results indicate that in TNBC patients with CCR5 overexpression, T Cell-DCs crosstalk involving IFN- $\gamma$  and IL12 is activated, which means these patients would show better therapeutic efficacy to anti-PD1 immunotherapy. Since the expression of SIGLEC and YTHDF1 are higher in not basal-like breast cancer subtype, these genes might be more promising therapeutic targets for non-basal-like subtype or breast cancer patients with low CCR5 expression.

## DISCUSSION

Great success has been achieved in cancer management in recent years. For instance, use implanted 3D-Printed vertebral bodies with robotic stereotactic radiotherapy



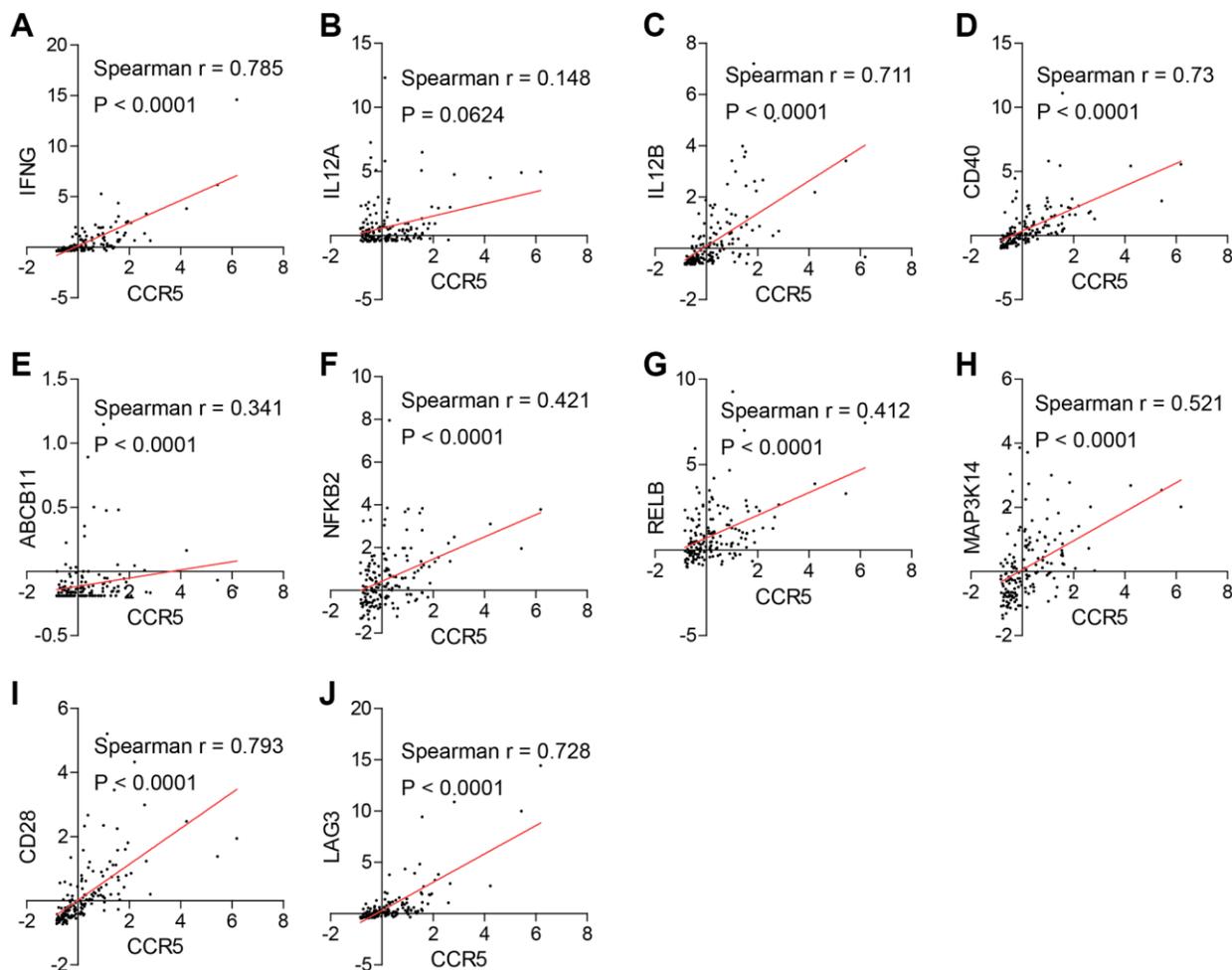
**Figure 7.** Kaplan-Meier analyses indicate that patients in low risk group have better OS than those in high risk group (blue represents low risk group and yellow represents high risk group, Logrank  $p < 0.0001$ ).

for spinal tumor treatment [13]; encapsulate irinotecan (CPT-11) into micelle-based nanoparticles for a better efficacy in cancer therapy [14]; immunotherapy using immune checkpoint inhibitors has been applied in several solid tumors including TNBC [15, 16]; development of allele-specific K-RasG12C inhibitors for the treatment of oncogenic KRAS mutant in different cancer types [17]. However, the current situation is far from satisfied. It is still crucial to clarify the molecular mechanisms underlying cancer progression and explore novel prognostic biomarkers and therapeutic targets.

Despite intensive studies on CCR5 in recent years, the roles of CCR5 in cancer remain elusive. Previously, we and our collaborators demonstrated that CCR5 is associated with better overall survival of several cancer types including breast cancer [12]. However, the expression pattern and prognostic value of CCR5 in different breast cancer subtypes and underlying mechanistic insights still needs to be clarified.

In this study, we showed that CCR5 is overexpressed in TNBC compared to non-TNBC or normal control and is associated with better prognosis of TNBC. CCR5 expression is positively correlated with tumor immune cell infiltration and tumor immune response related pathways. Multi-omics data of TNBC were compared based on CCR5 expression levels and CCR5 associated genomic and proteomic changes were identified. CCR5 overexpression was associated with better OS in TNBC patients with TP53 mutation. We also summarized the latest findings on ICB efficacy related genes and explored the association between CCR5 and those genes. These results indicate that CCR5 is a potential tumor repressor gene and individualized therapeutic strategy could be established based on multi-omics background and expression pattern of ICB related genes. Finally, several drugs that could potentially upregulated CCR5 expression were suggested.

Previously, Pestell et al. reported that CCR5 antagonists maraviroc could reduce invasion and



**Figure 8.** (A, C–J) CCR5 is positively correlated with good prognostic markers of anti-immune checkpoint therapies such as IFNG, IL12B, CD40, ABCB11, NFkB2, RELB, MAP3K14, CD28 and LAG3. (B) The correlation between the expression of CCR5 and IL12A does not have statistical significance ( $p = 0.0624$ ).

metastasis of basal breast cancer cells *in vitro* and *in vivo*. And they suggested that CCR5 antagonists could be used as a therapeutic option for mitigating the risk of metastasis in patients with the basal breast cancer subtype [8]. Our results indicated that CCR5 is associated with longer OS of TNBC patients with TP53 mutation and is positively correlated with tumor immune response. These results seem to be inconsistent with Pestell's results, which could be explained by drawbacks of their experiment design that merely using cancer cell lines and immunodeficient (NOD/SCID) mice model. Moreover, our study is supported by another group from The Scripps Research Institute, which showed that CCR5 expression in both CD4+ and CD8+ T cells was necessary to activate cancer immune response that might have implications for cancer treatment in patients with CCR5 deficiency.

We also summarized the latest findings on the molecular basis of ICB efficacy and discussed the association between CCR5 and those signatures. It is reported that effective anti-PD1 therapy requires the crosstalk between T cells and intratumoral dendritic cells (DCs) [18]. Specifically, anti-PD1 mAb could activate T cells and induce IFN- $\gamma$  production, which further induced IL-12 production by DCs. Effective anti-PD1 therapy requires IL-12 produced by DCs to license effector T cell responses in cancer patients. Agonizing non-canonical NF- $\kappa$ B pathway can induce IL-12 production through DCs activation and enhance anti-PD1 therapy. Moreover, IL-12 and T cell co-stimulator CD28 are required to achieve maximal IFN- $\gamma$  response. Lieping Chen et al. reported in *Cell* [19] that FGL1 was a major immune inhibitory ligand of LAG-3 and blockade of the FGL1-LAG-3 interaction could enhance anti-tumor immunity, suggesting that FGL1-LAG-3 pathway was an important immune evasion mechanism and a potential target for immunotherapy. Three months later, Lieping Chen group identified SIGLEC15 as a critical immune suppressor and its expression was mutually exclusive to PDL1, which implicated its potential therapeutic value in cancer patients especially for those who failed to respond to anti-PDL1 therapy [20]. Chuan He et al. reported a novel immune evasion mechanism in *Nature* that the N6-methyladenosine (m6A) marked transcripts encoding lysosomal proteases were recognized and bounded by YTHDF1 in DCs, which promoted translation of lysosomal proteases for excessive neoantigen degradation, thereby mitigating neoantigen-specific tumor immunity. Furthermore, the efficacy of anti-PDL1 therapy was enhanced in *Ythdf1*<sup>-/-</sup> mice, suggesting YTHDF1 as a promising target for immunotherapy [21]. Here, we show that CCR5 is positively correlated IFN- $\gamma$ , IL12B and key non-

canonical NF- $\kappa$ B pathway genes such as CD40, ABCB11, NFKB2, RELB and MAP3K14. It is also associated with CD28 and LAG3, while there are no correlation between CCR5 and IL12A, FGL1, SIGLEC15 and YTHDF1. These results suggest that CCR5 may increase the efficacy of anti-PD1 therapy through activating T cell-DCs crosstalk.

Furthermore, we find several drugs that can upregulated CCR5 expression by exploring CTD database. For instance, Cisplatin, cyclophosphamide, Oxaliplatin, Topotecan and Clofibrate can promote CCR5 expression, which implicates their potential application in managing TNBC patients with low CCR5 expression (Supplementary Figure 10).

In summary, CCR5 is overexpressed in TNBC and is associated with better prognosis of TNBC with TP53 mutation. Potential mechanisms may include activation of certain tumor suppressors while repressing some oncogenic pathways such as YAP1. Activation of effector T cell may also account for CCR5 related tumor immune response. All these data suggest that CCR5 is a prognostic marker and potential therapeutic target for TNBC with TP53 mutation. Further wet lab experiments and clinical trials are warranted.

## MATERIALS AND METHODS

### Ethics statement

All the data used in this study were downloaded from publicly available sources. The Research Ethics Committee of Zhejiang Provincial people's Hospital and National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital waived the requirement for ethical approval.

### Data source

Gene expression data for non-basal-like/basal-like or non-TNBC/TNBC comparison were obtained from bc-GenExMiner database. Gene expression data of CCR5 and other immune related genes in TNBC were downloaded from The Cancer Genome Atlas (TCGA: <http://cancergenome.nih.gov/>). Mutation, Methylation, Protein expression and copy number alteration data of TNBC were also obtained from TCGA. Data for survival analyses were downloaded from KMplot [22] and TCGA. All other expression data were obtained from Gene Expression Omnibus (GEO) [23]. Specifically, GSE70947 [24] was used for comparing expression levels between breast cancer and paired normal control; GSE47561 [25] ( $N = 1570$ ) was used for Gene Set Enrichment Analysis (GSEA) [26]; GSE96061 [27] was used for ROC (Receiver operating

characteristic) curve analysis. Chemical-gene interaction data was downloaded from The Comparative Toxicogenomics Database (CTD base) [28]. The abundance data of infiltrated immune cells in TCGA TNBC data was obtained from xCell [29].

### Bioinformatics and statistical analyses

Heat map and clustering analysis were performed using MeV software (<http://mev.tm4.org>). GSEA was performed to show the functional enrichment of CCR5 in breast cancer. Immune infiltration analysis was using TIMER [30]. R 4.0.0 (R Foundation for Statistical Computing (<http://www.r-project.org/>)) or GraphPad Prism 5.01 (GraphPad Software, Inc. (<http://www.graphpad.com>)) were utilized to perform all other statistical analyses. Cox regression and related survival analysis were performed using ‘survival’ and ‘survminer’ packages [31, 32]. Correlation analysis and visualization were performed using ‘corrplot’ package [33]. Standard statistical tests including paired *t*-test, fisher exact test and independent samples *t*-test were employed in the data analyses. Adjust *P* value was corrected for multiple comparisons using the Benjamini and Hochberg's false discovery rate [34]. Significance was defined as a *P* value < 0.05.

### Availability of data

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### AUTHOR CONTRIBUTIONS

Xin Wang and Yong Han contributed to manuscript drafting and revising. Xin Wang, Jiamin Peng and Yong Han contributed to data analysis and interpretation. Yong Han and Jie He contributed to the conception of the study and final approval.

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

### FUNDING

This work was supported by National Natural Science Foundation of China, Grant Number: 81702774. Zhejiang Provincial Administration of Traditional Chinese Medicine, Grant Number: 2016ZB018. Natural Science Foundation of Zhejiang Province, No. LY17H160065. Zhejiang Provincial Medical and Healthy Science Foundation (no. 2014KYA077). Major Project of Medical and Health Science and Technology in Zhejiang Province (WKJ-ZJ-1715).

### REFERENCES

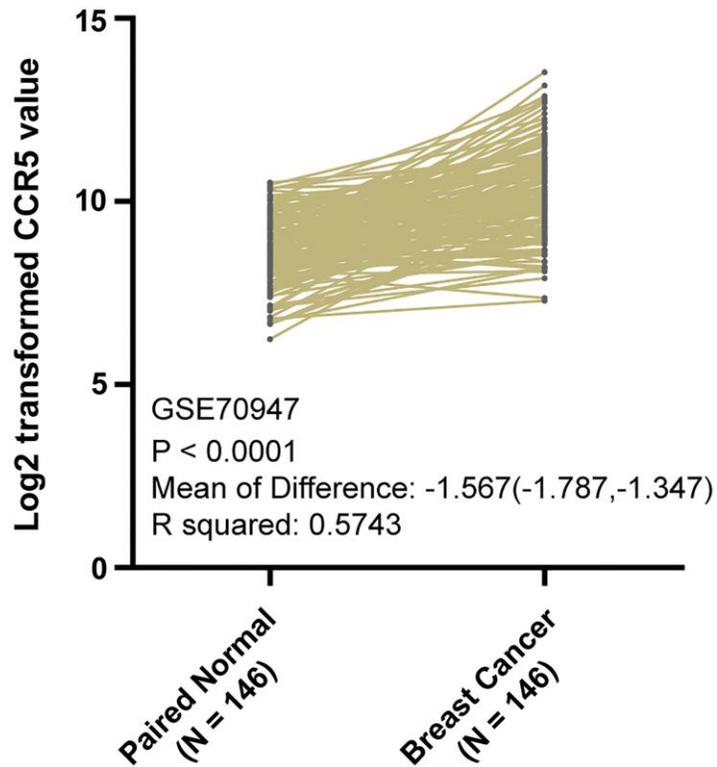
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:33538338
2. Feng M, Pan Y, Kong R, Shu S. Therapy of Primary Liver Cancer. *Innovation (N Y).* 2020; 1:100032. <https://doi.org/10.1016/j.xinn.2020.100032> PMID:32914142
3. Gu C, Shi X, Dai C, Shen F, Rocco G, Chen J, Huang Z, Chen C, He C, Huang T, Chen C. RNA m<sup>6</sup>A Modification in Cancers: Molecular Mechanisms and Potential Clinical Applications. *Innovation (N Y).* 2020; 1:100066. <https://doi.org/10.1016/j.xinn.2020.100066> PMID:34557726
4. Yang X, Weng X, Yang Y, Zhang M, Xiu Y, Peng W, Liao X, Xu M, Sun Y, Liu X. A combined hypoxia and immune gene signature for predicting survival and risk stratification in triple-negative breast cancer. *Aging (Albany NY).* 2021; 13:19486–509. <https://doi.org/10.18632/aging.203360> PMID:34341184
5. Jin L, Luo C, Wu X, Li M, Wu S, Feng Y. LncRNA-HAGLR motivates triple negative breast cancer progression by regulation of WNT2 via sponging miR-335-3p. *Aging (Albany NY).* 2021; 13:19306–16. <https://doi.org/10.18632/aging.203272> PMID:34375306
6. Liu N, Wang X, Li X, Lv X, Xie H, Guo Z, Wang J, Dou G, Du Y, Song D. Identification of effective natural PIK3CA H1047R inhibitors by computational study. *Aging (Albany NY).* 2021; 13:20246–57. <https://doi.org/10.18632/aging.203409> PMID:34415239
7. Wei X, Nielsen R. CCR5-Δ32 is deleterious in the homozygous state in humans. *Nat Med.* 2019; 25:909–10. <https://doi.org/10.1038/s41591-019-0459-6> PMID:31160814. Retraction in: *Nat Med.* 2019; 25:1796. <https://doi.org/10.1038/s41591-019-0637-6> PMID:31595084
8. Velasco-Velázquez M, Jiao X, De La Fuente M, Pestell TG, Ertel A, Lisanti MP, Pestell RG. CCR5 antagonist blocks metastasis of basal breast cancer cells. *Cancer Res.* 2012; 72:3839–50. <https://doi.org/10.1158/0008-5472.can-11-3917> PMID:22637726

9. Pervaiz A, Zepp M, Mahmood S, Ali DM, Berger MR, Adwan H. CCR5 blockage by maraviroc: a potential therapeutic option for metastatic breast cancer. *Cell Oncol (Dordr)*. 2019; 42:93–106. <https://doi.org/10.1007/s13402-018-0415-3> PMID:30456574
10. González-Martín A, Mira E, Mañes S. CCR5 in cancer immunotherapy: More than an "attractive" receptor for T cells. *Oncoimmunology*. 2012; 1:106–8. <https://doi.org/10.4161/onci.1.1.17995> PMID:22720226
11. Aldinucci D, Casagrande N. Inhibition of the CCL5/CCR5 Axis against the Progression of Gastric Cancer. *Int J Mol Sci*. 2018; 19:1477. <https://doi.org/10.3390/ijms19051477> PMID:29772686
12. Zhang J, Wang J, Qian Z, Han Y. CCR5 is Associated With Immune Cell Infiltration and Prognosis of Lung Cancer. *J Thorac Oncol*. 2019; 14:e102–3. <https://doi.org/10.1016/j.jtho.2018.12.037> PMID:31027743
13. Zhuang H, Wei F, Jiang L, Wang Y, Liu Z. Assessment of Spinal Tumor Treatment Using Implanted 3D-Printed Vertebral Bodies with Robotic Stereotactic Radiotherapy. *Innovation (N Y)*. 2020; 1:100040. <https://doi.org/10.1016/j.xinn.2020.100040> PMID:34557713
14. Li C, Xu J, Gan Y, Liang XJ. Innovative Irinotecan-Loaded Nanomicelles Will Enter Phase I Clinical Trial in 2021. *Innovation (N Y)*. 2020; 1:100057. <https://doi.org/10.1016/j.xinn.2020.100057> PMID:34557721
15. Lu Z, Peng Z, Liu C, Wang Z, Wang Y, Jiao X, Li J, Shen L. Current Status and Future Perspective of Immunotherapy in Gastrointestinal Cancers. *Innovation (N Y)*. 2020; 1:100041. <https://doi.org/10.1016/j.xinn.2020.100041> PMID:34557714
16. O'Meara TA, Tolaney SM. Tumor mutational burden as a predictor of immunotherapy response in breast cancer. *Oncotarget*. 2021; 12:394–400. <https://doi.org/10.18632/oncotarget.27877> PMID:33747355
17. Jiao D, Yang S. Overcoming Resistance to Drugs Targeting *KRAS*<sup>G12C</sup> Mutation. *Innovation (N Y)*. 2020; 1:100035. <https://doi.org/10.1016/j.xinn.2020.100035> PMID:32939510
18. Garris CS, Arlauckas SP, Kohler RH, Trefny MP, Garren S, Piot C, Engblom C, Pfirschke C, Siwicki M, Gungabeesoon J, Freeman GJ, Warren SE, Ong S, et al. Successful Anti-PD-1 Cancer Immunotherapy Requires T Cell-Dendritic Cell Crosstalk Involving the Cytokines IFN- $\gamma$  and IL-12. *Immunity*. 2018; 49:1148–61.e7. <https://doi.org/10.1016/j.immuni.2018.09.024> PMID:30552023
19. Wang J, Sanmamed MF, Datar I, Su TT, Ji L, Sun J, Chen L, Chen Y, Zhu G, Yin W, Zheng L, Zhou T, Badri T, et al. Fibrinogen-like Protein 1 Is a Major Immune Inhibitory Ligand of LAG-3. *Cell*. 2019; 176:334–47.e12. <https://doi.org/10.1016/j.cell.2018.11.010> PMID:30580966
20. Wang J, Sun J, Liu LN, Flies DB, Nie X, Toki M, Zhang J, Song C, Zarr M, Zhou X, Han X, Archer KA, O'Neill T, et al. Siglec-15 as an immune suppressor and potential target for normalization cancer immunotherapy. *Nat Med*. 2019; 25:656–66. <https://doi.org/10.1038/s41591-019-0374-x> PMID:30833750
21. Han D, Liu J, Chen C, Dong L, Liu Y, Chang R, Huang X, Liu Y, Wang J, Dougherty U, Bissonnette MB, Shen B, Weichselbaum RR, et al. Anti-tumour immunity controlled through mRNA m<sup>6</sup>A methylation and YTHDF1 in dendritic cells. *Nature*. 2019; 566:270–4. <https://doi.org/10.1038/s41586-019-0916-x> PMID:30728504
22. Györfy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*. 2010; 123:725–31. <https://doi.org/10.1007/s10549-009-0674-9> PMID:20020197
23. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013; 41:D991–5. <https://doi.org/10.1093/nar/gks1193> PMID:23193258
24. Quigley DA, Tahiri A, Lüders T, Riis MH, Balmain A, Børresen-Dale AL, Bukholm I, Kristensen V. Age, estrogen, and immune response in breast adenocarcinoma and adjacent normal tissue. *Oncoimmunology*. 2017; 6:e1356142. <https://doi.org/10.1080/2162402X.2017.1356142> PMID:29147603
25. Ur-Rehman S, Gao Q, Mitsopoulos C, Zvelebil M. ROCK: a resource for integrative breast cancer data analysis. *Breast Cancer Res Treat*. 2013; 139:907–21. <https://doi.org/10.1007/s10549-013-2593-z> PMID:23756628
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub

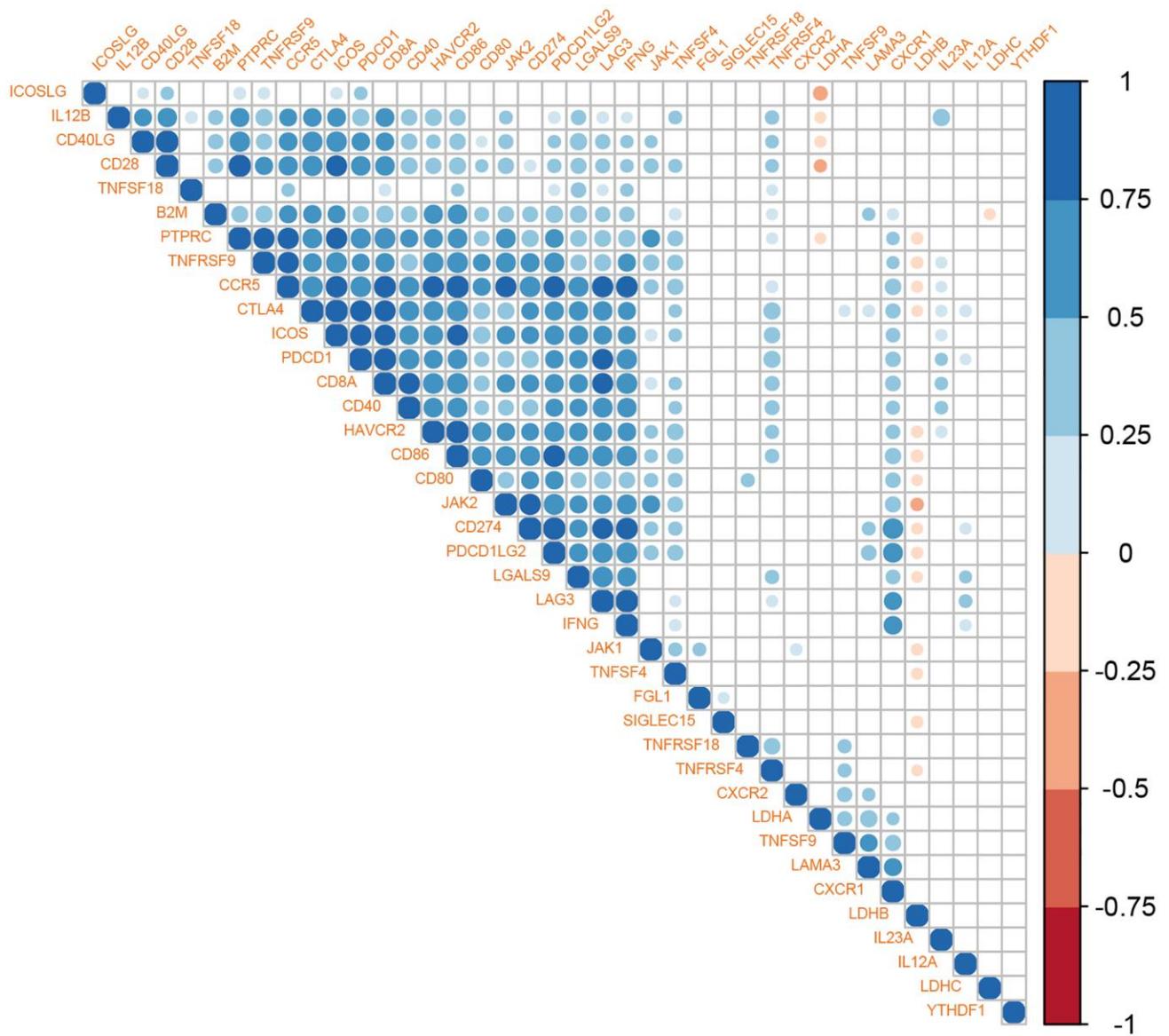
- TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50.  
<https://doi.org/10.1073/pnas.0506580102>  
PMID:16199517
27. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, Hodi FS, Martín-Algarra S, Mandal R, Sharfman WH, Bhatia S, Hwu WJ, Gajewski TF, et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell*. 2017; 171:934–49.e916.  
<https://doi.org/10.1016/j.cell.2017.09.028>  
PMID:29033130
28. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wieggers J, Wieggers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res*. 2019; 47:D948–54.  
<https://doi.org/10.1093/nar/gky868>  
PMID:30247620
29. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017; 18:220.  
<https://doi.org/10.1186/s13059-017-1349-1>  
PMID:29141660
30. Arcolia V, Journe F, Wattier A, Leteurtre E, Renaud F, Gabius HJ, Remmelink M, Decaestecker C, Rodriguez A, Boutry S, Laurent S, Saussez S. Galectin-1 is a diagnostic marker involved in thyroid cancer progression. *Int J Oncol*. 2017; 51:760–70.  
<https://doi.org/10.3892/ijo.2017.4065>  
PMID:28677745
31. Therneau TM. A Package for Survival Analysis in R. R package version 3.2-13. 2021. <https://CRAN.R-project.org/package=survival>.
32. Kassambara A, Kosinski M, Biecek P, Fabian S. survminer: Survival Analysis and Visualization. 2021. <https://rpkgs.datanovia.com/survminer/index.html>.
33. Wei T, Simko V. R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.90). 2021. <https://github.com/taiyun/corrplot>.
34. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol*. 1995; 57:289–300.  
<https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

SUPPLEMENTARY MATERIALS

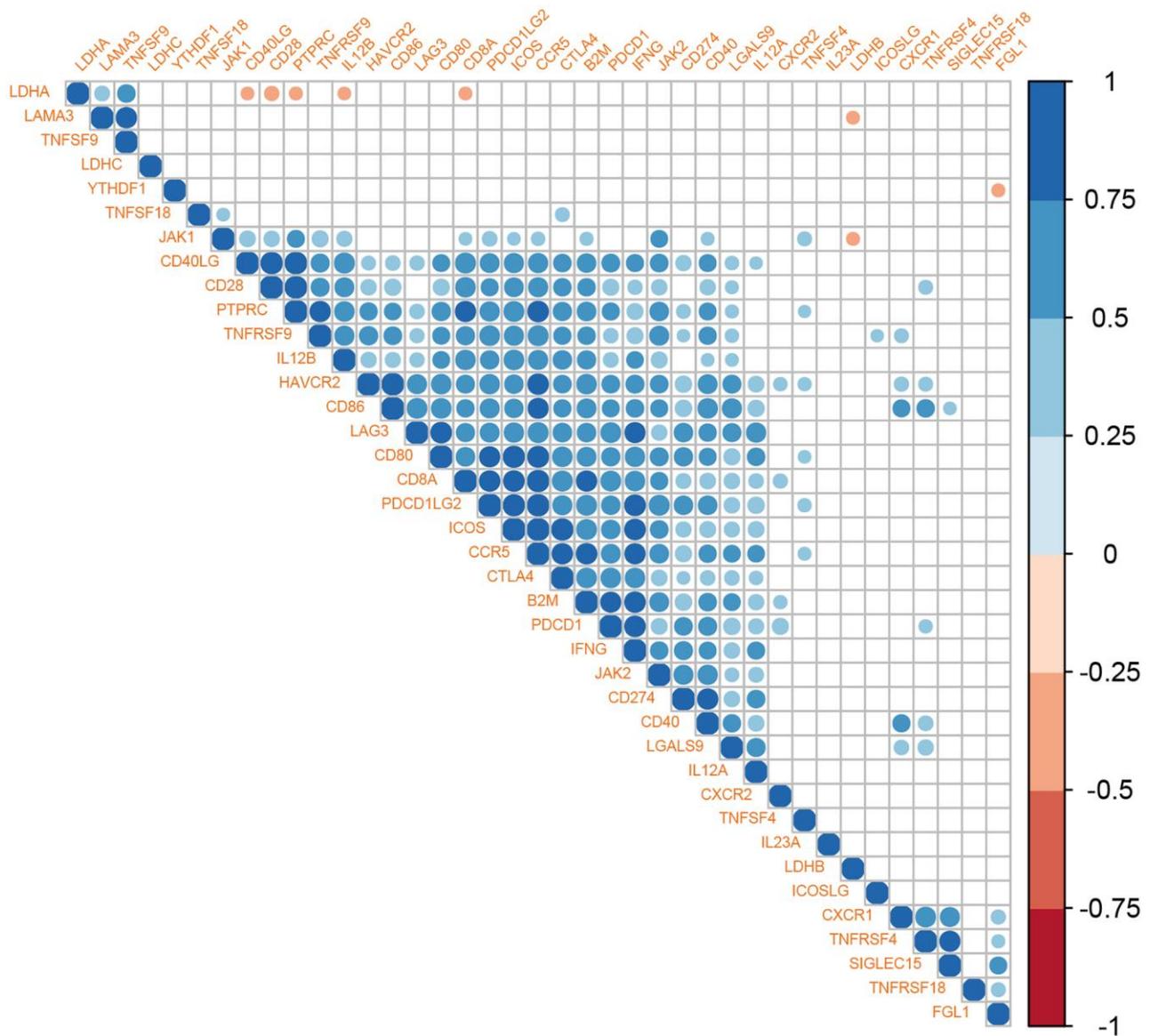
Supplementary Figures



Supplementary Figure 1. CCR5 is overexpressed in breast cancer compared to matched normal control.



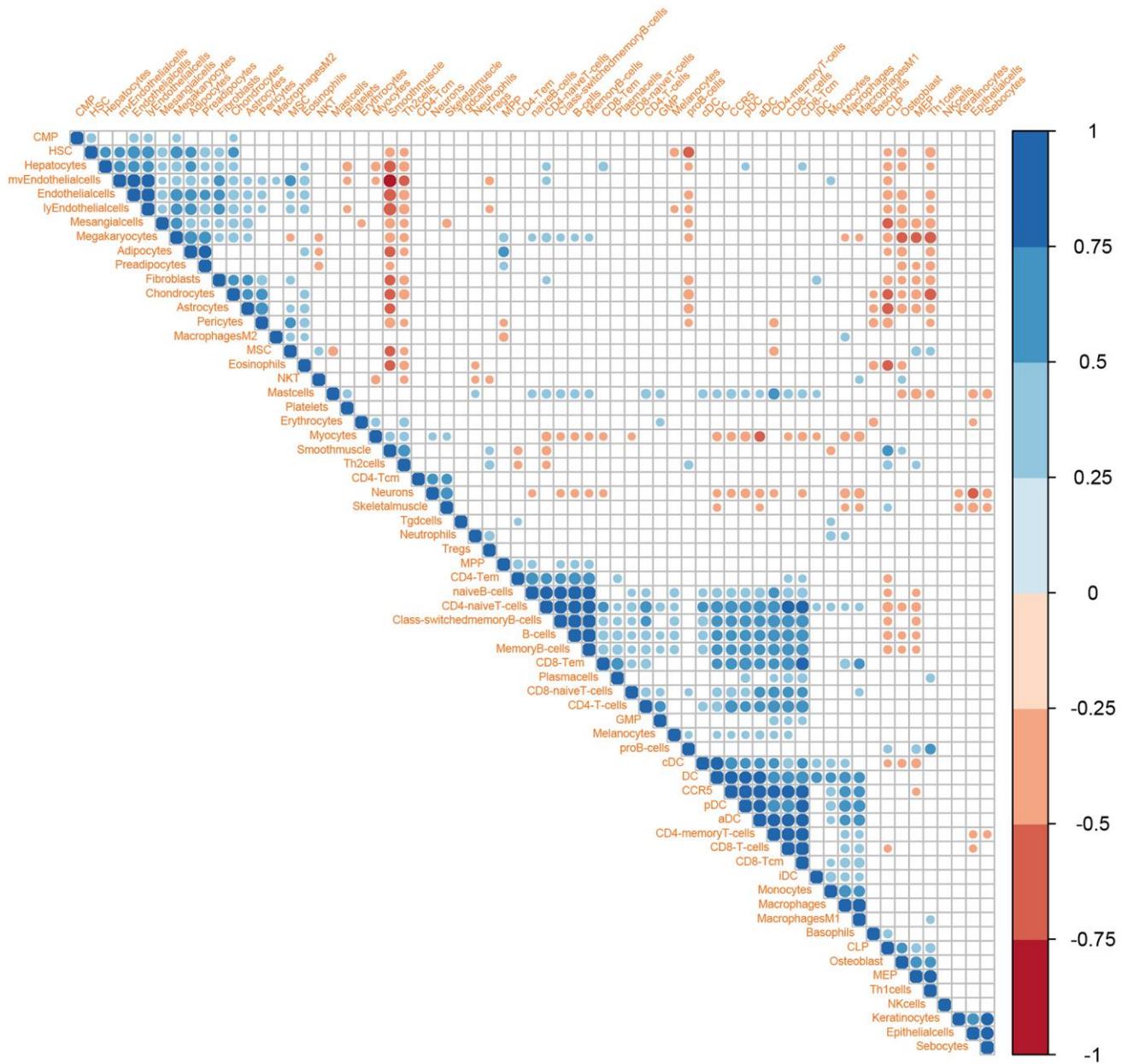
Supplementary Figure 2. The correlation between CCR5 expression and immune checkpoint markers in TNBC patients with P53 mutation.



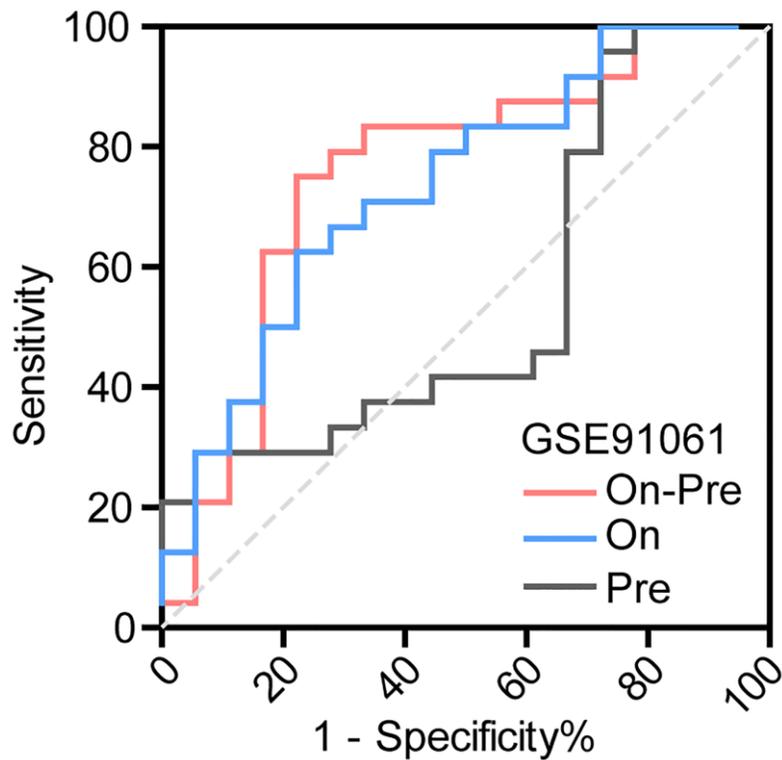
Supplementary Figure 3. The correlation between CCR5 expression and immune checkpoint markers in TNBC patients with wildtype P53.



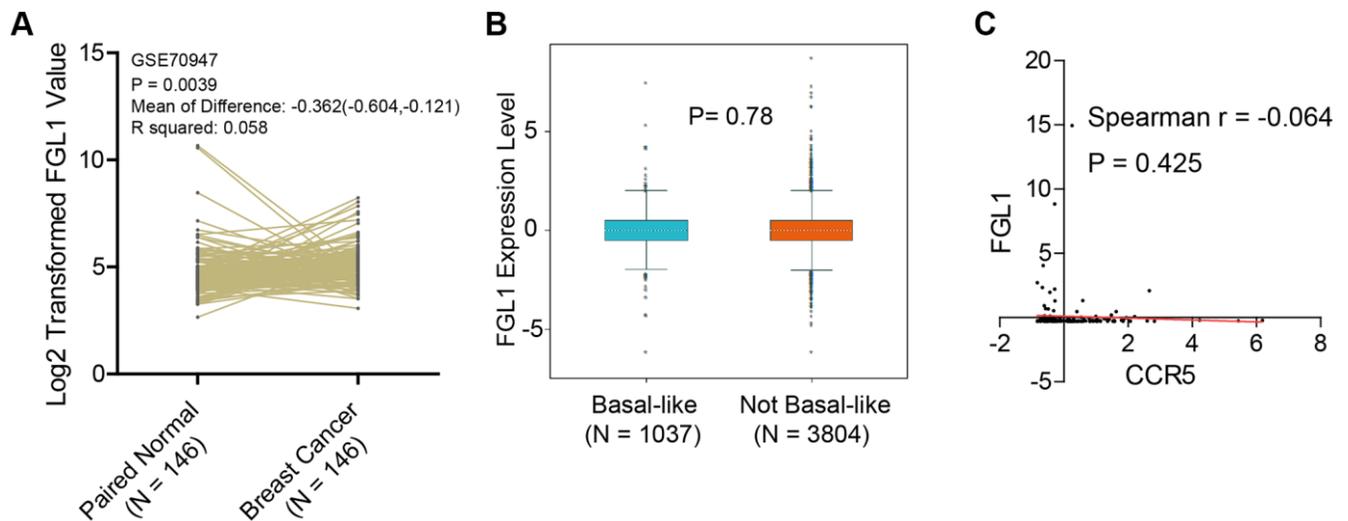
Supplementary Figure 4. The correlation between CCR5 expression and the abundance of tumor infiltrated immune cells in TNBC patients with P53 mutation.



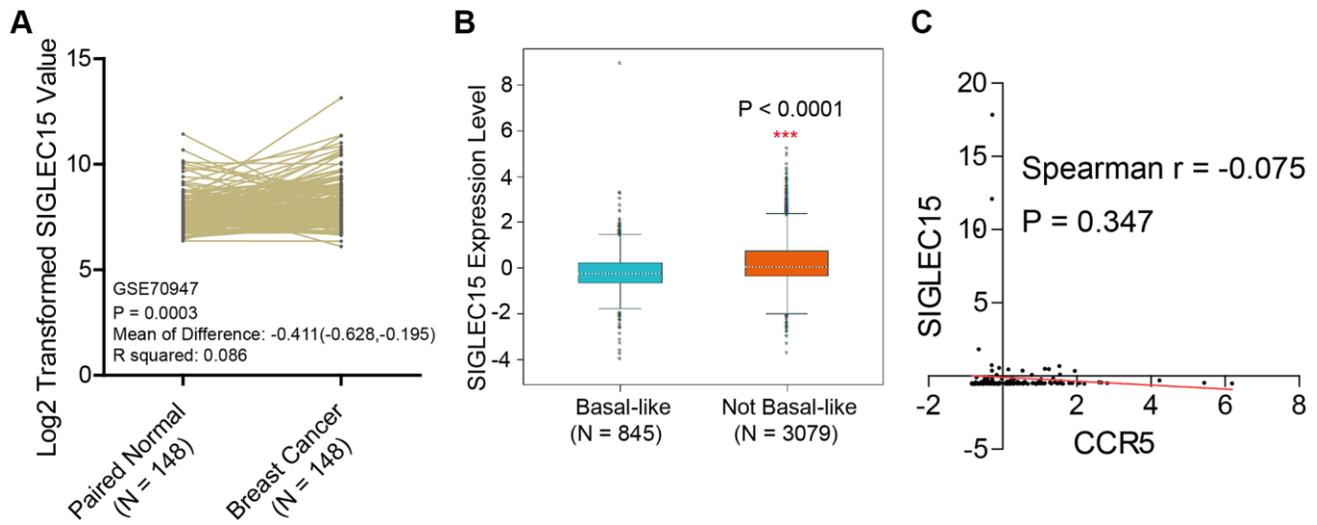
Supplementary Figure 5. The correlation between CCR5 expression and the abundance of tumor infiltrated immune cells in TNBC patients with wildtype P53.



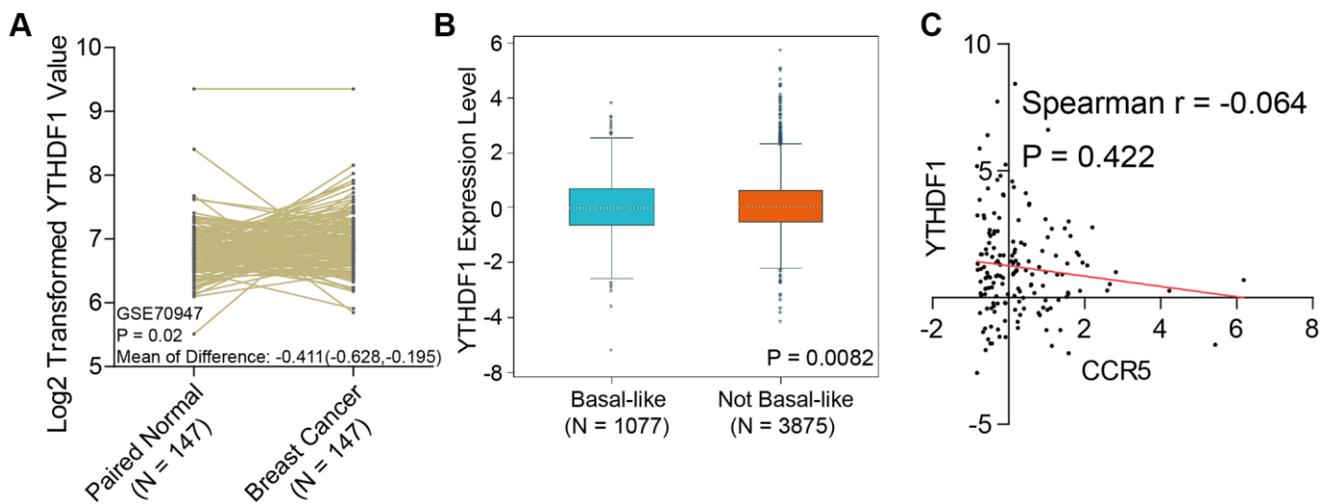
**Supplementary Figure 6. ROC curves of CCR5 in pre-treatment, on-treatment and on-treatment minus pre-treatment.** AUC are 0.5509, 0.7338 and 0.7546, respectively.



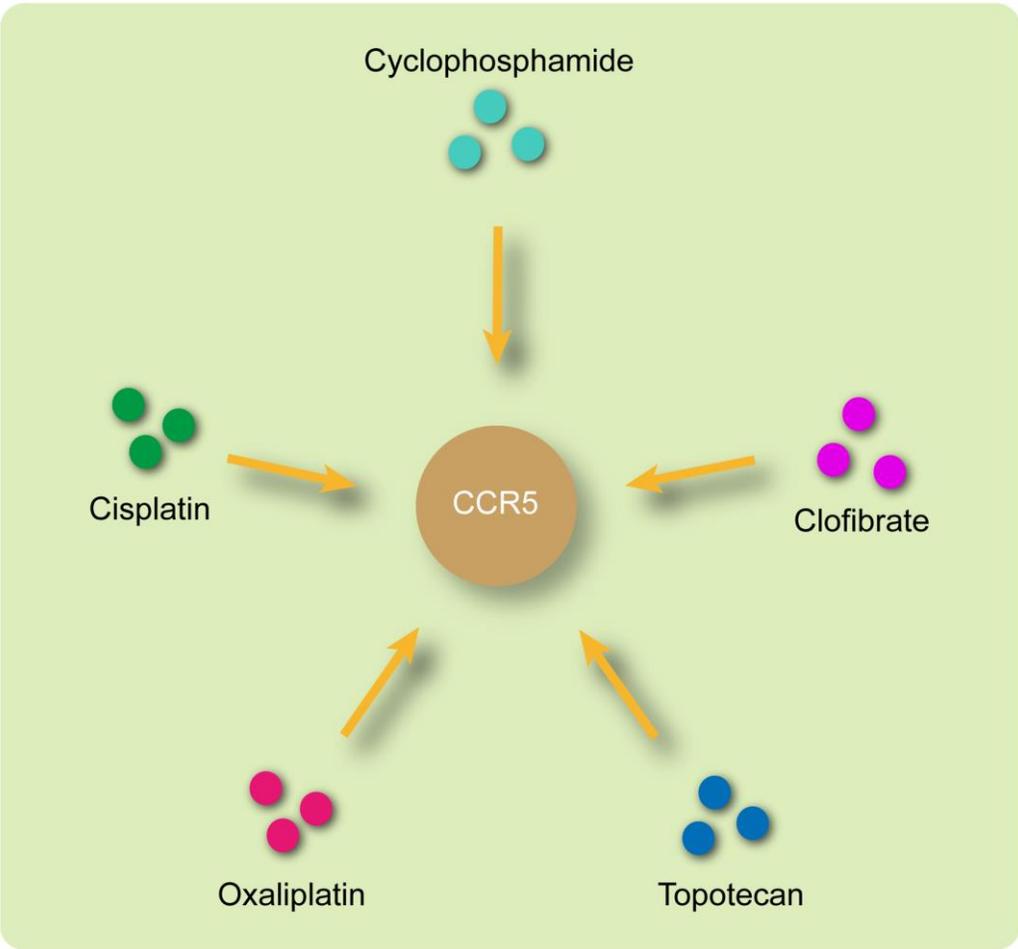
**Supplementary Figure 7. (A)** FGL1 is overexpressed in breast cancer compared to matched normal control. **(B)** There is no statistical difference of FGL1 expression between basal-like subtype and not basal-like subtypes. **(C)** There is no correlation between the expression of CCR5 and FGL1.



**Supplementary Figure 8.** (A) SIGLEC15 is overexpressed in breast cancer compared to matched normal control. (B) SIGLEC15 expression is significantly lower in basal-like subtype compared to not basal-like subtypes. (C) There is no correlation between the expression of CCR5 and SIGLEC15.



**Supplementary Figure 9.** (A) YTHDF1 is overexpressed in breast cancer compared to matched normal control. (B) YTHDF1 expression is significantly lower in basal-like subtype compared to not basal-like subtypes. (C) There is no correlation between the expression of CCR5 and SIGLEC15.



Supplementary Figure 10. Potential drugs that could up-regulate CCR5 expression.

## Supplementary Table

Please browse Full Text version to see the data of Supplementary Table 1.

**Supplementary Table 1. Differentially protein changes between CCR5 high and low groups.**

# THRSP identified as a potential hepatocellular carcinoma marker by integrated bioinformatics analysis and experimental validation

Yuxi Ding<sup>1,2,3,\*</sup>, Xiaoling Liu<sup>1,2,3,\*</sup>, Yue Yuan<sup>1,2,3</sup>, Yunjian Sheng<sup>1,2,3</sup>, Decheng Li<sup>1,2,3</sup>,  
Suvash Chandra Ojha<sup>1,2,3</sup>, Changfeng Sun<sup>1,2,3</sup>, Cunliang Deng<sup>1,2,3</sup>

<sup>1</sup>The Department of Infectious Diseases, The Affiliated Hospital of Southwest Medical University, Luzhou 646000, China

<sup>2</sup>The Department of Tuberculosis, The Affiliated Hospital of Southwest Medical University, Luzhou 646000, China

<sup>3</sup>Laboratory of Infection and Immunity, The Affiliated Hospital of Southwest Medical University, Luzhou 646000, China

\*Co-first author

**Correspondence to:** Changfeng Sun, Cunliang Deng; **email:** [sun\\_cf88@swmu.edu.cn](mailto:sun_cf88@swmu.edu.cn), [dengcunl@swmu.edu.cn](mailto:dengcunl@swmu.edu.cn)

**Keywords:** THRSP, hepatocellular carcinoma, prognosis, invasion, migration

**Received:** January 30, 2021

**Accepted:** January 17, 2022

**Published:** February 23, 2022

**Copyright:** © 2022 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Hepatocellular carcinoma (HCC) is the most common malignant liver tumor with high mortality and poor prognosis worldwide. This study aimed to identify hub genes and investigate the underlying molecular mechanisms in HCC progression by integrated bioinformatics analysis and experimental validation. Based on the Gene Expression Omnibus (GEO) databases and The Cancer Genome Atlas (TCGA), 12 critical differential co-expression genes were identified between tumor and normal tissues. Via survival analysis, we found higher expression of LCAT, ACSM3, IGF1, SRD5A2, THRSP and ACADS was associated with better prognoses in HCC patients. Among which, THRSP was selected for the next investigations. We found that THRSP mRNA expression was negatively correlated with its methylation and closely associated with clinical characteristics in HCC patients. Moreover, THRSP expression had a negative correlation with the infiltration levels of several immune cells (e.g., B cells and CD4+ T cells). qRT-PCR verified that THRSP was lower expressed in HCC tissues and cell lines compared with control. Silencing of THRSP promoted the migration, invasion, proliferation, and inhibited cell apoptosis of HCCLM and Huh7 cell lines. Decreased expression of THRSP promoted HCC progression by NF- $\kappa$ B, ERK1/2, and p38 MAPK signaling pathways. In conclusion, THRSP might serve as a novel biomarker and therapeutic target of HCC.

## INTRODUCTION

Primary liver cancer is one of the most frequently diagnosed malignant tumors and the third leading cause of cancer-related mortality worldwide, with an estimated 906,000 new cases and 830,000 deaths in 2020 [1]. Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer (accounting for 75–85%) [1]. Due to the lack of understanding about the complex carcinogenic mechanisms and efficient therapeutic targets, the 5-year survival rate for HCC

patients remains poor [2]. Thus, there is an urgent need to seek promising targets and elaborate on the underlying molecular mechanisms involved in HCC progression.

The major public databases such as GEO (<http://www.ncbi.nlm.nih.gov/geo/>) and TCGA (<https://portal.gdc.cancer.gov/>), containing gene expression profiles, provide an opportunity to screen the differentially expressed genes (DEGs) related to the carcinogenesis and development of HCC [3, 4]. The

cancer progression is regulated by the key modulators of gene-gene interaction networks, thus the Weighted Gene Co-expression Network Analysis (WGCNA) and protein-protein interaction (PPI) network analysis have been widely used to screen co-expressed genes that drive cancers [5, 6].

In this study, we performed differential gene expression analysis, WGCNA, and PPI network analysis to screen crucial differential co-expression genes associated with hepatocarcinogenesis based on the GEO and TCGA databases. Via 5-year survival analysis, we found that six genes (LCAT, ACSM3, IGF1, SRD5A2, THRSP, and ACADS) were associated with the prognosis of HCC patients. THRSP (thyroid hormone-responsive, also known as Spot 14 or S14) was originally identified in 1982 owing to its significant and rapid induction by thyroid hormone and it had been reported to have great effects on the tissue-specific regulation of lipid metabolism [7, 8]. Some studies demonstrated that THRSP was relatively abundant in liver, white and brown adipose, and lactating mammary tissues and it was associated with nonalcoholic fatty liver disease [9, 10]. THRSP was strongly expressed in most lipogenic breast cancers, and high expression of THRSP predicted a high recurrence rate of primary invasive breast cancers. THRSP mediated lipogenic effects of progesterin, and THRSP knockdown disrupted lipid synthesis and induced apoptosis of breast cancer cells [11]. Another study reported that over-expression of THRSP increased medium-chain fatty acids synthesis and cell proliferation, but reduced tumor metastasis [12]. However, THRSP was found to be down-regulated in HCC tissues, and the decreased expression of THRSP was associated with worse prognosis in our study. To understand the roles of THRSP in HCC progression, we further analyzed the biological function and clinical implications of THRSP via integrated bioinformatics analysis. Moreover, we performed qPCR and immunohistochemical experiments to explore the mRNA and protein expression of THRSP in HCC tissues and cells. We also did Western blotting, CCK-8, Transwell, wound scratch and flow cytometry assays to investigate the function and molecular mechanism of THRSP in HCC.

## RESULTS

### Identification of crucial modules by WGCNA

The “WGCNA” package was used to group genes into modules by the average linkage hierarchical clustering. In this study, the soft powers  $\beta = 3$  and 5 were selected as the soft-thresholding to ensure scale-free networks (Figure 1A, 1B), and 11 modules in the TCGA-LIHC (Figure 1C) and 8 modules in the GEO

datasets (Figure 1D) were generated. The heatmaps (Figure 1E, 1F) of module-trait relationships were plotted to identify modules most significantly correlated with clinical features (normal and tumor). We found the brown modules in the TCGA-LIHC (containing 2057 co-expression genes) and GEO datasets (containing 2145 co-expression genes) had the highest association with tumor tissues (brown module in TCGA-LIHC:  $r = 0.69$ ,  $p = 7e-60$ ; brown module in GEO datasets:  $r = 0.86$ ,  $p = 4e-56$ ), which were selected as modules of interest for the subsequent analysis.

### Identification of DEGs and differentially co-expressed genes

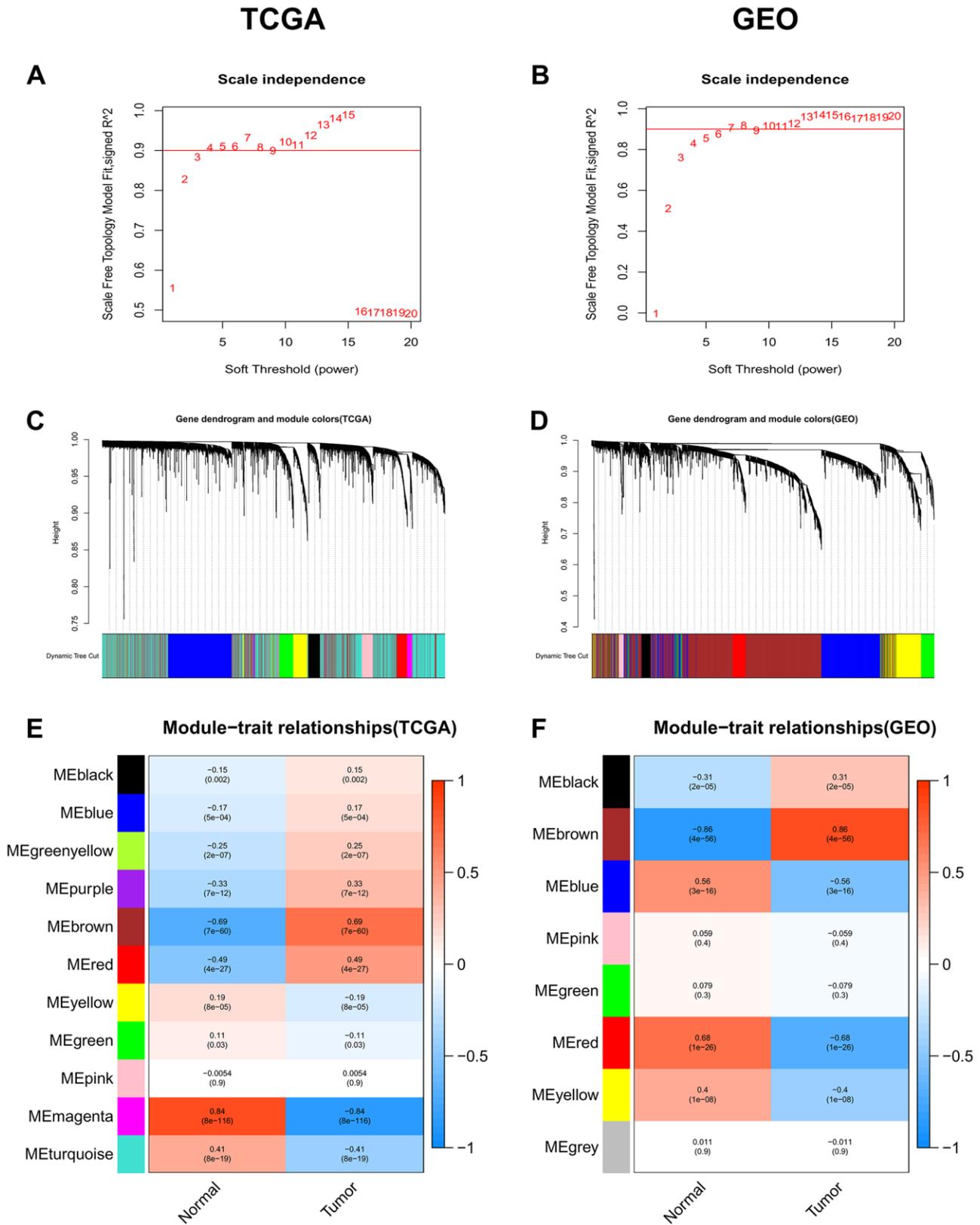
After normalization of the microarrays, 2705 differentially expressed genes (DEGs) between the HCC and normal tissues from the TCGA dataset (TCGA\_diff) and 567 DEGs from the GEO datasets (GEO\_diff) were screened by the “limma” package in R. Then, a Venn diagram was performed to examine the intersection among the DEGs and co-expressed genes of key modules. As shown in Figure 2A, 60 differentially co-expressed genes were finally obtained.

### PPI network and hub genes

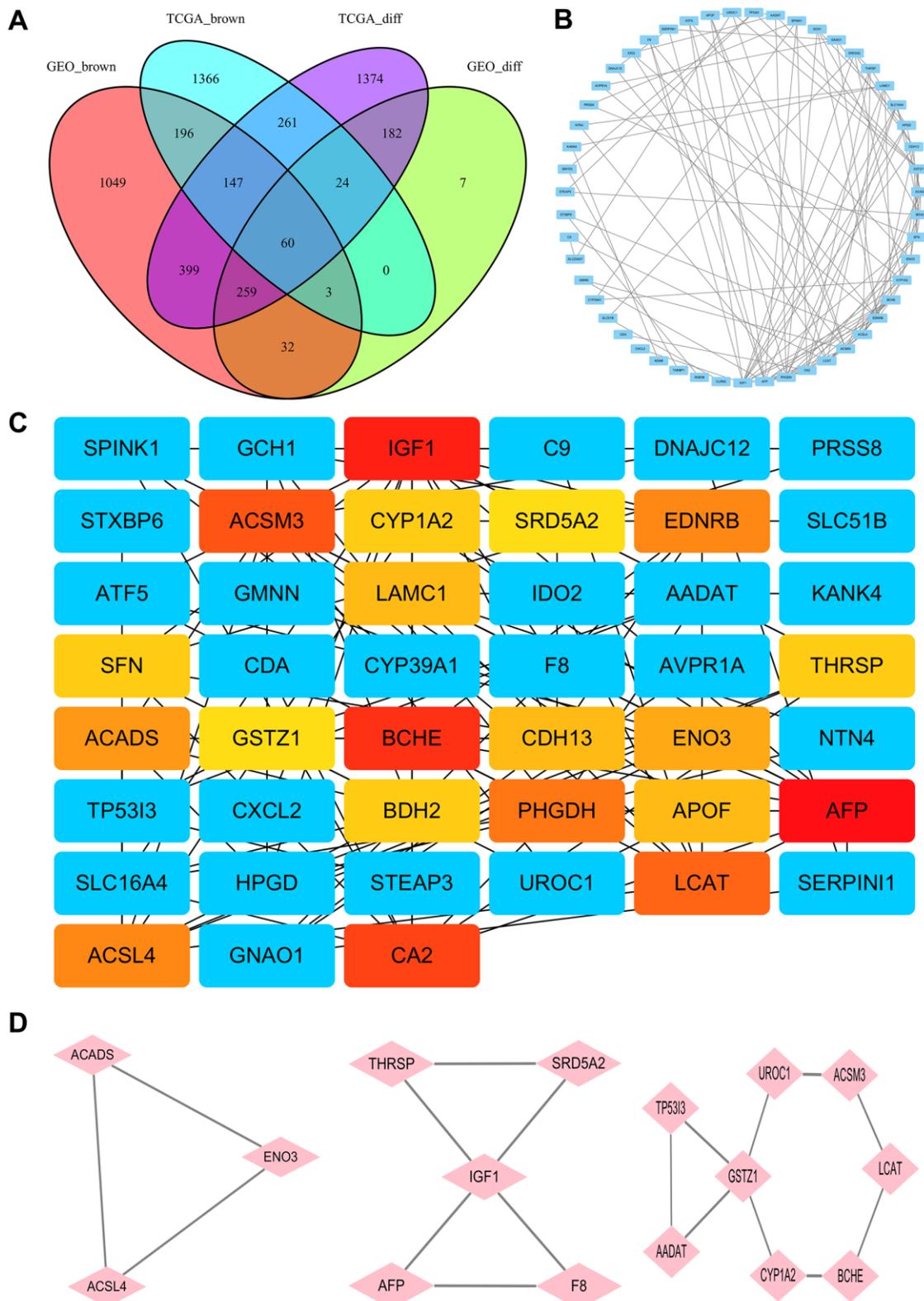
The PPI network of the 60 differentially co-expressed genes was constructed in the STRING database (Figure 2B). As showed in Figure 2C, the core genes were extracted from the PPI network by the MCC algorithms via the CytoHubba plug-in. Meanwhile, the significant modules of the 60 differential co-expression genes were established using the MCODE plug-in in Cytoscape (Figure 2D). Combining the above two algorithms, AFP, IGF1, BCHE, ACSM3, LCAT, ACSL4, ACADS, ENO3, CYP1A2, THRSP, GSTZ1, and SRD5A2 were finally selected as hub genes. Among them, AFP and ACSL4 were up-regulated and the other 10 hub genes were down-regulated in HCC tissues (Figure 3). To evaluate the prognostic values of the hub genes in HCC patients, a 5-year survival analysis was performed by Kaplan-Meier plotter. As shown in Figure 4, higher expression of LCAT, ACSM3, IGF1, SRD5A2, THRSP, and ACADS was associated with a better 5-year overall survival ( $p < 0.05$ ).

### THRSP was down-regulated in HCC tissues and cell lines

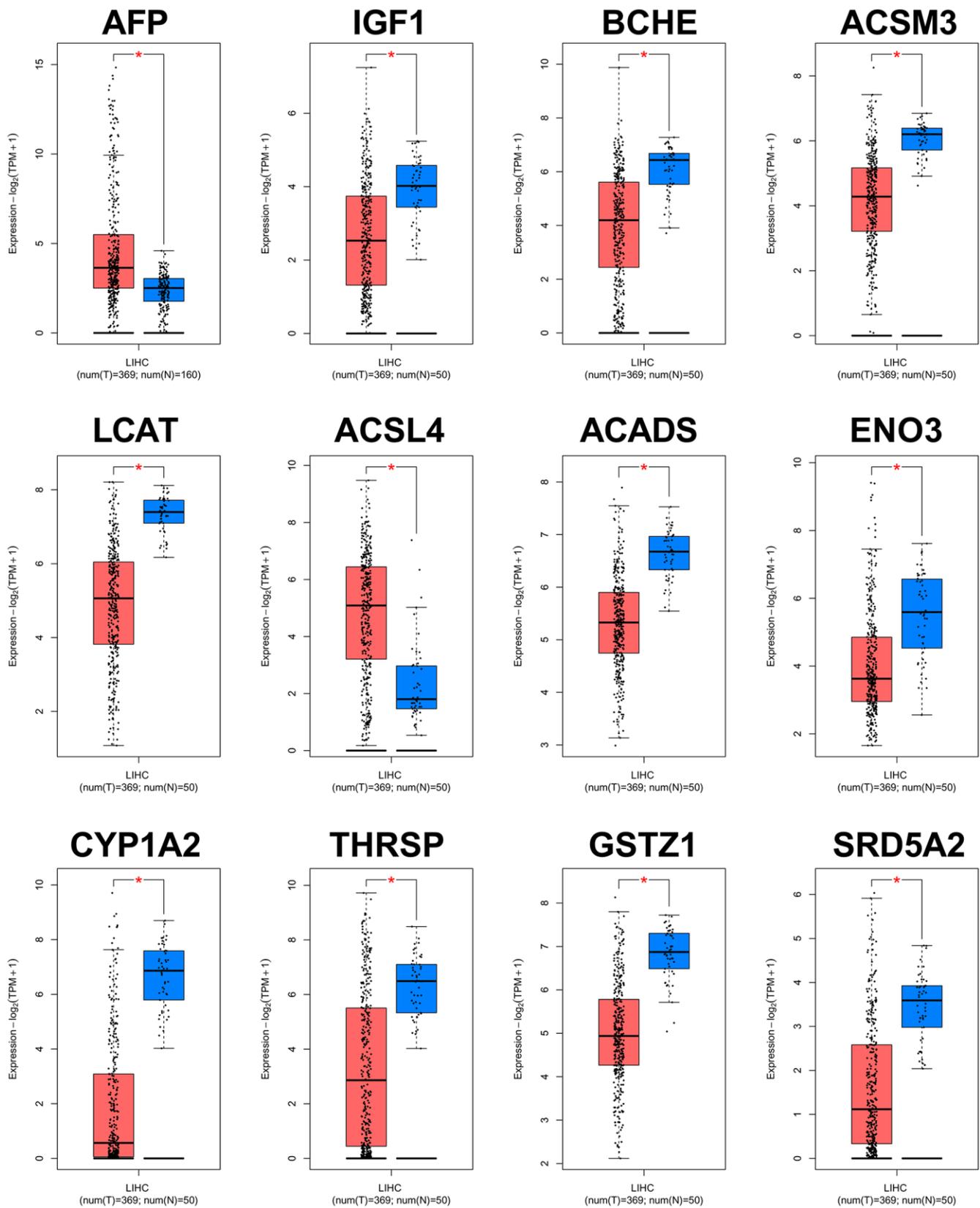
To verify the expression of THRSP in HCC tissues and cell lines, the quantitative real-time PCR (qRT-PCR) and immunohistochemistry (IHC) assays were performed in this study. The results of qRT-PCR verified that THRSP mRNA expression was remarkably lower in HCC tissues



**Figure 1. Identification of co-expression modules associated with the clinical traits in the TCGA-LIHC dataset and two GEO datasets. (A, B)** Analysis of the scale-free fit index for various soft-thresholding powers ( $\beta$ ). **(C, D)** Dendrogram of all genes in the TCGA-LIHC dataset or GEO datasets clustered based on the 1-TOM matrix. **(E, F)** Correlation between modules and clinical traits (normal and tumor). Each cell contains the corresponding correlation coefficient (the upper number) and the P-value (the lower number).



**Figure 2. Identification of hub genes.** (A) The Venn diagram for selection of the differential co-expression genes among DEG lists and co-expression modules. (B) PPI network of the intersection genes between DEG lists and co-expression modules. Each blue node represents a gene. Edges among nodes indicate interaction associations between genes. (C) Identification of the core genes from the PPI network by MCC algorithm. Darker colors refer to higher MCC scores. (D) The top three significant modules of the PPI network were evaluated in MOCDE. Pink nodes represent genes in corresponding modules.



**Figure 3. Validation of expression levels of the 12 hub genes in HCC and normal tissues using GEPIA.** \*P<0.01 is considered statistically significant. Tumor tissues are shown in red, and normal liver tissues are shown in blue.

as well as in HCCLM3 and Huh-7 cell lines compared with normal groups (Figure 5A, 5B). The results of IHC indicated that THRSP was down-regulated in most of HCC samples (8/10) (Figure 5C).

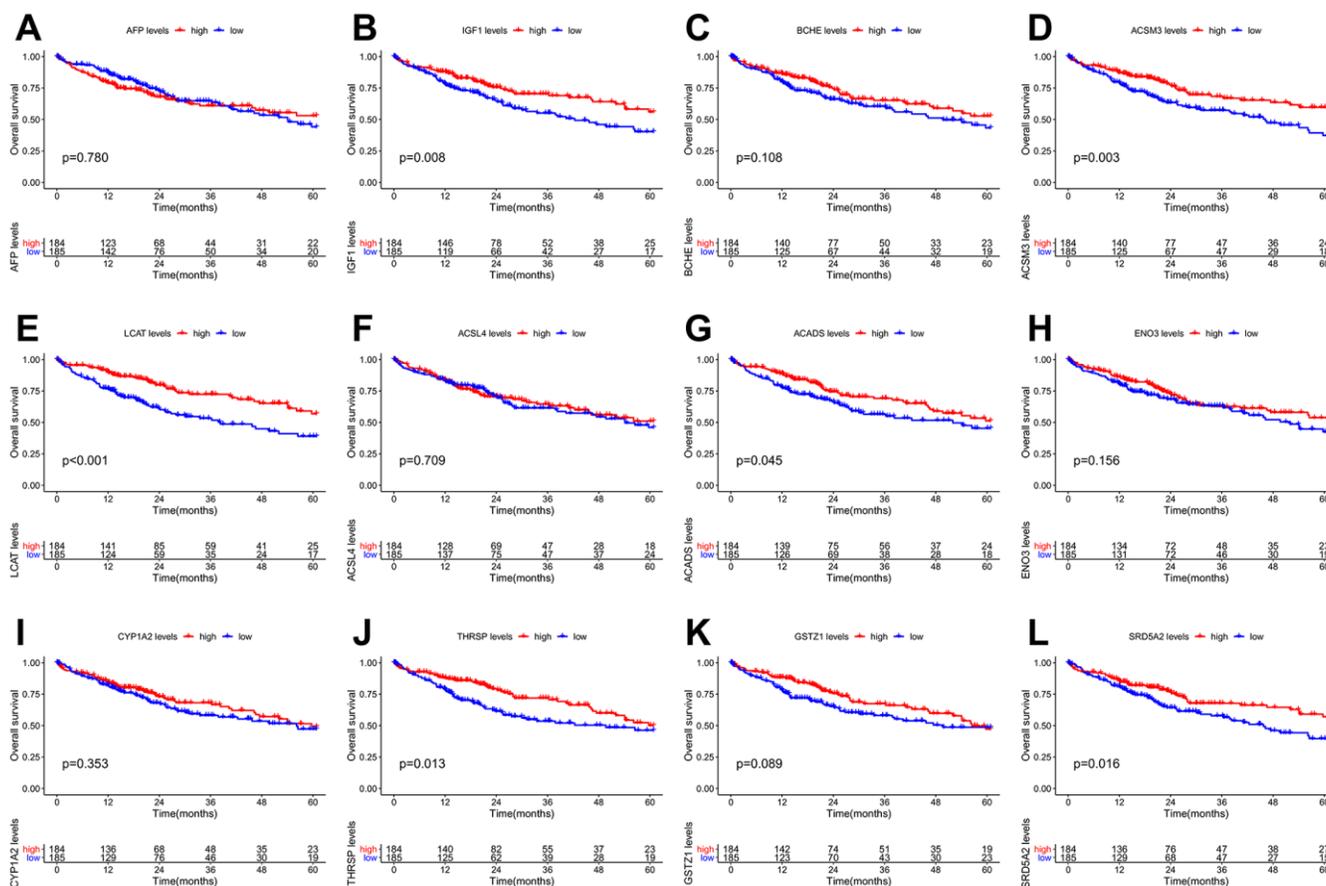
### The correlation between THRSP expression and clinical characteristics

Based on data from Illumina HumanMethylation 450 platform, we found a significant negative correlation ( $r = -0.81$ ,  $p < 0.001$ ) between THRSP mRNA expression and methylation (Supplementary Figure 1A). Four THRSP promoter CpG sites were shown in Supplementary Figure 1B. The Spearman correlation analysis demonstrated that CpG sites were negatively correlated with the expression of THRSP (Supplementary Figure 1C–1F). Wilcox test was used to compare the difference in THRSP expression or methylation in groups divided by age, family history of cancer, gender, grade, Ishak fibrosis score and race. The details were shown in Supplementary Figures 2, 3. Then, the HCC patients were dichotomized into two groups

(“low” or “high”) based on their THRSP expression levels or THRSP methylation levels using the median values as the cutoff point. The chi-square test was used to evaluate the correlation of THRSP expression or THRSP methylation with clinical characteristics. As listed in Table 1, THRSP mRNA expression or methylation was closely associated with the clinical indicators including age, family history of cancer, Ishak fibrosis score and gender et al.

### Relationship of THRSP with immune infiltration based on TIMER 2.0

As illustrated in the scatter plots (Supplementary Figure 4), THRSP expression was negatively correlated with infiltrating levels of B cells, CD4+ T cells, dendritic cells, and positively correlated with CD8+ T cells, but was not correlated with macrophages or neutrophils. In addition, we investigated the correlations of THRSP expression with the gene markers of various immune immune-infiltrating cells (including T cell, B cell, monocyte, neutrophil, dendritic cell, tumor-associated



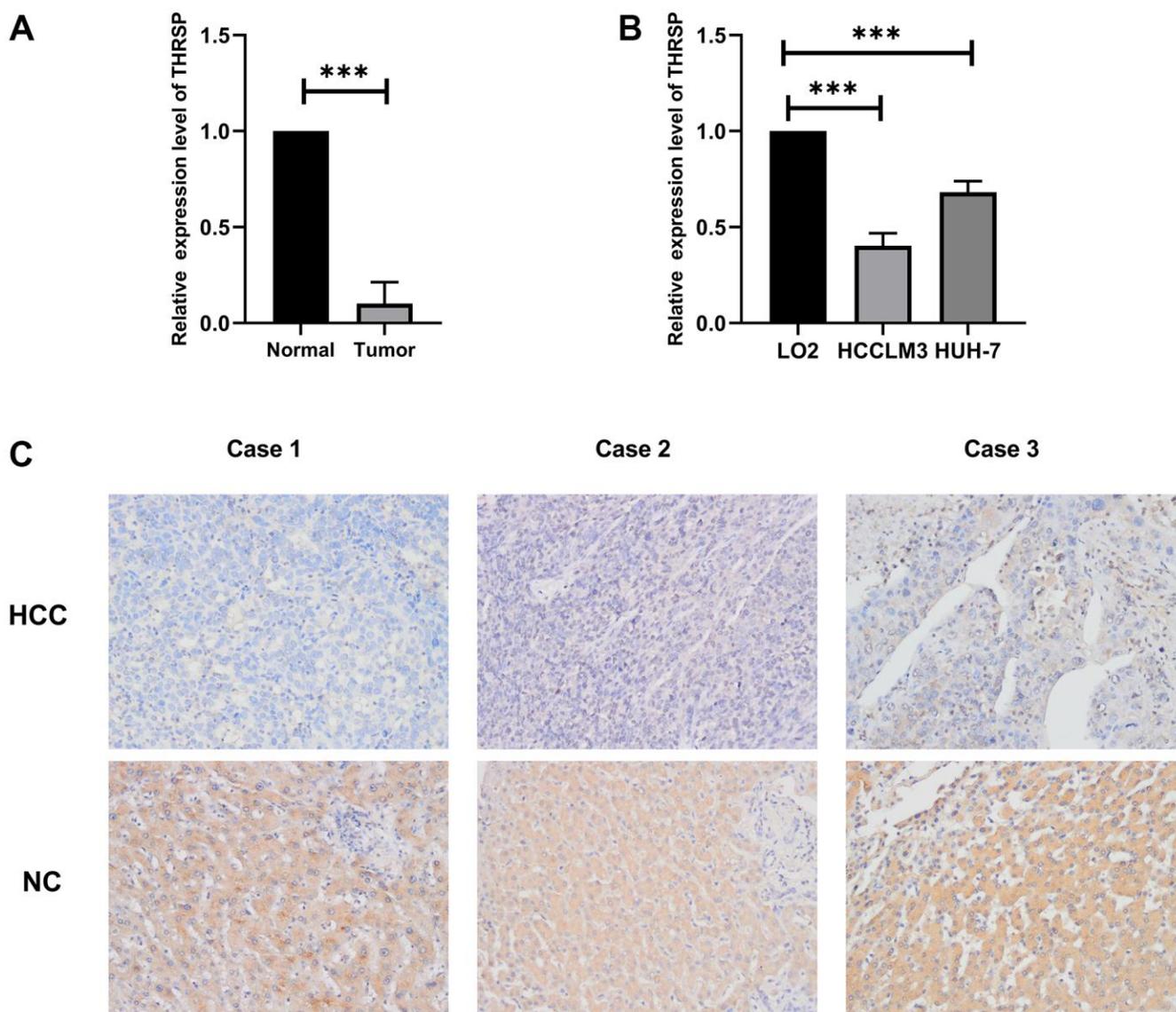
**Figure 4. Overall survival (OS) Kaplan-Meier curves of the 12 hub genes in HCC patients. (A) AFP,  $p = 0.780$ . (B) IGF1,  $p = 0.008$ . (C) BCHE,  $p = 0.108$ . (D) ACSM3,  $p = 0.003$ . (E) LCAT,  $p < 0.001$ . (F) ACSL4,  $p = 0.709$ . (G) ACADS,  $p = 0.045$ . (H) ENO3,  $p = 0.156$ . (I) CYP1A2,  $p = 0.353$ . (J) THRSP,  $p = 0.013$ . (K) GSTZ1,  $p = 0.089$ . (L) SRD5A2,  $p = 0.016$ .**

macrophage and different types of functional T cells) in HCC with the TIMER 2.0 database. As listed in Supplementary Table 1, THRSP expression was negatively associated with most of the gene markers.

### Silencing of THRSP promoted HCC progression

To explore the functional roles of THRSP in HCC, three candidate siRNAs of THRSP (THRSP-siRNAs) were transfected into HCCLM3 and Huh-7 cells. qRT-PCR was performed to evaluate the inhibition efficiency of the three THRSP-siRNAs. Three THRSP-siRNAs significantly inhibited THRSP

expression compared with the negative control (Figure 6A). The si-THRSP-3 was selected for the following experiments due to its robust silencing efficiency. The CCK-8 assays indicated that silencing of THRSP could promote proliferation of HCCLM3 and Huh7 cells (Figure 6B, 6C). The flow cytometry assays indicated that the si-THRSP group presented a significantly higher percentage of HCCLM3 cells in the S phase and the G2/M phase, but a lower percentage of cells in the G0/G1 phase compared with the control group (Figure 6D–6F). Transwell and wound scratch assays indicated that silencing of THRSP also promoted migration and invasion of HCC cells (Figure 7A–7C).



**Figure 5. The expression of THRSP at mRNA and protein levels.** (A, B) The mRNA expression of THRSP in HCC tissues and cells compared with control examined by RT-qPCR. (C) The protein expression of THRSP in HCC tissues and the adjacent normal tissues examined by immunohistochemical (200×).

**Table 1. Correlations between THRSP expression/methylation and clinical features.**

Clinical features	THRSP expression		P value	THRSP methylation		P value	
	Low (%)	High (%)		Low (%)	High (%)		
Age	<=65	128(68.45)	107(57.53)	0.0378	108(57.75)	127(68.28)	0.0457
	>65	59(31.55)	79(42.47)		79(42.25)	59(31.72)	
Cancer Status	Tumor-free	79(55.24)	83(58.04)	0.7204	85(59.44)	77(53.85)	0.4036
	With tumor	64(44.76)	60(41.96)		58(40.56)	66(46.15)	
Grade	G1	16(8.65)	39(21.2)	6.00E-04	39(21.08)	16(8.7)	3.00E-04
	G2	86(46.49)	92(50)		94(50.81)	84(45.65)	
	G3	74(40)	50(27.17)		49(26.49)	75(40.76)	
	G4	9(4.86)	3(1.63)		3(1.62)	9(4.89)	
Stage	Stage I	78(44.57)	95(54.29)	0.1955	94(53.71)	79(45.14)	0.2653
	Stage II	48(27.43)	39(22.29)		41(23.43)	46(26.29)	
	Stage III	45(25.71)	40(22.86)		39(22.29)	46(26.29)	
	Stage IV	4(2.29)	1(0.57)		1(0.57)	4(2.29)	
Family history of cancer	NO	114(70.81%)	96(59.63%)	0.0467	96(59.63)	114(70.81)	0.0467
	YES	47(29.19%)	65(40.37%)		65(40.37)	47(29.19)	
Living status	Alive	117(62.57)	126(67.74)	0.3472	128(68.45)	115(61.83)	0.2175
	Dead	70(37.43)	60(32.26)		59(31.55)	71(38.17)	
Ishak fibrosis score	0	28(25.93)	47(43.93)	0.0421	48(44.44)	27(25.23)	0.058
	1,2	18(16.67)	13(12.15)		14(12.96)	17(15.89)	
	3,4	17(15.74)	11(10.28)		12(11.11)	16(14.95)	
Gender	5,6	45(41.67%)	36(33.64%)		34(31.48%)	47(43.93%)	
	Female	71(37.97)	50(26.74)	0.0271	48(25.67)	73(39.04)	0.008
	Male	116(62.03)	137(73.26)		139(74.33)	114(60.96)	
Race	Asian	96(53.04)	64(35.36)	9.00E-04	64(35.36)	96(53.04)	9.00E-04
	Black	4(2.21)	13(7.18)		13(7.18)	4(2.21)	
	White	81(44.75)	104(57.46)		104(57.46)	81(44.75)	
expression	Low	-	-	-	26(13.9%)	161(86.1%)	< 0.001
expression	High	-	-	-	161(86.1%)	26(13.9%)	
methylation	Low	26(13.9%)	161(86.1%)	< 0.001	-	-	-
methylation	High	161(86.1%)	26(13.9%)		-	-	

**Silencing of THRSP inhibited apoptosis of HCC cells**

In this study, cells were strained with Annexin V/PI and subjected to flow cytometry to determine the apoptotic cells. The results indicated that silencing of THRSP could inhibit HCC cell apoptosis (Figure 8A, 8B). qRT-PCR and Western blotting assays were performed to detect the cell apoptosis-related molecules, including bax, bcl-2 and caspase 3. The results indicated that silencing of THRSP significantly reduced the expression of bax and caspase 3, while enhanced the expression of bcl-2 (Figure 8C, 8D).

**THRSP regulated HCC cell progression by modulating MAPK/NF-κB signaling pathway**

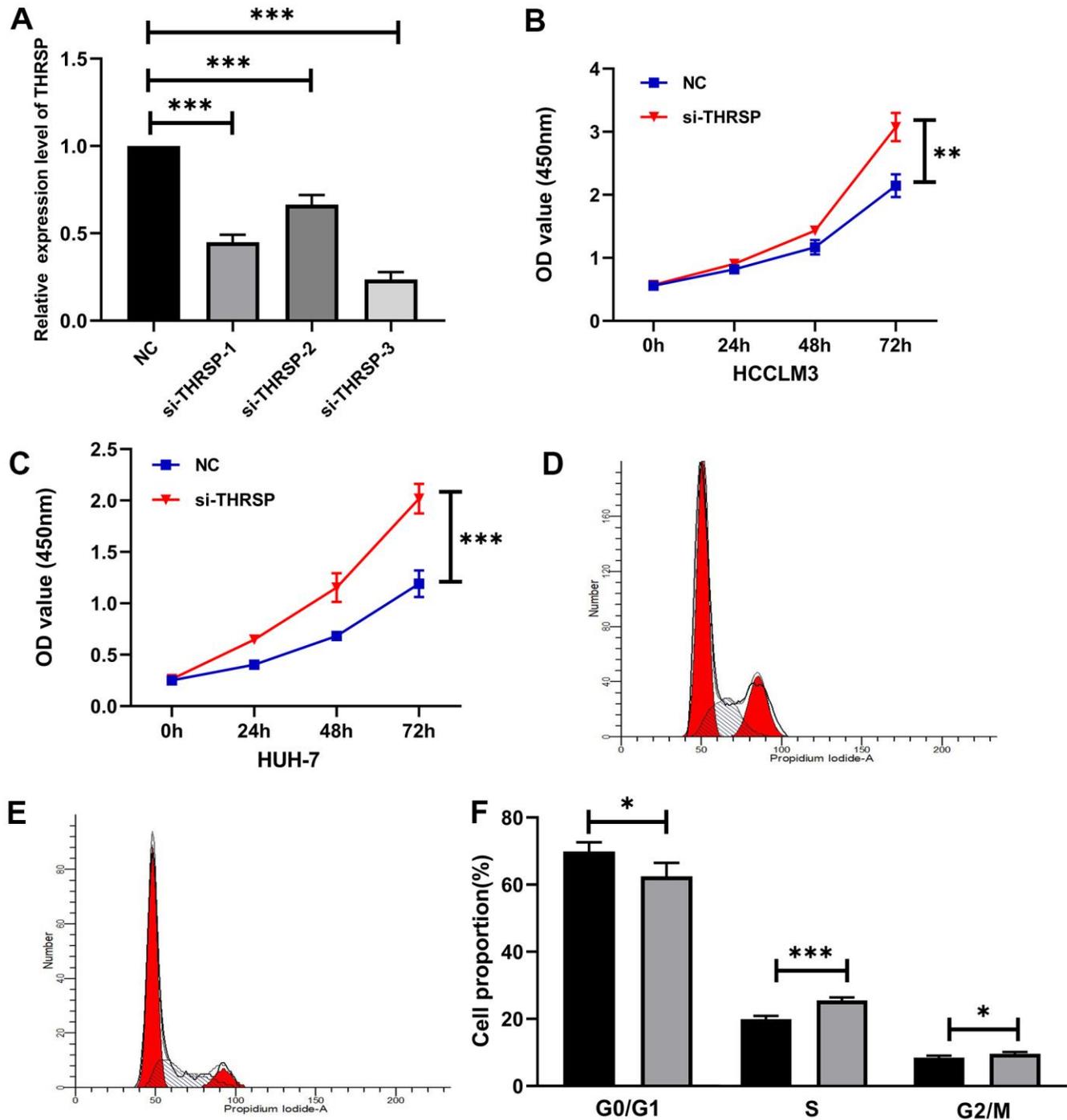
To further understand the molecular mechanism by which si-THRSP promoted the migration and invasion

of HCC cells, we explored the potential signaling pathways including the NF-κB and MAPK signaling pathways by Transwell and wound healing assays. As shown in Figure 8H–8J, compared with negative control, the migration and invasion of HCC cells transfected with si-THRSP or si-NC were inhibited after being treated with BAY-11-7082 (NF-κB inhibitor), AG-126 (ERK1/2 inhibitor), SB203580 (p38 MAPK inhibitor) and SP600125 (JNK inhibitor). And there was no difference between the si-THRSP group and the si-NC group when the cells were treated with BAY-11-7082, AG-126 and SB203580, indicating that the cell migration and invasion caused by the siRNA-induced silencing of THRSP might depend on the NF-κB, ERK1/2 and p38 MAPK signaling pathways. qRT-PCR and Western blotting assays were performed to assess the expression of MAPK/NF-κB pathway-related molecules (p65, p-p65,

p38, p-p38, erk1/2 and p-erk1/2). The results showed that the silencing of THRSP increased the phosphorylation of ERK1/2, p38 MAPK and p65, and enhanced the expression of p65, p38, and erk1/2 (Figure 8E, 8F). Each of the protein bands were showed in Figure 8G.

## DISCUSSION

HCC is highly malignant with a poor prognosis [13]. Despite advances in radiotherapy, chemotherapy, and surgical resection over the past decades, the 5-year survival rate of HCC remains frustrating [14]. There is



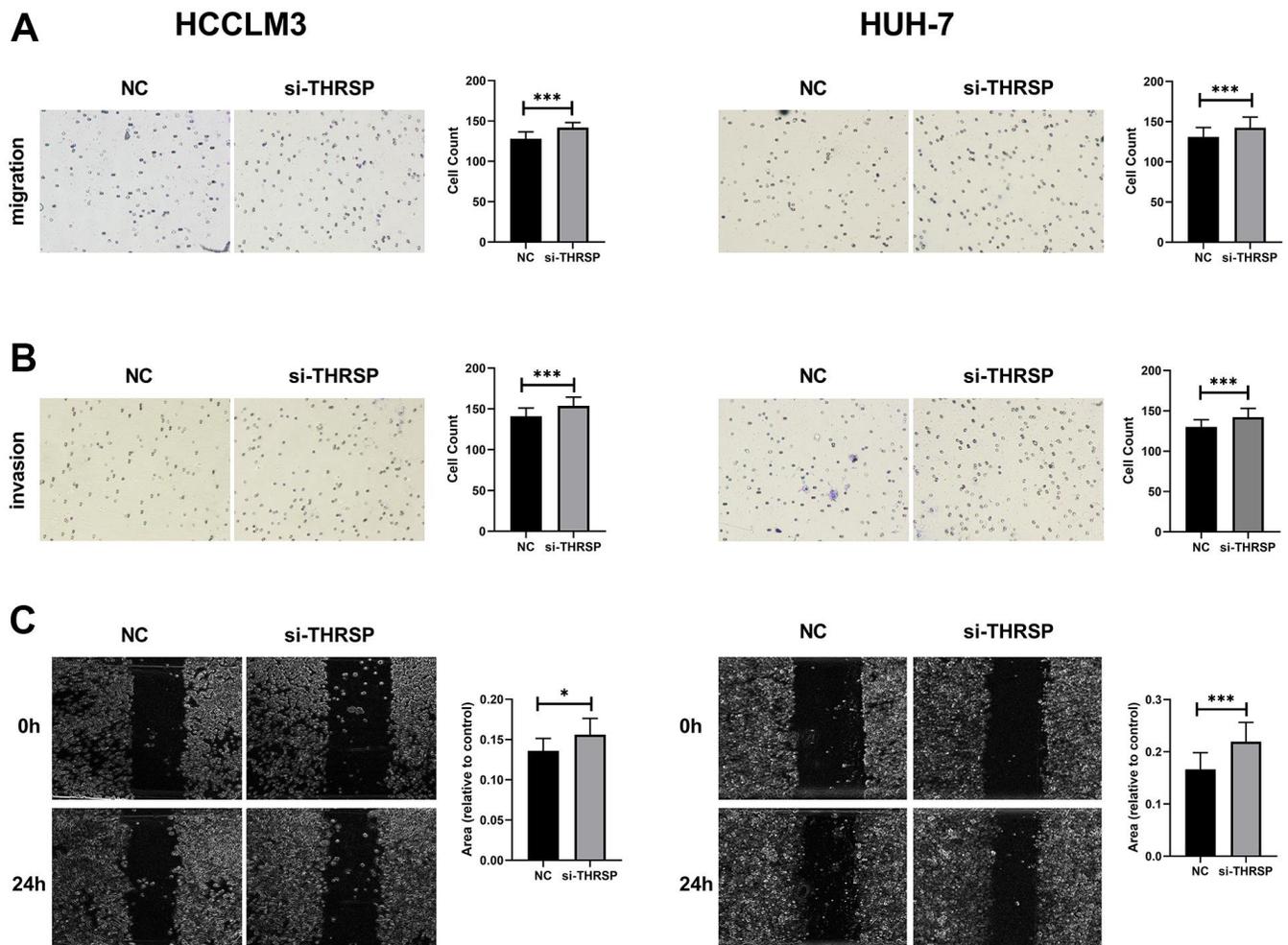
**Figure 6. The effect of THRSP expression on proliferation and cell cycle of HCC cells.** (A) The efficiency of the THRSP silencing determined by qRT-PCR. (B, C) The proliferation of HCCLM3 and Huh-7 cells examined by CCK-8. (D–F) The cell cycle assay detected by flow cytometry. NC group: black bars; si-THRSP group: grey bars. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

an urgent need to identify effective molecular targets to improve diagnostic and therapeutic approaches for HCC. Herein, integrated bioinformatic analysis was performed to identify effective molecular targets on two GEO datasets and the TCGA database. Six genes (namely LCAT, ACSM3, IGF1, SRD5A2, THRSP, and ACADS) were finally defined as survival-related hub genes. The roles of THRSP in HCC progression have rarely been reported. Here, we aimed to investigate biological functions and underlying mechanisms of THRSP in regulating HCC.

In this study, we systematically analyzed the mRNA expression, epigenetic modifications, immune significance, and clinical value of THRSP in HCC by bioinformatics analysis. The results indicated that the expression of THRSP was negatively correlated with its methylation and closely correlated with several clinical characteristics in HCC patients. The HCC patients with

higher expression of THRSP have better 5-year survival. In addition, THRSP expression was negatively correlated with most of the immune cells, and it might play an important role in the tumor microenvironment of HCC. The further function experiments implicated that silencing of THRSP could promote cell proliferation, migration, invasion and cell division, and inhibited apoptosis of HCC cells. NF- $\kappa$ B, ERK1/ERK2, and p38 MAPK signaling pathways were vital for THRSP-mediated HCC progression. Eventually, we concluded that THRSP may be a promising therapeutic target for HCC.

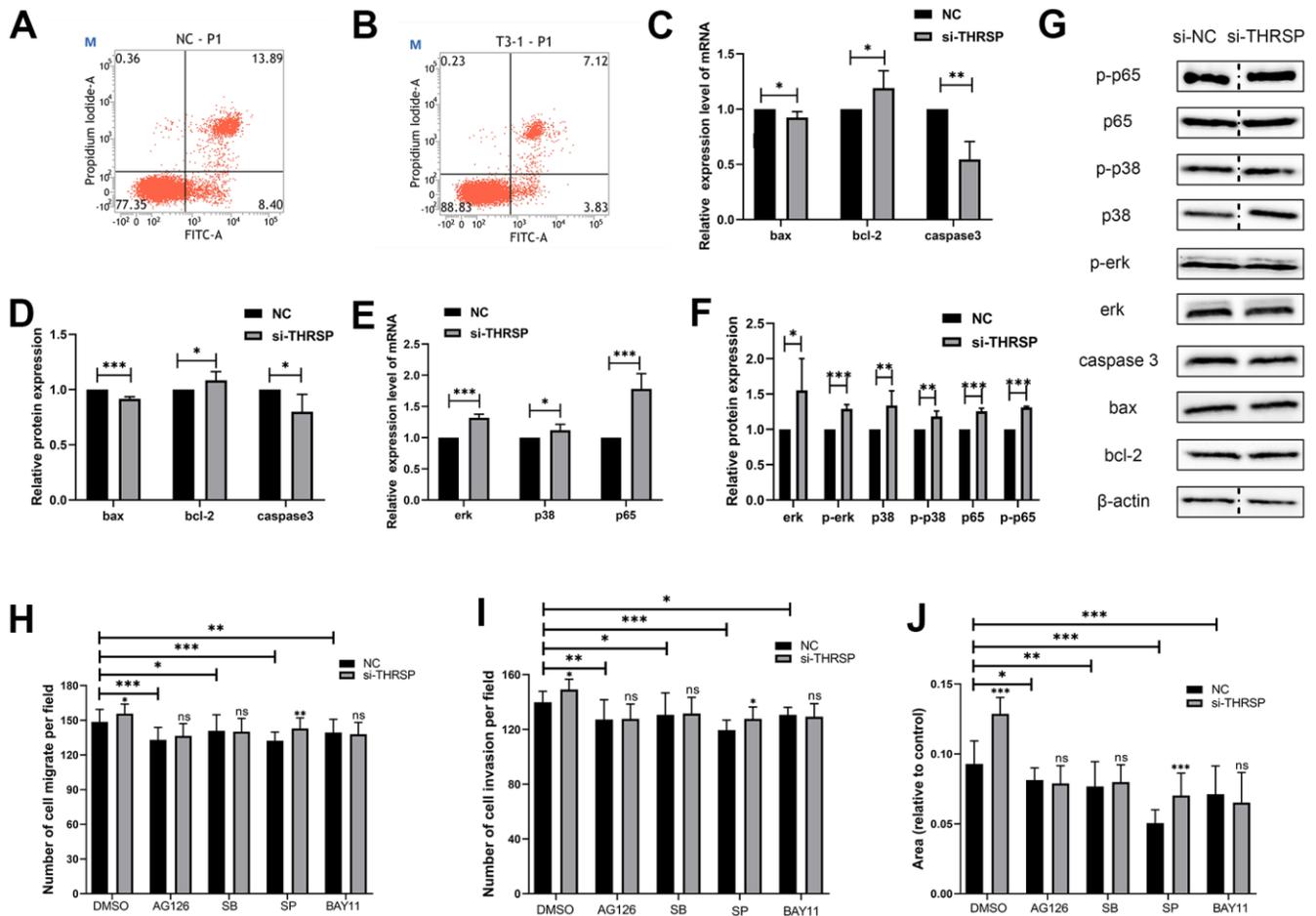
For the other hub genes related to the prognosis of HCC, LCAT (Lecithin-cholesterol acyltransferase) is a plasma enzyme involved in reverse cholesterol transport (RCT) and high-density lipoprotein (HDL) metabolism and has been reported to play an important role in many other cancers, such as breast cancer [15], Hodgkin



**Figure 7. Transwell and wound healing assays.** (A, B) The migration and invasion of HCCLM3 and Huh-7 cells detected by Transwell assays (magnification: 200 $\times$ ). (C) The migration ability of HCCLM3 and Huh-7 cells examined by wound healing assays (magnification: 40 $\times$ ).

lymphoma [16], and ovarian cancer [17]. The previous studies had reported the significantly low expression and high DNA methylation of LCAT in HCC patients [18–20]. Besides, LCAT plays a crucial role in the conversion of liver cirrhosis into HCC [21]. ACSM3, as one member of the acyl-CoA synthetase medium-chain family, was found to be frequently down-regulated in HCC patients exhibiting high AFP levels, high ALT levels, large tumors, and multiple nodules. On the contrary, higher ACSM3 expression was always associated with a better prognosis and may hinder metastasis of HCC by downregulating phosphorylation of WNK1 and AKT [22]. IGF-1 (growth factor-1) has been widely reported that its expression decreased sharply in patients with chronic liver disease such as steatosis, nonalcoholic steatohepatitis, chronic hepatitis C, cirrhosis, and HCC [23–29]. The reason may be that most of the circulating levels of IGF-1 were synthesized by the liver

[30, 31]. IGF1 synthesis decreases when hepatitis or liver necrosis occurs. In addition, a prospective cohort study demonstrated that IGF-1 can be an independent predictor of survival or recurrence in early HCC [32]. IGF1 was also demonstrated to play an important role in the cellular function aspects of hepatocarcinogenesis and could be a therapeutic target against HCC [33–35]. For instance, Sorafenib could inhibit macrophage-induced growth of hepatoma cells by disrupting IGF1 secretion [34]. SRD5A2, also known as steroid 5- $\alpha$ -reductase 2, encodes a microsomal protein. As a membrane-associated enzyme, it catalyzes the transformation of testosterone to dihydrotestosterone (DHT). SRD5A2 is highly expressed in androgen-sensitive tissues such as the prostate and the expression of SRD5A2 is associated with the progression of prostate cancer [36–39]. To date, some reports revealed that SRD5A2 polymorphism may be associated with liver cancer, and it might serve as a robust diagnosis or



**Figure 8. The effect of THRSP expression on cell apoptosis and channel regulation in HCC.** (A, B) The apoptosis of HCLM3 cells detected by flow cytometry. (C–F) The expression of apoptosis-related molecules (bax, bcl-2 and caspase 3) and MAPK/NF- $\kappa$ B pathway-related molecules (erk, p-erk, p38, p-p38, p65 and p-p65) examined by qRT-PCR or Western blotting assays. (G) Each of the protein bands. The dividing lines (dashed lines) indicated that the grouping of images were from different parts of the same gel. (H–J) Invasion and migration rates were analyzed when treated with AG-126, SB203580, SP600125 and BAY-11-7082 inhibitors.

prognosis marker for the diagnosis of HCC [40–42]. ACADS, namely acyl-CoA dehydrogenase short-chain, encodes a tetrameric mitochondrial flavoprotein and catalyzes the initial step of the mitochondrial fatty acid beta-oxidation pathway. It was identified as a potential biomarker in colon adenocarcinoma and bladder cancer [43, 44]. A previous study demonstrated that ACADS was significantly down-regulated in HCC tissues and was regulated by DNA methylation, which played a key role in promoting the proliferation and metastasis of HCC [45].

In summary, by a series of comprehensive bioinformatics analyses, our study screened six significant survival-related hub genes. Among them, we found a novel biomarker (THRSP) associated with HCC development. The experimental results showed that lower expression of THRSP can promote the progression of HCC cells. Therefore, THRSP has the potential to be a valuable therapeutic target for HCC.

## MATERIALS AND METHODS

### Datasets preprocessing

Two gene expression profiles GSE84005 and GSE121248 were obtained from the GEO database. The GSE84005 dataset, including 38 tumor tissues and paired 38 normal tissues from HCC patients, was based on the GPL5175 platform (Affymetrix Human Exon 1.0 ST Array). The GSE121248 dataset, including 70 HCC tissue samples and 37 adjacent non-tumor tissue samples, was based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array). Two datasets were merged by Perl 5.3 (available online: <http://www.perl.org/>) to increase the sample size. The merged dataset was batch-normalized by “limma” and “sva” packages in R 3.6.3 (<https://www.r-project.org/>) to remove batch effects.

Besides, the RNA-sequencing (RNA-seq) data of 373 HCC and 49 normal samples and the corresponding clinical information was downloaded from the TCGA database. As recommended by the package “edgeR” in R, genes with low read counts (count per million (cpm)  $\leq 1$ ) were omitted. Gene expression was calculated and normalized to RPKM (reads per kilobase per million) values using function “rpkm” in the “edgeR” package. Moreover, DNA promoter methylation data (Methylation 450k, including 430 samples) was downloaded from the TCGA database via the UCSC Xena browser (<https://xenabrowser.net/>).

### WGCNA analysis

The “WGCNA” package in R was applied to construct co-expression networks and to explore the key

modules of highly relevant genes by cluster analysis for relating modules to sample traits. In this study, the gene expression profiles of TCGA-LIHC and the merged dataset of GSE84005 and GSE121248 were respectively used to construct WGCNA. Briefly: an adjacency matrix was created by Pearson’s correlations between each of the gene pairs. Next, the adjacency matrix was utilized to erect a scale-free co-expression network based on the soft threshold power  $\beta$  which was selected using the pickSoftThreshold function [46]. Subsequently, the adjacency matrix was converted into a topological overlap matrix (TOM) as well as the corresponding dissimilarity (1-TOM). Then, module identification was conducted using the dynamic tree Cut approach by average linkage hierarchical clustering based on the TOM-based dissimilarity measure with the parameters of minModuleSize of 50, deepSplit value of 2, and mergeCutHeight of 0.25 for the genes dendrogram. Afterward, the correlation between module eigengenes (MEs) and the clinical trait information was calculated by the module-trait relationship analysis of WGCNA to identify the clinically significant modules in a co-expression network. Finally, modules with a high correlation coefficient were selected for further analysis.

### Identification of differentially co-expressed genes

The “limma” package in R was used to filter the DEGs between the HCC and normal samples in the TCGA-LIHC and the merged dataset of GSE84005 and GSE121248. The adjusted P-value (adj. P)  $< 0.05$  and  $|\log_2\text{foldchange (FC)}| > 1$  was set as the criteria of DEGs. Then, the overlapping genes between DEG lists and co-expression genes from significant modules were screened out using the “VennDiagram” package in R.

### Identification of hub genes

The PPI network was constructed in the STRING (<http://string-db.org>) and visualized by the Cytoscape software (Cytoscape\_v3.8.0, <https://cytoscape.org/>) [47]. The maximal clique centrality (MCC) analysis was performed to extract the candidate hub genes with the top20 MCC values in the PPI network using the cytoHubba plug-in [48]. Meanwhile, the Molecular Complex Detection (MCODE) [49] plug-in of Cytoscape was implemented to find significant PPI modules with degree cut-off  $\geq 2$ , node score cut-off  $\geq 0.2$ , K-core  $\geq 2$ , and max. depth =100. Finally, the overlapping genes obtained from the MCC analysis and MCODE analysis were regarded as hub genes.

## The gene expression and prognostic analysis of the hub genes

The differential expression analysis of the hub genes between HCC and normal tissues was performed based on the GEPIA2 database. The 5-year overall survival (OS) analysis of these hub genes was performed using the “survival” and “survminer” packages in R [50].

## The correlation between gene expression and clinical features

According to the median value of gene expression or methylation, the HCC patients were divided into the low- or high- group. The chi-square test was utilized to investigate the correlation of gene expression as well as methylation with clinical characteristics. The correlation between DNA methylation and gene expression in HCC samples from the TCGA database was examined using the Spearman correlation coefficient and visualized by “ggplot2” and “ggpubr” packages in R.

## Immune infiltrate analysis

The online tool TIMER 2.0 (<https://cistrome.shinyapps.io/timer/>) was used for immune infiltrate analysis [51]. The abundance of six types of immune cells (including CD8+ T cells, CD4+ T cells, B cells, neutrophils, macrophages, and dendritic cells) were computed by TIMER algorithm. In addition, we exploited the correlation between gene expression and the gene markers of different kinds of immune cells. The immune gene markers of interest used in this study were referred to the previous studies [52–56].

## Cell culture

The human normal liver cell line (LO2) and the human hepatoma cell lines (HCCLM3 and HUH-7) were purchased from Procell Life Science and Technology Co., Ltd. (Wuhan, China). The cells were cultured in DMEM (Servicebio Technology Co., Ltd, Wuhan, China) supplemented with 10% Fetal Bovine Serum (FBS, G-CLONE, Beijing, China) and maintained in a incubator with 5% CO<sub>2</sub>-humidified atmosphere at 37° C.

## Total RNA extraction and qPCR

The quantitative real-time polymerase chain reaction (qRT-PCR) was employed to detect the expressions of THRSP. GAPDH was served as a reference gene. Total RNA was extracted from tissues and cells by the TRIzol reagent (G-CLONE, Beijing, China). The expression of

THRSP was determined by the SweScript RT I First Strand cDNA Synthesis Kit with gDNA Remover and the SYBR Green qPCR Master Mix (High ROX) (Servicebio Technology Co., Wuhan, China) according to the manufacturer’s protocol. The qRT-PCR was performed on the StepOne Plus Real-Time PCR Systems. The primers used in this study were as follows: THRSP forward: 5’-CAGGTGCTAACCAAGCGTTAC-3’, THRSP reverse: 5’-CAGAAGGCTGGG GATCATCA-3’; GAPDH forward: 5’-GGACCTGACC TGCCGTCTAG-3’, GAPDH reverse: 5’-GTAGCCCA GGATGCCCTTGA-3’.

## IHC analysis

10 pairs of HCC tissues and paracancerous tissues were fixed in formalin, dehydrated, and embedded in paraffin. The paraffin sections were deparaffinized for antigen retrieval and treated with 3% hydrogen peroxide for blocking peroxidase activity, with 3% bovine serum albumin (BSA) for serum sealing. Afterward, the paraffin sections were incubated with primary THRSP antibody (Guangzhou Alexan Biotech Co., Ltd., China) overnight at 4° C, and then with HRP-conjugated secondary antibody for 50 min at room temperature. 3,3’-Diaminobenzidine (DAB) liquid substrate was used for staining and the hematoxylin solution was used for nucleus counterstaining. Finally, after dehydration and mounting, a microscope was used to acquire images of the staining of tissues.

## Transfection of small interfering RNA

Three short interfering RNAs (siRNAs) of THRSP (T1: 5’-ACACCTACTTCACCATGCT-3’; T2: 5’-CCAGGA AATGACGGGACAA-3’; T3: 5’-CATGCACCTCACCG AGAAA-3’) and negative control siRNA (si-NC) were purchased from RiboBio Co., Ltd. (Guangzhou, China). HCCLM3 and Huh-7 cells (2×10<sup>5</sup> per well) were inoculated on 24-well plates for 24h and then transfected with 25pmol of the RNA duplex according to the manufacturer’s protocol of Lipofectamine 2000 (Invitrogen, Grand Island, NY, USA). After 24h, the transfected cells were harvested for the following experiments.

## CCK-8 assay

The transfected HCCLM3 and Huh-7 cells were seeded into 96-well plates (3×10<sup>3</sup> cells/well) and incubated for 24h, 48h and 72h. At each time point, 10μl of CCK-8 reagent (Guangzhou Alexan Biotech Co., Ltd., China) was dripped into each well and the cells were cultured for an extra 4h. A micro-plate reader was used to detect the absorbance at 450nm to evaluate cell proliferation.

## Transwell and scratch wound healing assays

Transwell assays were used to determine the migration and invasion ability of the HCC cells. The transfected HCCLM3 cells and Huh-7 cells in 200 $\mu$ l serum-free DMEM (5 $\times$ 10<sup>4</sup> cells/well) were plated into the upper chamber and 600 $\mu$ l of complete medium was added in the lower chamber. For cell invasion assay, the Transwell chamber was coated with Matrigel (Beijing Solarbio Science and Technology Co., Ltd., China). After incubating at 37° C for 24h, the remaining cells in the upper chamber were removed by a cotton swab. The cells that have invaded to the lower surface of the filter were fixed with 4% (v/v) neutral formaldehyde solution (Servicebio, Wuhan, China) for 30min and stained with 0.1% Crystal violet for 30min. Finally, the cells in ten random microscope fields of each filter were counted.

The transfected HCCLM3 and Huh-7 cells (5 $\times$ 10<sup>5</sup> cells per well) were seeded into 24-well plates. When the cell confluence reaches 90%, a scratch wound was created using a sterilized pipette tip (200 $\mu$ l) on confluent cells. The images of wounds were acquired with a phase-contrast light (40 $\times$ ) at 0h and 24h. The heal area of each scratch wound was determined by ImageJ.

To further understand the molecular mechanism, the potential signaling pathways including the NF- $\kappa$ B and MAPK signaling pathways were detected by Transwell and wound healing assays. Briefly: post of transfection for 48h, the cells were harvested and pre-treated with BAY-11-7082 (5 $\mu$ M, NF- $\kappa$ B inhibitor), AG-126 (10 $\mu$ M, ERK1/2 inhibitor), SB203580 (10 $\mu$ M, p38 MAPK inhibitor) and SP600125 (50 $\mu$ M, JNK inhibitor) for 2h in serum-free DMEM medium. Then, the cells were cultured in the DMEM medium containing inhibitors for 24 hours after being seeded into Transwell chambers or culture wells. The cells treated with dimethyl sulfoxide (DMSO) were used as the negative control.

## Flow cytometry

For the cell cycle assays, after 48h of transfection, the HCCLM3 cells were trypsinized, and washed with cold phosphate-buffered saline (PBS) and then fixed in 70% cold ethanol at 4° C for 24 h. After centrifuging and washing, the cells were stained with 500 $\mu$ l PI buffer (50 $\mu$ g/mL, containing RNase, Beijing Leagene Biotechnology Co., Ltd., China) at 37° C in the dark for 30 min. The cell cycle distribution was determined by the Flow cytometer after PI staining. The apoptosis analysis was performed following the instruction of Annexin-V Apoptosis Detection kit (Jiangsu KeyGEN BioTECH Co., Ltd. China). Briefly, 72h post-transfection, the trypsinized cells were washed with

cold PBS twice and re-suspended in binding buffer, and then stained with Annexin V-FITC and PI at room temperature for 5-15 min in the dark. The apoptotic cells were analyzed by flow cytometry within an hour.

## Western blotting

The total protein was extracted with the nucleoprotein and cytoplasmic protein extraction kit (Jiangsu Keygen Biotech Co., Ltd., China) and quantified by the BCA protein assay kit (Beijing Bomaide Gene Technology Co., Ltd., China). After that, the protein solution was subjected to electrophoresis, detached via SDS-PAGE (12% gels), and transferred to PVDF membranes. The PVDF membranes were blocked with 5% skim milk powder in TBST (Tris buffered saline with 0.5% Tween 20) for 2 h, and then incubated with primary antibodies overnight at 4° C. After washing with TBST for 5 times, the PVDF membranes were incubated with secondary antibody at room temperature for an hour. The protein bands were exposed by enhanced chemiluminescent (ECL) substrate kit (Labgic Technology Co., Ltd. Hefei, China) and analyzed by ImageJ software. The Anti-Bcl-2 antibody, Anti-Caspase-3 antibody, Anti-Bax antibody, Anti-NF- $\kappa$ B p65 antibody and Anti-NF- $\kappa$ B p65 (phospho S536) antibody were purchased from Abcam. The p38 MAPK, Phospho-p38 MAPK (Thr180/Tyr182), P44/42 MAPK (Erk1/2), and Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) were purchased from Cell Signaling Technology.  $\beta$ -actin was severed as the internal control and anti- $\beta$ -actin was purchased from Labgic Technology Co., Ltd.

## Ethics approval

The manuscript has been approved by the ethics committee of the affiliated hospital of Southwest Medical University (Ethical ID: KY2021142).

## Abbreviations

HCC: hepatocellular carcinoma; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; DEGs: differentially expressed genes; WGCNA: Weighted Gene Co-expression Network Analysis; PPI: protein-protein interaction; THRSP: thyroid hormone-responsive; LIHC: liver hepatocellular carcinoma; STRING: search tool for the retrieval of interacting genes; MCC: maximal clique centrality; MCODE: molecular complex detection; OS: overall survival; GEPIA: Gene Expression Profiling Analysis; FDR: false discovery rate; qRT-PCR: quantitative real-time polymerase chain reaction; TIMER: Tumor Immune Estimation Resource; siRNA: small interfering RNA; DMSO: Dimethyl sulfoxide; NC group: normal control group; PBS: phosphate-buffered saline; SDS-PAGE:

sodium dodecyl sulfate polyacrylamide gel electrophoresis; PVDF: polyvinylidene fluoride; TBST: tris-buffer saline tween; ECL: enhanced chemiluminescence.

## AUTHOR CONTRIBUTIONS

Yuxi Ding, Changfeng Sun, and Cunliang Deng conceptualized the project and developed the methodology. Xiaoling Liu, Changfeng Sun, Yuxi Ding, Yue Yuan and Decheng Li performed the experiments. Yuxi Ding and Xiaoling Liu analyzed the data, prepared figures and/or tables. Yuxi Ding wrote the manuscript. Changfeng Sun, Xiaoling Liu, Yunjian Sheng, and Suvash Chandra Ojha revised the manuscript. All authors have read and approved the final version of the manuscript for publication.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This study was financed by grants from the scientific research project of the Sichuan Provincial Health and Family Planning Commission (18PJ340); the Luzhou Municipal People's Government-Southwest Medical University Science and Technology Strategic Cooperation Applied Basic Research Project (2018LZXNYD-ZK29); Scientific Research Project for Young Researchers of Southwest Medical University (2017-ZRQN-103); the Affiliated Hospital of Southwest Medical University Research Grant (16237 and 16251). We would like to thank Tiejun Zhou and Jun Liu of the Pathology Department of Southwest Medical University for the guidance on pathology.

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:33538338
2. Dutta R, Mahato RI. Recent advances in hepatocellular carcinoma therapy. *Pharmacol Ther.* 2017; 173:106–17. <https://doi.org/10.1016/j.pharmthera.2017.02.010> PMID:28174094
3. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer

- Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; 45:1113–20. <https://doi.org/10.1038/ng.2764> PMID:24071849
4. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002; 30:207–10. <https://doi.org/10.1093/nar/30.1.207> PMID:11752295
  5. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559. <https://doi.org/10.1186/1471-2105-9-559> PMID:19114008
  6. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017; 45:D362–8. <https://doi.org/10.1093/nar/gkw937> PMID:27924014
  7. Seelig S, Jump DB, Towle HC, Liaw C, Mariash CN, Schwartz HL, Oppenheimer JH. Paradoxical effects of cycloheximide on the ultra-rapid induction of two hepatic mRNA sequences by triiodothyronine (T3). *Endocrinology.* 1982; 110:671–3. <https://doi.org/10.1210/endo-110-2-671> PMID:6276146
  8. Cunningham BA, Moncur JT, Huntington JT, Kinlaw WB. "Spot 14" protein: a metabolic integrator in normal and neoplastic cells. *Thyroid.* 1998; 8:815–25. <https://doi.org/10.1089/thy.1998.8.815> PMID:9777755
  9. Chella Krishnan K, Kurt Z, Barrere-Cain R, Sabir S, Das A, Floyd R, Vergnes L, Zhao Y, Che N, Charugundla S, Qi H, Zhou Z, Meng Y, et al. Integration of Multi-omics Data from Mouse Diversity Panel Highlights Mitochondrial Dysfunction in Non-alcoholic Fatty Liver Disease. *Cell Syst.* 2018; 6:103–115.e7. <https://doi.org/10.1016/j.cels.2017.12.006> PMID:29361464
  10. Jump DB, Narayan P, Towle H, Oppenheimer JH. Rapid effects of triiodothyronine on hepatic gene expression. Hybridization analysis of tissue-specific triiodothyronine regulation of mRNAs14. *J Biol Chem.* 1984; 259:2789–97. [https://doi.org/10.1016/S0021-9258\(17\)43215-7](https://doi.org/10.1016/S0021-9258(17)43215-7) PMID:6199351
  11. Kinlaw WB, Quinn JL, Wells WA, Roser-Jones C, Moncur JT. Spot 14: A marker of aggressive breast cancer and a potential therapeutic target. *Endocrinology.* 2006; 147:4048–55. <https://doi.org/10.1210/en.2006-0463> PMID:16809441

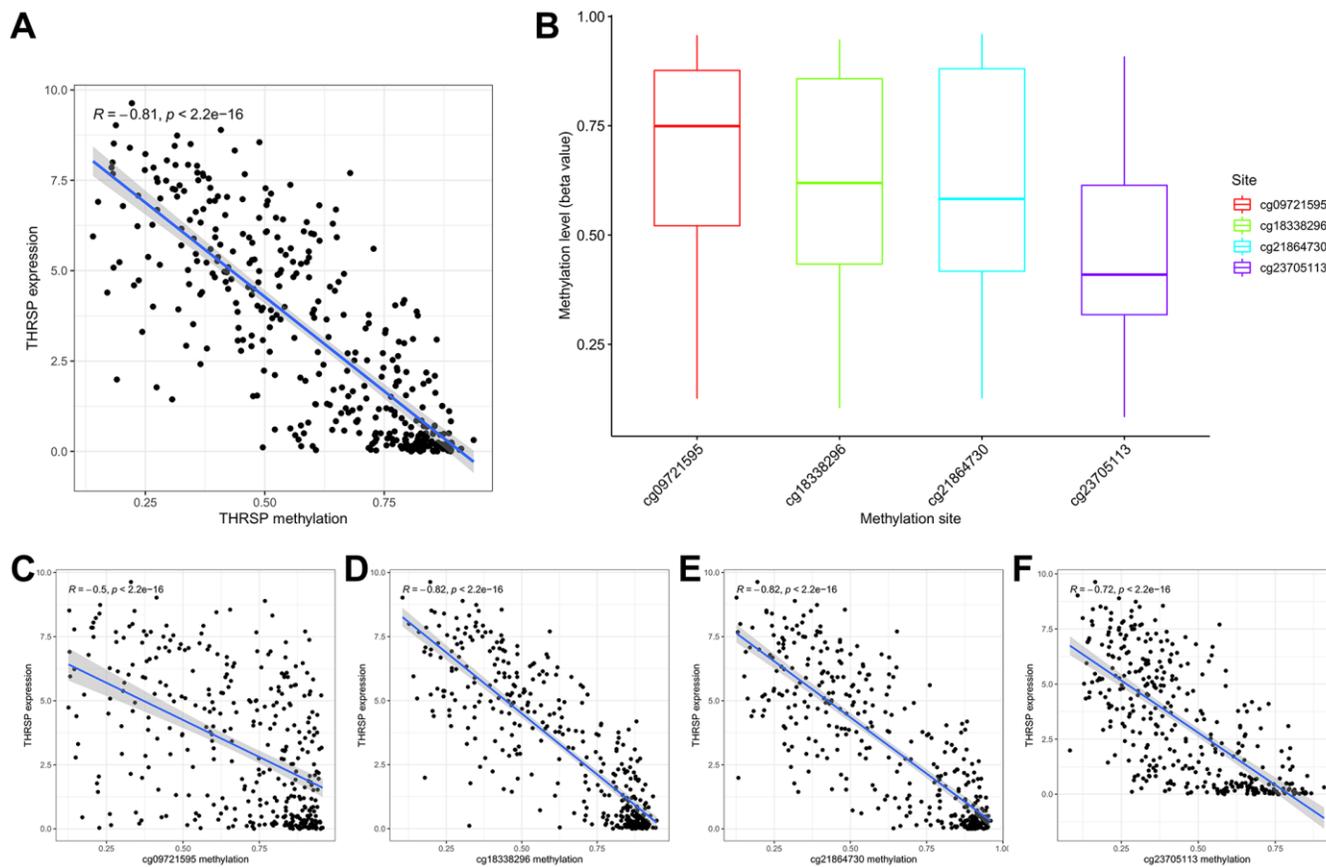
12. Wellberg EA, Rudolph MC, Lewis AS, Padilla-Just N, Jedlicka P, Anderson SM. Modulation of tumor fatty acids, through overexpression or loss of thyroid hormone responsive protein spot 14 is associated with altered growth and metastasis. *Breast Cancer Res.* 2014; 16:481.  
<https://doi.org/10.1186/s13058-014-0481-z>  
PMID:[25472762](https://pubmed.ncbi.nlm.nih.gov/25472762/)
13. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87–108.  
<https://doi.org/10.3322/caac.21262> PMID:[25651787](https://pubmed.ncbi.nlm.nih.gov/25651787/)
14. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang XS, Bannon F, Ahn JV, Johnson CJ, Bonaventure A, Marcos-Gragera R, Stiller C, Azevedo e Silva G, et al, and CONCORD Working Group. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet.* 2015; 385:977–1010.  
[https://doi.org/10.1016/S0140-6736\(14\)62038-9](https://doi.org/10.1016/S0140-6736(14)62038-9)  
PMID:[25467588](https://pubmed.ncbi.nlm.nih.gov/25467588/)
15. Subbaiah PV, Liu M, Witt TR. Impaired cholesterol esterification in the plasma in patients with breast cancer. *Lipids.* 1997; 32:157–62.  
<https://doi.org/10.1007/s11745-997-0020-5>  
PMID:[9075205](https://pubmed.ncbi.nlm.nih.gov/9075205/)
16. Cooke AL, Morris J, Melchior JT, Street SE, Jerome WG, Huang R, Herr AB, Smith LE, Segrest JP, Remaley AT, Shah AS, Thompson TB, Davidson WS. A thumbwheel mechanism for APOA1 activation of LCAT activity in HDL. *J Lipid Res.* 2018; 59:1244–55.  
<https://doi.org/10.1194/jlr.M085332> PMID:[29773713](https://pubmed.ncbi.nlm.nih.gov/29773713/)
17. Russell MR, Graham C, D'Amato A, Gentry-Maharaj A, Ryan A, Kalsi JK, Ainley C, Whetton AD, Menon U, Jacobs I, Graham RL. A combined biomarker panel shows improved sensitivity for the early detection of ovarian cancer allowing the identification of the most aggressive type II tumours. *Br J Cancer.* 2017; 117:666–74.  
<https://doi.org/10.1038/bjc.2017.199> PMID:[28664912](https://pubmed.ncbi.nlm.nih.gov/28664912/)
18. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, Lin Y, Wang D, Yang X, Zheng Y, Sang X, Zhao H. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. *Theranostics.* 2019; 9:7251–67.  
<https://doi.org/10.7150/thno.31155> PMID:[31695766](https://pubmed.ncbi.nlm.nih.gov/31695766/)
19. Zheng Y, Liu Y, Zhao S, Zheng Z, Shen C, An L, Yuan Y. Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma. *Cancer Manag Res.* 2018; 10:6079–96.  
<https://doi.org/10.2147/CMAR.S181396>  
PMID:[30538557](https://pubmed.ncbi.nlm.nih.gov/30538557/)
20. Ouyang G, Yi B, Pan G, Chen X. A robust twelve-gene signature for prognosis prediction of hepatocellular carcinoma. *Cancer Cell Int.* 2020; 20:207.  
<https://doi.org/10.1186/s12935-020-01294-9>  
PMID:[32514252](https://pubmed.ncbi.nlm.nih.gov/32514252/)
21. Jiang CH, Yuan X, Li JF, Xie YF, Zhang AZ, Wang XL, Yang L, Liu CX, Liang WH, Pang LJ, Zou H, Cui XB, Shen XH, et al. Bioinformatics-based screening of key genes for transformation of liver cirrhosis to hepatocellular carcinoma. *J Transl Med.* 2020; 18:40.  
<https://doi.org/10.1186/s12967-020-02229-8>  
PMID:[32000807](https://pubmed.ncbi.nlm.nih.gov/32000807/)
22. Ruan HY, Yang C, Tao XM, He J, Wang T, Wang H, Wang C, Jin GZ, Jin HJ, Qin WX. Downregulation of ACSM3 promotes metastasis and predicts poor prognosis in hepatocellular carcinoma. *Am J Cancer Res.* 2017; 7:543–53.  
PMID:[28401010](https://pubmed.ncbi.nlm.nih.gov/28401010/)
23. Buzzelli G, Dattolo P, Pinzani M, Brocchi A, Romano S, Gentilini P. Circulating growth hormone and insulin-like growth factor-I in nonalcoholic liver cirrhosis with or without superimposed hepatocarcinoma: evidence of an altered circadian rhythm. *Am J Gastroenterol.* 1993; 88:1744–8.  
PMID:[8213718](https://pubmed.ncbi.nlm.nih.gov/8213718/)
24. Kaseb AO, Morris JS, Hassan MM, Siddiqui AM, Lin E, Xiao L, Abdalla EK, Vauthey JN, Aloia TA, Krishnan S, Abbruzzese JL. Clinical and prognostic implications of plasma insulin-like growth factor-1 and vascular endothelial growth factor in patients with hepatocellular carcinoma. *J Clin Oncol.* 2011; 29:3892–9.  
<https://doi.org/10.1200/JCO.2011.36.0636>  
PMID:[21911725](https://pubmed.ncbi.nlm.nih.gov/21911725/)
25. Luo SM, Tan WM, Deng WX, Zhuang SM, Luo JW. Expression of albumin, IGF-1, IGFBP-3 in tumor tissues and adjacent non-tumor tissues of hepatocellular carcinoma patients with cirrhosis. *World J Gastroenterol.* 2005; 11:4272–6.  
<https://doi.org/10.3748/wjg.v11.i27.4272>  
PMID:[16015705](https://pubmed.ncbi.nlm.nih.gov/16015705/)
26. García-Galiano D, Sánchez-Garrido MA, Espejo I, Montero JL, Costán G, Marchal T, Membrives A, Gallardo-Valverde JM, Muñoz-Castañeda JR, Arévalo E, De la Mata M, Muntané J. IL-6 and IGF-1 are independent prognostic factors of liver steatosis and non-alcoholic steatohepatitis in morbidly obese patients. *Obes Surg.* 2007; 17:493–503.  
<https://doi.org/10.1007/s11695-007-9087-1>  
PMID:[17608262](https://pubmed.ncbi.nlm.nih.gov/17608262/)
27. Stuver SO, Kuper H, Tzonou A, Lagiou P, Spanos E, Hsieh CC, Mantzoros C, Trichopoulos D. Insulin-like growth factor 1 in hepatocellular carcinoma and

- metastatic liver cancer in men. *Int J Cancer*. 2000; 87:118–21.  
[https://doi.org/10.1002/1097-0215\(20000701\)87:1<118::AID-IJC17>3.0.CO;2-W](https://doi.org/10.1002/1097-0215(20000701)87:1<118::AID-IJC17>3.0.CO;2-W) PMID:10861461
28. Mazziotti G, Sorvillo F, Morisco F, Carbone A, Rotondi M, Stornaiuolo G, Precone DF, Cioffi M, Gaeta GB, Caporaso N, Carella C. Serum insulin-like growth factor I evaluation as a useful tool for predicting the risk of developing hepatocellular carcinoma in patients with hepatitis C virus-related cirrhosis: a prospective study. *Cancer*. 2002; 95:2539–45.  
<https://doi.org/10.1002/cncr.11002> PMID:12467068
29. Su TS, Liu WY, Han SH, Jansen M, Yang-Fen TL, P'eng FK, Chou CK. Transcripts of the insulin-like growth factors I and II in human hepatoma. *Cancer Res*. 1989; 49:1773–7.  
PMID:2466561
30. Yakar S, Liu JL, Stannard B, Butler A, Accili D, Sauer B, LeRoith D. Normal growth and development in the absence of hepatic insulin-like growth factor I. *Proc Natl Acad Sci USA*. 1999; 96:7324–9.  
<https://doi.org/10.1073/pnas.96.13.7324>  
PMID:10377413
31. Daughaday WH. The possible autocrine/paracrine and endocrine roles of insulin-like growth factors of human tumors. *Endocrinology*. 1990; 127:1–4.  
<https://doi.org/10.1210/endo-127-1-1>  
PMID:2163304
32. Cho EJ, Lee JH, Yoo JJ, Choi WM, Lee MJ, Cho Y, Lee DH, Lee YB, Kwon JH, Yu SJ, Lee JM, Suh KS, Kim K, et al. Serum insulin-like growth factor-I level is an independent predictor of recurrence and survival in early hepatocellular carcinoma: a prospective cohort study. *Clin Cancer Res*. 2013; 19:4218–27.  
<https://doi.org/10.1158/1078-0432.CCR-12-3443>  
PMID:23757355
33. Xia Q, Han T, Yang P, Wang R, Li H, Zhang J, Zhou X. MicroRNA-28-5p Regulates Liver Cancer Stem Cell Expansion via IGF-1 Pathway. *Stem Cells Int*. 2019; 2019:8734362.  
<https://doi.org/10.1155/2019/8734362>  
PMID:31885628
34. Sprinzl MF, Puschnik A, Schlitter AM, Schad A, Ackermann K, Esposito I, Lang H, Galle PR, Weinmann A, Heikenwälder M, Protzer U. Sorafenib inhibits macrophage-induced growth of hepatoma cells by interference with insulin-like growth factor-1 secretion. *J Hepatol*. 2015; 62:863–70.  
<https://doi.org/10.1016/j.jhep.2014.11.011>  
PMID:25463538
35. Chen YM, Qi S, Perrino S, Hashimoto M, Brodt P. Targeting the IGF-Axis for Cancer Therapy: Development and Validation of an IGF-Trap as a Potential Drug. *Cells*. 2020; 9:E1098.  
<https://doi.org/10.3390/cells9051098> PMID:32365498
36. Makridakis NM, Ross RK, Pike MC, Crocitto LE, Kolonel LN, Pearce CL, Henderson BE, Reichardt JK. Association of mis-sense substitution in SRD5A2 gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA. *Lancet*. 1999; 354:975–8.  
[https://doi.org/10.1016/S0140-6736\(98\)11282-5](https://doi.org/10.1016/S0140-6736(98)11282-5)  
PMID:10501358
37. Luo J, Dunn TA, Ewing CM, Walsh PC, Isaacs WB. Decreased gene expression of steroid 5 alpha-reductase 2 in human prostate cancer: implications for finasteride therapy of prostate carcinoma. *Prostate*. 2003; 57:134–9.  
<https://doi.org/10.1002/pros.10284>  
PMID:12949937
38. Iehlé C, Radvanyi F, Gil Diez de Medina S, Ouafik LH, Gérard H, Chopin D, Raynaud JP, Martin PM. Differences in steroid 5alpha-reductase iso-enzymes expression between normal and pathological human prostate tissue. *J Steroid Biochem Mol Biol*. 1999; 68:189–95.  
[https://doi.org/10.1016/S0960-0760\(99\)00030-8](https://doi.org/10.1016/S0960-0760(99)00030-8)  
PMID:10416833
39. Nam RK, Toi A, Vesprini D, Ho M, Chu W, Harvie S, Sweet J, Trachtenberg J, Jewett MA, Narod SA. V89L polymorphism of type-2, 5-alpha reductase enzyme gene predicts prostate cancer presence and progression. *Urology*. 2001; 57:199–204.  
[https://doi.org/10.1016/S0090-4295\(00\)00928-6](https://doi.org/10.1016/S0090-4295(00)00928-6)  
PMID:11164181
40. Tsunedomi R, Ogawa Y, Iizuka N, Sakamoto K, Tamesa T, Moribe T, Oka M. The assessment of methylated BASP1 and SRD5A2 levels in the detection of early hepatocellular carcinoma. *Int J Oncol*. 2010; 36:205–12.  
PMID:19956849
41. Rossi L, Leverì M, Gritti C, De Silvestri A, Zavaglia C, Sonzogni L, Silvestri L, Civardi E, Mondelli MU, Silini EM. Genetic polymorphisms of steroid hormone metabolizing enzymes and risk of liver cancer in hepatitis C-infected patients. *J Hepatol*. 2003; 39:564–70.  
[https://doi.org/10.1016/S0168-8278\(03\)00355-6](https://doi.org/10.1016/S0168-8278(03)00355-6)  
PMID:12971967
42. Moribe T, Iizuka N, Miura T, Stark M, Tamatsukuri S, Ishitsuka H, Hamamoto Y, Sakamoto K, Tamesa T, Oka M. Identification of novel aberrant methylation of BASP1 and SRD5A2 for early diagnosis of hepatocellular carcinoma by genome-wide search. *Int J Oncol*. 2008; 33:949–58.  
PMID:18949357

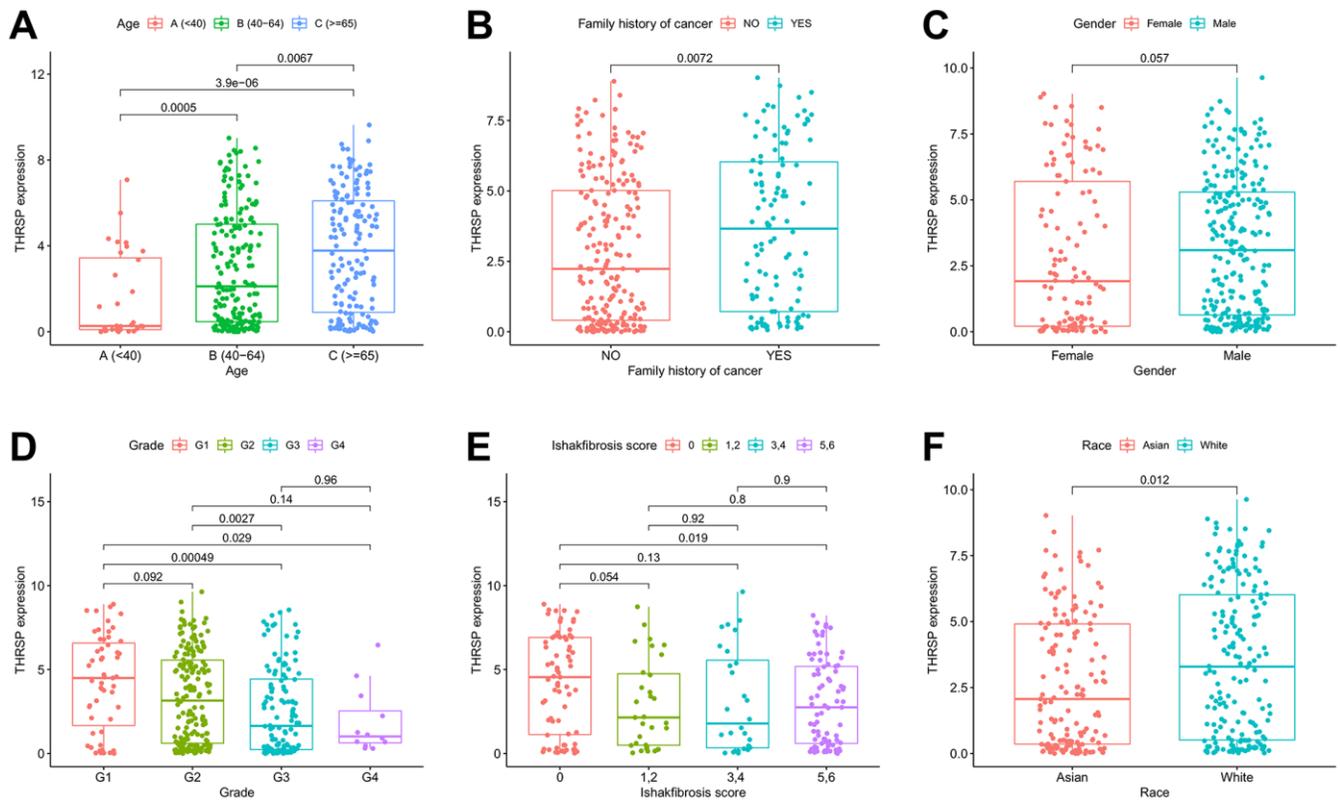
43. Liu CP, Zhang JH, Zheng SC, Liu J, Guo JC. A novel clinical multidimensional transcriptome signature predicts prognosis in bladder cancer. *Oncol Rep.* 2018; 40:2826–35. <https://doi.org/10.3892/or.2018.6677> PMID:30226624
44. Pang B, Xu X, Lu Y, Jin H, Yang R, Jiang C, Shao D, Liu Y, Shi J. Prediction of new targets and mechanisms for quercetin in the treatment of pancreatic cancer, colon cancer, and rectal cancer. *Food Funct.* 2019; 10:5339–49. <https://doi.org/10.1039/C9FO01168D> PMID:31393490
45. Chen D, Feng X, Lv Z, Xu X, Lu Y, Wu W, Wu H, Liu H, Cao L, Ye S, Chen J, Wu J. ACADS acts as a potential methylation biomarker associated with the proliferation and metastasis of hepatocellular carcinomas. *Aging (Albany NY).* 2019; 11:8825–44. <https://doi.org/10.18632/aging.102292> PMID:31652420
46. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLOS Comput Biol.* 2008; 4:e1000117. <https://doi.org/10.1371/journal.pcbi.1000117> PMID:18704157
47. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics.* 2014; 47:8.13.1–24. <https://doi.org/10.1002/0471250953.bi0813s47> PMID:25199793
48. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014 (Suppl 4); 8:S11. <https://doi.org/10.1186/1752-0509-8-S4-S11> PMID:25521941
49. Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J Cardiovasc Magn Reson.* 2012; 14:83. <https://doi.org/10.1186/1532-429X-14-83> PMID:23199362
50. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res.* 2010; 1:274–8. <https://doi.org/10.4103/0974-7788.76794> PMID:21455458
51. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020; 48:W509–14. <https://doi.org/10.1093/nar/gkaa407> PMID:32442275
52. Gu Y, Li X, Bi Y, Zheng Y, Wang J, Li X, Huang Z, Chen L, Huang Y, Huang Y. CCL14 is a prognostic biomarker and correlates with immune infiltrates in hepatocellular carcinoma. *Aging (Albany NY).* 2020; 12:784–807. <https://doi.org/10.18632/aging.102656> PMID:31927532
53. Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, Geller MA, Odunsi K, Beechem J, Fling SP. Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer.* 2017; 5:18. <https://doi.org/10.1186/s40425-017-0215-8> PMID:28239471
54. Liu B, Fan Y, Song Z, Han B, Meng Y, Cao P, Tan K. Identification of DRP1 as a prognostic factor correlated with immune infiltration in breast cancer. *Int Immunopharmacol.* 2020; 89:107078. <https://doi.org/10.1016/j.intimp.2020.107078> PMID:33049497
55. Zhong L, Xie L, Yang Z, Li L, Song S, Cao D, Liu Y. Prognostic value of S1PR1 and its correlation with immune infiltrates in breast and lung cancers. *BMC Cancer.* 2020; 20:766. <https://doi.org/10.1186/s12885-020-07278-2> PMID:32799825
56. Lin X, Wang Z, Yang G, Wen G, Zhang H. YTHDF2 correlates with tumor immune infiltrates in lower-grade glioma. *Aging (Albany NY).* 2020; 12:18476–500. <https://doi.org/10.18632/aging.103812> PMID:32986017

## SUPPLEMENTARY MATERIALS

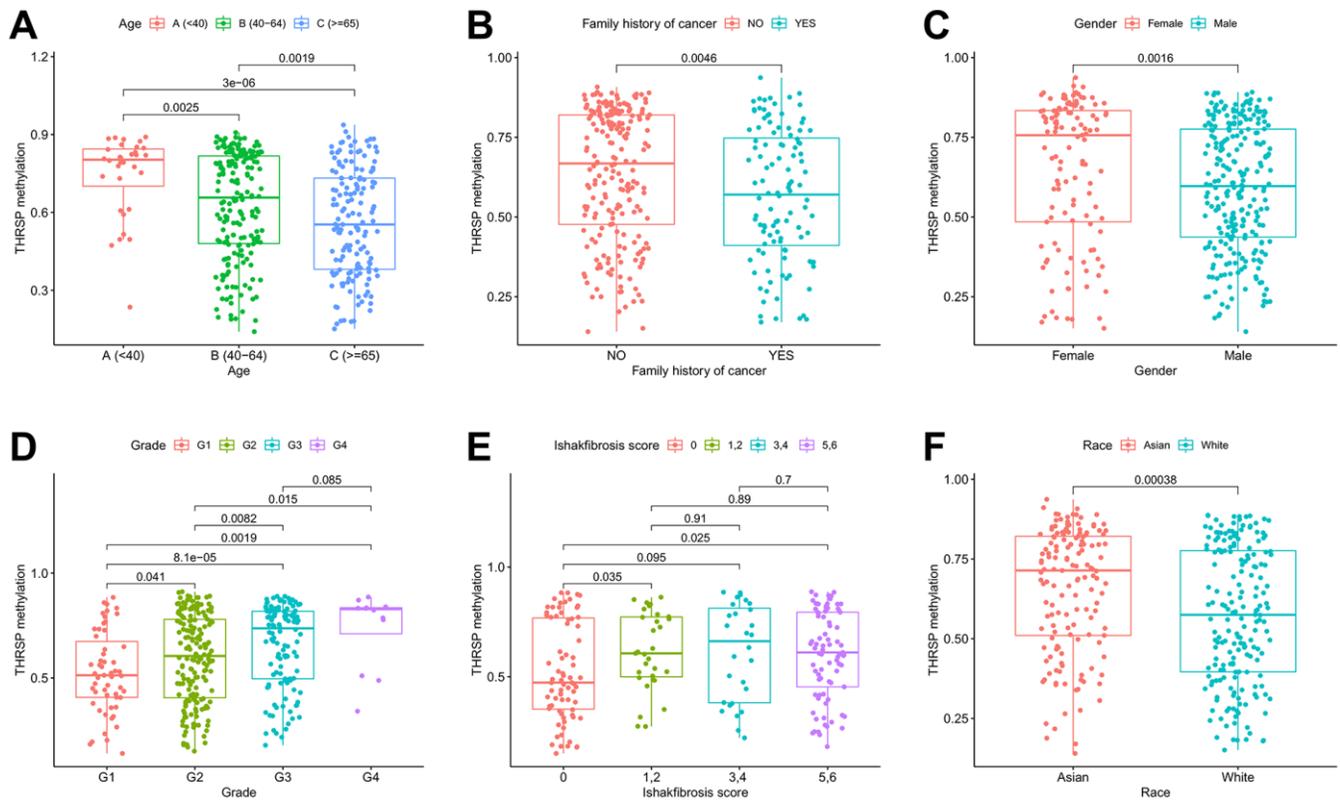
### Supplementary Figures



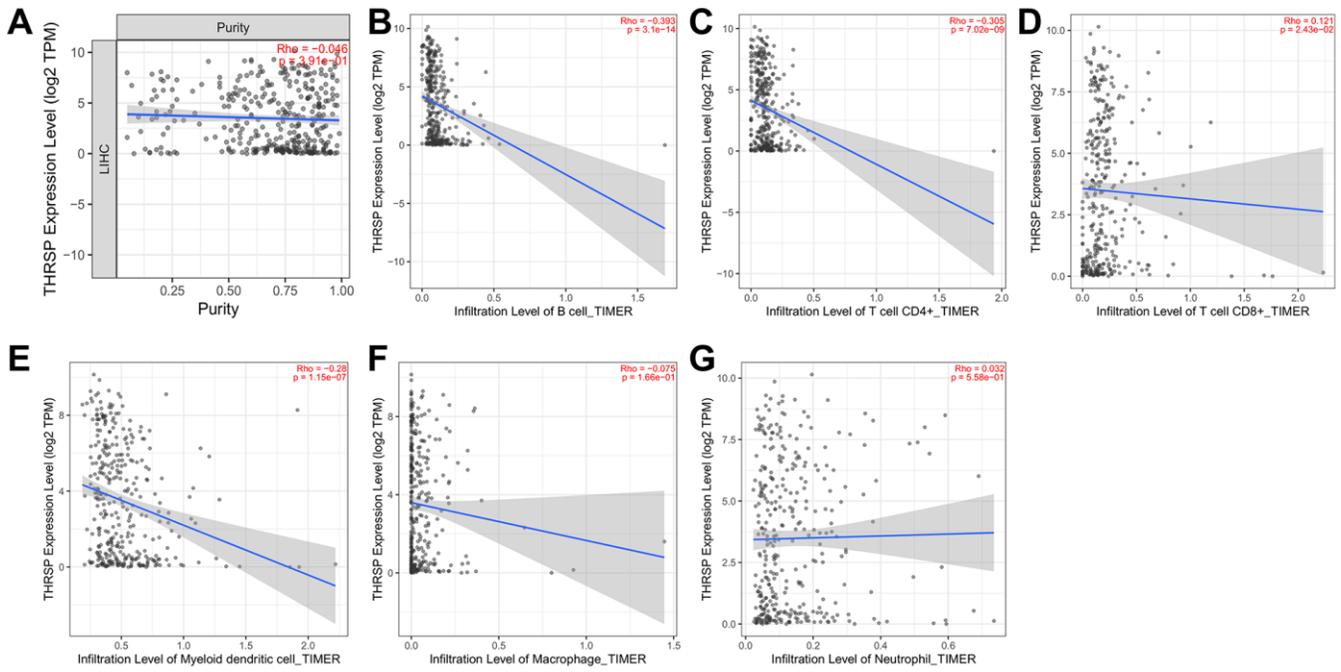
**Supplementary Figure 1. Relationship between THRSR mRNA expression and DNA methylation.** (A) THRSR expression was negatively correlated with DNA methylation ( $R = -0.81, p < 2.2e-16$ ). (B) The distribution of four THRSR DNA promoter CpG sites. (C–F) THRSR expression was negatively correlated with the methylation levels of the four CpG sites. From the left to right were cg09721595 ( $R = -0.5, p < 2.2e-16$ ), cg18338296 ( $R = -0.82, p < 2.2e-16$ ), cg21864730 ( $R = -0.82, p < 2.2e-16$ ) and cg23705113 ( $R = -0.72, p < 2.2e-16$ ).



**Supplementary Figure 2. The correlation between THRSF expression and the clinical features.** THRSF mRNA expression was stratified by (A) age, (B) family history of cancer, (C) gender, (D) grade, (E) Ishak fibrosis score, and (F) race.



**Supplementary Figure 3. The correlation between THRSF methylation and the clinical features.** THRSF methylation was stratified by (A) age, (B) family history of cancer, (C) gender, (D) grade, (E) Ishak fibrosis score, and (F) race.



**Supplementary Figure 4. The correlation between THRSP expression and immune infiltration levels in HCC. (A)** No association between THRSP expression and tumor purity ( $R = -0.046$ ,  $p = 0.391$ ). **(B–G)** THRSP expression was significantly negatively correlated with infiltrating levels of B cells, CD4+ T cells, and dendritic cells, positively correlated with CD8+ T cells, but not correlated with macrophages or neutrophils.

## Supplementary Table

Supplementary Table 1. Correlations between THRSP and related genes and markers of immune cells.

Description gene makers		LIHC			
		None		Purity	
		Cor	p	Cor	p
CD8+ T cell	CD8A	-0.067	0.198	-0.099	0.066
	CD8B	-0.078	0.134	-0.109	*
	CD3D	-0.152	**	-0.183	***
T cell (general)	CD3E	-0.105	*	-0.16	**
	CD2	-0.122	*	-0.176	**
	CD6	-0.111	*	-0.163	**
	CD3G	-0.108	*	-0.13	*
	CD19	-0.118	*	-0.14	**
B cell	KIAA0125	-0.126	*	-0.166	**
	SPIB	-0.242	***	-0.265	***
	PNOC	-0.161	**	-0.22	***
Monocyte	CD79A	-0.132	*	-0.181	***
	CD86	-0.143	**	-0.198	***
	CSF1R	-0.108	*	-0.167	**
Macrophages	CD68	-0.184	***	-0.218	***
	CD84	-0.004	0.946	-0.047	0.388
	MS4A4A	-0.034	0.512	-0.084	0.12
TAM	CCL2	0.039	0.452	0.033	0.541
	CD68	-0.184	***	-0.218	***
	IL10	-0.052	0.319	-0.093	0.086
M1	CSF1R	-0.108	*	-0.167	**
	IRF5	-0.061	0.241	-0.065	0.228
	NOS2	0.126	*	0.106	*
M2	PTGS2	-0.006	0.907	-0.028	0.605
	CD163	0.022	0.67	-0.012	0.831
	VSIG4	0.065	0.211	0.04	0.455
Neutrophils	MS4A4A	-0.034	0.512	-0.084	0.12
	MRC1	0.162	**	0.144	**
	CEACAM8	-0.06	0.246	-0.077	0.152
	ITGAM	0.052	0.322	0.03	0.578
	CCR7	-0.013	0.809	-0.049	0.362
	FCGR3B	0.154	**	0.154	**
	SIGLEC5	-0.11	*	-0.17	**
	CSF3R	-0.138	**	-0.19	***
	HLA-DPB1	-0.151	**	-0.208	***
	HLA-DQB1	-0.13	*	-0.177	***
Dendritic cell	HLA-DRA	-0.091	0.08	-0.139	**
	HLA-DPA1	-0.101	0.052	-0.157	**
	BDCA-1	-0.14	**	-0.17	**
	BDCA-4	-0.192	***	-0.194	***
	ITGAX	-0.111	*	-0.162	**
	HSD11B1	0.523	***	0.522	***
	CD209	-0.017	0.743	-0.04	0.46

	STAT4	0.002	0.97	0.002	0.974
	STAT1	-0.155	**	-0.167	**
Th1	TBX21	-0.012	0.82	-0.05	0.351
	IFNG	-0.047	0.362	-0.078	0.149
	TNF	-0.103	*	-0.151	**
	STAT5A	-0.126	*	-0.137	*
	IL13	0.287	***	0.297	***
Th2	GATA3	-0.079	0.127	-0.124	*
	STAT6	-0.099	0.056	-0.108	*
	CXCR4	-0.186	***	-0.222	***
Th1-like	BHLHE40	0.305	***	0.3	***
	RORC	0.251	***	0.285	***
Th17	CCR6	-0.313	***	-0.33	***
	CCR8	-0.06	0.251	-0.09	0.096
Treg	STAT5B	0.006	0.908	0.006	0.917
	TGFB1	-0.212	***	-0.246	***
Resting Treg T-cell	IL2RA	-0.139	**	-0.189	***
	FOXP3	0.211	***	0.198	***
Effective Treg T-cell	CTLA4	-0.224	***	-0.269	*
	TNFRSF9	-0.103	*	-0.129	*
Naïve T-cell	TCF7	-0.15	**	-0.17	**
Effective memory T-cell	DUSP4	-0.27	***	-0.328	***
	ITGAE	-0.443	***	-0.45	***
Resistant memory T-cell	CXCR6	-0.072	0.167	-0.119	*
	MYADM	-0.235	***	-0.243	***
	HAVCR2	-0.127	*	-0.185	***
	TIGIT	-0.11	*	-0.153	**
	LAYN	-0.129	*	-0.149	**
Exhausted T-cell	PDCD1	-0.209	***	-0.247	***
	CTLA4	-0.224	***	-0.269	***
	LAG3	-0.114	*	-0.121	*
	PTGER4	-0.124	*	-0.149	**

LHC, Liver hepatocellular carcinoma. None, correlation without adjustment. Purity, correlation adjusted by purity.  
P-value: 0 ≤ \*\*\* < 0.001 ≤ \*\* < 0.01 ≤ \* < 0.05.

# Polo-like kinases as potential targets and PLK2 as a novel biomarker for the prognosis of human glioblastoma

Yiming Ding<sup>1,\*</sup>, Hanjie Liu<sup>2,\*</sup>, Chuanbao Zhang<sup>1</sup>, Zhaoshi Bao<sup>1</sup>, Shuqing Yu<sup>1</sup>

<sup>1</sup>Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

<sup>2</sup>Department of Neurosurgery, Beijing Neurosurgical Institute, Capital Medical University, Beijing, China

\*Equal contribution and co-first authors

**Correspondence to:** Shuqing Yu; email: [yushuqing@bjtth.org](mailto:yushuqing@bjtth.org)

**Keywords:** polo-like kinases, glioblastoma, prognosis, biomarker, ON1231320

**Received:** July 19, 2021

**Accepted:** February 28, 2022

**Published:** March 7, 2022

**Copyright:** © 2022 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

The most prevalent malignant central nervous system (CNS) cancer is glioblastoma multiforme (GBM). PLKs (polo-like kinases) are a kind of serine-threonine kinase that modulate DNA replication, mitosis, and stress responses. PLKs in GBM need to be better studied and examined in terms of their expression, function, along with prognostic significance. Using an existing publicly available data set, we evaluated the expression level and prognostic relevance of PLKs in GBM patients at the molecular level. The biological processes along with cascades of the screened gene were predicted using the functional enrichment of Gene Set Enrichment Analysis, Gene Ontology, and Kyoto Encyclopedia of Genes and Genomes pathways. The data illustrated that PLK1/3/4 contents were greater in GBM tissues than in non-tumorous tissues, but PLK2/5 expression levels were lower. PLK2 expression was also linked to patient outcome in GBM. Our findings imply that PLKs might be useful molecular indicators as well as prospective treatment targets for GBM. A PLK2 inhibitor has been studied for the first time in a glioma cell in this work. In glioma cells, ON1231320 has anticancer effects. Finally, a summary of PLK inhibitors is presented, along with projections for future progress.

## INTRODUCTION

GBM is the most frequent malignant CNS cancer in adults, which accounts for 55% of all gliomas [1]. Surgical resection, radiation and chemotherapy constitutes the current standard of treatment. However, the median time of survival is about 18 months, the survival rate is still poor [2]. Thus, treatment of GBM remains challenging. Based on high-throughput genetic, genomic, and epigenetic data, several key molecules have been identified that contribute to GBM carcinogenesis and the development of targeted therapies for individual subtypes. However, targeted therapies for specific mutations or subtypes have mostly failed due to the complexity of molecular heterogeneity within tumors [3].

Polo-like kinases (PLKs) were first found to play an indispensable role in the mitosis of *Drosophila*

*melanogaster* [4]. Five members of PLK family have been discovered, including PLK1, PLK2, PLK3, PLK4, and PLK5 [5]. PLKs are important regulatory factors of the cell cycle in non-tumorous and cancer cells [6]. Recent advances in PLK1 research have dramatically improved our comprehension of its modulation, targets along with function. PLK1 plays an indispensable role in the regulation of the cell cycle, including activation of the APC/C (anaphase-promoting complex/cyclosome), entry into mitosis, assembly of the bipolar spindle, sister chromatid splitting, and centrosome maturation [7]. Despite remarkable advances in the research of PLK1, the functional roles of other members of the PLK family, particularly PLK2 along with PLK3, are still unknown. PLK2 along with PLK3 have been established as key mediators of DNA damage or oxidative stress in cancer cells. A growing body of research evidence suggests that PLKs and the tumor

repressor p53 interact in cancer cells [7, 8]. PLKs are aberrantly expressed in tumor cells, thereby promoting abnormal proliferation of cancer cells [9]. PLKs show aberrant expression in multiple cancers and are linked to poor prognosis. These findings have the potential to provide proof for anti-cancer medication development targets. Drug manufacturers have also invested in the research of PLKs in order to identify novel cancer medicines [6]. PLK1 inhibitors have now been tested clinically and found to be beneficial in the treatment of individuals with cancer [7].

Herein, we comprehensively purposed to evaluate the expression, function along with prognostic value of PLKs in GBM. Based on existing gene expression or copy number variants published online, we conducted a detailed analysis of PLK expression, as well as mutations in individuals with GBM. Furthermore, we performed validation using The Cancer Genome Atlas (TCGA) along with the Chinese Glioma Genome Atlas (CGGA) data resources to determine the expression, potential function, and prognostic value of PLKs in GBM.

## RESULTS

### PLK expression levels in patients with GBM

We compared the expression levels of PLKs in cancer with those in non-tumorous samples via the ONCOMINE data resources (Figure 1A). The transcript expression levels of PLK1/3/4 were remarkably elevated in individuals with GBM, and that of PLK2 was remarkably downregulated in patients with GBM. PLK5 expression was not remarkably different between GBM and non-tumorous tissues. All PLK mRNA expression levels were not remarkable in low-grade glioma (LGG) tissues. The fold change and *p*-values of the PLKs are shown in Table 1. Through the Gene Expression Profiling Interactive Analysis (GEPIA2) cohort, we compared the transcript expression levels of PLKs between GBM and non-tumorous tissues (Figure 1B). The data illustrated that the contents of PLK1/3/4 were higher in GBM tissues than in non-tumorous tissues, whilst the contents of PLK2 and PLK5 were lower in GBM (Figure 1C).

As shown in the CGGA data resource, the contents of PLK1/2/3/4 rose as the World Health Organization (WHO) grades of glioma increased. However, the expression levels of PLK5 decreased as the WHO grades of glioma increased. (Figure 1D).

Besides, immuno-histochemistry (IHC) staining abstracted from the Human Protein Atlas (HPA) data resource exhibited the contents of PLKs as shown in Figure 2A. PLK1/3/4 were over-expressed in GBM in

contrast with non-tumorous tissues. On the contrary, PLK2 was more highly expressed in non-tumorous tissues than in GBM. PLK5 was expressed in neither GBM nor non-tumorous tissues (Figure 2A). Besides, immuno-histochemistry illustrated that the staining index score of PLK1-4 in AII was frequently less than two. In sGBMIVs, the staining index score was frequently more than four (Figure 2B). The data illustrated that the content of PLK1-4 rose in tandem with rising tumor grades, confirming our prediction.

We analyzed the PLK alterations using the cBioPortal online tool for GBM. PLK alterations included the following forms: deep deletion, amplification, mRNA high, and mRNA low. Two or more PLK alterations were seen in 29% of the samples (136 samples) (Figure 3A and 3B). In addition, using the same tool, we found that PLK alterations typically occur in the primary GBM (Figure 3C).

### Relationship of PLK contents with prognosis in GBM

We further assessed the pivotal efficiency of PLKs in the survival of individuals with GBM. Kaplan–Meier Plotter tools were adopted to assess the relationship of PLK transcript contents with the survival of patients with GBM using GraphPad Prism (Figure 4A). The Kaplan–Meier curve along with the log-rank test analyses revealed that the decreased PLK2 expression level was remarkably linked to overall survival (OS) ( $p < 0.05$ ). The individuals with GBM with high mRNA contents of the PLK2 were predicted to have poor OS. While PLK1/3/4/5 showed no prognostic significance ( $p > 0.05$ ). This result was also validated in the CGGA data resource (Figure 4B). The high contents of PLK2 were remarkably linked to poor prognosis ( $p < 0.05$ ). Specifically, we retrieved clinical information from the TCGA data resource, and K-M analysis showed that high PLK2 expression likewise predicted poor patient prognosis ( $p < 0.05$ ).

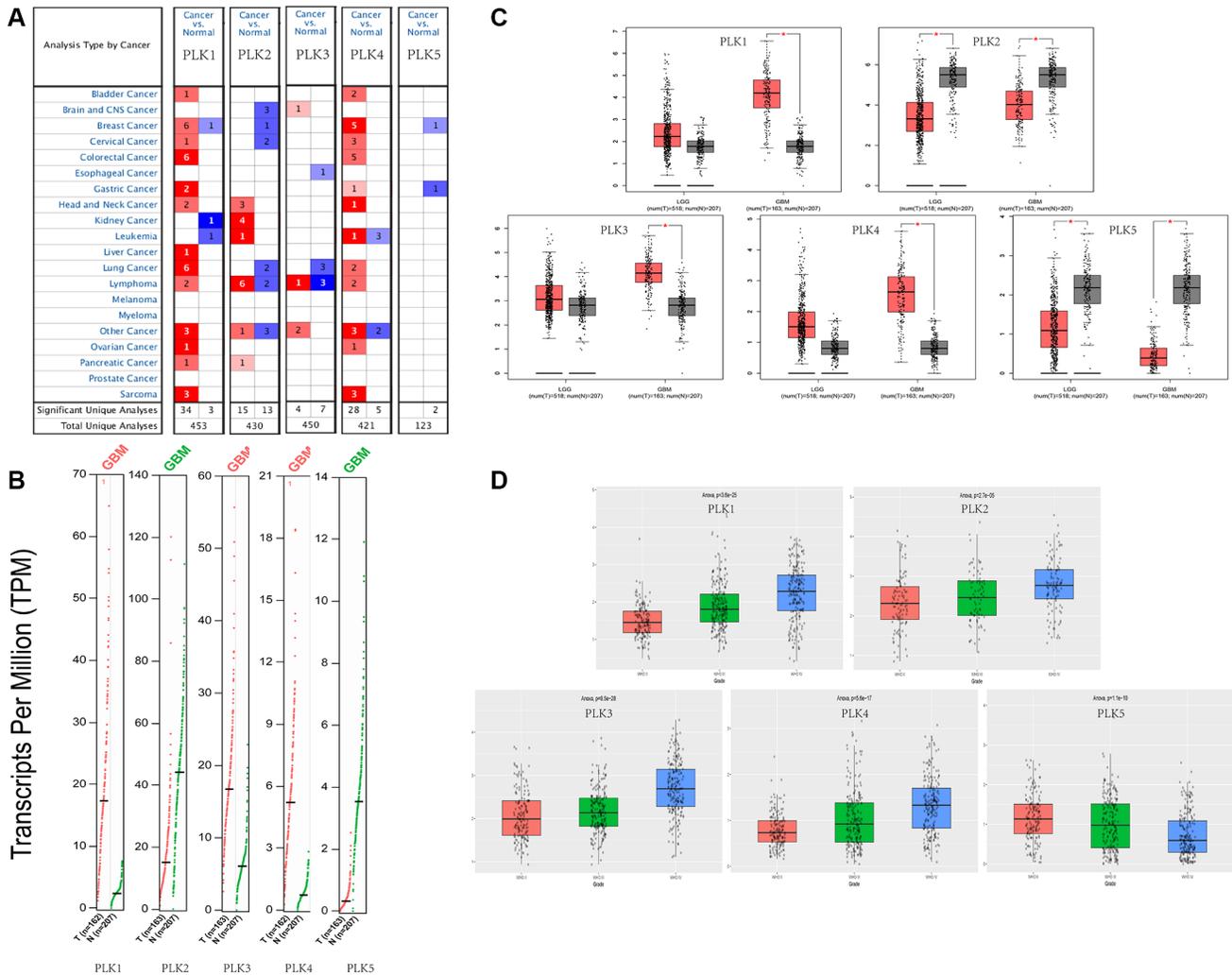
We evaluated the associations between age, gender, isocitrate dehydrogenase 1 (IDH1) status, PLK expression levels, and patients' OS in univariate along with multivariate Cox regression analyses. The univariate data revealed that age, IDH1 status, and PLK 2/3 expression were remarkably linked to patients' OS. Multivariate analysis revealed that age and PLK2 expression were remarkably linked to patients' OS (Table 2).

### Relationship between PLK mRNA levels and GBM

We also computed the correlations among PLKs by analyzing their transcript expression levels via the

**Table 1. PLK expression levels between different types of gliomas and brain tissues (ONCOMINE database).**

	Types of glioma vs. brain	Fold change	P value	t-test	Ref
PLK1	Glioblastoma	1.276	2.21E-05	6.828	Murat
	Astrocytoma	-1.073	0.674	-0.472	Rickman
PLK2	Glioblastoma	-1.7	9.03E-06	-6.079	Bredel
	Astrocytoma	2.186	0.035	2.308	Shai
PLK3	Glioblastoma	4.056	1.42E-05	4.706	Pomerooy
	Astrocytoma	-1.224	0.89	-1.353	Rickman
PLK4	Glioblastoma	1.47	4.92E-04	5.809	Murat
	Astrocytoma	-1.556	0.979	-2.257	Rickman
PLK5	Glioblastoma	1.736	0.077	1.922	Lee
	astrocytoma	1.255	0.05	1.723	Sun



**Figure 1. Expression levels of PLKs in TCGA and CGGA databases. (A)** Expression levels of PLKs in patients with GBM. **(B)** Gene Expression Profile shows the expression of PLKs between GBM (red) and normal tissue (green) via GEPIA2. **(C)** Box plot shows the expression of PLKs between brain tumor (red) and normal tissue (grey) via GEPIA2. The “\*” indicates that the *p*-value is statistically significant. **(D)** Box plot shows the expression of PLKs in different WHO classes via CGGA database.

**Table 2. Univariate and multivariate Cox proportional hazards regression analyses of prognostic factors for GBM.**

	Univariate analysis			Multivariate analysis		
	HR	95% CI	<i>p</i>	HR	95% CI	<i>p</i>
<b>Age</b>						
<64	4.036	1.019–1.054	5.43E-05*	3.183	1.0131–1.056	0.00146*
≥65						
<b>Gender</b>						
Female	−0.773	0.5828–1.264	0.439	−0.747	0.5612–1.296	0.45509
Male						
<b>IDH1_status</b>						
Mutant	−2.898	0.1035–0.6455	0.00376*	−0.408	0.2755–2.329	0.68352
Wild type						
<b>PLK1</b>						
High	0.023	0.6902–1.461	0.982	−0.518	0.5498–1.416	0.60434
Low						
<b>PLK2</b>						
High	3.236	1.285–2.777	0.00121*	3.033	1.2583–2.913	0.00242*
Low						
<b>PLK3</b>						
High	2.033	1.014–2.164	0.0421*	1.564	0.919–2.123	0.11775
Low						
<b>PLK4</b>						
High	−0.202	0.6583–1.405	0.84	−0.124	0.6111–1.543	0.90136
Low						
<b>PLK5</b>						
High	1.918	0.9918–2.133	0.0551	1.429	0.8961–2.016	0.15288
Low						

**Table 3. Analysis of mutual exclusivity between PLKs (cBioPortal).**

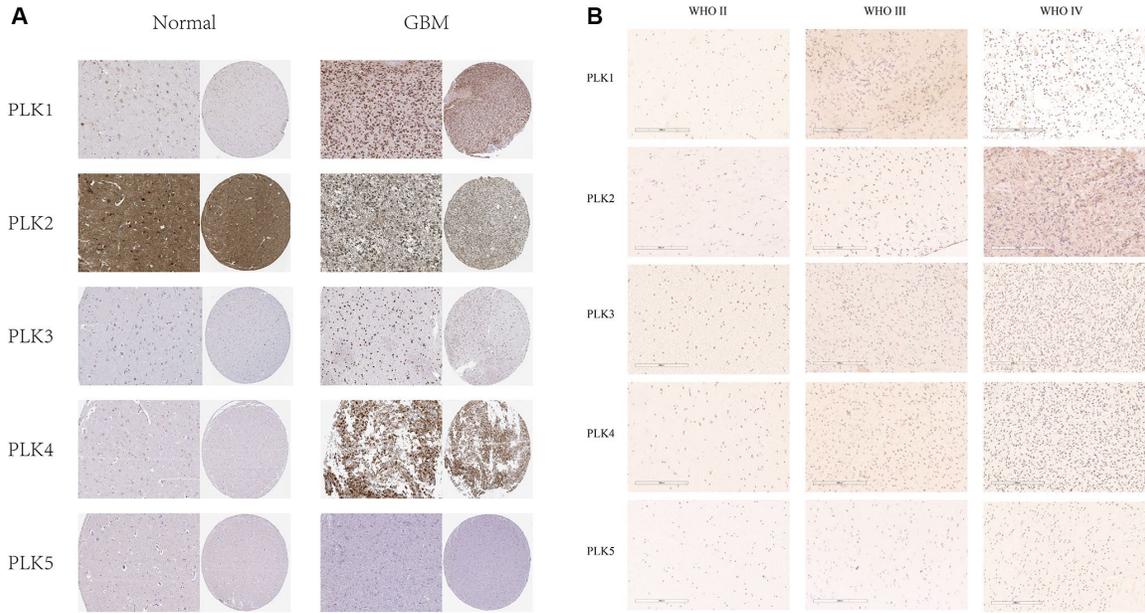
A	B	Log2 Odds Ratio	<i>p</i> -Value	<i>q</i> -Value	Tendency
PLK1	PLK4	>3	<0.001	0.003	Co-occurrence
PLK1	PLK3	<−3	0.114	0.568	Mutual exclusivity
PLK2	PLK3	0.7	0.324	0.9	Co-occurrence
PLK1	PLK2	0.497	0.433	0.9	Co-occurrence
PLK3	PLK4	−0.837	0.501	0.9	Mutual exclusivity
PLK2	PLK4	0	0.634	0.9	Mutual exclusivity
PLK2	PLK5	<−3	0.818	0.9	Mutual exclusivity
PLK3	PLK5	<−3	0.855	0.9	Mutual exclusivity
PLK1	PLK5	<−3	0.882	0.9	Mutual exclusivity
PLK4	PLK5	<−3	0.9	0.9	Mutual exclusivity

The analysis tested 10 pairs between the 5 tracks in the OncoPrint. Odds Ratio: Quantifies how strongly the presence or absence of alterations in A are associated with the presence or absence of alterations in B in the selected samples. *p*-Value: Derived from one-sided Fisher Exact Test. *q*-Value: Derived from Benjamini-Hochberg FDR correction procedure. Tendency: Log2 ratio >0: Tendency towards co-occurrence; Log2 ratio ≤0: Tendency towards mutual exclusivity; *q*-Value <0.05: Significant association.

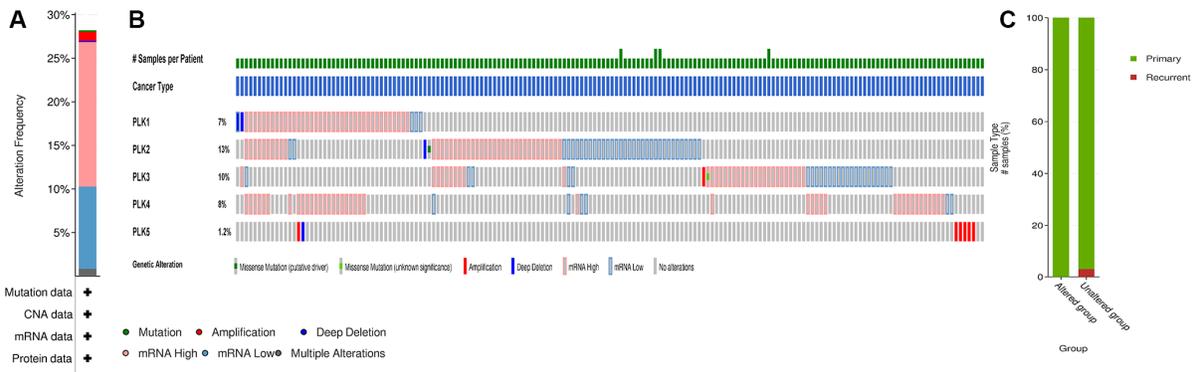
TCGA and CGGA data resources for GBM, and Pearson’s correction was conducted. The data exhibited remarkable correlations in several PLKs. PLK5 was negatively correlated with PLK1, PLK2, PLK3, and PLK4. A positive relationship was also evident between PLK1 and PLK4 (Figure 5). As shown in Table 3, the analysis of mutual exclusivity between PLKs found that the relationship between PLK1 and PLK4 is characterized by co-occurrence (Table 3).

**Predicted roles and cascades of the changes in PLKs factors and their frequently altered adjacent genes individuals with GBM**

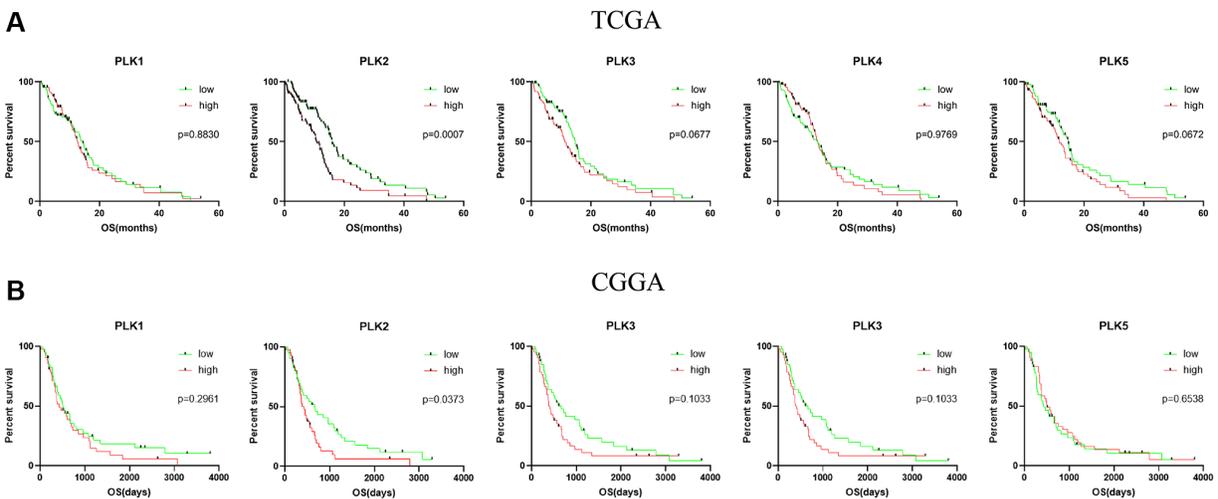
Using GeneMANIA in Cytoscape, we evaluated 100 genes linked to PLKs and built a network. We explored functional enrichment of GO (Gene Ontology) terms along with KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways in the network to elucidate the



**Figure 2.** (A) IHC illustration of PLK expression levels in GBM (HPA). (B) Immunohistochemical staining of PLKs in gliomas.



**Figure 3.** PLK mutation analysis in GBM (cBioPortal). (A) Frequency of alteration in each query in the detailed cancer types. (B) Overview of genetic alteration per sample in each query gene. (C) Explore clinical sample type comparisons among groups of samples as defined by the query.



**Figure 4.** Kaplan-Meier analysis of the prognostic value of PLKs. (A) The TCGA databases. (B) The CGGA databases.

biological processes (BPs) and cascades of the genes. The top ten BP, CC, and MF items were chosen (Figure 6).

Nuclear division, segregation of mitotic nuclear division chromosome, modulation of mitotic cell cell cycle phase transition, segregation of nuclear chromosome, organization of mitotic spindle, segregation of chromatid, cell transition of cycle G2/M phase, along with segregation of mitotic sister chromatid are all BPs. The condensed chromosome in the spindle, the spindle pole midbody, the chromosomal region, the condensed chromosome centromeric region, the kinetochore, the condensed chromosome kinetochore, the chromosome centromeric region, and the mitotic spindle are all parts of the condensed chromosome centromeric region. Tubulin docking, microtubule docking, motor activity, protein C-terminus docking, magnesium ion docking, histone kinase activity, histone serine kinase activity, and cyclin-dependent protein kinase activity constituted the MFs. The cell cycle, p53 signaling cascade, oocyte meiosis, progesterone-triggered oocyte maturation, cellular senescence, human T-cell leukemia virus 1 infection, and FoxO signaling cascade were all highly enriched in the PLK network (Figure 6).

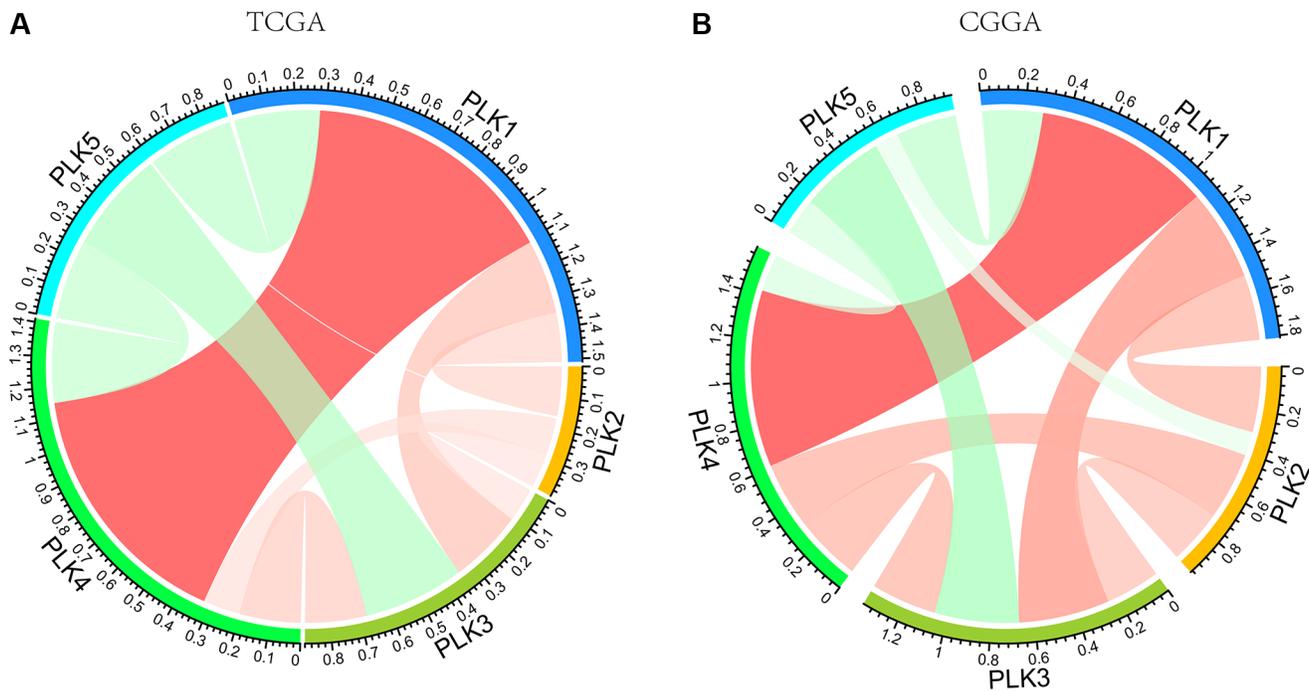
In addition, GSEA (Gene Set Enrichment Analysis) illustrated that the high expression of PLK1 was linked to the cell cycle, the low expression of PLK2 was linked to the NOTCH signaling cascade, the high expression of PLK3 was linked to apoptosis, the high expression of PLK4 was linked with the cell cycle, and the low

expression of PLK5 was linked to the NOTCH signaling cascade (Figure 7A). Finally, p53 signaling cascade was found to play a key role by overlapping KEGG cascades between PLKs' network and GSEA's results (Figure 7B).

**ON1231320 suppresses glioma cell proliferation and induces cleaved poly (ADP-ribose) polymerase (PARP) expression**

We found for the first time that ON1231320 dampened the growth of human glioma cells. The Cell Counting Kit-8 (CCK-8) assay was done after inoculating U251MG and U87MG cells with the specified levels of ON1231320 for 24, 48 and 72 h, respectively. The data illustrated that inoculation with 200 nm ON1231320 caused remarkable glioma cell growth dampening at 24, 48 h and 72 h (Figure 8A). The colony generation of U251MG and U87MG cells after 100 nM ON1231320 inoculation was remarkably lower in contrast with the controls, and the colony generation of glioma cells diminished with increasing ON1231320 levels (Figure 8B and 8C). The data illustrated that ON1231320 remarkably dampened the growth of glioma cells in a time- along with dose-dependent approach.

We carried out western blot assays to further demonstrate the manner in which ON1231320 enhances apoptosis in glioma cells. The data illustrated that the expression of cleaved PARP was elevated and the content of PLK2 was decreased in U251 and U87MG cells after ON1231320 treatment (Figure 9A and 9B).



**Figure 5.** PLK correlations with each other using TCGA (A) and CGGA (B) databases.

While the expression of PLK1/3/4/5 was not changed. Meanwhile ON1231320 promoted cleaved PARP expression and dampened the expression of PLK2 in a dose-dependent approach. These data suggest that ON1231320 shows its antitumor properties in glioma cells.

### ON1231320 inhibits glioma cell proliferation *in vivo*

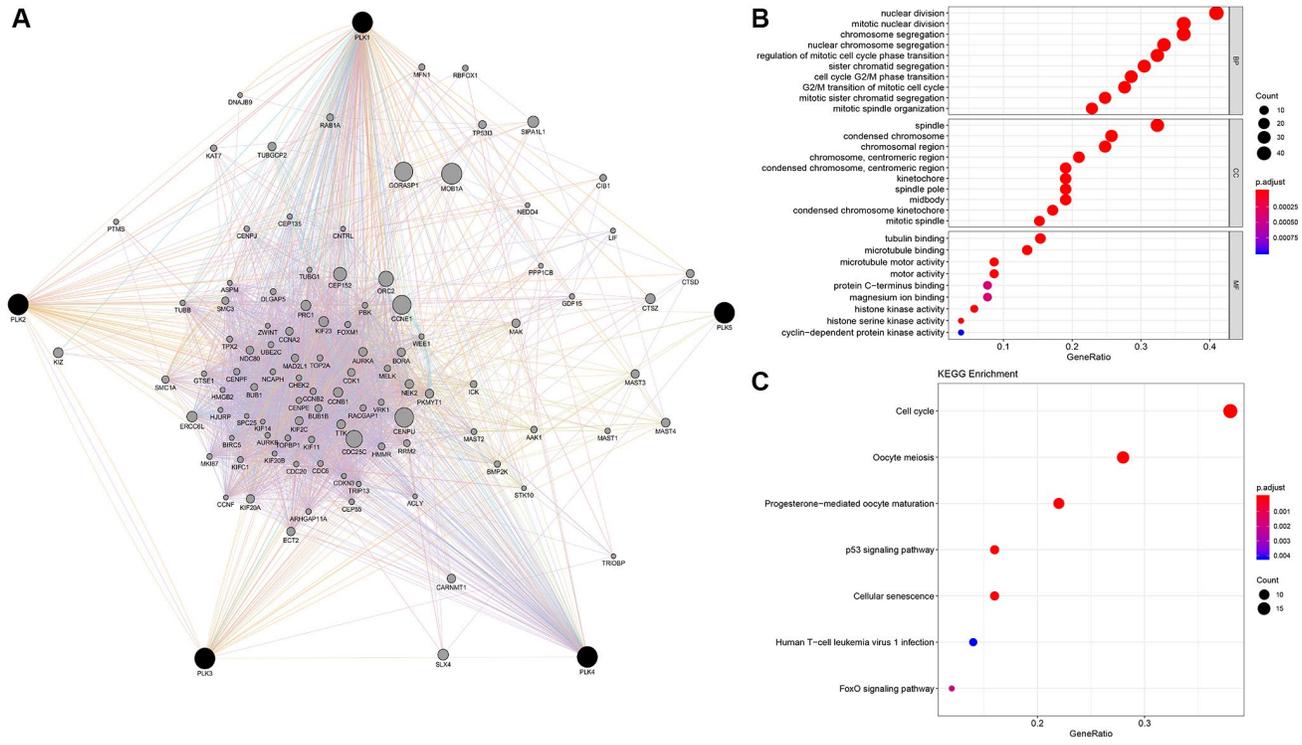
U87MG cells were subcutaneously inoculated into BALB/c Nude mice to test the effectiveness of ON1231320 *in vivo* utilizing a tumor xenograft model. ON1231320 was given daily (q.d.) through intraperitoneal inoculation at a dosage of 50 mg/kg body weight after the tumors had formed in the animals for two weeks. In contrast with the controls, ON1231320 inoculation resulted in a remarkable reduction in tumor growth (Figure 9C and 9D). ON1231320's high *in vivo* anticancer effectiveness in mice bodes well for its development for therapeutic usage.

## DISCUSSION

PLKs have been reported in common cranial brain tumors. This is the first study to explore the PLK expression along with prognosis in GBM, as well as the signaling cascades they may be involved in. It provides a diagnostic basis by contributing to existing studies on

glioblastoma-targeted therapies. In addition, this paper systematically evaluates the mechanism of the PLK family in GBM and provides an important foundation for individualized GBM treatment in the future.

PLK1 is currently one of the most intensively studied molecular markers in all kinds of tumors. PLK1 is overexpressed in various cancer, including breast cancer, non-small cell lung cancer, colorectal cancer, prostate cancer, pancreatic cancer, melanoma, ovarian cancer, non-Hodgkin's lymphomas, and acute myeloid leukemia (AML) [10]. PLK1 has also been increasingly reported in GBM. Lee et al. found that PLK1 mRNA was more highly expressed than non-tumorous human astrocytes (HAs) and identified PLK1 as a critical factor for the survival of brain cancer cells and brain tumor initiating cells (BTICs) [11]. PLK1 dampening blocked tumorsphere formation in all BTIC lines examined [11]. The present study confirmed that PLK1 is positively correlated with the malignant progression of gliomas. Likewise, Zhao et al. identified PLK1 as a key factor in the conversion of LGGs to secondary glioblastoma (sGBM) and observed that patients from the high-expression PLK1 group were more sensitive to common chemotherapies during clinical treatment. The findings of Pezuk et al. suggested that PLK1 may be a promising biomarker for the treatment of GBMs [12]. The latest research indicates that TMZ treatment activated PLK1-related signaling cascades and that PLK1 may be

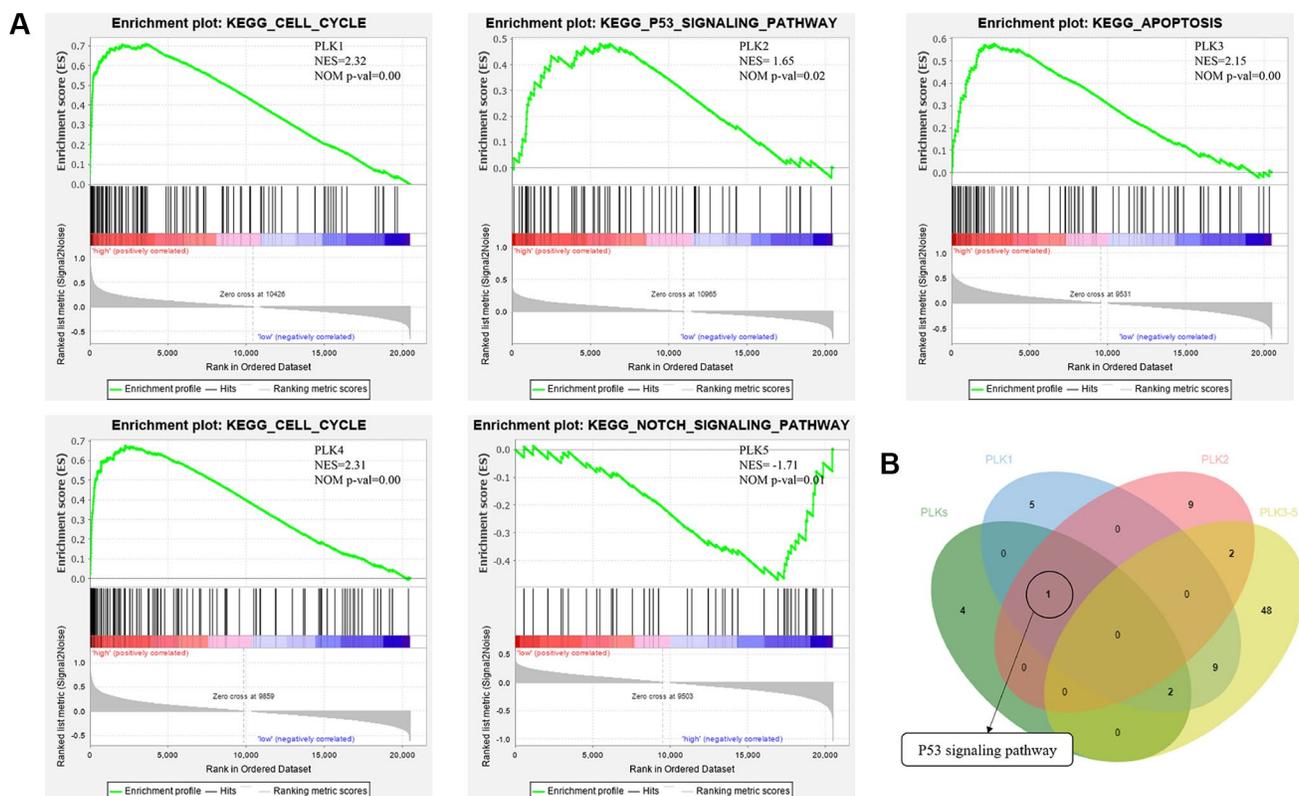


**Figure 6. PLK network and the functional enrichment of GO terms and KEGG pathways. (A)** The network of PLKs. **(B)** The functional enrichment of GO. **(C)** The functional enrichment of KEGG pathways.

involved in TMZ tolerance [12]. Current small molecule PLK1 inhibitors include Rigosertib, Volasertib, GSK 461364, GW843682, PLHS-Pmab, PPG, and Poloxin [10]. In addition, PLK1 may be a tumor recognition antigen for T lymphocytes, which has important implications in tumor immunotherapy [13].

The expression of PLK2 was found to decrease with CpG methylation [14]. PLK2 methylation in ovarian cancer observed to correlate with patients treated with chemotherapy [15]. In the present study, PLK2 was found to be remarkably reduced in GBM in contrast with non-tumorous tissues, which may result from PLK2 methylation in GBM. However, low PLK2 expression seems to suggest a good prognosis for patients. Similarly, in the study by Xia et al., the researchers found that PLK2 methylation resulted in its reduced expression in GBM patients [16]. Furthermore, low PLK2 expression has been found to correlate with better prognosis and its possible use as a prognostic biomarker. Most importantly, the present study found that PLK2 may work as an independent predictor in the PLK family and demonstrates promise as a novel molecular target. In univariate and multivariate analysis, high contents of PLK2 were remarkably linked to OS time. Furthermore, high expression of PLK2 had a poor prognosis of patients with GBM. Therefore, PLK2 can

be identified as an independent predictor in contrast with PLK1 in GBM. Current PLK2 inhibitors, ON1231320, has been demonstrated to be a selective repressor of PLK2, with no dampening activity against PLK1, PLK3 and PLK4. In contrast with the control group, PLK2 inhibitor treatment remarkably restrained tumor growth and there were no remarkable toxic effects in the treatment group [17]. ON1231210 is effective in different tumor cell lines, but not in non-tumorous human fibroblasts. The recent development of PLK2 ATP competitive inhibitors, has produces promising and specific molecules. Other examples of selective inhibitors of PLK2 have been also reported, thus demonstrating that efficient drug development strategy for the identification of PLK2 inhibitors have been reached by the scientific community [18]. This study is the first to use a PLK2 inhibitor in a glioma cell. Our results show that ON1231320 dampens GBM cell growth in a dose- along with time-dependent approach. Cleaved PARP is considered a marker of apoptosis [19]. The data illustrated that the expression of cleaved PARP was increased in U251MG and U87MG cells after ON1231320 treatment. Our data illustrated that ON1231320 dampened GBM cell growth in a dose- and time-dependent manner. *In vivo*, the inoculation with ON1231320 resulted in remarkable dampening of tumor growth in contrast with the control



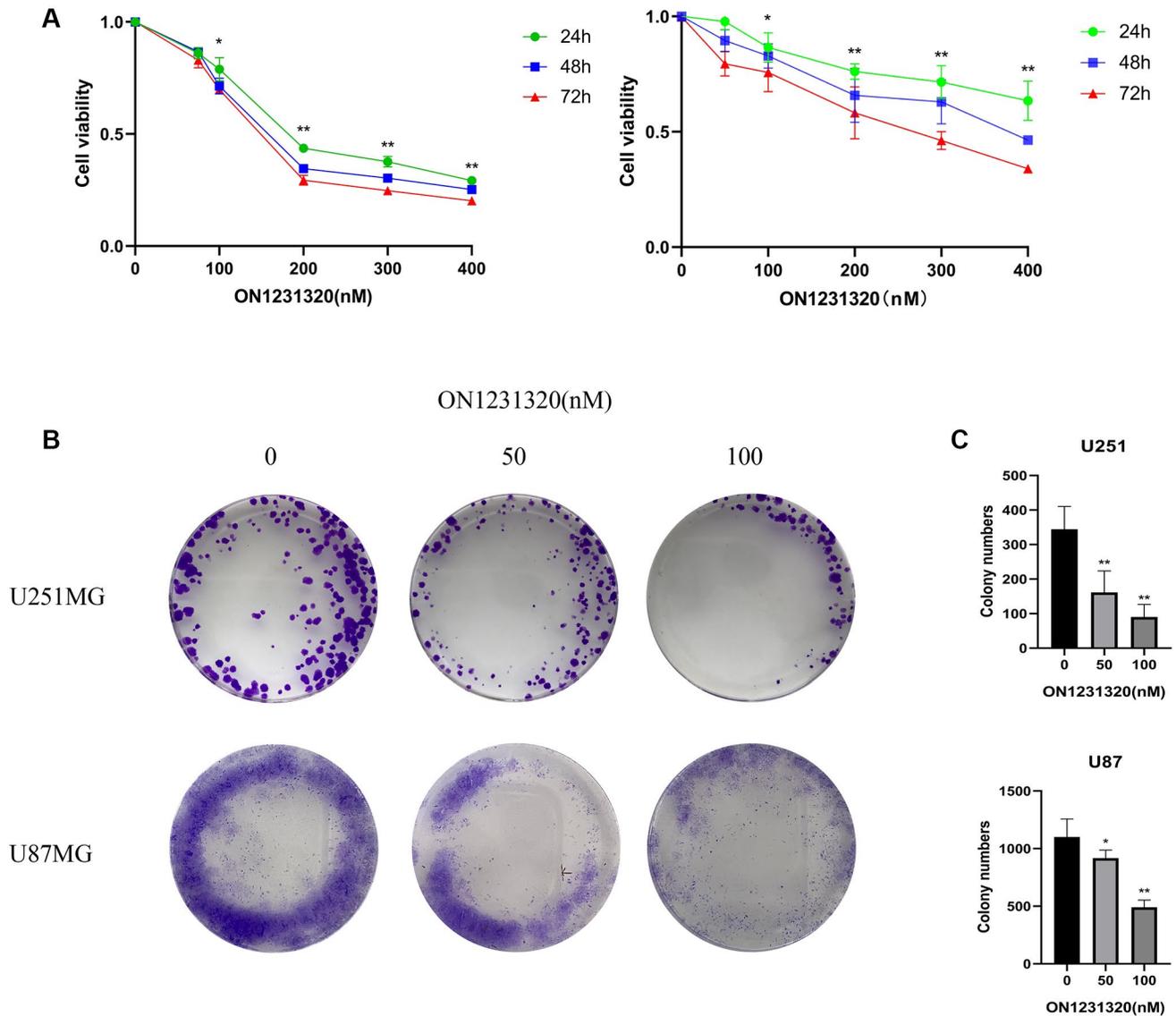
**Figure 7.** (A) Functional analysis of PLKs in TCGA cohort using GSEA. (B) Overlapping KEGG pathways between PLKs' network and GSEA's results.

group. And then, ON1231320 shows its antitumor properties in glioma cells.

PLK3 is often lowly expressed in lung, head and neck cancers [20]. PLK3 is poorly studied in gliomas and GBM. PLK3 was found to be differentially expressed between IDH mutations carriers and wild-type carriers and was linked to LGG patient survival [21]. PLK3 regulates the cell cycle and is also a mediator of apoptosis [22]. This observation was consistent with the GSEA results. This study showed that PLK3 is remarkably upregulated in GBM, but the biological role of PLK3 in GBM needs further validation. PLK2 and PLK3 recognize a very similar consensus sequence [23]. However, the low number of PLK2 and PLK3 substrates identified so far could not be sufficient to

highlight differences between two kinase. Moreover,  $\alpha$ -synuclein C-terminal peptide is phosphorylated *in vitro* by PLK2 with a higher efficiency than by PLK3, thus suggesting some differences in substrates preference. Nevertheless, it must be considered that a similar consensus sequence is not sufficient to hypothesize redundant cellular functions for these kinases, as kinase specificity is the results of multiple mechanisms. Different spatial and temporal expression/activation of the two kinases or specific substrate binding sites could indeed account for distinct PLK2 and PLK3 cellular functions [18].

SAK-a and SAK-b are two isoforms of PLK4. PLK4 increases during the G1/S transition, then persists until late M phase, and finally decreases in early G1 phase.



**Figure 8.** (A) The U251MG cells were treated under different concentrations of ON1231320 and cultured for 24, 48 and 72 h, respectively. (B) Colony formation assay showing the sensitizing effects on GBM cells after apcin treatment. (C) Quantitative results of Colony formation assay (\* $p < 0.05$  vs. the control group, \*\* $p < 0.01$  vs. the control group).

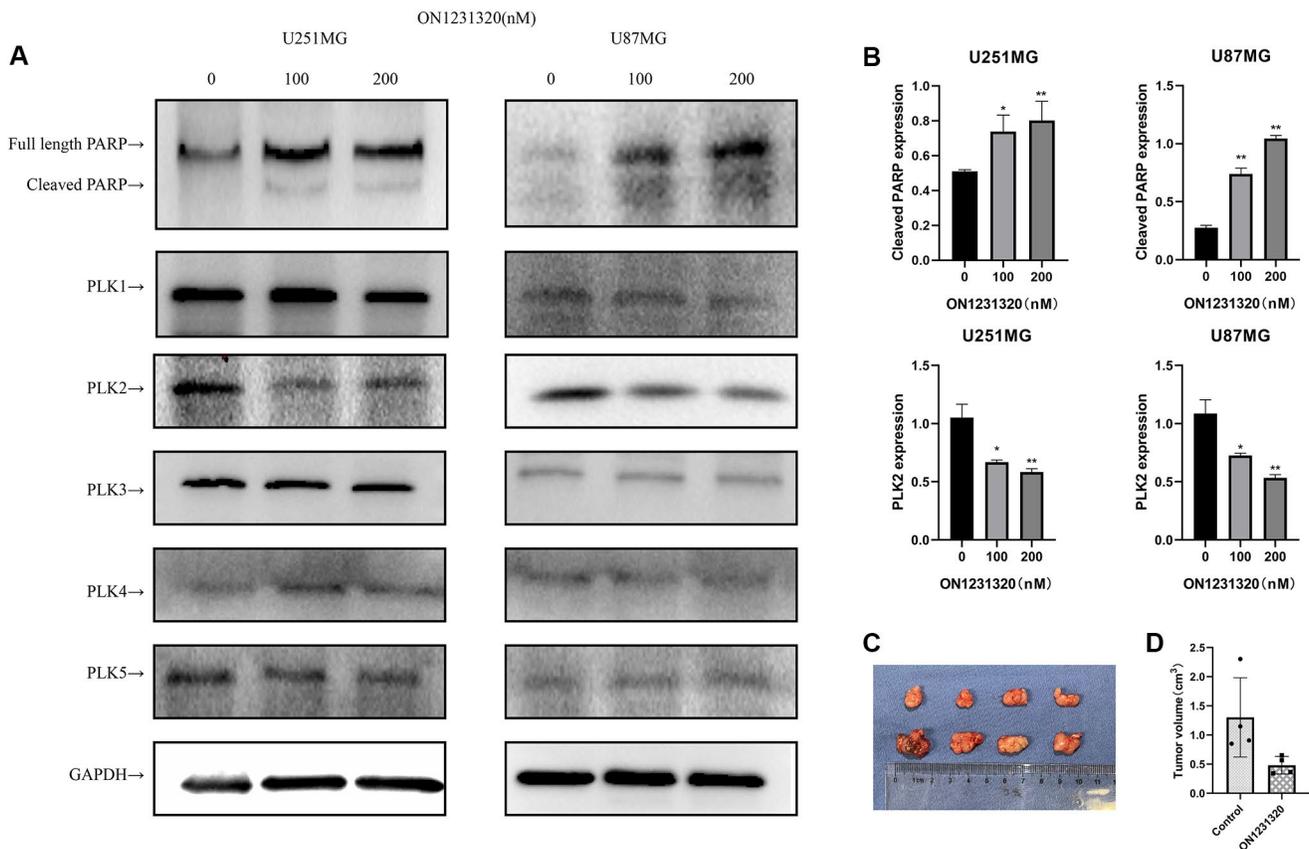
PLK4 is overexpressed in colon and breast cancers and downregulated in hepatocellular carcinoma [10].

Expression contents of PLK1 and PLK4 were elevated in contrast with non-tumorous mucosa [24]. This observation appears to be consistent with the PLK1/PLK4 co-occurrence described in Table 3. Wang et al. [25] observed elevated PLK4 expression in HGG patients, which was linked to poor prognosis. Zhang et al. discovered that the contents of PLK4 was remarkably correlated with glioma grade and inversely linked to overall survival of individuals with high-grade gliomas [26]. TMZ sensitivity was enhanced by depletion of PLK4 [26]. PLK4 inhibitor CFI400945 improves the sensitivity of GBM cells to TMZ [26].

PLK5 lacks a kinase domain in contrast with other PLK family members and therefore has no catalytic activity [10]. Interestingly, PLK5 is expressed primarily in the brains of humans. PLK5 is frequently silenced in astrocytoma and glioblastoma due to hypermethylation [27]. PLK5 in this study appeared to be potentially correlated with other PLK family members. Indeed, overexpression of PLK5 promotes tumor cell apoptosis. Thus, PLK5 is a kinase-deficient polo box domain-

containing protein with exclusive neurological function and brain tumor suppressor activity [27].

Structurally, the general structure of PLK1-4 includes an N-terminal serine/threonine kinase domain along with a C-terminal polo-box domain (PBD), but the domain truncated in PLK5, without the T-loop [28]. The C-terminus contains two PBDs in Plk1-3 and Plk5, while there is only one PBD in PLK4 [29]. In Figure 5, there is a high positive correlation between PLK1 and PLK4. This may be due to the structural similarity between PLK1 and PLK4, which leads to their functional consistency. Pearson correlation analysis and mutual exclusivity analysis showed a remarkable positive correlation between PLK1 and PLK4. Zeng et al. also found the same conclusion in human non-small cell lung cancer [30]. The cross talk of PLK1 with PLK4 needs to be further demonstrated. PLK1 and PLK4 expression levels are higher in leukemia cells than in non-tumorous cells and have promise as being new targets for cancer therapy [28]. Therefore, theoretically combined dampening of PLK1 and PLK4 seems to be an effective treatment of GBM. It has been shown that the ATP-binding site of kinases is a target for the design of inhibitors that can be used to inhibit



**Figure 9.** (A) The expression of cleaved PARP and PLKs in ON1231320 treated cells was detected using western blotting. (B) Quantitative results of immunoblots. (C) ON1231320 inhibits tumor growth *in vivo*. (D) Quantitative results of tumor volumes. \* $P < 0.05$  and \*\* $P < 0.001$  vs. the control groups.

kinase. The kinase domain of PLK1 is similar in sequence to PLK2, PLK3 and PLK4, the development of a specific, ATP-competitive PLK1 inhibitor remains challenging. It is therefore conceivable that inhibitors of PLK1, scarcely discriminate between PLK1–PLK3. Unfortunately, PLK4 is not dampened by inhibitors of PLK1, and it is possible that PLK4 is very different in contrast with other PLK family members in terms of structure [7]. There is a need to further understand which gene is dominant in this correlation, to explore whether there is a common signaling cascade between PLK1 and PLK4, and to suppress the high expression of both PLK1 and PLK4 by inhibiting a target gene. Considering the expression of PLK4 in GBM, it may be possible to control GBM progression more effectively if PLK1-4 can be dampened simultaneously. The alteration of PLKs in GBM was mainly expressed in amplified form (Figure 3), which explains that PLKs are still mostly upregulated with increasing glioma grade. Furthermore, PLK alterations usually occur in primary GBM, suggesting an important role for PLKs in primary GBM.

P53 plays an indispensable role in transcriptional regulation of BP, such as DNA repair, cell cycle arrest, senescence, and apoptosis [31]. PLKs-associated KEGG cascade analysis also shows correlation with p53 signaling cascade. This means that the p53 signaling cascade could be the key signaling cascade in PLKs network. In addition, PLK2 may contribute to the poor prognosis of glioma patients through other cascades.

Herein, our results suggest that the high expression levels of PLK1/3/4 and low expression levels of PLK2/5 play important roles in the development of GBM. PLK family can be an important molecular marker to identify the malignancy of GBM. Furthermore, our findings suggest that PLKs are potential therapeutic targets for GBM. In particular, low PLK2 expression presents a potential prognostic marker for improving GBM patient survival and prognostic accuracy.

## **MATERIALS AND METHODS**

### **ONCOMINE analysis**

PLK transcription levels in various malignancies were analyzed using gene expression array datasets from ONCOMINE, an open cancer microarray data repository. The tumor analyses were in contrast with the non-tumorous tissue of the same kind. The PLK transcript contents in clinical cancer specimens were in contrast with those in non-tumorous controls using the student's *t*-test, which yielded a *p*-value. 0.01 and 2 were chosen as the *p*-value cut-off and fold change, respectively.

### **Gene expression profiling interactive analysis 2(GEPIA2) dataset**

GEPIA2 is a newly established interactive web resource that employs a standard processing pipeline to explore the RNA sequencing expression data of 9,736 tumors along with 8,587 non-tumorous samples from the TCGA, as well as GTEx projects (<http://GEPIA2.cancer-pku.cn/>) [10]. For tumor/non-tumorous differential expression and correlation analysis, GEPIA2 was employed. The approach for estimating differential expression constitutes one-way ANOVA, with disease status (tumorous or non-tumorous) as the variable.

### **cBioPortal**

The cBio Cancer Genomics Portal (<http://cbioportal.org>) was created primarily to resolve the data integration challenges that large-scale cancer genomics research suffer [32]. We identified somatic mutations, transcript expression changes, along with copy number changes in a set of cases using OncoPrints. Based on the cancer types summary, various PLK alteration forms were observed in GMB, including mRNA low, mutation amplification, mRNA high, deep deletion, and multiple alterations. The difference between the altered and unaltered groups in the subtypes was determined via clinical comparison. To understand the mutual exclusivity between PLKs, the samples with a variation on a gene were used as a set to analyze whether two genes were mutually exclusive or co-occurring in a tumor by analyzing the sets corresponding to the two genes. A positive value here suggests that alterations in these genes co-occur in the same samples, while a negative value suggests that alterations in these genes are mutually exclusive and occur in different samples.

### **TCGA data**

We acquired 699 samples from TCGA, which included gender, WHO grade, age, and pathologic diagnosis. The molecular analysis results were obtained using the Genomic Data Commons (GDC) [33].

### **CGGA data**

The CGGA data resource (<http://www.cgga.org.cn/>) is a user-friendly web tool for data storage and analysis that allows you to explore brain tumor data sets consisting of over 2,000 samples from Chinese cohorts [34]. This study utilized CGGA tools for grade-related differential expression, and correlation analyses. In addition, 693 samples were downloaded for mRNA sequencing.

## HPA data

The Human Protein Atlas data resource (<https://www.proteinatlas.org/>) involves mapping all human proteins in cells, tissues, along with organs by incorporating multiple omics approaches, for instance antibody-centered imaging, mass spectrometry-centered proteomics, transcriptomics, as well as systems biology [35]. The HPA data set was adopted to verify PLKs at the translational level. The score of protein expression was calculated via immuno-histochemistry data that was manually rated for staining intensity (negative, weak, moderate, or strong) and the proportion of stained cells (25%, 25–75%, or >75%).

## Functional enrichment analyses

GO along with KEGG analyses were also adopted to assess the functional roles of the PLK network. The cluster Profiler program in R statistical software was adopted to conduct these gene functional enrichment analyses (The R Foundation, Vienna, Austria).  $P < 0.05$  served as the cutoff for GO along with KEGG enrichment analyses. The findings of the GO along with KEGG assessments were visualized using the R tool GO plot.

## GSEA

GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) constitutes a statistical approach for determining if an a priori defined list of genes exhibits remarkably different, concordant differences between two biological states. The GSEA v4.0 software was adopted to assess the normalized enrichment score (NES) along with false discovery rate (FDR) to validate the remarkable changes [36].

## Cox proportional hazards regression analysis

Univariate along with multivariate Cox proportional hazards regression assessments were adopted to explore predictive clinico-pathological variables. The hazard ratio for each category and its 95% CI are utilized to express the results for noteworthy prognostic indicators.  $P < 0.05$  was regarded as statistically significant. R statistical program was used to conduct the statistical analysis.

## GeneMANIA in Cytoscape

GeneMANIA utilizes a vast database of functional relationship data to uncover other genes that are connected to a set of input genes. Protein along with genetic relationships, cascades, co-expression, co-localization, as well as protein domain similarity are all examples of association data [16]. GeneMANIA may be

accessed using the Cytoscape app. Using GeneMANIA, we created a network of 100 genes connected to PLKs.

## Cell culture

The Chinese Academy of Sciences provided U251MG (Shanghai, China). The cells were propagated in DMEM (Gibco; Thermo Fisher Scientific, USA) enriched with 10% FBS (Gibco; Thermo Fisher Scientific, USA) under 37°C along with 5% CO<sub>2</sub> conditions.

## Cell viability assay

The CCK8 kit was adopted to assess cell viability. To begin, 5000 cells/well were inoculated in 96-well plates. The cells were inoculated with diverse dosages of ON1231320 and propagated for 24, 48, and 72 hours after an overnight incubation. After the inoculation, each well was introduced with 100L DMEM and 10L CCK-8 and incubated for two hours. Finally, utilizing an Infinite M200 PRO plate reader, the absorbance was taken at 450 nm (Tecan, Switzerland).

## Colony forming cell assay

On 6-well culture plates, 500 cells/well were planted and grown in DMEM enriched with 10% FBS. The cells were inoculated with the appropriate agents and grown at 37°C along with 5% CO<sub>2</sub> conditions for ten days. The colonies were then enumerated after being stained with 0.1 percent crystal violet. Three independent tests were performed on each group of colonies. ImageJ software was adopted to evaluate the data.

## Western blot assay

Cell lysing was done with the RIPA buffer (APPLYGEN, Beijing, China) enriched with a cocktail of protease inhibitor. Fractionation of proteins was done with a suitable dosage of gel (BioSci: 8012011), followed by blotting onto nitrocellulose membranes. After that, 3% BSA in TBST was adopted to block the membranes for one hour at RT (room temperature). Thereafter, we overnight inoculated the samples with the primary antibody of the matching antigen at 4°C, and then rinsed thrice in TBST. Afterwards, we inoculated the samples for one hour with the fluorescently linked secondary antibody at RT. Using the Odyssey infrared imaging equipment (LI-COR, Lincoln, NE), distinct protein bands were identified after three rinses in TBST. For sample loading and standardization, GAPDH was employed as an internal control.

## Immunohistochemistry

PLKs were stained immuno-histochemically on 5-μm unstained tissue slices, which were subsequently

deparaffinized, then re-hydrated. Tissue slices were treated for 30 minutes at 100°C in a steamer harboring 10 mM citrate buffer (pH 6.0) to assess antigenicity. To suppress activity of endogenous peroxidase, the segments were sub-merged in ethanol enriched with 3% hydrogen peroxidase for 20 minutes. Sections were inoculated overnight at 4°C with a 1:100 dilution of primary anti-PLKs (Santa Cruz, CA, USA). The tissue slices were then rinsed in PBS and inoculated for 30 minutes with an anti-rabbit secondary antibody before being inoculated with the streptavidin horseradish peroxidase complex. DAB was used to develop the sections, which were then counterstained with hematoxylin. At 400 magnifications, 15–20 fields of the sections were examined. Four grades categorized the fraction of positively-stained tumor cells: >30% positive tumor cells (3); 10–30% positive tumor cells (2); <10% positive tumor cells (1); no positive tumor cells (0). Four grades categorized the intensity of staining: no staining (0), light-yellow-weak staining (1); yellowish brown-moderate staining (2), and brown-strong staining (3). The staining index was computed as the fraction of positively stained tumor cells multiplied by the intensity of staining. High PLKs expression was described as a staining index score >4, whilst low expression was characterized as a staining index score 2, and mid expression was defined as others score.

### Nude mouse xenograft

$5 \times 10^6$  U87 cells were subcutaneously inoculated into the right axillae of 6 weeks old BALB/c Nude female mice. Once the tumors have grown in the mice for 2 weeks, we stratified the mice at random into treatment groups ( $n = 4$ ) and were given daily, either placebo, or the specified dosages of ON1231320. The formula ( $\text{long} \times \text{short}^2$ )/2 was employed to compute tumor volumes.

### Statistical analysis

For all statistical analyses, SPSS v20.0 was utilized. The unpaired Student's *t*-test was adopted to uncover differentially expressed genes. For the statistical examination of the correlation between two independent variables, the 2 test was used. The survival distributions were estimated using Kaplan-Meier survival analysis. Using GraphPad Prism, the log-rank approach was adopted to assess the statistical significance of the stratified survival groups.

### Data accessibility

These data were derived from the following resources available in the public domain: TCGA database (<https://portal.gdc.cancer.gov/>) and CGGA database (<http://www.cgga.org.cn/>).

### Author contributions

YM Ding and HJ Liu conducted experiments, performed the data analyses and wrote the manuscript. ZS Bao and CB Zhang contributed significantly in manuscript revision. SQ Yu conceived and designed the study. All authors read and approved the final manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

### ACKNOWLEDGMENTS

We would like to thank Siyuan Chen, Lei He, Xinyu Song for their assistance.

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

### FUNDING

The study was supported by General Technological Program of Peking Education Commission (KM201810025023).

### REFERENCES

1. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, Villano JL. Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiol Biomarkers Prev.* 2014; 23:1985–96. <https://doi.org/10.1158/1055-9965.EPI-14-0275> PMID:25053711
2. Di Carlo DT, Cagnazzo F, Benedetto N, Morganti R, Perrini P. Multiple high-grade gliomas: epidemiology, management, and outcome. A systematic review and meta-analysis. *Neurosurg Rev.* 2019; 42:263–75. <https://doi.org/10.1007/s10143-017-0928-7> PMID:29138949
3. Lee E, Yong RL, Paddison P, Zhu J. Comparison of glioblastoma (GBM) molecular classification methods. *Semin Cancer Biol.* 2018; 53:201–11. <https://doi.org/10.1016/j.semcancer.2018.07.006> PMID:30031763
4. Llamazares S, Moreira A, Tavares A, Girdham C, Spruce BA, Gonzalez C, Karess RE, Glover DM, Sunkel CE. polo encodes a protein kinase homolog required for mitosis in *Drosophila*. *Genes Dev.* 1991; 5:2153–65. <https://doi.org/10.1101/gad.5.12a.2153> PMID:1660828

5. Barr FA, Silljé HH, Nigg EA. Polo-like kinases and the orchestration of cell division. *Nat Rev Mol Cell Biol.* 2004; 5:429–40.  
<https://doi.org/10.1038/nrm1401>  
PMID:[15173822](https://pubmed.ncbi.nlm.nih.gov/15173822/)
6. Zitouni S, Nabais C, Jana SC, Guerrero A, Bettencourt-Dias M. Polo-like kinases: structural variations lead to multiple functions. *Nat Rev Mol Cell Biol.* 2014; 15:433–52.  
<https://doi.org/10.1038/nrm3819>  
PMID:[24954208](https://pubmed.ncbi.nlm.nih.gov/24954208/)
7. Strebhardt K. Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy. *Nat Rev Drug Discov.* 2010; 9:643–60.  
<https://doi.org/10.1038/nrd3184>  
PMID:[20671765](https://pubmed.ncbi.nlm.nih.gov/20671765/)
8. Xie S, Wu H, Wang Q, Cogswell JP, Husain I, Conn C, Stambrook P, Jhanwar-Uniyal M, Dai W. PLK3 functionally links DNA damage to cell cycle arrest and apoptosis at least in part via the p53 pathway. *J Biol Chem.* 2001; 276:43305–12.  
<https://doi.org/10.1074/jbc.M106050200>  
PMID:[11551930](https://pubmed.ncbi.nlm.nih.gov/11551930/)
9. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer.* 2009; 9:153–66.  
<https://doi.org/10.1038/nrc2602>  
PMID:[19238148](https://pubmed.ncbi.nlm.nih.gov/19238148/)
10. Liu X. Targeting Polo-Like Kinases: A Promising Therapeutic Approach for Cancer Treatment. *Transl Oncol.* 2015; 8:185–95.  
<https://doi.org/10.1016/j.tranon.2015.03.010>  
PMID:[26055176](https://pubmed.ncbi.nlm.nih.gov/26055176/)
11. Lee C, Fotovati A, Triscott J, Chen J, Venugopal C, Singhal A, Dunham C, Kerr JM, Verreault M, Yip S, Wakimoto H, Jones C, Jayanthan A, et al. Polo-like kinase 1 inhibition kills glioblastoma multiforme brain tumor cells in part through loss of SOX2 and delays tumor progression in mice. *Stem Cells.* 2012; 30:1064–75.  
<https://doi.org/10.1002/stem.1081>  
PMID:[22415968](https://pubmed.ncbi.nlm.nih.gov/22415968/)
12. Liu N, Hu G, Wang H, Li Z, Guo Z. PLK1 inhibitor facilitates the suppressing effect of temozolomide on human brain glioma stem cells. *J Cell Mol Med.* 2018; 22:5300–10.  
<https://doi.org/10.1111/jcmm.13793>  
PMID:[30133120](https://pubmed.ncbi.nlm.nih.gov/30133120/)
13. Park JS, Sohn HJ, Park GS, Chung YJ, Kim TG. Induction of antitumor immunity using dendritic cells electroporated with Polo-like kinase 1 (Plk1) mRNA in murine tumor models. *Cancer Sci.* 2011; 102:1448–54.  
<https://doi.org/10.1111/j.1349-7006.2011.01974.x>  
PMID:[21545375](https://pubmed.ncbi.nlm.nih.gov/21545375/)
14. Syed N, Smith P, Sullivan A, Spender LC, Dyer M, Karran L, O'Nions J, Allday M, Hoffmann I, Crawford D, Griffin B, Farrell PJ, Crook T. Transcriptional silencing of Polo-like kinase 2 (SNK/PLK2) is a frequent event in B-cell malignancies. *Blood.* 2006; 107:250–6.  
<https://doi.org/10.1182/blood-2005-03-1194>  
PMID:[16160013](https://pubmed.ncbi.nlm.nih.gov/16160013/)
15. Coley HM, Hatzimichael E, Blagden S, McNeish I, Thompson A, Crook T, Syed N. Polo Like Kinase 2 Tumour Suppressor and cancer biomarker: new perspectives on drug sensitivity/resistance in ovarian cancer. *Oncotarget.* 2012; 3:78–83.  
<https://doi.org/10.18632/oncotarget.332>  
PMID:[22289679](https://pubmed.ncbi.nlm.nih.gov/22289679/)
16. Xia X, Cao F, Yuan X, Zhang Q, Chen W, Yu Y, Xiao H, Han C, Yao S. Low expression or hypermethylation of PLK2 might predict favorable prognosis for patients with glioblastoma multiforme. *PeerJ.* 2019; 7:e7974.  
<https://doi.org/10.7717/peerj.7974>  
PMID:[31763067](https://pubmed.ncbi.nlm.nih.gov/31763067/)
17. Reddy MV, Akula B, Jatiani S, Vasquez-Del Carpio R, Billa VK, Mallireddigari MR, Cosenza SC, Venkata Subbaiah DR, Bharathi EV, Pallela VR, Ramkumar P, Jain R, Aggarwal AK, Reddy EP. Discovery of 2-(1H-indol-5-ylamino)-6-(2,4-difluorophenylsulfonyl)-8-methylpyrido[2,3-d]pyrimidin-7(8H)-one (7ao) as a potent selective inhibitor of Polo like kinase 2 (PLK2). *Bioorg Med Chem.* 2016; 24:521–44.  
<https://doi.org/10.1016/j.bmc.2015.11.045>  
PMID:[26762835](https://pubmed.ncbi.nlm.nih.gov/26762835/)
18. Cozza G, Salvi M. The Acidophilic Kinases PLK2 and PLK3: Structure, Substrate Targeting and Inhibition. *Curr Protein Pept Sci.* 2018; 19:728–45.  
<https://doi.org/10.2174/1389203719666180124095405>  
PMID:[29366414](https://pubmed.ncbi.nlm.nih.gov/29366414/)
19. Agarwal A, Mahfouz RZ, Sharma RK, Sarkar O, Mangrola D, Mathur PP. Potential biological role of poly (ADP-ribose) polymerase (PARP) in male gametes. *Reprod Biol Endocrinol.* 2009; 7:143.  
<https://doi.org/10.1186/1477-7827-7-143>  
PMID:[19961617](https://pubmed.ncbi.nlm.nih.gov/19961617/)
20. Dai W, Li Y, Ouyang B, Pan H, Reissmann P, Li J, Wiest J, Stambrook P, Gluckman JL, Noffsinger A, Bejarano P. PRK, a cell cycle gene localized to 8p21, is downregulated in head and neck cancer. *Genes Chromosomes Cancer.* 2000; 27:332–6.  
[https://doi.org/10.1002/\(sici\)1098-2264\(200003\)27:3<332::aid-gcc15>3.0.co;2-k](https://doi.org/10.1002/(sici)1098-2264(200003)27:3<332::aid-gcc15>3.0.co;2-k)  
PMID:[10679924](https://pubmed.ncbi.nlm.nih.gov/10679924/)

21. Pang FM, Yan H, Mo JL, Li D, Chen Y, Zhang L, Liu ZQ, Zhou HH, Wu J, Li X. Integrative analyses identify a DNA damage repair gene signature for prognosis prediction in lower grade gliomas. *Future Oncol.* 2020; 16:367–82. <https://doi.org/10.2217/fo-2019-0764> PMID:32065545
22. Helmke C, Becker S, Strebhardt K. The role of Plk3 in oncogenesis. *Oncogene.* 2016; 35:135–47. <https://doi.org/10.1038/onc.2015.105> PMID:25915845
23. Johnson EF, Stewart KD, Woods KW, Giranda VL, Luo Y. Pharmacological and functional comparison of the polo-like kinase family: insight into inhibitor and substrate specificity. *Biochemistry.* 2007; 46:9551–63. <https://doi.org/10.1021/bi7008745> PMID:17655330
24. Macmillan JC, Hudson JW, Bull S, Dennis JW, Swallow CJ. Comparative expression of the mitotic regulators SAK and PLK in colorectal cancer. *Ann Surg Oncol.* 2001; 8:729–40. <https://doi.org/10.1007/s10434-001-0729-6> PMID:11597015
25. Wang J, Zuo J, Wang M, Ma X, Gao K, Bai X, Wang N, Xie W, Liu H. Polo-like kinase 4 promotes tumorigenesis and induces resistance to radiotherapy in glioblastoma. *Oncol Rep.* 2019; 41:2159–67. <https://doi.org/10.3892/or.2019.7012> PMID:30816483
26. Zhang Z, Wang Z, Huang K, Liu Y, Wei C, Zhou J, Zhang W, Wang Q, Liang H, Zhang A, Wang G, Zhen Y, Han L. PLK4 is a determinant of temozolomide sensitivity through phosphorylation of IKBKE in glioblastoma. *Cancer Lett.* 2019; 443:91–107. <https://doi.org/10.1016/j.canlet.2018.11.034> PMID:30529153
27. de Cárcer G, Escobar B, Higuero AM, García L, Ansón A, Pérez G, Mollejo M, Manning G, Meléndez B, Abad-Rodríguez J, Malumbres M. Plk5, a polo box domain-only protein with specific roles in neuron differentiation and glioblastoma suppression. *Mol Cell Biol.* 2011; 31:1225–39. <https://doi.org/10.1128/MCB.00607-10> PMID:21245385
28. Goroshchuk O, Kolosenko I, Vidarsdottir L, Azimi A, Palm-Apergi C. Polo-like kinases and acute leukemia. *Oncogene.* 2019; 38:1–16. <https://doi.org/10.1038/s41388-018-0443-5> PMID:30104712
29. Jang YJ, Lin CY, Ma S, Erikson RL. Functional studies on the role of the C-terminal domain of mammalian polo-like kinase. *Proc Natl Acad Sci U S A.* 2002; 99:1984–9. <https://doi.org/10.1073/pnas.042689299> PMID:11854496
30. Zeng Y, Li N, Liu W, Zeng M, Cheng J, Huang J. Analyses of expressions and prognostic values of Polo-like kinases in non-small cell lung cancer. *J Cancer Res Clin Oncol.* 2020; 146:2447–60. <https://doi.org/10.1007/s00432-020-03288-6> PMID:32627077
31. Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol.* 2008; 9:402–12. <https://doi.org/10.1038/nrm2395> PMID:18431400
32. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; 2:401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID:22588877
33. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, Staudt LM. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med.* 2016; 375:1109–12. <https://doi.org/10.1056/NEJMp1607591> PMID:27653561
34. Wang Y, Qian T, You G, Peng X, Chen C, You Y, Yao K, Wu C, Ma J, Sha Z, Wang S, Jiang T. Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping. *Neuro Oncol.* 2015; 17:282–8. <https://doi.org/10.1093/neuonc/nou130> PMID:25031032
35. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015; 347:1260419. <https://doi.org/10.1126/science.1260419> PMID:25613900
36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102:15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID:16199517

# Centrosomal protein 290 is a novel prognostic indicator that modulates liver cancer cell ferroptosis via the *Nrf2* pathway

Yiru Shan<sup>1,\*</sup>, Guang Yang<sup>2,\*</sup>, Qihong Lu<sup>3,\*</sup>, Xiangyu Hu<sup>4,\*</sup>, Dongwei Qi<sup>4,\*</sup>, Yehan Zhou<sup>5</sup>, Yin Xiao<sup>1</sup>, Li Cao<sup>6,&</sup>, Fuhua Tian<sup>1,&</sup>, Qi Pan<sup>4</sup>

<sup>1</sup>Department of Oncology, Jiulongpo People's Hospital of Chongqing, Chongqing, P.R. China

<sup>2</sup>Department of Urology Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P.R. China

<sup>3</sup>Department of Orthopaedics, Jiulongpo People's Hospital of Chongqing, Chongqing, P.R. China

<sup>4</sup>Department of Dermatology, Chongqing Hospital of Traditional Chinese Medicine, Chongqing, P.R. China

<sup>5</sup>Department of Pathology, Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, P.R. China

<sup>6</sup>Department of Patient Service Center, Jiulongpo People's Hospital of Chongqing, Chongqing, P.R. China

\*Equal contribution

**Correspondence to:** Li Cao, Fuhua Tian, Qi Pan; email: [1364819149@qq.com](mailto:1364819149@qq.com), <https://orcid.org/0000-0002-2301-2472>; [2638860139@qq.com](mailto:2638860139@qq.com), <https://orcid.org/0000-0002-8637-6951>; [qi\\_pan@vip.sina.com](mailto:qi_pan@vip.sina.com), <https://orcid.org/0000-0002-9498-1270>

**Keywords:** ferroptosis, hepatocellular carcinoma, CEP290, Nrf2 signaling pathway

**Received:** November 24, 2021

**Accepted:** March 1, 2022

**Published:** March 10, 2022

**Copyright:** © 2022 Shan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Ferroptosis is an iron-dependent form of cell death. In spite of its significance in pathogenesis and disease progression, ferroptotic signal transduction in HBV-HCC has not been fully explained. Here, four HCC open-source datasets were downloaded from the GEO repository. Cox regression and LASSO models were established to prioritize novel prognostic candidate biomarkers, and the results were verified *in vitro* and *in vivo*. We identified 633 common DEGs in both of the bulk RNA-Seq expression profiles. Next, based upon the TCGA-LIHC cohort, a prognostic signature consisting of nine genes was extracted from 633 shared DEGs, and the specificity and sensitivity of the signature were evaluated in both training and validation datasets. This signature showed that the high-risk group had a worse prognosis than the low-risk group. CEP290 was discovered among the prognostic signature genes, and its expression notably correlated with survival, AFP level, TNM stage and vascular invasion. We confirmed expression of CEP290 in eight pairs of HCC tissues and diverse liver cancer cell lines. CEP290 knockdown reduced proliferation, migration and invasion in Hep3B liver cancer cells while Fe<sup>2+</sup> and malondialdehyde levels were elevated. Mechanically, co-immunoprecipitation showed an interaction between CEP290 and Nrf2 proteins, and biological phenotypes of Hep3B cells under CEP290 interference were rescued by Nrf2 activator. Furthermore, CEP290 silencing considerably blocked protein expression of Nrf2 pathway members. Finally, suppression of CEP290 effectively inhibited tumor growth *in vivo*. The above results shed light on the important role of CEP290 in ferroptosis and present an important implication for HCC progression.

## INTRODUCTION

Hepatocellular carcinoma (HCC) is the 4<sup>th</sup> most common cause of cancer-associated mortality worldwide [1], and

hepatitis B virus (HBV) infection accounts for the majority of cases [2]. In developing countries such as China, the HCC incidence rate has been increasing continuously [3, 4]. At present, accurate biomarkers for

HCC prognosis prediction have not been identified. Therefore, identifying predictive biomarkers is critical for the assessing the HCC survival rate [5, 6].

Great progress has been made in the analysis of transcriptomes with high-throughput sequencing technologies, and these studies have revealed detailed gene expression patterns in diverse cancers [7]. Bioinformatic analysis of gene expression profiles facilitates biomarker identification. For example, centrosomal protein 131 (*CEP131*) was recognized as a novel substrate of *PLK4* that facilitated centrosome amplification and colon cancer development [8]; overexpression of centrosomal protein 70 (*CEP70*) stimulated the growth of pancreatic cancer cells by inducing abnormal centrosomes and disorganized microtubules [9]. These studies revealed that centrosomal proteins could be potential biomarkers and therapeutic targets and played critical roles in diverse cancers.

Ferroptosis is a form of iron-dependent cell death, which is different from autophagy, necrosis and apoptosis [10]. Due to iron-dependent accumulation of lipid peroxidation, ferroptosis is regulated by a specific set of proteins involved in many cancer signaling pathways related to iron metabolism [11, 12]. *Nrf2*, a core gene for the oxidant stress response, induces the activation of its downstream genes that function against tumor ferroptosis [13]. Moreover, ferroptosis can also be triggered by the aberration of the glutathione (GSH)-glutathione peroxidase 4 (*GPX4*) antioxidant systems, indicating that *GPX4* is the vital link of antiperoxidant defense [14, 15]. However, the potential ferroptosis-related regulatory mechanism remains unknown.

In the present work, we demonstrate that centrosomal protein 290 (*CEP290*), which regulates cancer cell ferroptosis, growth, migration and invasion through the *Nrf2* signaling pathway, is a reliable biomarker for prognosis prediction. Bioinformatics analysis in combination with *in vivo* and *in vitro* experiments represents a more effective approach for exploring the molecular mechanism of HCC.

## MATERIALS AND METHODS

### HCC cases

For this study, we acquired eight HCC specimens and paired normal liver specimens from HBV-associated HCC patients who underwent surgical treatment at Sichuan Cancer Hospital. The Research Ethics Committee of Sichuan Cancer Hospital approved the research design. All patients signed informed consent for participation. After collection, each specimen was

frozen immediately in liquid nitrogen until processed for subsequent assays.

### HCC RNA-Seq dataset acquisition

To identify differentially expressed genes (DEGs) between HBV-associated HCC specimens and non-carcinoma liver specimens (NL), this work made use of two open-source datasets, GSE22058 (n=192) [16] and GSE54238 (n=23) [17], from the GEO database using GPL6793 together with the GPL16955 platform. One 10×Genomics single-cell RNA sequencing dataset (GSE103867) including resected tissue from a primary HCC patient and HBV-negative liver cancer Huh1 and Huh7 cell lines [18] was downloaded from the Gene Expression Omnibus (GEO) repository.

### DEG recognition

After mean background correction of multi chips, we normalized quantiles and calculated the expression levels of the microarray matrix by Affymetrix to obtain specific gene expression data. In addition, we adopted Limma and Bayesian statistics to construct a linear model. This study selected DEGs from HCC specimens upon the Log<sub>2</sub> (fold change, FC) > 0.6 and *p* < 0.05 thresholds.

### Annotation of biological function

We used Gene Ontology (GO) analysis as the primary reference for the annotation of genes or gene products or for the interpretation of high-throughput genomic and transcriptomic analysis results [19]. We also used DAVID database (<https://david.ncifcrf.gov/>) for mapping a user-defined gene for related biological annotations, which plays an important role in the successful analysis of genes post-HTS [20]. The present work adopted DAVID for GO analysis to examine DEG effects. *p* < 0.05 stood for statistical significance.

### HCC prognosis nomogram construction

The survival rate was calculated based on Kaplan–Meier (KM) curve analysis, and significant differences across diverse survival curves were determined by the log-rank test. In addition, the Cox proportional hazards model was utilized for univariate and multivariate analyses. Later, the LASSO-Cox method was adopted to reduce the dimensionality, and we selected significant HCC prognostic genes to construct a prognostic model based on Cox regression [21]. The LASSO approach was adopted to establish and validate a nomogram [22]. Afterwards, we drew receiver operating characteristic (ROC) curves and then determined the area under the curve (AUC) values to analyze the specificity and sensitivity of our model [23].

## Immunohistochemical (IHC) staining

Eight HBV-HCC and matched normal specimens were dehydrated, paraffin embedded, sliced into 4- $\mu$ m sections, deparaffinized with xylene and rehydrated with a series of ethanol solutions at room temperature. Then, antigen recovery was performed using sodium citrate, endogenous peroxidase activity was blocked using 3% H<sub>2</sub>O<sub>2</sub>, and specimens were blocked using 5% bovine serum albumin (BSA) at room temperature for 30 min. Subsequently, specimens were incubated with anti-CEP290 primary antibody (No. 22490-1-AP; Proteintech, China) for 12 h at 4° C, followed by an additional 2 h of incubation with HRP-labeled secondary antibodies at room temperature. Color development was conducted using a Cell and Tissue Staining HRP-DAB kit (Beyond) according to manufacturer's protocols. Images were acquired using an Orthophoto microscope ( $\times$ 100). IHC results were semi-quantitatively analyzed by using 'image-pro\_plus' software. IOD values and region area were measured by selecting both control and target regions. We determined the average density (IOD/area) within both regions.

## Cell culture and treatment

Huh7 and Hep3B liver cancer cells, together with immortalized non-cancerous THLE-2 cells, were provided by the Cell Bank of the Chinese Academy of Sciences. All cells were cultured in RPMI-1640 medium (Gibco, USA) supplemented with 10% fetal bovine serum (FBS; Gibco, USA) and 1% penicillin-streptomycin at 37° C with 5% CO<sub>2</sub>.

For inhibition of CEP290 expression, CEP290-siRNAs and the relevant negative control (NC) siRNAs were acquired from RIBOBIO (Guangzhou, China). siRNA sequences used in the present work are shown below: si-CEP290 #1: ID stB0013686A; si-CEP290 #2: ID stB0013686B. Cells ( $1 \times 10^5$ /well) were plated into 6-well plates and cultured until they reached 70%–80% confluence. Lipofectamine 2000 (Invitrogen, USA) was used to transfect siRNA into cells at a final concentration of 100  $\mu$ M according to manufacturer's protocols. Target gene levels were determined at 48 h post-transfection.

## RNA isolation and qRT-PCR

The RNAiso Plus kit (Takara, USA) was used to extract total RNA according to manufacturer's protocols. Complementary DNA was prepared from the extracted RNA by a PrimeScript™ RT Reagent kit with gDNA Eraser (Takara, USA) by reverse transcription under the following conditions: 15 min at 37° C, 5 s at 85° C and

5 min at 4° C. The CFX96 Touch Real-Time PCR system (Bio-Rad, USA) and SYBR Premix Ex Taq II (Takara, USA) were used for qPCR under the following conditions: 30 s at 95° C; followed by 5 s at 95° C and 30 s at 60° C for 40 cycles. The 2- $\Delta\Delta$ Cq approach was employed to quantify relative gene expression levels, with GAPDH as the endogenous reference. The primers used in the present work are shown below: for GAPDH, 5'-CTTTGGTATCGTGGAAGGACTC-3' (forward), 5'-GTAGAGGCAGGGATGATGTTCT-3' (reverse); for CEP290, 5'-GATGCTCACCGAACAAAGTAGAAC-3' (forward), 5'-ATGAGTCTGTTGAGAAAGGGTTG-3' (reverse). All the reactions were performed in triplicate.

## Cell proliferation analysis and colony formation experiment

Cell proliferation was determined with Cell Counting Kit-8 (CCK8, Dojindo, Japan) according to manufacturer's instructions. In brief, cells ( $4 \times 10^3$ /well) were plated in 96-well plates and cultured for 0, 24, 48 and 72 h. Afterwards, CCK-8 solution was added to each well and incubated for 1.5 h in the dark. Thereafter, we determined the living cell count by measuring the absorbance (OD) at 450 nm.

Cells were seeded into 6-well plates (600 cells/well for the si-CEP290 cell model and the si-CEP290 + Oltipraz cell model) using three replicate wells for each group, and were cultured in 5% CO<sub>2</sub> incubator at 37° C. After one week, the cells were fixed with 4% paraformaldehyde for 30 minutes and incubated with crystal violet. After rinsed with PBS three times, the clones were recorded and the number of clones was calculated.

## Scratch assay

Migration of the treated or untreated Hep3B cells was evaluated with an *in vitro* scratch assay. In brief, cells were grown in 6-well plates until they reached 100% density. A sterile pipette tip was utilized to make a wound in the cell monolayer, detached cells were removed by rinsing three times with PBS, and cells were further cultured in serum-free medium for 24 h. Images were taken at 0 and 24 h.

## Transwell assay

The migration and invasion of Hep3B cells were examined by Transwell assay. The treated or untreated Hep3B cells were resuspended in serum-free medium. For measuring cell migration, we first added 100  $\mu$ L as-prepared cell suspension into the Matrigel (BD Biosciences)-coated upper chamber, and 600  $\mu$ L medium that contained 10% FBS was added to the

lower chamber. Then,  $4 \times 10^4$  cells were added into the upper chamber and fixed using 4% paraformaldehyde for 15 min. Twenty-four hours later, we removed the cells from the chamber. Then, the cells were stained with crystal violet (0.1%) for a period of 10 min. Then, we selected cells located in the inner layer and randomly chose 3 fields of view (FOVs) in each sample to determine the number of penetrating cells.

### Iron assay

Relative iron levels within cell lysates were assessed using an iron assay kit (No. DIFE-250; BioAssay Systems, USA) according to manufacturer's instructions.

### Lipid peroxidation analysis

A lipid peroxidation assay kit (No. A003-1-2; Nanjing Jiancheng College of Biotechnology, China) was used to assess malondialdehyde (MDA) content within cell lysates according to manufacturer's instructions.

### Western blotting (WB) analysis

Total proteins were extracted from cells, and protein levels were measured using a BCA protein assay kit (Beyotime, Shanghai, China). Proteins were separated by 8% SDS-PAGE and transferred to PVDF membranes. After blocking in 5% skimmed milk, the membranes were incubated with primary antibodies at 4° C overnight. Next day, the membranes were rinsed with TBST three times, incubated for 1 h with secondary antibody at ambient temperature, rinsed with TBST three times and visualized by ECL (Wanleibio, Shenyang, China). The following antibodies *CEP290* (ab85728, Abcam, USA), *Nrf2*, *NQO1*, *HO-1* (Proteintech, Wuhan, China), *FTH1* (Cell Signaling Technology, USA) and  $\beta$ -actin (DianyinBio, Shanghai, China) were used in this study.

### Xenograft assay in nude mice

Athymic nude (BALB/c) mice were acquired from Chongqing Hospital of Traditional Chinese Medicine. Six-week-old male mice were injected with Hep3B cells and *CEP290*-deficient Hep3B (*CEP290*-KO) cells ( $2 \times 10^6$  cells) in the right flanks into the subdermal space (n=5 per group). Tumors were measured with a caliper on days 0, 4, 8 and 14, and the volume was estimated according to the formula: volume = tumor length in mm  $\times$  width<sup>2</sup> in mm  $\times$  0.5236. Fourteen days post-injection, all animals were euthanized and the tumor tissues were collected. All animal procedures were carried out following the Guidelines for Animal Experiments released by the Chinese government.

### Statistical analysis

Data were displayed as mean  $\pm$  SD. GraphPad Prism analysis software was used for statistical analyses. One-way ANOVA and t-test were performed to assess differences of 2 groups, while those across multiple groups were analyzed through LSD-*t* test and one-way ANOVA. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; #  $p < 0.05$ ; ##  $p < 0.01$ .

### Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## RESULTS

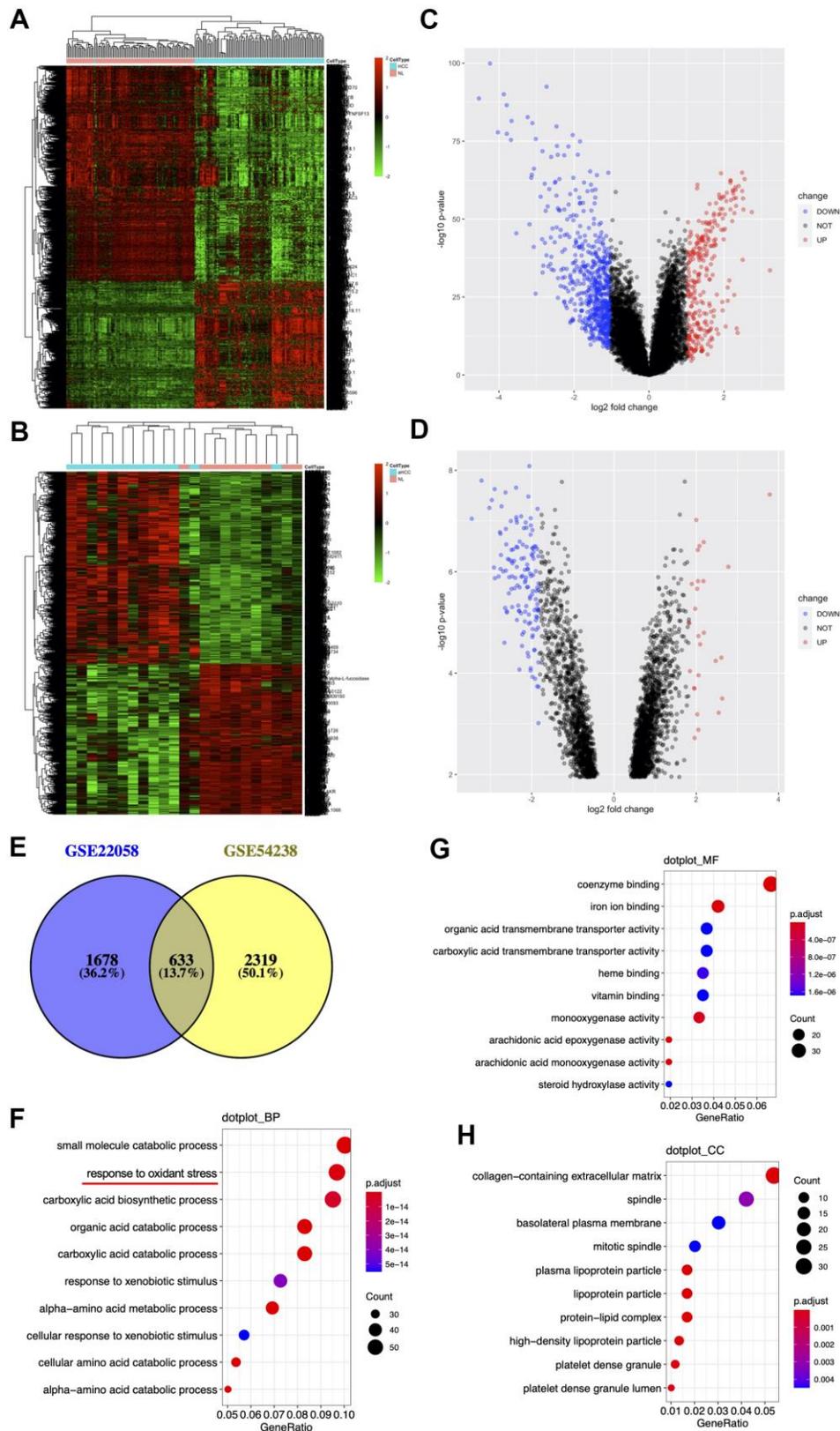
### Shared DEGs identified in the 2 HCC datasets

Two open-source HBV-related HCC datasets GSE54238 (n=23) and GSE22058 (n=192) were downloaded from the GEO repository. DEG expression data were investigated using the Limma R package in both datasets. We identified 2,311 DEGs between 96 HCC specimens and paired non-cancerous specimens from the GSE22058 dataset with the threshold set at  $p < 0.05$  and  $|\text{Log}_2(\text{FC})| > 0.6$ , with 859 genes upregulated and 1,452 downregulated (Figure 1A, 1B). 2,952 DEGs were identified from the GSE54238 dataset when 13 HCC specimens were compared with 10 non-cancerous specimens, with 1,658 genes upregulated and 1,294 downregulated (Figure 1C, 1D). There were 633 shared DEGs between the two datasets (Figure 1E).

To better understand the biological roles of 633 shared DEGs, we used the 'clusterProfiler' [24] function in the R package for functional annotations. GO terms can be divided into three categories, biological processes (BPs), molecular functions (MFs) and cellular components (CCs). For BPs, DEGs were mostly associated with 'small molecule catabolic process', 'response to oxidative stress', and 'carboxylic acid biosynthetic process' (Figure 1F). As for MF, DEGs were mostly related to 'coenzyme binding', 'iron ion binding' and 'organic acid transmembrane transporter activity' (Figure 1G). Concerning CC, those DEGs were mostly associated with 'collagen-containing extracellular matrix', 'spindle' and 'basolateral plasma membrane' (Figure 1H).

### HCC prognosis prediction model

HCC data were obtained from the TCGA-LIHC cohort and randomly divided into training (n=183) and validation (n=182) cohorts. To identify the most representative genes associated with HCC survival from



**Figure 1. Identification and enrichment of DEGs from the GSE54238 and GSE22058 datasets.** (A, B) DEGs hierarchical clustering. Data are the log<sub>2</sub> HCC-to-normal intensity ratio. Green and red represent significantly downregulated and upregulated genes ( $p < 0.05$ ), respectively. aHCC, advanced hepatocellular carcinoma. (C, D) Volcano plot representing the distribution of DEGs. (E) DEGs shared in two datasets. (F–H) Ten significant GO terms.

shared DEGs, we imported 633 genes by removing the overfitted genes using the LASSO regression algorithm and then analyzed them by multivariate Cox proportional risk regression (Figure 2A). Through this maneuver, we obtained nine risk genes from the training cohort to construct a prognostic model and determined the risk score of every specimen according to their expression and regression coefficients [25]. The median risk score in the training set was 0.9519; therefore, this value was selected for dividing specimens into high- and low-risk groups (Figure 2B center). The survival distribution in the training cohort is presented in Figure 2B (Figure 2B bottom). The variations in the expression of the nine risk genes in the heatmap conformed to the respective risk scores of the prognostic model (Figure 2B top). For the validation cohort, the median risk score was 0.9664. This study verified the prognostic prediction performance of the risk score and came to similar conclusions (Figure 2C). According to KM curve analysis, the high-risk group in the training cohort had a poorer prognosis than the low-risk group (Figure 2D;  $p=4.722e-07$ ). Similarly, the high-risk group in the validation cohort had worse survival compared with the low-risk group (Figure 2E;  $p=9.493e-03$ ). In addition, we drew ROCs and determined the AUC values of the two cohorts (0.811 and 0.688, respectively) (Figure 2F, 2G).

### Prioritization of candidate genes

To better identify the most informative gene among those nine risk genes, we examined the signature genes for their clinical implications. According to primary screening, centrosomal protein 290 (*CEP290*), which was not reported in previous studies, has remarkable clinical value. *CEP290* was highly expressed in HCC specimens from the TCGA-LIHC cohort ( $n=419$ ) (Figure 3A), and the ROC curve showed that upregulation of *CEP290* showed good diagnostic performance for HCC (Figure 3B; AUC=0.783;  $p<0.0001$ ). Combined with other HCC clinical indicators, we found that *CEP290* overexpression was significantly related to TNM stage, vascular invasion and the alpha fetoprotein (AFP) level (Figure 3C–3E), whereas there was no significant difference in Child-Pugh grade or histologic grade (Figure 3F, 3G). Subsequently, we analyzed *CEP290* levels in eight HCC specimens and matched non-cancerous specimens. As suggested by the IHC assay, *CEP290* protein expression was significantly increased in HBV-associated HCC compared with matched non-cancerous specimens (Figure 3H). In addition, the prognostic value of *CEP290* for disease-free survival of HCC patients was evaluated by a KM curve analysis according to the best threshold (Figure 3I; cutoff

point=14.8). We found that TCGA LIHC samples with *CEP290* overexpression had worse disease-free survival (DFS) (Figure 3I; HR=1.682,  $p=8.95e-03$ ). To further validate the prognostic prediction performance of *CEP290* following radical hepatectomy, a nomogram was built base on TNM stage and *CEP290* level, the two independent factors for disease-free interval (DFI) discovered from multivariate analysis (Figure 3J). Clinicians were able to estimate 1-, 3- and 5-year disease-free survival based on the summation of each score from different prognostic factors incorporated in the nomogram. Calibration plots and bootstrap C-index were determined for the internal validation of the nomogram. Superior 1-, 3- and 5-year disease-free survival were estimated based on the bootstrapped calibration plots relative to the ideal model (Figure 3K). As a result, the above findings suggested that *CEP290* upregulation was a candidate biomarker to predict liver cancer prognosis.

We next extracted RNA for qRT-PCR and protein for WB from THLE-2, Huh7 and Hep3B cell lines. *CEP290* expression was elevated in Hep3B and Huh7 cell lines relative to THLE-2 cells when examined by qRT-PCR and WB (Figure 4A–4C). These results suggested *CEP290* might play an important role in liver cancer. Although *CEP290* mRNA was more abundantly expressed in Hep3B cells compared with Huh7 cells, the protein expression in Hep3B cells was lower than that in Huh7 cells. As a result, we chose Hep3B cells for subsequent experiments.

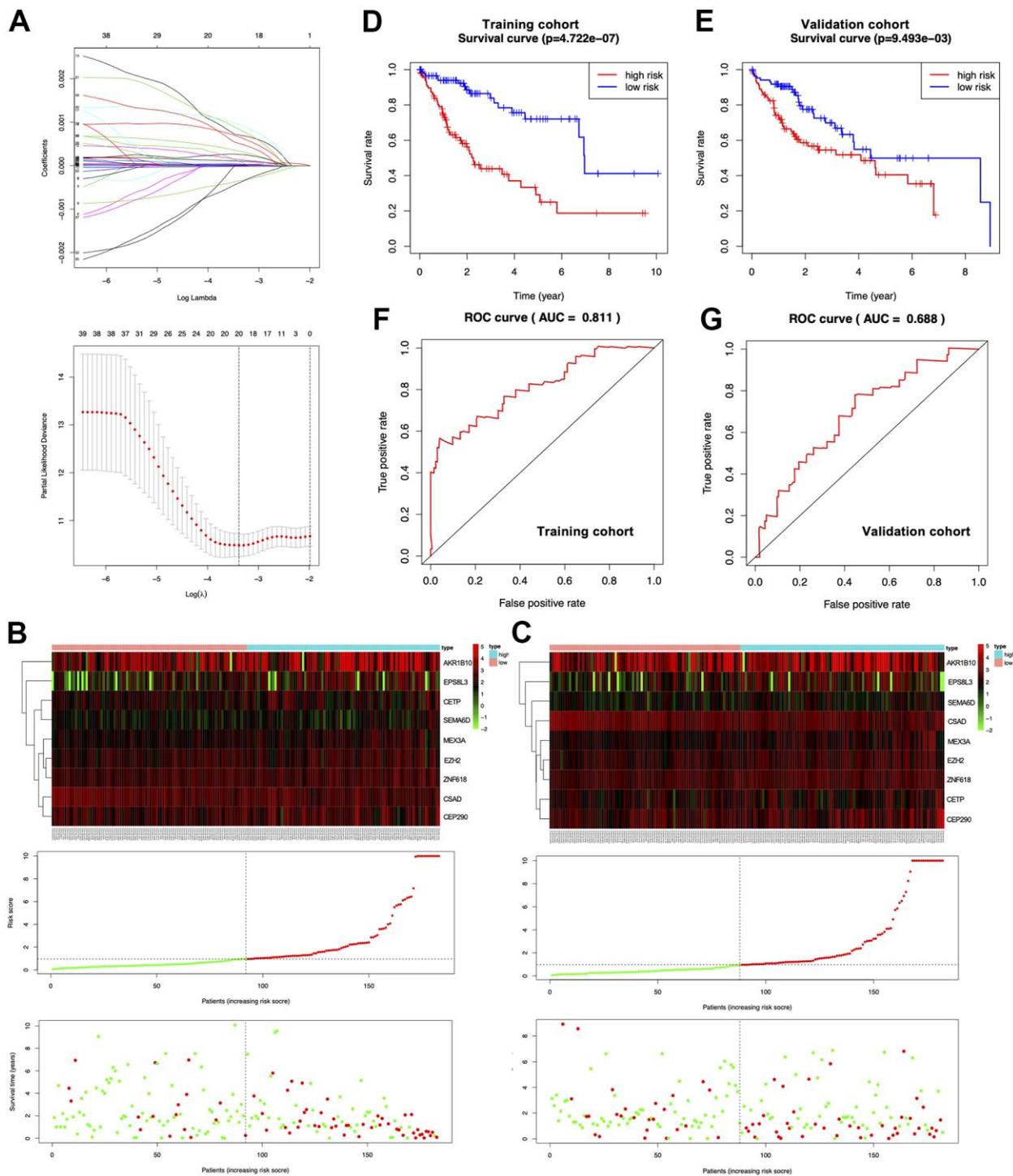
### *CEP290* knockdown suppressed liver cancer cell malignant phenotypes and ferroptosis

To elucidate the role of *CEP290* in HCC, we transfected Hep3B cells with small interfering RNA (siRNA) specific to *CEP290* (siRNA #1 and #2). *CEP290* mRNA levels were markedly reduced in Hep3B cells transfected with si-*CEP290*#1 and #2, indicating that *CEP290* was successfully silenced (Figure 4D, 4E). Si-*CEP290* exhibited superior activity and was used for subsequent analysis. Compared to NC-siRNA, Hep3B cells transfected with *CEP290*-siRNA showed significantly decreased proliferation (Figure 4F). In addition, a colony formation assay was detected. *CEP290* inhibition reduced the number of colonies formed (Figure 4G, 4H). Furthermore, *CEP290*-silenced cells exhibited decreased migration and invasion compared with NC-siRNA cells (Figure 5A–5C). Therefore, we concluded that *CEP290* played an essential role in the invasion and migration of Hep3B cells.

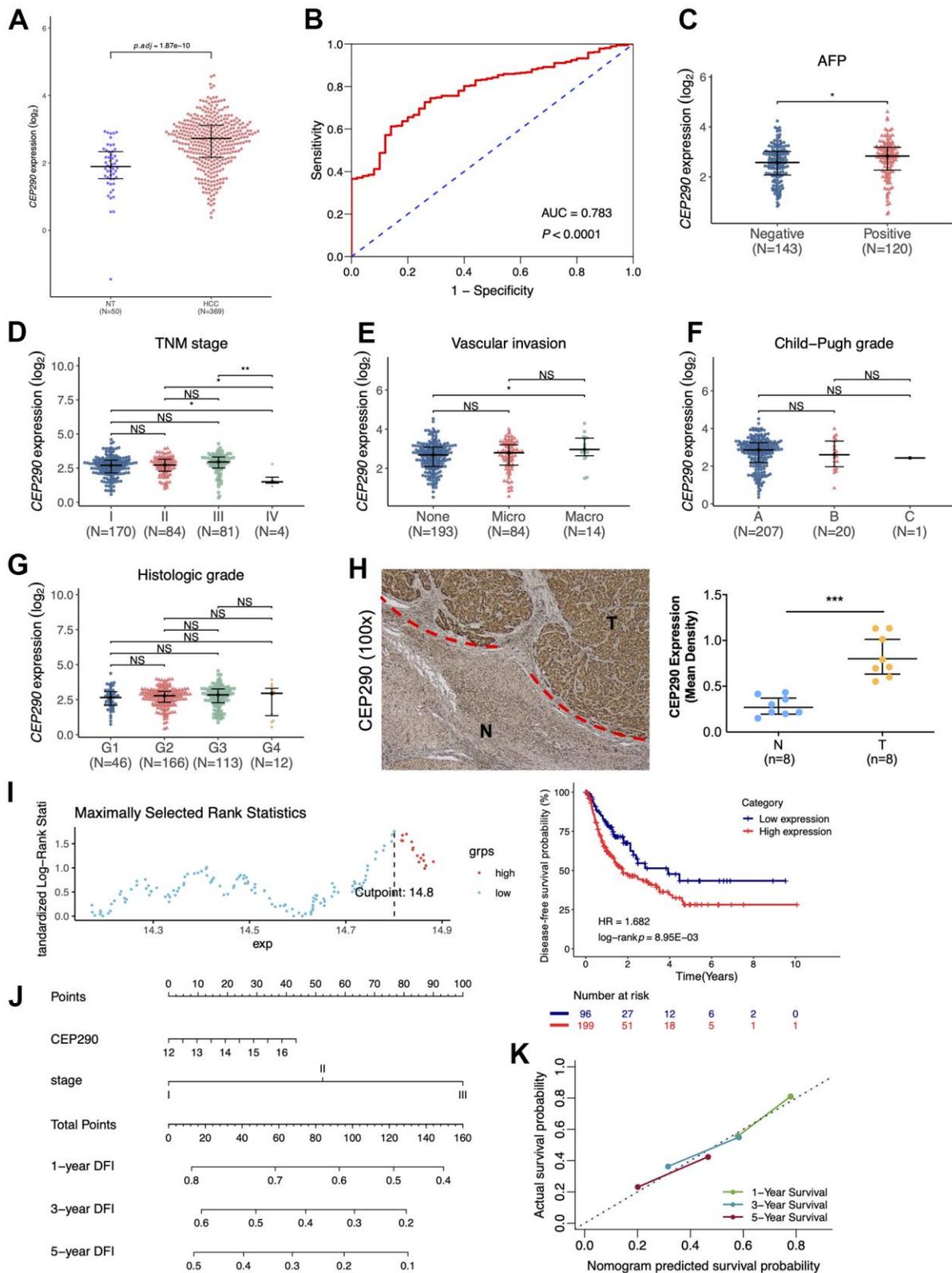
Next, to determine the mode of cell death, we treated Hep3B cells with the apoptosis inhibitor Z-DEVD-FMK, autophagy inhibitor 3-MA, and ferroptosis

inhibitor ferrostatin-1. The results demonstrated that both apoptosis and autophagy inhibition had only marginal impact on the cell death induced by *CEP290* depletion. However, the ferroptosis inhibitor strikingly

reduced melanoma cell death induced by the suppression of *CEP290* (Figure 5D). The above results suggested that *CEP290* inhibition promoted ferroptosis in liver cancer cells (Figure 5E). Subsequently, we



**Figure 2. Establishment of the prognosis prediction nomogram.** (A) Signature gene number determined through LASSO analysis. (B, C) The top, middle and bottom represent prognostic signature gene expression within the high- or low-risk group, risk score distribution, and patient survival of both cohorts, respectively. (D, E) KM curves for both cohorts. (F, G) ROC curves for both cohorts.



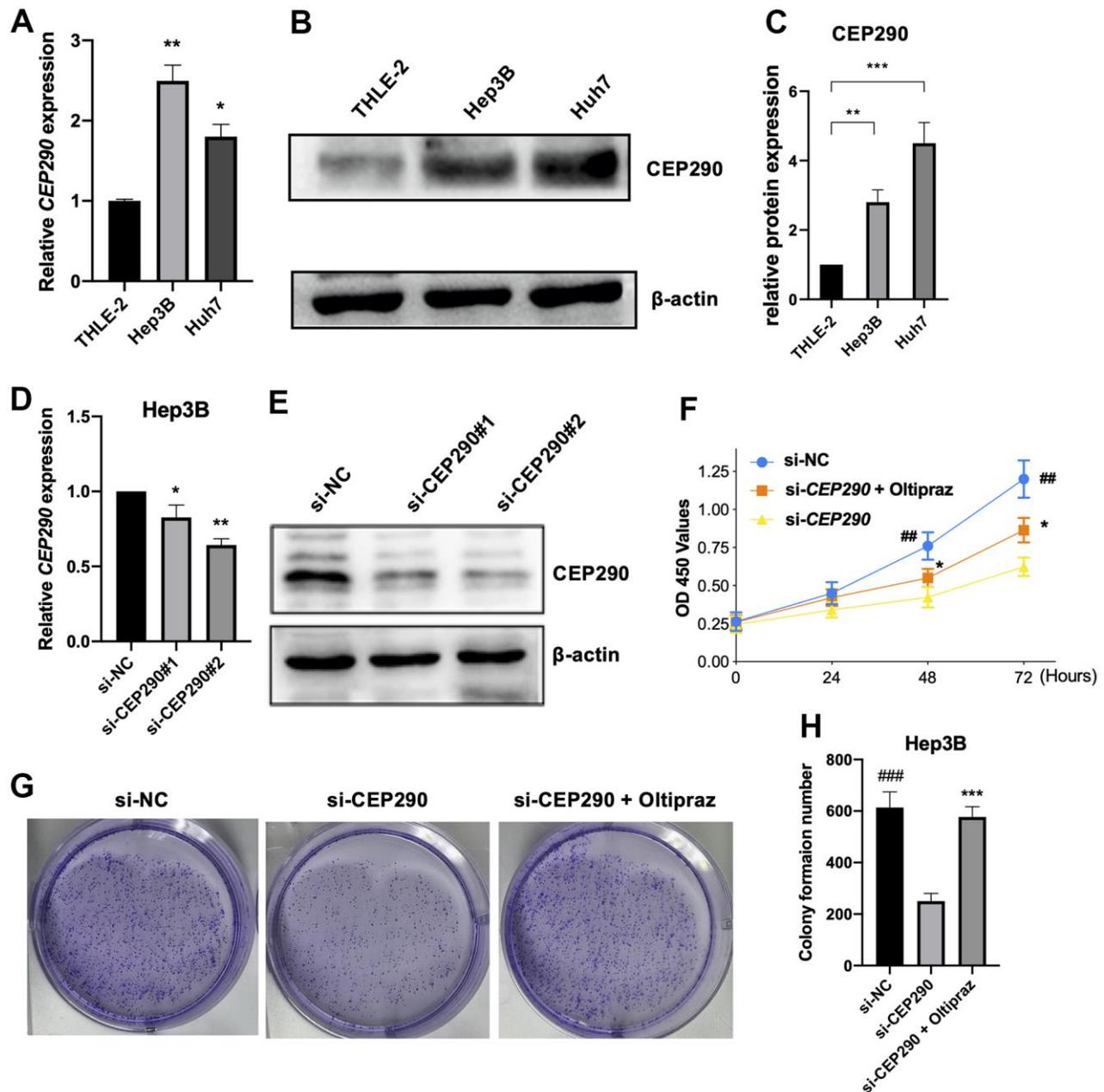
**Figure 3. Clinical significance of *CEP290*.** (A) The *CEP290* mRNA expression level is shown in TCGA-LIHC data (n=419). (B) Validation of the predictive value of *CEP290* upregulation in diagnosing HCC based on the ROC curve. (C–G) *CEP290* levels in comparison with AFP, TMN stage, vascular invasion, Child-Pugh grade and histologic grade. (H) Typical images (left) as well as quantification (right) of the IHC staining of *CEP290* in the 8 HCC specimens and matched non-cancerous specimens. (I) DFS analysis based on the X-tile plots threshold. (J) Prognosis nomogram for HCC cases after surgery. (K) Calibration curve for the nomogram used to predict 1-, 3- and 5-year DFS.

evaluated ferroptosis-related markers such as lipid peroxidation (MDA) and iron accumulation upon *CEP290* interference to determine the association of *CEP290* with ferroptosis in Hep3B cells. *CEP290* silencing increased the intracellular  $Fe^{2+}$  level (Figure 5F). Furthermore, the accumulation of MDA, the representative end product of lipid peroxidation, was analyzed in Hep3B cells depleted of *CEP290*. *CEP290* inhibition enhanced MDA accumulation in Hep3B cells

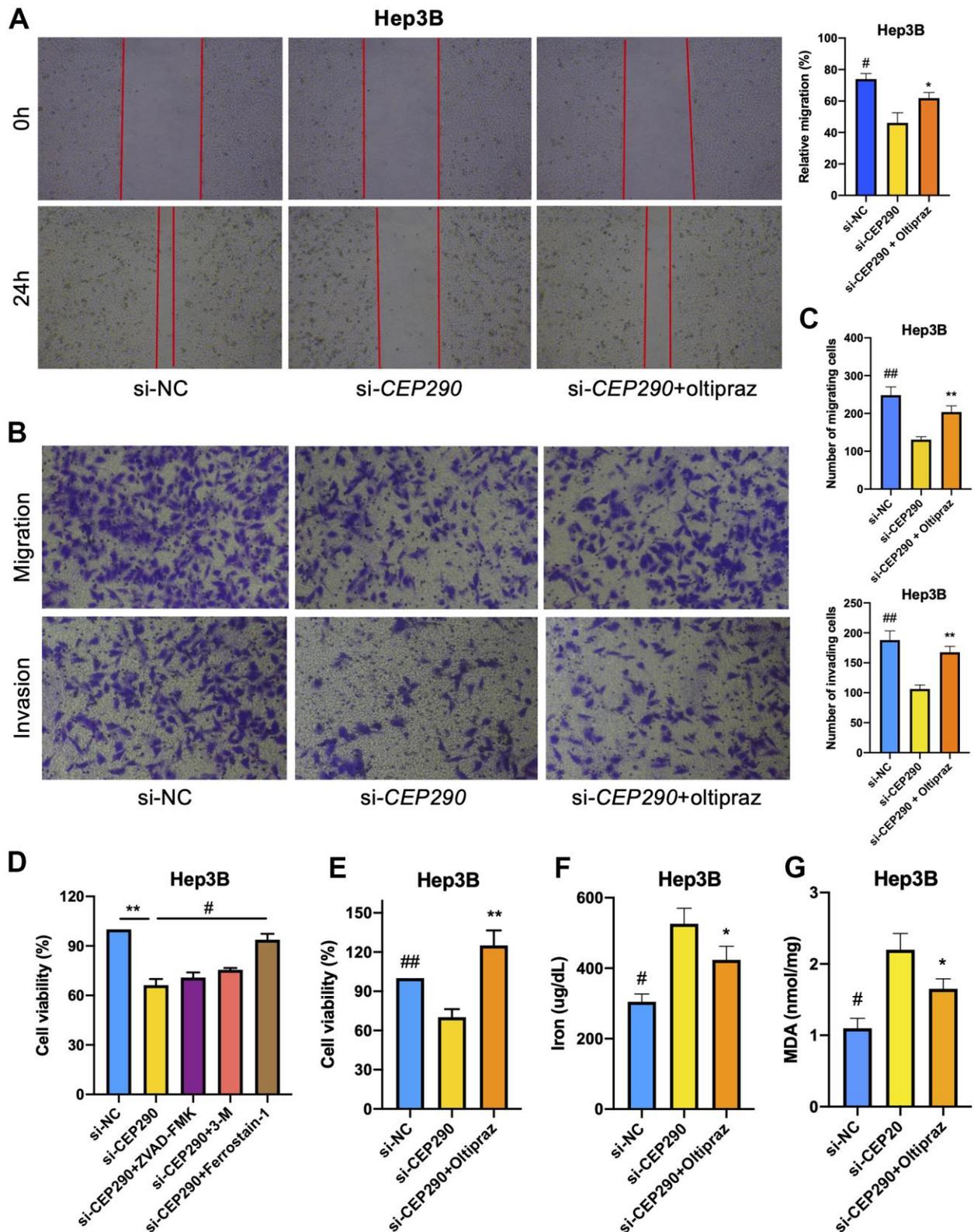
(Figure 5G). The above findings suggested a possible role of *CEP290* as a ferroptosis regulating factor that modulates MDA and  $Fe^{2+}$  contents in Hep3B cells.

#### *CEP290* knockdown inhibited *Nrf2* signaling pathway

Enrichment analysis suggested that (Figure 1F) the oxidant-stress response played an important role in HCC development. Therefore, we conducted correlation



**Figure 4. Verification of *CEP290* expression levels in liver cancer cells.** (A, B) *CEP290* levels in HCC cells were measured by qRT-PCR and WB. (C) The quantification of protein expression of *CEP290* in diverse cell lines. (D, E) *CEP290* interference reduced mRNA and protein expression in Huh7 cells. (F) Cell proliferation measured by CCK-8 assay after *CEP290* interference. (G, H) Representative images (left) and quantification (right) of colony formation capacity. (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; ## $p < 0.01$ , ### $p < 0.001$ ).

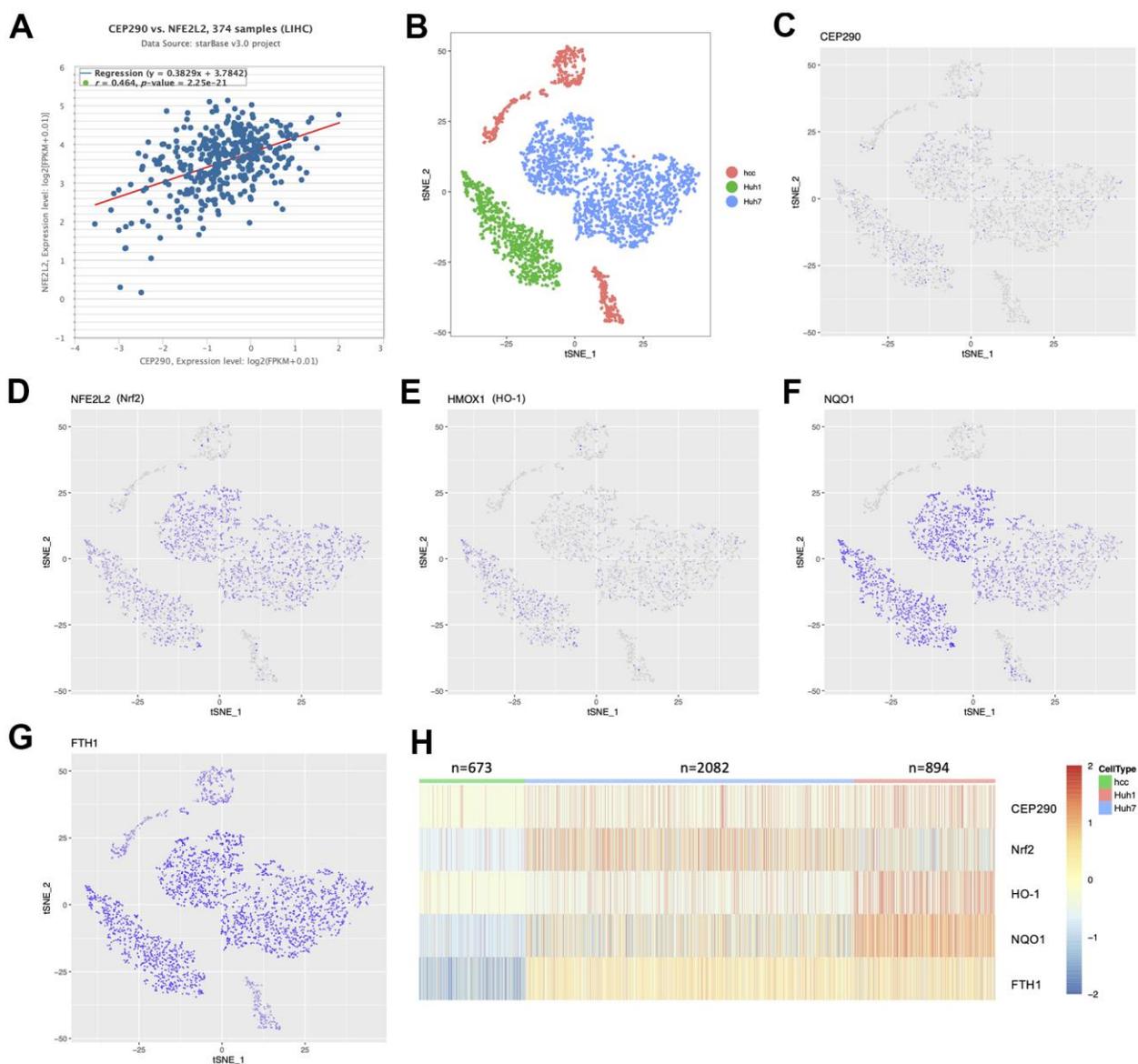


**Figure 5. Effect of *CEP290* interference on liver cancer cell ferroptosis and malignant phenotypes.** (A) Scratch assay. Scratches were made in a confluent culture, and cell migration into the scratch wounds was documented. (B, C) Transwell assays. After treatment, the migration and invasion of Hep3B cells were analyzed. (D) Hep3B cell viability was evaluated after si-*CEP290*, incorporate with treatment of Z-DEVD-FMK (100  $\mu$ M), 3-MA (10 mM), and ferrostatin-1 (100  $\mu$ M). (E–G) Cell viability, ferrous iron and MDA levels in Hep3B cells treated with si-*CEP290* and 75  $\mu$ M oltpiraz (\* si-*CEP290* vs. si-*CEP290*+oltpiraz, \*\* $p$  < 0.01, \* $p$  < 0.05; # si-NC vs. si-*CEP290*, ## $p$  < 0.01, # $p$  < 0.05).

analysis of the TCGA-LIHC RNA-Seq cohort (n=374) to investigate the relationship of *CEP290* with the oxidant-stress core gene *Nrf2*. *CEP290* expression showed a positive correlation with *Nrf2* (Figure 6A;  $r=0.464$ ,  $p=2.25e-21$ ), which suggests *CEP290* may be functionally related to the *Nrf2*-mediated oxidant stress response. Next, we sought to determine the expression landscape of *CEP290* and the genes in the *Nrf2* signal transduction pathway in HCC cells using 10×Genomics single-cell RNA sequencing data (GSE103867) [18]. The tSNE clustering analysis clearly distinguished between resected HCC (hcc), Huh1 and Huh7 cells (Figure 6B). However, *CEP290*, *HO-1*, *Nrf2*, *NQO1*

and *FTH1* levels were basically consistent (Figure 6C–6H), implying that the *Nrf2* pathway activation may be associated with *CEP290* overexpression.

Then, we explored the potential mechanism how *CEP290* might regulate *Nrf2* expression. A previous article reported that *Nrf2* is a vital transcription factor [26]; hence, we hypothesized that *CEP290* may interact with *Nrf2*, promote *Nrf2* transcription and activate the *Nrf2* signaling pathway. To test this, we first evaluated the *CEP290-Nrf2* interaction within Hep3B cells by conducting CoIP assays. Since the siRNA approach may not be sufficient to show the effect of *CEP290*



**Figure 6. The expression distribution between *CEP290* and *Nrf2* pathway members in single-cell profiling.** (A) Co-expression correlation analysis using the TCGA-LIHC RNA-Seq cohort (n=374). (B) t-SNE map shows three liver cancer cell classes. Each dot represents one cell. (C–H) Profile of target expression (scRNA-Seq) based on the t-SNE plot from Figure 6B.

depletion in CoIP experiments clearly, we generated a cell line (Hep3B *CEP290*<sup>-/-</sup>) with a homozygous knockout of *CEP290* using the CRISPR/Case9 gene editing system. CoIP results confirmed the interaction between *CEP290* and *Nrf2* (Figure 7A, 7B). Western blot results showed that nuclear translocation of *Nrf2* in *CEP290*-KO Hep3B cells was reduced (Figure 6C), and *CEP290* silencing significantly downregulated the expression levels of *Nrf2* signaling pathway members (*FTH1*, *Nrf2*, *NQO1*, *HO-1*) (Figure 7D, 7E). These observations demonstrated that *CEP290* promoted the transcription of *Nrf2* pathway genes through *Nrf2*.

We next used oltipraz, an activator of *Nrf2*, to determine whether the inhibitory effect of *CEP290* could be circumvented by *Nrf2* activation. Hep3B cells were treated at different concentrations to determine an appropriate dose, and we found that 75  $\mu$ M oltipraz had the optimal effect as shown in Figure 7F. Administration of oltipraz completely reversed the effect of *CEP290* silencing on cell growth, migration, invasion and ferroptosis (Figures 4F–4H, 5A–5C, 5E–5G). Furthermore, oltipraz treatment upregulated the expression levels of the *Nrf2* pathway genes (*Nrf2*, *HO-1*, *NQO1* and *FTH1*) (Figure 7D, 7E). These findings suggested that *CEP290* regulated biological behaviors and ferroptosis in HCC cells by activating the *Nrf2* pathway.

### Loss of *CEP290* suppresses tumor growth *in vivo*

Since *CEP290* silencing suppressed HCC malignant behaviors *in vitro*, we investigated whether *CEP290* knockout inhibited tumor growth *in vivo*. *CEP290* knockout (*CEP290*-KO) Hep3B cells ( $2 \times 10^6$  per injection site) were injected into nude mice subcutaneously. Compared to the control group, *CEP290*-KO cells formed smaller tumors (Figure 7G, 7H). Collectively, these results suggest that *CEP290* plays an important role in the regulation of HCC proliferation in a subcutaneous xenograft model.

## DISCUSSION

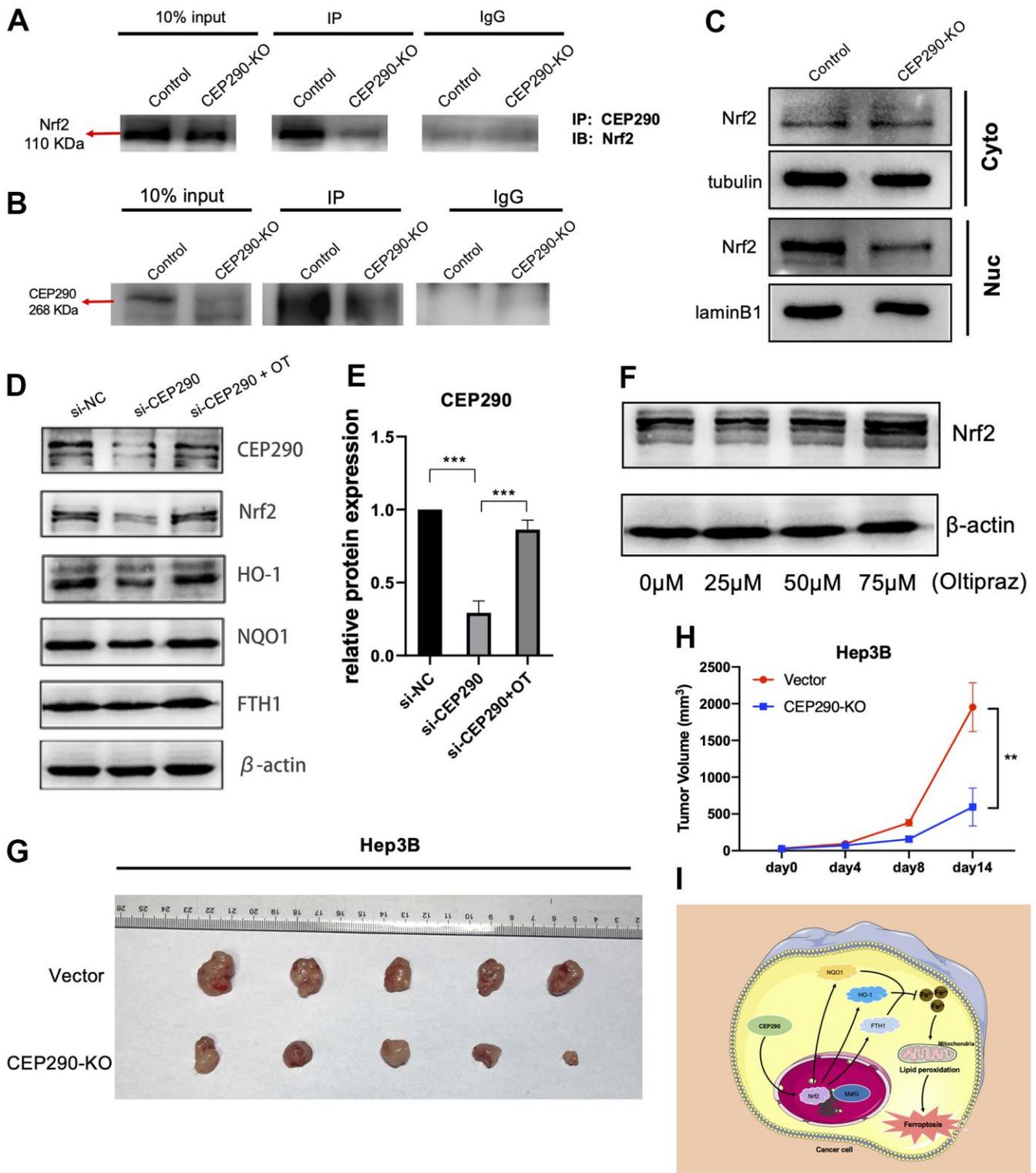
In the present work, we downloaded gene expression profile datasets GSE54238 and GSE22058 and identified 633 shared DEGs. Through a series of bioinformatic analyses and *in vitro* and *in vivo* assays, we identified *CEP290* as a novel candidate biomarker of HCC prognosis, demonstrated that *CEP290* played important functions in the growth, migration, invasion and ferroptosis of HCC cells and explored its potential mechanism of action. This study provides strong evidence that *CEP290* functions in regulating ferroptosis in liver cancer cells via the *Nrf2* signaling pathway (Figure 7I).

Ferroptosis is a novel iron-dependent form of cell death [27] and is regulated by diverse signal transduction pathways in cancers [28]. *SLC7A11* regulated cancer cell ferroptosis through glucose- and glutamine-dependency [29]. *GLRX5* knockdown stimulated the iron-starvation response and increased intracellular Fe<sup>2+</sup> through iron-responsive element-binding protein, thus inducing ferroptosis [30]. However, the regulatory mechanism of ferroptosis remains ambiguous.

As evident from recent findings, centrosomal protein (*CEP*) plays an important role in carcinogenesis. For example, *CEP55* overexpression increases cancer cell stemness and enhances tumor formation by activating the PI3K/AKT pathway [31]. *TACC3*, a key centrosomal protein, was upregulated in prostate cancer, and its silencing inhibited tumor growth [32]. *CEP290* was previously reported to be involved in cell ciliogenesis [33]. Deletion of *CEP290* blocked the formation of cilia by directly recruiting *DAZ* and zinc finger protein 1 *DZIF1* [34] or by disrupting the formation and subcellular distribution of the protein complex *PCM-1* [35]. However, the biological function of *CEP290* is virtually unexplored in cancers. In our study, *CEP290* was identified from a prognostic signature screening. It is overexpressed in the TCGA-LIHC cohort, HCC tissues and liver cancer cell lines. Using multivariate analyses, we found that *CEP290* expression in TCGA-LIHC data could serve as an independent indicator of poor prognosis. In addition, overexpression of *CEP290* was also correlated with the AFP level, TNM stage and vascular invasion, suggesting that *CEP290* could be used as a prognostic marker.

To clarify the contribution of *CEP290* to liver cancer progression, we examined the effect of *CEP290* depletion both *in vitro* and *in vivo*. In Hep3B cells, the suppression of *CEP290* expression inhibited cell proliferation, migration and invasion and induced ferroptosis. Moreover, *CEP290* knockout repressed Hep3B cell tumor formation *in vivo*. These findings proved that *CEP290* contributed to hepatocarcinogenesis and key implication for the development of therapeutic tactics of HCC.

*Nrf2*, a core gene for the oxidant stress response, activates its downstream genes and would promote cancer progression [36], and therefore, *Nrf2* is a candidate therapeutic target to treat cancer [37, 38]. According to published literature, *Nrf2* knockdown inhibited HCC development [39, 40], which suggested that *Nrf2* induces HCC by activating downstream targets. Our results have confirmed the interaction between *CEP290* and *Nrf2* proteins. *CEP290* silencing blocked the nuclear translocation of *Nrf2* and reduced



**Figure 7. CEP290 knockout inhibits tumor growth *in vivo*.** (A, B) The interaction between CEP290 and Nrf2 was analyzed by WB and immunoprecipitation in Hep3B cells. (C) Nrf2 expression in nuclear extracts after CEP290 knockout. (D) Following CEP290 silencing, the protein expression of Nrf2 and the respective downstream members (HO-1, FTH1 and NQO1) were detected by WB, and (E) the protein levels of CEP290 were quantified. (F) Selection of optimal oltipraz dose. (G) Gross appearances of tumors. Male nude mice (six weeks old, n = 5 for each group) were subcutaneously injected with Hep3B and CEP290-KO Hep3B cells ( $2 \times 10^6$ ) on the right flank. The mouse tumors were collected 2 weeks after implantation. (H) Xenograft tumor sizes were determined in mice bearing tumors. (I) A proposed model of the CEP290/Nrf2 axis activation in HCC. All results are presented as the mean  $\pm$  SD; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

the expression of *Nrf2* pathway members. Administration of *Nrf2* activator oltipraz elevated the expression of *Nrf2* pathway members and induced malignant phenotypes of liver cancer cells in *CEP290*-depleted cancer cells. These data revealed that *CEP290* modulated ferroptosis and malignant phenotypes of liver cancer cells by regulating the *Nrf2* signaling pathway.

## CONCLUSIONS

In summary, the present work established DEGs from 2 independent HCC datasets and identified *CEP290* as a predictor of patient survival. *CEP290* is involved in the ferroptosis of HCC cells via the *Nrf2* pathway. The present study provides a novel direction in deciphering the mechanism of HCC initiation.

## AUTHOR CONTRIBUTIONS

Project conception and planning: Q.P.; Main experiments and data analysis: Y.R.S., G.Y., Q.H.L., X.Y.H., D.W.Q. and Y.H.Z.; Manuscript preparation: Q.P., Y.R.S. and Y.X.; Critical review and discussion: Q.P., F.H.T. and L.C.

## CONFLICTS OF INTEREST

All authors declared that they have no conflicts of interest in this work.

## FUNDING

This work was sponsored by National Natural Science Foundation of China (Grant nos. 81904218 to Q.P. and 81704090 to X.Y.H.) and Natural Science Foundation of Chongqing, China (cstc2019jcyj-msxmX0095 to Q.P.).

## Editorial note

&This corresponding author has a verified history of publications using a personal email address for correspondence.

## REFERENCES

1. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. 2019; 16:589–604. <https://doi.org/10.1038/s41575-019-0186-y> PMID:31439937
2. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer*. 2006; 118:3030–44. <https://doi.org/10.1002/ijc.21731> PMID:16404738
3. Liu X, Baecker A, Wu M, Zhou JY, Yang J, Han RQ, Wang PH, Liu AM, Gu X, Zhang XF, Wang XS, Su M, Hu X, et al. Raw Garlic Consumption and Risk of Liver Cancer: A Population-Based Case-Control Study in Eastern China. *Nutrients*. 2019; 11:2038. <https://doi.org/10.3390/nu11092038> PMID:31480423
4. Wu J, Yang S, Xu K, Ding C, Zhou Y, Fu X, Li Y, Deng M, Wang C, Liu X, Li L. Patterns and Trends of Liver Cancer Incidence Rates in Eastern and Southeastern Asian Countries (1983-2007) and Predictions to 2030. *Gastroenterology*. 2018; 154:1719–28.e5. <https://doi.org/10.1053/j.gastro.2018.01.033> PMID:29549041
5. Sandquist M, Wong HR. Biomarkers of sepsis and their potential value in diagnosis, prognosis and treatment. *Expert Rev Clin Immunol*. 2014; 10:1349–56. <https://doi.org/10.1586/1744666X.2014.949675> PMID:25142036
6. Chistiakov DA, Myasoedova VA, Grechko AV, Melnichenko AA, Orekhov AN. New biomarkers for diagnosis and prognosis of localized prostate cancer. *Semin Cancer Biol*. 2018; 52:9–16. <https://doi.org/10.1016/j.semcancer.2018.01.012> PMID:29360504
7. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc*. 2015; 2015:951–69. <https://doi.org/10.1101/pdb.top084970> PMID:25870306
8. Kim DH, Ahn JS, Han HJ, Kim HM, Hwang J, Lee KH, Cha-Molstad H, Ryoo IJ, Jang JH, Ko SK, Yang JO, Lee HG, Lee S, et al. Cep131 overexpression promotes centrosome amplification and colon cancer progression by regulating Plk4 stability. *Cell Death Dis*. 2019; 10:570. <https://doi.org/10.1038/s41419-019-1778-8> PMID:31358734
9. Xie S, Qin J, Liu S, Zhang Y, Wang J, Shi X, Li D, Zhou J, Liu M. Cep70 overexpression stimulates pancreatic cancer by inducing centrosome abnormality and microtubule disorganization. *Sci Rep*. 2016; 6:21263. <https://doi.org/10.1038/srep21263> PMID:26893288
10. Stockwell BR, Friedmann Angeli JP, Bayir H, Bush AI, Conrad M, Dixon SJ, Fulda S, Gascón S, Hatzios SK, Kagan VE, Noel K, Jiang X, Linkermann A, et al. Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease. *Cell*. 2017; 171:273–85. <https://doi.org/10.1016/j.cell.2017.09.021>

PMID:[28985560](https://pubmed.ncbi.nlm.nih.gov/28985560/)

11. Li J, Cao F, Yin HL, Huang ZJ, Lin ZT, Mao N, Sun B, Wang G. Ferroptosis: past, present and future. *Cell Death Dis.* 2020; 11:88. <https://doi.org/10.1038/s41419-020-2298-2> PMID:[32015325](https://pubmed.ncbi.nlm.nih.gov/32015325/)
12. Liang C, Zhang X, Yang M, Dong X. Recent Progress in Ferroptosis Inducers for Cancer Therapy. *Adv Mater.* 2019; 31:e1904197. <https://doi.org/10.1002/adma.201904197> PMID:[31595562](https://pubmed.ncbi.nlm.nih.gov/31595562/)
13. Liu P, Wu D, Duan J, Xiao H, Zhou Y, Zhao L, Feng Y. NRF2 regulates the sensitivity of human NSCLC cells to cystine deprivation-induced ferroptosis via FOCAD-FAK signaling pathway. *Redox Biol.* 2020; 37:101702. <https://doi.org/10.1016/j.redox.2020.101702> PMID:[32898818](https://pubmed.ncbi.nlm.nih.gov/32898818/)
14. Ursini F, Maiorino M. Lipid peroxidation and ferroptosis: The role of GSH and GPx4. *Free Radic Biol Med.* 2020; 152:175–85. <https://doi.org/10.1016/j.freeradbiomed.2020.02.027> PMID:[32165281](https://pubmed.ncbi.nlm.nih.gov/32165281/)
15. Bersuker K, Hendricks JM, Li Z, Magtanong L, Ford B, Tang PH, Roberts MA, Tong B, Maimone TJ, Zoncu R, Bassik MC, Nomura DK, Dixon SJ, Olzmann JA. The CoQ oxidoreductase FSP1 acts parallel to GPX4 to inhibit ferroptosis. *Nature.* 2019; 575:688–92. <https://doi.org/10.1038/s41586-019-1705-2> PMID:[31634900](https://pubmed.ncbi.nlm.nih.gov/31634900/)
16. Burchard J, Zhang C, Liu AM, Poon RT, Lee NP, Wong KF, Sham PC, Lam BY, Ferguson MD, Tokiwa G, Smith R, Leeson B, Beard R, et al. microRNA-122 as a regulator of mitochondrial metabolic gene network in hepatocellular carcinoma. *Mol Syst Biol.* 2010; 6:402. <https://doi.org/10.1038/msb.2010.58> PMID:[20739924](https://pubmed.ncbi.nlm.nih.gov/20739924/)
17. Yuan S, Wang J, Yang Y, Zhang J, Liu H, Xiao J, Xu Q, Huang X, Xiang B, Zhu S, Li L, Liu J, Liu L, Zhou W. The Prediction of Clinical Outcome in Hepatocellular Carcinoma Based on a Six-Gene Metastasis Signature. *Clin Cancer Res.* 2017; 23:289–97. <https://doi.org/10.1158/1078-0432.CCR-16-0395> PMID:[27449498](https://pubmed.ncbi.nlm.nih.gov/27449498/)
18. Zheng H, Pomyen Y, Hernandez MO, Li C, Livak F, Tang W, Dang H, Greten TF, Davis JL, Zhao Y, Mehta M, Levin Y, Shetty J, et al. Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma. *Hepatology.* 2018; 68:127–40. <https://doi.org/10.1002/hep.29778> PMID:[29315726](https://pubmed.ncbi.nlm.nih.gov/29315726/)
19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat Genet.* 2000; 25:25–9. <https://doi.org/10.1038/75556> PMID:[10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
20. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003; 4:P3. <https://doi.org/10.1186/gb-2003-4-5-p3> PMID:[12734009](https://pubmed.ncbi.nlm.nih.gov/12734009/)
21. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J.* 2010; 52:70–84. <https://doi.org/10.1002/bimj.200900028> PMID:[19937997](https://pubmed.ncbi.nlm.nih.gov/19937997/)
22. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* 2008; 26:1364–70. <https://doi.org/10.1200/JCO.2007.12.9791> PMID:[18323559](https://pubmed.ncbi.nlm.nih.gov/18323559/)
23. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol.* 2017; 17:53. <https://doi.org/10.1186/s12874-017-0332-6> PMID:[28388943](https://pubmed.ncbi.nlm.nih.gov/28388943/)
24. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–7. <https://doi.org/10.1089/omi.2011.0118> PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
25. Zeng D, Zhou R, Yu Y, Luo Y, Zhang J, Sun H, Bin J, Liao Y, Rao J, Zhang Y, Liao W. Gene expression profiles for a prognostic immunoscore in gastric cancer. *Br J Surg.* 2018; 105:1338–48. <https://doi.org/10.1002/bjs.10871> PMID:[29691839](https://pubmed.ncbi.nlm.nih.gov/29691839/)
26. Shelton P, Jaiswal AK. The transcription factor NF-E2-related factor 2 (Nrf2): a protooncogene? *FASEB J.* 2013; 27:414–23. <https://doi.org/10.1096/fj.12-217257> PMID:[23109674](https://pubmed.ncbi.nlm.nih.gov/23109674/)
27. Ma Y, Han F, Min J, Lin W. Energy metabolism as a regulator of ferroptosis. *Cell Cycle.* 2020; 19:2960–2. <https://doi.org/10.1080/15384101.2020.1838781> PMID:[33112197](https://pubmed.ncbi.nlm.nih.gov/33112197/)
28. Battaglia AM, Chirillo R, Aversa I, Sacco A, Costanzo F, Biamonte F. Ferroptosis and Cancer: Mitochondria Meet the “Iron Maiden” Cell Death. *Cells.* 2020; 9:1505. <https://doi.org/10.3390/cells9061505> PMID:[32575749](https://pubmed.ncbi.nlm.nih.gov/32575749/)
29. Koppula P, Zhuang L, Gan B. Cystine transporter SLC7A11/xCT in cancer: ferroptosis, nutrient dependency, and cancer therapy. *Protein Cell.* 2021; 12:599–620. <https://doi.org/10.1007/s13238-020-00789-5>

- PMID:[33000412](#)
30. Lee J, You JH, Shin D, Roh JL. Inhibition of Glutaredoxin 5 predisposes Cisplatin-resistant Head and Neck Cancer Cells to Ferroptosis. *Theranostics*. 2020; 10:7775–86.  
<https://doi.org/10.7150/thno.46903>  
PMID:[32685019](#)
  31. Tandon D, Banerjee M. Centrosomal protein 55: A new paradigm in tumorigenesis. *Eur J Cell Biol*. 2020; 99:151086.  
<https://doi.org/10.1016/j.ejcb.2020.151086>  
PMID:[32646645](#)
  32. Qie Y, Wang L, Du E, Chen S, Lu C, Ding N, Yang K, Xu Y. TACC3 promotes prostate cancer cell proliferation and restrains primary cilium formation. *Exp Cell Res*. 2020; 390:111952.  
<https://doi.org/10.1016/j.yexcr.2020.111952>  
PMID:[32156598](#)
  33. Rachel RA, Li T, Swaroop A. Photoreceptor sensory cilia and ciliopathies: focus on CEP290, RPGR and their interacting proteins. *Cilia*. 2012; 1:22.  
<https://doi.org/10.1186/2046-2530-1-22>  
PMID:[23351659](#)
  34. Wu Z, Pang N, Zhang Y, Chen H, Peng Y, Fu J, Wei Q. CEP290 is essential for the initiation of ciliary transition zone assembly. *PLoS Biol*. 2020; 18:e3001034.  
<https://doi.org/10.1371/journal.pbio.3001034>  
PMID:[33370260](#)
  35. Kim J, Krishnaswami SR, Gleeson JG. CEP290 interacts with the centriolar satellite component PCM-1 and is required for Rab8 localization to the primary cilium. *Hum Mol Genet*. 2008; 17:3796–805.  
<https://doi.org/10.1093/hmg/ddn277>  
PMID:[18772192](#)
  36. Hayes JD, Dinkova-Kostova AT, Tew KD. Oxidative Stress in Cancer. *Cancer Cell*. 2020; 38:167–97.  
<https://doi.org/10.1016/j.ccell.2020.06.001>  
PMID:[32649885](#)
  37. Shin SM, Yang JH, Ki SH. Role of the Nrf2-ARE pathway in liver diseases. *Oxid Med Cell Longev*. 2013; 2013:763257.  
<https://doi.org/10.1155/2013/763257>  
PMID:[23766860](#)
  38. Xue D, Zhou C, Shi Y, Lu H, Xu R, He X. Nuclear transcription factor Nrf2 suppresses prostate cancer cells growth and migration through upregulating ferroportin. *Oncotarget*. 2016; 7:78804–12.  
<https://doi.org/10.18632/oncotarget.12860>  
PMID:[27788496](#)
  39. Bartolini D, Dallaglio K, Torquato P, Piroddi M, Galli F. Nrf2-p62 autophagy pathway and its response to oxidative stress in hepatocellular carcinoma. *Transl Res*. 2018; 193:54–71.  
<https://doi.org/10.1016/j.trsl.2017.11.007>  
PMID:[29274776](#)
  40. Raghunath A, Sundarraj K, Arfuso F, Sethi G, Perumal E. Dysregulation of Nrf2 in Hepatocellular Carcinoma: Role in Cancer Progression and Chemoresistance. *Cancers (Basel)*. 2018; 10:481.  
<https://doi.org/10.3390/cancers10120481>  
PMID:[30513925](#)

# Pan-cancer analysis revealed the significance of the GTPBP family in cancer

Yiming Hu<sup>1,\*</sup>, Liang Chen<sup>2,\*</sup>, Qikai Tang<sup>3,\*</sup>, Wei Wei<sup>2,\*</sup>, Yuan Cao<sup>4</sup>, Jiaheng Xie<sup>5</sup>, Jing Ji<sup>1</sup>

<sup>1</sup>College of Pharmacy, Jiangsu Ocean University, Lianyungang, Jiangsu, China

<sup>2</sup>Department of General Surgery, Fuyang Hospital Affiliated to Anhui Medical University, Fuyang, Anhui, China

<sup>3</sup>Department of Neurosurgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

<sup>4</sup>Fourth School of Clinical Medicine, Nanjing Medical University, Nanjing, Jiangsu, China

<sup>5</sup>Department of Burn and Plastic Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

\*These authors are the joint first authors

Correspondence to: Jing Ji; email: [jjing@jou.edu.cn](mailto:jjing@jou.edu.cn)

Keywords: cancer, biomarker, immunotherapy, pan-cancer analysis, GTPBP

Received: October 14, 2021

Accepted: March 1, 2022

Published: March 23, 2022

Copyright: © 2022 Hu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Background:** At present, cancer is still one of the principal diseases to represent a serious danger to human health. Although research on the pathogenesis and treatment of cancer is progressing rapidly, the current knowledge on this topic is far from sufficient. Some tumors with poor prognoses lack effective prognostic biomarkers.

**Methods:** Firstly, the Wilcoxon test was used to analyse the expression of GTPBP1-GTPBP10 in cancerous and normal tissues. Subsequently, we explored the expression of GTPBP1-10 in cancer by way of a paired *t*-test and plotted the survival curve using KM and univariate Cox regression analysis to explore the relationship between GTPBP1-10 and the prognosis of cancer. We then explored the significance of the GTPBP family in the tumor microenvironment.

**Results:** The results showed that many members of the GTPBP family are differentially expressed in a variety of cancers and alter the prognosis of a number of cancers. Members of the GTPBP family may serve as novel prognostic markers for these tumors. Moreover, members of the GTPBP family are correlated with the immune microenvironment of tumors, which is valuable in terms of adding to our understanding of the mechanisms of tumor genesis. Finally, we identified drugs showing a high correlation with GTPBP family members, which are therefore conducive to the development of GTPBP family member-based treatment regimens.

**Conclusions:** The 10 members of the GTPBP family have prognostic value in multiple tumor types and are associated with the immune microenvironment. Our study may provide a reference for the diagnosis and treatment of tumors.

## INTRODUCTION

Cancer, as one of the principal diseases endangering the health of all mankind, has placed a heavy burden on the development of the social economy and has had a profound effect on people's everyday lives [1]. Cancer cells replicate unrestrictedly and, by activating multiple signaling pathways, promote the growth of blood

vessels and the reediting of the immune system to form a microenvironment suitable for tumor growth [2]. Exploring the physiology of cancer and the cancer-immune microenvironment can lead to an improvement in patient prognoses. Most cancers lack effective biomarkers [3]. Therefore, it is of great significance to search for effective prognostic indicators and explore their roles in cancer physiology and the tumor microenvironment.

GTP was previously thought to be a form of cellular energy storage, and the function of GTPase was previously thought to be mainly related to energy supply [4]. However, studies in the past 50 years have revealed a number of additional functions [4]. Some GTP-binding proteins (GTPBP) with GTPase activity—including Ras, Rho and Rab—have been shown to play an important role in cancer [5]. This suggests that GTPBP may be a key factor in human diseases, particularly cancer [5]. GTPBP1-10 is a class of proteins exhibiting GTPase activity, but the research on them remains unclear [6]. It has been suggested that they may be related to mRNA monitoring and quality control of ribosome translation [7]. However, there has hitherto been no systematic study conducted on the role of GTPBP1-10 in various cancer types.

In this study, we conducted a pan-cancer exploration of GTPBP1-10, including expression analysis, survival analysis, immune infiltration analysis, methylation correlation analysis, and prediction of GTPBP4-related transcription factors. Our results may provide new ideas for the future diagnosis and treatment of cancer.

## METHODS

### Data download

From the Xena Browser website (<https://xenabrowser.net/datapages/>) we downloaded data needed for analysis, including RNA-Seq (HTSeq-FPKM), clinical data (phenotype and survival data), immune subtypes, DNA methylation induced tumor stem cell properties (DNAss), and mRNA induced tumor stem cell properties (RNAss) data. We downloaded data of the following 33 cancer types: Acute Myeloid Leukemia (LAML, 151 samples), Adrenocortical carcinoma (ACC, 79 samples), Cholangiocarcinoma (CHOL, 45 samples), Bladder Urothelial Carcinoma (BLCA, 430 samples), Breast invasive carcinoma (BRCA, 1217 samples), Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC, 309 samples), Colon adenocarcinoma (COAD, 512 samples), Uterine Corpus Endometrial Carcinoma (UCEC, 583 samples), Esophageal carcinoma (ESCA, 173 samples), Glioblastoma multiforme (GBM, 73 samples), Head and Neck squamous cell carcinoma (HNSC, 546 samples), Kidney Chromophobe (KICH, 89 samples), Kidney renal clear cell carcinoma (KIRC, 607 samples), Kidney renal papillary cell carcinoma (KIRP, 321 samples), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC, 48 samples), Liver hepatocellular carcinoma (LIHC, 424 samples), Brain Lower Grade Glioma (LGG, 529 samples), Lung adenocarcinoma (LUAD, 585 samples), Lung squamous cell carcinoma (LUSC, 550 samples), Skin Cutaneous Melanoma (SKCM, 472

samples), Mesothelioma (MESO, 86 samples), Uveal Melanoma (UVM, 80 samples), Ovarian serous cystadenocarcinoma (OV, 379 samples), Pancreatic adenocarcinoma (PAAD, 182 samples), Pheochromocytoma and Paraganglioma (PCPG, 186 samples), Prostate adenocarcinoma (PAD, 551 samples), Rectum adenocarcinoma (READ, 177 samples), Sarcoma (SARC, 265 samples), Stomach adenocarcinoma (STAD, 407 samples), Testicular Germ Cell Tumors (TGCT, 156 samples), Thymoma (THYM, 121 samples), Thyroid carcinoma (THCA, 568 samples), Uterine Carcinosarcoma (USC, 56 samples). We used the CellMiner website (<https://discover.nci.nih.gov/cellminer/home.do>) to download the transcriptome data and drug sensitivity data.

### Pan-cancer analysis

We first selected only those tumor types with a sample size of more than 5; then we analyzed the significance of the difference in expression of GTPBP1-10 between cancer and normal tissues using the Wilcoxon test. To calculate the fold change, the mean amount of expression in the tumor sample was divided by the mean amount of expression in the normal sample; we presented this in the form of a heat map, with high expression shown in red and low expression in blue. Subsequently, we explored the expression of GTPBP1-10 in cancer via a paired *t*-test and plotted the survival curve using KM and univariate Cox analysis to explore the relationship between GTPBP1-10 and the prognosis for cancer.

### Analysis of tumor microenvironment

The levels of stromal cell and immune cell infiltration in the tumor microenvironment were analyzed using R software, according to the immune score and stromal score of the transcriptome expression matrix. The characteristics of tumor stem cells and the level of immune infiltration in the tumor microenvironment were analyzed using the data of tumor stem cell characteristics downloaded from UCSC and the related data of immune subtypes (C1, C2, C3, C4, C5 and C6). The Spearman correlation coefficient was then used to further analyze the correlation between the mRNA expression of the GTPBP family and tumor stem cell properties.

### Prediction of transcription factors

Potential transcription factors of GTPBP4 were predicted by multiple sites—including CHEA, Encode, Jaspar, MotifMap, Transfac and Trurust—and potential transcription factors whose total predicted positive results from the sites were recorded. Then, we uploaded the prediction results to Cytoscape for

visualization. Next, we predicted the potential promoter region of GTPBP4. The sequence location of the sense strand of GTPBP4 is 988434-1019932, and the upstream from -2000 to -1900 of GTPBP is the potential promoter region. The predicted promoter region is 986434~988533. On the Jaspas website (a database of transcription factor binding profiles), we compared the sequence of the potential promoter of the target gene to the transcription factors previously predicted by three or more sites, and we set a threshold of 90%.

### Prediction of drug sensitivity associated with the GTPBP family in cancers

We used the DREIMT database to explore the association between drug sensitivity and GTPBP family members in various cancers, so as to provide a reference for GTPBP-based treatment.

### Correlation analysis of GTPBP family and mTOR pathway

Since the GTPBP family is also involved in the mTOR pathway, we then analyzed the link between these two. The genes relating to the mTOR signaling pathway were downloaded from the GSEA website (<https://www.gsea-msigdb.org>). The Spearman correlation test was used to analyze the correlation between the GTPBP family and the mTOR pathway.  $P < 0.05$  was considered statistically significant.

### Statistical analysis

The chi-squared and Wilcoxon tests were used for exploring the differences between different groups. Cox regression analysis was used to examine the degree of risk and the prognosis of gene expression in each cancer type. The Spearman correlation was used to test the correlation between GTPBP and the immune score, interstitial score, etc. The statistical significance is defined when the two-paired  $p < 0.05$ .

### Data availability statement

The Xena Browser website (<https://xenabrowser.net/datapages/>); The CELLMINER website (<https://discover.nci.nih.gov/cellminer/home.do>).

### Availability of supporting data

All data is available in UCSC, TCGA and other databases.

## RESULTS

The flow chart of our work is shown in Figure 1.

### Ridgeline plot revealed the central trend of GTPBPs in the dataset

The ridgeline plot showed the fluctuation of the expression of GTPBP in the data (Figure 2A). The gradient of a mountain represents the degree of

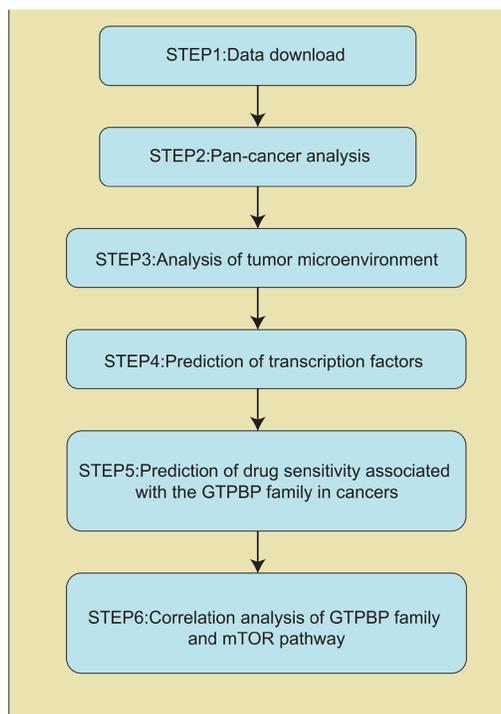


Figure 1. Schematic representation of the experimental design.

discreteness between different sets of data. The steeper the mountain, the more concentrated the data distribution is, and the fewer discrete values there are in the data. The less steep the mountain, the more fragmented the data. Our study found that the dispersion of GTPBP expression in the data set was not high, which was of great research value.

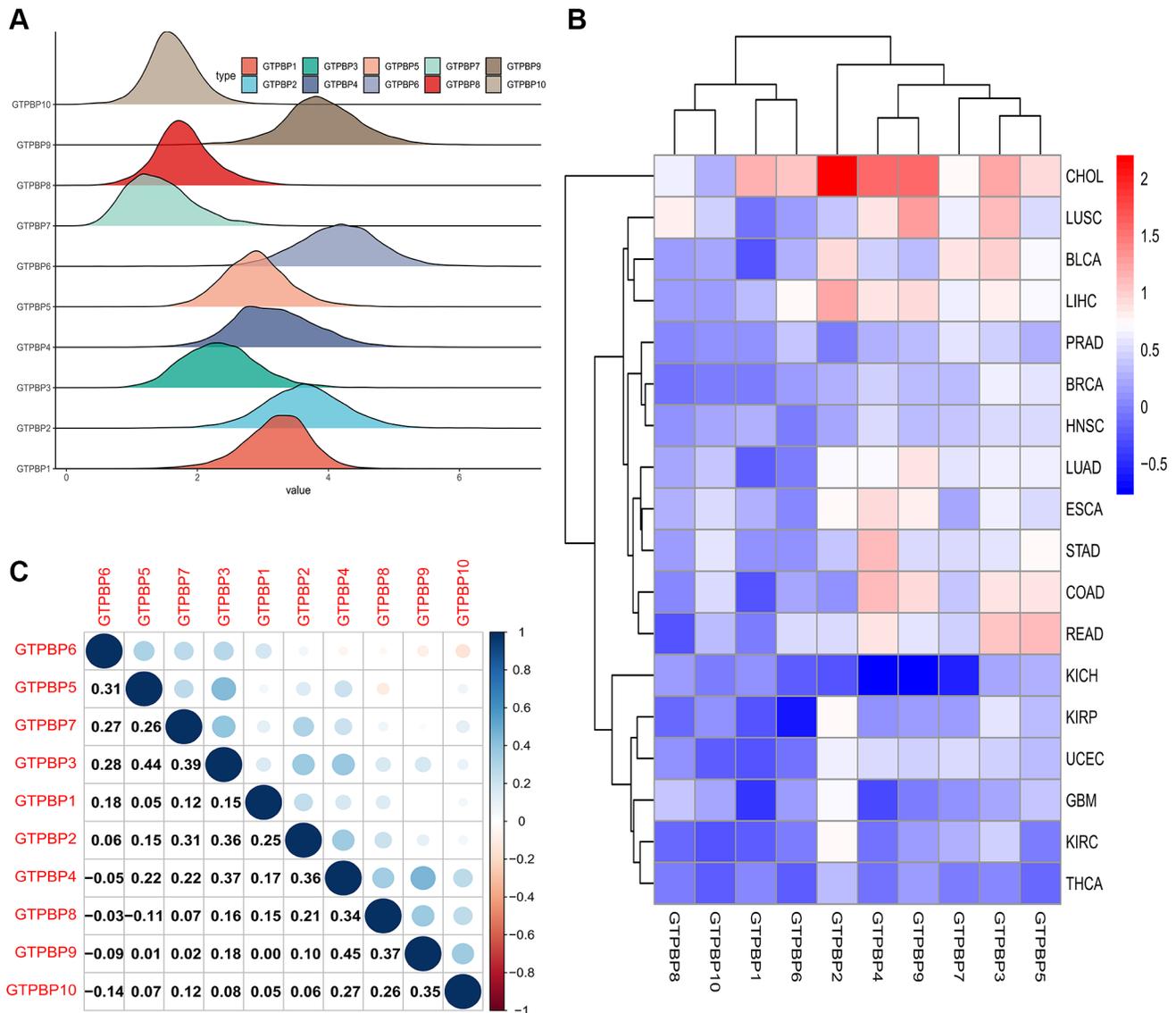
### Overview of the expression of GTPBP family in multiple cancers and normal tissues

Since the above analysis revealed the value of further exploration of GTPBP family members, we conducted

expression analysis in order to clarify their expression in tumors. The heat map provides an overview of the differential expression of the GTPBP family in each cancer compared to normal tissue (Figure 2B). Red represents upregulated expression, blue represents downregulated expression, and the shade of the color indicates the degree of difference.

### Analysis of correlations between members of the GTPBP family in multiple cancers

Next, we conducted correlation analysis among members of the GTPBP family to explore their modes



**Figure 2. Expression of GTPBP family across cancers and the relationship between the GTPBP members.** (A) The Ridgeline Plot showed the fluctuation of GTPBPs expression in the data: The steeper the mountain, the more concentrated the data distribution is, and there are few discrete values between the data. The flatter the mountain, the more fragmented the data. (B) The heat map of the expression of the GTPBP family in each cancer compared to normal tissue: red represents upregulated expression, blue represents downregulated expression, and the shade of the color represents the degree of difference. (C) The graph of the correlation values among the GTPBP members.

of action. Through the Spearman correlation test, we can calculate the correlation between each member of the GTPBP family and visualize it with the use of R software. The correlation graph shows the correlation values among the members (Figure 2C). This is meaningful for us to more fully understand the GTPBP family's mode of action.

### Expression of GTPBP family across cancers

In order to further investigate the significance of GTPBP in various tumors, we analyzed the expression of the GTPBP family in a range of tumors (Figure 3A–3J). We found that GTPBP protein was differentially expressed in a variety of cancers, such as BLCA, CHOL

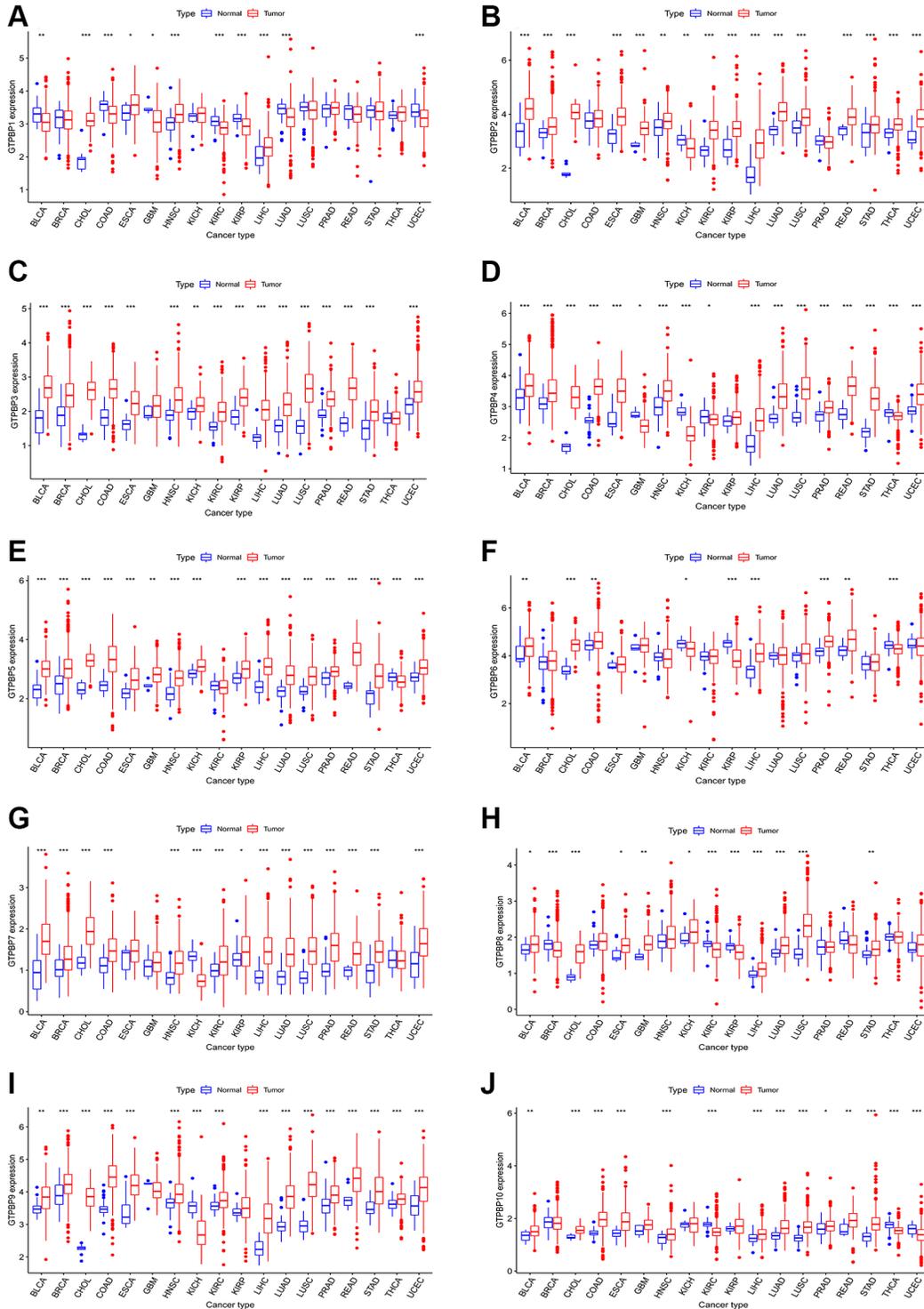


Figure 3. Expression of GTPBP family across cancers. (A–J) The expression of GTPBP family in cancers compared with normal tissues.

and LUAD. Moreover, their expression in cancer mainly trends towards upregulation, suggesting that they may play an important role in cancer and may indeed promote it.

### Survival analysis

Since members of the GTPBP family are differentially expressed in multiple tumors, we subsequently performed survival analysis to assess their prognostic value. We found GTPBP expression to be associated with the prognoses of various types of cancer (Figures 4 and 5). Tumors of prognostic value to members of the

GTPBP family—and their survival curves—are shown above. We can see that the GTPBP family is a potent prognostic marker for a variety of tumors. Subsequently, univariate Cox regression was used to analyze the survival of the GTPBP family in cancers; these results also revealed the GTPBP family to be a prognostic marker for a variety of cancers (Figure 6).

### Immune infiltration analysis

The tumor microenvironment plays an important role in the genesis and prognosis of tumors. In order to

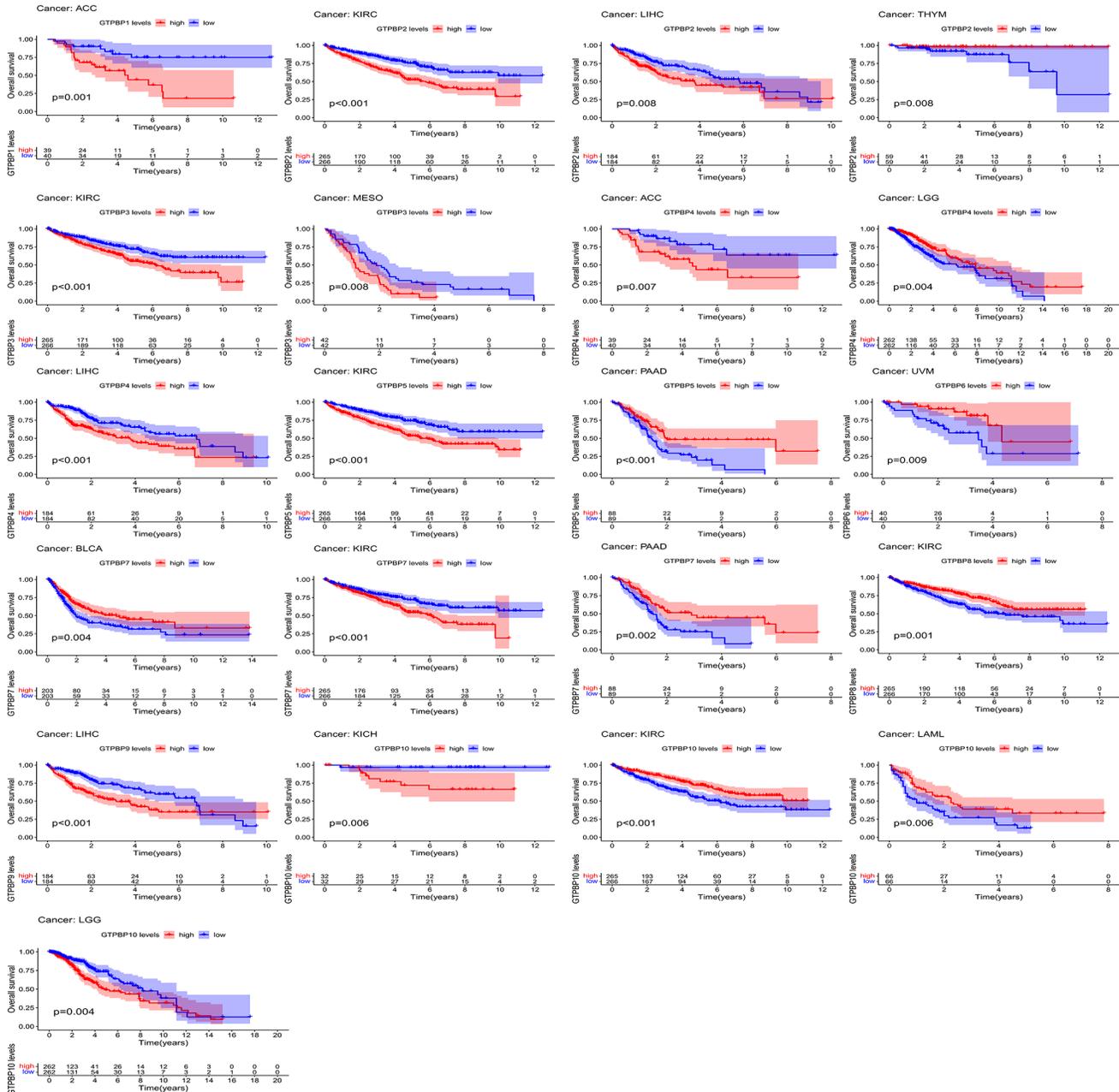


Figure 4. Survival analysis revealed that GTPBP expression is associated with the prognosis of various types of cancer ( $p < 0.05$ ).

further understand the mechanism and significance of the GTPBP family, we conducted immune microenvironment analysis. We found that the expressions of GTPBP1-10 were different in different immune subtypes ( $P < 0.05$ , Figure 7A), suggesting that GTPBP protein may be related to immunity. In addition, we observed a correlation between GTPB1-10 and the immune score (Figure 7B). Most of the GTPBP proteins displayed a negative correlation with the immune score in cancer, while GTPBP1, 2 and 8 showed a positive correlation with the immune score

in cancer. We also found that GTPBP1-10 was most negatively correlated with stromal scores in cancer (Figure 7C), while being positively correlated with RNAss and DNAss (Figure 7D, 7E). The expressions of GTPBP1-10 in immune subtype activation and proliferation are shown in Supplementary Figure 1. The results showed that GTPBP1, GTPBP2, GTPBP3, GTPBP4, GTPBP6, GTPBP7, GTPBP9 and GTPBP10 were differentially expressed in the activation and proliferation of different immune subtypes ( $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ).

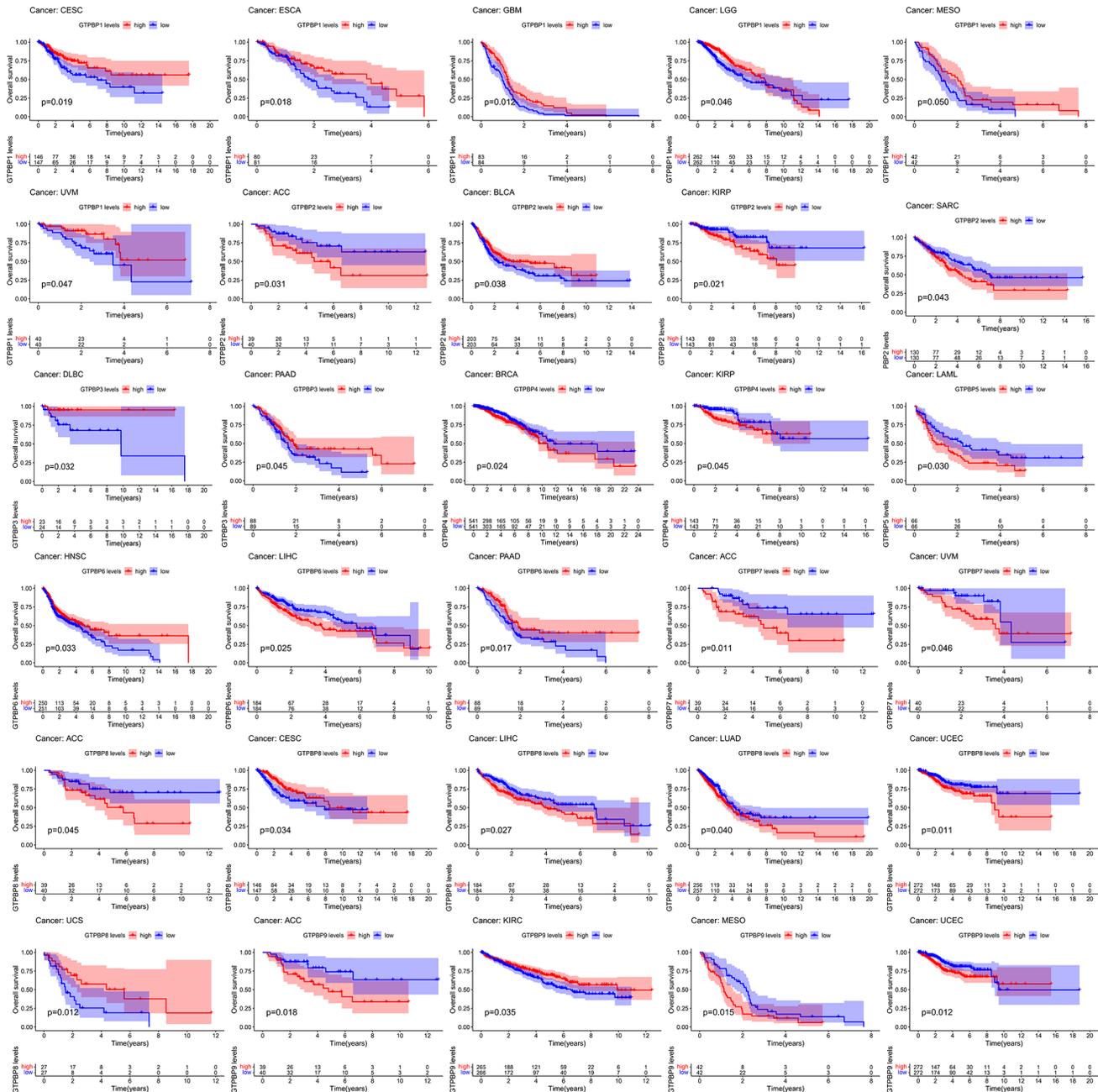


Figure 5. Survival analysis revealed that GTPBP expression is associated with the prognosis of various types of cancer ( $p < 0.05$ ).

## Prediction of drug sensitivity associated with the GTPBP family in cancers

In order to facilitate the development of therapies based on the GTPBP family, we performed drug sensitivity analysis to identify the drugs most associated with the GTPBP family. The results showed that many drugs are related to the GTPBP family; these findings are a useful reference for GTPBP family-based therapy. Among others, Nelarabine, Vorinostat, Allopurinol, LEe-011 and Hydroxyurea were all found to be potentially effective drugs for GTPBP mutation (Figure 8).

## Predicting transcription factors of GTPBP4

Finally, we predicted the transcription factors of GTPBP4. JUN and FLI-1 can be detected by the three

algorithms as possible transcription factors of GTPBP4; these are marked in red in Figure 9A. Of the other potential transcription factors, those that can be detected by two algorithms are labeled in pink, and those that can only be detected by one algorithm are labeled in blue. The correlation of GTPBP4 with FLI-1 (Figure 9B) and JUN (Figure 9C) was also investigated. Furthermore, we constructed the gene modules of FLI-1 (Figure 9D) and JUN (Figure 9E). Since JUN's correlation was stronger, we predicted JUN's promoter sequence in the GTPBP gene (Figure 9F).

## Correlation analysis of GTPBP family and mTOR pathway

Since the GTPBP family is also involved in the mTOR pathway, we then analyzed the link between the two.

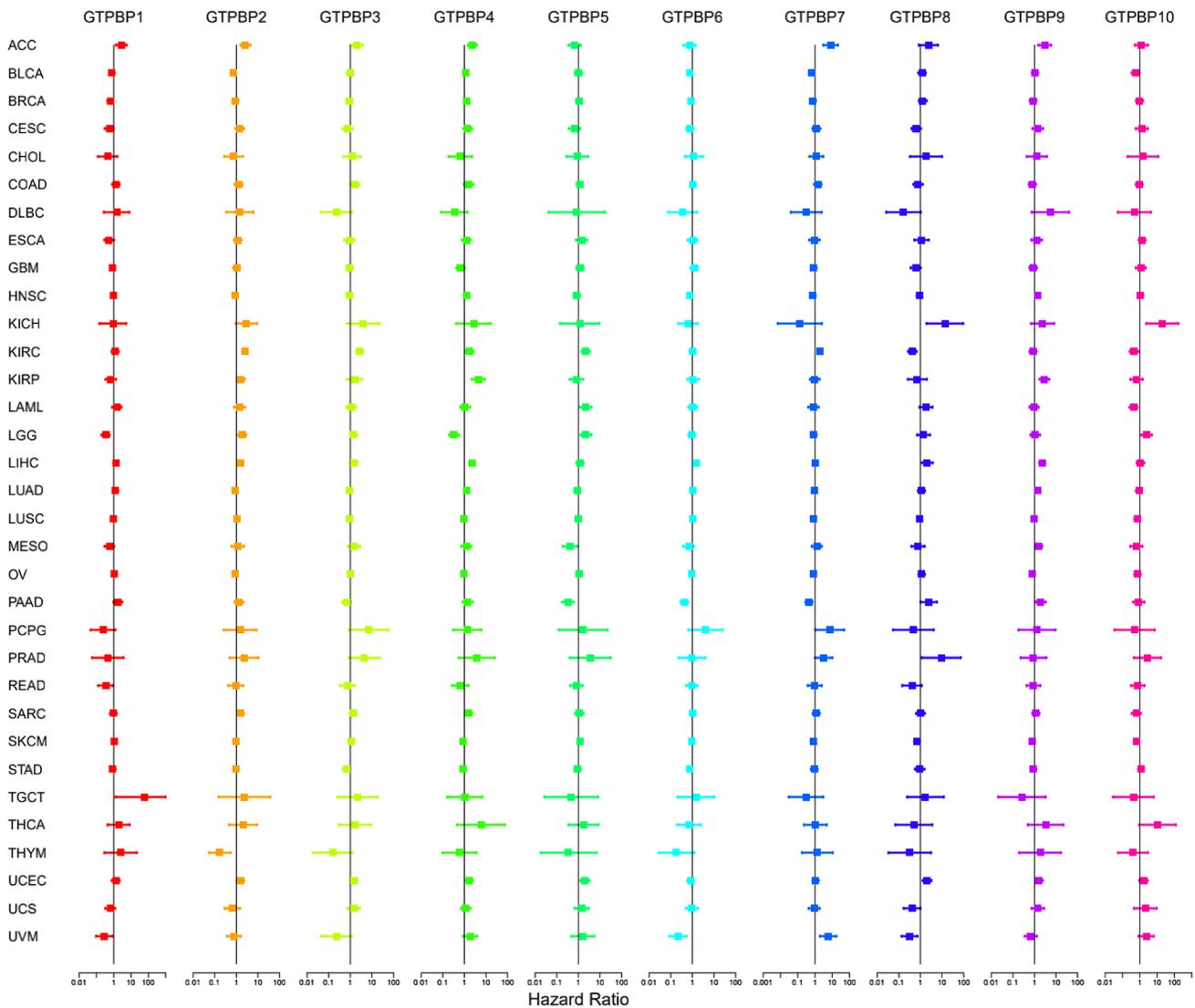


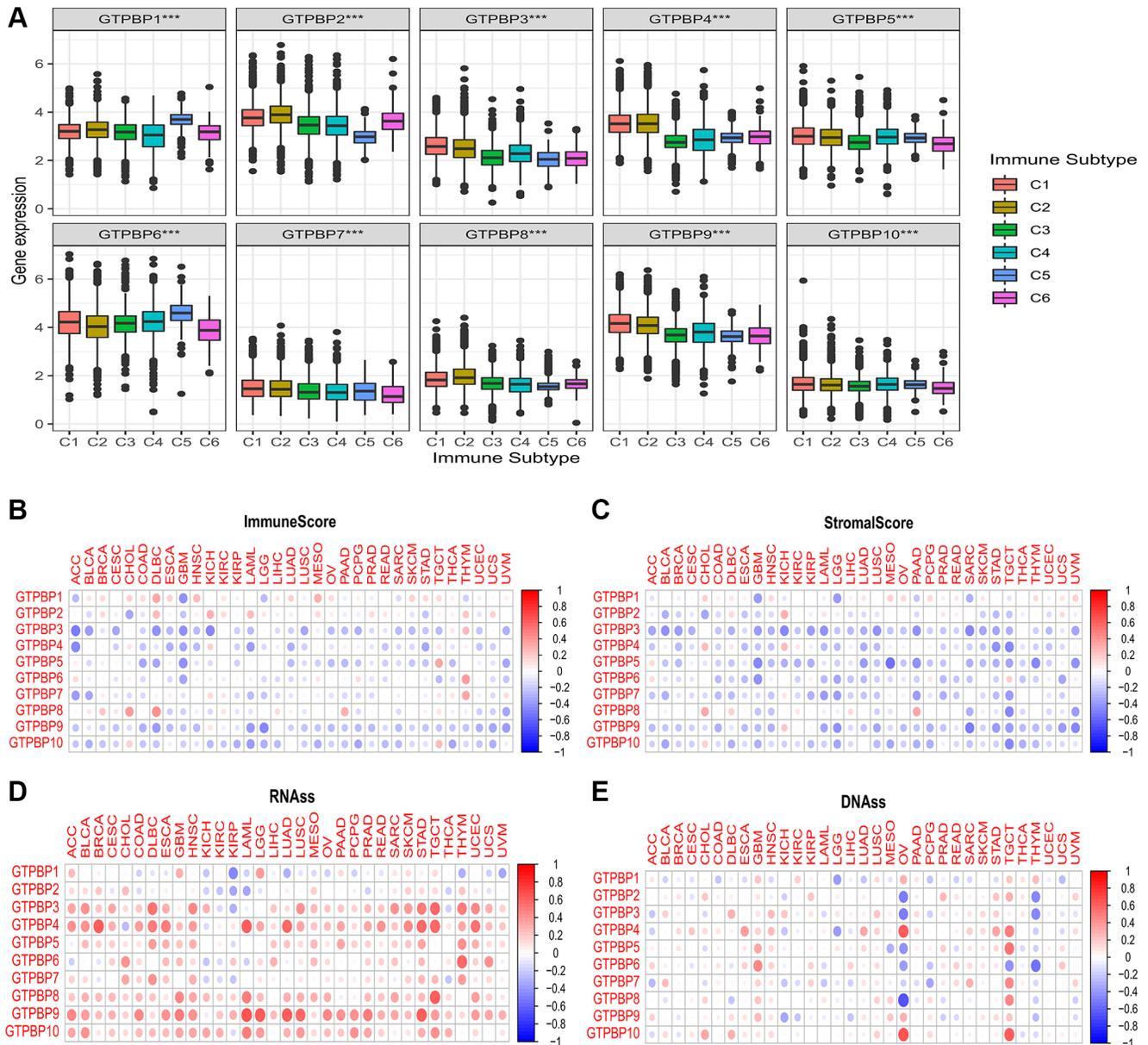
Figure 6. Univariate Cox regression was used to analyze the survival of the GTPBP family in cancers, and the results showed that the GTPBP family is a prognostic marker for a variety of cancers.

The results showed that GTPBP1, GTPBP2, GTPBP3, GTPBP4, GTPBP7, GTPBP8, GTPBP9 and GTPBP10 were correlated with the mTOR signaling pathway (Figure 10).

## DISCUSSION

The complex mechanisms in cancer cells have been extensively studied by researchers, but the available information is still far from sufficient [8]. The highly

metabolically active nature of cancer, the over-activation of growth signals, and the unrestricted replication of cells have led us to ask the following questions: what on earth promotes the occurrence of cancer [9–11]? How do you kill cancer cells by inhibiting these mechanisms? In addition, tumor-induced changes in the immune system are equally complex [12]. Cancer-induced immune-reediting enables tumor cells to acquire immune tolerance and contributes to the development of drug resistance [13].



**Figure 7. Significance of the GTPBP family in the tumor microenvironment.** (A) The expressions of GTPBP1-10 were different in different immune subtypes ( $P < 0.05$ ), suggesting that GTPBP protein may be related to immunity. (B) The correlation between GTPBP1-10 and immune score: Red represents positive correlation, blue represents negative correlation. (C) The correlation between GTPBP1-10 and stromal score: Red represents positive correlation, blue represents negative correlation. (D) The correlation between GTPBP1-10 and mRNAs induced tumor stem cell properties (RNAss). (E) The correlation between GTPBP1-10 and DNA methylation induced tumor stem cell properties (DNAss).

The exploration of the immune microenvironment of tumors is well under way. How to awaken the sleepy immune system has become a hot topic in cancer treatment.

In this era of precision medicine for cancer, precision biomarkers are needed [14]. These biomarkers often play an important role in the physiology of cancer, and changes in their expression can affect the prognoses of patients [14]. Therefore, it is of profound significance to locate these valuable biomarkers and to explore their functions and behavioural patterns.

In our study, we performed a pan-cancer analysis of the GTPBP family. The results showed that 10 members of

the GTPBP family were expressed differently in many cancers, such as urothelial carcinoma, invasive ductal carcinoma, bile duct carcinoma and colon carcinoma, with a trend for high expression. Accordingly, their expression may guide the prognosis of several of these cancers. For example, the high expression of GTPBP4 is a poor prognostic factor for clear cell renal cell carcinoma and hepatocellular carcinoma ( $P < 0.05$ ). Some of these prognostic markers remained significant after univariate Cox regression. Moreover, GTPBP1-10 were differentially expressed in different immune subtypes and correlated with the immune score, stromal score, RNASS and DNASS of various cancers. Since GTPBP4 has been reported as being of extreme significance in hepatocellular carcinoma, we further

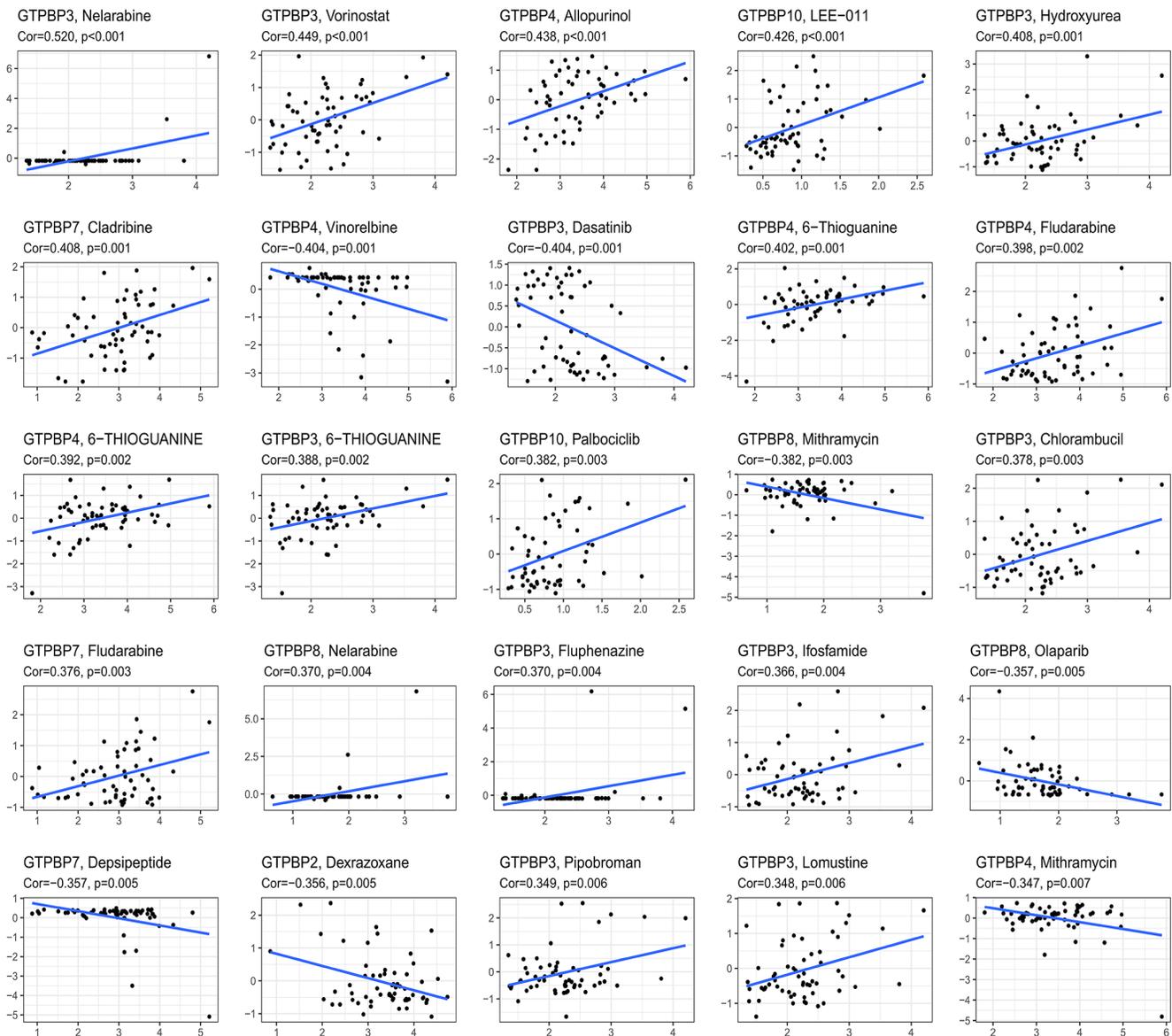
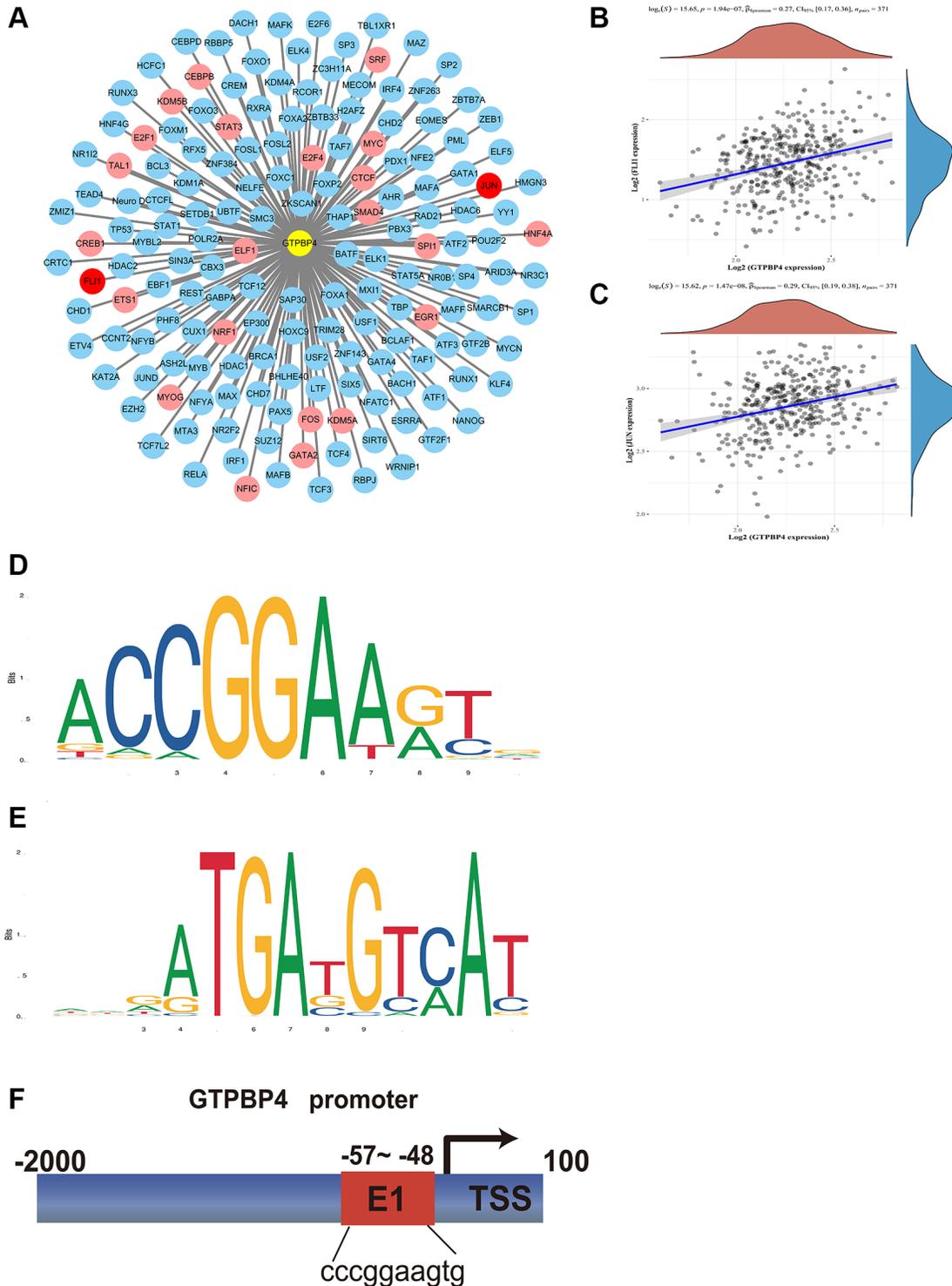


Figure 8. Nelarabine, Vorinostat, Allopurinol, LEE-011, Hydroxyurea, etc., are all potentially effective drugs for GTPBP family mutation (Figure 7).

studied the transcription factors of GTPBP4. Firstly, we constructed a schema map of the potential transcription factors of GTPBP4; it was found that JUN and FLI-1

could be detected by all three algorithms, indicating that JUN and FLI-1 are highly likely to be transcription factors of GTPBP4.



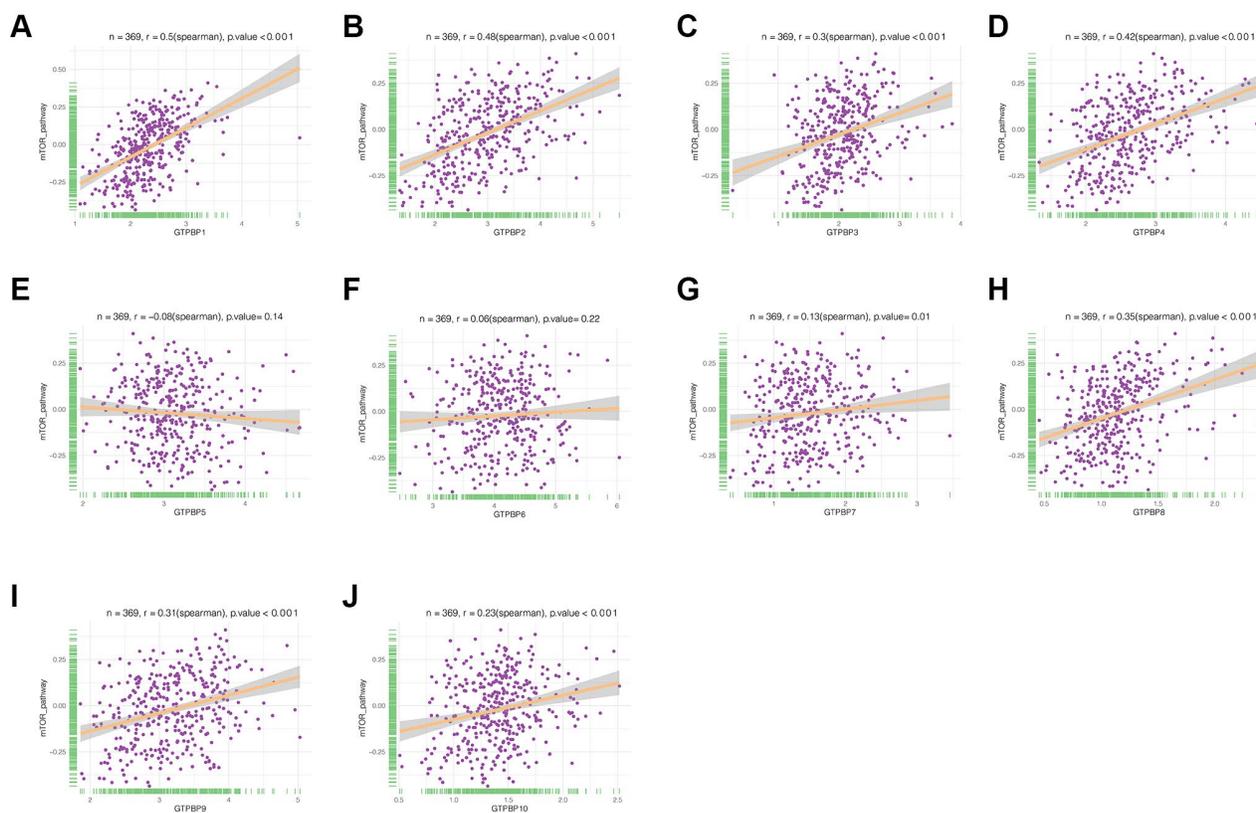
**Figure 9. Transcription factors analysis of GTPBP4.** (A) Potential transcription factors of GTPBP4 were predicted by multiple sites including CHEA, Encode, Jaspas, MotifMap, Transfac, and Trustrust, and potential transcription factors whose total predicted positive results of the sites were recorded: Three sites measured transcription factors in red, two sites in pink, and one site in blue. (B) Correlation analysis of GTPBP4 and FLI-1. (C) Correlation analysis of GTPBP4 and JUN. (D) Sequence motif construction of transcription factor FLI-1. (E) Sequence motif construction of transcription factor JUN. (F) Promoter site prediction of GTPBP4.

Our study identified a new class of markers of significance to 18 types of cancer: GTPBP1-10. They are helpful for guiding the prognosis of cancer and boosting the progress of treatment. Cancer seriously hinders social and economic development as well as negatively affecting people's quality of life. New biomarkers are urgently needed. Most of these 18 types of cancer have a relatively poor prognosis and are often accompanied by distant metastasis. Therefore, finding an effective marker and exploring its mechanism of action may be instrumental in improving this situation.

Immunoediting is one of the hallmarks of cancer [15]. It is precisely because cancer can inhibit host immunity and promote the immune tolerance of tumor cells that cancer can continue to grow without being killed by human immune cells [16]. Therapeutic regimens based on tumor immunity are progressing well. Cytokines such as interferon and interleukin have proven effective in the treatment of many tumors by regulating their immune function [17]. In addition, the discovery of immune checkpoints was a major development. Immune checkpoint-based immunotherapies such as PD-1/PD-L1 inhibitors have improved outcomes in patients with a variety of tumors (melanoma, for instance) [18]. Therefore, having found an effective prognostic marker, it is necessary to explore its

significance in tumor immunity. Our study revealed a strong association between GTPBP1-10 and six immune subtypes, which could have significant implications for GTPBP-based immunotherapy. At the same time, GTPBP1-10-related immune scores and stromal scores in different types of tumors help to more clearly define the immune landscape.

The tumor is a highly heterogeneous group of cells [19]. Genetic, epigenetic and tumor microenvironment heterogeneity are the main sources of tumor heterogeneity [20]. It is of great value to study the heterogeneity of tumors and explore potential treatments. Currently, cancer stem cells (CSCs) are recognized as a group of self-renewing and differentiated cells in tumors [21]. The reprogramming of CSCs leads to a change in the degree of differentiation and therefore may be a source of tumor heterogeneity [22]. By targeting CSCs, the plasticity and heterogeneity inherent in these cells can be successfully overcome; consequently, it represents a promising therapeutic strategy [23]. Therefore, we also investigated the role of GTPBP1-10 in tumor stem cell properties, including mRNA induced tumor stem cell properties (RNAss) and DNA methylation induced tumor stem cell properties (DNAss). Our results could serve as a reference for future CSC-based therapies.



**Figure 10. Correlation analysis of GTPBP family and mTOR pathway.** (A–J) GTPBP1, GTPBP2, GTPBP3, GTPBP4, GTPBP7, GTPBP8, GTPBP9, GTPBP10 were correlated with mTOR signaling pathway.

Until now, many studies have suggested the significance of the GTPBP family in tumors without exploring it in depth. Liu et al. found that GTPBP2 expression was upregulated in NSCLC and correlated with lymph node metastasis [24]. Zhang et al. found that high GTPBP4 expression is a poor prognostic factor for lung adenocarcinoma [25]. It stands to reason, therefore, that the GTPBP family warrants further exploration in tumor research. GTPBP4 has now been described as having a possible role in the development of cancer, particularly liver cancer [26]. With this in mind, we predicted the transcription factor of GTPBP4 as well as its binding site. It is well known that the binding of promoters and transcription factors is the key element of gene transcription and the basis of gene expression [27]. The prediction of a transcription factor of a key gene not only enriches our understanding of its mode of action but also facilitates the development of transcription-based therapies, providing an alternative approach to cancer treatment.

In summary, our study explored the role of the GTPBP family in 18 types of cancer. The results of this study can not only provide good prognostic indicators for these tumors but also facilitate further exploration of the immune microenvironment based on the GTPBP family and the characteristics of tumor stem cells, which is undoubtedly beneficial for the treatment of tumors. In addition, we predicted the transcription factor of GTPBP4, enriching our understanding of its function in the process. However, our study also has some limitations; we lack *in vivo* and *in vitro* experiments to verify the function of GTPBP1-10. We shall strive to expand the scope of our study in the future.

## Abbreviations

LAML: Acute Myeloid Leukemia; ACC: Adrenocortical carcinoma; CHOL: Cholangiocarcinoma; BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD: Colon adenocarcinoma; UCEC: Uterine Corpus Endometrial Carcinoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; LIHC: Liver hepatocellular carcinoma; LGG: Brain Lower Grade Glioma; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; SKCM: Skin Cutaneous Melanoma; MESO: Mesothelioma; UVM: Uveal Melanoma; OV: Ovarian serous cystadenocarcinoma; PAAD: Pancreatic

adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SARC: Sarcoma; STAD: Stomach adenocarcinoma; TGCT: Testicular Germ Cell Tumors; THYM: Thymoma; THCA: Thyroid carcinoma; USC: Uterine Carcinosarcoma.

## AUTHOR CONTRIBUTIONS

Yiming Hu, Jiaheng Xie, Liang Chen: acquisition of data, analyzing and interpretation of data, drafting the article; Yiming Hu, Yuan Cao, Jing Ji: designing, revising, and guiding the study. The authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

We are very grateful for data provided by databases such as TCGA, GEO.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

## FUNDING

This Research was funded by 2021 Graduate Research and Practice Innovation Program of Jiangsu Ocean University (Nos.KYCX2021-021), the Natural Science Foundation of China (Nos.82104174), the Open Foundation of Jiangsu Key Laboratory of Marine Pharmaceutical Compound Screening (Nos. HY202103) and the Natural Science Foundation of Jiangsu Higher Education Institutions of China (Nos.20KJB350008).

## REFERENCES

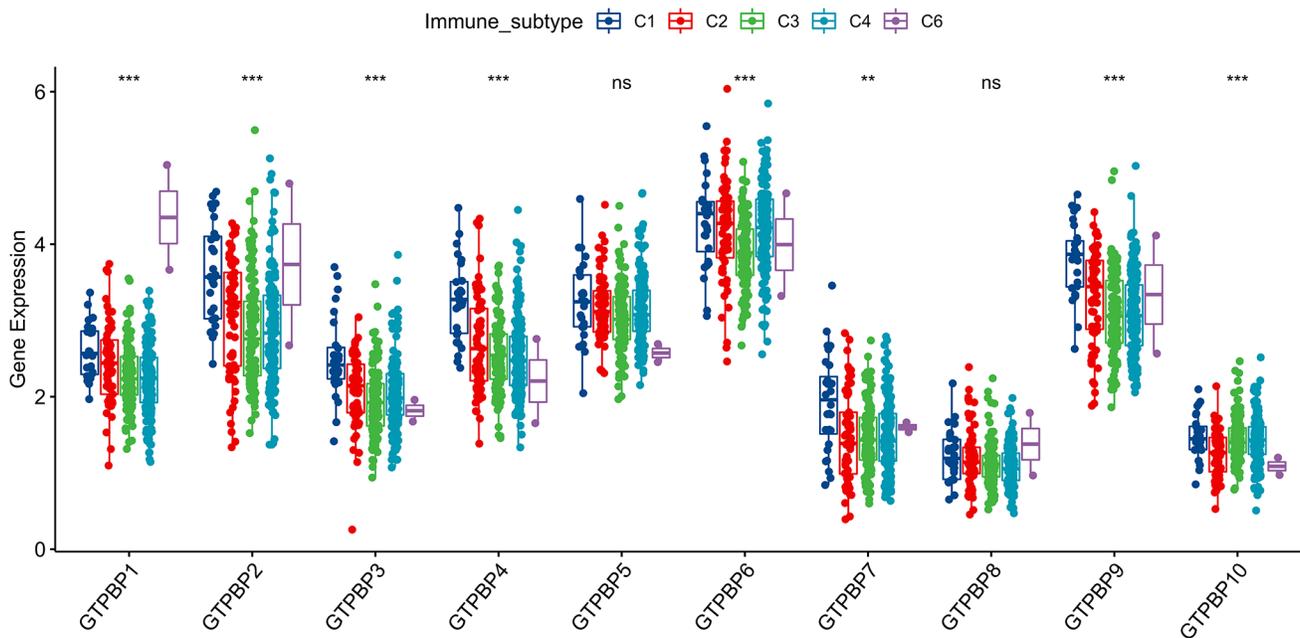
1. Cross D, Burmester JK. Gene therapy for cancer treatment: past, present and future. *Clin Med Res.* 2006; 4:218–27. <https://doi.org/10.3121/cmr.4.3.218> PMID:16988102
2. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000; 100:57–70. [https://doi.org/10.1016/s0092-8674\(00\)81683-9](https://doi.org/10.1016/s0092-8674(00)81683-9) PMID:10647931
3. Wu L, Qu X. Cancer biomarker detection: recent achievements and challenges. *Chem Soc Rev.* 2015; 44:2963–97. <https://doi.org/10.1039/c4cs00370e> PMID:25739971
4. Svensmark JH, Brakebusch C. Rho GTPases in cancer: friend or foe? *Oncogene.* 2019; 38:7447–56.

- <https://doi.org/10.1038/s41388-019-0963-7>  
PMID:[31427738](https://pubmed.ncbi.nlm.nih.gov/31427738/)
5. Kazanietz MG, Caloca MJ. The Rac GTPase in Cancer: From Old Concepts to New Paradigms. *Cancer Res.* 2017; 77:5445–51.  
<https://doi.org/10.1158/0008-5472.CAN-17-1456>  
PMID:[28807941](https://pubmed.ncbi.nlm.nih.gov/28807941/)
6. Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. *Physiol Rev.* 2001; 81:153–208.  
<https://doi.org/10.1152/physrev.2001.81.1.153>  
PMID:[11152757](https://pubmed.ncbi.nlm.nih.gov/11152757/)
7. Vetter IR. Interface analysis of small GTP binding protein complexes suggests preferred membrane orientations. *Biol Chem.* 2017; 398:637–51.  
<https://doi.org/10.1515/hsz-2016-0287>  
PMID:[28002022](https://pubmed.ncbi.nlm.nih.gov/28002022/)
8. Berger MF, Mardis ER. The emerging clinical relevance of genomics in cancer medicine. *Nat Rev Clin Oncol.* 2018; 15:353–65.  
<https://doi.org/10.1038/s41571-018-0002-6>  
PMID:[29599476](https://pubmed.ncbi.nlm.nih.gov/29599476/)
9. Cao Y, Xie J, Chen L, Hu Y, Zhai L, Yuan J, Suo L, Shen Y, Ye R, Li J, Gong Z, Dong Y, Bao W, et al. Construction and Validation of a Novel Pyroptosis-Related Gene Signature to Predict the Prognosis of Uveal Melanoma. *Front Cell Dev Biol.* 2021; 9:761350.  
<https://doi.org/10.3389/fcell.2021.761350>  
PMID:[34901006](https://pubmed.ncbi.nlm.nih.gov/34901006/)
10. Ying X, Chen L, Xie J, Hu Y, Wu Q, Cao L, Yu H. ANXA1 (Annexin A1) regulated by MYC (MYC proto-oncogene) promotes the growth of papillary thyroid carcinoma. *Bioengineered.* 2021; 12:9251–65.  
<https://doi.org/10.1080/21655979.2021.1996511>  
PMID:[34723715](https://pubmed.ncbi.nlm.nih.gov/34723715/)
11. Chen L, Lin YH, Liu GQ, Huang JE, Wei W, Yang ZH, Hu YM, Xie JH, Yu HZ. Clinical Significance and Potential Role of *LSM4* Overexpression in Hepatocellular Carcinoma: An Integrated Analysis Based on Multiple Databases. *Front Genet.* 2022; 12:804916.  
<https://doi.org/10.3389/fgene.2021.804916>  
PMID:[35096017](https://pubmed.ncbi.nlm.nih.gov/35096017/)
12. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity.* 2013; 39:1–10.  
<https://doi.org/10.1016/j.immuni.2013.07.012>  
PMID:[23890059](https://pubmed.ncbi.nlm.nih.gov/23890059/)
13. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, Lichtor T, Decker WK, Whelan RL, Kumara HMC, Signori E, Honoki K, Georgakilas AG, et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin Cancer Biol.* 2015; 35:S185–98.  
<https://doi.org/10.1016/j.semcancer.2015.03.004>  
PMID:[25818339](https://pubmed.ncbi.nlm.nih.gov/25818339/)
14. Zou J, Wang E. Cancer Biomarker Discovery for Precision Medicine: New Progress. *Curr Med Chem.* 2019; 26:7655–71.  
<https://doi.org/10.2174/0929867325666180718164712>  
PMID:[30027846](https://pubmed.ncbi.nlm.nih.gov/30027846/)
15. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity.* 2004; 21:137–48.  
<https://doi.org/10.1016/j.immuni.2004.07.017>  
PMID:[15308095](https://pubmed.ncbi.nlm.nih.gov/15308095/)
16. O'Donnell JS, Teng MWL, Smyth MJ. Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat Rev Clin Oncol.* 2019; 16:151–67.  
<https://doi.org/10.1038/s41571-018-0142-8>  
PMID:[30523282](https://pubmed.ncbi.nlm.nih.gov/30523282/)
17. Yang Y. Cancer immunotherapy: harnessing the immune system to battle cancer. *J Clin Invest.* 2015; 125:3335–7.  
<https://doi.org/10.1172/JCI83871>  
PMID:[26325031](https://pubmed.ncbi.nlm.nih.gov/26325031/)
18. Daassi D, Mahoney KM, Freeman GJ. The importance of exosomal PDL1 in tumour immune evasion. *Nat Rev Immunol.* 2020; 20:209–15.  
<https://doi.org/10.1038/s41577-019-0264-y>  
PMID:[31965064](https://pubmed.ncbi.nlm.nih.gov/31965064/)
19. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol.* 2018; 15:81–94.  
<https://doi.org/10.1038/nrclinonc.2017.166>  
PMID:[29115304](https://pubmed.ncbi.nlm.nih.gov/29115304/)
20. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature.* 2013; 501:328–37.  
<https://doi.org/10.1038/nature12624>  
PMID:[24048065](https://pubmed.ncbi.nlm.nih.gov/24048065/)
21. Dawood S, Austin L, Cristofanilli M. Cancer stem cells: implications for cancer therapy. *Oncology (Williston Park).* 2014; 28:1101–7.  
PMID:[25510809](https://pubmed.ncbi.nlm.nih.gov/25510809/)
22. Pan Y, Ma S, Cao K, Zhou S, Zhao A, Li M, Qian F, Zhu C. Therapeutic approaches targeting cancer stem cells. *J Cancer Res Ther.* 2018; 14:1469–75.  
[https://doi.org/10.4103/jcrt.JCRT\\_976\\_17](https://doi.org/10.4103/jcrt.JCRT_976_17)  
PMID:[30589025](https://pubmed.ncbi.nlm.nih.gov/30589025/)
23. Barbato L, Bocchetti M, Di Biase A, Regad T. Cancer Stem Cells and Targeting Strategies. *Cells.* 2019; 8:926.  
<https://doi.org/10.3390/cells8080926>  
PMID:[31426611](https://pubmed.ncbi.nlm.nih.gov/31426611/)

24. Jie L, Cong L, Conghui W, Ying G. GTPBP2 positively regulates the invasion, migration and proliferation of non-small cell lung cancer. *J Cancer*. 2021; 12:3819–26.  
<https://doi.org/10.7150/jca.48340>  
PMID:[34093790](https://pubmed.ncbi.nlm.nih.gov/34093790/)
25. Zhang Z, Wang J, Mao J, Li F, Chen W, Wang W. Determining the Clinical Value and Critical Pathway of GTPBP4 in Lung Adenocarcinoma Using a Bioinformatics Strategy: A Study Based on Datasets from The Cancer Genome Atlas. *Biomed Res Int*. 2020; 2020:5171242.  
<https://doi.org/10.1155/2020/5171242>  
PMID:[33134380](https://pubmed.ncbi.nlm.nih.gov/33134380/)
26. Liu WB, Jia WD, Ma JL, Xu GL, Zhou HC, Peng Y, Wang W. Knockdown of GTPBP4 inhibits cell growth and survival in human hepatocellular carcinoma and its prognostic significance. *Oncotarget*. 2017; 8:93984–97.  
<https://doi.org/10.18632/oncotarget.21500>  
PMID:[29212203](https://pubmed.ncbi.nlm.nih.gov/29212203/)
27. Zabidi MA, Stark A. Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet*. 2016; 32:801–14.  
<https://doi.org/10.1016/j.tig.2016.10.003>  
PMID:[27816209](https://pubmed.ncbi.nlm.nih.gov/27816209/)

## SUPPLEMENTARY MATERIALS

### Supplementary Figure



Supplementary Figure 1. The expressions of GTPBP1-10 in immune subtype activation and proliferation (\*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).

# F12 as a reliable diagnostic and prognostic biomarker associated with immune infiltration in papillary thyroid cancer

Jun-Hua Luo<sup>1</sup>, Xiu-Xia Zhang<sup>1</sup>, Wu-Hui Sun<sup>2</sup>

<sup>1</sup>Department of Thyroid and Breast Surgery, Linping Campus, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 311100, Zhejiang, China

<sup>2</sup>Department of Thyroid Surgery, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310009, Zhejiang, China

**Correspondence to:** Wu-Hui Sun; **email:** [sunwuhui@zju.edu.cn](mailto:sunwuhui@zju.edu.cn)

**Keywords:** coagulation factor XII, papillary thyroid cancer, biomarker, diagnosis, prognosis

**Received:** February 18, 2022

**Accepted:** April 14, 2022

**Published:** April 28, 2022

**Copyright:** © 2022 Luo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Objective:** To explore the function of coagulation factor XII (F12) in papillary thyroid cancer (PTC).

**Materials and Methods:** We assessed F12 expression and its relationship with overall survival (OS) in various cancers using TIMER and TISIDB databases. Further, we evaluated the mRNA and protein expression levels of F12 in PTC via different bioinformatics tools. The receiver operating characteristic (ROC) curve was applied to determine the diagnostic value of F12 in PTC. Then, the Kaplan-Meier plotter and Cox regression analyses were performed to examine the prognostic significance of F12. The possible mechanism of F12 in PTC was investigated through enrichment analyses. Finally, the correlation between F12 expression and immune cell infiltration was analyzed using TCGA data.

**Results:** This study revealed the clinical significance of F12 in various cancers. Higher mRNA ( $P < 0.001$ ) and protein expressions of F12 were observed in PTC compared with normal tissues. Besides, F12 expression exhibited high diagnostic performance in PTC and its overexpression served as an independent predictor for the poor OS ( $P < 0.05$ ). Enrichment analyses results showed that F12 was mainly involved in metabolism-associated pathways. Additionally, F12 expression was significantly linked to immune cell infiltration levels, especially macrophage infiltration.

**Conclusions:** F12 might be a reliable diagnostic and prognostic biomarker for PTC. Moreover, F12 expression might affect the OS of PTC patients via regulating metabolic pathways.

## INTRODUCTION

Thyroid cancer is one of the most common endocrine malignancies with increasing morbidity [1]. It is estimated that there were 95,030 newly diagnosed cases of thyroid cancer and 22,070 deaths in 1990, which increased to 255,490 new cases and 41,240 deaths in 2017 [2]. Papillary thyroid cancer (PTC) is the most predominant subtype of thyroid cancer, accounting for approximately 85% of all cases [3]. Radical surgery, thyroid-stimulating hormone suppression, and <sup>131</sup>I

therapy are the traditional approaches for the PTC treatment, displaying relatively satisfactory efficacy [4]. Although the 5-year survival rate of PTC patients is greater than 97% and the 10-year survival rate of such patients reached 85%, recurrence was found in roughly 15% and 28% of the patients within 10 years and 27 years, respectively after the initial treatment [5–8]. The majority of patients with PTC possess indolent progression, while overdiagnosis and overtreatment are common problems related to indolent diseases, which may increase the risk of injury to patients [9, 10]. Some

biomarkers such as HBME-1 and thyroglobulin have been used for the diagnosis of PTC in clinical practice, whereas, these biomarkers presented low specificity and sensitivity [11, 12]. Therefore, it is imperative to develop more effective and sensitive biomarkers for improving the diagnosis and prognosis of PTC.

Coagulation factor XII (F12) also called the Hageman factor, is a single-chain zymogen with a molecular weight of about 80kDa [13]. It is produced by hepatocytes in the liver and consists of a heavy chain (353 residues) and a light chain (243 residues) [14, 15]. F12 is a circulating serine protease, which is activated as F12a by plasma kallikrein and negatively charged surfaces [16, 17]. The contact system driven by F12 can be both prothrombotic via activation of F11 to trigger the intrinsic pathway of coagulation, and be proinflammatory by activating the kallikrein-kinin pathway to release the peptide bradykinin [18]. The binding of bradykinin to the G-protein-coupled kinin B2 receptor initiates intracellular signaling pathways that may accelerate angiogenesis and cancer growth [17, 19]. F12 has been demonstrated to be involved in various diseases. It colocalized with A $\beta$  plaques in the brain of patients with Alzheimer's disease and in brain lesions of those with multiple sclerosis [20, 21]. F12 was also found to participate in stroke development and the inhibition of F12 might offer a selective strategy to prevent this disease [22]. In addition, low F12 plasma levels were observed in patients with colorectal, gastrointestinal, and lung cancer [23–25]. However, little research has been done on the role of F12 in PTC.

Herein, we adopted TIMER and TISIDB databases to analyze the clinical significance of F12 in pan-cancer. Then, the expression analysis of F12 in PTC was performed through multiple bioinformatics tools. The correlation between F12 expression and PTC patient prognosis was analyzed using the TCGA data. The enrichment analyses were applied to reveal the potential molecular mechanism for F12 in PTC. Moreover, we assessed the association of F12 expression with immune cell infiltration levels. This study identified the potential role of F12 in PTC occurrence and progression and provided that F12 might be a reliable therapeutic target for PTC treatment.

## MATERIALS AND METHODS

### Expression and survival analyses of F12 in pan-cancer

TIMER (<https://cistrome.shinyapps.io/timer/>) is a comprehensive resource for the systematic analysis of immune infiltrates across diverse cancer types. The database was used to analyze the expression levels of

F12 in different kinds of tumors with Wilcoxon tests. Next, the associations between F12 expression and overall survival (OS) across human cancers were assessed using the TISIDB (<http://cis.hku.hk/TISIDB/>), which is an integrated repository portal for tumor-immune system interactions.

### The genetic status and expression level of F12 in PTC

Firstly, the cBioportal database (<https://www.cbioportal.org/>) was adopted to obtain the genetic alteration information by selecting “Papillary Thyroid Carcinoma (TCGA, Cell 2014)” with 496 total samples and entering “F12” for the query. Then, gene expression profiles and relevant clinical data of GDC TCGA-thyroid cancer were retrieved from the UCSC Xena (<https://xenabrowser.net/>). Totally 498 PTC samples with complete survival and expression data were enrolled for further analyses. In this dataset, the statistical significance of the F12 expression levels was evaluated in PTC tissues and adjacent normal tissues using unpaired and paired t-tests. The immunohistochemical images of F12 protein in PTC and healthy thyroid tissues were obtained from the HPA (<https://www.proteinatlas.org/>) database. The protein expression is scored with regard to staining intensity (strong, moderate, weak, or negative) and the fraction of stained cells (>75%, 25-75%, or <25%). Subsequently, all the included patients were divided into two groups according to the median expression of F12. The relationship between F12 expression and clinicopathological characteristics was analyzed using Pearson's chi-square test.

### Survival analysis

The diagnostic value of F12 in PTC was initially analyzed by calculating the area under the curve (AUC) of the receiver operating characteristic (ROC) curves. Using the “maxstat” packages in the R program, the 498 PTC patients were divided into high- and low- F12 expression groups by setting the best cutoff value as the parameter. Then, F12 mRNA expression and different population OS curves were plotted by the Kaplan-Meier method using “survival” packages in the R program. Log-rank tests were employed to evaluate the survival differences between two expression groups. The ROC curve analyses were also conducted to determine the value of F12 in predicting the survival status of the PTC patients. Moreover, univariate and multivariate Cox regression analyses were performed to assess the correlation between OS and clinical pathological factors in PTC patients. Male was a reference level for gender, while tumor-free for cancer status, and stage 1 for the stage.

## Identification of F12-related genes and enrichment analyses

Based on the median value of F12 expression, 498 PTC patients were classified into high- and low- F12 expression groups. Then, the “limma” package in R was used to identify the differentially expressed genes (DEGs) between the two expression groups. Genes with  $|\log_2 \text{foldchange (FC)}| > 1$  and  $P < 0.05$  were considered statistically significant for the DEGs. A PPI network of the top 200 DEGs was constructed in the STRING database (<https://cn.string-db.org/>) and subsequently was visualized by Cytoscape software. After that, the top 200 DEGs were subjected to gene ontology (GO) annotations and Kyoto Encyclopedia of Genes and Genomes pathway analyses using the R package “clusterProfile”. The GO annotations included biological process (BP), cellular component (CC), and molecular function (MF). The cutoff value for significant function and pathway screening was set at  $P < 0.05$  and false discovery rate (FDR)  $< 0.25$ .

Further, 498 PTC samples were divided into high and low expression groups using F12 expression median level as a cutoff criterion. The gene set enrichment analysis (GSEA) was performed to illustrate the significant survival difference between the two expression groups. The gene set was permuted 1000 times and the expression level of F12 was used as a phenotypic label. A nominal p-value  $< 0.05$  and an FDR q-value  $< 0.25$  were considered to be statistically significant.

## Immune cell infiltration analysis

The MCP-counter algorithm was employed to estimate the correlation of eight immune cells and two stromal cells' infiltration levels with F12 mRNA expression. In addition, the immune response of 22 tumor-infiltrating immune cells was measured to assess their association with F12 mRNA expression by the CIBERSORT which is a deconvolution algorithm based on gene expression [26]. Samples from TCGA were divided into high F12 (50%) and low F12 (50%) expression groups to compare the level of immune cell infiltration.

## Statistical analysis

All statistical analyses were performed in SPSS software (version 23.0) and packages of R (version 3.6.3). Image J software was used to quantify the F12 protein content in PTC and normal thyroid tissues. Pearson's correlation test was employed to evaluate the association of F12 expression with the immune cell infiltration levels and the expression of the markers of macrophages.  $P < 0.05$  was considered to be statistically significant.

## RESULTS

### The clinical value of F12 in pan-cancer

To identify the role of F12 in pan-cancer, the TIMER database was used to evaluate the expression levels of F12 in various kinds of cancers. Higher F12 expression was observed in most cancers including BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LUAD, LUSC, PRAD, READ, SKCM, STAD, THCA, and UCEC (all  $P < 0.01$ ) (Figure 1A). Then, we evaluated the association of F12 expression with OS across human cancers through the TISIDB database. As shown in Figure 1B, the increased expression of F12 led to a shorter OS time of patients with ACC, LUAD, MESO, SKCM, THCA, and UVM; however, LIHC patients with higher F12 expression had a favorable prognosis. These results indicated that F12 might play an important role in the occurrence and progression of various tumors including thyroid cancer.

### Genetic status and expression level of F12 in PTC

Due to its aberrant expression in thyroid cancer, we analyzed its possible change from the aspect of genetic alterations. We firstly retrieved the alteration frequency of F12 in PTC using the cBioportal website. As exhibited in Figure 2A, mRNA high and amplification were major status. By analyzing the copy number of F12 gene in PTC, diploid and gain were primary status, indicating the elevated levels of F12 (Figure 2B). But surprisingly, F12 was not mutated in PTC (Figure 2C).

Moreover, the TCGA data were applied to assess the F12 mRNA expression levels in PTC. The unpaired t-test ( $P = 4.9e-14$ ) (Figure 3A) and paired t-test ( $P = 1.3e-6$ ) (Figure 3B) presented higher F12 expression in PTC tissues than that in normal tissues. Following this, the protein levels of F12 were determined by the HPA database and the amount of F12 proteins was quantified by Image J software. As expected, higher F12 protein expression was observed in PTC tissue compared with the normal thyroid tissue. The staining was not detected and the intensity was weak in normal thyroid tissue, while medium staining and moderate intensity were found in PTC tissue (Figure 3C). These findings suggested that F12 overexpression might be essential in PTC tumorigenesis.

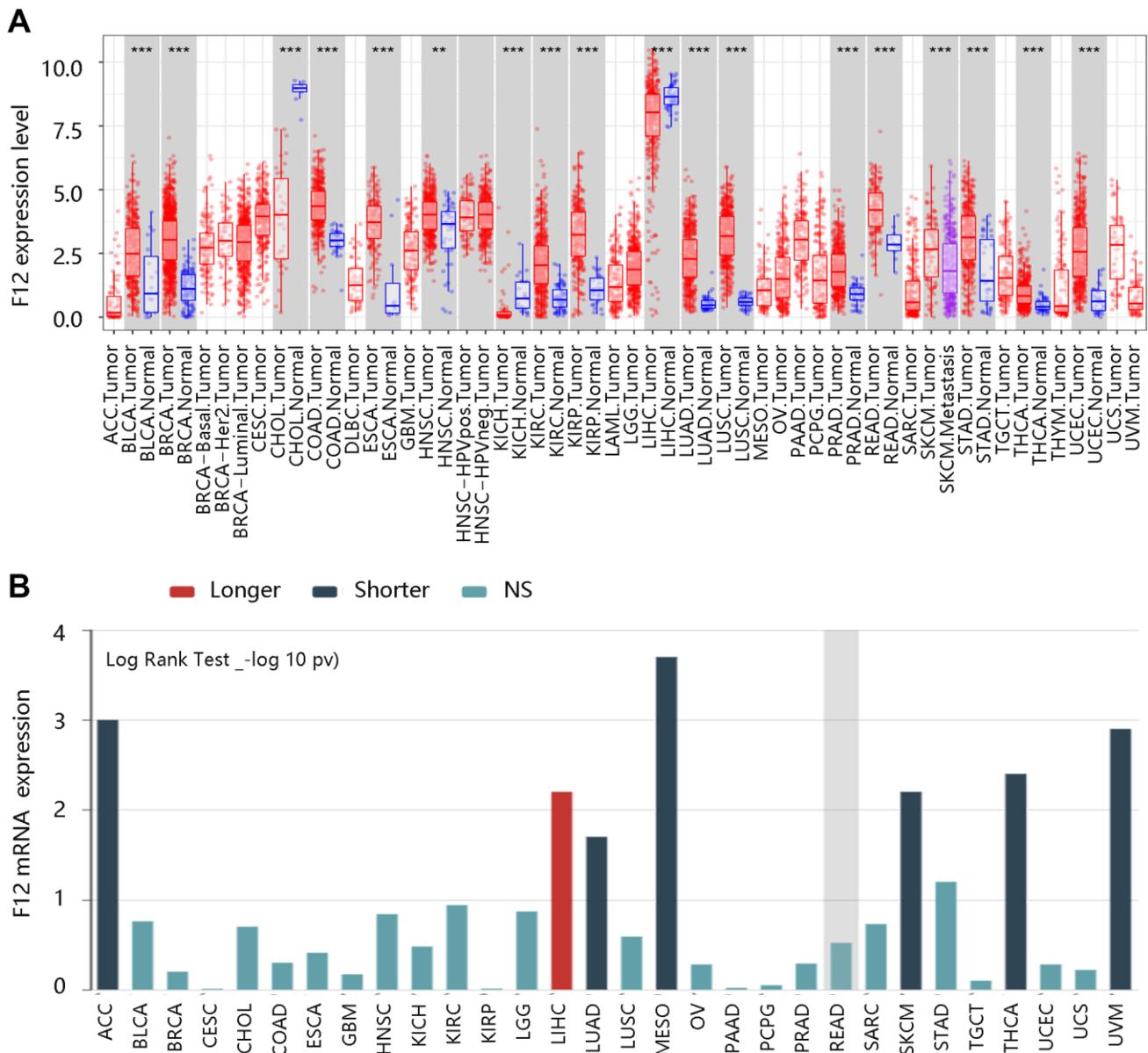
After characterizing the remarkable differential expression of F12 in PTC and normal thyroid tissues, we evaluated the clinical factors affecting its expression. As shown in Table 1, age, gender, cancer status, and stage were all not significantly linked to F12 mRNA expression (all  $P > 0.05$ ).

## F12 had a high diagnostic efficiency

To determine the value of F12 in diagnosis, the ROC curves were generated using the TCGA data. The entire AUC for F12 was 0.81 (95% confidence interval [95% CI]: 0.75-0.86), implying that F12 was capable of distinguishing normal individuals from PTC patients (Figure 4A). Besides, subgroup analysis demonstrated that the AUC values for stage I-II and stage III-IV groups were 0.77 and 0.89, respectively (Figure 4B, 4C). These results suggested that F12 had reliable diagnostic efficiency in discriminating the PTC patients and healthy subjects.

## High F12 expression predicted poor OS in PTC

To explore the effect of F12 on OS of PTC patients, survival curves were drawn using the Kaplan-Meier method. Notably, PTC patients with higher F12 expression had worse OS than those with low F12 expression ( $P < 0.001$ ) (Figure 5A). The ROC curve result showed that the entire AUC was 0.718 (95% CI: 0.581-0.854) (Figure 5B). After considering the time factor, we found that F12 still had satisfactory performance in predicting OS status of PTC patients (concordance index [C-index]: 0.755; the AUCs of 1-, 3-, and 5-year were 0.85, 0.75, and 0.79, respectively)



**Figure 1. The clinical value of F12 in pan-cancer. (A)** The expression analysis of F12 in various cancers via TIMER.  $^{**}P < 0.01$ ;  $^{***}P < 0.001$ . **(B)** The survival analysis F12 across human cancers via TISIDB.

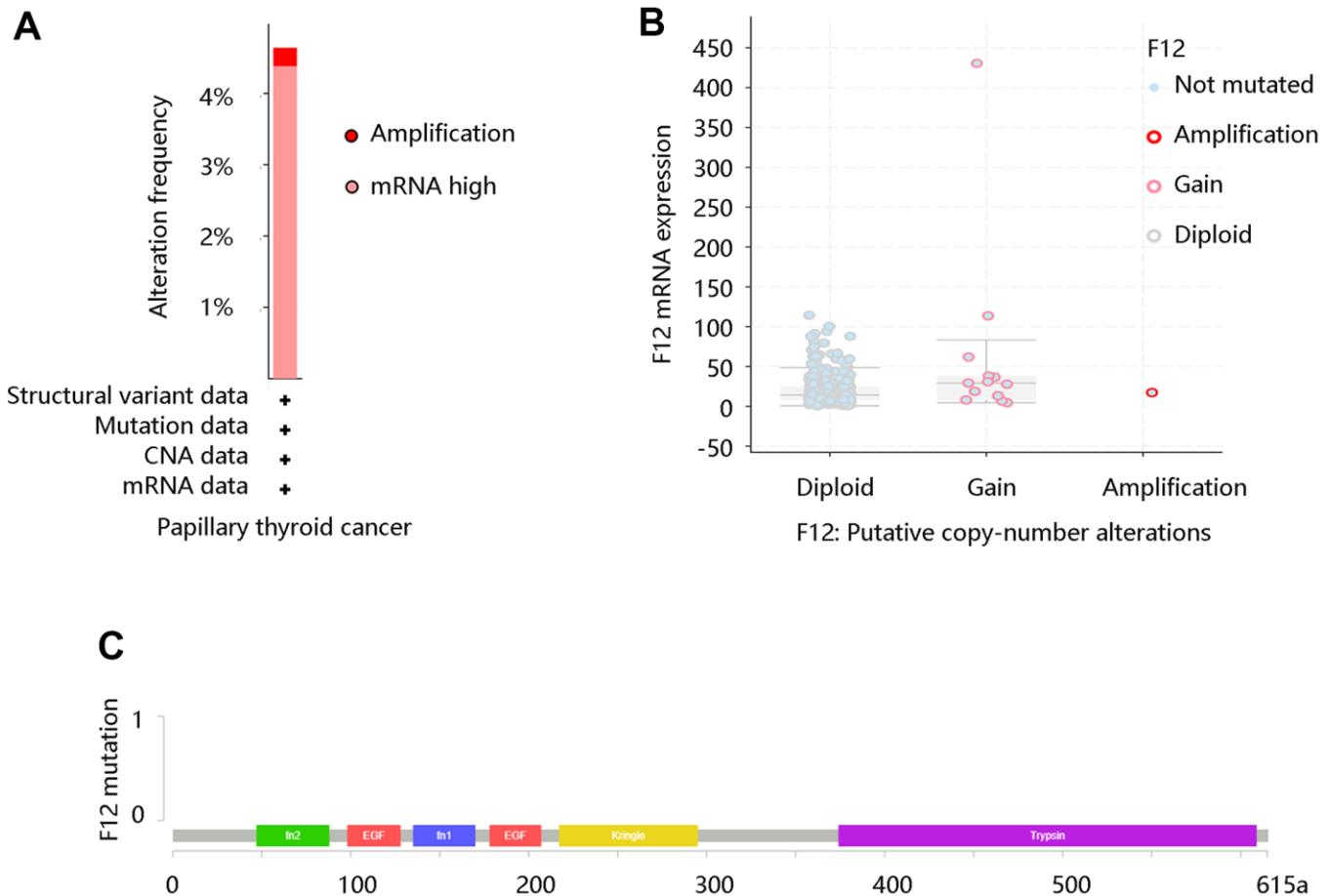
(Figure 5C). Time-dependent ROC analysis in stage III-IV PTC patients showed that the AUCs were 0.87, 0.79, and 0.85 for 1-, 3-, and 5-year, separately (C-index: 0.778) (Figure 5D). The time-dependent ROC analysis was not performed in patients at stage I-II due to fewer death samples. These findings revealed the potential value of F12 in predicting the survival status of PTC patients.

To further elucidate the effect of F12 overexpression on patient OS, we performed survival analyses on PTC patients with restricted clinicopathological characteristics. Due to no death occurring in patients less than 55 years, the survival analysis was not carried out in this subgroup. The elevated expression level of F12 had a significant relation with unfavorable OS in patients at age  $\geq 55$  years ( $P < 0.001$ ) (Figure 6A). In addition, high F12 expression led to the poor OS in subgroups of females, males, patients at stage 1+2, and stage 3+4 (all  $P < 0.01$ ) (Figure 6B–6E). F12 expression had a significant impact on the OS of PTC patients regardless of cancer status (all  $P < 0.05$ ) (Figure 6F, 6G).

To determine the independent prognostic value of F12 in PTC, Cox regression analyses were performed using the TCGA data. The univariate Cox regression analysis revealed the OS-related variables including age (Hazard ratio [HR] =1.160;  $P < 0.001$ ), cancer status (HR =30.615,  $P = 0.005$ ), stage 3 (HR =10.050,  $P = 0.004$ ), stage 4 (HR =20.622,  $P < 0.001$ ), and F12 (HR =4.894,  $P < 0.001$ ), while gender ( $P = 0.180$ ) and stage 2 ( $P = 0.070$ ) had no relationship with OS (Table 2). In the multivariate Cox regression analysis, age (HR =1.134,  $P = 0.004$ ) and F12 (HR =39.477,  $P = 0.035$ ) still had remarkable relationship with OS of PTC patients (Table 3). Therefore, high expression of F12 could independently predict poorer OS among PTC patients.

### Identification of F12-related genes and enrichment analyses

To explore the biological functions of F12 in PTC, we obtained the DEGs between high- and low- F12 expression groups for functional enrichment analysis. According to the selection criterion, the top 200 DEGs

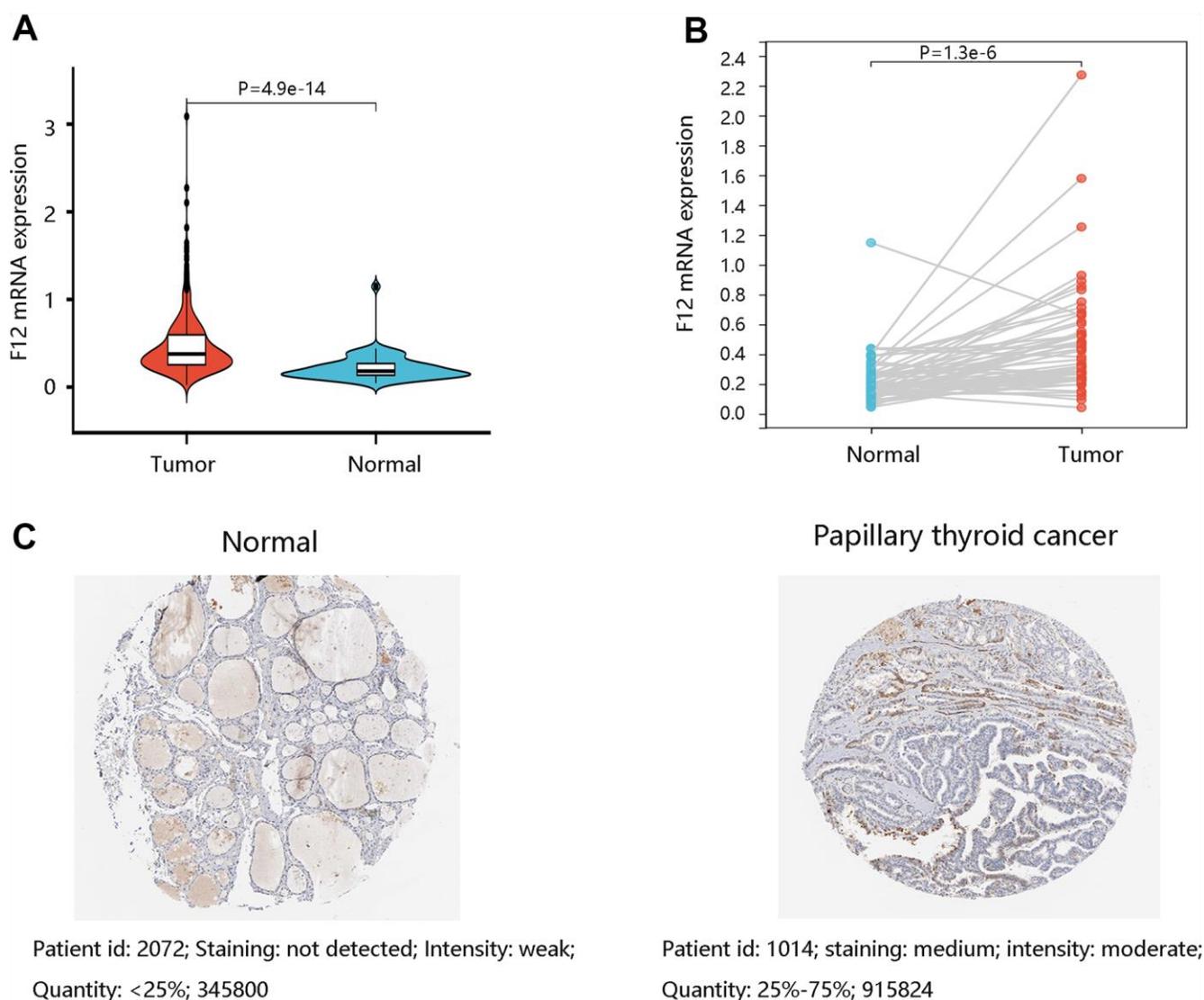


**Figure 2. The genetic status of F12 in PTC.** (A) The alteration frequency of F12 in PTC. (B) The copy number alteration of F12 in PTC. (C) The F12 mutation in PTC. PTC, papillary thyroid cancer.

were selected for analysis, and the PPI network of these 200 DEGs was constructed using the STRING database, visualized by Cytoscape software (Figure 7A). Then, the selected genes were used for GO annotation and KEGG pathway analyses. As for BP, the genes were mainly enriched in protein localization to membrane, organic cyclic compound catabolic process, and RNA catabolic process (Figure 7B). For CC, they were mainly involved in ribonucleoprotein complex, anchoring junction, and ribosomal subunit (Figure 7C). The major MFs were RNA binding, structural; molecule activity, and structural constituent of ribosome (Figure 7D). The KEGG pathways that they participated in were oxidative phosphorylation, thyroid hormone synthesis, antigen processing and presentation, glutathione

metabolism, and ferroptosis, which were closely related to tumor progression (Figure 7E).

To further reveal the underlying mechanism of F12 involved in PTC, GSEA was performed to exhibit the significantly enriched KEGG pathways in high F12 expression phenotype. The top 5 significant pathways were glycerophospholipid metabolism (NES=-1.755, FDR=0.0872), RNA polymerase (NES=-1.7989, FDR=0.0881), tyrosine metabolism (NES=-1.8032, FDR=0.1032), glycerolipid metabolism (NES=-1.7042, FDR=0.1055), and peroxisome (NES=-1.7153, FDR=0.1057) (Figure 8). Thus, F12 might participate in the development of PTC through the regulation of these metabolic pathways.



**Figure 3. The expression levels of F12 in PTC.** (A) The F12 mRNA expression in PTC and normal thyroid tissues. (B) The F12 mRNA expression in paired PTC samples. (C) The F12 protein expression in normal thyroid and PTC tissues. Image J software was used to quantify the protein levels of F12. PTC, papillary thyroid cancer.

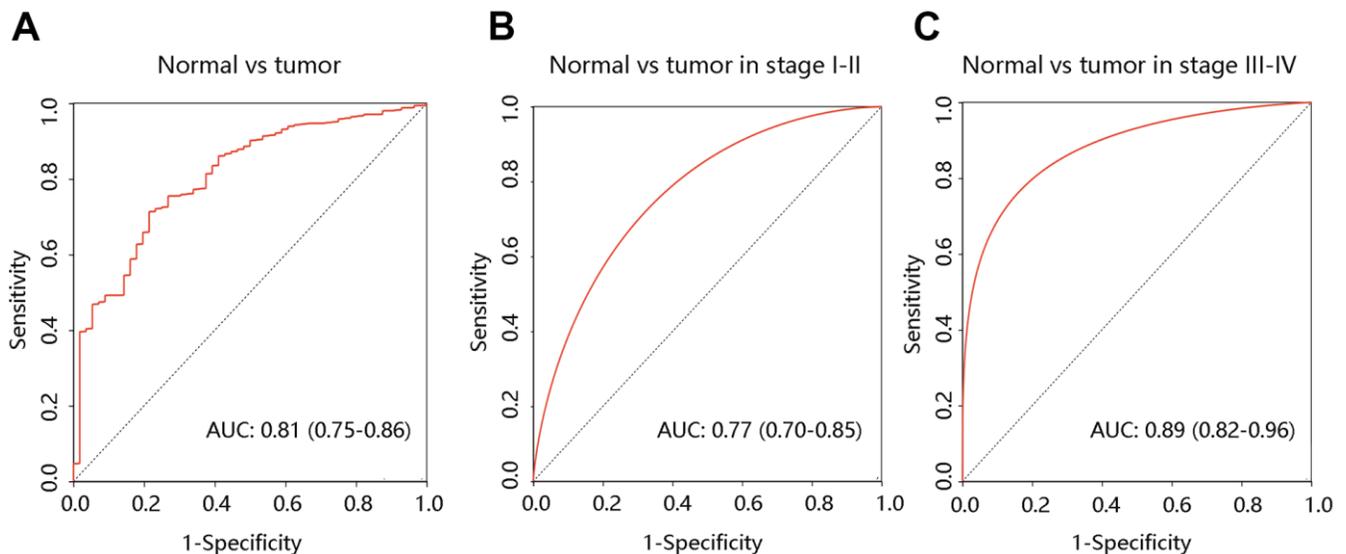
**Table 1. Relationship between F12 mRNA expression and clinicopathological characteristics in papillary thyroid cancer.**

Characteristics	F12 low (%)	F12 high (%)	$\chi^2$	P-value
Age			2.946	0.086
<55	176 (52.7)	158 (47.3)		
≥55	73 (44.5)	91 (55.5)		
Gender			0.041	0.839
Female	184 (50.3)	182 (49.7)		
Male	65 (49.2)	67 (50.8)		
Cancer status			0.001	0.980
Tumor free	200 (49.8)	202 (50.2)		
With tumor	14 (50)	14 (50)		
Stage			4.618	0.202
Stage 1	153 (54.3)	129 (45.7)		
Stage 2	22 (43.1)	29 (56.9)		
Stage 3	49 (44.5)	61 (55.5)		
Stage 4	24 (46.2)	28 (53.8)		

**F12 expression was associated with immune cell infiltration**

To assess the possible effect of F12 on various immune cell types in the PTC microenvironment, we first adopted the MCP-counter to assess the association of eight immune cells and two stromal cells' infiltration levels with the F12 mRNA expression in PTC (Figure 9A). The F12 mRNA expression was significantly related to T cells, cytotoxic lymphocytes,

B lineage, monocytic lineage, and endothelial cells (all P <0.05) (Figure 9B–9F). Following this, the CIBERSORT was used to calculate the fractions of 22 immune cells between high and low F12 expression groups in PTC samples. As shown in Figure 10A, the fractions of B cells naïve, B cells memory, T cells CD8, T cells CD4 memory resting, T cells CD4 memory activated, macrophages M1, macrophages M2, dendritic cells resting in low F12 expression group were significantly different from the high F12

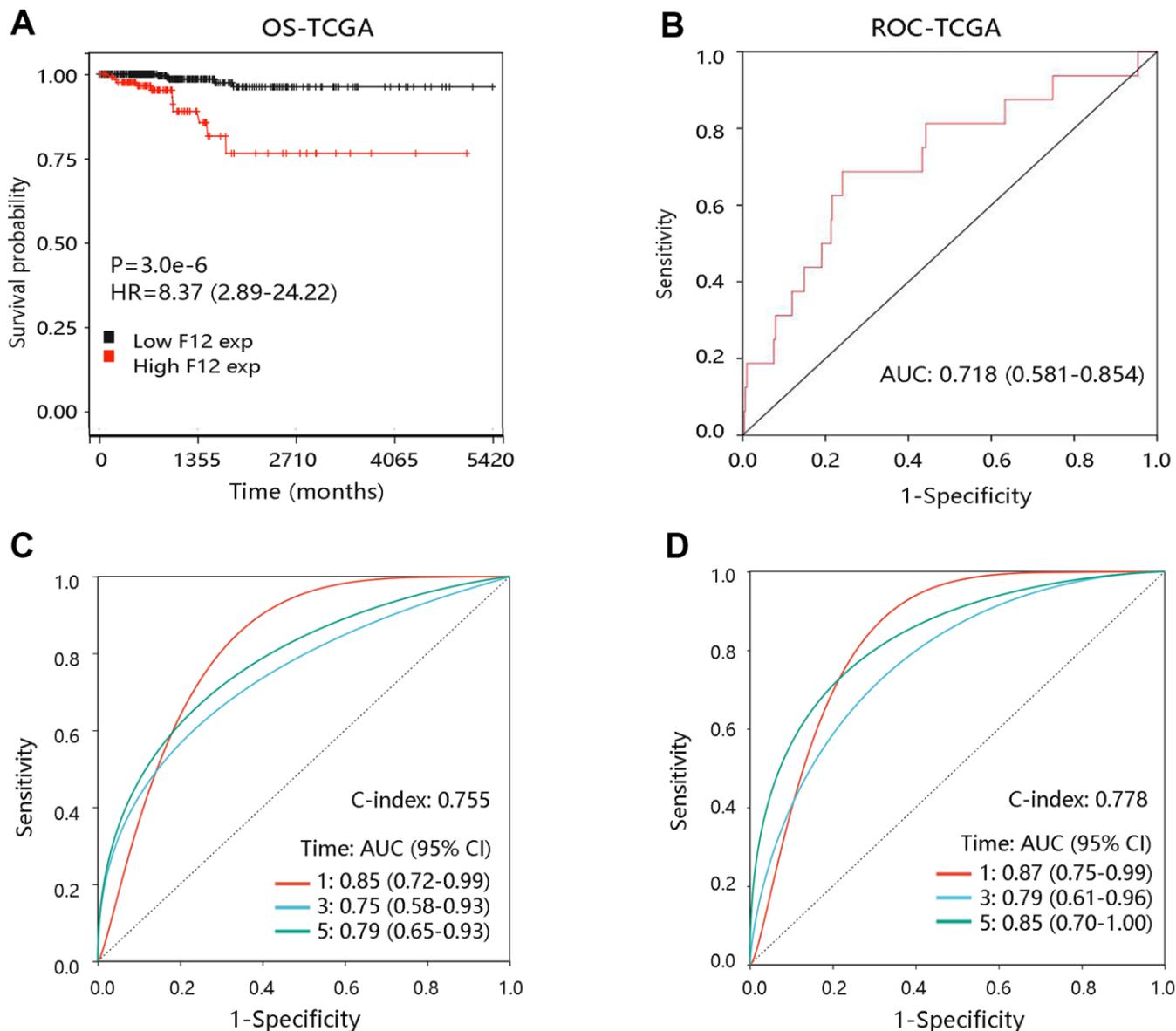


**Figure 4. Diagnostic value of F12 in PTC.** The ROC curves of F12 expression in PTC vs normal tissues (A) and different clinical stages (B, C). PTC, papillary thyroid cancer; ROC, receiver operating characteristic; AUC, the area under the curve.

expression group (all  $P < 0.05$ ). Interestingly, the M1 macrophage infiltration level was low, while that of M2 macrophages was high in the high F12 expression group. This suggested that high expression of F12 might promote the polarization of macrophages, which was intimately associated with the immunosuppressive state of the tumor [27]. Moreover, the TIMER web tool was used to validate the relationship between F12 expression and macrophage infiltration level with different algorithms. Figure 10B showed that F12 was negatively correlated with macrophage infiltration

level ( $P = 7.39 \times 10^{-3}$ ). Besides, F12 expression was negatively related to the M1 infiltration level but positively linked to the M2 infiltration level (all  $P < 0.001$ ) (Figure 10C, 10D).

Further, analysis using GDC TCGA data showed that the upregulated expression of F12 significantly correlated with the M1 macrophage marker IRF5 expression ( $P = 8.2 \times 10^{-4}$ ) (Figure 11A), but was not significantly related to M1 macrophage marker NOS2 and M2 macrophage marker MS4A4A (all  $P > 0.05$ )

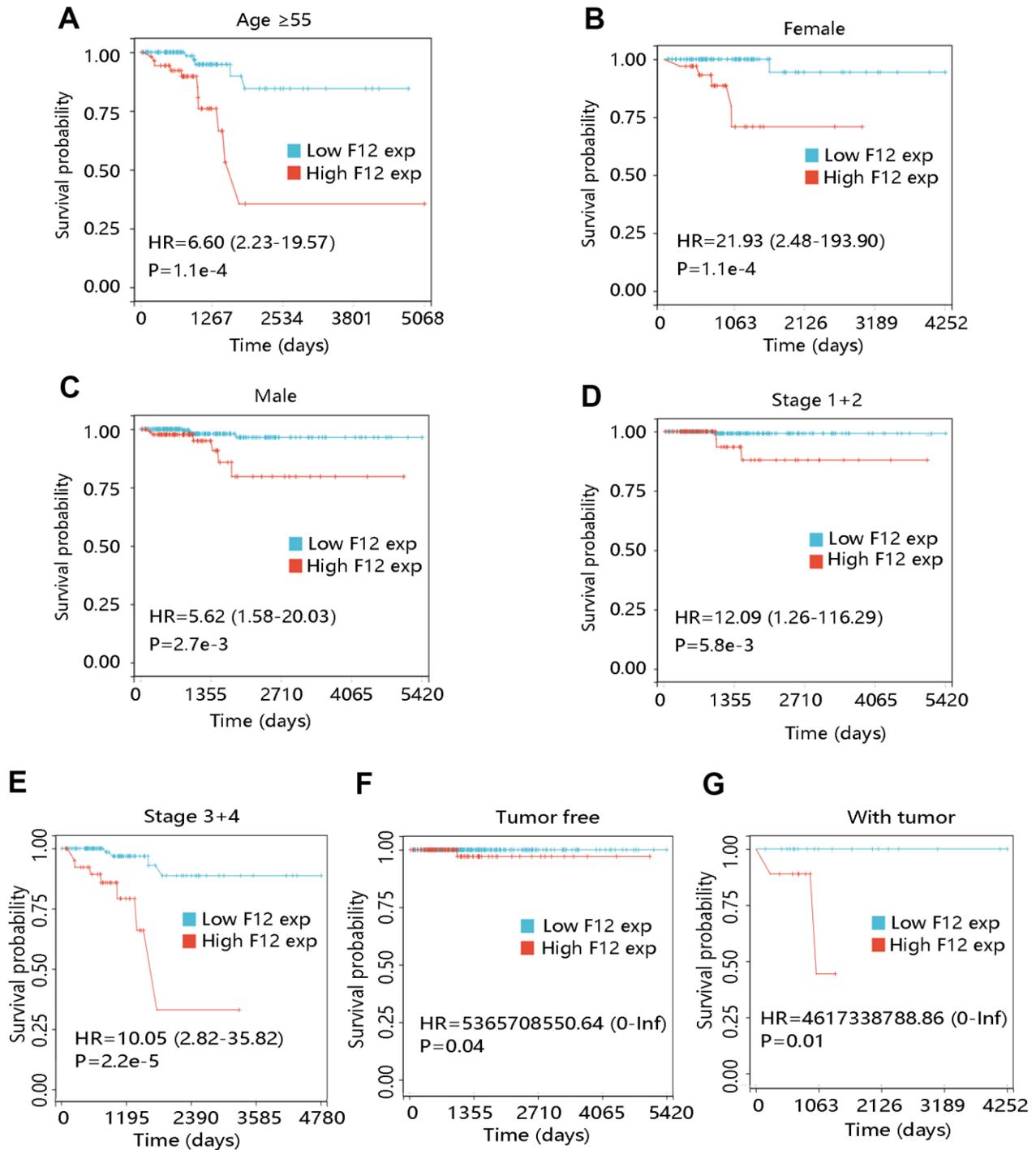


**Figure 5. The prognostic value of F12 in PTC.** (A) The association of F12 expression with OS of PTC patients. (B) The ROC curve of F12 expression for predicting the survival status. (C) Time-dependent ROC analysis in all PTC patients. (D) Time-dependent ROC analysis in PTC patients at stage III-IV. PTC, papillary thyroid cancer; ROC, receiver operating characteristic; OS, overall survival; HR, hazard ratio; AUC, area under the curve.

(Figure 11B, 11C). However, a significant positive relationship was observed between F12 expression and M2 macrophage marker TGFB1 expression ( $P=8.1e-7$ ) (Figure 11D). The above findings indicated that F12 expression was correlated with immune cell infiltration levels and might promote macrophages polarization.

## DISCUSSION

F12 serves as a plasma serine protease when autoactivated into F12a upon exposure to plasma kallikrein and negatively charged surfaces [16]. F12 is involved in human atherosclerotic lesions and F12 serum



**Figure 6. Effect of F12 expression on overall survival of papillary thyroid cancer patients with restricted clinicopathological characteristics. (A) Age  $\geq 55$ . (B) Female. (C) Male. (D) Stage 1+2. (E) Stage 3+4. (F) Tumor free. (G) With tumor.**

**Table 2. Univariate Cox regression analysis of overall survival in papillary thyroid cancer patients.**

Characteristics	HR	95% CI-l	95% CI-u	P-value
Age	1.160	1.101	1.222	<0.001
Gender	2.005	0.725	5.543	0.180
Cancer status	30.615	2.775	337.813	0.005
Stage 2	6.137	0.864	43.582	0.070
Stage 3	10.050	2.085	48.441	0.004
Stage 4	20.622	3.951	107.644	<0.001
F12	4.894	2.346	10.208	<0.001

Abbreviations: HR, hazard ratio; 95% CI-l, 95% confidence interval-lower; 95% CI-u, 95% confidence interval-upper.

**Table 3. Multivariate Cox regression analysis of overall survival in papillary thyroid cancer patients.**

Characteristics	HR	95% CI-l	95% CI-u	P-value
Age	1.134	1.041	1.235	0.004
Gender	0.087	0.000	74.469	0.479
Cancer status	1.961	0.000	3038836.204	0.926
Stage 2	0.000	0.000	1.218E+233	0.962
Stage 3	0.001	0.000	10242.999	0.395
Stage 4	0.000	0.000	2.962E+28	0.825
F12	39.477	1.305	1194.594	0.035

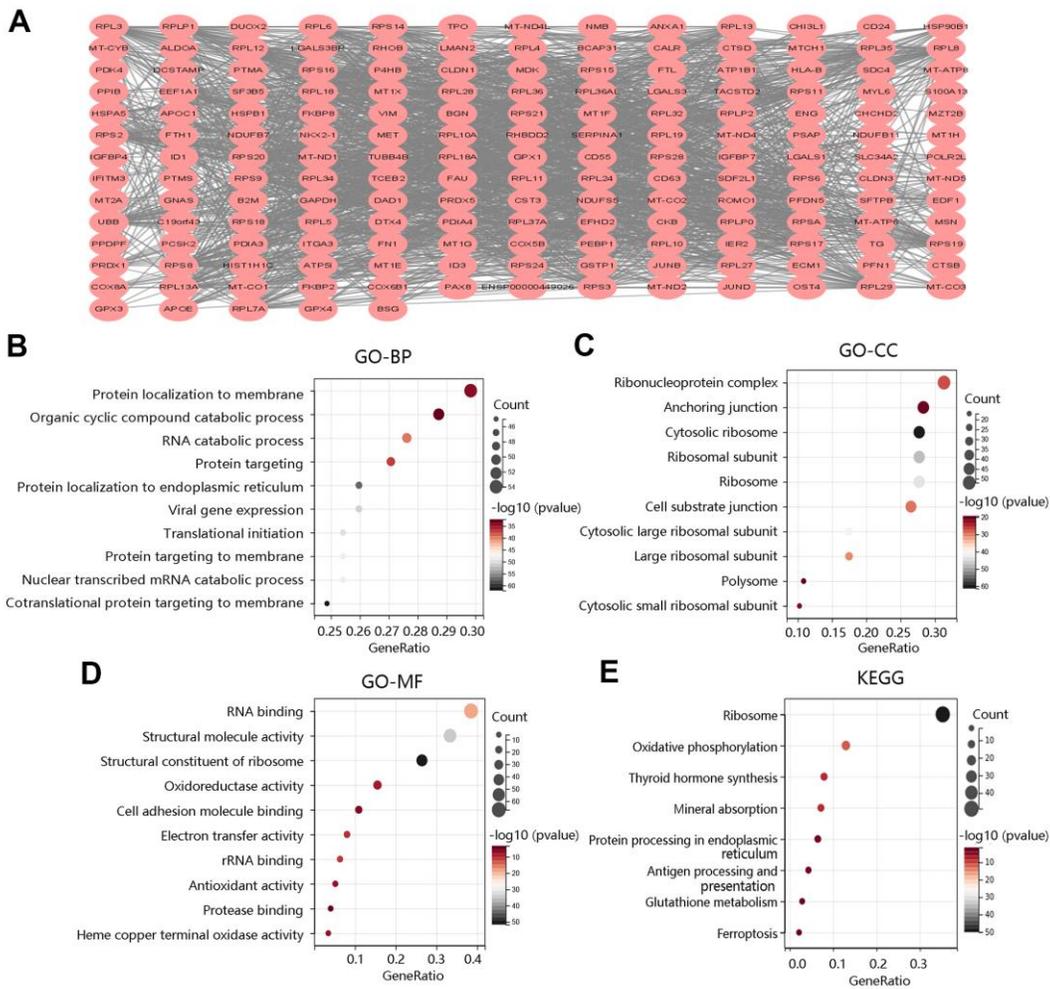
Abbreviations: HR, hazard ratio; 95% CI-l, 95% confidence interval-lower; 95% CI-u, 95% confidence interval-upper.

levels were upregulated in early atherosclerosis in low-density lipoprotein [28]. It has also been demonstrated that F12 was related to the risk of coronary heart disease, diabetes, and sepsis [29–32]. Additionally, the elevated expression level of F12 was found in the peritoneum of epithelial ovarian cancer and the interaction of F12 and macrophages lead to cancer invasion and metastasis [33], however, it has not been extensively studied in various tumors. In this study, we identified the clinical significance of F12 in tumors and further identified its independent prognostic function in PTC. Moreover, F12 expression was associated with immune cell infiltrates especially macrophages. Our study identified F12 as a potential biomarker for predicting the prognosis of PTC patients.

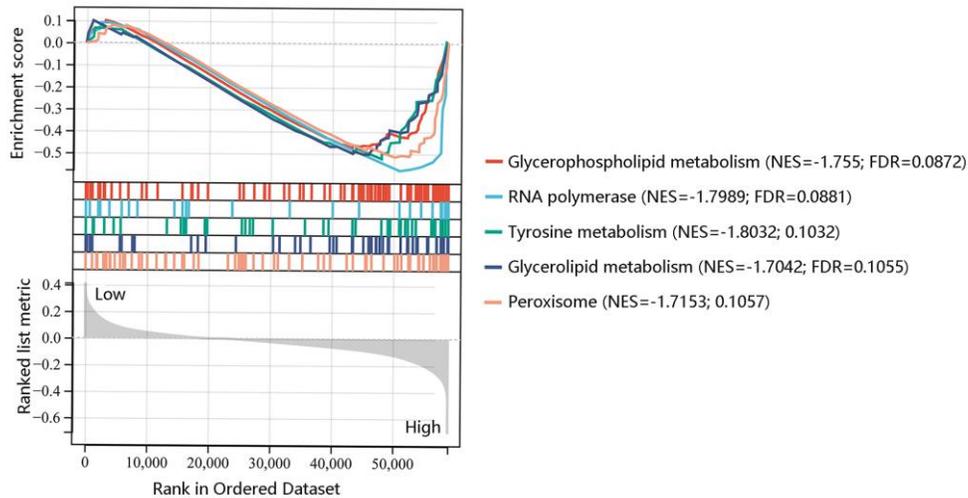
Firstly, we investigated the clinical significance of F12 in pan-cancer and found that F12 was highly expressed in most cancers including BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LUAD, LUSC, PRAD, READ, SKCM, STAD, THCA, and UCEC. F12 expression was closely related to the prognosis of many

cancers, indicating it has essential clinical significance in various tumors. We focused on the role of F12 in PTC in the following research.

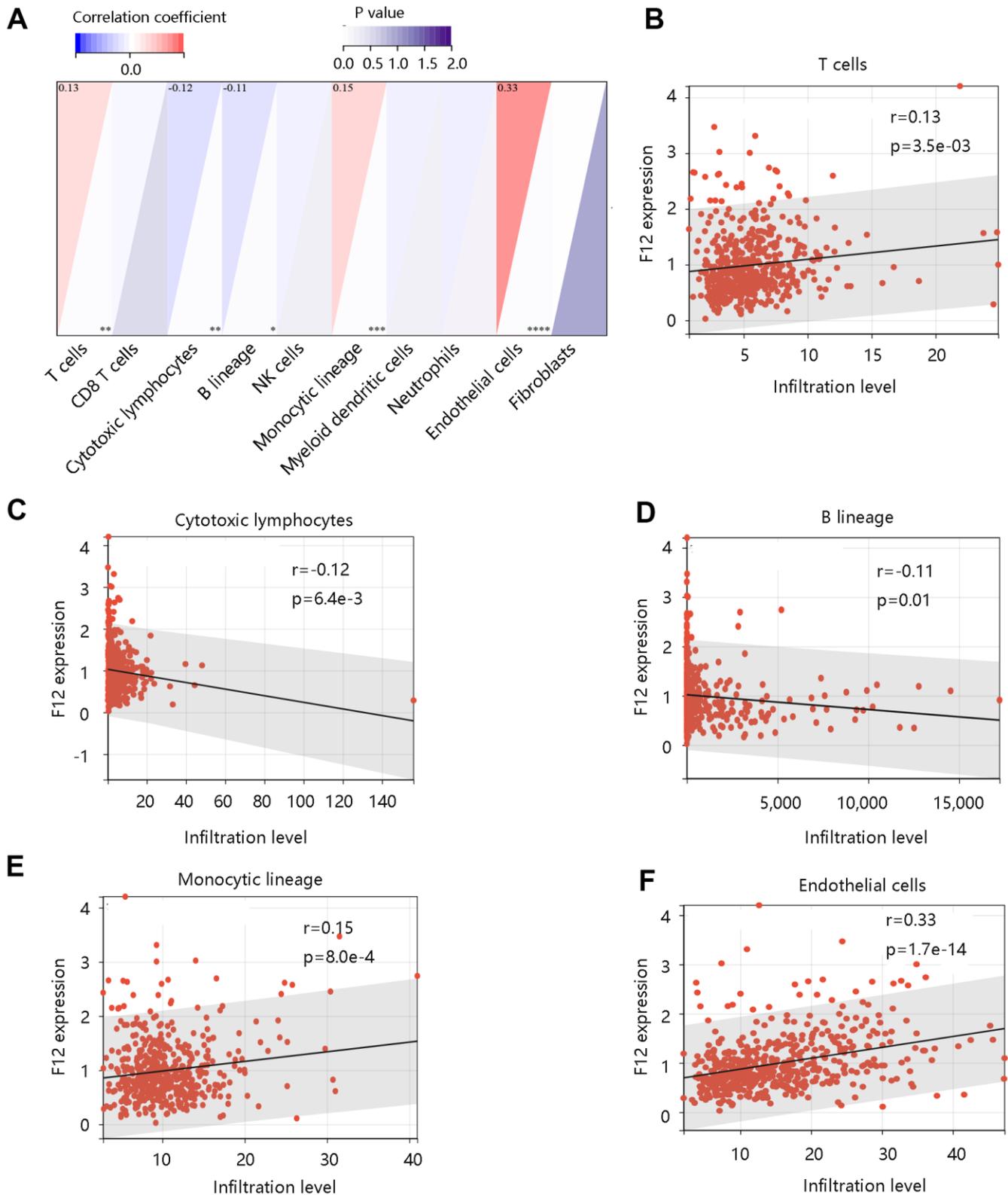
We analyzed the F12 alteration frequency, copy number, and mutation in PTC, which indicated the overexpression of F12 in PTC. Using TCGA data, both unpaired and paired t-tests confirmed the upregulation of F12 mRNA expression in PTC. And the immunohistochemical staining results presented a higher F12 protein level in PTC tissue compared with the normal thyroid tissue, suggesting that F12 might be associated with carcinogenesis. Besides, ROC curves showed that F12 expression had high sensitivity and specificity with AUC values close to 1, indicating that F12 had a favorable diagnostic value in distinguishing PTC patients from healthy subjects. Besides, F12 overexpression contributed to the poor OS in PTC patients and Cox regression analysis exhibited F12 as an independent prognostic factor for predicting OS. These results indicated that F12 might be a robust diagnostic and prognostic biomarker in PTC. F12a was reported to



**Figure 7. Functional enrichment analysis of top 200 DEGs between high and low F12 expression groups. (A)** Protein-protein interaction network of top 200 DEGs. Gene ontology annotations of **(B)** biological process, **(C)** cellular component, and **(D)** molecular function. **(E)** KEGG pathway. DEGs, differentially expressed genes.



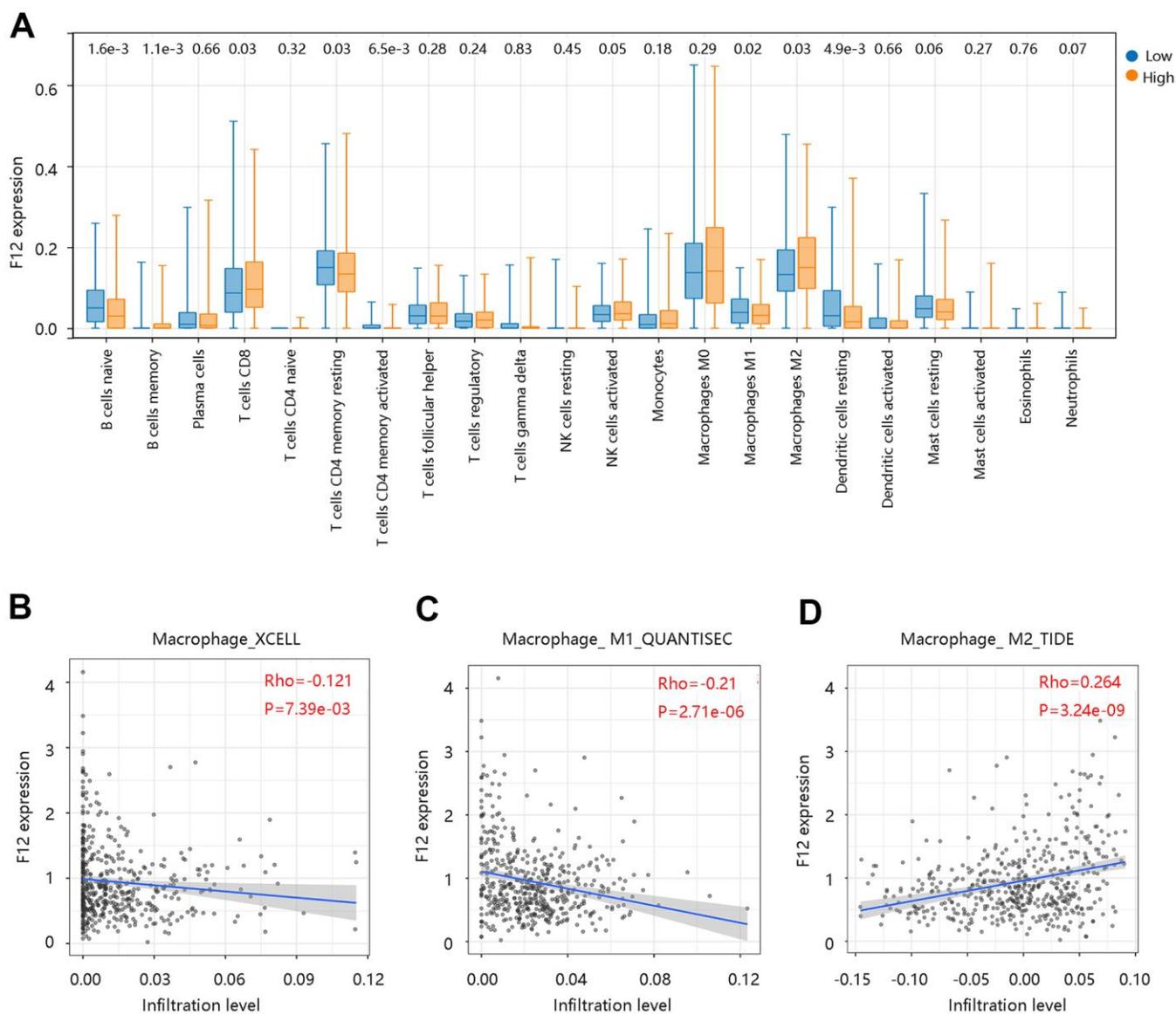
**Figure 8. Gene set enrichment analysis of the five most significant pathways in high F12 expression phenotype.**



**Figure 9. Correlation between F12 and immune cell infiltration in papillary thyroid cancer by MCP-counter.** (A) The association of F12 with immune and stromal cell infiltration level. The significant relationship between F12 expression and (B) T cells, (C) cytotoxic lymphocytes, (D) B lineage, (E) monocytic lineage, and (F) endothelial cells.

initiate the intrinsic coagulation pathway, thereby contributing to fibrin clot formation [34]. The vascular endothelial growth factor (VEGF) expressed by the tumor cells induces fibrin deposition, which served as a provisional matrix for migration of inflammatory and endothelial cells, and stromal fibroblasts into the healing wound [35–38]. On the other hand, fibrin deposited in adjacent tumor vasculature and tumor macrophages with tissue factors would initiate the coagulation pathways in a vicious cycle [39, 40]. Importantly, fibrin has been reported to promote cancer cell growth and migration [41]. Besides, fibrin matrices

were essential in tumor angiogenesis [42, 43]. Therefore, the authors speculated that F12 might affect the progression of PTC patients by triggering the coagulation pathway. The complement system activated by F12a is the primary immune response system in the blood circulation system, participating in the innate immune response [44]. The innate immune system exerts a major influence on the occurrence and proliferation of cancers, having an impact on the survival of cancer patients [45, 46]. This might be another explanation for the unfavorable prognosis of PTC patients with high F12 expression.

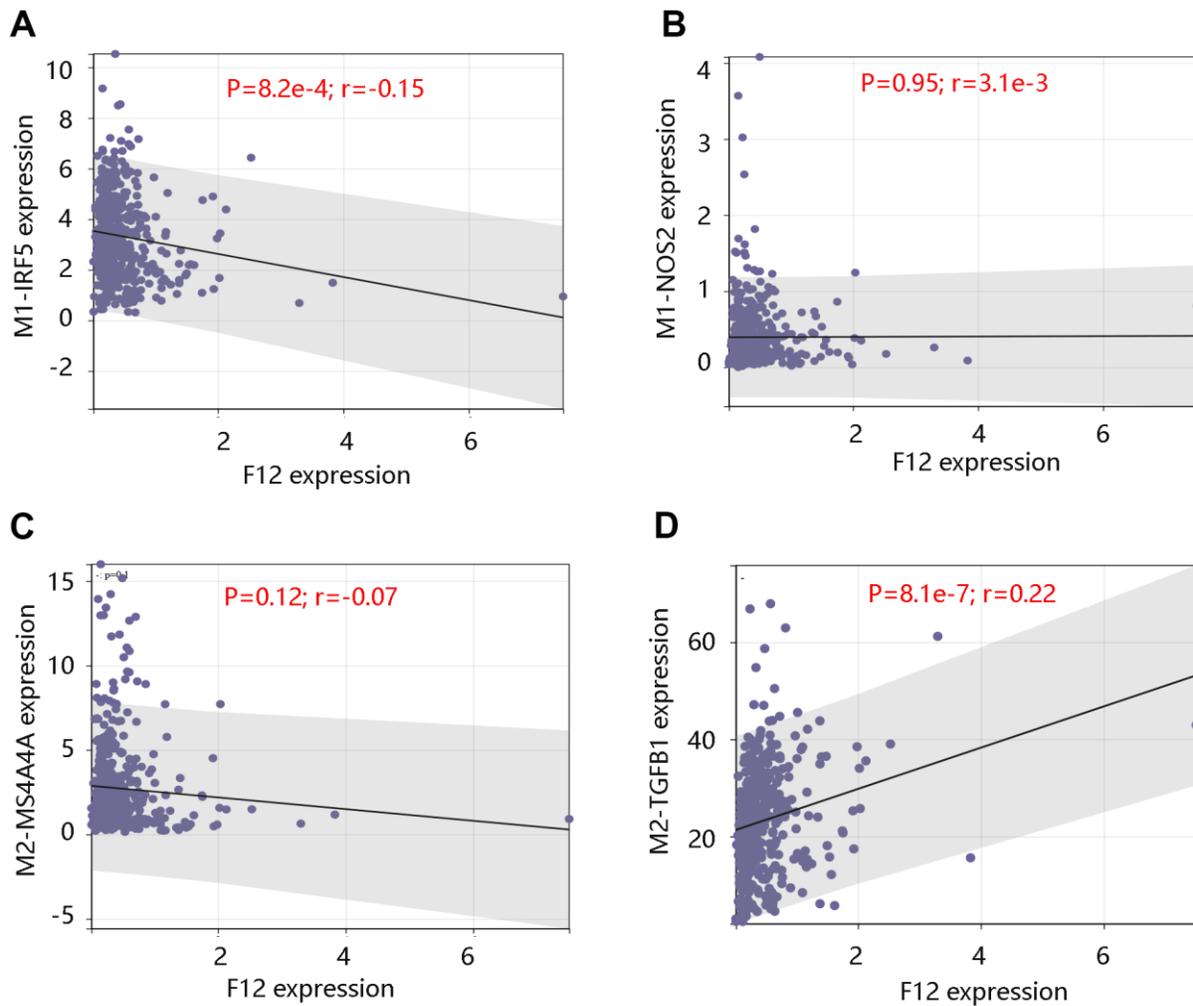


**Figure 10. Correlation between F12 and immune cell infiltration in papillary thyroid cancer by CIBERSORT.** (A) Immune cell infiltration level in the low- and high- F12 expression group. (B) Association of F12 expression with macrophage infiltration levels. (C) Association of F12 expression with M1 macrophage infiltration level. (D) Association of F12 expression with M2 macrophage infiltration level.

To explore the underlying mechanism of F12 involved in PTC, the KEGG pathway analysis of DEGs between the high- and low- F12 expression groups was performed. These DEGs were mainly enriched in oxidative phosphorylation, thyroid hormone synthesis, antigen processing and presentation, glutathione metabolism, and ferroptosis. The GSEA result showed that glycerophospholipid metabolism, RNA polymerase, tyrosine metabolism, glycerolipid metabolism, and peroxisome were the major significant pathways in high F12 expression phenotype. Glutamine metabolism is able to inhibit glutathione synthesis and induce ferroptosis. The inhibition of glutamine metabolism can suppress ferroptosis [47]. Moreover, ferroptosis represents a crucial role in the development and treatment of various cancers, which can be both stimulatory and inhibitory. Ferroptosis induces inflammation-related immunosuppressive tumor microenvironment to facilitate tumor growth; however, ferroptosis inducers including

Erastin cause ferroptosis, inhibiting the progression of tumor [48–51]. Therefore, F12 might affect the progression of PTC through metabolic pathways, especially glutamine metabolism.

Tumor microenvironment immune cells constitute a key factor of tumor tissues and exhibit a significant role in predicting the survival status and therapeutic efficacy of cancer patients [52–54]. M2 macrophages promote tumor growth through the promotion of angiogenesis, secretion of growth factors, and suppression of adaptive immunity [55, 56]. Whereas, M1 macrophages stimulate the immune response via secreting proinflammatory cytokines [57]. By analyzing the relationship between F12 expression and immune cell infiltration, we found that the fractions of many immune cells in the high F12 expression group vary from the low F12 expression group. Interestingly, the M2 macrophage infiltration level was notably higher in the high F12 expression group,



**Figure 11. Correlation between F12 expression and macrophage markers.** Relationship between F12 expression and M1 macrophage markers (A) IRF5 and (B) NOS2. Relationship between F12 expression and M2 macrophage markers (C) MS4A4A and (D) TGFB1.

while a downregulated M1 macrophage infiltration level was observed in this group. This provided new insight into the correlation between F12 overexpression and shorter OS time of PTC patients. However, the data we analyzed were obtained from the public datasets, which requires further validation by *in vitro* and *in vivo* experiments.

In summary, F12 expression had high diagnostic efficacy and its overexpression served as an independent prognostic predictor for PTC patients. In addition, F12 expression was involved in metabolic pathways especially glutathione metabolism in PTC. Our findings also suggested that F12 might play a vital role in immune cell infiltration.

### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### AUTHOR CONTRIBUTIONS

(I) Conception and design: Jun-Hua Luo and Xiu-Xia Zhang. (II) Collection and assembly of data: Wu-Hui Sun. (III) Data analysis and interpretation: All authors. (IV) Manuscript writing: All authors.

### CONFLICTS OF INTEREST

The authors report no conflicts of interest.

### REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:31912902
2. Deng Y, Li H, Wang M, Li N, Tian T, Wu Y, Xu P, Yang S, Zhai Z, Zhou L, Hao Q, Song D, Jin T, et al. Global Burden of Thyroid Cancer From 1990 to 2017. *JAMA Netw Open.* 2020; 3:e208759. <https://doi.org/10.1001/jamanetworkopen.2020.8759> PMID:32589231
3. Fagin JA, Wells SA Jr. Biologic and Clinical Perspectives on Thyroid Cancer. *N Engl J Med.* 2016; 375:1054–67. <https://doi.org/10.1056/NEJMra1501993> PMID:27626519
4. Verburg FA, Van Santen HM, Luster M. Pediatric papillary thyroid cancer: current management challenges. *Onco Targets Ther.* 2016; 10:165–75. <https://doi.org/10.2147/OTT.S100512> PMID:28096684
5. Schneider DF, Chen H. New developments in the diagnosis and treatment of thyroid cancer. *CA Cancer J Clin.* 2013; 63:374–94. <https://doi.org/10.3322/caac.21195> PMID:23797834
6. Riesco-Eizaguirre G, Galofré JC, Grande E, Zafón Llopis C, Ramón y Cajal Asensio T, Navarro González E, Jiménez-Fonseca P, Santamaría Sandi J, Gómez Sáez JM, Capdevila J. Spanish consensus for the management of patients with advanced radioactive iodine refractory differentiated thyroid cancer. *Endocrinol Nutr.* 2016; 63:e17–24. <https://doi.org/10.1016/j.endonu.2015.08.007> PMID:26601805
7. Luo J, Zhang B, Cui L, Liu T, Gu Y. FMO1 gene expression independently predicts favorable recurrence-free survival of classical papillary thyroid cancer. *Future Oncol.* 2019; 15:1303–11. <https://doi.org/10.2217/fon-2018-0885> PMID:30757917
8. Grogan RH, Kaplan SP, Cao H, Weiss RE, Degroot LJ, Simon CA, Embia OM, Angelos P, Kaplan EL, Schechter RB. A study of recurrence and death from papillary thyroid cancer with 27 years of median follow-up. *Surgery.* 2013; 154:1436–46. <https://doi.org/10.1016/j.surg.2013.07.008> PMID:24075674
9. Lin JS, Bowles EJ, Williams SB, Morrison CC. Screening for Thyroid Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA.* 2017; 317:1888–903. <https://doi.org/10.1001/jama.2017.0562> PMID:28492904
10. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. *Nat Rev Endocrinol.* 2016; 12:646–53. <https://doi.org/10.1038/nrendo.2016.110> PMID:27418023
11. Ito Y, Yoshida H, Tomoda C, Miya A, Kobayashi K, Matsuzuka F, Kakudo K, Kuma K, Miyauchi A. HBME-1 expression in follicular tumor of the thyroid: an investigation of whether it can be used as a marker to diagnose follicular carcinoma. *Anticancer Res.* 2005; 25:179–82. PMID:15816536
12. Kebebew E, Reiff E. Patients with differentiated thyroid cancer have a venous gradient in thyroglobulin levels. *Cancer.* 2007; 109:1078–81. <https://doi.org/10.1002/cncr.22505> PMID:17279579
13. Pixley RA, Colman RW. Factor XII: Hageman factor. *Methods Enzymol.* 1993; 222:51–65. [https://doi.org/10.1016/0076-6879\(93\)22007-3](https://doi.org/10.1016/0076-6879(93)22007-3) PMID:8412813

14. McMichael M. New models of hemostasis. *Top Companion Anim Med.* 2012; 27:40–5.  
<https://doi.org/10.1053/j.tcam.2012.07.005>  
PMID:23031454
15. Stavrou E, Schmaier AH. Factor XII: what does it contribute to our understanding of the physiology and pathophysiology of hemostasis & thrombosis. *Thromb Res.* 2010; 125:210–5.  
<https://doi.org/10.1016/j.thromres.2009.11.028>  
PMID:20022081
16. Schmaier AH. The elusive physiologic role of Factor XII. *J Clin Invest.* 2008; 118:3006–9.  
<https://doi.org/10.1172/JCI36617>  
PMID:18725991
17. Renné T, Schmaier AH, Nickel KF, Blombäck M, Maas C. *In vivo* roles of factor XII. *Blood.* 2012; 120:4296–303.  
<https://doi.org/10.1182/blood-2012-07-292094>  
PMID:22993391
18. Wu Y. Contact pathway of coagulation and inflammation. *Thromb J.* 2015; 13:17.  
<https://doi.org/10.1186/s12959-015-0048-y>  
PMID:25949215
19. Ikeda Y, Hayashi I, Kamoshita E, Yamazaki A, Endo H, Ishihara K, Yamashina S, Tsutsumi Y, Matsubara H, Majima M. Host stromal bradykinin B2 receptor signaling facilitates tumor-associated angiogenesis and tumor growth. *Cancer Res.* 2004; 64:5178–85.  
<https://doi.org/10.1158/0008-5472.CAN-03-3589>  
PMID:15289322
20. Yasuhara O, Walker DG, McGeer PL. Hageman factor and its binding sites are present in senile plaques of Alzheimer's disease. *Brain Res.* 1994; 654:234–40.  
[https://doi.org/10.1016/0006-8993\(94\)90484-7](https://doi.org/10.1016/0006-8993(94)90484-7)  
PMID:7987673
21. Göbel K, Pankratz S, Asaridou CM, Herrmann AM, Bittner S, Merker M, Ruck T, Glumm S, Langhauser F, Kraft P, Krug TF, Breuer J, Herold M, et al. Blood coagulation factor XII drives adaptive immunity during neuroinflammation via CD87-mediated modulation of dendritic cells. *Nat Commun.* 2016; 7:11626.  
<https://doi.org/10.1038/ncomms11626>  
PMID:27188843
22. Kleinschnitz C, Stoll G, Bendszus M, Schuh K, Pauer HU, Burfeind P, Renné C, Gailani D, Nieswandt B, Renné T. Targeting coagulation factor XII provides protection from pathological thrombosis in cerebral ischemia without interfering with hemostasis. *J Exp Med.* 2006; 203:513–8.  
<https://doi.org/10.1084/jem.20052458>  
PMID:16533887
23. Battistelli S, Stefanoni M, Lorenzi B, Dell'Avanzato R, Varrone F, Pascucci A, Petrioli R, Vittoria V. Coagulation factor levels in non-metastatic colorectal cancer patients. *Int J Biol Markers.* 2008; 23:36–41.  
<https://doi.org/10.5301/IBM.2008.4255>  
PMID:28207105
24. Roeise O, Sivertsen S, Ruud TE, Bouma BN, Stadaas JO, Aasen AO. Studies on components of the contact phase system in patients with advanced gastrointestinal cancer. *Cancer.* 1990; 65:1355–9.  
[https://doi.org/10.1002/1097-0142\(19900315\)65:6<1355::aid-cnrcr2820650618>3.0.co;2-1](https://doi.org/10.1002/1097-0142(19900315)65:6<1355::aid-cnrcr2820650618>3.0.co;2-1)  
PMID:1689607
25. Pan J, Qian Y, Weiser P, Zhou X, Lu H, Studelska DR, Zhang L. Glycosaminoglycans and activated contact system in cancer patient plasmas. *Prog Mol Biol Transl Sci.* 2010; 93:473–95.  
[https://doi.org/10.1016/S1877-1173\(10\)93020-2](https://doi.org/10.1016/S1877-1173(10)93020-2)  
PMID:20807657
26. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015; 12:453–7.  
<https://doi.org/10.1038/nmeth.3337> PMID:25822800
27. Ying X, Wu Q, Wu X, Zhu Q, Wang X, Jiang L, Chen X, Wang X. Epithelial ovarian cancer-secreted exosomal miR-222-3p induces polarization of tumor-associated macrophages. *Oncotarget.* 2016; 7:43076–87.  
<https://doi.org/10.18632/oncotarget.9246>  
PMID:27172798
28. Evangelho JS, Casali KR, Campos C, De Angelis K, Veiga AB, Rigatto K. Hypercholesterolemia magnitude increases sympathetic modulation and coagulation in LDLr knockout mice. *Auton Neurosci.* 2011; 159:98–103.  
<https://doi.org/10.1016/j.autneu.2010.07.010>  
PMID:20719578
29. Zito F, Drummond F, Bujac SR, Esnouf MP, Morrissey JH, Humphries SE, Miller GJ. Epidemiological and genetic associations of activated factor XII concentration with factor VII activity, fibrinopeptide A concentration, and risk of coronary heart disease in men. *Circulation.* 2000; 102:2058–62.  
<https://doi.org/10.1161/01.cir.102.17.2058>  
PMID:11044420
30. Miller GJ, Esnouf MP, Burgess AI, Cooper JA, Mitchell JP. Risk of coronary heart disease and activation of factor XII in middle-aged men. *Arterioscler Thromb Vasc Biol.* 1997; 17:2103–6.  
<https://doi.org/10.1161/01.atv.17.10.2103>  
PMID:9351378
31. Carr ME. Diabetes mellitus: a hypercoagulable state. *J Diabetes Complications.* 2001; 15:44–54.  
[https://doi.org/10.1016/s1056-8727\(00\)00132-x](https://doi.org/10.1016/s1056-8727(00)00132-x)  
PMID:11259926

32. Jansen PM, Pixley RA, Brouwer M, de Jong IW, Chang AC, Hack CE, Taylor FB Jr, Colman RW. Inhibition of factor XII in septic baboons attenuates the activation of complement and fibrinolytic systems and reduces the release of interleukin-6 and neutrophil elastase. *Blood*. 1996; 87:2337–44.  
<https://doi.org/10.1182/blood.V87.6.2337.bloodjournal8762337> PMID:[8630396](https://pubmed.ncbi.nlm.nih.gov/8630396/)
33. Wang R, Zhang T, Ma Z, Wang Y, Cheng Z, Xu H, Li W, Wang X. The interaction of coagulation factor XII and monocyte/macrophages mediating peritoneal metastasis of epithelial ovarian cancer. *Gynecol Oncol*. 2010; 117:460–6.  
<https://doi.org/10.1016/j.ygyno.2010.02.015> PMID:[20233624](https://pubmed.ncbi.nlm.nih.gov/20233624/)
34. Maas C, Renné T. Regulatory mechanisms of the plasma contact system. *Thromb Res*. 2012 (Suppl 2); 129:S73–6.  
<https://doi.org/10.1016/j.thromres.2012.02.039> PMID:[22398015](https://pubmed.ncbi.nlm.nih.gov/22398015/)
35. Senger DR, Van de Water L, Brown LF, Nagy JA, Yeo KT, Yeo TK, Berse B, Jackman RW, Dvorak AM, Dvorak HF. Vascular permeability factor (VPF, VEGF) in tumor biology. *Cancer Metastasis Rev*. 1993; 12:303–24.  
<https://doi.org/10.1007/BF00665960> PMID:[8281615](https://pubmed.ncbi.nlm.nih.gov/8281615/)
36. Dvorak HF. Vascular permeability to plasma, plasma proteins, and cells: an update. *Curr Opin Hematol*. 2010; 17:225–9.  
<https://doi.org/10.1097/MOH.0b013e3283386638> PMID:[20375889](https://pubmed.ncbi.nlm.nih.gov/20375889/)
37. Clark RA, Folkvord JM, Wertz RL. Fibronectin, as well as other extracellular matrix proteins, mediate human keratinocyte adherence. *J Invest Dermatol*. 1985; 84:378–83.  
<https://doi.org/10.1111/1523-1747.ep12265466> PMID:[2582060](https://pubmed.ncbi.nlm.nih.gov/2582060/)
38. Dvorak HF, Nagy JA, Berse B, Brown LF, Yeo KT, Yeo TK, Dvorak AM, van de Water L, Sioussat TM, Senger DR. Vascular permeability factor, fibrin, and the pathogenesis of tumor stroma formation. *Ann N Y Acad Sci*. 1992; 667:101–11.  
<https://doi.org/10.1111/j.1749-6632.1992.tb51603.x> PMID:[1309029](https://pubmed.ncbi.nlm.nih.gov/1309029/)
39. Ornstein DL, Zacharski LR, Memoli VA, Kisiel W, Kudryk BJ, Hunt J, Rousseau SM, Stump DC. Coexisting macrophage-associated fibrin formation and tumor cell urokinase in squamous cell and adenocarcinoma of the lung tissues. *Cancer*. 1991; 68:1061–7.  
[https://doi.org/10.1002/1097-0142\(19910901\)68:5<1061::aid-cnrcr2820680525>3.0.co;2-d](https://doi.org/10.1002/1097-0142(19910901)68:5<1061::aid-cnrcr2820680525>3.0.co;2-d) PMID:[1913476](https://pubmed.ncbi.nlm.nih.gov/1913476/)
40. Wojtukiewicz MZ, Zacharski LR, Memoli VA, Kisiel W, Kudryk BJ, Rousseau SM, Stump DC. Indirect activation of blood coagulation in colon cancer. *Thromb Haemost*. 1989; 62:1062–6.  
PMID:[2694422](https://pubmed.ncbi.nlm.nih.gov/2694422/)
41. Henke CA, Roongta U, Mickelson DJ, Knutson JR, McCarthy JB. CD44-related chondroitin sulfate proteoglycan, a cell surface receptor implicated with tumor cell invasion, mediates endothelial cell migration on fibrinogen and invasion into a fibrin matrix. *J Clin Invest*. 1996; 97:2541–52.  
<https://doi.org/10.1172/JCI118702> PMID:[8647947](https://pubmed.ncbi.nlm.nih.gov/8647947/)
42. Bootle-Wilbraham CA, Tazzyman S, Thompson WD, Stirk CM, Lewis CE. Fibrin fragment E stimulates the proliferation, migration and differentiation of human microvascular endothelial cells *in vitro*. *Angiogenesis*. 2001; 4:269–75.  
<https://doi.org/10.1023/a:1016076121918> PMID:[12197472](https://pubmed.ncbi.nlm.nih.gov/12197472/)
43. Kołodziejczyk J, Ponczek MB. The role of fibrinogen, fibrin and fibrin(ogen) degradation products (FDPs) in tumor progression. *Contemp Oncol (Pozn)*. 2013; 17:113–9.  
<https://doi.org/10.5114/wo.2013.34611> PMID:[23788975](https://pubmed.ncbi.nlm.nih.gov/23788975/)
44. Gialeli C, Gungor B, Blom AM. Novel potential inhibitors of complement system and their roles in complement regulation and beyond. *Mol Immunol*. 2018; 102:73–83.  
<https://doi.org/10.1016/j.molimm.2018.05.023> PMID:[30217334](https://pubmed.ncbi.nlm.nih.gov/30217334/)
45. Corrales L, Matson V, Flood B, Spranger S, Gajewski TF. Innate immune signaling and regulation in cancer immunotherapy. *Cell Res*. 2017; 27:96–108.  
<https://doi.org/10.1038/cr.2016.149> PMID:[27981969](https://pubmed.ncbi.nlm.nih.gov/27981969/)
46. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. 2013; 14:1014–22.  
<https://doi.org/10.1038/ni.2703> PMID:[24048123](https://pubmed.ncbi.nlm.nih.gov/24048123/)
47. Gao M, Monian P, Quadri N, Ramasamy R, Jiang X. Glutaminolysis and Transferrin Regulate Ferroptosis. *Mol Cell*. 2015; 59:298–308.  
<https://doi.org/10.1016/j.molcel.2015.06.011> PMID:[26166707](https://pubmed.ncbi.nlm.nih.gov/26166707/)
48. Ryan MB, Corcoran RB. Therapeutic strategies to target RAS-mutant cancers. *Nat Rev Clin Oncol*. 2018; 15:709–20.  
<https://doi.org/10.1038/s41571-018-0105-0> PMID:[30275515](https://pubmed.ncbi.nlm.nih.gov/30275515/)
49. Ingold I, Berndt C, Schmitt S, Doll S, Poschmann G, Buday K, Roveri A, Peng X, Porto Freitas F, Seibt T, Mehr L, Aichler M, Walch A, et al. Selenium Utilization by GPX4 Is Required to Prevent Hydroperoxide-Induced Ferroptosis. *Cell*. 2018; 172:409–22.e21.

- <https://doi.org/10.1016/j.cell.2017.11.048>  
PMID:29290465
50. Rojo de la Vega M, Chapman E, Zhang DD. NRF2 and the Hallmarks of Cancer. *Cancer Cell*. 2018; 34:21–43.  
<https://doi.org/10.1016/j.ccell.2018.03.022>  
PMID:29731393
51. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, Campbell K, Cano A, Casanova J, Christofori G, Dedhar S, Derynck R, Ford HL, et al, and EMT International Association (TEMPTIA). Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2020; 21:341–52.  
<https://doi.org/10.1038/s41580-020-0237-9>  
PMID:32300252
52. Vitale I, Manic G, Coussens LM, Kroemer G, Galluzzi L. Macrophages and Metabolism in the Tumor Microenvironment. *Cell Metab*. 2019; 30:36–50.  
<https://doi.org/10.1016/j.cmet.2019.06.001>  
PMID:31269428
53. Liu C, Zhou X, Long Q, Zeng H, Sun Q, Chen Y, Wu D, Liu L. Small extracellular vesicles containing miR-30a-3p attenuate the migration and invasion of hepatocellular carcinoma by targeting SNAP23 gene. *Oncogene*. 2021; 40:233–45.
- <https://doi.org/10.1038/s41388-020-01521-7>  
PMID:33110233
54. Xu F, Zhan X, Zheng X, Xu H, Li Y, Huang X, Lin L, Chen Y. A signature of immune-related gene pairs predicts oncologic outcomes and response to immunotherapy in lung adenocarcinoma. *Genomics*. 2020; 112:4675–83.  
<https://doi.org/10.1016/j.ygeno.2020.08.014>  
PMID:32798715
55. Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002; 420:860–7.  
<https://doi.org/10.1038/nature01322>  
PMID:12490959
56. Lewis CE, Pollard JW. Distinct role of macrophages in different tumor microenvironments. *Cancer Res*. 2006; 66:605–12.  
<https://doi.org/10.1158/0008-5472.CAN-05-4005>  
PMID:16423985
57. Choo YW, Kang M, Kim HY, Han J, Kang S, Lee JR, Jeong GJ, Kwon SP, Song SY, Go S, Jung M, Hong J, Kim BS. M1 Macrophage-Derived Nanovesicles Potentiate the Anticancer Efficacy of Immune Checkpoint Inhibitors. *ACS Nano*. 2018; 12:8977–93.  
<https://doi.org/10.1021/acsnano.8b02446>  
PMID:30133260

# Integrative analysis of expression, prognostic significance and immune infiltration of RFC family genes in human sarcoma

Gen Wu<sup>1,2,\*</sup>, Jian Zhou<sup>1,\*</sup>, Xi Zhu<sup>3</sup>, Xianzhe Tang<sup>4</sup>, Jie Liu<sup>5</sup>, Qiong Zhou<sup>5</sup>, Ziyuan Chen<sup>6</sup>, Tang Liu<sup>1</sup>, Wanchun Wang<sup>1</sup>, Xungang Xiao<sup>4,&</sup>, Tong Wu<sup>7</sup>

<sup>1</sup>Department of Orthopedics, The Second Xiangya Hospital of Central South University, Changsha 410011, Hunan, China

<sup>2</sup>Clinical Medicine Eight-Year Program, Central South University, Changsha 410013, Hunan Province, China

<sup>3</sup>Department of Internal Medicine III, University Hospital, Ludwig Maximilian University, Munich 81377, Germany

<sup>4</sup>Department of Orthopedics, Chenzhou No.1 People's Hospital, Chenzhou 423000, Hunan, China

<sup>5</sup>Department of Cardiology, The Fourth Hospital of Changsha, Changsha 410006, Hunan, China

<sup>6</sup>Department of Orthopedics, The First People's Hospital of Changde City, Changde 415003, Hunan, China

<sup>7</sup>Department of Emergency, The First Hospital of Changsha, Changsha 410005, Hunan, China

\*Equal contribution

**Correspondence to:** Xungang Xiao, Tong Wu; **email:** [xxg8088@163.com](mailto:xxg8088@163.com), [tong.wu0727@hotmail.com](mailto:tong.wu0727@hotmail.com), <https://orcid.org/0000-0003-3107-702X>

**Keywords:** RFC, prognosis, expression, KEGG, bioinformatics analysis

**Received:** February 8, 2022

**Accepted:** April 13, 2022

**Published:** April 29, 2022

**Copyright:** © 2022 Wu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Objective:** To reveal the expression and prognostic value of replication factor C family genes (RFCs) in patients with sarcoma.

**Results:** The results showed that the mRNA expression levels of RFC2, RFC3, RFC4, and RFC5 were increased in sarcoma tissues. In addition, Cancer Cell Line Encyclopedia (CCLE) dataset analysis indicated that RFC1, RFC2, RFC3, RFC4, and RFC5 were elevated expressed in sarcoma cell lines. Moreover, Gene Expression Profiling Interactive Analysis (GEPIA) and Kaplan-Meier Plotter showed that highly expressed RFC2-5 were associated with poor overall survival (OS) or relapse-free survival (RFS) in sarcoma patients. The results of the Tumor Immune Estimation Resource (TIMER) database indicated that the expression of RFCs was negatively correlated with the infiltration of CD4+ T cells and macrophages.

**Conclusions:** There were significant differences in the expression of RFCs between normal tissue and sarcoma tissue, and RFC2, RFC3, RFC4, and RFC5 might be promising prognostic biomarkers for sarcoma.

**Methods:** The expression of RFCs was analyzed using the ONCOMINE dataset and GEPIA dataset. CCLE dataset was used to assess the expression of RFCs in the cancer cell line. The prognostic value of RFCs was evaluated by GEPIA and Kaplan-Meier analysis. Furthermore, the association between RFCs and their co-expressed genes were explored via ONCOMINE and GEPIA datasets. We used the TIMER dataset to analyze the immune cell infiltration of RFCs in sarcoma.

## INTRODUCTION

The replication factor C (RFC, activator 1) was first purified from the extracts of HeLa cells in human

cervical cancer, participates as an important host factor in the replication of DNA [1, 2]. As a primer identification factor for DNA polymerase, RFC is a DNA binding protein with a specific structure and

function [3]. *In vivo*, RFC plays an essential in cell biology cycles as a regulatory protein [4]. In humans, RFC is reported as a complex consisting of RFC1 (140 kDa), RFC2 (40 kDa), RFC3 (38 kDa), RFC4 (37 kDa) and RFC5 (36 kDa) subunits [5]. The binding of the five subunits determines the physiological function of RFC. According to reports [6, 7], RFC can participate in excision repair and mismatch repair of damaged DNA by initiating signal transduction downstream of the checkpoint at the site of DNA damage by binding to the cell cycle checkpoint protein. In addition, RFC can load DNA polymerase and proliferating cell nuclear antigen (PCNA) onto the primer-bound DNA template to form a DNA-RFC-PCNA-DNA polymerase complex. And then, the polymerase complex extended along with the DNA template in the presence of deoxynucleotides (dNTPs), via the action of human single-stranded DNA binding protein (hSSB) [4]. As for interacting partners with a variety of proteins, not only are RFC factors involved in multiple processes in the normal cell cycle, but RFC factors also play an essential role in the transcription and proliferation of tumor cells.

Further studies indicated that in the RFC family, different subunits have different roles in the cell cycle [4]. RFC1 DNA-binding domain contains the main, and of PCNA interacts directly with, involved in DNA synthesis, DNA repair, and cell cycle. Unlike other subunits, RFC1 is rarely reported to have a relationship with sarcoma. In the studies of Tang [8] and Pennaneach [9], it is pointed out that RFC1 can promote cell survival after DNA damage through the retinoblastoma (Rb) pathway, which is related to Hutchinson-Gilford Progeria Syndrome (HGPS). According to reports [10], RFC2 is one of the important components of the RFC complex that can unload PCNA and inhibit DNA polymerase activity, it is highly expressed in some sarcoma tissues and cells. RFC2, as a key gene, was upregulated in metastatic samples from Ewing's sarcoma patients [11]. Meanwhile, bioinformatics analysis showed that the up regulation of this key gene reduced the overall survival rate of Ewing's sarcoma patients. As the dominant gene in the 13q13 amplicon, RFC3 is considered to be an oncogene or anti-oncogene in different cancers based on its cellular and histological characteristics [12]. Recently, the study suggested that RFC3 is regulated by a series of miRNAs including miR-802 [13]. At the same time, it is reported that the up-regulated expression of miR-802 is shown in osteosarcoma tissues and promotes cell proliferation by targeting p27 in U27 OS and MG-63 cells [14]. Hence, RFC3 is also closely related to the cell proliferation of sarcoma tissue. In the DNA damage checkpoint pathway, RFC4 plays an important role and can enhance the anti-tumor activity of DNA-damaging chemotherapeutics [15]. A study has pointed out that

changes in cell cycle regulation occur in several types of cancer, including osteosarcoma [16]. RFC4 interacts with CDK1, MAD2L1, NDC80, and BUB1, and acts on cell mitosis and cell cycle [13]. RFC5 is a necessary subunit to open the PCNA clamp during DNA replication. RFC5 participates in the repair and regulation of mismatches, nucleotide excision, cell cycle, and DNA double helix damage, [17, 18]. Studies have suggested that RFC5 is significantly up-regulated in various cancer tissues or cells, and its expression increases as the disease progress [19–22]. However, the specific role of RFC5 in sarcoma is rarely expressed in more detail. So far, the expression program, functional role in sarcoma tissues, and impact on the prognosis of sarcoma patients by RFC5 are still poorly known.

In sarcoma patients, the pathological features conferred by RFC with different expression levels and their prognostic impact in these patients have been reported [4, 13]. To the best of our knowledge, there is still no research using bioinformatics to analyze the role of the RFC family in sarcoma. In our study, we summarized the expression and mutations of RFC genes in sarcoma to further analyze their process, latent function, and prognosis of sarcoma transcription levels.

## RESULTS

### The transcription level of RFCs in patients with sarcoma

In mammalian cells, there have been identified five kinds of RFC factors. In the ONCOMINE database, the transcription level of RFC in cancer tissues was different from that in normal tissues (Figure 1). The mRNA transcription level of RFCs showed a significant difference between normal and sarcoma patients, except RFC1. In Detwiller Sarcoma's dataset [23], RFC2, with a fold change of 3.287, was overexpressed in Fibrosarcoma (Table 1). In the database of Detwiller sarcoma [23], RFC3 expressed an increase in fibroids with a multiple change of 3.184. Detwiller sarcoma's dataset [23] showed that RFC3 expression factor with the increased expression: the change of RFC3 in Round Cell Liposarcoma was 3.588, the change of RFC3 in patients with Synovial Sarcoma was 2.548, and the change of patients with Leiomyosarcoma was 2.624 (Table 1). In Barretina Sarcoma's dataset [24], RFC3 was over-expressed than normal in the following sarcomas: 2.413 in myxoid/round cell liposarcoma, 2.257 in myxofibrosarcoma, 2.514 in leiomyosarcoma, and 2.539 in pleomorphic liposarcoma.

Detwiller Sarcoma's dataset [23] suggests that RFC4 over-expression was found in Leiomyosarcoma with a

fold change of 7.003, RFC4 over-expression was found in Pleomorphic Liposarcoma with a fold change of 3.658, RFC4 over-expression was found in Malignant Fibrous Histiocytoma with a fold change of 4.337, and RFC4 over-expression was found in Fibrosarcoma with of a fold change of 3.579. In Barretina Sarcoma's dataset [24], RFC4 was overexpressed in Leiomyosarcoma with a fold change of 7.827. Barretina Sarcoma's dataset [24] also indicated that RFC4 overexpression is found in Pleomorphic Liposarcoma with a fold change of 4.682. RFC4 over-expression was found in Myxofibrosarcoma with a fold change of 4.518, in Myxoid/Round Cell Liposarcoma with a

change of 3.952, and in Dedifferentiated Liposarcoma with a change of 3.099.

In the 2 databases, there were significant differences in mRNA transcription levels of RFC5. In Barretina Sarcoma's dataset [24], RFC5 over-expression was found in Myxofibrosarcoma with a change of 2.033 compared with normal, and in Pleomorphic Liposarcoma with a change of 2.097. In Detwiller Sarcoma's dataset [23], RFC5 over-expression was found in Leiomyosarcoma with a fold change of 5.371, in Fibrosarcoma with a change of 3.255, and in Malignant Fibrous Histiocytoma with a change of 4.134.

Analysis Type by Cancer	RFC1		RFC2		RFC3		RFC4		RFC5	
	Cancer vs. Normal									
Bladder Cancer							1		1	
Brain and CNS Cancer	1	1	1	1	1	1	2		2	3
Breast Cancer			1				2	1	1	
Cervical Cancer	1				3		4		4	
Colorectal Cancer	1		3		17		11			
Esophageal Cancer	1				1		2		1	
Gastric Cancer			1		4		1			
Head and Neck Cancer		1	1		1		6		4	
Kidney Cancer			2				2			
Leukemia	1				1	4				1
Liver Cancer	1				1		3			1
Lung Cancer		1			1		5		4	
Lymphoma					4		3		3	
Melanoma										
Myeloma			1							
Other Cancer		5	2		1		3	1	4	
Ovarian Cancer		1	1		1		2		2	
Pancreatic Cancer										
Prostate Cancer										
Sarcoma			1		8		9		5	
Significant Unique Analyses	5	9	13	2	44	5	55	2	30	5
Total Unique Analyses	436		413		465		403		434	

Figure 1. The transcription level of RFCs in patients with sarcoma.

**Table 1. The significant changes of rfc expression in transcription level between different types of sarcoma, NA: not available.**

Gene ID	Types of sarcoma vs. normal	Fold change	P value	t test	References
<b>RFC1</b>	NA	NA	NA	NA	NA
<b>RFC2</b>	Fibrosarcoma vs. Normal	3.287	6.42E-5	4.779	Detwiller Sarcoma
	Fibrosarcoma vs. Normal	3.184	5.00E-7	7.154	Detwiller Sarcoma
	Round Cell Liposarcoma vs. Normal	3.588	6.65E-7	7.393	Detwiller Sarcoma
	Synovial Sarcoma vs. Normal	2.548	6.13E-6	6.366	Detwiller Sarcoma
<b>RFC3</b>	Leiomyosarcoma vs. Normal	2.624	5.72E-5	5.351	Detwiller Sarcoma
	Myxoid/Round Cell Liposarcoma vs. Normal	2.413	5.49E-13	12.673	Barretina Sarcoma
	Myxofibrosarcoma vs. Normal	2.257	3.37E-11	9.037	Barretina Sarcoma
	Leiomyosarcoma vs. Normal	2.514	6.20E-10	8.700	Barretina Sarcoma
	Pleomorphic Liposarcoma vs. Normal	2.539	9.98E-8	7.117	Barretina Sarcoma
	Leiomyosarcoma vs. Normal	7.003	1.06E-9	10.790	Detwiller Sarcoma
	Pleomorphic Liposarcoma vs. Normal	3.658	4.03E-7	7.948	Detwiller Sarcoma
	Malignant Fibrous Histiocytoma vs. Normal	4.337	1.15E-7	7.444	Detwiller Sarcoma
<b>RFC4</b>	Fibrosarcoma vs. Normal	3.579	6.63E-7	6.866	Detwiller Sarcoma
	Leiomyosarcoma vs. Normal	7.827	1.46E-17	16.192	Barretina Sarcoma
	Pleomorphic Liposarcoma vs. Normal	4.682	4.85E-15	14.216	Barretina Sarcoma
	Myxofibrosarcoma vs. Normal	4.518	1.33E-15	17.566	Barretina Sarcoma
	Myxoid/Round Cell Liposarcoma vs. Normal	3.952	5.27E-12	18.791	Barretina Sarcoma
	Dedifferentiated Liposarcoma vs. Normal	3.099	3.56E-12	14.411	Barretina Sarcoma
	Myxofibrosarcoma vs. Normal	2.003	2.59E-13	10.719	Barretina Sarcoma
	Pleomorphic Liposarcoma vs. Normal	2.097	7.02E-9	7.985	Barretina Sarcoma
<b>RFC5</b>	Leiomyosarcoma vs. Normal	5.371	7.61E-6	6.193	Detwiller Sarcoma
	Fibrosarcoma vs. Normal	3.255	1.73E-5	5.368	Detwiller Sarcoma
	Malignant Fibrous Histiocytoma vs. Normal	4.134	8.80E-5	4.686	Detwiller Sarcoma

### Relationship between the mRNA transcription levels and the clinical pathological parameters in RFC in sarcoma patients

We use the Gene Expression Profiling Interactive Analysis (GEPIA) dataset (<http://gepia.cancer-pku.cn/>) to compare different mRNA expression levels of RFCs in sarcoma and normal samples. The results showed that RFC2, RFC4, and RFC5 were upregulated in sarcoma patients, while the high expression levels of RFC1 and RFC3 were both with no significance. (Figure 2A–2G).

### Expression of RFC transforming factors in sarcoma cell lines

Through the Cancer Cell Line Encyclopedia, we expanded our preclinical human cancer model of detailed annotation process (<https://www.broadinstitute.org/ccle>). The expressions of RFC1-5 were high in sarcoma cell lines (Figures 3A–3E).

### Prognostic value of RFCs in sarcoma

We investigated the prognostic analysis of RFC1-5 in sarcoma using the plotter tool in the GEPIA and Kaplan

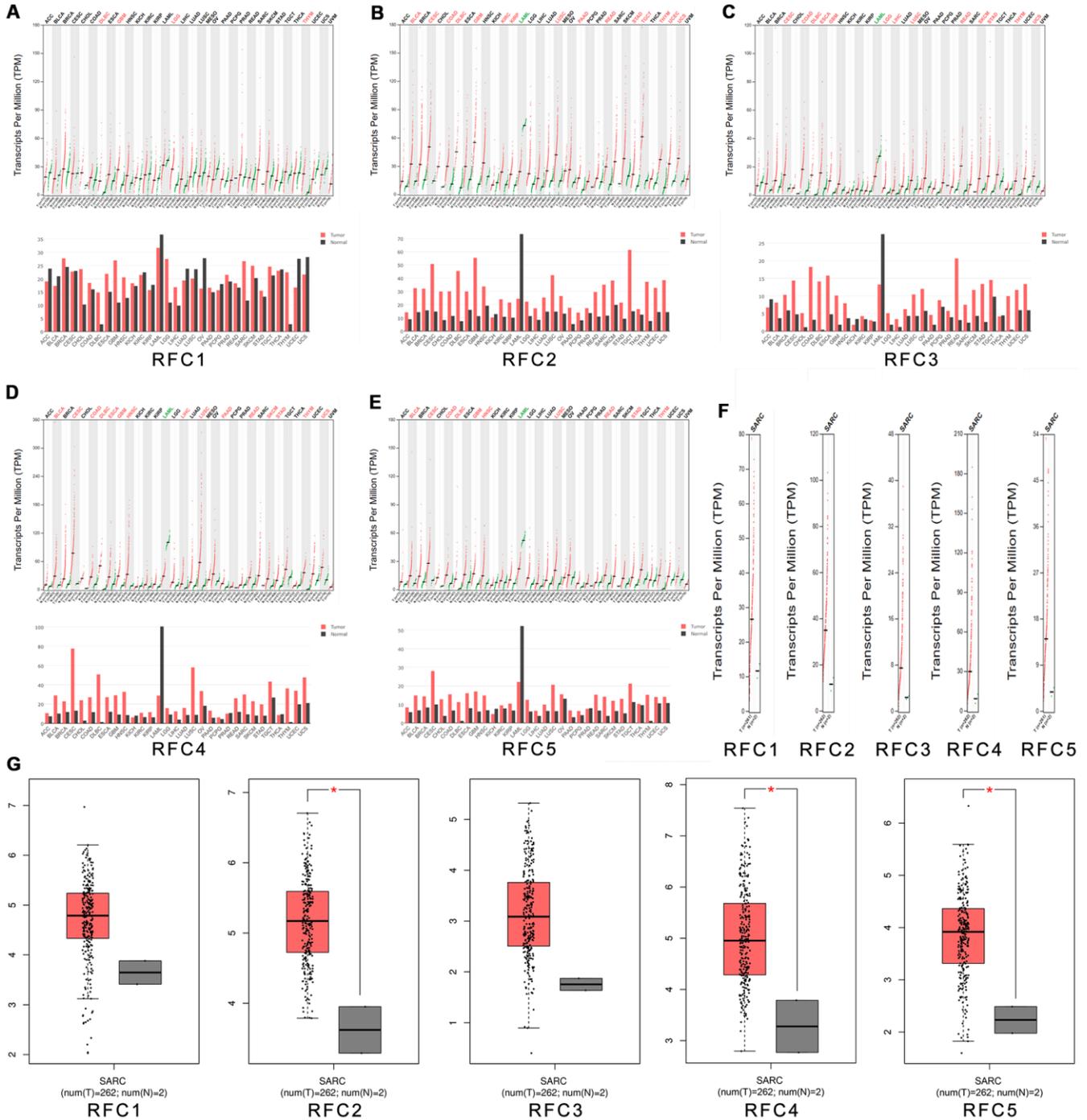
Meier databases (Kaplan Meier plotter). Interestingly, in these two databases, poor overall survival (OS) and disease-free survival (DFS) of sarcoma were related to the upregulation of RFC1, but with meaningless (Figure 4). The results, however, of the database suggested that high expression of RFC2 and RFC4 were associated with the poor DFS and RFS in sarcoma (Figure 4B, 4D), with statistical differences. Nevertheless, increased RFC3 and RFC5 mRNA levels were associated with poor OS and RFS in sarcoma (Figure 4A, 4C, 4D).

### Co-expressed RFC genes and the correction between RFCs in sarcoma

Analyzed genes co-expressed with RFC1, in Chen's study [25], we found that RFC1 has been positively corrected by AKAP13, DCLK1, GLB1, DOCK2, CLTC, LOC100128361, MGC11082, CXorf65, and SLC01A2. And then we analyzed genes co-expressed with RFC2 in the study of Stossi [26], the results showed that RFC2 has been positively corrected by MRPS12, RBP1, MARS, SHMT2, NDUFAF3, CSNK2B, CDK16, DNAJB1, PDLIM4, MFAP2, SF3B4, SMAGP, CKB, TLE2, MAPKAPK3, FLII, HIP1R, ARHGDI1, and TERF2. Analyzed genes co-expressed with RFC3 in the

study of Schaefer [27], we found that RFC3 has been positively corrected by MTCH2, CCDC86, TRAPPC3, LRRC59, SMCR7L, DDX3X, PNO1, PCMT1, EIF4E, GLRX3, ARPC4, SLC25A1, DDA1, SNAP23, API5, CLIC4, and VAMP3. Genes co-expressed with RFC4 were described in Chibon's study [28], the results

showed that RFC4 was positively corrected with MCM2, RMI1, NCAPG2, EZH2, FANCI, ZNF367, ATAD2, TYMS, RNASEH2A, ASF1B, and DTL. Genes co-expressed with RFC5 in the study of Chen [25], and we found RFC5 was positively corrected with ORC1L, RFC2, MRTO4, SDHIB, TMEM48, PPIH,



**Figure 2. The expression of RFCs in sarcoma.** (A) The expression of RFC1 in pan-cancer. (B) The expression of RFC2 in pan-cancer. (C) The expression of RFC3 in pan-cancer. (D) The expression of RFC4 in pan-cancer. (E) The expression of RFC5 in pan-cancer. (F, G) The expression of RFCs in SARC.



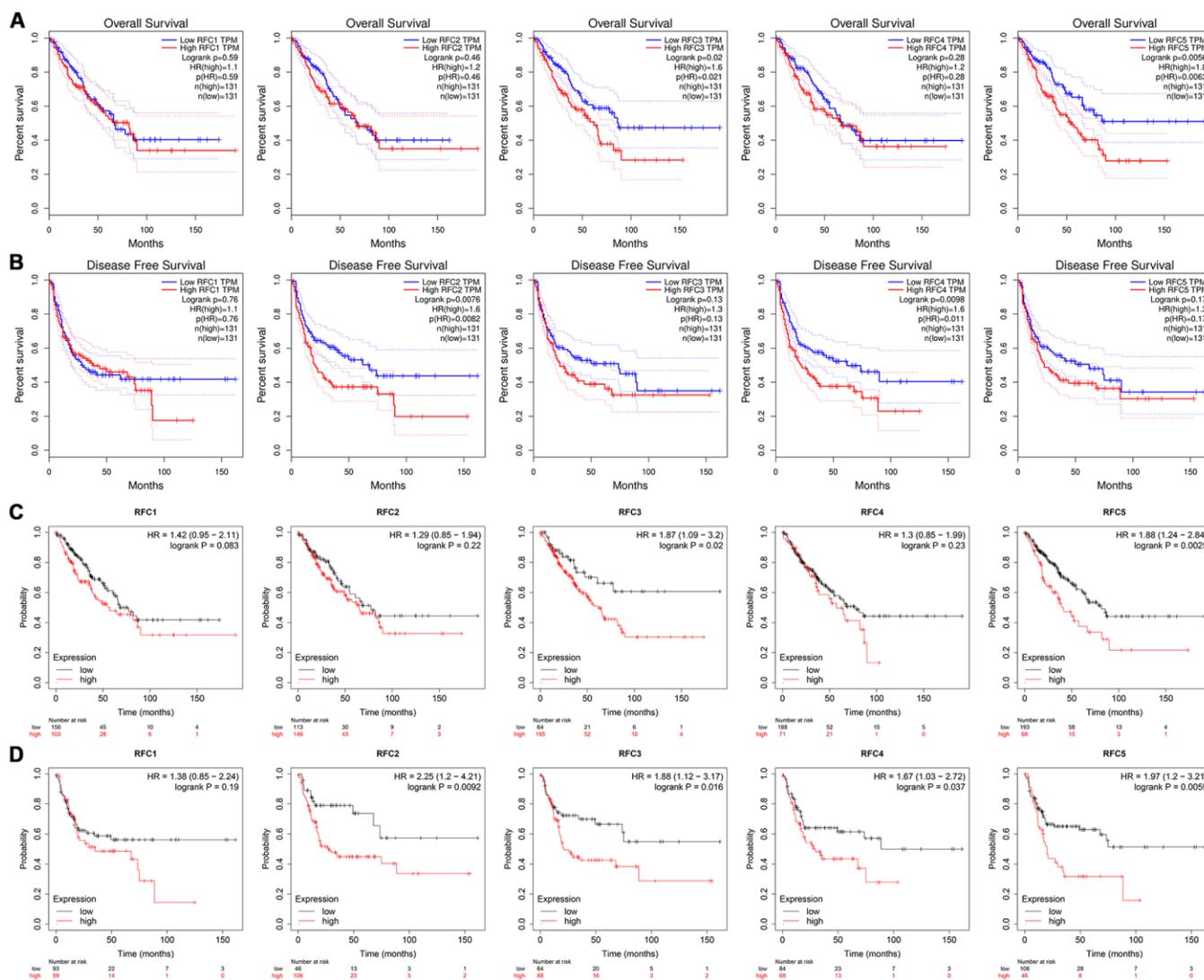
$p < 0.05$ ) (Figure 5B). RFC3 was both positively corrected by RFC4 ( $R = 0.59, p < 0.05$ ), and RFC5 ( $R = 0.52, p < 0.05$ ). RFC4 was positively corrected by RFC5 ( $R = 0.5, p < 0.05$ ) (Figure 5B).

### Association between RFCs and infiltration levels of immune cells in sarcoma

We explored the relationship between the expression level of RFCs and immune cell infiltration in sarcoma using the TIMER database. The results showed that RFC1-5 were all positively correlated with tumor purity, negatively correlated with the infiltration of CD4+ T cell and macrophage, with significance. (Figure 6).

## DISCUSSION

The role of RFC factor dysregulation in many types of cancer has been reported [29–32]. Similarly, we found that mRNA expression levels of RFC2, RFC3, RFC4, and RFC5 from RFC family genes were upregulated in sarcoma tissues. The imbalance of RFC may lead to the disorder of the cell cycle, which is no longer regulated by normal physiological mechanisms, and may further lead to the cancerization of normal tissues. RFC1 is generally considered to be an important part of DNA replication and repair in the RFC family [33]. Moggs et al. [34] found that when the expression of RFC1 is inhibited, the proliferation of estrogen receptor-negative breast cancer cells is also inhibited. Bermudez et al. [35]



**Figure 4. The prognostic value of mRNA level of RFC factors in sarcoma patients. (A, B)** The prognostic value of mRNA level of RFC factors in sarcoma patients, analyzed by GEPIA. **(C, D)** The prognostic value of mRNA level of RFC factors in sarcoma patients, analyzed by Kaplan-Meier plotter.

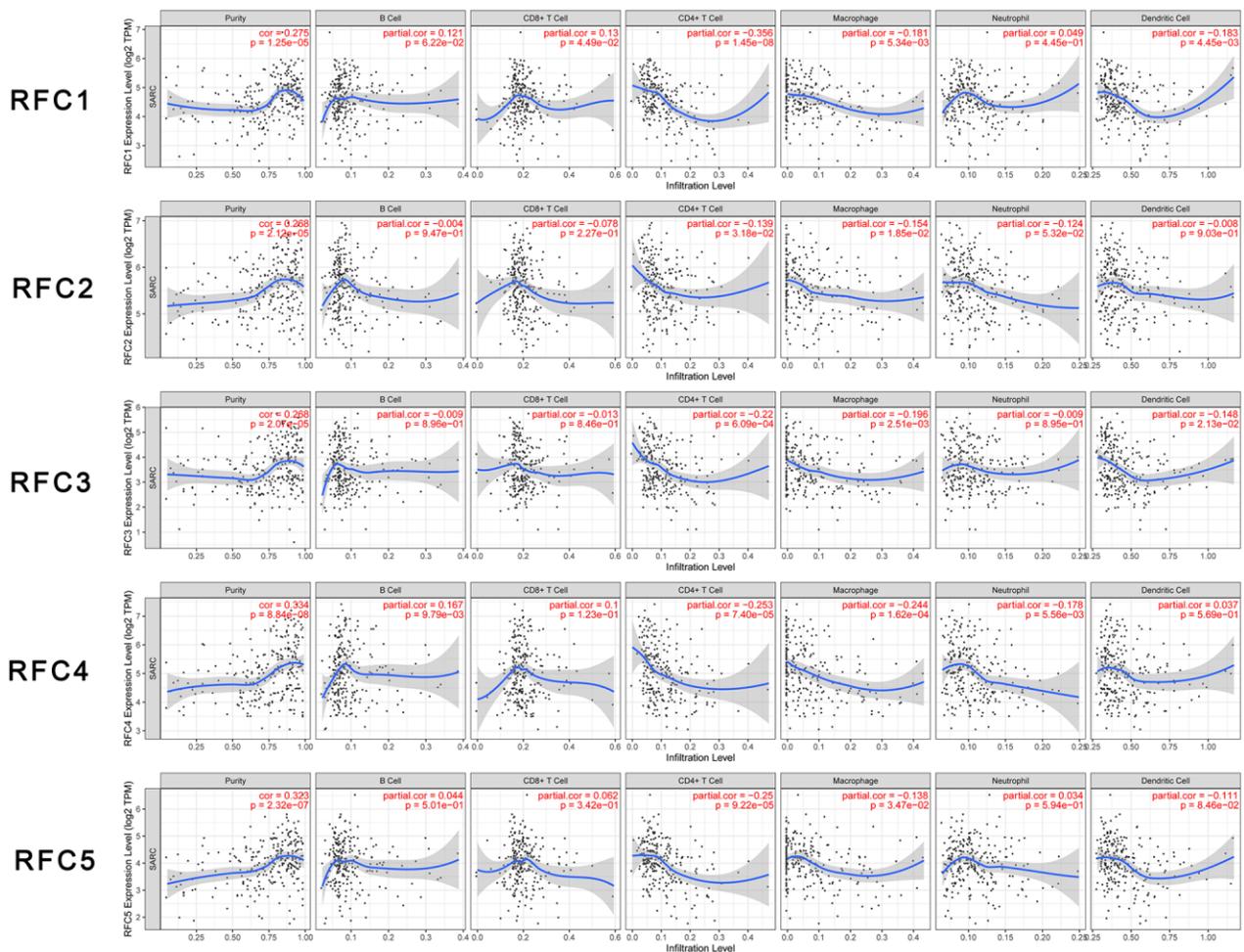


shows a high expression status, we can still consider that RFC1 gene is highly expressed in sarcoma cells. Using the GEPIA and Kaplan Meier plotter databases, we determined the prognostic value of the RFC1 gene in sarcoma, and although high expression of RFC1 was associated with poor OS and DFS in sarcoma patients, it was not statistically significant, so RFC1 expression did not predict sarcoma patients related prognosis.

Some evidence suggested that RFC2 was a key gene, and its upregulation was related to the metastasis and prognosis of Ewing's sarcoma [11]. Xiong et al. [37] showed that RFC2 was closely related to nasopharyngeal carcinoma, and the up regulation of RFC2 expression was obviously high in patients with nasopharyngeal carcinoma than normal. Meanwhile, other studies have shown that the expression level of RFC2 was also higher in the normal tissues than in patients with glioblastoma [38]. There was a study pointed out that in patients with choriocarcinoma, high

expression of RFC2 may have the effect of predicting the prognosis of the disease [39]. Therefore, RFC2 may play an oncogene in a variety of malignant tumors. Until now, there is no definite research showing the specific effect of RFC2 in sarcoma. In our study, the expression of RFC2 in fibrosarcoma was 3.287 times higher than in normal tissues. Through the CCLE database, we found that RFC2 was also up-regulated in sarcoma cell lines. Meanwhile, in the analysis of GEPIA and Kaplan Meier plotter, we also found that patients with sarcoma with high expression of RFC2 had poor disease-free survival (DFS), which was statistically significant. Therefore, RFC2 not only plays an oncogene role in fibrosarcoma but also has a latent function in predicting disease outcomes in sarcoma.

In a study on osteosarcoma [13], it was pointed out that RFC3, CDK1, MAD2L1, NDC80, BUB1, etc. jointly participated in the related links of the disease prognosis



**Figure 6. Relationship between differentially expressed RFC genes and immune cell infiltration.** The immune cells we analyzed included B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells.

of osteosarcoma. According to the study of patients with acute myeloid leukemia (AML) [40], RFC3 was found to be highly expressed in their tumor cells. In addition, in the study of ovarian cancer patients [12], the average survival time of patients with high RFC3 expression levels was only 7.7 months, while that of patients with normal expression was as long as 92.9 months. Highly expressed RFC3 has been confirmed to be a predictive gene in Kaposi's sarcoma [41], breast cancer [42], esophageal adenocarcinoma [43], and hepatocellular carcinoma [44]. In summary, this means that RFC3 may act as an oncogene in cancers. Consistent with our analysis, in this study, the RFC3 high expression was found in sarcoma tissue. Using the CCLE database, the results suggested that RFC3 was also highly expressed in human sarcoma. In ONCOMINE, the mRNA expression level of RFC3 was up-regulated, which was specifically reflected in fibrosarcoma, leiomyosarcoma, mucosal fibrosarcoma, smooth muscle sarcoma, pleomorphic liposarcoma, and so on. Meanwhile, the results illustrated that highly expression RFC3 was associated with poor OS in patients with sarcoma, which was statistically significant.

Previous study has shown that RFC4 is highly expressed in cancer tissues such as hepatocellular carcinoma [15], non-small cell lung cancer (NSCLC) [45], prostate cancer [46], breast cancer [20], and cervical cancer [47]. In addition, some scholars further proposed that RFC4 may serve as a potential prognostic biomarker and therapeutic target [4]. In general, RFC4 has been discussed and studied in various cancers, but its exact function in sarcoma has not been described yet. At the same time, highly expressed RFC4 was related to the RFS difference in sarcoma, and there was statistical significance, that is, the DFS of sarcoma patients with high expression of RFC4 was worse. It seemed that RFC4 not only acted as an oncogene in sarcoma patients but also had a certain predictive effect on the prognosis of sarcoma patients.

RFC5 can repair mismatches, DNA double helix damage, and nucleotide excision during the cell cycle [18]. These biological characteristics have been confirmed in research to be related to the progression of cancer [17]. Similar to other subunits of the RFC family, RFC5 is expressed in a variety of cancers, such as head and neck squamous cell carcinoma [48], prostate cancer [49], cervical cancer [50], and diffuse large B-cell lymphoma (DLBCL) [51] than the normal. Hence the exact function of RFC5 in sarcoma is still inconclusive. In our study, highly expressed RFC5 was significantly correlated with poor OS in sarcoma patients, and there was statistical significance. It suggests that patients with high expression of rfc5 sarcoma usually have poor overall survival.

The expression levels of the five RFC family members were negatively correlated with the infiltration of CD4+ T cells and macrophages. Immune cells in the tumor microenvironment can promote or inhibit tumor activity, which is considered to be an important determinant of clinical outcomes and immunotherapy. The RFCs expression levels were negatively correlated with the infiltration level of CD4+ T cells and macrophages. Past studies have shown that CD4+ T cells played active roles in anti-tumor immunity. CD4+ T cells could target tumor cells in a variety of ways, directly eliminate tumor cells through the cytolytic mechanism, or indirectly interact with tumor cells by regulating the tumor immune microenvironment [52]. This indicated that the expression of RFCs might affect the development of sarcoma by regulating the infiltration of immune cells, thereby affecting the prognosis of patients. This study could provide more detailed immune information for the immunotherapy of sarcoma patients.

After understanding the heterogeneity and complexity of the molecular biology of sarcoma, this study systematically analyzed the expression and predictive value of RFC in sarcoma. Our results indicated that increased expression of RFC2-5 in sarcoma tissue might play an important role in the occurrence of sarcoma and could be used as a potential indicator of diagnosis. The expression level of RFC2-5 has predictive effects on the survival period of patients with sarcoma. Thus, transcriptional RFC2-5 are latent prognostic markers to improve the survival and prognostic accuracy of sarcoma.

However, such a study still has some limitations. First, it can only respond to some relationships between RFC factors and sarcomas, and the role of some subunits in sarcomas under a partially defined pathological classification, so it cannot make a deeper analysis. Second, disadvantages still exist in the precision of treatment and prediction, but some genomic, as well as proteomic evidence, can be provided for the study of the corresponding sarcomas. Further studies also need to combine immunohistochemical as well as correlative analysis of tumor cytology, which will facilitate the diagnosis and treatment of sarcomas with greater precision. Third, the ONCOMINE database had been taken offline on January 17, 2022, so the figure with case numbers in specific subtype vs 'normal' cases cannot be obtained.

## CONCLUSIONS

In conclusion, this finding systematically showed the expression of the RFC gene in sarcoma tissue and the prognosis effect on sarcoma patients. Our results

indicate that increased expression of RFC2-5 in sarcoma tissue may show an essential role in the occurrence of sarcoma and can be used as latent indicators for diagnosing sarcoma. Expression levels of RFC2-5 have a predictive effect on the total survival (OS) or disease-free survival (DFS) of sarcoma patients. Therefore, transcription of RFC2-5 is a potential prognostic marker for improving survival and prognostic accuracy in sarcoma patients. The results of this study may provide new ideas for diagnosis and prognosis in sarcoma patients to select potential prognostic biomarkers.

## MATERIALS AND METHODS

### ONCOMINE analysis

ONCOMINE (<https://www.oncomine.com/>) is an online microarray database focused on cancer. In different cancers, it can analyze the transcript levels of the RFC gene family. Comparison of mRNA expression of the RFC gene family in clinical cancer tissue specimens and mRNA expression in normal controls was performed using the student's t-test for significance of the mean difference. Respectively, the cut-off was defined as 0.01, and the fold change of p-value was 2.

### GEPIA dataset

Gene expression analysis interactive analysis dataset (GEPIA), a newly developed interactive online gene bioinformatics analysis platform, is a way of RNA sequencing analysis. The database contains data from 9736 tumor tissues and 8587 normal tissues from The Cancer Genome Atlas (TCGA) as well as Genotype-Tissue Expression (GTEx) (<http://gepia.cancer-pku.cn/>). GEPIA shows a series of custom-made features such as gene differential expression analysis of tumor tissue versus normal tissue, but also correlation analysis depending on the type of pathological stage of cancer, analysis of patient survival, similar genetic tests, correlation analysis, and dimension reduction analysis.

### CCLC dataset

CCLC (Cancer Cell Line Encyclopedia), an oncogenomics research project, is led by the Broad Institute of MIT at Harvard. Similar to the above databases, CCLC is also a large, publicly available tumor genome database. It collects and collates profiling data of 1457 tumor cell lines, 84434 genomes, which includes gene expression data, chromosomal copy data as well as massively parallel sequencing data. In sarcoma cell lines, we illustrated the expression levels of the RFC gene family using the CCLC database.

### Kaplan-Meier plotter

The Kaplan Meier plotter is a publicly available online database (<https://www.kmplot.com>) constructed based on microarray and RNA SEQ data from public databases such as TCGA, geo, and EGA. The prognostic value of signal transducer and activator of transcription protein (STAT), as well as mRNA expression, was assessed using this online database containing expression data and survival information for the analyzed genes as well as 259 clinical sarcoma patients. To analyze the overall survival (OS) and progression-free survival (RFS), of patients with sarcoma, those samples were divided into two groups by median expression (high expression vs. low expression) and evaluated by Kaplan Meier survival plots with hazard ratios (HR, hazard ratio) with 95% confidence intervals (CI, confidence interval) and log-rank P values. TIMER database

Through the TIMER database, we evaluated the effect of RFC expression level on tumor immune cell infiltration. TIMER database contains the data of tumor-infiltrating immune cells in more than 10000 samples of 32 types of cancers from TCGA [53]. On this basis, we analyzed the effects of the expression level of the RFC gene family on six common immune cell infiltration levels: CD4 + T cells, CD8 + T cells, B cells, neutrophils, dendritic cells, and macrophages.

### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## AUTHOR CONTRIBUTIONS

Jian Zhou, Gen Wu, and Wanchun Wang conceived, designed, and conducted the experiments. Gen Wu and Tong Wu wrote the paper. Gen Wu and Tong Wu collected the data. Ziyuan Chen, Tang Liu, Gen Wu, and Wanchun Wang edited the paper. Gen Wu and Tong Wu provided the research guide. Xi Zhu, Xianzhe Tang, Jie Liu, Qiong Zhou and Xungang Xiao revised the manuscript. Wanchun Wang supervised this project. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

The authors would like to thank all co-investigators, and colleagues who made this study possible.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## FUNDING

This work was supported by the Chenzhou Municipal Science and Technology Bureau Science and Technology Development Program Project (Grant No. yfzx201910), Mittal Innovation Project of Central South University (Grant No. GCX20190879Y), the Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2018zzts930), the Central South University Sports Medicine Scholarship and the National College Students' Innovation, and Entrepreneurship Training Program (Grant No. 201710422116).

## Editorial note

<sup>&</sup>This corresponding author has a verified history of publications using a personal email address for correspondence.

## REFERENCES

- Okumura K, Nogami M, Taguchi H, Dean FB, Chen M, Pan ZQ, Hurwitz J, Shiratori A, Murakami Y, Ozawa K. Assignment of the 36.5-kDa (RFC5), 37-kDa (RFC4), 38-kDa (RFC3), and 40-kDa (RFC2) subunit genes of human replication factor C to chromosome bands 12q24.2-q24.3, 3q27, 13q12.3-q13, and 7q11.23. *Genomics*. 1995; 25:274–8. [https://doi.org/10.1016/0888-7543\(95\)80135-9](https://doi.org/10.1016/0888-7543(95)80135-9) PMID:7774928
- Tsurimoto T, Stillman B. Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication *in vitro*. *Mol Cell Biol*. 1989; 9:609–19. <https://doi.org/10.1128/mcb.9.2.609-619.1989> PMID:2565531
- Zhou Y, Hingorani MM. Impact of individual proliferating cell nuclear antigen-DNA contacts on clamp loading and function on DNA. *J Biol Chem*. 2012; 287:35370–81. <https://doi.org/10.1074/jbc.M112.399071> PMID:22902629
- Li Y, Gan S, Ren L, Yuan L, Liu J, Wang W, Wang X, Zhang Y, Jiang J, Zhang F, Qi X. Multifaceted regulation and functions of replication factor C family in human cancers. *Am J Cancer Res*. 2018; 8:1343–55. PMID:30210909
- Bowman GD, O'Donnell M, Kuriyan J. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature*. 2004; 429:724–30. <https://doi.org/10.1038/nature02585> PMID:15201901
- Majka J, Burgers PM. The PCNA-RFC families of DNA clamps and clamp loaders. *Prog Nucleic Acid Res Mol Biol*. 2004; 78:227–60. [https://doi.org/10.1016/S0079-6603\(04\)78006-X](https://doi.org/10.1016/S0079-6603(04)78006-X) PMID:15210332
- Sakato M, O'Donnell M, Hingorani MM. A central swivel point in the RFC clamp loader controls PCNA opening and loading on DNA. *J Mol Biol*. 2012; 416:163–75. <https://doi.org/10.1016/j.jmb.2011.12.017> PMID:22197374
- Tang H, Hilton B, Musich PR, Fang DZ, Zou Y. Replication factor C1, the large subunit of replication factor C, is proteolytically truncated in Hutchinson-Gilford progeria syndrome. *Aging Cell*. 2012; 11:363–5. <https://doi.org/10.1111/j.1474-9726.2011.00779.x> PMID:22168243
- Pennaneach V, Salles-Passador I, Munshi A, Brickner H, Regazzoni K, Dick F, Dyson N, Chen TT, Wang JY, Fotedar R, Fotedar A. The large subunit of replication factor C promotes cell survival after DNA damage in an LxCxE motif- and Rb-dependent manner. *Mol Cell*. 2001; 7:715–27. [https://doi.org/10.1016/s1097-2765\(01\)00217-9](https://doi.org/10.1016/s1097-2765(01)00217-9) PMID:11336696
- Rakhee G. Replication factor C-40 (RFC40/RFC2) as a prognostic marker and target in estrogen positive and negative and triple negative breast cancer. 2015.
- Li G, Zhang P, Zhang W, Lei Z, He J, Meng J, Di T, Yan W. Identification of key genes and pathways in Ewing's sarcoma patients associated with metastasis and poor prognosis. *Onco Targets Ther*. 2019; 12:4153–65. <https://doi.org/10.2147/OTT.S195675> PMID:31213834
- Shen H, Cai M, Zhao S, Wang H, Li M, Yao S, Jiang N. Overexpression of RFC3 is correlated with ovarian tumor development and poor prognosis. *Tumour Biol*. 2014; 35:10259–66. <https://doi.org/10.1007/s13277-014-2216-2> PMID:25030735
- Peng C, Yang Q, Wei B, Yuan B, Liu Y, Li Y, Gu D, Yin G, Wang B, Xu D, Zhang X, Kong D. Investigation of crucial genes and microRNAs in conventional osteosarcoma using gene expression profiling analysis. *Mol Med Rep*. 2017; 16:7617–24. <https://doi.org/10.3892/mmr.2017.7506> PMID:28944822
- Cao ZQ, Shen Z, Huang WY. MicroRNA-802 promotes osteosarcoma cell proliferation by targeting p27. *Asian Pac J Cancer Prev*. 2013; 14:7081–4. <https://doi.org/10.7314/apjcp.2013.14.12.7081> PMID:24460254

15. Arai M, Kondoh N, Imazeki N, Hada A, Hatsuse K, Matsubara O, Yamamoto M. The knockdown of endogenous replication factor C4 decreases the growth and enhances the chemosensitivity of hepatocellular carcinoma cells. *Liver Int.* 2009; 29:55–62.  
<https://doi.org/10.1111/j.1478-3231.2008.01792.x>  
PMID:[18492021](https://pubmed.ncbi.nlm.nih.gov/18492021/)
16. Champeris Tsaniras S, Kanellakis N, Symeonidou IE, Nikolopoulou P, Lygerou Z, Taraviras S. Licensing of DNA replication, cancer, pluripotency and differentiation: an interlinked world? *Semin Cell Dev Biol.* 2014; 30:174–80.  
<https://doi.org/10.1016/j.semcdb.2014.03.013>  
PMID:[24641889](https://pubmed.ncbi.nlm.nih.gov/24641889/)
17. Ryu DS, Baek GO, Kim EY, Kim KH, Lee DS. Effects of polysaccharides derived from *Orostachys japonicus* on induction of cell cycle arrest and apoptotic cell death in human colon cancer cells. *BMB Rep.* 2010; 43:750–5.  
<https://doi.org/10.5483/BMBRep.2010.43.11.750>  
PMID:[21110919](https://pubmed.ncbi.nlm.nih.gov/21110919/)
18. Liu SM, Chen W, Wang J. Distinguishing between cancer cell differentiation and resistance induced by all-trans retinoic acid using transcriptional profiles and functional pathway analysis. *Sci Rep.* 2014; 4:5577.  
<https://doi.org/10.1038/srep05577> PMID:[24993014](https://pubmed.ncbi.nlm.nih.gov/24993014/)
19. Varghese S, Xu H, Bartlett D, Hughes M, Pingpank JF, Beresnev T, Alexander HR Jr. Isolated hepatic perfusion with high-dose melphalan results in immediate alterations in tumor gene expression in patients with metastatic ocular melanoma. *Ann Surg Oncol.* 2010; 17:1870–7.  
<https://doi.org/10.1245/s10434-010-0998-z>  
PMID:[20221901](https://pubmed.ncbi.nlm.nih.gov/20221901/)
20. Fatima A, Tariq F, Malik MF, Qasim M, Haq F. Copy Number Profiling of MammaPrint™ Genes Reveals Association with the Prognosis of Breast Cancer Patients. *J Breast Cancer.* 2017; 20:246–53.  
<https://doi.org/10.4048/jbc.2017.20.3.246>  
PMID:[28970850](https://pubmed.ncbi.nlm.nih.gov/28970850/)
21. Loo SK, Ab Hamid SS, Musa M, Wong KK. DNMT1 is associated with cell cycle and DNA replication gene sets in diffuse large B-cell lymphoma. *Pathol Res Pract.* 2018; 214:134–43.  
<https://doi.org/10.1016/j.prp.2017.10.005>  
PMID:[29137822](https://pubmed.ncbi.nlm.nih.gov/29137822/)
22. Tu S, Zhang H, Yang X, Wen W, Song K, Yu X, Qu X. Screening of cervical cancer-related hub genes based on comprehensive bioinformatics analysis. *Cancer Biomark.* 2021; 32:303–15.  
<https://doi.org/10.3233/CBM-203262> PMID:[34151839](https://pubmed.ncbi.nlm.nih.gov/34151839/)
23. Detwiler KY, Fernando NT, Segal NH, Ryeom SW, D'Amore PA, Yoon SS. Analysis of hypoxia-related gene expression in sarcomas and effect of hypoxia on RNA interference of vascular endothelial cell growth factor A. *Cancer Res.* 2005; 65:5881–9.  
<https://doi.org/10.1158/0008-5472.CAN-04-4078>  
PMID:[15994966](https://pubmed.ncbi.nlm.nih.gov/15994966/)
24. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Socci ND, Weir BA, Ho A, Chiang DY, Reva B, Mermel CH, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat Genet.* 2010; 42:715–21.  
<https://doi.org/10.1038/ng.619> PMID:[20601955](https://pubmed.ncbi.nlm.nih.gov/20601955/)
25. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, Lai KM, Ji J, Dudoit S, Ng IO, Van De Rijn M, Botstein D, Brown PO. Gene expression patterns in human liver cancers. *Mol Biol Cell.* 2002; 13:1929–39.  
<https://doi.org/10.1091/mbc.02-02-0023>  
PMID:[12058060](https://pubmed.ncbi.nlm.nih.gov/12058060/)
26. Stossi F, Barnett DH, Frasor J, Komm B, Lyttle CR, Katzenellenbogen BS. Transcriptional profiling of estrogen-regulated gene expression via estrogen receptor (ER) alpha or ERbeta in human osteosarcoma cells: distinct and common target genes for these receptors. *Endocrinology.* 2004; 145:3473–86.  
<https://doi.org/10.1210/en.2003-1682>  
PMID:[15033914](https://pubmed.ncbi.nlm.nih.gov/15033914/)
27. Schaefer KL, Eisenacher M, Braun Y, Brachwitz K, Wai DH, Dirksen U, Lanvers-Kaminsky C, Juergens H, Herrero D, Stegmaier S, Koscielniak E, Eggert A, Nathrath M, et al. Microarray analysis of Ewing's sarcoma family of tumours reveals characteristic gene expression signatures associated with metastasis and resistance to chemotherapy. *Eur J Cancer.* 2008; 44:699–709.  
<https://doi.org/10.1016/j.ejca.2008.01.020>  
PMID:[18294840](https://pubmed.ncbi.nlm.nih.gov/18294840/)
28. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, Tirode F, Lucchesi C, de Reynies A, Kauffmann A, Bui B, Terrier P, Bonvalot S, Le Cesne A, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med.* 2010; 16:781–7.  
<https://doi.org/10.1038/nm.2174>  
PMID:[20581836](https://pubmed.ncbi.nlm.nih.gov/20581836/)
29. Deng J, Zhong F, Gu W, Qiu F. Exploration of Prognostic Biomarkers among Replication Factor C Family in the Hepatocellular Carcinoma. *Evol Bioinform Online.* 2021; 17:1176934321994109.  
<https://doi.org/10.1177/1176934321994109>  
PMID:[33628006](https://pubmed.ncbi.nlm.nih.gov/33628006/)
30. Ji Z, Li J, Wang J. Up-regulated RFC2 predicts unfavorable progression in hepatocellular carcinoma.

- Hereditas. 2021; 158:17.  
<https://doi.org/10.1186/s41065-021-00179-9>  
PMID:[34022962](https://pubmed.ncbi.nlm.nih.gov/34022962/)
31. Zhao X, Wang Y, Li J, Qu F, Fu X, Liu S, Wang X, Xie Y, Zhang X. RFC2: a prognosis biomarker correlated with the immune signature in diffuse lower-grade gliomas. *Sci Rep*. 2022; 12:3122.  
<https://doi.org/10.1038/s41598-022-06197-5>  
PMID:[35210438](https://pubmed.ncbi.nlm.nih.gov/35210438/)
  32. Zhang J, Wang L, Xie X. RFC4 promotes the progression and growth of Oral Tongue squamous cell carcinoma *in vivo* and *in vitro*. *J Clin Lab Anal*. 2021; 35:e23761.  
<https://doi.org/10.1002/jcla.23761> PMID:[33783864](https://pubmed.ncbi.nlm.nih.gov/33783864/)
  33. Maradeo ME, Garg A, Skibbens RV. Rfc5p regulates alternate RFC complex functions in sister chromatid pairing reactions in budding yeast. *Cell Cycle*. 2010; 9:4370–8.  
<https://doi.org/10.4161/cc.9.21.13634>  
PMID:[20980821](https://pubmed.ncbi.nlm.nih.gov/20980821/)
  34. Moggs JG, Murphy TC, Lim FL, Moore DJ, Stuckey R, Antrobus K, Kimber I, Orphanides G. Anti-proliferative effect of estrogen in breast cancer cells that re-express ERalpha is mediated by aberrant regulation of cell cycle genes. *J Mol Endocrinol*. 2005; 34:535–51.  
<https://doi.org/10.1677/jme.1.01677> PMID:[15821115](https://pubmed.ncbi.nlm.nih.gov/15821115/)
  35. Bermudez VP, Lindsey-Boltz LA, Cesare AJ, Maniwa Y, Griffith JD, Hurwitz J, Sancar A. Loading of the human 9-1-1 checkpoint complex onto DNA by the checkpoint clamp loader hRad17-replication factor C complex *in vitro*. *Proc Natl Acad Sci USA*. 2003; 100:1633–8.  
<https://doi.org/10.1073/pnas.0437927100>  
PMID:[12578958](https://pubmed.ncbi.nlm.nih.gov/12578958/)
  36. Celic I, Verreault A, Boeke JD. Histone H3 K56 hyperacetylation perturbs replisomes and causes DNA damage. *Genetics*. 2008; 179:1769–84.  
<https://doi.org/10.1534/genetics.108.088914>  
PMID:[18579506](https://pubmed.ncbi.nlm.nih.gov/18579506/)
  37. Xiong S, Wang Q, Zheng L, Gao F, Li J. Identification of candidate molecular markers of nasopharyngeal carcinoma by tissue microarray and *in situ* hybridization. *Med Oncol*. 2011 (Suppl 1); 28:S341–8.  
<https://doi.org/10.1007/s12032-010-9727-5>  
PMID:[21057896](https://pubmed.ncbi.nlm.nih.gov/21057896/)
  38. Kun S, Duan Q, Liu G, Lu JM. Prognostic value of DNA repair genes based on stratification of glioblastomas. *Oncotarget*. 2017; 8:58222–30.  
<https://doi.org/10.18632/oncotarget.17452>  
PMID:[28938550](https://pubmed.ncbi.nlm.nih.gov/28938550/)
  39. Cui JQ, Shi YF, Zhou HJ. [Expression of RFC2 and PCNA in different gestational trophoblastic diseases]. *Ai Zheng*. 2004; 23:196–200.  
PMID:[14960244](https://pubmed.ncbi.nlm.nih.gov/14960244/)
  40. Hatfield KJ, Reikvam H, Bruserud Ø. Identification of a subset of patients with acute myeloid leukemia characterized by long-term *in vitro* proliferation and altered cell cycle regulation of the leukemic cells. *Expert Opin Ther Targets*. 2014; 18:1237–51.  
<https://doi.org/10.1517/14728222.2014.957671>  
PMID:[25200484](https://pubmed.ncbi.nlm.nih.gov/25200484/)
  41. Sun Q, Tsurimoto T, Juillard F, Li L, Li S, De León Vázquez E, Chen S, Kaye K. Kaposi's sarcoma-associated herpesvirus LANA recruits the DNA polymerase clamp loader to mediate efficient replication and virus persistence. *Proc Natl Acad Sci USA*. 2014; 111:11816–21.  
<https://doi.org/10.1073/pnas.1404219111>  
PMID:[25071216](https://pubmed.ncbi.nlm.nih.gov/25071216/)
  42. He ZY, Wu SG, Peng F, Zhang Q, Luo Y, Chen M, Bao Y. Up-Regulation of RFC3 Promotes Triple Negative Breast Cancer Metastasis and is Associated With Poor Prognosis Via EMT. *Transl Oncol*. 2017; 10:1–9.  
<https://doi.org/10.1016/j.tranon.2016.10.004>  
PMID:[27888707](https://pubmed.ncbi.nlm.nih.gov/27888707/)
  43. Lockwood WW, Thu KL, Lin L, Pikor LA, Chari R, Lam WL, Beer DG. Integrative genomics identified RFC3 as an amplified candidate oncogene in esophageal adenocarcinoma. *Clin Cancer Res*. 2012; 18:1936–46.  
<https://doi.org/10.1158/1078-0432.CCR-11-1431>  
PMID:[22328562](https://pubmed.ncbi.nlm.nih.gov/22328562/)
  44. Yao Z, Hu K, Huang H, Xu S, Wang Q, Zhang P, Yang P, Liu B. shRNA-mediated silencing of the RFC3 gene suppresses hepatocellular carcinoma cell proliferation. *Int J Mol Med*. 2015; 36:1393–9.  
<https://doi.org/10.3892/ijmm.2015.2350>  
PMID:[26397132](https://pubmed.ncbi.nlm.nih.gov/26397132/)
  45. Erdogan E, Klee EW, Thompson EA, Fields AP. Meta-analysis of oncogenic protein kinase Ciota signaling in lung adenocarcinoma. *Clin Cancer Res*. 2009; 15:1527–33.  
<https://doi.org/10.1158/1078-0432.CCR-08-2459>  
PMID:[19223491](https://pubmed.ncbi.nlm.nih.gov/19223491/)
  46. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res*. 2002; 62:4499–506.  
PMID:[12154061](https://pubmed.ncbi.nlm.nih.gov/12154061/)
  47. Bachtiry B, Boutros PC, Pintilie M, Shi W, Bastianutto C, Li JH, Schwock J, Zhang W, Penn LZ, Jurisica I, Fyles A, Liu FF. Gene expression profiling in cervical cancer: an exploration of intratumor heterogeneity. *Clin Cancer Res*. 2006; 12:5632–40.  
<https://doi.org/10.1158/1078-0432.CCR-06-0357>  
PMID:[17020965](https://pubmed.ncbi.nlm.nih.gov/17020965/)

48. Martinez I, Wang J, Hobson KF, Ferris RL, Khan SA. Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. *Eur J Cancer*. 2007; 43:415–32. <https://doi.org/10.1016/j.ejca.2006.09.001> PMID:[17079134](https://pubmed.ncbi.nlm.nih.gov/17079134/)
49. Barfeld SJ, East P, Zuber V, Mills IG. Meta-analysis of prostate cancer gene expression data identifies a novel discriminatory signature enriched for glycosylating enzymes. *BMC Med Genomics*. 2014; 7:513. <https://doi.org/10.1186/s12920-014-0074-9> PMID:[25551447](https://pubmed.ncbi.nlm.nih.gov/25551447/)
50. Liu D, Zhang XX, Xi BX, Wan DY, Li L, Zhou J, Wang W, Ma D, Wang H, Gao QL. Sine oculis homeobox homolog 1 promotes DNA replication and cell proliferation in cervical cancer. *Int J Oncol*. 2014; 45:1232–40. <https://doi.org/10.3892/ijo.2014.2510> PMID:[24970368](https://pubmed.ncbi.nlm.nih.gov/24970368/)
51. Peng WX, Han X, Zhang CL, Ge L, Du FY, Jin J, Gong AH. FoxM1-mediated RFC5 expression promotes temozolomide resistance. *Cell Biol Toxicol*. 2017; 33:527–37. <https://doi.org/10.1007/s10565-017-9381-1> PMID:[28185110](https://pubmed.ncbi.nlm.nih.gov/28185110/)
52. Kim HJ, Cantor H. CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. *Cancer Immunol Res*. 2014; 2:91–8. <https://doi.org/10.1158/2326-6066.CIR-13-0216> PMID:[24778273](https://pubmed.ncbi.nlm.nih.gov/24778273/)
53. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res*. 2017; 77:e108–10. <https://doi.org/10.1158/0008-5472.CAN-17-0307> PMID:[29092952](https://pubmed.ncbi.nlm.nih.gov/29092952/)

# Transcriptional expressions of hsa-mir-183 predicted target genes as independent indicators for prognosis in bladder urothelial carcinoma

Ming Li<sup>1,\*</sup>, Da-Ming Xu<sup>1,\*</sup>, Shu-Bin Lin<sup>1</sup>, Zheng-Liang Yang<sup>1</sup>, Teng-Yu Xu<sup>1</sup>, Jin-Huan Yang<sup>1</sup>, Ze-Xin Lin<sup>1</sup>, Ze-Kai Huang<sup>1</sup>, Jun Yin<sup>2,3</sup>

<sup>1</sup>Division of Urological Surgery, Second Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

<sup>2</sup>Division of Hematology, Second Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

<sup>3</sup>Department of Clinical Laboratory Medicine, Second Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

\*Equal contribution as co-first authors

**Correspondence to:** Jun Yin; email: [jyin@stu.edu.cn](mailto:jyin@stu.edu.cn)

**Keywords:** hsa-mir-183, predicted target gene, TP53, prognosis, bladder urothelial carcinoma

**Received:** November 3, 2021

**Accepted:** April 22, 2022

**Published:** May 3, 2022

**Copyright:** © 2022 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Objective:** To uncover novel prognostic and therapeutic targets for BLCA, our study is the first to investigate the role of hsa-mir-183 and its up-regulated predicted target genes in bladder urothelial carcinoma.

**Methods:** To address this issue, our study explored the roles of hsa-mir-183 predicted target genes in the prognosis of BLCA via UALCAN, Metascape, Kaplan-Meier plotter, Human Protein Atlas, TIMER2.0, cBioPortal and Genomics of Drug Sensitivity in Cancer databases.

**Results:** High transcriptional expressions of PDCD6, GNG5, PHF6 and MAL2 were markedly relevant to favorable OS in BLCA patients, whereas SLC25A15 and PTDSS1 had opposite expression significance. Additionally, high transcriptional expression of PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1 were significantly correlated with BLCA individual cancer stages and molecular subtypes. Furthermore, high mutation rate of PDCD6, MAL2, SLC25A15 and PTDSS1 were observed. Finally, TP53 mutation of PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1 has guiding significance for drug selection in BLCA.

**Conclusions:** PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1 could be the advanced independent indicators for prognosis of BLCA patients, and TP53-mutation might be a biomarker for drug option in BLCA patients.

## INTRODUCTION

Bladder Urothelial Carcinoma (BLCA) is one of the most common urinary system tumors with the unfavorable mortality and increased incidence in the world [1, 2]. The histological features of BLCA are the transformation between urothelium and malignant cells, involving a series process of polymorphic nuclei,

hyperchromatic nuclei, nucleolar herniation, and an increased number of uncontrolled mitoses [3]. Recently, the detailed molecular subtypes of BLCA have been widely explored, namely neuronal, basal squamous, luminal, luminal infiltrated and luminal papillary, which has important prognostic and therapeutic implications. Studies on BLCA have suggested that BLCA *in situ* without treatment may lead to progression to muscular

invasive disease (occur in 50%) and disease relapse (occur in 90%) [4, 5]. New insights on the pathogenesis, invasion and metastasis of BLCA have been ongoing in recent studies, however, the research on the gene molecular basis of bladder cancer has been limited. Providing crucial insight into the gene level characteristics of BLCA will help to better reveal the mechanism of pathogenesis and discover more significant diagnostic or prognostic biomarkers and therapeutic options.

MicroRNA (miRNA), a small noncoding RNA with the function of negative regulating gene expression, plays an important role in the occurrence and development of various tumors [6–8]. The functions of miRNA are not to directly block the expression of genes, but to make a slightly and specifically adjustment to the expression of target genes in post-transcriptional control [9, 10]. MiRNA plays an important regulatory role in the biological processes of normal human cells, including cell division cycle, cell growth, cell apoptosis, cell differentiation and cell response. Almost all tumors have different types of miRNA expression, which are either up-regulated or down-regulated, thus providing new opportunities for miRNA and their predicted target genes to be used as the potential diagnostic or prognostic biomarkers. Hsa-mir-183 has been shown to frequently involve methylation events in human hepatocellular carcinoma and was an unfavorable factor for survival [11]. Huang et al., found that hsa-mir-183 was an overexpressed miRNA in non-small-cell lung cancer and was positively correlated with tumor invasiveness [12]. Additionally, hsa-mir-183 has been reported to be markedly correlated with overall survival for colon adenocarcinoma patients [13]. However, the role of hsa-mir-183 and its predicted target genes in BLCA remains unclear.

To uncover novel prognostic and therapeutic targets for BLCA, our study is the first to investigate the role of hsa-mir-183 and its up-regulated predicted target genes and evaluate the significance of TP53-mutation related drug option in BLCA patients.

## MATERIALS AND METHODS

To estimate the prognostic and therapeutic significance of hsa-mir-183 and its up-regulated predicted target genes in BLCA patients, the clinical data, transcriptional expression data of miRNA and predicted target genes were collected from UALCAN (<http://ualcan.path.uab.edu>). The Metascape was applied for GO and KEGG analysis of predicted target genes (<https://metascape.org>). The Overall Survival (OS) data of mRNA expression of predicted target genes were collected from Kaplan-Meier plotter

(<http://www.kmplot.com/>). The protein expression data were collected from Human Protein Atlas (<https://www.proteinatlas.org>). The data of pan cancer analysis of predicted target genes were collected from TIMER2.0 (<http://timer.cistrome.org>). Gene mutations of predicted target genes in BLCA and their OS and Progression Free Survival (PFS) were analyzed by cBioPortal (<http://www.cbioportal.org>). Finally, the drug sensitivity and resistance of TP53 mutation in BLCA were analyzed by Genomics of Drug Sensitivity in Cancer (<https://www.cancerrxgene.org>).

## Ethics approval

The study protocols were conducted according to the principles of the Declaration of Helsinki and were approved by the Scientific and Medical Ethical Committee of the Second Affiliated Hospital of Shantou University Medical College.

## RESULTS

### Over-expression of different miRNA in BLCA patients

The result of top 10 over-expressed miRNA in BLCA patients were shown by TCGA miRNA analysis (Data from UALCAN) (Figure 1). The expression of hsa-mir-183 in patients with BLCA were significantly increased compared to normal tissue ( $P = 1.62E-12$ ). The expression of hsa-mir-183 were also significantly increased based on patient's gender or age (Figure 2). There were significant differences in the Comparison of Normal-vs-Male ( $P = 1.62E-12$ ), Normal-vs-Female ( $P = 1.63E-8$ ), Normal-vs-Age (41–60 Yrs) ( $P < 1E-12$ ), Normal-vs-Age (61–80 Yrs) ( $P < 1E-12$ ) and Normal-vs-Age (81–100 Yrs) ( $P = 7.55E-9$ ).

### Expression of predicted target genes of hsa-mir-183 in BLCA patients

There were 247 predicted genes of hsa-mir-183 (Data from TargetScan, microRNA.org and miRDB), including 15 up-regulated genes, 74 down-regulated genes and 158 unchanged genes in BLCA. 15 up-regulated genes were selected to study in our research (Table 1). The expression of all up-regulated predicted genes of hsa-mir-183 in patients with BLCA were significantly increased compared to normal tissue (Figure 3).

### Predicted functions and pathways of up-expressed predicted target genes of hsa-mir-183 and their 10 frequently altered neighbor genes

Functions and pathways of up-expressed predicted target genes of hsa-mir-183 and their 10 frequently

altered neighbor genes were analyzed by GO and KEGG in Metascape. As the Enrichment heatmap showed that pathways included hsa04726: Serotonergic synapse, hsa05219: Bladder cancer, GO:0007169: transmembrane receptor protein tyrosine kinase signaling

pathway, GO:0048471: perinuclear region of cytoplasm, GO:0008324: cation transmembrane transporter activity, GO:1990234: transferase complex and GO:1905477: positive regulation of protein localization to membrane (Figure 4).

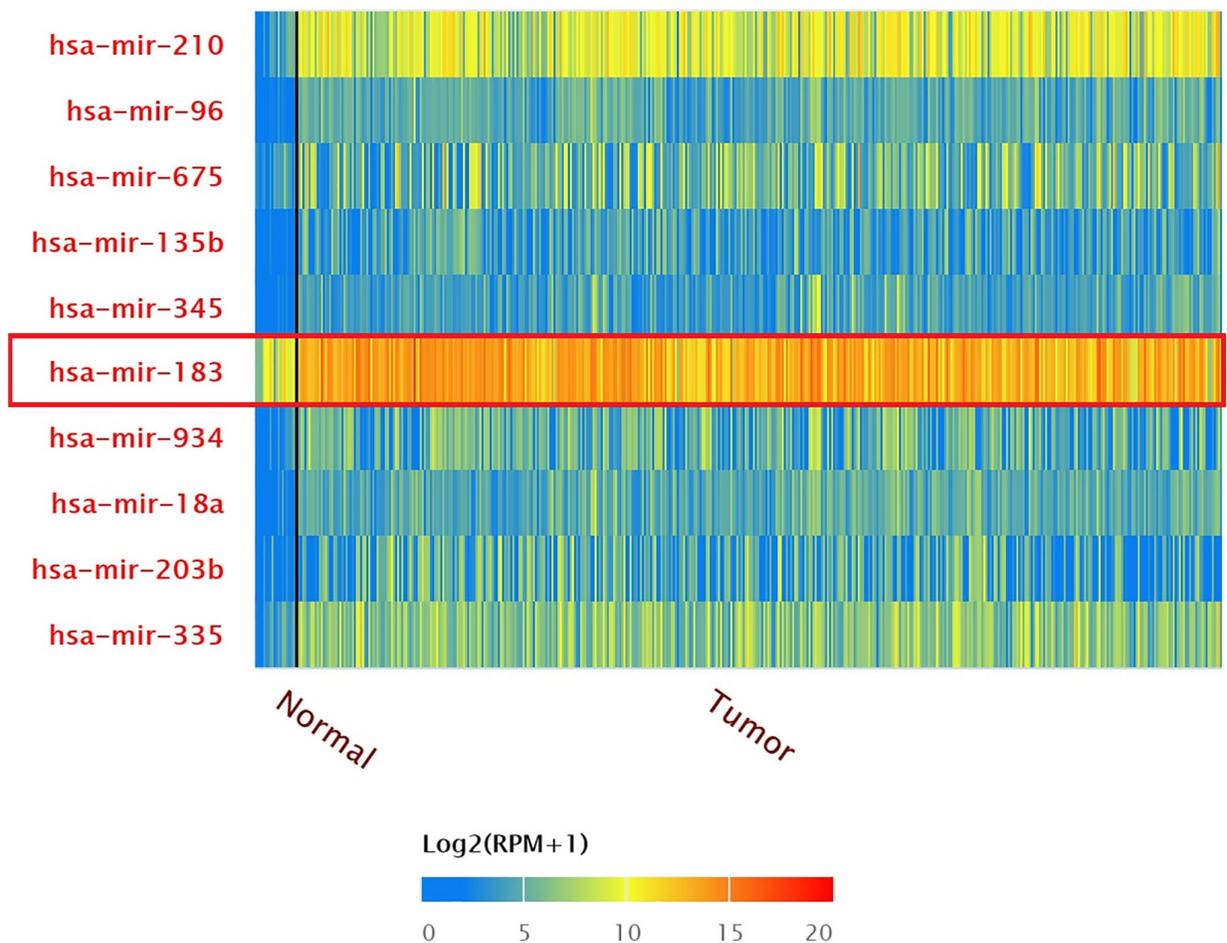


Figure 1. Top 10 over-expressed miRNAs in BLCA.

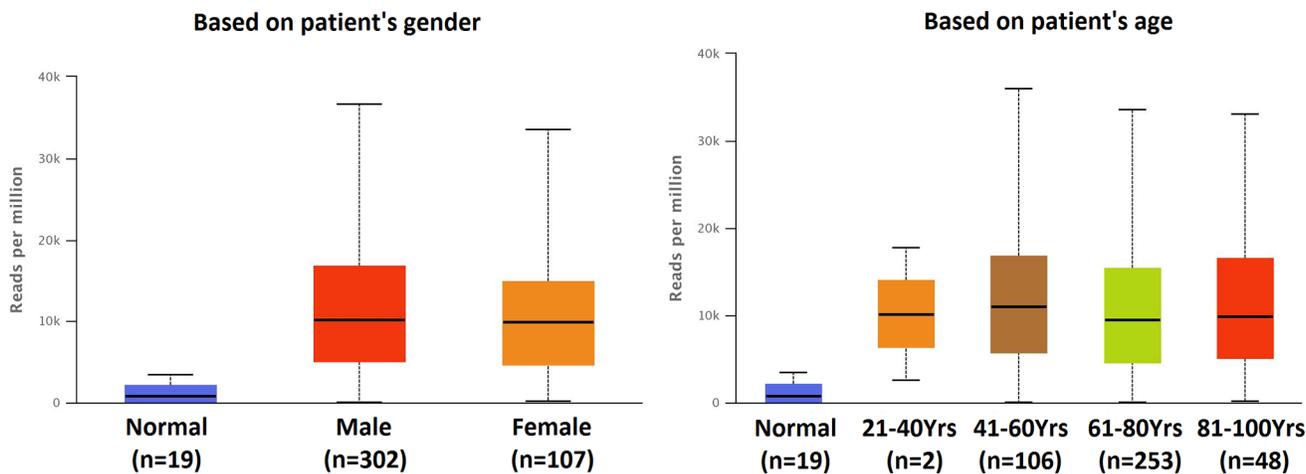


Figure 2. Expression of hsa-mir-183 in BLCA based on patient's gender and age.

**Table 1. Up-expressed predicted target genes of hsa-mir-183 in BLCA (TargetScan, microRNA.org and miRDB).**

Target gene	Description	Expression in BLCA
ENY2	ENY2, transcription and export complex 2 subunit	Up-expression
PLAG1	PLAG1 zinc finger	Up-expression
PDCD6	programmed cell death 6	Up-expression
GNG5	G protein subunit gamma 5	Up-expression
SLC6A6	solute carrier family 6 member 6	Up-expression
CELSR3	cadherin EGF LAG seven-pass G-type receptor 3	Up-expression
SLC22A23	solute carrier family 22 member 23	Up-expression
SPATS2	spermatogenesis associated serine rich 2	Up-expression
CDK5R1	cyclin dependent kinase 5 regulatory subunit 1	Up-expression
PHF6	PHD finger protein 6	Up-expression
MAL2	mal, T cell differentiation protein 2	Up-expression
ATP2C1	ATPase secretory pathway Ca <sup>2+</sup> transporting 1	Up-expression
NRAS	NRAS proto-oncogene, GTPase	Up-expression
SLC25A15	solute carrier family 25 member 15	Up-expression
PTDSS1	phosphatidylserine synthase 1	Up-expression

### Prognostic value of mRNA expression of up-expressed predicted target genes of hsa-mir-183 in BLCA patients

Prognostic value of mRNA expression of up-expressed predicted target genes of hsa-mir-183 were analyzed by Kaplan-Meier plotter (Figure 5). It was showed that mRNA expressions of PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1 were significantly associated with BLCA patients' prognosis. According to the database analysis form Kaplan-Meier plotter, PDCD6 (HR = 0.55, 95% CI: 0.4–0.75, and  $P = 0.00013$ ), GNG5 (HR = 0.73, 95% CI: 0.54–0.89, and  $P = 0.034$ ), PHF6 (HR = 0.68, 95% CI: 0.51–0.92, and  $P = 0.011$ ) and MAL2 (HR = 0.68, 95% CI: 0.5–0.91, and  $P = 0.01$ ) seem to be the favorable factor for BLCA patients' prognosis. SLC25A15 (HR = 1.68, 95% CI: 1.24–2.27, and  $P = 0.00062$ ) and PTDSS1 (HR = 1.41, 95% CI: 1.01–1.96, and  $P = 0.041$ ) more likely to be the unfavorable factor for BLCA patients' prognosis. In addition, ENY2, PLAG1, SLC6A6, CELSR3, SLC22A23, SPATS2, CDK5R1, ATP2C1 and NRAS were not prognostic in BLCA patients.

### Protein expression and pan cancer analysis of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) in BLCA patients

Protein expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) in BLCA were explored by Human Protein Atlas

(Figure 6). The proteins of PDCD6, GNG5 and PHF6 were not or low expressed in BLCA tissues, whereas medium expressions of them were found in normal urinary bladder tissues. It is noteworthy that the protein expression of MAL2 and SLC25A15 could not be detected in both normal urinary bladder tissues and BLCA tissues. Additionally, the protein expression of PTDSS1 in tumor tissues was stronger than that in normal tissues. Furthermore, we performed a pan-cancer analysis of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) by using TIMER (Figure 7).

### Relationship between mRNA expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) and clinicopathological parameters in BLCA patients

The association of mRNA expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) with clinicopathological parameters of BLCA patients were analyzed by UALCAN, containing BLCA individual cancer stages and molecular subtypes (Figures 8 and 9). Figure 8 showed that the mRNA expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) were significantly correlated with BLCA individual cancer stages. PDCD6, GNG5, PHF6 and MAL2 as favorable factor for BLCA patients, the high mRNA expressions of them tended to be in stage 1 or 2, whereas the high mRNA expressions of unfavorable factor of SLC25A15 and PTDSS1 tended to be in stage 3

or 4. Similarly, Figure 9 showed that the mRNA expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) were significantly correlated with BLCA molecular subtypes. The high mRNA expressions of PDCD6, GNG5, PHF6 and MAL2 tended to be the molecular subtypes of luminal and luminal papillary, whereas the high mRNA expressions of SLC25A15 and PTDSS1 tended to be the molecular subtypes of neuronal and basal squamous.

### Gene mutations of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) in BLCA

Gene mutations of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) in BLCA and their OS and PFS were analyzed by cBioPortal (Figure 10). Figure 10A showed that PDCD6, MAL2, SLC25A15, PTDSS1 were prone to mutation in

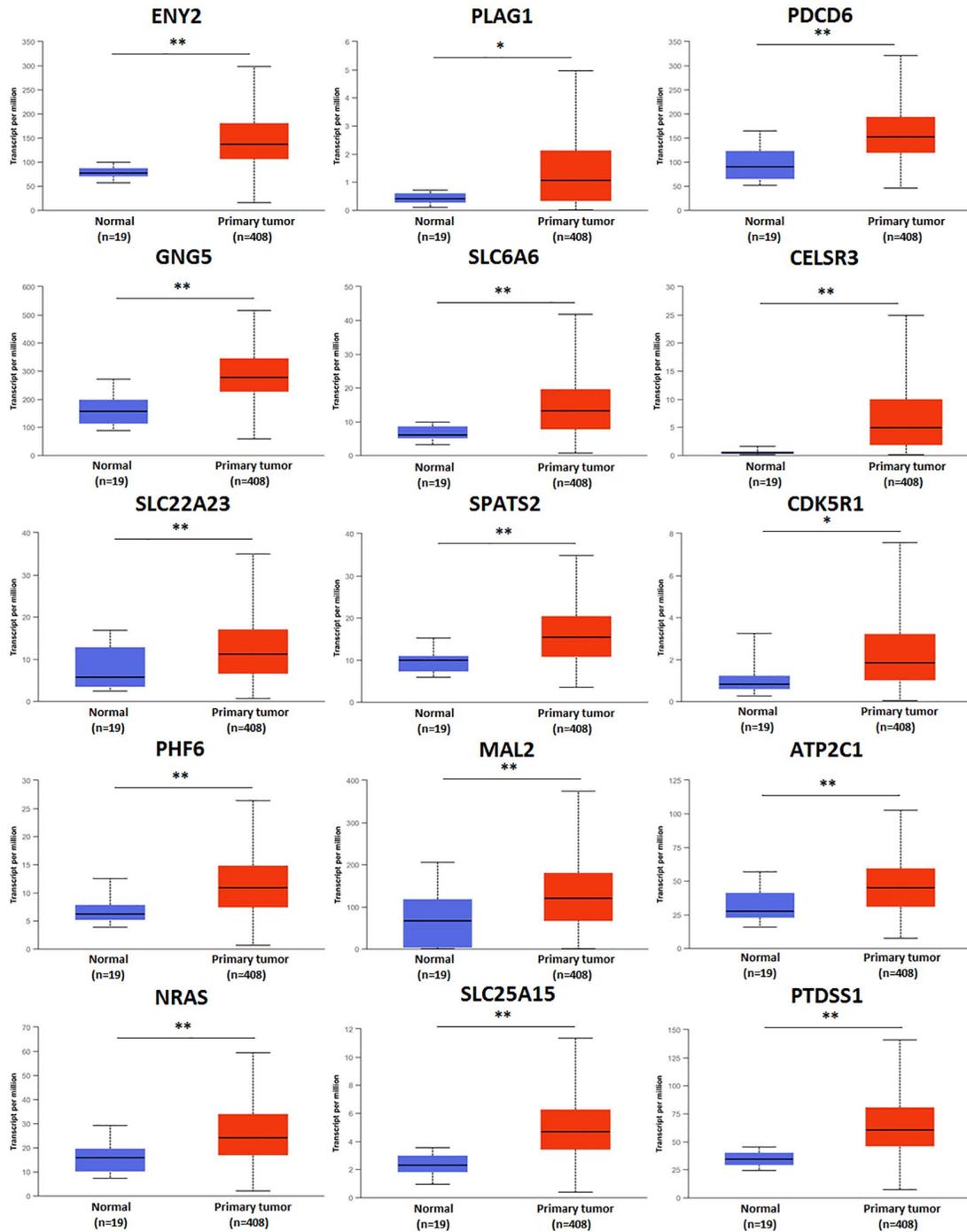


Figure 3. Transcriptional expression of predicted target genes of hsa-mir-183 in BLCA based on TCGA sample types. (\*\* $P < 0.001$ , \* $P < 0.01$ ).

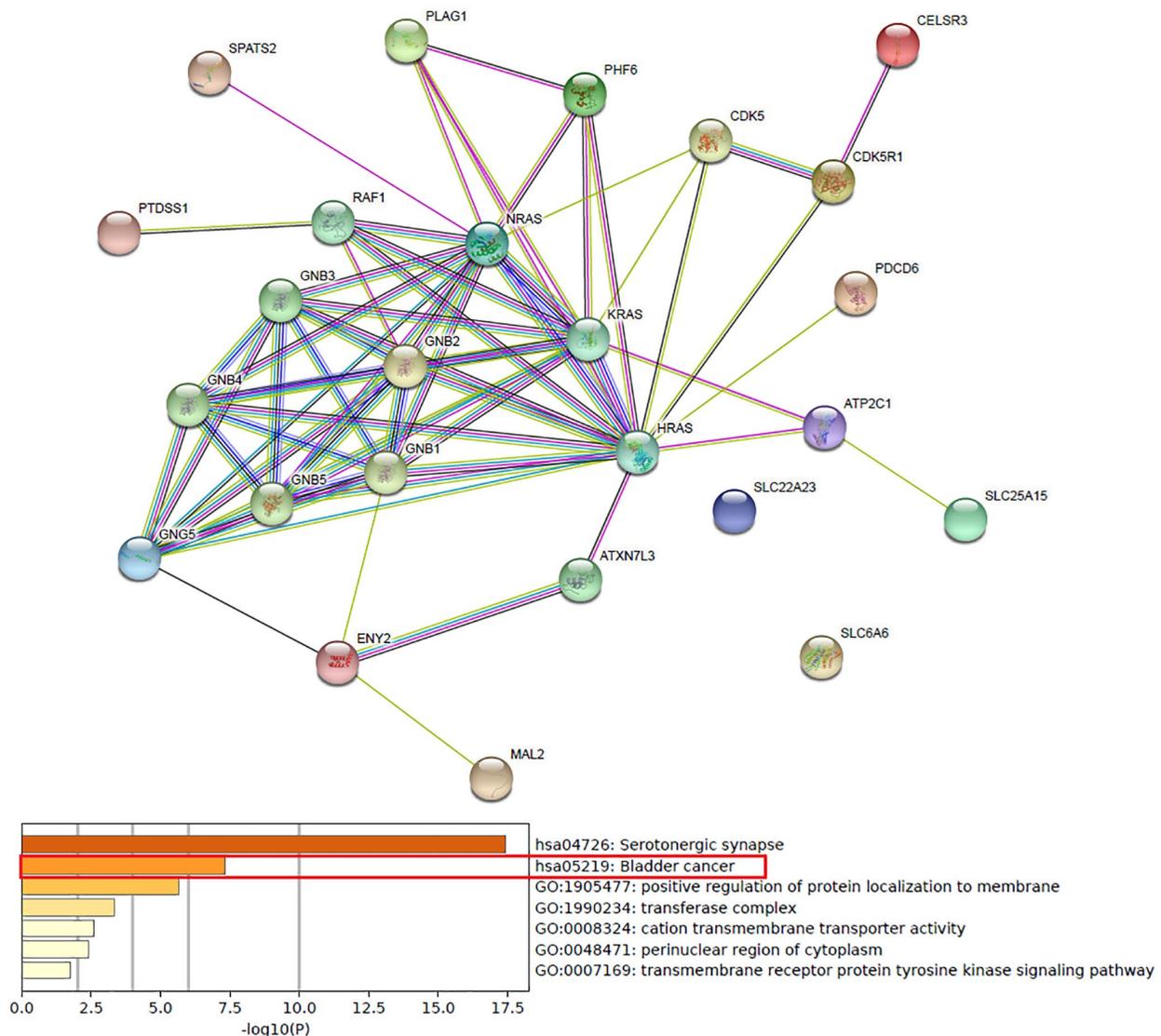
BLCA patients. Figure 10B and 10C revealed the association between the gene mutation and BLCA patients' OS and DFS. In OS analyze, median months overall (95% CI) of unaltered group was 36.43, while it was significantly shorter in PHF6 (1.87) and SLC25A15 (17.62). Similarly, in PFS analyze, Median Months Progression Free (95% CI) of unaltered group was 32.58, while it was significantly shorter in PHF6 (1.87), MAL2 (5.52) and SLC25A15 (17.52).

**TP53 mutation of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) and its association with drug selection in BLCA**

TP53 mutation of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) were analyzed by UALCAN. The expression of TP53

mutation of GNG5, MAL2, SLC25A15, PTDSS1 in patients with BLCA were significantly increased compared to non-mutation group, and all predicted target genes were markedly increased compared to normal group (Figure 11). In addition, types of TP53 mutation were analyzed by cBioPortal (Figure 12) from Bladder Cancer (MSK/TCGA, 2020). Mutation frequency of TP53 was 51.7% from NM\_000546 | ENST00000269305 CCDS11118 | P53\_HUMAN.

As was showed in Figure 12, TP53 driver had the highest probability of mutation, including Missense (197), Truncating (80) and Splice (17). Furthermore, the drug sensitivity and resistance of TP53 mutation in BLCA were analyzed by Genomics of Drug Sensitivity in Cancer (Figure 13). Nutlin-3a (-) drug target was MDM2 and Nutlin-3a (-) drug target pathway was p53



**Figure 4. Predicted functions and pathways of up-expressed predicted target genes of hsa-mir-183 and their 10 frequently altered neighbor genes (Metascape).**

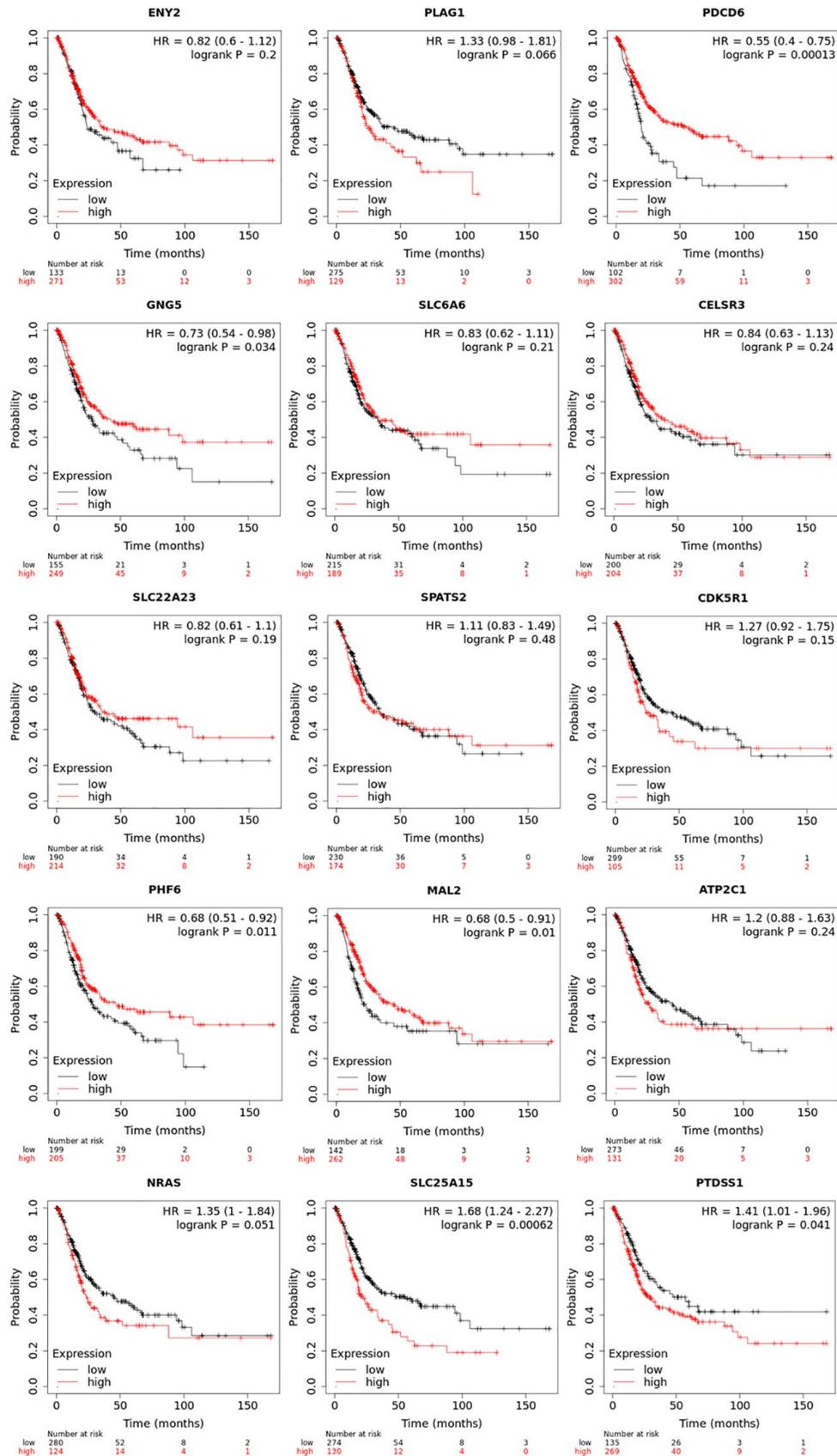


Figure 5. Prognostic value of transcriptional expression of up-expressed predicted target genes of hsa-mir-183 in BLCA (Kaplan-Meier Plotter).

pathway in BLCA. As was showed in Figure 13 and Table 2, Nutlin-3a (-) had a significant increased resistance in TP53 mutation ( $P = 2.12e-06$ ) of BLCA patients and the IC50 of Nutlin-3a (-) was significantly increased in TP53 mutation group ( $P < 0.001$ ). TWS119 drug target was GSK3 and TWS119 drug target pathway was WNT signaling in BLCA. As was showed in Figure 13 and Table 2, TWS119 had a significant increased sensitivity in TP53 mutation ( $P = 5.95e-06$ ) of BLCA patients and the IC50 of TWS119 was significant decreased in TP53 mutation group ( $P < 0.001$ ). Mitomycin-C drug target was DNA crosslinker and Mitomycin-C drug target pathway was DNA replication in BLCA. As was showed in Figure 13 and Table 2, Mitomycin-C had a significant increased sensitivity in TP53 mutation ( $P = 0.000322$ ) of BLCA

patients and the IC50 of Mitomycin-C was significant decreased in TP53 mutation group ( $P = 0.001$ ).

## DISCUSSION

BLCA was the most common histopathological type of bladder cancer, and its occurrence and development were affected by multiple complex factors such as gene transcription, gene regulation, gene mutation and epigenetic inheritance [14, 15]. Each miRNA had a series of related target genes, and abnormal expression of miRNA often led to abnormal transcription of target genes, which was related to the occurrence of various cancers [16]. Although hsa-mir-183 have been reported to play an essential role in liver cancer, lung cancer and other tumors, the specific role of hsa-mir-

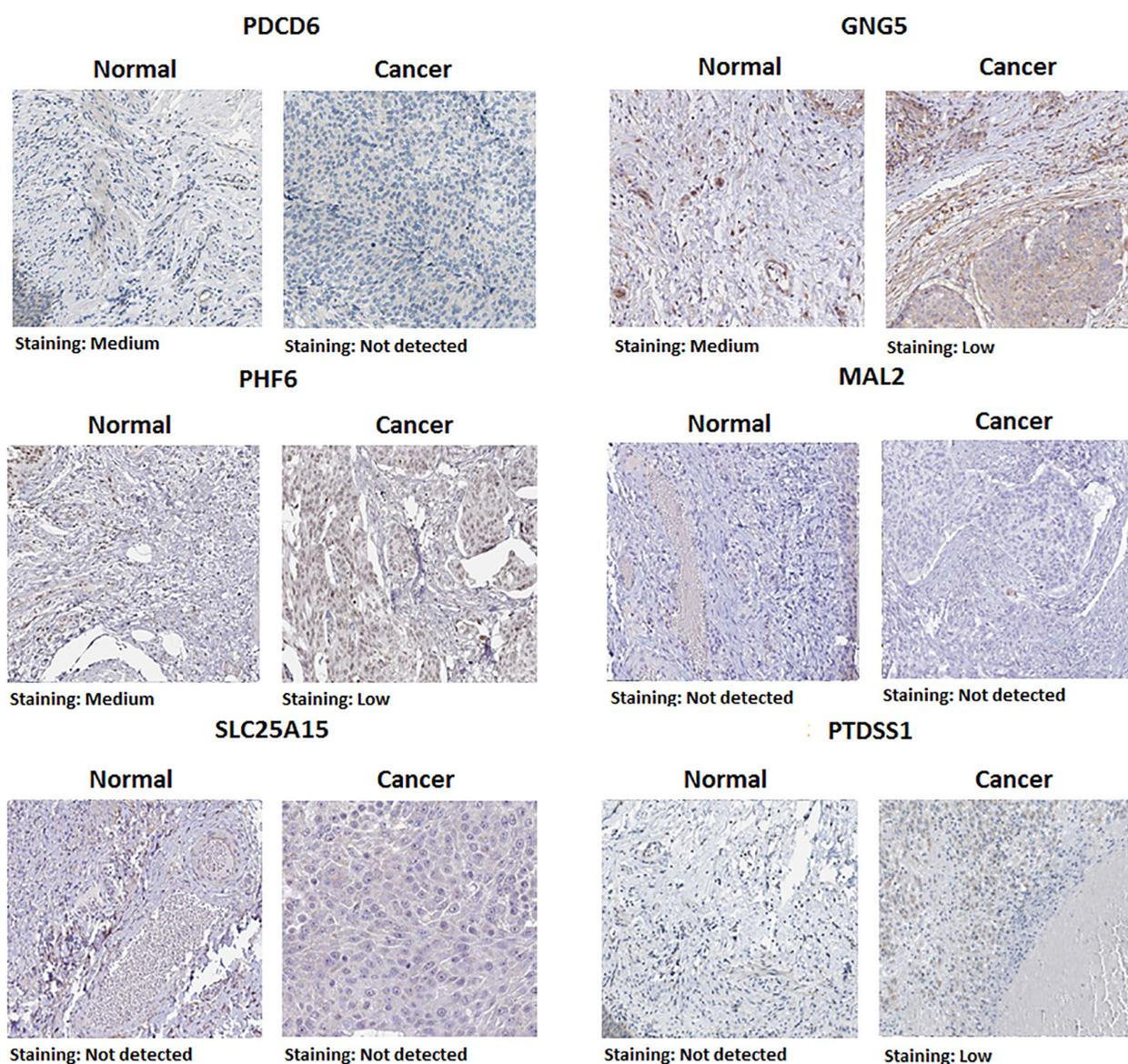


Figure 6. Representative immunohistochemistry images of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) in normal urinary bladder tissues and BLCA tissues (Human Protein Atlas).

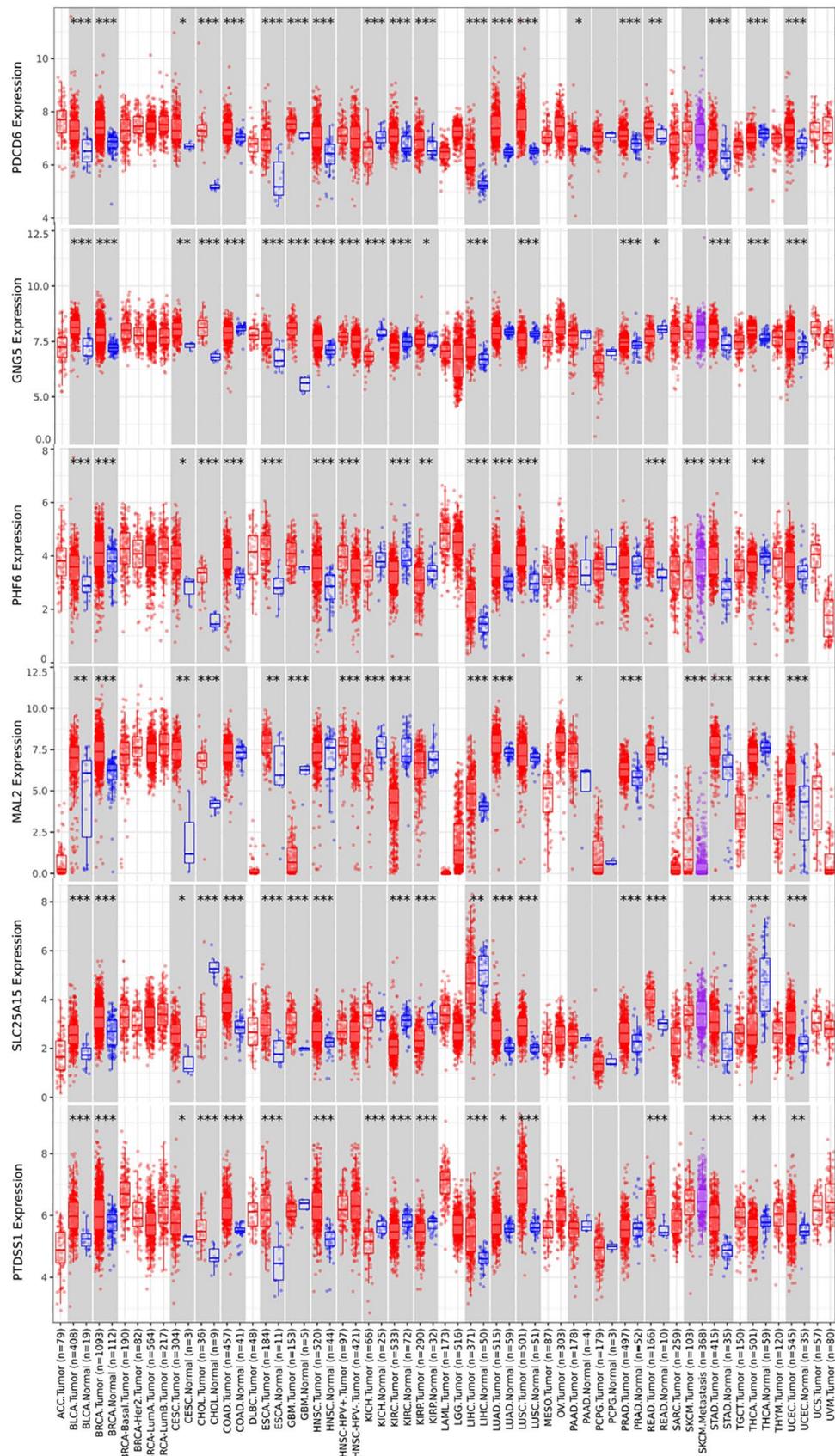
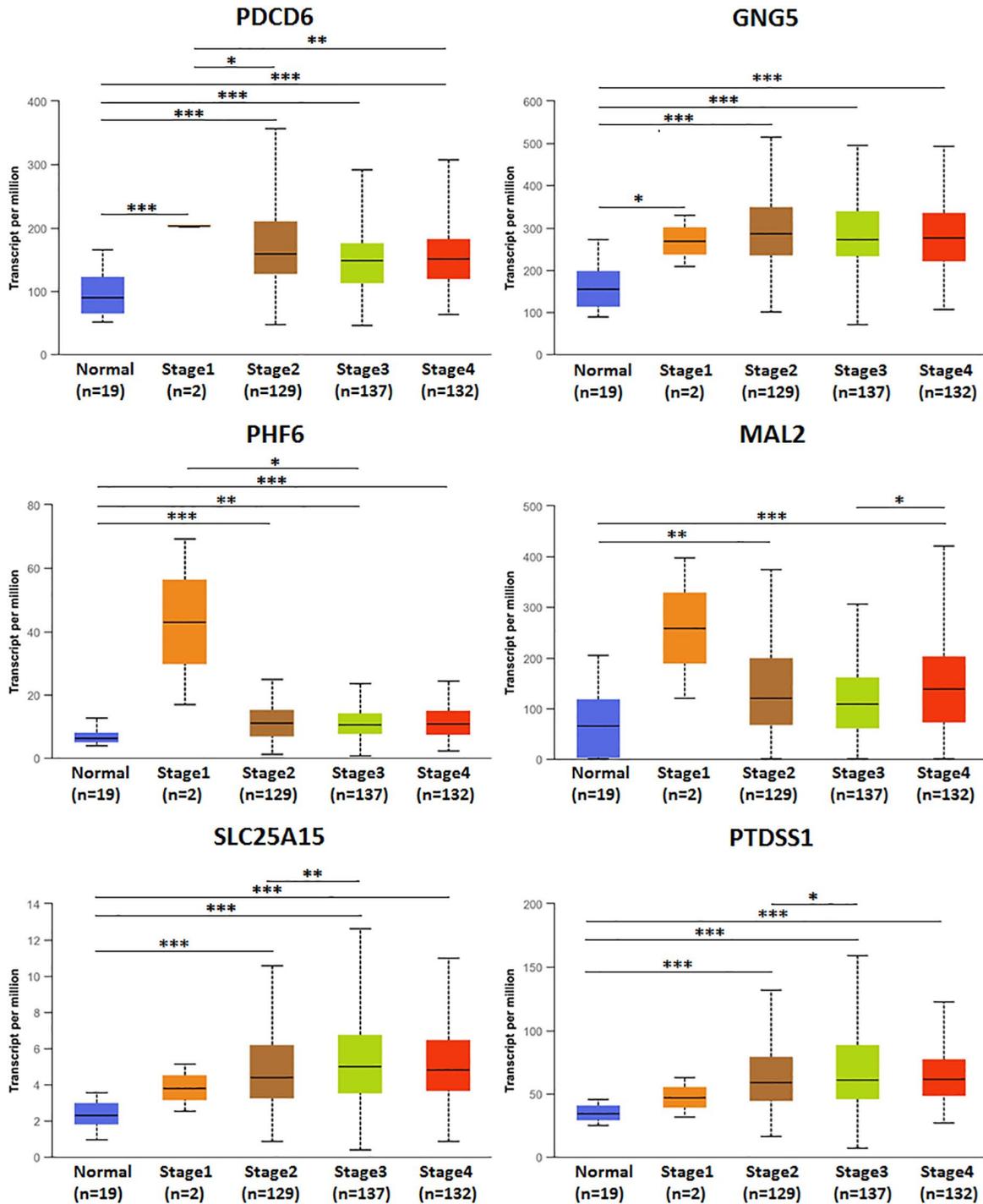


Figure 7. Pan cancer analysis of predicted target genes (PCD6, GNG5, PHF6, MAL2, SLC25A15 and PTSS1) (TIMER). (\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ ).

**Table 2. Association between TP53 mutation and drug option of BLCA.**

Drug	Drug Target	IC50 effect	Effect size	P-value	FDR%	Tissue
TWS119	GSK3	Increased sensitivity	-3.63	5.95e-06	0.00714	BLCA
Mitomycin-C	DNA crosslinker	Increased sensitivity	-2.53	0.000322	0	BLCA
Nutlin-3a (-)	MDM2	Increased resistance	4.08	2.12e-06	0.00255	BLCA



**Figure 8. Relationship between mRNA expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTSS1) and BLCA individual cancer stages. (\*\**P* < 0.001, \*\**P* < 0.01, \**P* < 0.05).**

183 and its predicted target genes in BLCA remain to be explored. In our study, the transcriptional expressions, mutations and Prognosis of hsa-mir-183 and its predicted target genes in BLCA have been elucidated.

Our study found that up-expressed hsa-mir-183 and over transcriptional expressions of hsa-mir-183

predicted target genes (ENY2, PLAG1, PDCD6, GNG5, SLC6A6, CELSR3, SLC22A23, SPATS2, CDK5R1, PHF6, MAL2, ATP2C1, NRAS, SLC25A15 and PTDSS1) were observed in BLCA patients. Moreover, predicted functions and pathways of up-expressed predicted target genes of hsa-mir-183 and their 10 frequently altered neighbor genes indicated a link

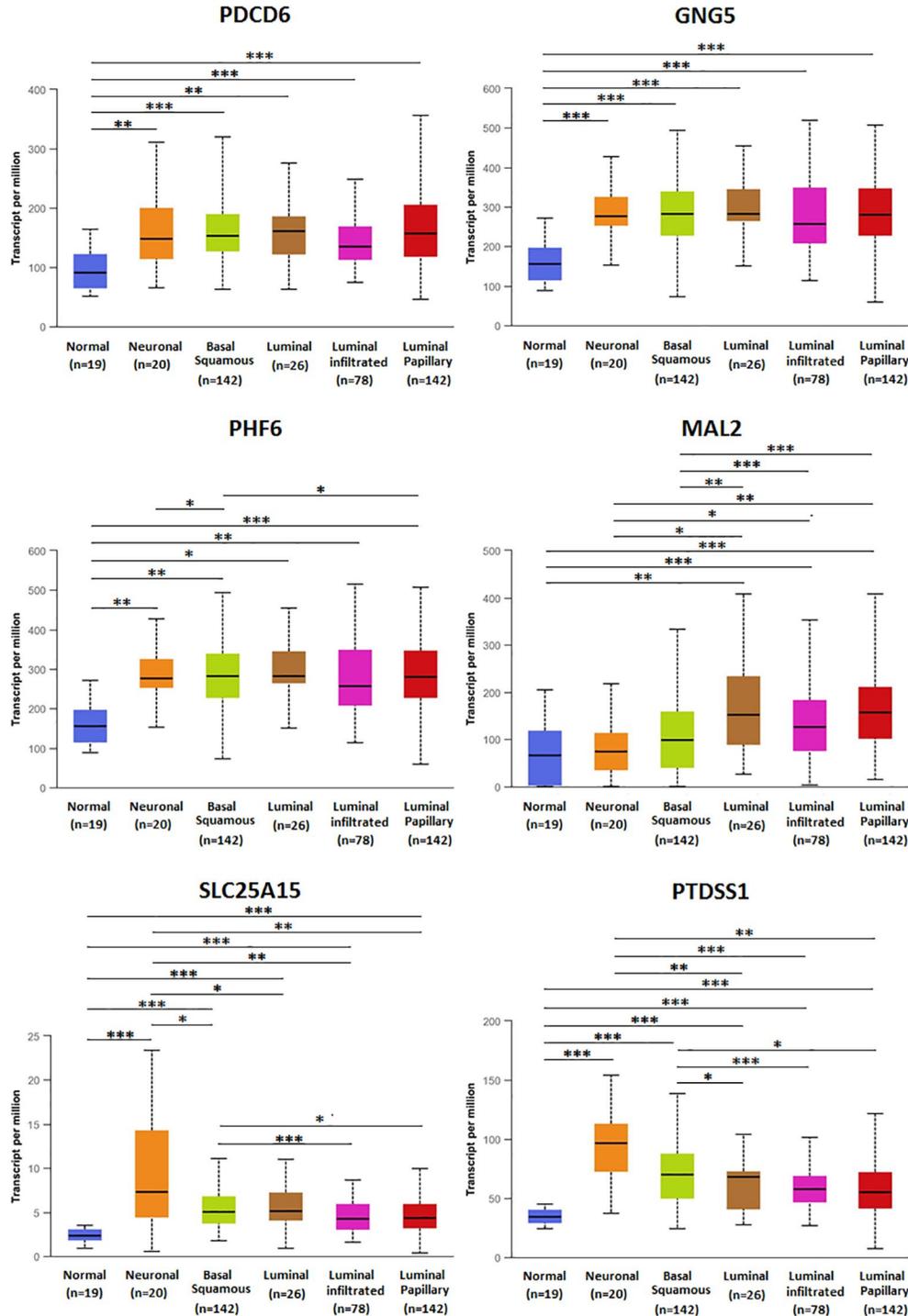


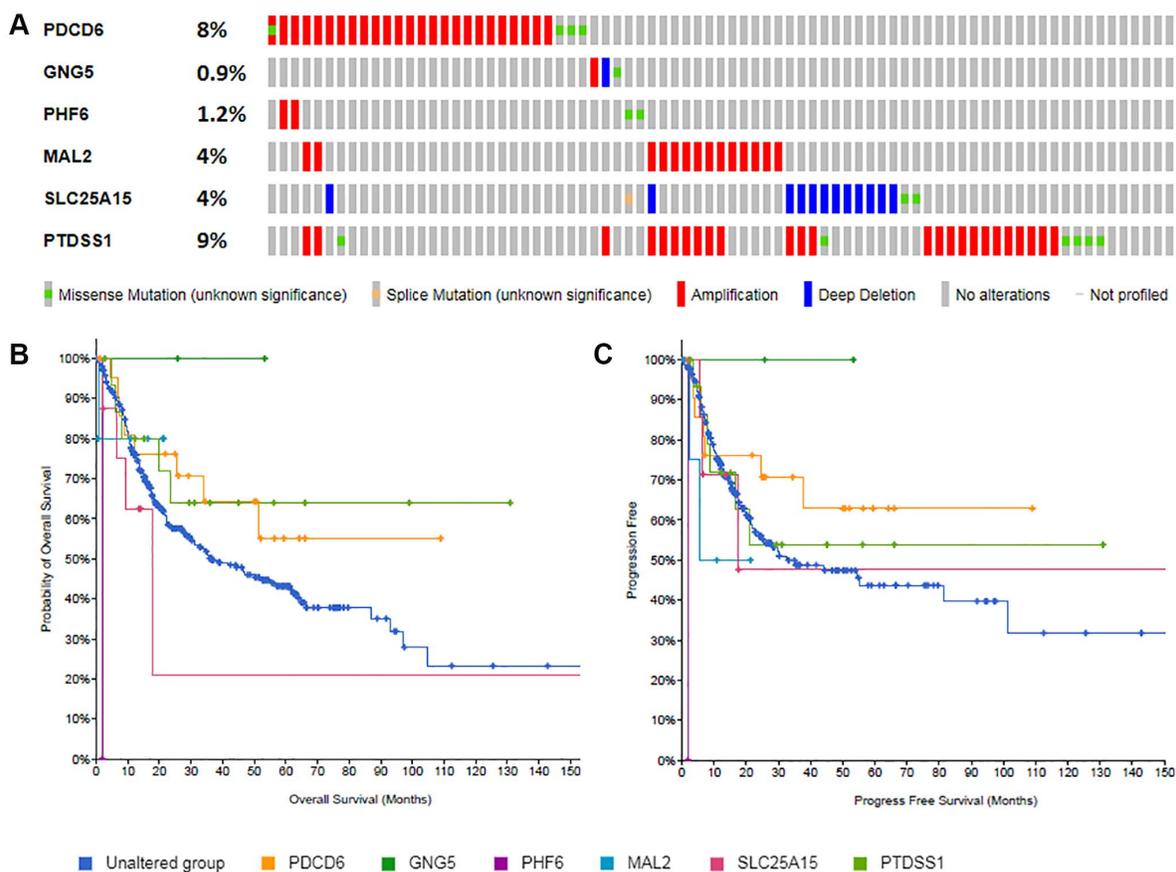
Figure 9. Relationship between mRNA expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) and BLCA molecular subtypes. (\*\*\*)  $P < 0.001$ , (\*\*)  $P < 0.01$ , (\*)  $P < 0.05$ .

with bladder cancer. Besides, high transcriptional expressions of PDCD6, GNG5, PHF6 and MAL2 were markedly relevant to favorable OS in BLCA patients, whereas high transcriptional expression of SLC25A15 and PTDSS1 were markedly relevant to unfavorable OS in BLCA patients. Additionally, ENY2, PLAG1, SLC6A6, CELSR3, SLC22A23, SPATS2, CDK5R1, ATP2C1 and NRAS were not prognostic in BLCA patients. However, protein expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) were inconsistent with the transcriptional expression. The mRNA levels were significantly increased than that observed in protein levels. There are two possible explanations for this phenomenon. miRNA do not limit gene translation but lead to rapid destruction of newborn peptides [17]. Alternatively, miRNA may negatively affect the initiation, elongation and termination of transcription, while the number of ribosomes remains unchanged, resulting in a decreased rate of protein completion [17]. Multivariate analysis indicated that transcriptional expression of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) were significantly correlated with BLCA individual cancer stages and molecular subtypes. Furthermore, high mutation rates of PDCD6, MAL2,

SLC25A15 and PTDSS1 were observed in BLCA patients and some genetic alteration were significantly relevant to OS and PFS. Finally, TP53 mutation of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1) has guiding significance for drug selection in BLCA.

### PDCD6

Programmed cell death 6, can encode the calcium binding protein, which is belonging to the penta-EF-hand protein family. High expression of PDCD6 was relevant to unfavorable prognosis in colorectal cancer patients via activating c-Raf, MEK and ERK pathway [18]. Inhibition of PDCD6 expression can reduce the rate of tumor metastasis [19]. PDCD6 can induce migration and invasion of tumor cell in ovarian cancer causing unfavorable clinical outcome to patients [20]. However, the expression of GG genotype of PDCD6 has been reported to reduce the risk of lung cancer [21]. Similarly, up-expressed PDCD6 played the role of inhibiting cell growth in cervical carcinoma cells [22]. These results suggested that PDCD6 may play different roles in tumorigenesis and prognosis. In our study, the transcriptional expression of PDCD6 was significantly



**Figure 10.** Gene mutations of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) and their association with OS and PFS in BLCA patients (A: Genetic alterations; B: Overall survival; C: Progress free survival).

increased in BLCA and suggested a favorable prognosis. The function of MAPK pathway is to induce cell hyperplasia and invasion, but it can also increase the occurrence of cell apoptosis or inhibit cell invasion. Different effects of PDCD6 predominantly depend on the signal strength and the microenvironment of abnormal

signal activation [23]. In addition, the downstream MAPK signal is mainly regulated by the upstream c-Raf signal. It has been proved that PDCD6-c-Raf-MAPK pathway exists in tumors in recent research [18]. Therefore, we speculated that tumor microenvironment or signal specificity lead to different roles of PDCD6 in tumors.

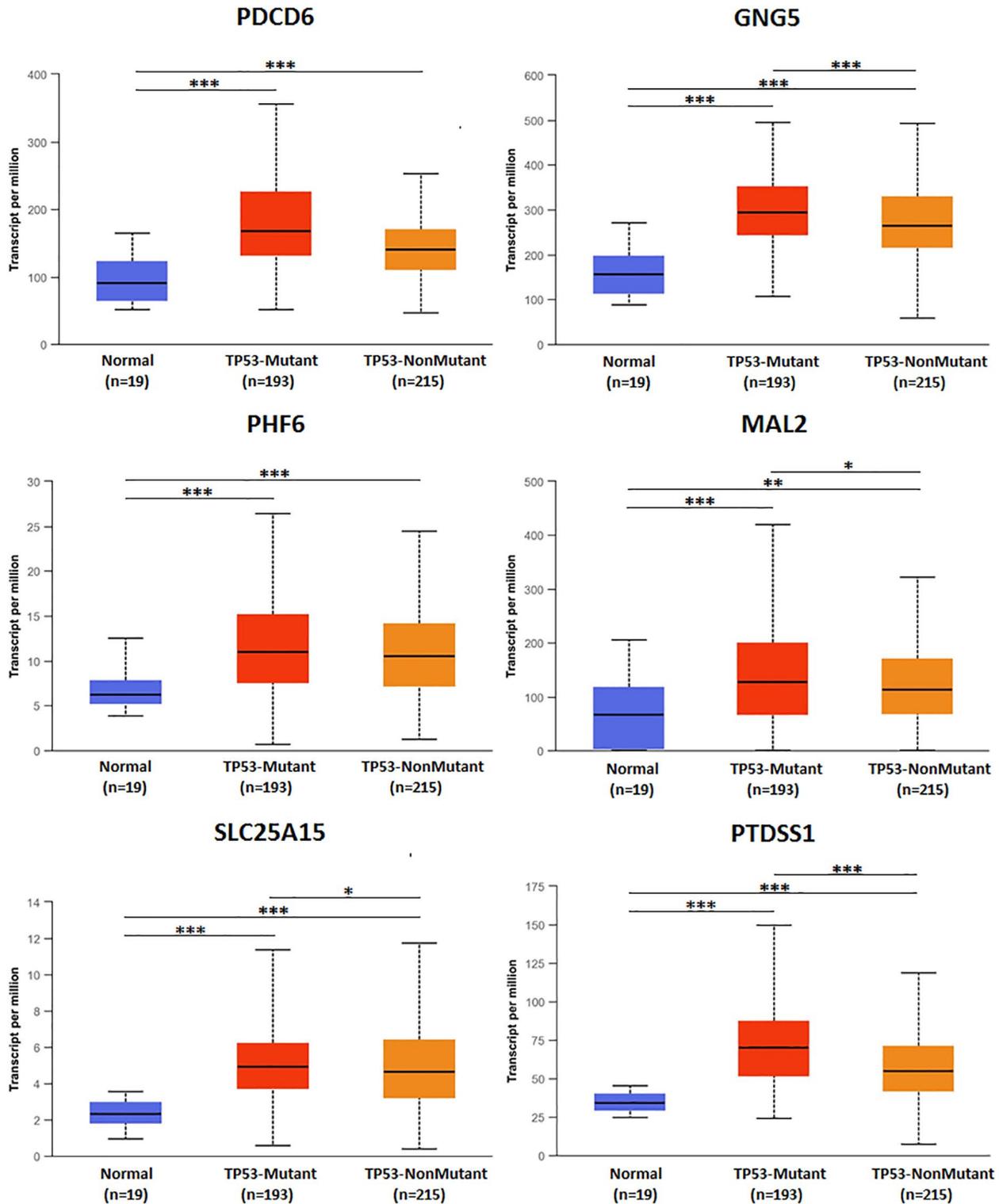
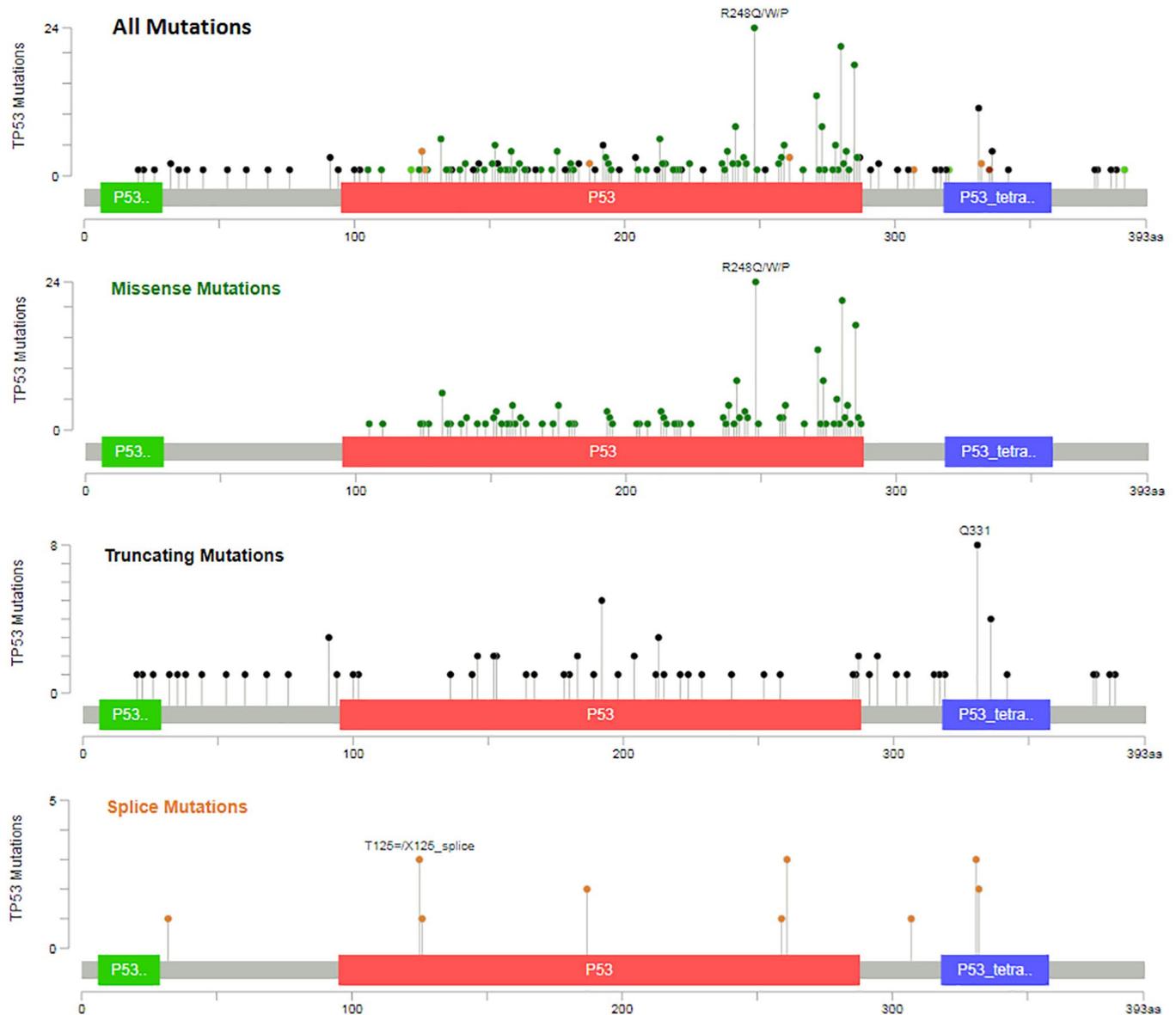


Figure 11. TP53 mutation of predicted target genes (PDCD6, GNG5, PHF6, MAL2, SLC25A15, PTDSS1) in BLCA.

## GNG5

G protein subunit gamma 5, can regulate the information from a variety of receptors. High expression of GNG5 was relevant to unfavorable prognosis in gliomas [24]. At present, there are few studies on GNG5 in other tumors. It was reported that GNG5 may regulate the occurrence and development of tumors by stimulating the PI3K/Akt signal pathway [25]. Akt is activated by upstream PI3K, leading to

phosphorylation, which finally promotes tumor cell growth. To our knowledge, this is the first study to explore the role of GNG5 in BLCA. Transcriptional expression of GNG5 was found to be significantly associated with BLCA individual cancer stages, BLCA molecular subtypes and OS. Of note, higher expression of GNG5 indicated a longer possibility of survival. Therefore, GNG5 may act as an advanced indicator of the stages, types and prognosis for BLCA which has not been reported in previous researches.



Mutation types and corresponding color codes are as follows

- **Missense Mutations**(putative driver)
- **Missense Mutations**(unknown significance)
- **Truncating Mutations**(putative driver)
- **Other Mutations**: All other types of mutations
- **Truncating Mutations**(unknown significance)
- **Inframe Mutations**(putative driver)
- **Inframe Mutations**(unknown significance)
- **Splice Mutations**(putative driver)
- **Splice Mutations**(unknown significance)
- **Fusion Mutations**

Figure 12. Types of TP53 mutation from bladder cancer (MSK/TCGA, 2020).

## PHF6

PHD finger protein 6, has a function of transcriptional regulation by encoding domain with 2 PHD-type zinc fingers in protein. The high expression of PHF6 can play a role of tumor suppressor in patients with acute lymphoblastic leukemia and as a favorable prognostic factor [26–28]. However, other study has come up with an inconsistent view that PHF6 was involved in proliferation, migration, apoptosis and metabolism in hepatocellular carcinoma and its deficiency can inhibit above mentioned processes in tumor cells [29]. Wang et al., found that PHF6-null can inhibit cell proliferation and arrest cells in the G(2)/M phase, suggesting that loss of PHF6 may result in accumulation of DNA damage [30]. Hsu et al., reported that the deficiency of PHF6 can enhance cell self-renewal and carcinogenic potential [31]. Together, these findings indicated that whether PHF6 was a tumor suppressor or a tumor protein may depend on the specific context in which it acted. In our study, high transcriptional expression of PHF6 was significantly related to high probability of survival in BLCA which were consistent with the research results in acute lymphoblastic leukemia.

## MAL2

Mal, T cell differentiation protein 2, can encode multitransmembrane protein, which is a member of MAL proteolipid family. MAL2 was over-expressed in many human cancers, including breast, stomach, liver, pancreatic, colon, kidney, prostate and ovarian cancers [32–36]. Zheng et al., found that MAL2 was markedly relevant to grade, stage and the Gleason score of prostate cancer and it can promote tumor progression

via the Notch pathway [33]. Fang et al., showed that Inhibition of MAL2 expression can significantly increase the cytotoxicity of CD8<sup>+</sup> T lymphocytes, thereby inhibiting the growth and metastasis of breast cancer [35]. Zhang et al., reported that overexpressed MAL2 can phosphorylate ERK1/2, which may promote pancreatic cancer progression [37]. In BLCA patients, MAL2 was also highly expressed, but it was associated with a good prognosis. At present, the specific signaling pathway that MAL2 participates in bladder cancer is not clear, so we speculate that up-expressed MAL2 may not be the initiating factor of tumorigenesis, but may lead to changes in cell morphology, signal transduction, migration and metabolism by altering protein distribution, thus promoting tumor transformation [38], which is bidirectional.

## SLC25A15

Solute carrier family 25 member 15, can encode the protein transporter ornithine, which is a member of mitochondrial carrier family. Ji et al., found that SLC25A15 was up-expressed in melanoma patients and was negatively associated with OS and DFS [39]. SLC25A15 was involved in the regulation of cell proliferation and apoptosis in cells and over-expression can reverse the role of tumor suppressor genes in prostate cancer [40]. At present, most of the studies on SLC25A15 are focused on hyperornithinemia-hyperammonemia-homocitrullinuria [41], and there are few reports on tumor. SLC25A15 is primarily responsible for transporting ornithine from the cytoplasm to the mitochondrial matrix, which is an important step in regulating the urea cycle [42]. Whether SLC25A15 has other potential signaling

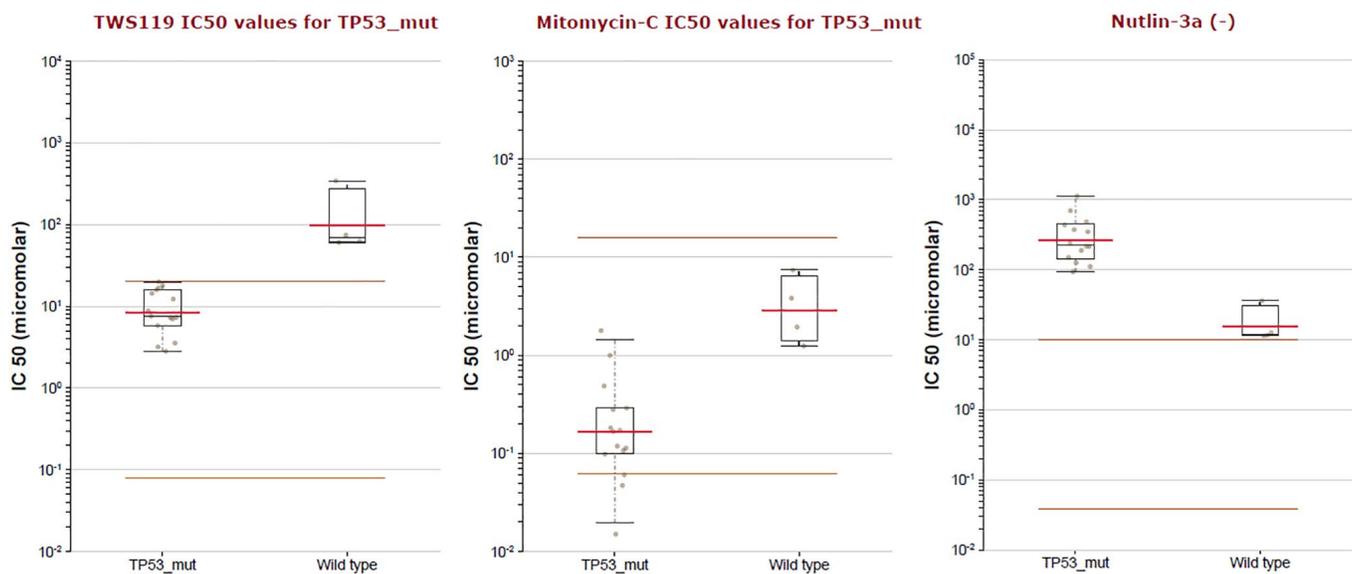


Figure 13. The IC<sub>50</sub> of TWS119, Mitomycin-C and Nutlin-3a (-) in TP53 mutation of BLCA patients.

pathways involved in tumor progression. However, by comparing recent researches with our present study, it can be preliminarily speculated that SLC25A15 is an unfavorable prognostic signal for BLCA, which may act as a new biomarker to assist the diagnosis and prognosis of BLCA.

### PTDSS1

Phosphatidylserine synthase 1, has a function to promote the synthesis of phosphatidylserine. Li et al., found that over-expression of PTDSS1 in astrogloma involved in lipid metabolism [43]. Wang et al., shown that PTDSS1 was an oncogene of lung adenocarcinoma, and overexpression was significantly associated with low survival [44]. PTDSS1 can participate in the biosynthesis of phosphatidylserine by directing the synthesis of a key enzyme, phosphatidylserine synthase, which was an important part of phosphatidylserine metabolism [44]. PTDSS1-mediated phosphatidylserine signaling has been shown to be severely disrupted in metabolism of tumor cell and pathogenesis of autoimmune disorders [45]. Tumor progression and metastasis are greatly influenced by the interaction between tumor cells and their living microenvironment. TAM receptors exist on the surface of tumor cells, and an environment rich in phosphatidylserine can provide a platform for activation of TAM receptors [46]. In our study, high transcriptional expression of PTDSS1 was significantly related to low probability of survival in BLCA which indicated that PTDSS1-mediated phosphatidylserine signaling was one of the pathogenesis of BLCA.

### CONCLUSION

To sum up, we concluded that PDCD6, GNG5, PHF6, MAL2, SLC25A15 and PTDSS1 could be the advanced independent indicators for prognosis of BLCA patients, and TP53-mutation might be a biomarker for drug option in BLCA patients.

### AUTHOR CONTRIBUTIONS

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

### REFERENCES

1. Wullweber A, Strick R, Lange F, Sikic D, Taubert H, Wach S, Wullich B, Bertz S, Weyerer V, Stoehr R, Breyer J, Burger M, Hartmann A, et al. Bladder Tumor Subtype Commitment Occurs in Carcinoma *In Situ* Driven by Key Signaling Pathways Including ECM Remodeling. *Cancer Res.* 2021; 81:1552–66. <https://doi.org/10.1158/0008-5472.CAN-20-2336> PMID:[33472889](https://pubmed.ncbi.nlm.nih.gov/33472889/)
2. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? *Cancer Commun (Lond).* 2019; 39:22. <https://doi.org/10.1186/s40880-019-0368-6> PMID:[31030667](https://pubmed.ncbi.nlm.nih.gov/31030667/)
3. Akhtar M, Al-Bozom IA, Ben Gashir M, Taha NM, Rashid S, Al-Nabet ADM. Urothelial Carcinoma In Situ (CIS): New Insights. *Adv Anat Pathol.* 2019; 26:313–9. <https://doi.org/10.1097/PAP.000000000000239> PMID:[31149909](https://pubmed.ncbi.nlm.nih.gov/31149909/)
4. Williamson SR, Montironi R, Lopez-Beltran A, MacLennan GT, Davidson DD, Cheng L. Diagnosis, evaluation and treatment of carcinoma in situ of the urinary bladder: the state of the art. *Crit Rev Oncol Hematol.* 2010; 76:112–26. <https://doi.org/10.1016/j.critrevonc.2010.01.005> PMID:[20097572](https://pubmed.ncbi.nlm.nih.gov/20097572/)
5. Casey RG, Catto JW, Cheng L, Cookson MS, Herr H, Shariat S, Witjes JA, Black PC. Diagnosis and management of urothelial carcinoma in situ of the lower urinary tract: a systematic review. *Eur Urol.* 2015; 67:876–88. <https://doi.org/10.1016/j.eururo.2014.10.040> PMID:[25466937](https://pubmed.ncbi.nlm.nih.gov/25466937/)
6. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A.* 2004; 101:2999–3004. <https://doi.org/10.1073/pnas.0307323101> PMID:[14973191](https://pubmed.ncbi.nlm.nih.gov/14973191/)
7. Melo SA, Esteller M. Dysregulation of microRNAs in cancer: playing with fire. *FEBS Lett.* 2011; 585:2087–99. <https://doi.org/10.1016/j.febslet.2010.08.009> PMID:[20708002](https://pubmed.ncbi.nlm.nih.gov/20708002/)
8. Chan JJ, Tay Y. Noncoding RNA:RNA Regulatory Networks in Cancer. *Int J Mol Sci.* 2018; 19:E1310. <https://doi.org/10.3390/ijms19051310> PMID:[29702599](https://pubmed.ncbi.nlm.nih.gov/29702599/)

9. Fabbri M, Calin GA. Epigenetics and miRNAs in human cancer. *Adv Genet.* 2010; 70:87–99.  
<https://doi.org/10.1016/B978-0-12-380866-0.60004-6>  
PMID:[20920746](https://pubmed.ncbi.nlm.nih.gov/20920746/)
10. Kunej T, Godnic I, Ferdin J, Horvat S, Dovc P, Calin GA. Epigenetic regulation of microRNAs in cancer: an integrated review of literature. *Mutat Res.* 2011; 717:77–84.  
<https://doi.org/10.1016/j.mrfmmm.2011.03.008>  
PMID:[21420983](https://pubmed.ncbi.nlm.nih.gov/21420983/)
11. Anwar SL, Krech T, Hasemeier B, Schipper E, Schweitzer N, Vogel A, Kreipe H, Buurman R, Skawran B, Lehmann U. *hsa-mir-183* is frequently methylated and related to poor survival in human hepatocellular carcinoma. *World J Gastroenterol.* 2017; 23:1568–75.  
<https://doi.org/10.3748/wjg.v23.i9.1568>  
PMID:[28321157](https://pubmed.ncbi.nlm.nih.gov/28321157/)
12. Huang R, Cho WC, Sun Y, Katie Chan KH. The Lung Cancer Associated MicroRNAs and Single Nucleotides Polymorphisms: a Mendelian Randomization Analysis. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020; 2020:2346–52.  
<https://doi.org/10.1109/EMBC44109.2020.9176344>  
PMID:[33018478](https://pubmed.ncbi.nlm.nih.gov/33018478/)
13. Wang WJ, Li HT, Yu JP, Han XP, Xu ZP, Li YM, Jiao ZY, Liu HB. A Competing Endogenous RNA Network Reveals Novel Potential lncRNA, miRNA, and mRNA Biomarkers in the Prognosis of Human Colon Adenocarcinoma. *J Surg Res.* 2019; 235:22–33.  
<https://doi.org/10.1016/j.jss.2018.09.053>  
PMID:[30691798](https://pubmed.ncbi.nlm.nih.gov/30691798/)
14. Kim WJ, Bae SC. Molecular biomarkers in urothelial bladder cancer. *Cancer Sci.* 2008; 99:646–52.  
<https://doi.org/10.1111/j.1349-7006.2008.00735.x>  
PMID:[18377416](https://pubmed.ncbi.nlm.nih.gov/18377416/)
15. Han Y, Chen J, Zhao X, Liang C, Wang Y, Sun L, Jiang Z, Zhang Z, Yang R, Chen J, Li Z, Tang A, Li X, et al. MicroRNA expression signatures of bladder cancer revealed by deep sequencing. *PLoS One.* 2011; 6:e18286.  
<https://doi.org/10.1371/journal.pone.0018286>  
PMID:[21464941](https://pubmed.ncbi.nlm.nih.gov/21464941/)
16. Mishra S, Yadav T, Rani V. Exploring miRNA based approaches in cancer diagnostics and therapeutics. *Crit Rev Oncol Hematol.* 2016; 98:12–23.  
<https://doi.org/10.1016/j.critrevonc.2015.10.003>  
PMID:[26481951](https://pubmed.ncbi.nlm.nih.gov/26481951/)
17. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* 2006; 20:515–24.  
<https://doi.org/10.1101/gad.1399806>  
PMID:[16510870](https://pubmed.ncbi.nlm.nih.gov/16510870/)
18. Wang X, Wu F, Wang H, Duan X, Huang R, Tuersuntuoheti A, Su L, Yan S, Zhao Y, Lu Y, Li K, Yao J, Luo Z, et al. PDCD6 cooperates with C-Raf to facilitate colorectal cancer progression via Raf/MEK/ERK activation. *J Exp Clin Cancer Res.* 2020; 39:147.  
<https://doi.org/10.1186/s13046-020-01632-9>  
PMID:[32746883](https://pubmed.ncbi.nlm.nih.gov/32746883/)
19. Zhang L, Chen X, Liu B, Han J. MicroRNA-124-3p directly targets PDCD6 to inhibit metastasis in breast cancer. *Oncol Lett.* 2018; 15:984–90.  
<https://doi.org/10.3892/ol.2017.7358>  
PMID:[29387242](https://pubmed.ncbi.nlm.nih.gov/29387242/)
20. Zhou B, Zhang P, Tang T, Zhang K, Wang Y, Song Y, Liao H, Zhang L. Prognostic value of PDCD6 polymorphisms and the susceptibility to bladder cancer. *Tumour Biol.* 2014; 35:7547–54.  
<https://doi.org/10.1007/s13277-014-2010-1>  
PMID:[24792888](https://pubmed.ncbi.nlm.nih.gov/24792888/)
21. Shen TC, Chang WS, Hsia TC, Li HT, Chen WC, Tsai CW, Bau DT. Contribution of programmed cell death 6 genetic variations, gender, and smoking status to lung cancer. *Onco Targets Ther.* 2019; 12:6237–44.  
<https://doi.org/10.2147/OTT.S205544>  
PMID:[31496727](https://pubmed.ncbi.nlm.nih.gov/31496727/)
22. Liu X. Up-regulation of miR-20a by HPV16 E6 exerts growth-promoting effects by targeting PDCD6 in cervical carcinoma cells. *Biomed Pharmacother.* 2018; 102:996–1002.  
<https://doi.org/10.1016/j.biopha.2018.03.154>  
PMID:[29710555](https://pubmed.ncbi.nlm.nih.gov/29710555/)
23. Burotto M, Chiou VL, Lee JM, Kohn EC. The MAPK pathway across different malignancies: a new perspective. *Cancer.* 2014; 120:3446–56.  
<https://doi.org/10.1002/cncr.28864>  
PMID:[24948110](https://pubmed.ncbi.nlm.nih.gov/24948110/)
24. Yang B, Han ZY, Wang WJ, Ma YB, Chu SH. GNG5 is an unfavourable independent prognostic indicator of gliomas. *J Cell Mol Med.* 2020; 24:12873–8.  
<https://doi.org/10.1111/jcmm.15923>  
PMID:[33000557](https://pubmed.ncbi.nlm.nih.gov/33000557/)
25. Jin R, Jin YY, Tang YL, Yang HJ, Zhou XQ, Lei Z. GPNMB silencing suppresses the proliferation and metastasis of osteosarcoma cells by blocking the PI3K/Akt/mTOR signaling pathway. *Oncol Rep.* 2018; 39:3034–40.  
<https://doi.org/10.3892/or.2018.6346>  
PMID:[29620278](https://pubmed.ncbi.nlm.nih.gov/29620278/)
26. Alexander TB, Gu Z, Iacobucci I, Dickerson K, Choi JK, Xu B, Payne-Turner D, Yoshihara H, Loh ML, Horan J, Buldini B, Basso G, Elitzur S, et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature.* 2018; 562:373–9.  
<https://doi.org/10.1038/s41586-018-0436-0>  
PMID:[30209392](https://pubmed.ncbi.nlm.nih.gov/30209392/)

27. Xiao W, Bharadwaj M, Levine M, Farnhoud N, Pastore F, Getta BM, Hultquist A, Famulare C, Medina JS, Patel MA, Gao Q, Lewis N, Pichardo J, et al. *PHF6* and *DNMT3A* mutations are enriched in distinct subgroups of mixed phenotype acute leukemia with T-lineage differentiation. *Blood Adv.* 2018; 2:3526–39. <https://doi.org/10.1182/bloodadvances.2018023531> PMID: [30530780](https://pubmed.ncbi.nlm.nih.gov/30530780/)
- Noushin Farnhoud,<sup>2</sup> Friederike Pastore,<sup>3</sup> Bartlomiej M. Getta,<sup>4</sup> Anne Hultquist,<sup>3</sup> Christopher Famulare,<sup>2</sup> Juan S. Medina,<sup>2</sup> Minal A. Patel,<sup>2</sup> Qi Gao,<sup>1</sup> Natasha Lewis,<sup>1</sup> Janine Pichardo,
28. Kurzer JH, Weinberg OK. *PHF6* Mutations in Hematologic Malignancies. *Front Oncol.* 2021; 11:704471. <https://doi.org/10.3389/fonc.2021.704471> PMID: [34381727](https://pubmed.ncbi.nlm.nih.gov/34381727/)
29. Yu Q, Zhou J, Jian Y, Xiu Z, Xiang L, Yang D, Zeng W. MicroRNA-214 suppresses cell proliferation and migration and cell metabolism by targeting PDK2 and PHF6 in hepatocellular carcinoma. *Cell Biol Int.* 2019. [Epub ahead of print]. <https://doi.org/10.1002/cbin.11207> PMID: [31329335](https://pubmed.ncbi.nlm.nih.gov/31329335/)
30. Wang J, Leung JW, Gong Z, Feng L, Shi X, Chen J. PHF6 regulates cell cycle progression by suppressing ribosomal RNA synthesis. *J Biol Chem.* 2013; 288:3174–83. <https://doi.org/10.1074/jbc.M112.414839> PMID: [23229552](https://pubmed.ncbi.nlm.nih.gov/23229552/)
31. Hsu YC, Chen TC, Lin CC, Yuan CT, Hsu CL, Hou HA, Kao CJ, Chuang PH, Chen YR, Chou WC, Tien HF. Phf6-null hematopoietic stem cells have enhanced self-renewal capacity and oncogenic potentials. *Blood Adv.* 2019; 3:2355–67. <https://doi.org/10.1182/bloodadvances.2019000391> PMID: [31395598](https://pubmed.ncbi.nlm.nih.gov/31395598/)
32. López-Coral A, Del Vecchio GJ, Chahine JJ, Kallakury BV, Tuma PL. MAL2-Induced Actin-Based Protrusion Formation is Anti-Oncogenic in Hepatocellular Carcinoma. *Cancers (Basel).* 2020; 12:E422. <https://doi.org/10.3390/cancers12020422> PMID: [32059473](https://pubmed.ncbi.nlm.nih.gov/32059473/)
33. Zheng C, Wang J, Zhang J, Hou S, Zheng Y, Wang Q. Myelin and lymphocyte protein 2 regulates cell proliferation and metastasis through the Notch pathway in prostate adenocarcinoma. *Transl Androl Urol.* 2021; 10:2067–77. <https://doi.org/10.21037/tau-21-244> PMID: [34159087](https://pubmed.ncbi.nlm.nih.gov/34159087/)
34. Eguchi D, Ohuchida K, Kozono S, Ikenaga N, Shindo K, Cui L, Fujiwara K, Akagawa S, Ohtsuka T, Takahata S, Tokunaga S, Mizumoto K, Tanaka M. MAL2 expression predicts distant metastasis and short survival in pancreatic cancer. *Surgery.* 2013; 154:573–82. <https://doi.org/10.1016/j.surg.2013.03.010> PMID: [23876361](https://pubmed.ncbi.nlm.nih.gov/23876361/)
35. Fang Y, Wang L, Wan C, Sun Y, Van der Jeught K, Zhou Z, Dong T, So KM, Yu T, Li Y, Eyvani H, Colter AB, Dong E, et al. MAL2 drives immune evasion in breast cancer by suppressing tumor antigen presentation. *J Clin Invest.* 2021; 131:140837. <https://doi.org/10.1172/JCI140837> PMID: [32990678](https://pubmed.ncbi.nlm.nih.gov/32990678/)
36. Li J, Li Y, Liu H, Liu Y, Cui B. The four-transmembrane protein MAL2 and tumor protein D52 (TPD52) are highly expressed in colorectal cancer and correlated with poor prognosis. *PLoS One.* 2017; 12:e0178515. <https://doi.org/10.1371/journal.pone.0178515> PMID: [28562687](https://pubmed.ncbi.nlm.nih.gov/28562687/)
37. Zhang B, Xiao J, Cheng X, Liu T. MAL2 interacts with IQGAP1 to promote pancreatic cancer progression by increasing ERK1/2 phosphorylation. *Biochem Biophys Res Commun.* 2021; 554:63–70. <https://doi.org/10.1016/j.bbrc.2021.02.146> PMID: [33780861](https://pubmed.ncbi.nlm.nih.gov/33780861/)
38. Goldenring JR. A central role for vesicle trafficking in epithelial neoplasia: intracellular highways to carcinogenesis. *Nat Rev Cancer.* 2013; 13:813–20. <https://doi.org/10.1038/nrc3601> PMID: [24108097](https://pubmed.ncbi.nlm.nih.gov/24108097/)
39. Ji SM. Overexpression of SLC25A15 is involved in the proliferation of cutaneous melanoma and leads to poor prognosis. *Med Sci (Paris).* 2018; 34:74–80. <https://doi.org/10.1051/medsci/201834f113> PMID: [30403179](https://pubmed.ncbi.nlm.nih.gov/30403179/)
40. Kong Z, Wan X, Lu Y, Zhang Y, Huang Y, Xu Y, Liu Y, Zhao P, Xiang X, Li L, Li Y. Circular RNA circFOXO3 promotes prostate cancer progression through sponging miR-29a-3p. *J Cell Mol Med.* 2020; 24:799–813. <https://doi.org/10.1111/jcmm.14791> PMID: [31733095](https://pubmed.ncbi.nlm.nih.gov/31733095/)
41. Martinelli D, Diodato D, Ponzi E, Monné M, Boenzi S, Bertini E, Fiermonte G, Dionisi-Vici C. The hyperornithinemia-hyperammonemia-homocitrullinuria syndrome. *Orphanet J Rare Dis.* 2015; 10:29. <https://doi.org/10.1186/s13023-015-0242-9> PMID: [25874378](https://pubmed.ncbi.nlm.nih.gov/25874378/)
42. Palmieri F. The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol Aspects Med.* 2013; 34:465–84. <https://doi.org/10.1016/j.mam.2012.05.005> PMID: [23266187](https://pubmed.ncbi.nlm.nih.gov/23266187/)

43. Li R, Tang X, Xu C, Guo Y, Qi L, Li S, Ren Q, Wu J, Chen D. Circular RNA NF1-419 Inhibits Proliferation and Induces Apoptosis by Regulating Lipid Metabolism in Astrogloma Cells. *Recent Pat Anticancer Drug Discov.* 2021. [Epub ahead of print].  
<https://doi.org/10.2174/1574892816666210729125802>  
PMID:[34376137](https://pubmed.ncbi.nlm.nih.gov/34376137/)
44. Wang YT, Lin MR, Chen WC, Wu WH, Wang FS. Optimization of a modeling platform to predict oncogenes from genome-scale metabolic networks of non-small-cell lung cancers. *FEBS Open Bio.* 2021; 11:2078–94.  
<https://doi.org/10.1002/2211-5463.13231>  
PMID:[34137202](https://pubmed.ncbi.nlm.nih.gov/34137202/)
45. Birge RB, Boeltz S, Kumar S, Carlson J, Wanderley J, Calianese D, Barcinski M, Brekken RA, Huang X, Hutchins JT, Freimark B, Empig C, Mercer J, et al. Phosphatidylserine is a global immunosuppressive signal in efferocytosis, infectious disease, and cancer. *Cell Death Differ.* 2016; 23:962–78.  
<https://doi.org/10.1038/cdd.2016.11>  
PMID:[26915293](https://pubmed.ncbi.nlm.nih.gov/26915293/)
46. Burstyn-Cohen T, Maimon A. TAM receptors, Phosphatidylserine, inflammation, and Cancer. *Cell Commun Signal.* 2019; 17:156.  
<https://doi.org/10.1186/s12964-019-0461-0>  
PMID:[31775787](https://pubmed.ncbi.nlm.nih.gov/31775787/)

# Expression status and prognostic significance of mitochondrial dynamics OPA3 in human ovarian cancer

Hsiao-Wen Tsai<sup>1,2,3</sup>, Chia-Jung Li<sup>1,2</sup>, Li-Te Lin<sup>1,2,3</sup>, An-Jen Chiang<sup>1,4</sup>, San-Nung Chen<sup>1</sup>, Zhi-Hong Wen<sup>5</sup>, Kuan-Hao Tsui<sup>1,2,3,4,6,7,8,9</sup>

<sup>1</sup>Department of Obstetrics and Gynaecology, Kaohsiung Veterans General Hospital, Kaohsiung 813, Taiwan

<sup>2</sup>Institute of Biopharmaceutical Sciences, National Sun Yat-Sen University, Kaohsiung 804, Taiwan

<sup>3</sup>Department of Obstetrics and Gynaecology, National Yang-Ming University School of Medicine, Taipei 112, Taiwan

<sup>4</sup>Institute of Biomedical Sciences, National Sun Yat-Sen University, Kaohsiung 804, Taiwan

<sup>5</sup>Department of Marine Biotechnology and Resources, National Sun Yat-Sen University, Kaohsiung 804, Taiwan

<sup>6</sup>Department of Obstetrics and Gynecology, Taipei Veterans General Hospital, Taipei 112, Taiwan

<sup>7</sup>Department of Pharmacy and Master Program, College of Pharmacy and Health Care, Tajen University, Pingtung County 907, Taiwan

<sup>8</sup>Department of Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan

<sup>9</sup>College of Health and Nursing, Meiho University, Pingtung 912, Taiwan

**Correspondence to:** Kuan-Hao Tsui; email: [khtsui@vghks.gov.tw](mailto:khtsui@vghks.gov.tw), <https://orcid.org/0000-0001-8090-8657>

**Keywords:** multi-omics, ovarian cancer, OPA3, immune infiltration

**Received:** December 31, 2021

**Accepted:** April 25, 2022

**Published:** May 4, 2022

**Copyright:** © 2022 Tsai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Early diagnosis of ovarian cancer and the discovery of prognostic markers can significantly improve survival and reduce mortality. OPA3 protein exists in a structure called mitochondria, which is the energy production center of cells, but its molecular and biological functions in ovarian cancer are still unclear. Here, the expression of OPA3 mRNA in ovarian cancer was estimated using TCGA, Oncomine, TIMER databases. We found that functional OPA3 activation caused by mutations and profound deletions predicted poor prognosis in OV patients. OPA3 was highly expressed in both OV tissues and cells compared to normal ovarian tissues/cells. High OPA3 expression is associated with poorer overall survival (OS). The association between OPA3 and immune infiltration of ovarian cancer was assessed by TIMER and CIBERSORT algorithms. OPA3 showed a strong correlation with various immune marker sets. Most importantly, pharmacogenetic analysis of OV cell lines revealed that OPA3 inactivation was associated with increased sensitivity to PFI-1, and WZ4003. Therefore, we investigated the clinical application of OPA3 to provide a basis for sensitive diagnosis, prognosis and targeted treatment of ovarian cancer.

## INTRODUCTION

With the continuous advancement of medical technology, ovarian cancer is still a troublesome cancer and frightening disease. The global incidence rate of ovarian cancer is 42 cases per 100,000 people, making it the second leading cause of death among gynecological cancers. [1, 2]. The high mortality rate makes ovarian cancer a worldwide public health issue. [3, 4]. Currently, due to the lack of sensitive tumor markers, early diagnosis of ovarian cancer is difficult work. Most

ovarian cancer patients are diagnosed at an advanced stage of cancer because they came for help only when they feel abdominal pain or ascites with clinical relevance [5–7]. Therefore, the discovery of new potential targets for early diagnosis of ovarian cancer is necessary and worthy.

Metastasis is the cause of most cancer progression, leading to clinical treatment failure and patient death. Cancer cells enter the blood or lymphatic vessels through intravascular vessels. After extravasation from

these vessels, they form clonal lesions in distant organs [8, 9]. Cell migration is a critical step in cancer metastasis, and several studies have demonstrated the role of mitochondria in cancer cell metastasis. For example, mitochondrial dynamics regulate the progression and distant metastasis of breast cancer cells. However, dynamic imbalances are more frequently observed in breast cancer cells with high metastasis [10]. More recently, additional evidence has emerged suggesting a regulatory role for mitochondrial dynamics in cancer metastasis [11–13]. Dysregulation of mitochondrial dynamics can lead to a malignant phenotype of cancer. Various lines of evidence link mitochondrial dynamics to cancer development and progression. Mitochondrial fusion and fission affect mitochondrial transport in lymphocytes and cancer cells. Dysregulation of mitochondrial fusion/ fission inhibits lymphocyte polarization and migration [14]. The above studies suggest that mitochondrial dynamics affect the progression of cancer.

The pathophysiology of ovarian cancer is not yet fully understood. The progression of malignant tumors is complex and involves interactions between oncogenes and the microenvironment [3, 7]. Recent studies have shown the association of oncogenes and ovarian cancers [15]. The optic atrophy 3 (OPA3) gene was first identified in patients with optic neuropathy, and OPA3 is located in the mitochondria and has a biological function in maintaining the shape and structure of the mitochondria. Previous study had found overexpression of OPA3 may cause mitochondria fragmentation and knockdown of OPA3 cause elongation of mitochondria in ARPE-19 cells [16]. Since the interaction between oncogenes and ovarian cancer is a key factor in tumor development, we tried to identify the role of OPA3 in ovarian cancer.

Therefore, our current study is dedicated to explore the impact of OPA3 in the diagnosis and prognosis of ovarian cancer patients. We further explored the role of OPA3 in ovarian cancer by systematically analyzing OPA3 expression and analyzing the potential mechanisms in the malignant transformation of ovarian cancer in multiple publicly available databases.

## **MATERIALS AND METHODS**

### **Oncomine**

The Oncomine platform is a publicly accessible online tumor-associated gene microarray database for collecting relevant gene expression profiles and related clinical information [17]. The Oncomine database includes approximately 200 biopsies analyzing OPA3 gene transcript levels in different tumors and normal tissues.

### **cBioPortal**

cBioPortal is used to perform interactive analysis of biomolecules in tumor tissues in the TCGA database. We used cBioPortal to mine gene set data and ovarian cancer gene variants. A hypothetical copy number of 370 cases identified using GISTIC 2.0 was deployed. In module comparison/survival, we analyzed OPA3 mutations, copy number variants (CNV) and gene coexistence in ovarian cancer.

### **GEPIA2**

OPA3 expression data in various human tumor tissues were obtained from the Gene Expression Profiling Interactive Analysis (GEPIA) public database [18, 19]. GEPIA2 was used to compare OPA3 expression levels in OV, and we used default parameters to analyze OPA3 expression at different OV stages.

### **Cells and cell culture**

In this study, multiple human ovarian cancer cell lines including: OC-117-VGH cells, OC-117-VGH cells, OCPC-2-VGH cells, OC-3-VGH cells, TOV-21G cells and NIH-OVCAR-3 cells (BCRC#6060, #60602, #60603, #60599, #60407, #605511, Hsinchu, Taiwan) were cultured in a humidified atmosphere supplemented with 1.5 g/L sodium bicarbonate and 10% FBS (ThermoFisher Scientific, MA, USA) at 37° C in 95% air and 5% CO<sub>2</sub>.

### **RNA extraction and real-time PCR**

RNA isolation of all samples was performed using EasyPrep Total RNA Kit (BIOTOOLS Co., Ltd., Taipei, Taiwan.), as indicated above. Next, 1 µg of total RNA was reverse transcribed using a ToolScript MMLV RT kit. (BIOTOOLS Co., Ltd.) in an Applied Biosystems™ (ABI 7500) under the following reaction conditions: 65° C for 5 min, 42° C for 60 min, and 70° C for 10 min. The resulting cDNAs were subjected to quantitative real-time PCR (qRT-PCR) analysis using a TOOLS 2X SYBR qPCR Mix (BIOTOOLS Co., Ltd.) in a StepOnePlus Real-Time PCR system. The conditions used included an initial step at 95° C for 10 min, followed by 40 cycles at 95° C for 15 s and a final step at 60° C for 1 min. Ct values were calculated using U6 (RNU6-1) as reference. Untreated samples were used as controls to determine the relative fold-changes in mRNA expression.

### **Human BRCA tissue microarray (TMA) and immunohistochemistry (IHC) analysis**

The IHC staining and scoring in this study was the same as our previously published method. We analyzed

OPA3 expression on tissue array slides (CJ2, SuperBioChips Laboratories, Korea) purchased from human ovarian cancer and measured and scored immunohistochemistry (IHC) as described in a previous publication [20]. To detect protein levels, *in situ* hybridization was performed with OPA3 antibody (A4995, ABclonal, MA, USA).

### Human Protein Atlas

We validated OPA3 levels in normal and tumor tissues using the Human Protein Atlas public database. This database is designed to map the biology of all human proteins in cells, tissues and organs by integrating various histological techniques including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics and systems.

### Tumor immune estimation resource (TIMER)

TIMER is a web-based resource for the systematic evaluation of the clinical impact of different immune cells on different cancer types. We used this database to compare OPA3 expression with changes in the ovarian tumor microenvironment [21]. We selected OPA3 in the TIMER database as an input for the detection of ovarian cancer under the Immune Association Module and analyzed the correlation of OPA3 with other immune cells.

### Statistical analyses

All data are shown as the mean  $\pm$  S.E.M. In the case of single mean comparisons, data were analyzed by Student's t-tests. For multiple comparisons, statistical difference was calculated by one-way ANOVA. *Post hoc* analysis was performed with Tukey's test when one-way ANOVA showed significant differences. Statistical analyses were carried out using Prism 8.0 (GraphPad Software). \* $p$  value  $< 0.05$ ; \*\* $p$  value  $< 0.01$ ; \*\*\* $p$  value  $< 0.001$ .

### Data availability statement

The dataset supporting the conclusions of this article is included within the article.

## RESULTS

### Characterization, mutation and copy number changes of OPA3 gene in OV

To identify changes in potential oncogenes in OV in the TCGA dataset. Among ovarian cancers, the majority of tumors were serous ovarian cancer ( $n = 1202$ ) and high-grade serous ovarian cancer ( $n = 531$ ) (Figure 1A). The

red dots in the "Locus Enrichment" diagram represent a gene and its associated location on the chromosome. The correlation between RNA level and CNV for each gene (Figure 1B). The transcriptional and translational expression levels of OPA3 in the reproductive system showed a high expression of OPA3 in the ovary (Figure 1C). We further analyzed the results from the cBioPortal database and showed that OPA3 expression in ovarian and cervical cancers was most significant in female reproductive cancers (Figure 1D). The same results were reflected in the survival rate in the ENSG00000125741.4 database, where the HR was similarly increased in ovarian cancers with high OPA3 expression (Figure 1E). In addition, OPA3 mRNA level was significantly higher in ovarian cancer than in normal tissue in both Hendrix and TCGA databases (Figure 1F, 1G).

### The genetic alteration landscape of OPA3 in OV

We then used the cBioPortal database to evaluate the mutation and frequency of OPA3 in OV tissues based on data from the Pan-Cancer Atlas database of OV patients obtained from TCGA. We found that 3% of OPA3 genes were mutated in various cancers (Figure 2A). We examined the genetic alterations of OPA3 in various tumor types in the TCGA dataset. We found that the top three most frequent OPA3 gene alterations were in cervical cancer tumors, BRCA, and ovarian cancer (Figure 2B), which also indicates that OPA3 is more frequently mutated in female cancers. We further explored the specific changes in each gene, with the highest number of mutated sites in the core structure of OPA3 (Figure 2C). The somatic cell copy alteration (sCNA) module allows users to compare the distribution of different cancer types by the sCNA status of genes from the TCGA database. For this purpose, we examined the ratio of Arm-level deletion, Diploid/Normal, and Arm-level gain of OPA3 genes.

We observed a high "Arm-level deletion" of OPA3 master regulators in ovarian cancer (Figure 3A). Next, we observed significant changes in OPA3 gain and loss in the CNV ratio distribution and box plot (cor. = 0.774) (Figure 3B). While OV patients were divided into low and high OPA3 groups using median expression thresholds, the expression difference between the two groups was significant ( $p < 0.049$ ) in Figure 3C. Kaplan-Meier survival analysis showed a shorter survival rate in the high OPA3 group than in the low OPA3 group (Figure 3C). We used TNM plots to analyze OPA3 expression from gene microarray data and RNA-seq data (Figure 3D, 3F) ( $p = 1.71e-02$ ,  $p = 3.2e-06$ ). We also analyzed the specificity of OPA3 in OV from both data (gene chip and RNA-seq). OPA3 expression in tumor biopsies was higher than the

percentage of normal biopsies at each of the major cut-off values in Figure 3E, 3G.

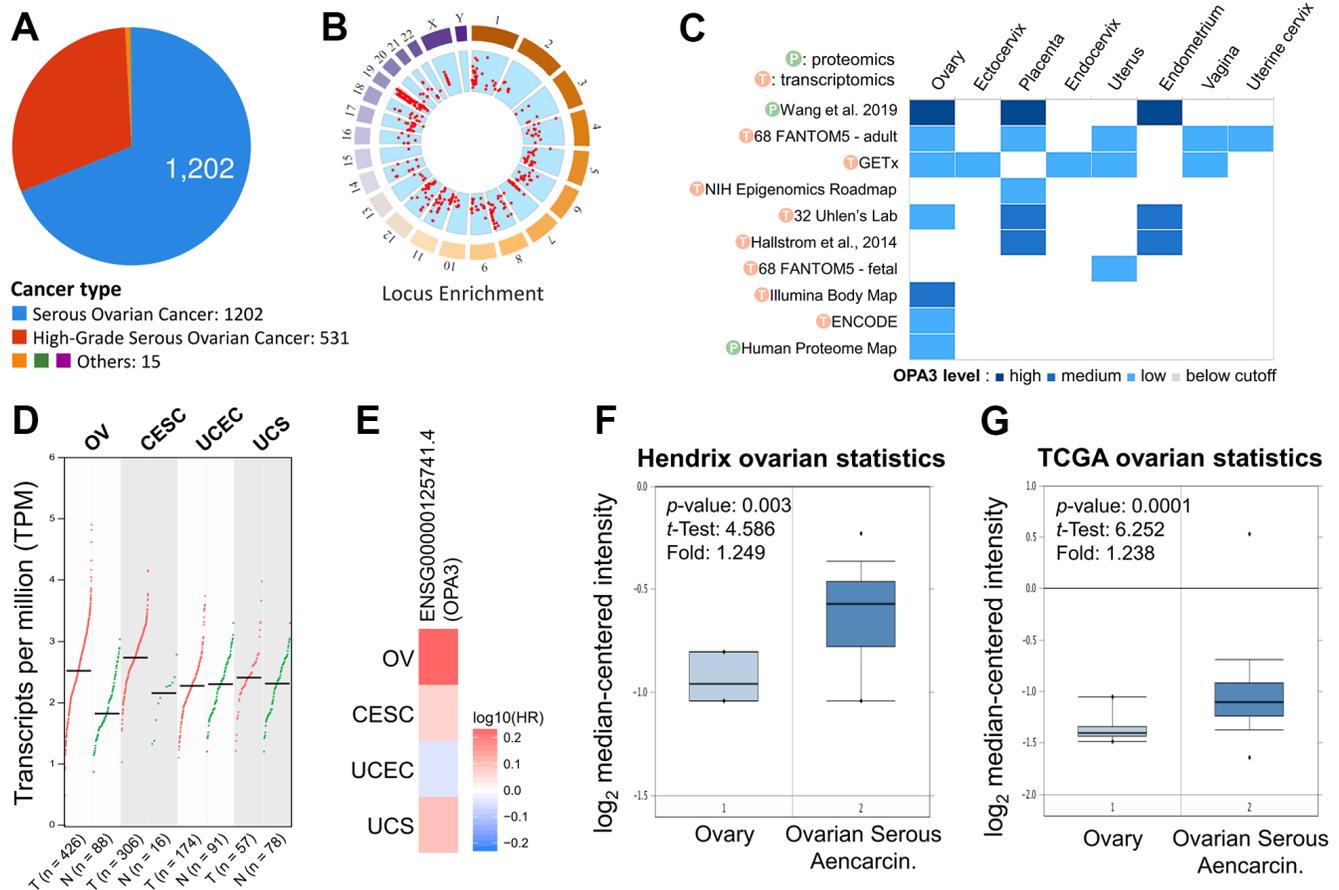
### Protein expression levels of OPA3 in OV patients

First, we identified OPA3 in OV patients through the Human Protein Atlas (HPA) database showing moderate staining levels (Figure 4A). Next, to further evaluate the accuracy of the HPA database, we used immunohistochemistry to assess the protein levels of OPA3 detected in TMA using human ovarian cancer tissue microarray (TMA) at different stages. Figure 4B shows the results of OPA3 protein expression in IHC-stained ovarian cancer tissues. The H-score of OPA3 showed a positive correlation increase with stage, and OPA3 protein expression was significantly higher in stage I-IV patients than in benign patients at different stages. OPA3 expression was higher in stage III+IV than in stages I and II (Figure 4C). The analysis of

clinical biopsies confirmed the same trend as the above-mentioned multi-histological database, with higher OPA3 expression in advanced stages of ovarian cancer. We also tested the endogenous levels of OPA3 in multiple ovarian cancer cell lines and the data indicated that OPA3 expression in various ovarian cancer cell lines was positively correlated with higher malignancy (Figure 4D).

### Analysis of differentially expressed genes correlated with OPA3 in OV

Next, we attempted to show the association between the efficacy of OPA3 in ovarian cancer and the expression of the target genes. We used statistical tests to calculate the two-way predictive and descriptive scores for each of the more than 16,000-17,000 genes. The results in Figure 5A indicate that 140 genes (red dots) had positive predictive and descriptive scores, while



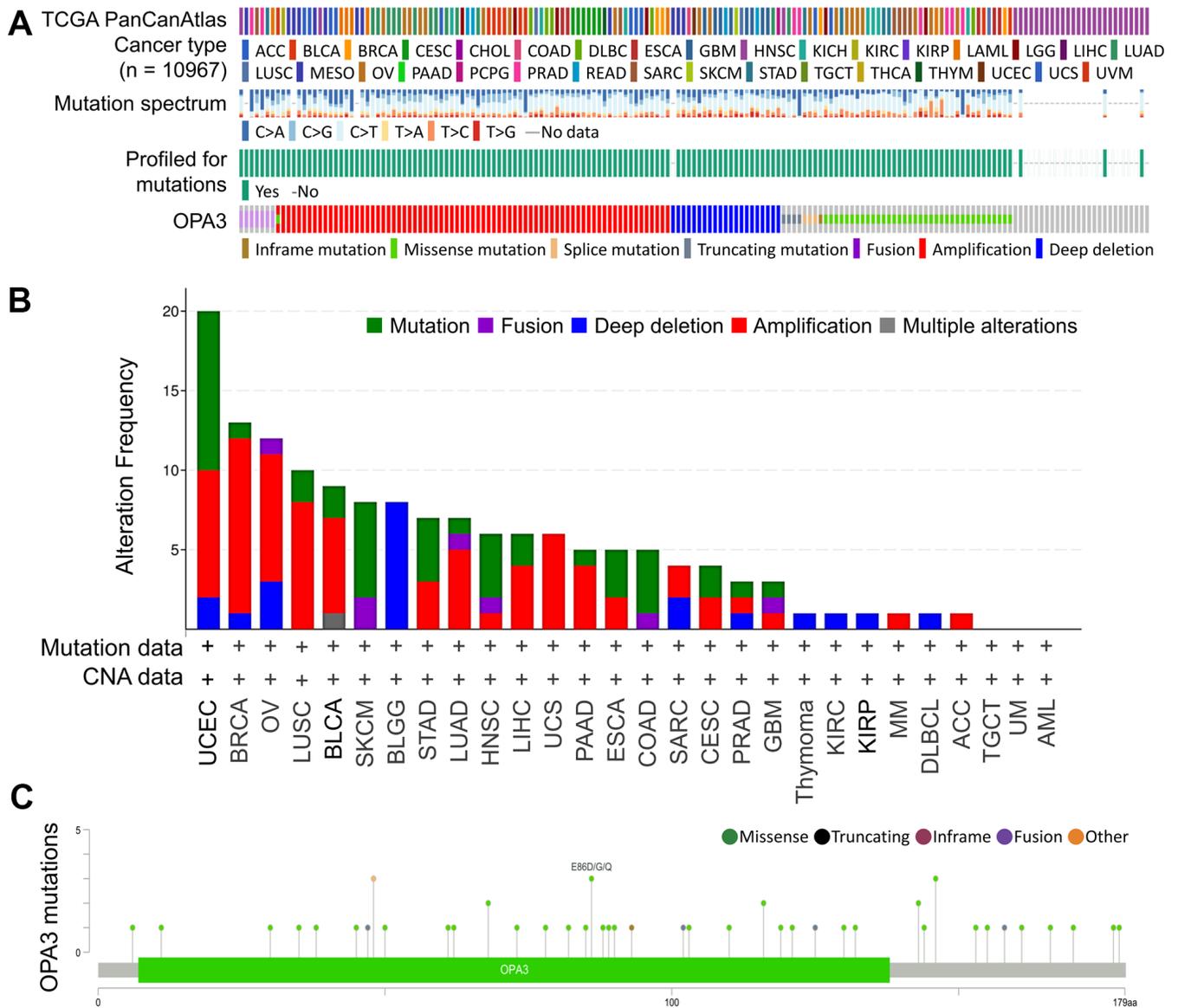
**Figure 1. Expression analysis with OPA3 in the reproductive system.** (A) Percentage of each ovarian cancer type in the TCGA dataset from cBioPortal. (B) Percentage of mutation counts and genomic changes due to copy number changes for different ovarian cancer types in the TCGA dataset. (C) RNA and protein levels of OPA3 in the reproductive system in the GETx database. (D, E) Gene transcriptional expression and survival of OPA3 in the female reproductive system. (F, G) Box plots of OPA3 mRNA levels in OV and normal tissues from Hendrix and TCGA ovarian Statistics.

58 genes (blue dots) had negative scores. The volcano map shows the genes positively and negatively associated with OPA3 (Figure 5B). The red dots are the clusters of genes positively associated with OPA3, while the green dots are the clusters of genes negatively associated with OPA3 ( $p < 0.01$  and  $FDR < 0.01$ ). We enriched the genes with more significant differences and classified them according to their biological functions in Figure 5B. The top 20 significant gene clusters positively and negatively associated with OPA3 by functional enrichment are shown (Figure 5C). The heat map shows the top 50 positively and 50 negatively

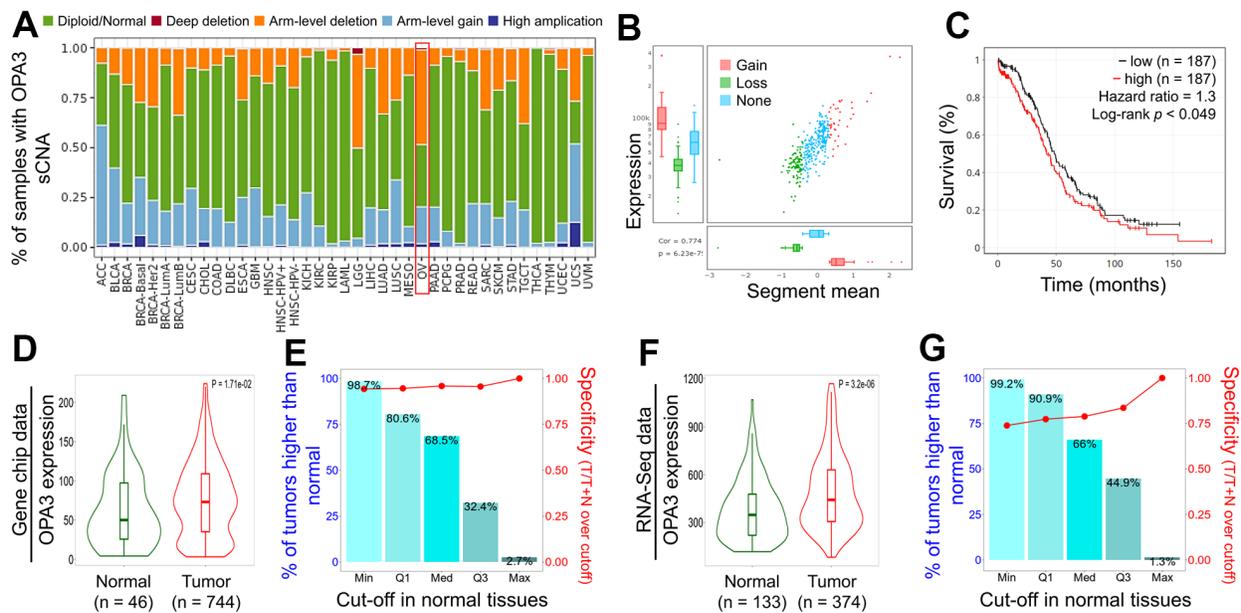
associated genes with the broad effect of OPA3 on the transcriptome (Figure 5D).

### Correlation of OPA3 expression with tumor microenvironment

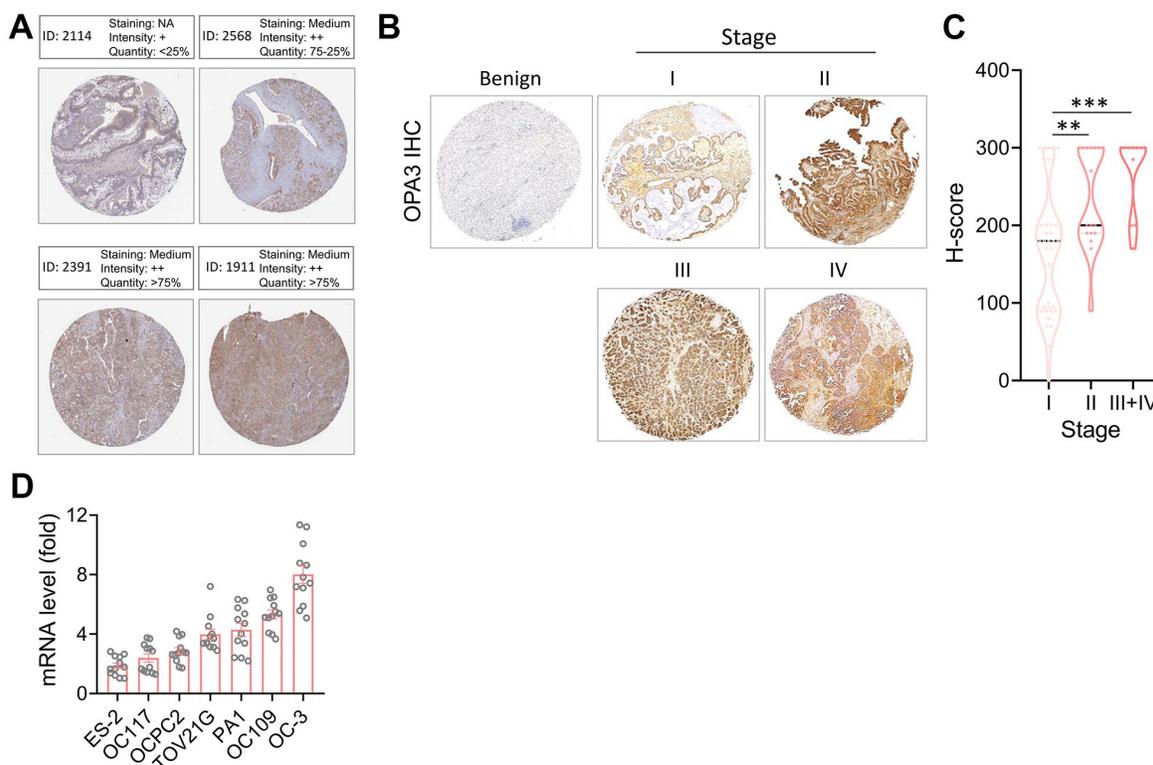
We used single-cell RNA sequencing datasets to analyze transcriptomic data and to clarify heterogeneous cell populations and reconstruct the microenvironment of cell development. The HPA database (10× Genomics) was used to explore the potential association of OPA3 with immune cells in the ovarian microenvironment. We



**Figure 2. Analysis of OPA3-related transcription factor variants in ovarian cancer.** (A) Frequency and type of OPA3 gene mutations in pancancer. (B) Distribution of copy number mutations of pan-oncogenes from TCGA. (C) OPA3 protein structural domain and location of specific mutations. The length of the line linking the mutation annotation to the protein indicates the number of samples with mutations.



**Figure 3. Relative expression and survival of OPA3 in OV tissues based on multiple databases.** (A) Illustration of the definition of somatic cell copy alteration (SCNA) at the OPA3 gene level. (B) Distribution and correlation of CNV in ovarian cancer are marked in red (gain) and green (loss) to visualize the distribution of log2 ratios. (C) Overall survival estimates of OPA3 mRNA levels from the Kaplan-Meier Mapper database. Violin plots (D, F) and box plots (E, G) of OPA3 gene expression from RNA sequencing data and gene microarray data.

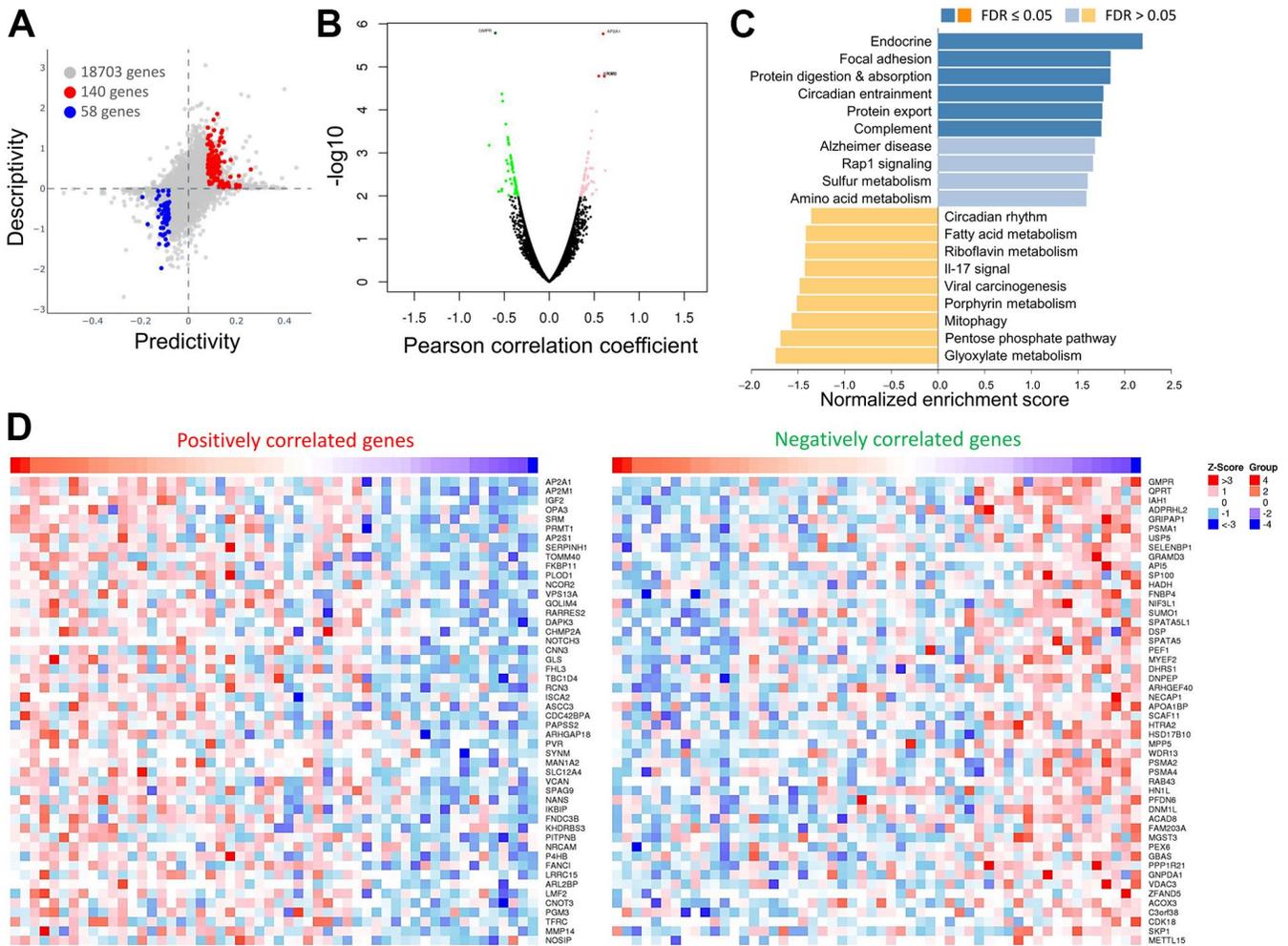


**Figure 4. Protein levels of OPA3 in ovarian carcinoma.** (A) Representative images of OPA3 IHC staining in ovarian cancer from the human Protein Atlas dataset. (B) Representative images of OPA3 expression at different stages of OV. (C) Violin plots of OPA3 expression levels in different stages. (D) RT-PCR was used to detect the expression levels of different ovarian cancer cells. \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Scale bar = 500  $\mu\text{m}$ .

obtained 18,547 single cells in this database as UMAP plots and bar graphs (Figure 6A). The specificity and distribution of OPA3 in different cell populations in the ovary were analyzed to determine the variation in gene number in single-cell types. The heatmap shown in the left panel indicates the expression of different biomarkers in multiple cell clusters in the ovarian tissue. We identified specific immune cells (green boxes) with high expression from the single-cell sequencing data (Figure 6B). Therefore, we further analyzed OPA3 and immune highly correlated genes in female reproductive cancers through different databases with highly correlated immune markers in Figure 6B. The results showed that OPA3 was highly associated with these immune signatures in ovarian cancer among female reproductive cancers (Figure 6C).

### Relationship between OPA3 and different immune marker sets

Our results from the TIMER analysis of the immune database showed that the absence or expansion of different forms of OPA3 copies was associated with the possibility of differential suppression of immune cell infiltration by many immune cells (Figure 7A). In ovarian cancer, OPA3 expression was associated with different types of immune cells. As shown in Figure 7A, Arm-level deletion of OPA3 was associated with B cells and CD8+ T cells, macrophage and dendritic cells; Arm-level gain of OPA3 was associated with CD4+ T cells and dendritic cells. We observed high expression levels of several cells in ovarian cancer in Figure 6 single-cell sequencing. Thus, we found that endothelial cells and dendritic cells are the two immune cell types



**Figure 5. Comparisons and enrichment analysis of gene expression profiles in OPA3 in OV. (A)** The predictability and descriptiveness between mRNA expression functions are plotted with ovarian cancer cell lines. **(B)** Analysis of differential gene expression associated with OPA3 in OV. **(C)** Functional enrichment analysis of OPA3 in OV. **(D)** Heatmap showing the top 50 genes are each significantly positively and negatively correlated with OPA3.

most closely correlated with OPA3 (Figure 7B). We further dissected the correlation between OPA3 and the immune biomarkers of smooth muscle cells, tumor-associated macrophages (TAM), M1 macrophages and M2 macrophages. OPA3 was positively correlated with smooth muscle cell infiltration in OV tissues but not in normal ovarian tissue.

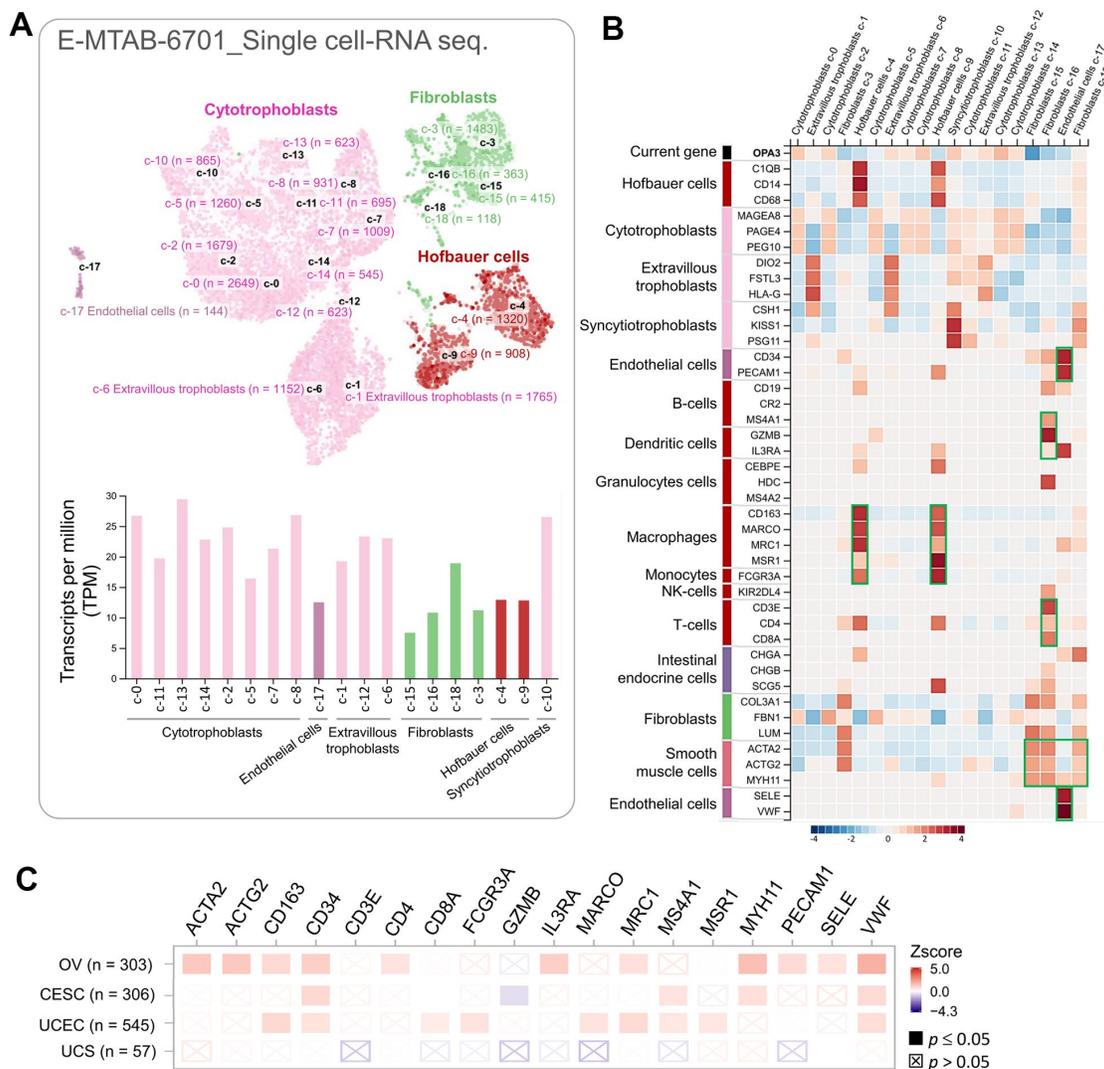
### Pharmacogenetic screening for potential drugs to inhibit OPA3 in OV

We attempted to retrieve the OPA3 gene library from the pharmacogenetic database for screening potential drugs for the therapy of ovarian cancer. As shown in Figure 8A, two of the 476 drugs were significant, including PFI-1 and WZ4003, which inhibit OPA3 overexpression. When

queried in the Q-omics database for the relationship between PFI-1 and WZ4003 in knocking down OPA3 co-expression gene features, we found that the relationship between PFI-1 and WZ4003 for CRISPR OPA3 knockdown We found a high sensitivity and positive correlation between PFI-1 and WZ4003 for CRISPR OPA3 knockdown (PFI-1:  $r = 0.342$ ; WZ4003:  $r = 0.532$ ) (Figure 8B, 8C). Therefore, PFI-1 and WZ4003 have anti-cancer potential to inhibit the growth of ovarian cancer cells with high OPA3 expression.

### DISCUSSION

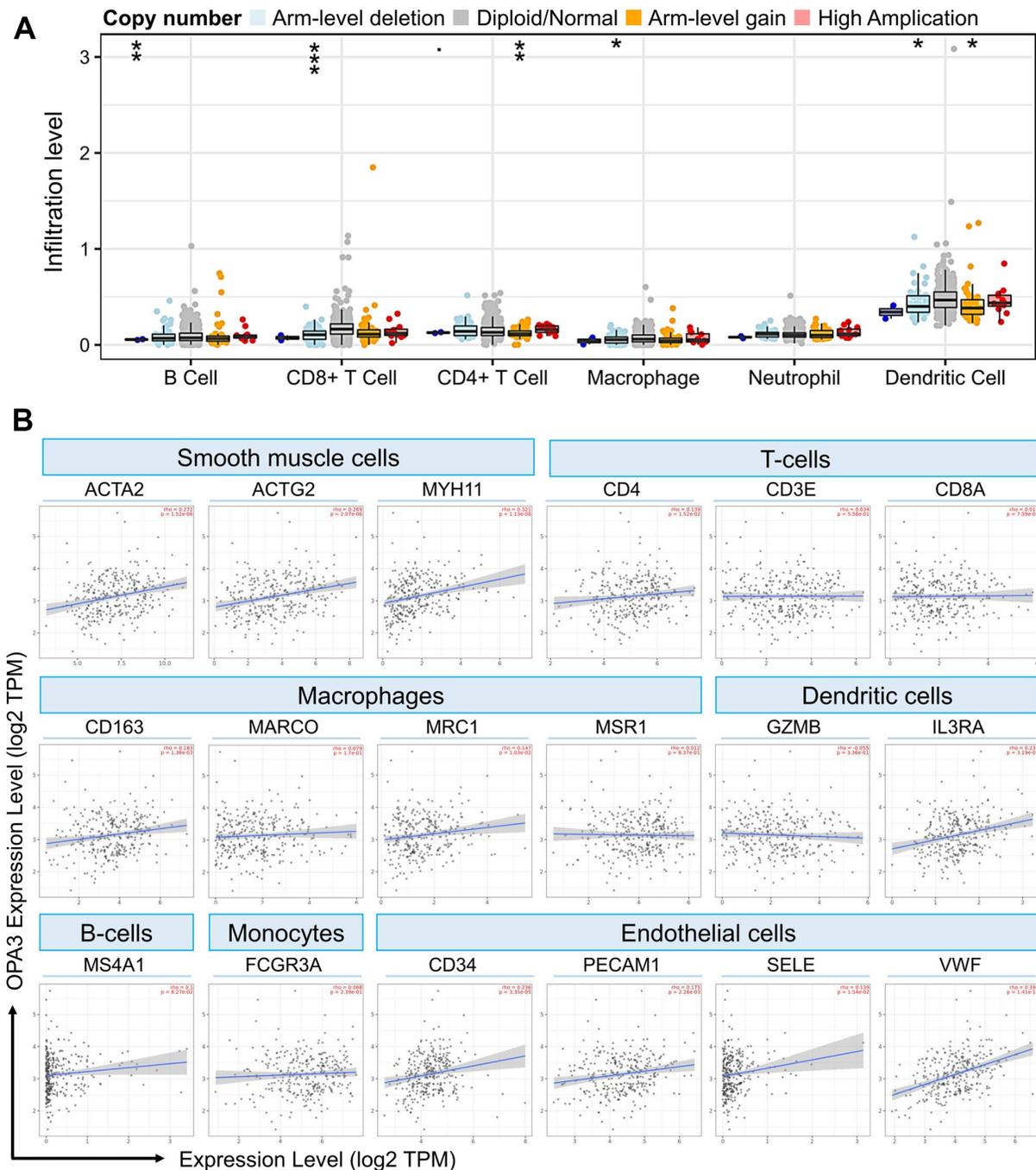
Cancer is a disease characterized by abnormal cell proliferation and metabolic dysregulation [22]. However, previous studies have pointed out that mitochondrial



**Figure 6. The correlation between OPA3 and immunization. (A)** Single cell-RNA sequencing in identified single cell type clusters in placental tissue as shown by UMAP plots and bar graphs. **(B)** Heatmap showing the expression of OPA3 gene and well-known cell type markers in different single cell type clusters of the tissue. **(C)** Relationship between immune cell infiltration and OPA3 expression.

dynamics have been shown to be related to these mechanisms, and evidence suggests that mitochondrial fusion and fission processes are closely associated with cancer progression in different types of cancer [23]. Mitochondrial dynamics balance in cancer cells plays a key role in cellular biogenesis and maintenance of mitochondrial morphology and antioxidation. The key to change in mitochondrial morphology is the fusion

and fission in dynamic equilibrium to maintain the constant level of intracellular nutrients [24]. In addition, cancer metabolism refers to the altered cellular metabolic pathways evident in cancer cells compared to normal cells and to the reprogramming and utilization of nutrients in the tumor microenvironment to increase the nutrient enrichment of cancer cells during metastasis and progression [25].



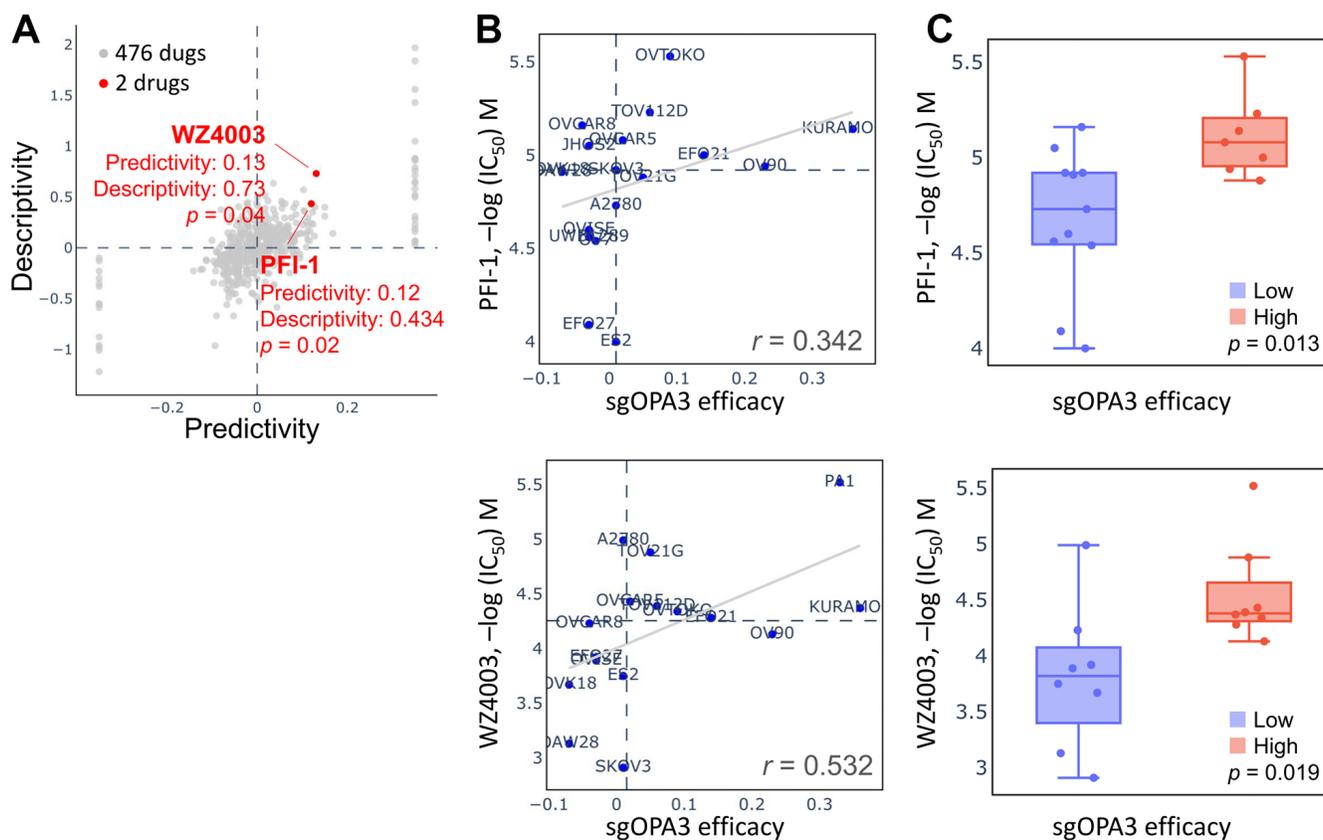
**Figure 7. OPA3 is closely related to immunity in OV.** (A) OPA3 copy number variable (CNV) affects infiltration levels of CD8+ T-cells, macrophages, neutrophils, and dendritic cells in OV. (B) Relationship between immune cell infiltration and OPA3 expression.

The growth of cancer cells requires suitable nutritional supplementation, therefore, cancer cells with high fatty acid synthesis rate and glycolytic activity in proliferating rapidly [26]. Metastatic cancer cells respond by increasing metabolic turnover rate to maintain the energy required for distant metastasis [27, 28]. In general, fragmented mitochondria have lower oxidative metabolic activity compared to tubular mitochondria. In contrast, the limited mitochondrial oxidation preserves glycolytic metabolites that can be used as a resource for cancer cell proliferation. Mitochondrial fission has also been associated with highly activated glycolysis in a variety of cancers [29].

Previous studies have noted that inhibition of DRP1 or in combination with chemotherapy reduces cancer cell proliferation and/or programmed cell death in several cancer types observed *in vitro* and *in vivo*, including pancreatic, colon, lung, hepatocellular carcinoma, and melanoma [30–33]. Furthermore, focusing on related proteins involved in mitochondrial dynamics are considered potential biomarkers of malignancy or therapeutic targets in different tumor types. Thus,

reduced fragmentation of mitochondria is correlated with genomic instability in breast cancer cell lines that normally have high levels of globular mitochondria [34–36]. The underlying mechanisms of cellular dysfunction and cell death induced by mitochondrial dynamic imbalance appear to be factors that severely affect genomic replication, oxidative stress, and mitotic abnormalities in cancer cells. However, once mitochondrial dynamics are imbalanced, mitochondrial dysfunction can also result, including mtDNA mutations and excessive accumulation of ROS. This genomic instability and abnormal mitochondrial function ultimately leads to the initiation of the mitochondrial apoptotic pathway. This also suggests that reducing excessive mitochondrial fission in cancer cells may enhance apoptosis, increase therapeutic sensitivity, and increase therapeutic efficacy by contributing to reduced drug tolerance in cancer cells [37].

There are few studies related to OPA3 and various cancers, and few studies have been conducted on this topic, it is an issue worth exploring in depth. Mitochondrial dynamics are balanced to enable cells to



**Figure 8. Drug sensitivity and cytotoxicity analysis in ovarian cancer cells.** (A) Use of the database to query OPA3 gene signatures and screen for potential drugs. (B) Drug sensitivity of sgOPA3 gene to PFI-1 and WZ4003 in OV cell lines. (C) OPA3 efficacy of PFI-1 and WZ4003 in inhibiting OV cancer cells.

cope with environmental changes and metabolic reprogramming, such as tumor development and remote cancer metastasis. OPA3 may promote cellular energy metabolism, and its upregulation in K-ras-driven cancers may be a mechanism to counteract the negative effects of K-ras on mitochondria to maintain energy homeostasis [38]. Thus, OPA3 may play an important role in ovarian cancer and has the potential to become a therapeutic target or biomarker. This study observed a high level of differential OPA3 gene amplification in pan-cancer. However, gene amplification is a process by which the genome of a specific protein-encoded gene selectively increases while other genes do not increase proportionally.

Finally, we constructed a multi-omics study to characterize OPA3 in ovarian cancer to provide a good predictor for prognosis. Single cell sequencing and network analysis confirmed the regulatory relationship and interaction between immunity and OPA3, which could also serve as a basis for future immunotherapy. However, although we provide clinical data to support our view, the insufficient sample size remains a limitation of this study. Moreover, future prospective and validation studies are necessary to perform to predict the accuracy of genetic features.

## CONCLUSIONS

In summary, we comprehensively analyzed the prognostic value and expression of OPA3 in human ovarian cancer. The high expression of OPA3 indicated that the prognosis of ovarian cancer was poor. Whether inhibition of the key player OPA3 could serve as a diagnostic basis or therapeutic target for prostate cancer and could more comprehensively inhibit the progression of the tumor microenvironment. Of course, further experiments are needed to test these hypotheses. There are relatively few studies on OPA3 in ovarian cancer. Our work may lay the foundation for future exploration of the biological function of OPA3 in tumors and its impact on clinical drug resistance.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This research was funded by the Ministry of Science Technology (MOST 109-2314-B-075B-002, MOST-109-2314-B-075B-014-MY2, MOST 110-2314-B-075B-006 and MOST 110-2314-B-075B-011) and Kaohsiung Veterans General Hospital (VGHKS109-103, 109-105, 109-106, 109-D07, 110-088, 110-143, 110-090, and 110-D06).

## REFERENCES

1. Chen T, Jansen L, Gondos A, Emrich K, Hollecsek B, Luttmann S, Waldmann A, Brenner H, and Gekid Cancer Survival Working Group. Survival of cervical cancer patients in Germany in the early 21st century: a period analysis by age, histology, and stage. *Acta Oncol.* 2012; 51:915–21.  
<https://doi.org/10.3109/0284186X.2012.708105>  
PMID:[22928692](https://pubmed.ncbi.nlm.nih.gov/22928692/)
2. Cho KR, Shih IM. Ovarian cancer. *Annu Rev Pathol.* 2009; 4:287–313.  
<https://doi.org/10.1146/annurev.pathol.4.110807.092246> PMID:[18842102](https://pubmed.ncbi.nlm.nih.gov/18842102/)
3. Barua A, Bitterman P, Abramowicz JS, Dirks AL, Bahr JM, Hales DB, Bradaric MJ, Edassery SL, Rotmensch J, Luborsky JL. Histopathology of ovarian tumors in laying hens: a preclinical model of human ovarian cancer. *Int J Gynecol Cancer.* 2009; 19:531–9.  
<https://doi.org/10.1111/IGC.0b013e3181a41613>  
PMID:[19509547](https://pubmed.ncbi.nlm.nih.gov/19509547/)
4. Kim A, Ueda Y, Naka T, Enomoto T. Therapeutic strategies in epithelial ovarian cancer. *J Exp Clin Cancer Res.* 2012; 31:14.  
<https://doi.org/10.1186/1756-9966-31-14>  
PMID:[22330607](https://pubmed.ncbi.nlm.nih.gov/22330607/)
5. Cobec IM, Sas I, Pirtea L, Cimpean AM, Moatar AE, Ceaușu RA, Raica M. Podoplanin as Key Player of Tumor Progression and Lymph Vessel Proliferation in Ovarian Cancer. *Anticancer Res.* 2016; 36:5265–72.  
<https://doi.org/10.21873/anticancer.11097>  
PMID:[27798887](https://pubmed.ncbi.nlm.nih.gov/27798887/)
6. Rosen DG, Yang G, Liu G, Mercado-Uribe I, Chang B, Xiao XS, Zheng J, Xue FX, Liu J. Ovarian cancer: pathology, biology, and disease models. *Front Biosci (Landmark Ed).* 2009; 14:2089–102.  
<https://doi.org/10.2741/3364> PMID:[19273186](https://pubmed.ncbi.nlm.nih.gov/19273186/)
7. Saad AF, Hu W, Sood AK. Microenvironment and pathogenesis of epithelial ovarian cancer. *Horm Cancer.* 2010; 1:277–90.  
<https://doi.org/10.1007/s12672-010-0054-2>  
PMID:[21761359](https://pubmed.ncbi.nlm.nih.gov/21761359/)
8. Hainaut P, Plymoth A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. *Curr Opin Oncol.* 2013; 25:50–1.  
<https://doi.org/10.1097/CCO.0b013e32835b651e>  
PMID:[23150341](https://pubmed.ncbi.nlm.nih.gov/23150341/)
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74.  
<https://doi.org/10.1016/j.cell.2011.02.013>  
PMID:[21376230](https://pubmed.ncbi.nlm.nih.gov/21376230/)

10. Zhao J, Zhang J, Yu M, Xie Y, Huang Y, Wolff DW, Abel PW, Tu Y. Mitochondrial dynamics regulates migration and invasion of breast cancer cells. *Oncogene*. 2013; 32:4814–24.  
<https://doi.org/10.1038/onc.2012.494> PMID:[23128392](https://pubmed.ncbi.nlm.nih.gov/23128392/)
11. Jung JU, Ravi S, Lee DW, McFadden K, Kamradt ML, Toussaint LG, Sitcheran R. NIK/MAP3K14 Regulates Mitochondrial Dynamics and Trafficking to Promote Cell Invasion. *Curr Biol*. 2016; 26:3288–302.  
<https://doi.org/10.1016/j.cub.2016.10.009> PMID:[27889261](https://pubmed.ncbi.nlm.nih.gov/27889261/)
12. Ko SH, Choi GE, Oh JY, Lee HJ, Kim JS, Chae CW, Choi D, Han HJ. Succinate promotes stem cell migration through the GPR91-dependent regulation of DRP1-mediated mitochondrial fission. *Sci Rep*. 2017; 7:12582.  
<https://doi.org/10.1038/s41598-017-12692-x> PMID:[28974722](https://pubmed.ncbi.nlm.nih.gov/28974722/)
13. Sun X, Cao H, Zhan L, Yin C, Wang G, Liang P, Li J, Wang Z, Liu B, Huang Q, Xing J. Mitochondrial fission promotes cell migration by Ca<sup>2+</sup> /CaMKII/ERK/FAK pathway in hepatocellular carcinoma. *Liver Int*. 2018; 38:1263–72.  
<https://doi.org/10.1111/liv.13660> PMID:[29210177](https://pubmed.ncbi.nlm.nih.gov/29210177/)
14. Campello S, Lacalle RA, Bettella M, Mañes S, Scorrano L, Viola A. Orchestration of lymphocyte chemotaxis by mitochondrial dynamics. *J Exp Med*. 2006; 203:2879–86.  
<https://doi.org/10.1084/jem.20061877> PMID:[17145957](https://pubmed.ncbi.nlm.nih.gov/17145957/)
15. Murphy MA, Wentzensen N. Frequency of mismatch repair deficiency in ovarian cancer: a systematic review This article is a US Government work and, as such, is in the public domain of the United States of America. *Int J Cancer*. 2011; 129:1914–22.  
<https://doi.org/10.1002/ijc.25835> PMID:[21140452](https://pubmed.ncbi.nlm.nih.gov/21140452/)
16. Ryu SW, Jeong HJ, Choi M, Karbowski M, Choi C. Optic atrophy 3 as a protein of the mitochondrial outer membrane induces mitochondrial fragmentation. *Cell Mol Life Sci*. 2010; 67:2839–50.  
<https://doi.org/10.1007/s00018-010-0365-z> PMID:[20372962](https://pubmed.ncbi.nlm.nih.gov/20372962/)
17. Chu PY, Tzeng YT, Chiu YH, Lin HY, Kuo CH, Hou MF, Li CJ. Multi-Omics Reveals the Immunological Role and Prognostic Potential of Mitochondrial Ubiquitin Ligase MARCH5 in Human Breast Cancer. *Biomedicines*. 2021; 9:1329.  
<https://doi.org/10.3390/biomedicines9101329> PMID:[34680446](https://pubmed.ncbi.nlm.nih.gov/34680446/)
18. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017; 45:W98–102.  
<https://doi.org/10.1093/nar/gkx247> PMID:[28407145](https://pubmed.ncbi.nlm.nih.gov/28407145/)
19. Li JY, Li CJ, Lin LT, Tsui KH. Multi-Omics Analysis Identifying Key Biomarkers in Ovarian Cancer. *Cancer Control*. 2020; 27:1073274820976671.  
<https://doi.org/10.1177/1073274820976671> PMID:[33297760](https://pubmed.ncbi.nlm.nih.gov/33297760/)
20. Li CJ, Lin LT, Chu PY, Chiang AJ, Tsai HW, Chiu YH, Huang MS, Wen ZH, Tsui KH. Identification of Novel Biomarkers and Candidate Drug in Ovarian Cancer. *J Pers Med*. 2021; 11:316.  
<https://doi.org/10.3390/jpm11040316> PMID:[33921660](https://pubmed.ncbi.nlm.nih.gov/33921660/)
21. Li CJ, Chiu YH, Chang C, Chang YI, Sheu JJ, Chiang AJ. Acetyl Coenzyme A Synthase 2 Acts as a Prognostic Biomarker Associated with Immune Infiltration in Cervical Squamous Cell Carcinoma. *Cancers (Basel)*. 2021; 13:3125.  
<https://doi.org/10.3390/cancers13133125> PMID:[34206705](https://pubmed.ncbi.nlm.nih.gov/34206705/)
22. Chen X, Qian Y, Wu S. The Warburg effect: evolving interpretations of an established concept. *Free Radic Biol Med*. 2015; 79:253–63.  
<https://doi.org/10.1016/j.freeradbiomed.2014.08.027> PMID:[25277420](https://pubmed.ncbi.nlm.nih.gov/25277420/)
23. Gogvadze V, Orrenius S, Zhivotovsky B. Mitochondria in cancer cells: what is so special about them? *Trends Cell Biol*. 2008; 18:165–73.  
<https://doi.org/10.1016/j.tcb.2008.01.006> PMID:[18296052](https://pubmed.ncbi.nlm.nih.gov/18296052/)
24. Chen H, Chan DC. Mitochondrial Dynamics in Regulating the Unique Phenotypes of Cancer and Stem Cells. *Cell Metab*. 2017; 26:39–48.  
<https://doi.org/10.1016/j.cmet.2017.05.016> PMID:[28648983](https://pubmed.ncbi.nlm.nih.gov/28648983/)
25. Palm W, Thompson CB. Nutrient acquisition strategies of mammalian cells. *Nature*. 2017; 546:234–42.  
<https://doi.org/10.1038/nature22379> PMID:[28593971](https://pubmed.ncbi.nlm.nih.gov/28593971/)
26. Jiang L, Deberardinis RJ. Cancer metabolism: When more is less. *Nature*. 2012; 489:511–2.  
<https://doi.org/10.1038/489511a> PMID:[23018962](https://pubmed.ncbi.nlm.nih.gov/23018962/)
27. Jiang L, Deberardinis R, Boothman DA. The cancer cell ‘energy grid’: TGF-β1 signaling coordinates metabolism for migration. *Mol Cell Oncol*. 2015; 2:e981994.  
<https://doi.org/10.4161/23723556.2014.981994> PMID:[27308459](https://pubmed.ncbi.nlm.nih.gov/27308459/)
28. Jiang L, Xiao L, Sugiura H, Huang X, Ali A, Kuro-o M, Deberardinis RJ, Boothman DA. Metabolic reprogramming during TGFβ1-induced epithelial-to-mesenchymal transition. *Oncogene*. 2015; 34:3908–16.

- <https://doi.org/10.1038/onc.2014.321>  
PMID:25284588
29. Senft D, Ronai ZA. Regulators of mitochondrial dynamics in cancer. *Curr Opin Cell Biol.* 2016; 39:43–52.  
<https://doi.org/10.1016/j.ceb.2016.02.001>  
PMID:26896558
30. Inoue-Yamauchi A, Oda H. Depletion of mitochondrial fission factor DRP1 causes increased apoptosis in human colon cancer cells. *Biochem Biophys Res Commun.* 2012; 421:81–5.  
<https://doi.org/10.1016/j.bbrc.2012.03.118>  
PMID:22487795
31. Qian W, Wang J, Van Houten B. The role of dynamin-related protein 1 in cancer growth: a promising therapeutic target? *Expert Opin Ther Targets.* 2013; 17:997–1001.  
<https://doi.org/10.1517/14728222.2013.823160>  
PMID:23888838
32. Rehman J, Zhang HJ, Toth PT, Zhang Y, Marsboom G, Hong Z, Salgia R, Husain AN, Wietholt C, Archer SL. Inhibition of mitochondrial fission prevents cell cycle progression in lung cancer. *FASEB J.* 2012; 26:2175–86.  
<https://doi.org/10.1096/fj.11-196543> PMID:22321727
33. Zou P, Liu L, Zheng LD, Payne KK, Manjili MH, Idowu MO, Zhang J, Schmelz EM, Cheng Z. Coordinated Upregulation of Mitochondrial Biogenesis and Autophagy in Breast Cancer Cells: The Role of Dynamin Related Protein-1 and Implication for Breast Cancer Treatment. *Oxid Med Cell Longev.* 2016; 2016:4085727.  
<https://doi.org/10.1155/2016/4085727>  
PMID:27746856
34. Losón OC, Song Z, Chen H, Chan DC. Fis1, Mff, MiD49, and MiD51 mediate Drp1 recruitment in mitochondrial fission. *Mol Biol Cell.* 2013; 24:659–67.  
<https://doi.org/10.1091/mbc.E12-10-0721>  
PMID:23283981
35. Wakabayashi J, Zhang Z, Wakabayashi N, Tamura Y, Fukaya M, Kensler TW, Iijima M, Sesaki H. The dynamin-related GTPase Drp1 is required for embryonic and brain development in mice. *J Cell Biol.* 2009; 186:805–16.  
<https://doi.org/10.1083/jcb.200903065>  
PMID:19752021
36. Yoon Y, Krueger EW, Oswald BJ, McNiven MA. The mitochondrial protein hFis1 regulates mitochondrial fission in mammalian cells through an interaction with the dynamin-like protein DLP1. *Mol Cell Biol.* 2003; 23:5409–20.  
<https://doi.org/10.1128/MCB.23.15.5409-5420.2003>  
PMID:12861026
37. Schneeberger M, Dietrich MO, Sebastián D, Imbernón M, Castaño C, Garcia A, Esteban Y, Gonzalez-Franquesa A, Rodríguez IC, Bortolozzi A, Garcia-Roves PM, Gomis R, Nogueiras R, et al. Mitofusin 2 in POMC neurons connects ER stress with leptin resistance and energy imbalance. *Cell.* 2013; 155:172–87.  
<https://doi.org/10.1016/j.cell.2013.09.003>  
PMID:24074867
38. Meng N, Glorieux C, Zhang Y, Liang L, Zeng P, Lu W, Huang P. Oncogenic K-ras Induces Mitochondrial OPA3 Expression to Promote Energy Metabolism in Pancreatic Cancer Cells. *Cancers (Basel).* 2019; 12:65.  
<https://doi.org/10.3390/cancers12010065>  
PMID:31881642