

Eight immune-related genes predict survival outcomes and immune characteristics in breast cancer

Han Xu^{1,*}, Gangjian Wang^{2,*}, Lili Zhu², Hong Liu², Bingjie Li²

¹The Department of Breast Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

²The Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

*Equal contribution

Correspondence to: Bingjie Li; email: bingjie.li@monash.edu

Keywords: breast cancer, immune checkpoint, immune cell, risk score, TCGA

Received: February 2, 2020

Accepted: July 6, 2020

Published: August 3, 2020

Copyright: Xu 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

Advancements in immunotherapy have improved our understanding of the immune characteristics of breast cancer. Here, we analyzed gene expression profiles and clinical data obtained from The Cancer Genome Atlas database to identify genes that were differentially expressed between breast tumor tissues and normal breast tissues. Comparisons with the Immunology Database and Analysis Portal (ImmPort) indicated that many of the identified differentially expressed genes were immune-related. Risk scores calculated based on an eight-gene signature constructed from these immune-related genes predicted both overall survival and relapse-free survival outcomes in breast cancer patients. The predictive value of the eight-gene signature was validated in different breast cancer subtypes using external datasets. Associations between risk score and breast cancer immune characteristics were also identified; *in vitro* experiments using breast cancer cell lines confirmed those associations. Thus, the novel eight-gene signature described here accurately predicts breast cancer survival outcomes as well as immune checkpoint expression and immune cell infiltration processes.

INTRODUCTION

Breast cancer is a life-threatening disease of increasing clinical concern worldwide [1]. It is classified into four molecular subtypes; luminal A, luminal B, triple-negative breast cancer (TNBC), and human epidermal growth factor receptor type 2 (HER2) positive [2, 3]. Treatment methods are very different for each subtype [4]. Endocrine therapy is effective for treating luminal A and B subtypes, which are therefore associated with good prognoses. In addition, HER-2-positive breast cancer is sensitive to chemotherapy and anti-HER-2 therapy. However, prognoses are poor for TNBC because neither endocrine nor anti-HER-2 therapies are effective in this subtype. Thus, effective treatments for TNBC are urgently needed [5]. TNBC is characterized by genomic instability, high mutation load, and high levels of immune infiltration. Some clinical studies have

therefore examined immunotherapy treatments for TNBC in recent years.

Immunotherapies can target both innate and adaptive immune mechanisms in the treatment of breast cancer [5, 6]. Immunotherapy techniques, especially those targeting PD1 and PDL-1, have been considered for use in clinical practice [7]. Although immunotherapy is a promising treatment method for breast cancer, many issues still need to be addressed. Immune evasion is a key problem in breast cancer immunotherapy, and it is further complicated by substantial differences in immune cell infiltration processes and immune response in breast cancer compared to other types of cancer [8, 9]. Additional research is needed to identify immune checkpoints and immune cell infiltration processes that could serve as treatment targets.

Some previously identified immune-related genes and cells with prognostic value in breast cancer patients might be effective immunotherapy targets. For example, the immune-related gene TGFBR2 predicts prognosis in estrogen receptor-negative patients after chemotherapy. Several other genes identified as potential targets for cancer treatment play important roles in immune responses [10–12]. In addition, colocalization of immune and breast cancer cells predicts prognosis in

breast cancer patients [13]. However, an immune-related gene signature that can accurately predict breast cancer survival outcomes and other clinical features would be greatly beneficial.

In this study, we explored whether immune-related genes influence clinical outcomes in breast cancer via immune-related mechanisms (Figure 1). First, we identified 4391 differentially expressed genes (DEGs),

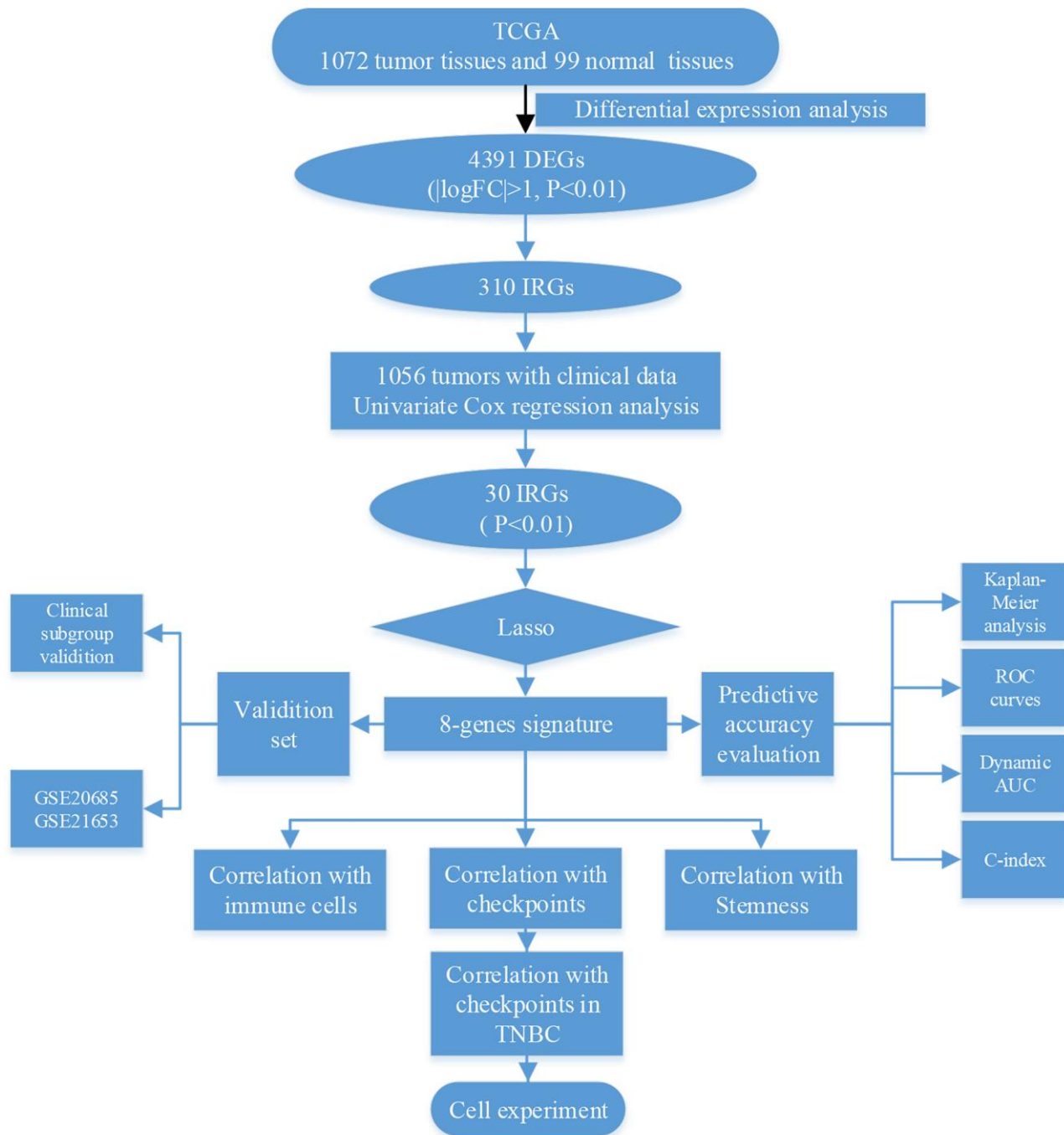


Figure 1. Schematic of research strategy.

of which 310 were immune-related (IRGs), in 1072 breast tumor and 99 normal breast tissues from the TCGA database. Univariate Cox regression analysis of clinical data obtained from 1056 breast cancer patients revealed that the 301 IRGs were statistically significant predictors of survival. Eight of the 301 IRGs were incorporated into a model capable of predicting breast cancer survival outcomes based on Lasso regression analysis. This eight-gene signature was validated in two other breast cancer datasets, and its ability to predict immune checkpoint expression and immune cell

infiltration was confirmed using breast cancer cell lines *in vitro*.

RESULTS

DEGs and IRGs in breast cancer

Gene expression data were downloaded from TCGA, and a total of 4391 differentially expressed genes (DEGs) were identified between breast tumor and normal breast tissues (Figure 2A). Out of these 4391

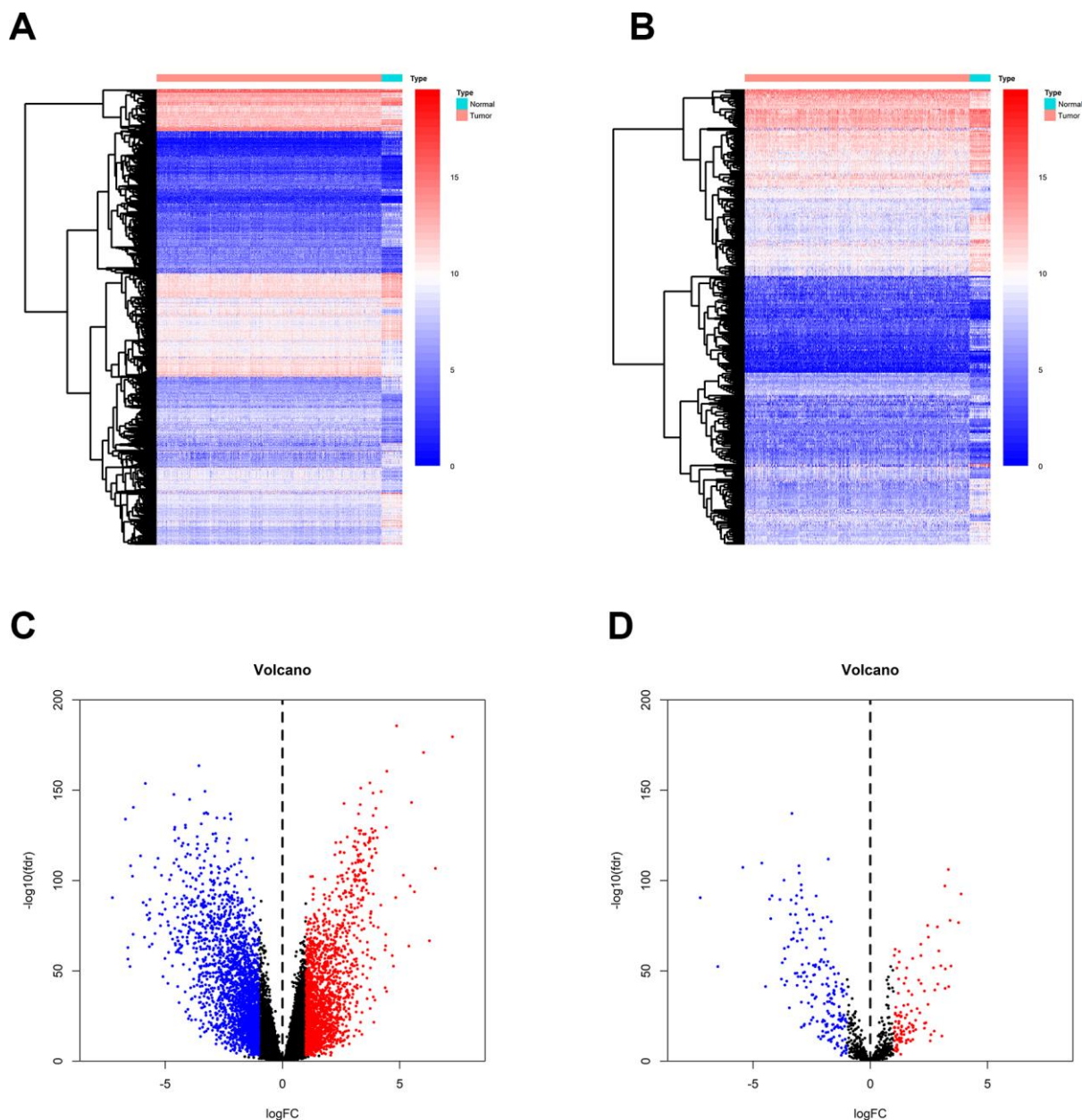


Figure 2. Identification of DEGs and IRGs between breast tumor and normal breast tissues from the TCGA database. (A, C) Heatmap and volcano plot of DEGs between breast tumor and normal breast tissues. **(B, D)** Heat map and volcano plot of IRGs between breast tumor and normal breast tissues.

DEGs, 2042 DEGs were over-expressed and 2349 were under-expressed in breast tumor tissues compared to normal tissues (Figure 2C). Using the ImmPort gene list, 310 of the DEGs were identified as immune-related genes IRGs (Figure 2B); of these, 195 genes were under-expressed and 115 were over-expressed in breast tumor tissues compared to normal tissues.

GO and KEGG enrichment analyses

To further explore their functions, the 310 IRGs that were differentially expressed between tumor and normal tissues were subjected to GO and KEGG enrichment

analyses. GO enrichment analysis indicated that the IRGs were enriched in the following five GO terms: inflammatory response, immune response, response to lipopolysaccharide, chemokine-mediated signaling pathway, and chemotaxis (Figure 3A). These GO terms are associated with immune functions, confirming that these DEGs are immune-related.

KEGG enrichment analysis indicated that the IRGs were enriched in the following five KEGG terms: cytokine-cytokine receptor interaction, chemokine signaling pathway, legionellosis, salmonella infection, and TNF signaling pathway (Figure 3B). These results

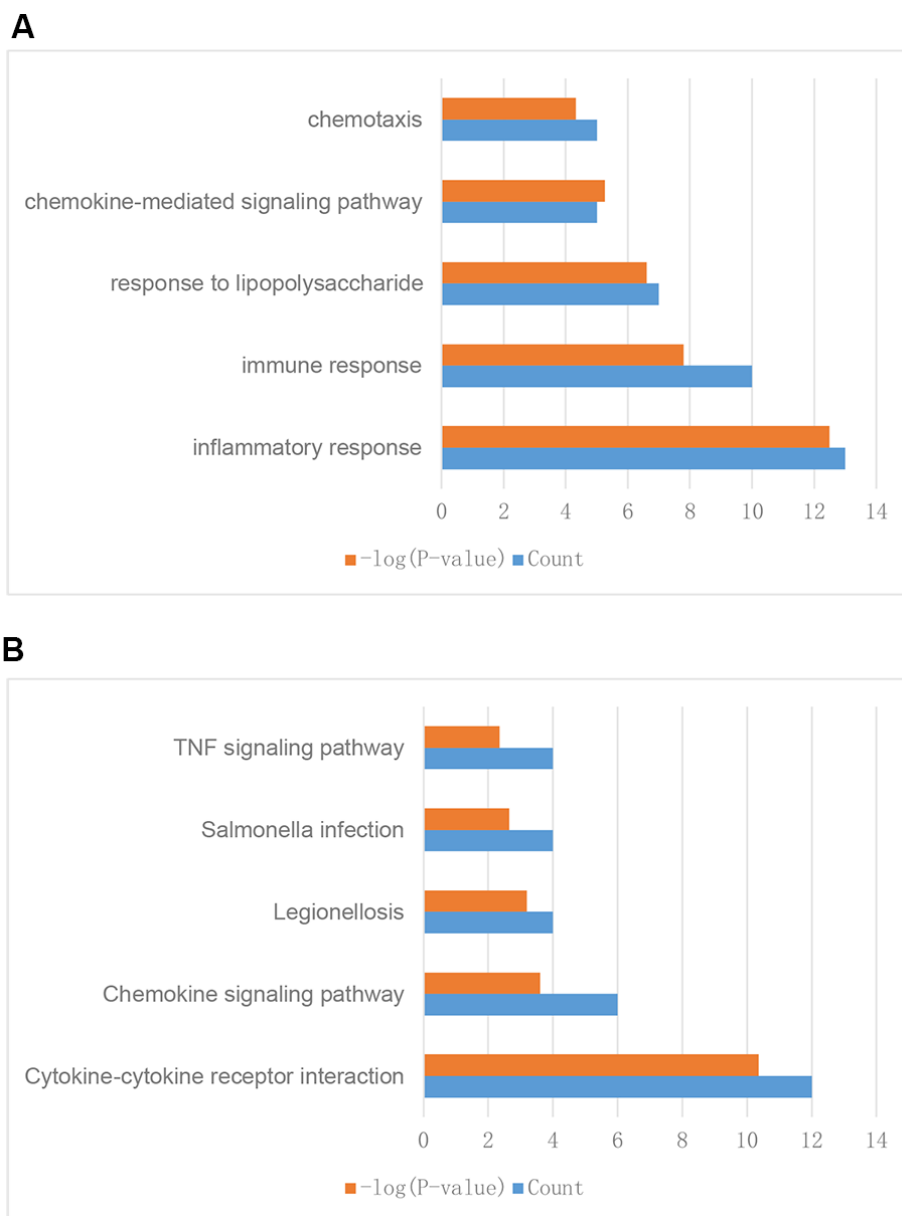


Figure 3. Enrichment analysis of differentially expressed IRGs between breast tumor and normal breast tissues. (A) GO enrichment analysis results. **(B)** KEGG enrichment analysis results.

suggested that these genes might have functions in cell interaction, infection, and other immune related pathways and further validated the GO enrichment analysis results.

Construction of the eight-IRG signature

The analysis process is depicted in Figure 1. As shown in the flow chart, the 310 IRGs were subjected to single factor Cox regression analysis. After considering the statistical significance of associations with OS, five IRGs were selected for further consideration. IRGs that were involved in breast cancer pathogenesis and progression were identified only among the 30 IRGs that were significantly associated with clinical outcomes (Table 1). Finally, using LASSO regression analysis, an eight-IRG signature was constructed in which risk score was calculated using the following formula (Figure 4A):

$$\begin{aligned} \text{risk score} = & (0.1775 \times \text{ULBP2}) - (0.1140 \\ & \times \text{ADRBI}) - (0.0231 \times \text{TSLP}) - (0.0324 \\ & \times \text{MIA}) + (0.1522 \times \text{IL27}) - (0.1013 \times \text{IFNE}) \\ & + (0.1413 \times \text{SCG2}) - (0.1017 \times \text{NR0B1}) \end{aligned}$$

Hazard ratios and expression levels for each of the eight genes as well as risk score distributions are shown in Figure 4B–4E. Breast cancer patients were divided into high-risk and low-risk groups using the median risk score as a cut-off point. OS and RFS were shorter in high-risk patients than in low-risk patients ($p < 0.001$) (Figure 5A and 4C). Time-dependent ROC curves indicated that the AUC for three-year and five-year OS were 0.753 and 0.720, while the AUC for three-year and five-year RFS were 0.643 and 0.603 (Figure 5B, 5D). Incorporation of the important clinical variables age, HER2/ER/PR status, stage, TP53 mutation status, therapy type, and risk score into a multivariate regression analysis revealed that risk score was an independent prognostic factor ($p = 0.003$).

Validation of the eight-gene signature

To further validate the predictive power of our model, we re-evaluated its prediction accuracy in two additional data sets from GEO, GSE20685 and GSE21653. KM curves and survival information revealed significant differences in survival outcomes between high-risk and low-risk patients, confirming the robustness of the eight-IRG signature (Supplementary Figure 1A, 1B).

Evaluating predictive accuracy of survival outcomes in breast cancer patients

To further explore the predictive capacity of the eight-gene signature, we used it to predict survival outcomes

of in different breast cancer patient subgroups (Supplementary Figure 1). Mutations in oncogenes and tumor suppressor genes contribute to malignant behavior in cancer cells [14], and TP53 mutations are very common in breast cancer [15]. Our results revealed a significant difference in survival between high-risk and low-risk patients regardless of TP53 mutation status (Supplementary Figure 1C, 1D). Survival analysis of the breast cancer patients with different disease stages indicated that survival outcomes were significantly worse for both stage I-II and stage III-IV high-risk breast cancer patients than for low-risk breast cancer patients (Supplementary Figure 1E, 1F). This demonstrated that the eight-gene signature could accurately predict survival outcomes in patients with different stages of breast cancer.

Next, we performed survival analysis for breast cancer patients of different pathological subtypes: ER positive or negative, PR positive or negative, and HER2 positive or negative (Supplementary Figure 1E–1L). Survival outcomes were significantly worse for high-risk patients than for low-risk patients regardless of ER and PR status as well as in HER2-negative patients, indicating that the eight-gene signature accurately predicted survival for these pathological types. In addition, although the p value for HER2-positive patients was greater than 0.05, there was an obvious trend towards poorer survival outcomes in high-risk patients compared to low-risk patients of this subtype.

Associations between eight-gene signature and adjuvant therapies

Adjuvant radiotherapy is often an important component of breast cancer treatment [16]. In addition, radiotherapy can not only reduce the risk of breast cancer recurrence, but also improve prognosis [17]. Targeted molecular therapy has also improved the prognosis of early and advanced stage breast cancer patients over the past 15 years [18]. To explore the relationship between the eight-gene signature and these two treatments, we conducted a subgroup analysis of low-risk and high-risk patients based on treatment type. The results showed that targeted molecular therapy had therapeutic benefits only in low-risk patients, while radiotherapy had therapeutic benefits only in high-risk patients (Figure 6).

Associations between eight-gene signature and degree of cancer stemness

Next, we tested associations between the eight-gene signature and levels of G1 phase, high PKH26, and low PKH26 cells in breast cancer patients using single cell sequencing data [19]. An increase in the population of

Table 1. General characteristics of breast cancer-specific immune-related genes.

Gene	logFC	FDR	HR	z-value	p-value
ULBP2	1.334234	1.44E-15	1.213258	3.787657	0.000152
ADRB1	-2.43947	6.59E-25	0.865577	-3.58362	0.000339
TSLP	-3.85142	3.11E-90	0.856824	-3.29241	0.000993
MIA	-2.20894	3.00E-14	0.901339	-3.13967	0.001691
IL27	1.777017	1.51E-37	1.213428	2.945851	0.003221
IFNE	-1.07249	7.13E-13	0.826754	-2.86661	0.004149
IL33	-3.65911	3.67E-63	0.90004	-2.77615	0.005501
ULBP1	2.464382	3.77E-32	1.118432	2.665098	0.007697
RXFP1	1.42523	2.38E-17	1.162835	2.630139	0.008535
SCG2	2.309767	1.41E-39	1.13489	2.616983	0.008871
SDC1	2.064759	1.23E-45	1.220534	2.586434	0.009697
IL17B	-2.70568	1.21E-47	0.87877	-2.57143	0.010128
CXCL1	-1.71462	1.07E-12	0.907549	-2.57002	0.010169
VEGF	3.344907	5.84E-42	1.093142	2.52823	0.011464
CXCL6	-1.64242	3.68E-13	0.901633	-2.49364	0.012644
NROB1	-1.87465	5.36E-20	0.88259	-2.48448	0.012974
CXCL2	-4.28926	2.02E-90	0.90192	-2.46882	0.013556
LIFR	-2.94302	2.24E-98	0.852014	-2.46877	0.013558
JUN	-1.55224	5.11E-40	0.826904	-2.42568	0.01528
LGR6	-2.50792	1.59E-23	0.916671	-2.42018	0.015513
TACR1	-3.49715	5.23E-64	0.903367	-2.40346	0.016241
CXCL14	-1.97421	1.45E-14	0.924163	-2.37137	0.017722
NGFR	-2.72487	1.26E-41	0.900896	-2.33917	0.019327
BMP5	-3.63872	7.19E-43	0.923655	-2.25647	0.024041
CXCL3	-2.30754	2.21E-36	0.900084	-2.20137	0.02771
C3	-1.39465	1.49E-20	0.888334	-2.16518	0.030374
SEMA3G	-2.91638	3.15E-91	0.873563	-2.07498	0.037988
TNFRSF8	-1.24776	6.80E-23	0.86757	-2.07424	0.038057
EDN3	-4.4653	4.90E-42	0.944592	-2.05427	0.039949
CCL23	-1.59865	1.44E-25	0.890331	-2.04987	0.040377

cells in the G1 cell cycle phase indicates increased proliferation of cancer cells. PKH26 is a biomarker of cancer cell proliferation; cell growth rates and cancer sameness are higher in cancer cells with lower PKH26 expression [19]. Our results revealed that G1 phase cell numbers were slightly increased, low PKH26 cell numbers were increased, and high PKH26 cell numbers were decreased in high-risk breast cancer patients (Figure 7A). Furthermore, risk score was positively correlated with low PKH26 cell numbers and negatively correlated with high PKH26 cell numbers (Figure 7B, 7C). These associations between the eight-gene signature and cancer cell stemness are consistent with the ability of the risk score to predict survival outcomes.

Associations between eight-gene signature and immune characteristics

Associations between risk score and breast cancer immune characteristics, such as immune checkpoints and immune cell infiltration, were examined (Figures 8, 9). Expression levels were significantly different for 13 of the 18 immune checkpoints tested between high-risk and low-risk breast cancer patients (Figure 8). Among these 13 immune checkpoints, PD1, PDL2, PDL1, B7H3, CTLA4, IDO1, LAG3, TIM3, CD28, ICOS, OX40, and X41BB were expressed at higher levels, while VSIR expression was lower, in high-risk patients compared to low-risk patients.

The CIBERSORT algorithm was then used to identify eight immune cells for which infiltration differed between high-risk and low-risk patients (Figure 9A). The results revealed that naïve B cells, CD8 T cells, resting CD4 memory T cells, and monocytes showed less infiltration, while T follicular helper cells and M0, M1, and M2 macrophages showed more infiltration, in the high-risk group. In an analysis using the xCell algorithm, which uses more detailed immune cell classifications compared to CIBERSORT, a total of 21 immune cells showed significant differences in

infiltration between low- and high-risk patients (Figure 9B). Five immune cells with the same definition were identified based on two algorithms, including naïve B cells, monocytes, and M0, M1, and M2 macrophages. The infiltration differences were consistent between the algorithms for four of these five immune cells; although infiltration changes for monocytes were inconsistent between the algorithms, these immune cells show very low infiltration levels overall. In general, the eight-gene signature is therefore predictive of changes in immune cell infiltration.

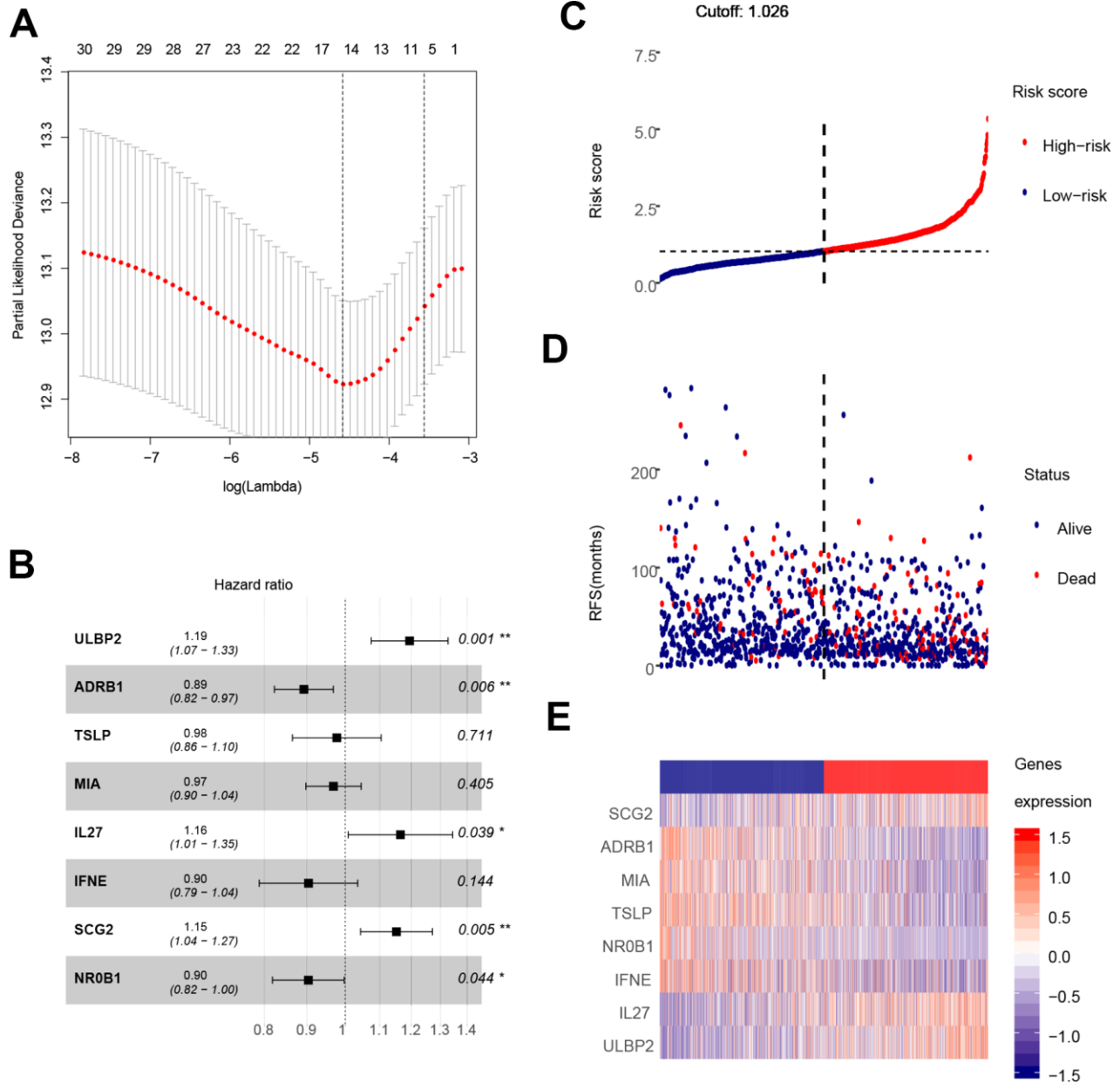


Figure 4. Construction of the eight-gene signature and performance analysis. (A) Construction of the eight-gene signature. (B) Hazard ratio of each gene in the eight-gene signature. (C) Risk score distribution and cut-off point. (D) Distribution of breast cancer patient survival outcomes. (E) Heat map showing expression levels of the eight genes in breast cancer patients.

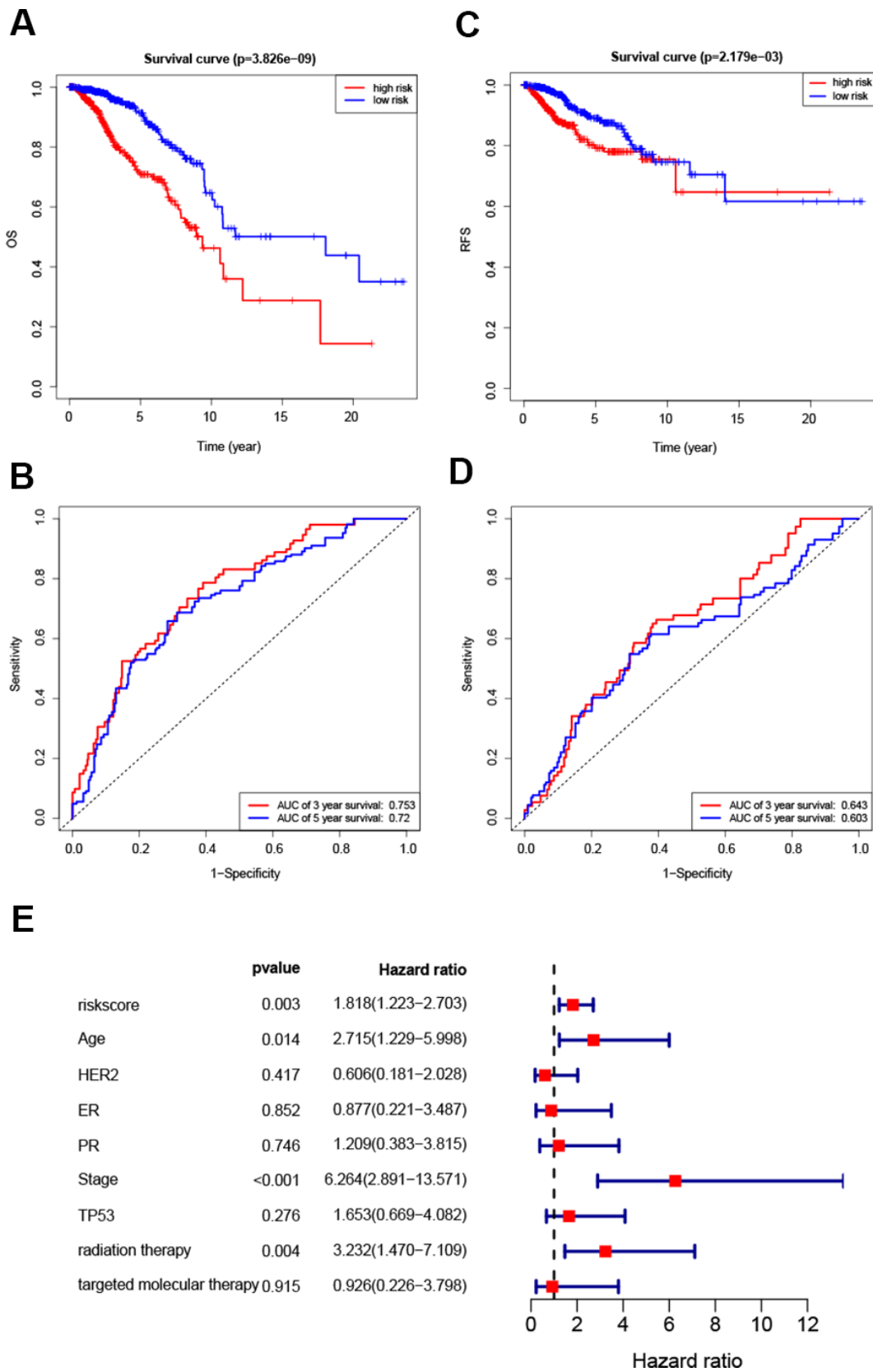


Figure 5. Survival analysis of high- and low-risk breast cancer patients. (A, B) Analysis of OS in high- and low-risk breast cancer patients. (C, D) Analysis of RFS in high- and low-risk breast cancer patients. (E) Hazard ratio of the eight-gene signature and important clinical variables.

Predictive value of the eight-gene signature in TNBC

Additional analyses of both survival outcomes and immune checkpoints were performed in the TNBC subgroup. The results indicated that high-risk non-TNBC patients had poorer survival outcomes than low-risk non-TNBC patients; although the p-value for this comparison was slightly greater than 0.05 in TNBC patients, this

trend would likely have reached statistical significance in a larger group of TNBC patients (Supplementary Figure 2A, 2B). Risk score also predicted expression of several immune checkpoints, including PD1, PDL1, PDL2, TIM3, CD28, ICOS, IL2RB, and 41BB, in TNBC patients, and its predictive value in these patients was similar to that observed in the overall breast cancer patient cohort (Supplementary Figure 2C, 2D).

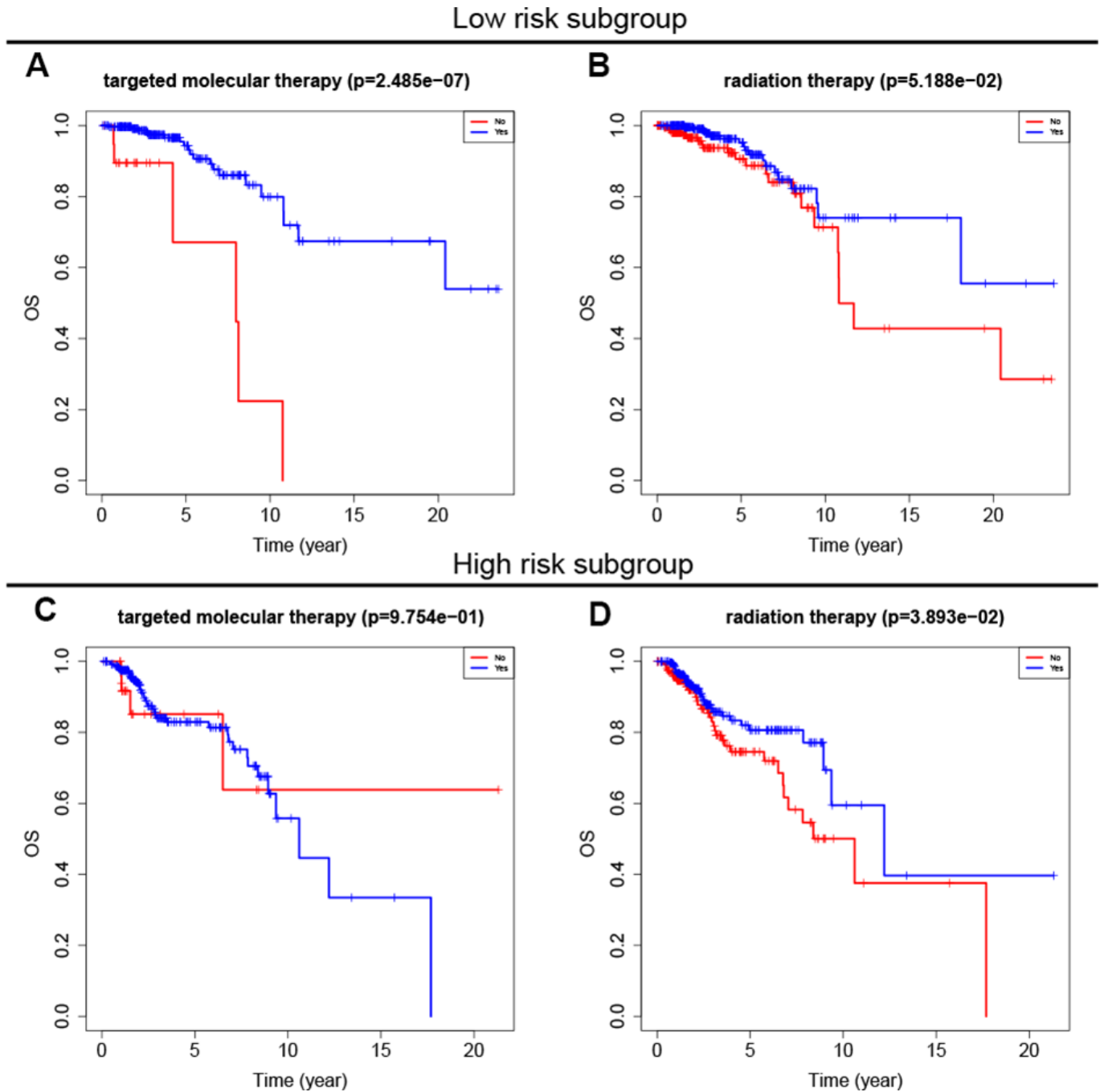


Figure 6. Survival analysis of adjuvant therapy in high- and low-risk patients. (A, B) Survival analysis of targeted molecular therapy and radiation in low-risk patients. (C, D) Survival analysis of targeted molecular therapy and radiation in high-risk patients.

Validation of eight-gene signature *in vitro*

Associations between the eight-gene immune signature and immune checkpoint levels were examined in four breast cancer cell lines (Figure 10). The results indicated that expression of the five immune checkpoints PD1, PDL1, B7H3, LAG-3, and OX40 tended to be lower in cells with higher risk scores. This agrees with our finding that higher risk scores predict higher

expression of four immune checkpoints in breast cancer patients.

DISCUSSION

In this study, we constructed an eight-gene IRG signature that predicted both survival and immune characteristics in breast cancer patients. Enrichment analysis confirmed that these eight genes were involved

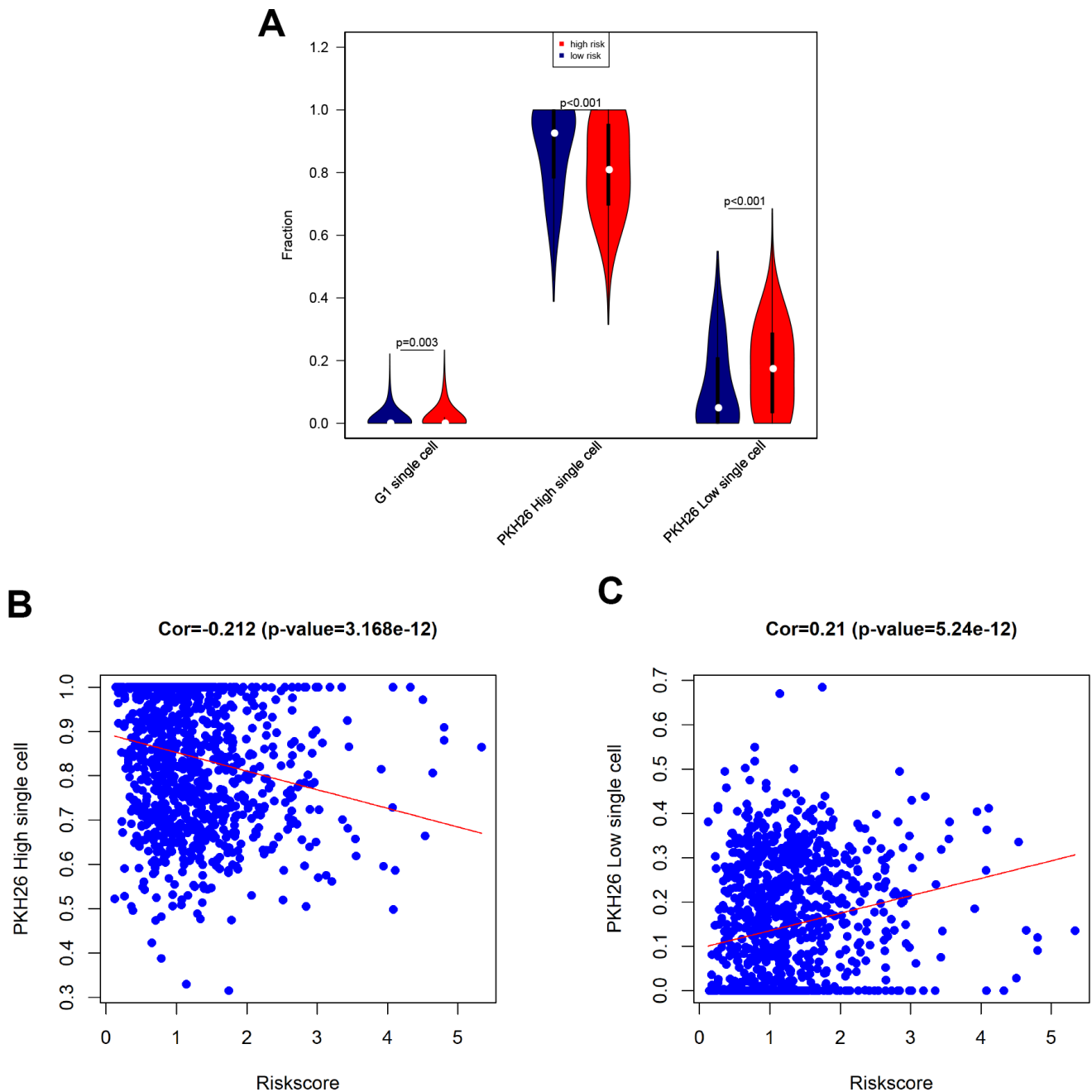


Figure 7. Associations between eight-gene signature and cancer stemness. (A) Infiltration of G1 phase, high PKH26, and low PKH26 single cells between high- and low-risk patients. **(B)** Analysis of associations between risk score and high PKH26 single cells. **(C)** Analysis of associations between risk score and low PKH26 single cells.

in immune responses, suggesting that they function by interacting with immune checkpoints or immune cells. Additionally, the gene signature was validated using datasets containing different types of breast cancers, indicating that it may be broadly applicable for many breast cancer patients.

Previous studies have demonstrated that some of the eight genes comprising our signature play important

roles in cancer pathology and clinical assessment. ULBP2, a ligand of NKG2D, is associated with poor prognosis in a number of human cancers, and surface expression of this protein is often lost in many human cancer cell types during NK cell-mediated cytotoxicity [20–22]. The TSLP signaling pathway interacts with other immune pathways and may promote survival of breast and pancreatic cancer cells, although its effects in breast cancer remain poorly understood [23–25]. The

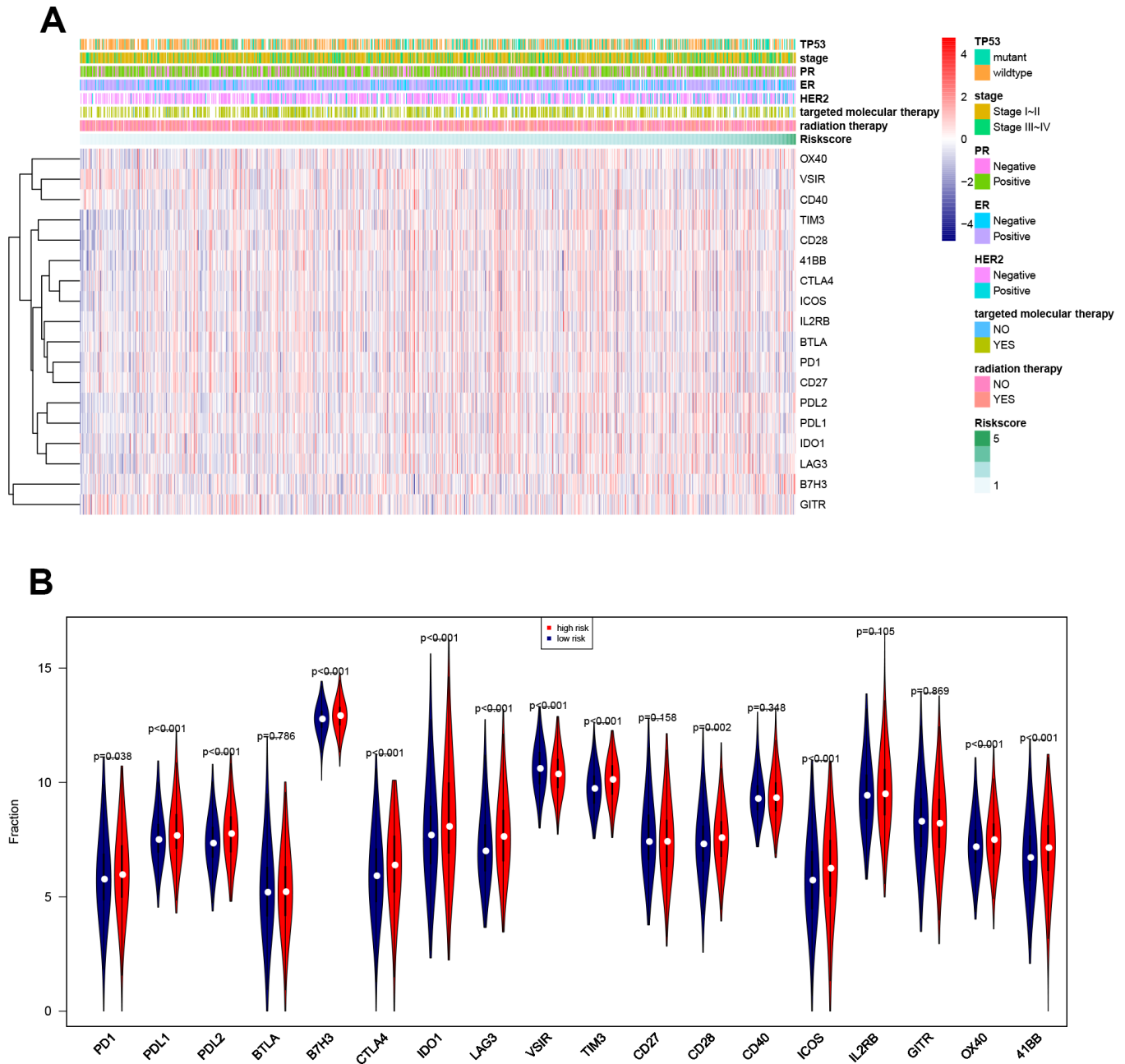


Figure 8. Associations between eight-gene signature and 18 immune checkpoints. (A) Heatmap showing associations between risk score, clinical variables, and 18 immune checkpoints. **(B)** Predictive value of the eight-gene signature for PD1, PDL2, PDL1, B7H3, CTLA4, IDO1, LAG3, VSIR, TIM3, CD28, ICOS, OX40, and X41BB.

MIA gene family is considered a useful marker for many types of cancers, and its upregulation has been associated with shorter progression free survival times [26, 27]. Here, we found that upregulation of TSLP and MIA was associated with better survival outcomes, which contradicts previous studies indicating that these genes interact with other pathways to promote proliferation and growth of breast cancer cells. IL27 has been identified as a potentially useful target for anti-

cancer clinical applications, probably due to its ability to regulate CD8+ T cells, natural killer cells, macrophages, and other immune checkpoints [28, 29]; our present findings regarding IL27 are consistent with these prior studies. NROB1 sensitizes lung cancer cells to chemotherapy and inhibits their invasive abilities [30, 31]; similar effects might explain the poorer survival outcomes observed here in breast cancer patients upon downregulation of this gene.

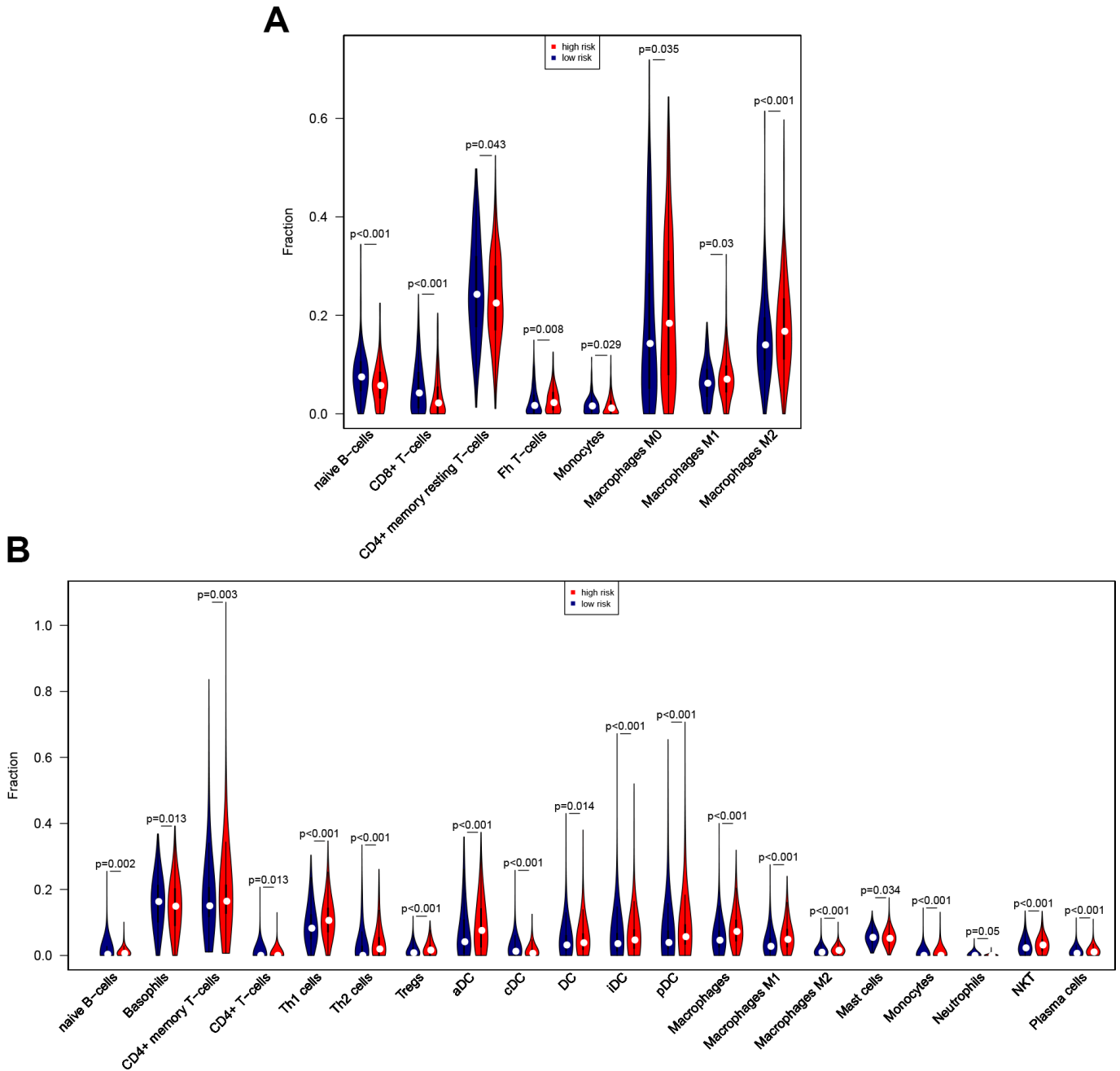


Figure 9. Associations between eight-gene signature and immune cell infiltration. (A) Predictive value of the eight-gene signature for 8 immune cells based on the CIBERSORT algorithm. **(B)** Predictive value of the eight-gene signature for 21 immune cells based on the xCell algorithm.

Although some of the genes included in our signature may not play important roles in clinical cancer pathology and assessment, they do play roles in other diseases. ADRB1 is implicated in cognitive neural diseases, possibly due to its role in neuroinflammatory processes, and is also associated with heart failure [32–34]. These effects might also have contributed to the survival outcomes seen here. IFNE is a member of the interferon family that inhibits proliferation in various cells by regulating NK cells and the JAK-STAT

pathway [35]. Upregulation of this gene in the present study might therefore contribute to improved survival outcomes. SCG2 may be a biomarker of bipolar disease and is known to regulate hypertension in humans [36], perhaps explaining the poorer survival outcomes observed here following upregulation of this gene.

In this study, we also identified pathways through which the eight IRGs might affect breast cancer outcomes. Several proteins from the TNF signaling

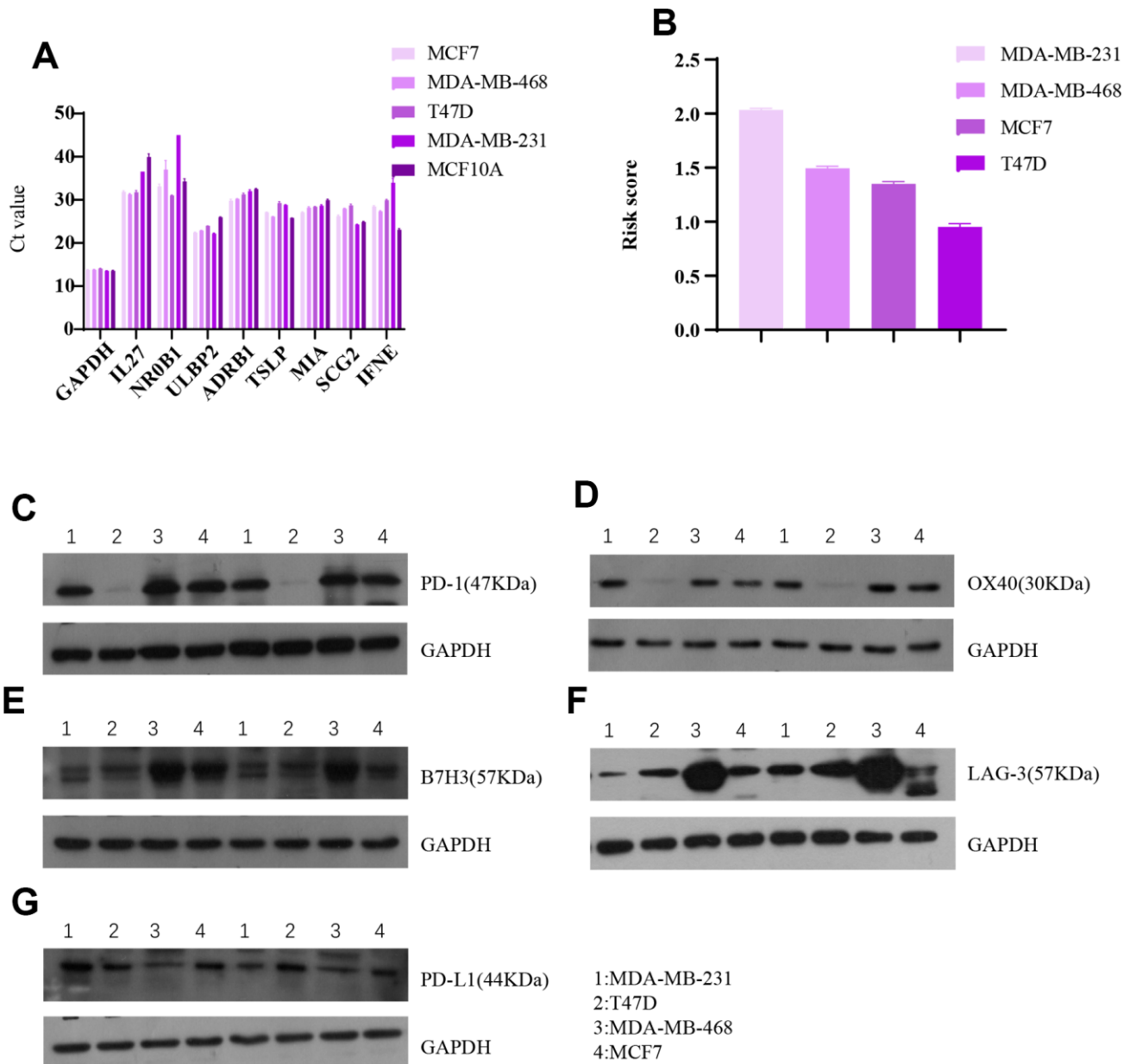


Figure 10. *In vitro* breast cancer cell line experiments validate the predictive value of the eight-gene signature for immune checkpoints. (A) Raw data (Ct value) from Rt-PCR analysis of the eight genes in four breast cancer cell lines and one normal breast cell line. **(B)** Risk scores of the four breast cell lines calculated based on Rt-PCR results. Western blot results for **(C)** PD1, **(D)** OX40, **(E)** B7H3, **(F)** LAG-3, and **(G)** PDL1 expression in the four breast cancer cell lines.

pathway play important roles in breast cancer and its treatment [37]. For example, TNF- α is implicated in immune responses in breast cancer and might therefore serve as a treatment target in triple negative breast cancer [38, 39]. Moreover, chemokine signaling pathway members, especially CCL5, are involved in the pathogenesis and development of breast cancer [40–42]. Although not all of the genes included in the enrichment analysis were incorporated into the eight-IRG signature, they were differentially expressed in breast tumor and normal tissues. Because the signature was constructed from these genes, the pathways in which they are enriched are relevant to the eight-IRG signature and may represent important differences between the two tissue types. These pathways might therefore reveal biological processes responsible for the immune functions of these genes as well as potential mechanisms that contribute to survival outcomes in breast cancer patients.

The eight-gene signature was capable of predicting both a number of immune checkpoints which may serve as biomarkers and the infiltration of immune cells that can act as therapeutic targets in breast cancer. Previous studies strongly support the use of PD1 and PDL1 as targets for breast cancer treatment [43, 44]. Overexpression of CTLA4 can increase numbers of Treg cells and thereby influence breast cancer pathogenesis and development [45]. In addition, tumor-associated macrophages are associated with poor prognosis in breast cancer patients [46], which is consistent with our present finding that increased infiltration of M0, M1, and M2 macrophages was associated with poorer survival outcomes. Evidence also suggests that CD4 and CD8 T cells can act as biomarkers and therapeutic targets for breast cancer treatment [47, 48], which is in keeping with our finding that higher risk scores based on the eight-IRG signature were associated with higher levels of CD8 and resting CD4 memory T cells. Since these immune checkpoints and cells can serve as targets for immunotherapy [49], the ability of our eight-IRG signature to predict these immune characteristics might prove valuable in the clinical setting.

Programmed death receptor 1 (PD1), which is mainly expressed in activated T lymphocytes and myeloid cells, and its ligand PD-L1 are important immunosuppressive molecules [50, 51]. The binding of PD1 to its ligand can inactivate T cells, leading to immune escape reactions in tumors [50]. Single or combined drug therapies using immune checkpoint inhibitors (ICIs) play anti-tumor roles by blocking the transmission of immunosuppressive signals, reactivating the immune response of T cells to tumors, and restoring immune activity in the tumor microenvironment [52]. The advent of immunotherapy has changed treatment regimens for many tumors, including breast cancer, and clinical trials

of TNBC inhibitors have yielded encouraging results. Higher risk scores based on our eight-IRG model tended to be associated with poorer survival outcomes in TNBC group, indicating that this gene signature might help predict prognosis in TNBC patients.

In conclusion, we developed an eight-gene signature using IRGs that were differentially expressed between breast tumor and normal breast tissues. This signature predicted breast cancer survival outcomes for various pathological types at different clinical stages. The genes included in the signature were also associated with immune checkpoint expression and immune cell infiltration. Our eight-gene signature therefore accurately predicted both immune characteristics and survival outcomes in breast cancer patients.

MATERIALS AND METHODS

Accessing gene expression data from TCGA

TCGA is a cancer gene expression database accessible to all cancer researchers and clinicians. We downloaded clinical data for 1056 breast cancer patients and mRNA level gene expression data for 1072 breast tumor tissues and 99 normal breast tissues from the TCGA database. Clinical data was reordered and is summarized in Table 2. Because TCGA is open access and contains publicly available, ethical approval is not required before use.

Differential expression analysis and identification of IRGs

ImmPort is a publicly available database accessible to professionals specializing in immunology [53]. Differential expression analysis of RNA-Seq data from the 1072 breast tumor and 99 normal breast tissues was conducted using the R package “limma” from Bioconductor [54]. We applied $|\log_{2}FC| < 1$ and $P < 0.01$ as the criteria to identify DEGs and determined which DEGs were also IRGs based on the IRG list downloaded from ImmPort.

Enrichment analysis

GO enrichment analysis was performed to identify biological functions, while KEGG enrichment analysis was used to identify both biological functions and pathways, associated with the DEGs [55, 56]. The R package “ClusterProfiler” was used for both enrichment analyses [57].

Construction and validation of the IRG signature

Cox regression is a widely used tool for survival analysis [58]. Based on differential expression data,

Table 2. Clinical characteristics of the patients in TCGA data sets.

Characteristics	Number(%)
Age	
<=60	586(55.5)
>60	470(44.5)
HER2	
Positive	108(14.7)
Negative	625(85.3)
ER	
Positive	777(77.0)
Negative	232(23.0)
PR	
Positive	676(67.2)
Negative	330(32.8)
Stage	
I~II	773(74.7)
III~IV	262(25.3)
TP53 status	
Wildtype	513(67.0)
mutant	253(33.0)
Radiation	
Yes	540(55.6)
No	432(44.4)
Targeted molecular therapy	
Yes	516(91.7)
No	47(8.3)
Survival status	
Survival	907(85.9)
Dead	149(14.1)
Relapse status	
Relapse-free	793(89.6)
Relapse	92(10.4)

30 IRGs were identified that contributed to survival outcomes in the 1056 breast cancer patients for which clinical data was available. Least Absolute Shrinkage and Selection Operator (Lasso) regression is a useful method for weighting model parameters and helps identify the most important variables to generate the best predictive model. The “glmnet” R package was used to carry out the LASSO Cox regression analysis [59]. This analysis identified an eight-gene signature that we used to construct a model that predicted both immune characteristics and clinical outcomes in breast cancer patients. Risk scores were calculated for each sample based on coefficients assigned to each prognostic IRG in the signature. The median risk score was used as a cut-off value for dividing training and validation group patients into high- and low-risk groups.

Performance analysis

The Kaplan-Meier (KM) survival curve is a powerful tool for analyzing patient survival outcomes [60]. In this study, the R package “survival” was used to generate the KM survival curve. A Receiver Operating Characteristic (ROC) curve is often used to evaluate the sensitivity and specificity of a model in predicting outcome events [61]. We used the R package “survival ROC” to conduct ROC analysis. Using the median risk score as a cut-off and plotting clinical outcome data for breast cancer patients against their risk scores, we generated an ROC curve and calculated area under curve (AUC) values for both 3- and 5-year survival. An AUC value between 0.5 and 0.7 indicates evidence of a successful model, values between 0.7 and 0.9 indicate

strong evidence of a successful model, and values greater than 0.9 indicate very strong evidence of a successful model.

Validation using the GEO database

The Gene Expression Omnibus (GEO) database is an open access database containing datasets from published projects [62, 63]. In our study, OS data for 328 patients from the GSE20685 dataset and RFS data for 249 patients from the GSE21653 dataset were used as validation groups [64–66].

Analysis of cancer stemness using single cell sequencing data

Single cell sequencing is a new method for generating sequencing profiles for specific cell types [67]. GSE124989 includes single cell sequencing data from three breast cancer cell subtypes, enabling analysis of degree of stemness in breast cancer cells [19]. The CIBERSORTX algorithm generates signatures from single cell sequencing data that allow the calculation of numbers of individual cell types from bulk RNA sequencing data [68]. We used the CIBERSORTX algorithm to analyze GSE124989 data and constructed signatures for the following breast cancer cell subtypes: G1 phase, high PKH26, and low PKH26 single cells. We then used the signatures of these three cell types to evaluate cancer stemness in breast cancer patients.

Evaluation of predictive accuracy among different clinical stages and pathological types of breast cancer

Clinical stage and pathological type are important factors that influence clinical decisions. We therefore tested the ability of the eight-gene signature to predict the survival outcomes in patients with different clinical stages and pathological types of breast cancer. Clinical stages were grouped as stage I-II and stage III-IV, while the pathological types were classified as ER positive or negative, PR positive or negative, and HER2 positive or negative. KM survival analysis was used to analyze clinical outcomes in the different breast cancer patient subgroups.

Associations between eight-gene signature and immune characteristics

Correlation analysis was then conducted to explore the eight-gene signature's ability to predict immune checkpoint expression and immune cell infiltration. Breast cancer patients were divided into high- and low-risk groups based on the cut-off risk score value before analyzing associations with immune characteristics.

We analyzed the correlation between the eight-gene signature and the expression of the 18 immune checkpoints identified as existing or potential targets for cancer immunotherapy. T-tests were used to compare the mean immune checkpoint expression levels between high- and low-risk patients.

The CIBERSORT algorithm is used to estimate the proportion of specific cell types based on bulk gene expression data [68]. LM22 is a leukocyte gene signature comprised of 547 genes that distinguishes 22 human immune cell subsets with high accuracy. The XCell algorithm calculates infiltration of 64 immune cells from based on RNA-seq data [69]. Using these algorithms, we evaluated amounts of immune cells belonging to these 22 and 64 immune cell subsets in each sample at a significance level of $p < 0.05$; only samples that exceeded that threshold were included in our study. Infiltration of the 22 and 64 immune cell subsets was compared between high- and low-risk breast cancer patients using t-tests.

***In vivo* validation of results**

In vitro experiments using MCF7, MDA-MB-468, T47D, and MDA-MB-231 breast cancer cell lines as well as the MCF10A normal breast cell line as external reference (all from Genechem, Shanghai) were conducted to further validate the immune-checkpoint prediction accuracy of the eight-gene signature. Real-time quantitative PCR (Rt-PCR) was performed to quantify expression of the eight genes in the signature using GAPDH as an internal reference gene (Table 3). Relative RNA expression levels for genes in the signature were calculated via the $2^{-\Delta\Delta C_t}$ method. Promega M-MLV and Trizol (Pufei, Shanghai) kits were used in this experiment, and primers were obtained from Ribobio (Guangzhou).

The R package “sva” was used for batch normalization of the Rt-PCR results with TCGA data. Risk scores were then calculated for each breast cancer cell line. Protein expression levels for the immune checkpoints PD1, B7H3, LAG-3, OX40, and PDL1, and the internal reference GAPDH in the breast cancer cell lines were assessed by western blot, and associations with risk score were examined. Antibodies for these proteins were purchased from Abcam (Shanghai).

Statistical analysis

Statistical analyses were conducted using the R program from the R project for statistical computing (<https://www.r-project.org/>) and SPSS. The R package “pheatmap” was used to plot heatmaps, and “ggpubr” was used to generate boxplots. Unless otherwise

Table 3. Primer sequences used in qPCR in the cell experiment.

Genes	Upstream primer sequence	Downstream primer sequence	Amplified fragment size (bp)
GAPDH	TGACTTCAACAGCGACACCCA	CACCCTGTTGCTGTAGCCAAA	121
ULBP2	CCGCTACCAAGATCCTTCTG	GGATGACGGTGATGTCATAGC	109
ADRB1	TCTCGGCCCTGGTGTCTT	GCCCGGTTGGTGACGAAGT	115
TSLP	CTAACCTTCAATCCCACCG	CTGAGTTTCCGAATAGCCT	108
MIA	CGAAGTTTGGGACTGGTTTAG	GGCAGACAGCAAGATGATGAC	179
IL27	CGCTTTGCGGAATCTCACC	AGGGCATGGAAGGGCTGAA	158
IFNE	AGCCGATGTCTGTTCTTTGTG	CCTCGGGCTTCTAAACTCTGT	108
SCG2	CTGAAGCAAAGACCCACTG	TGTACTCAAAGCCCTGAT	180
NR0B1	CCAAGGAGTACGCCTACCTCA	CATTTCCAGCATCATATCATCCA	272

indicated, $p < 0.05$ was used to indicate statistical significance.

Abbreviations

TCGA: The Cancer Genome Atlas; OS: overall survival; RFS: relapse-free survival; TNBC: triple-negative breast cancer; HER2: human epidermal growth factor receptor type 2; DEG: differentially expressed gene; IRG: immune-related gene; LASSO: Least Absolute Shrinkage and Selection Operator; ROC: Receiver Operating Characteristic; AUC: Area Under the Curve; GEO: Gene Expression Omnibus; Rt-PCR: Real-time quantitative PCR.

AUTHOR CONTRIBUTIONS

Han Xu: Conceptualization, supervision, writing original draft, methodology, data curation, formal analysis, manuscript review and editing. Gangjian Wang: Data curation, resources, manuscript review and editing, methodology, formal analysis. Lili Zhu: Methodology, resources, manuscript review and editing. Hong Liu: Methodology, resources, manuscript review and editing, fund acquisition. Bingjie Li: Conceptualization, supervision, writing original draft, formal analysis, manuscript review and editing.

ACKNOWLEDGMENTS

The results of this study are partially based on data available at the TCGA Research Network: <https://www.cancer.gov/tcga>.

CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

FUNDING

This study was funded by the Education Department of Henan Province (20A320043). The funders did not play

a role in research design, data collection and analysis, *in vitro* experiments, or manuscript preparation.

REFERENCES

- DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin.* 2017; 67:439–48. <https://doi.org/10.3322/caac.21412> PMID:28972651
- Yeo SK, Guan JL. Breast cancer: multiple subtypes within a tumor? *Trends Cancer.* 2017; 3:753–60. <https://doi.org/10.1016/j.trecan.2017.09.001> PMID:29120751
- O'Reilly EA, Gubbins L, Sharma S, Tully R, Guang MH, Weiner-Gorzel K, McCaffrey J, Harrison M, Furlong F, Kell M, McCann A. The fate of chemoresistance in triple negative breast cancer (TNBC). *BBA Clin.* 2015; 3:257–75. <https://doi.org/10.1016/j.bbacli.2015.03.003> PMID:26676166
- Zhang S, Wang J, Ghoshal T, Wilkins D, Mo YY, Chen Y, Zhou Y. lncRNA gene signatures for prediction of breast cancer intrinsic subtypes and prognosis. *Genes (Basel).* 2018; 9:65. <https://doi.org/10.3390/genes9020065> PMID:29373522
- Law AM, Lim E, Ormandy CJ, Gallego-Ortega D. The innate and adaptive infiltrating immune systems as targets for breast cancer immunotherapy. *Endocr Relat Cancer.* 2017; 24:X1. <https://doi.org/10.1530/ERC-16-0404e> PMID:28566432
- Varn FS, Mullins DW, Arias-Pulido H, Fiering S, Cheng C. Adaptive immunity programmes in breast cancer. *Immunology.* 2017; 150:25–34. <https://doi.org/10.1111/imm.12664> PMID:27564847
- Vonderheide RH, Domchek SM, Clark AS. Immunotherapy for breast cancer: what are we missing? *Clin Cancer Res.* 2017; 23:2640–46.

- <https://doi.org/10.1158/1078-0432.CCR-16-2569>
PMID:[28572258](https://pubmed.ncbi.nlm.nih.gov/28572258/)
8. Bates JP, Derakhshandeh R, Jones L, Webb TJ. Mechanisms of immune evasion in breast cancer. *BMC Cancer*. 2018; 18:556.
<https://doi.org/10.1186/s12885-018-4441-3>
PMID:[29751789](https://pubmed.ncbi.nlm.nih.gov/29751789/)
9. Emens LA. Breast cancer immunotherapy: facts and hopes. *Clin Cancer Res*. 2018; 24:511–20.
<https://doi.org/10.1158/1078-0432.CCR-16-3001>
PMID:[28801472](https://pubmed.ncbi.nlm.nih.gov/28801472/)
10. Lei J, Rudolph A, Moysich KB, Rafiq S, Behrens S, Goode EL, Pharoah PP, Seibold P, Fasching PA, Andrulis IL, Kristensen VN, Couch FJ, Hamann U, et al, and kConFab Investigators. Assessment of variation in immunosuppressive pathway genes reveals TGFBR2 to be associated with prognosis of estrogen receptor-negative breast cancer after chemotherapy. *Breast Cancer Res*. 2015; 17:18.
<https://doi.org/10.1186/s13058-015-0522-2>
PMID:[25849327](https://pubmed.ncbi.nlm.nih.gov/25849327/)
11. Fan P, Maximov PY, Curpan RF, Abderrahman B, Jordan VC. The molecular, cellular and clinical consequences of targeting the estrogen receptor following estrogen deprivation therapy. *Mol Cell Endocrinol*. 2015; 418:245–63.
<https://doi.org/10.1016/j.mce.2015.06.004>
PMID:[26052034](https://pubmed.ncbi.nlm.nih.gov/26052034/)
12. Huang S, Murphy L, Xu W. Genes and functions from breast cancer signatures. *BMC Cancer*. 2018; 18:473.
<https://doi.org/10.1186/s12885-018-4388-4>
PMID:[29699511](https://pubmed.ncbi.nlm.nih.gov/29699511/)
13. Maley CC, Koelble K, Natrajan R, Aktipis A, Yuan Y. An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. *Breast Cancer Res*. 2015; 17:131.
<https://doi.org/10.1186/s13058-015-0638-4>
PMID:[26395345](https://pubmed.ncbi.nlm.nih.gov/26395345/)
14. 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/)
15. Zhang G, Wang Y, Chen B, Guo L, Cao L, Ren C, Wen L, Li K, Jia M, Li C, Mok H, Chen X, Wei G, et al. Characterization of frequently mutated cancer genes in Chinese breast tumors: a comparison of Chinese and TCGA cohorts. *Ann Transl Med*. 2019; 7:179.
<https://doi.org/10.21037/atm.2019.04.23>
PMID:[31168460](https://pubmed.ncbi.nlm.nih.gov/31168460/)
16. Horton JK, Jagsi R, Woodward WA, Ho A. Breast cancer biology: clinical implications for breast radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018; 100:23–37.
<https://doi.org/10.1016/j.ijrobp.2017.08.025>
PMID:[29254776](https://pubmed.ncbi.nlm.nih.gov/29254776/)
17. Boyages J. Radiation therapy and early breast cancer: current controversies. *Med J Aust*. 2017; 207:216–22.
<https://doi.org/10.5694/mja16.01020> PMID:[28987136](https://pubmed.ncbi.nlm.nih.gov/28987136/)
18. Pondé N, Wildiers H, Awada A, de Azambuja E, Deliens C, Lago LD. Targeted therapy for breast cancer in older patients. *J Geriatr Oncol*. 2020; 11:380–88.
<https://doi.org/10.1016/j.jgo.2019.05.012>
PMID:[31171494](https://pubmed.ncbi.nlm.nih.gov/31171494/)
19. Jonasson E, Ghannoum S, Persson E, Karlsson J, Kroneis T, Larsson E, Landberg G, Ståhlberg A. Identification of breast cancer stem cell related genes using functional cellular assays combined with single-cell RNA sequencing in MDA-MB-231 cells. *Front Genet*. 2019; 10:500.
<https://doi.org/10.3389/fgene.2019.00500>
PMID:[31191614](https://pubmed.ncbi.nlm.nih.gov/31191614/)
20. Wang R, Sun PD. Natural killer cell-mediated shedding of ULBP2. *PLoS One*. 2014; 9:e91133.
<https://doi.org/10.1371/journal.pone.0091133>
PMID:[24614922](https://pubmed.ncbi.nlm.nih.gov/24614922/)
21. Shi L, Lin H, Li G, Sun Y, Shen J, Xu J, Lin C, Yeh S, Cai X, Chang C. Cisplatin enhances NK cells immunotherapy efficacy to suppress HCC progression via altering the androgen receptor (AR)-ULBP2 signals. *Cancer Lett*. 2016; 373:45–56.
<https://doi.org/10.1016/j.canlet.2016.01.017>
PMID:[26805759](https://pubmed.ncbi.nlm.nih.gov/26805759/)
22. Fernández-Messina L, Ashiru O, Agüera-González S, Reyburn HT, Valés-Gómez M. The human NKG2D ligand ULBP2 can be expressed at the cell surface with or without a GPI anchor and both forms can activate NK cells. *J Cell Sci*. 2011; 124:321–27.
<https://doi.org/10.1242/jcs.076042>
PMID:[21224393](https://pubmed.ncbi.nlm.nih.gov/21224393/)
23. Ziegler SF, Roan F, Bell BD, Stoklasek TA, Kitajima M, Han H. The biology of thymic stromal lymphopoietin (TSLP). *Adv Pharmacol*. 2013; 66:129–55.
<https://doi.org/10.1016/B978-0-12-404717-4.00004-4>
PMID:[23433457](https://pubmed.ncbi.nlm.nih.gov/23433457/)
24. Kuan EL, Ziegler SF. A tumor-myeloid cell axis, mediated via the cytokines IL-1 α and TSLP, promotes the progression of breast cancer. *Nat Immunol*. 2018; 19:366–74.
<https://doi.org/10.1038/s41590-018-0066-6>
PMID:[29556001](https://pubmed.ncbi.nlm.nih.gov/29556001/)
25. Ghirelli C, Sadacca B, Reyal F, Zollinger R, Michea P, Sirven P, Pattarini L, Martínez-Cingolani C, Guillot-Delost M, Nicolas A, Scholer-Dahirel A, Soumelis V. No evidence for TSLP pathway activity in human breast cancer. *Oncoimmunology*. 2016; 5:e1178438.

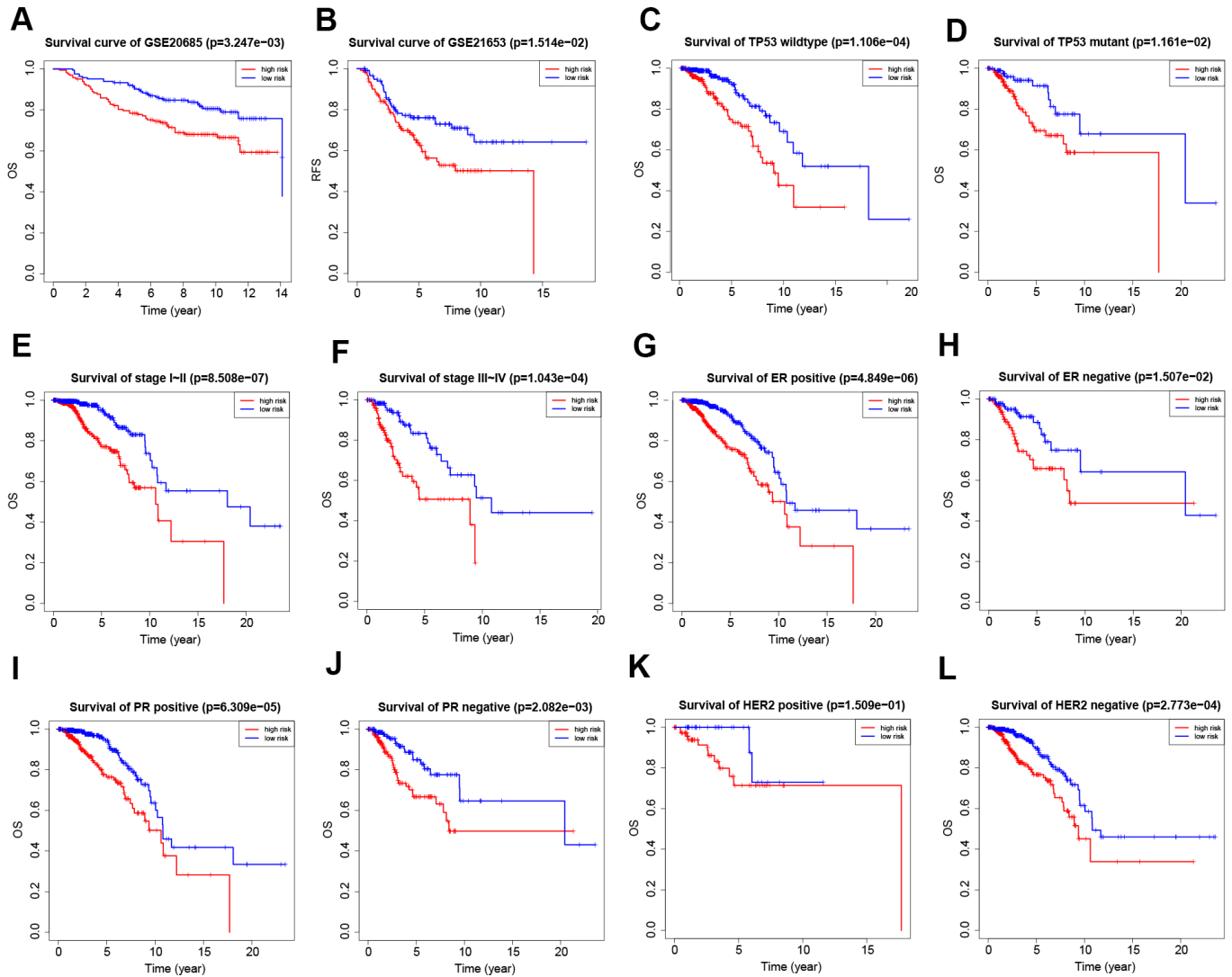
- <https://doi.org/10.1080/2162402X.2016.1178438>
PMID:[27622057](https://pubmed.ncbi.nlm.nih.gov/27622057/)
26. Sasahira T, Kirita T, Nishiguchi Y, Kurihara M, Nakashima C, Bosserhoff AK, Kuniyasu H. A comprehensive expression analysis of the MIA gene family in Malignancies: MIA gene family members are novel, useful markers of esophageal, lung, and cervical squamous cell carcinoma. *Oncotarget*. 2016; 7:31137–52.
<https://doi.org/10.18632/oncotarget.9082>
PMID:[27145272](https://pubmed.ncbi.nlm.nih.gov/27145272/)
27. Schmidt J, Riechers A, Stoll R, Amann T, Fink F, Spruss T, Gronwald W, König B, Hellerbrand C, Bosserhoff AK. Targeting melanoma metastasis and immunosuppression with a new mode of melanoma inhibitory activity (MIA) protein inhibition. *PLoS One*. 2012; 7:e37941.
<https://doi.org/10.1371/journal.pone.0037941>
PMID:[22666418](https://pubmed.ncbi.nlm.nih.gov/22666418/)
28. Yoshimoto T, Chiba Y, Furusawa J, Xu M, Tsunoda R, Higuchi K, Mizoguchi I. Potential clinical application of interleukin-27 as an antitumor agent. *Cancer Sci*. 2015; 106:1103–10.
<https://doi.org/10.1111/cas.12731>
PMID:[26132605](https://pubmed.ncbi.nlm.nih.gov/26132605/)
29. Liao KL, Bai XF, Friedman A. Mathematical modeling of interleukin-27 induction of anti-tumor T cells response. *PLoS One*. 2014; 9:e91844.
<https://doi.org/10.1371/journal.pone.0091844>
PMID:[24633175](https://pubmed.ncbi.nlm.nih.gov/24633175/)
30. Oda T, Tian T, Inoue M, Ikeda J, Qiu Y, Okumura M, Aozasa K, Morii E. Tumorigenic role of orphan nuclear receptor NROB1 in lung adenocarcinoma. *Am J Pathol*. 2009; 175:1235–45.
<https://doi.org/10.2353/ajpath.2009.090010>
PMID:[19644015](https://pubmed.ncbi.nlm.nih.gov/19644015/)
31. Lu Y, Liu Y, Liao S, Tu W, Shen Y, Yan Y, Tao D, Lu Y, Ma Y, Yang Y, Zhang S. Epigenetic modifications promote the expression of the orphan nuclear receptor NROB1 in human lung adenocarcinoma cells. *Oncotarget*. 2016; 7:43162–76.
<https://doi.org/10.18632/oncotarget.9012>
PMID:[27281610](https://pubmed.ncbi.nlm.nih.gov/27281610/)
32. Yi B, Jahangir A, Evans AK, Briggs D, Ravina K, Ernest J, Farimani AB, Sun W, Rajadas J, Green M, Feinberg EN, Pande VS, Shamloo M. Discovery of novel brain permeable and G protein-biased beta-1 adrenergic receptor partial agonists for the treatment of neurocognitive disorders. *PLoS One*. 2017; 12:e0180319.
<https://doi.org/10.1371/journal.pone.0180319>
PMID:[28746336](https://pubmed.ncbi.nlm.nih.gov/28746336/)
33. Lee HY, Chung WJ, Jeon HK, Seo HS, Choi DJ, Jeon ES, Kim JJ, Shin JH, Kang SM, Lim SC, Baek SH. Impact of the β -1 adrenergic receptor polymorphism on tolerability and efficacy of bisoprolol therapy in Korean heart failure patients: association between β adrenergic receptor polymorphism and bisoprolol therapy in heart failure (ABBA) study. *Korean J Intern Med*. 2016; 31:277–87.
<https://doi.org/10.3904/kjim.2015.043>
PMID:[26879662](https://pubmed.ncbi.nlm.nih.gov/26879662/)
34. Kang S, Hong X, Ruan CW, Yu P, Yu SS, Chen M, Zhang DF, Fan HM, Liu ZM. Effects of GRK5 and ADRB1 polymorphisms influence on systolic heart failure. *J Transl Med*. 2015; 13:44.
<https://doi.org/10.1186/s12967-015-0402-7>
PMID:[25638254](https://pubmed.ncbi.nlm.nih.gov/25638254/)
35. Nickerson ML, Witte N, Im KM, Turan S, Owens C, Misner K, Tsang SX, Cai Z, Wu S, Dean M, Costello JC, Theodorescu D. Molecular analysis of urothelial cancer cell lines for modeling tumor biology and drug response. *Oncogene*. 2017; 36:35–46.
<https://doi.org/10.1038/onc.2016.172> PMID:[27270441](https://pubmed.ncbi.nlm.nih.gov/27270441/)
36. Jakobsson J, Stridsberg M, Zetterberg H, Blennow K, Ekman CJ, Johansson AG, Sellgren C, Landén M. Decreased cerebrospinal fluid secretogranin II concentrations in severe forms of bipolar disorder. *J Psychiatry Neurosci*. 2013; 38:E21–26.
<https://doi.org/10.1503/jpn.120170> PMID:[23415276](https://pubmed.ncbi.nlm.nih.gov/23415276/)
37. Liu D, Wang X, Chen Z. Tumor necrosis factor- α , a regulator and therapeutic agent on breast cancer. *Curr Pharm Biotechnol*. 2016; 17:486–94.
<https://doi.org/10.2174/1389201017666160301102713> PMID:[26927216](https://pubmed.ncbi.nlm.nih.gov/26927216/)
38. Bauer D, Mazzi E, Soliman KF. Whole transcriptomic analysis of apigenin on TNF α immuno-activated MDA-MB-231 breast cancer cells. *Cancer Genomics Proteomics*. 2019; 16:421–31.
<https://doi.org/10.21873/cgp.20146> PMID:[31659097](https://pubmed.ncbi.nlm.nih.gov/31659097/)
39. Martínez-Reza I, Díaz L, García-Becerra R. Preclinical and clinical aspects of TNF- α and its receptors TNFR1 and TNFR2 in breast cancer. *J Biomed Sci*. 2017; 24:90.
<https://doi.org/10.1186/s12929-017-0398-9>
PMID:[29202842](https://pubmed.ncbi.nlm.nih.gov/29202842/)
40. Derossi DR, Amarante MK, Guembarovski RL, de Oliveira CE, Suzuki KM, Watanabe MA, de Syllos Cólus IM. CCL5 protein level: influence on breast cancer staging and lymph nodes commitment. *Mol Biol Rep*. 2019; 46:6165–70.
<https://doi.org/10.1007/s11033-019-05051-8>
PMID:[31691056](https://pubmed.ncbi.nlm.nih.gov/31691056/)
41. An G, Wu F, Huang S, Feng L, Bai J, Gu S, Zhao X. Effects of CCL5 on the biological behavior of breast

- cancer and the mechanisms of its interaction with tumor-associated macrophages. *Oncol Rep.* 2019; 42:2499–511.
<https://doi.org/10.3892/or.2019.7344> PMID:31578575
42. Kundu N, Ma X, Brox R, Fan X, Kochel T, Reader J, Tschammer N, Fulton A. The chemokine receptor CXCR3 isoform B drives breast cancer stem cells. *Breast Cancer (Auckl).* 2019; 13:1178223419873628.
<https://doi.org/10.1177/1178223419873628> PMID:31619923
43. Uhercik M, Sanders AJ, Owen S, Davies EL, Sharma AK, Jiang WG, Mokbel K. Clinical significance of PD1 and PDL1 in human breast cancer. *Anticancer Res.* 2017; 37:4249–54.
<https://doi.org/10.21873/anticancer.11817> PMID:28739716
44. Planes-Laine G, Rochigneux P, Bertucci F, Chrétien AS, Viens P, Sabatier R, Gonçalves A. PD-1/PD-L1 targeting in breast cancer: the first clinical evidences are emerging. A literature review. *Cancers (Basel).* 2019; 11:1033.
<https://doi.org/10.3390/cancers11071033> PMID:31336685
45. Khalife E, Khodadadi A, Talaeizadeh A, Rahimian L, Nemati M, Jafarzadeh A. Overexpression of regulatory T cell-related markers (FOXP3, CTLA-4 and GITR) by peripheral blood mononuclear cells from patients with breast cancer. *Asian Pac J Cancer Prev.* 2018; 19:3019–25.
<https://doi.org/10.31557/APJCP.2018.19.11.3019> PMID:30484986
46. Tuit S, Salvagno C, Kapellos TS, Hau CS, Seep L, Oestreich M, Klee K, de Visser KE, Ulas T, Schultze JL. Transcriptional signature derived from murine tumor-associated macrophages correlates with poor outcome in breast cancer patients. *Cell Rep.* 2019; 29:1221–35.e5.
<https://doi.org/10.1016/j.celrep.2019.09.067> PMID:31665635
47. Su S, Liao J, Liu J, Huang D, He C, Chen F, Yang L, Wu W, Chen J, Lin L, Zeng Y, Ouyang N, Cui X, et al. Blocking the recruitment of naive CD4⁺ T cells reverses immunosuppression in breast cancer. *Cell Res.* 2017; 27:461–82.
<https://doi.org/10.1038/cr.2017.34> PMID:28290464
48. Huang Y, Ma C, Zhang Q, Ye J, Wang F, Zhang Y, Hunborg P, Varvares MA, Hoft DF, Hsueh EC, Peng G. CD4⁺ and CD8⁺ T cells have opposing roles in breast cancer progression and outcome. *Oncotarget.* 2015; 6:17462–78.
<https://doi.org/10.18632/oncotarget.3958> PMID:25968569
49. Santa-Maria CA, Nanda R. Immune checkpoint inhibitor therapy in breast cancer. *J Natl Compr Canc Netw.* 2018; 16:1259–68.
<https://doi.org/10.6004/jnccn.2018.7046> PMID:30323094
50. Ren X, Wu H, Lu J, Zhang Y, Luo Y, Xu Q, Shen S, Liang Z. PD1 protein expression in tumor infiltrated lymphocytes rather than PDL1 in tumor cells predicts survival in triple-negative breast cancer. *Cancer Biol Ther.* 2018; 19:373–380.
<https://doi.org/10.1080/15384047.2018.1423919> PMID:29336717
51. Li CW, Lim SO, Chung EM, Kim YS, Park AH, Yao J, Cha JH, Xia W, Chan LC, Kim T, Chang SS, Lee HH, Chou CK, et al. Eradication of triple-negative breast cancer cells by targeting glycosylated PD-L1. *Cancer Cell.* 2018; 33:187–201.e10.
<https://doi.org/10.1016/j.ccell.2018.01.009> PMID:29438695
52. Zhang L, Wang XI, Ding J, Sun Q, Zhang S. The predictive and prognostic value of Foxp3⁺/CD25⁺ regulatory T cells and PD-L1 expression in triple negative breast cancer. *Ann Diagn Pathol.* 2019; 40:143–51.
<https://doi.org/10.1016/j.anndiagpath.2019.04.004> PMID:31096176
53. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS, Thomson E, Wiser J, Butte AJ. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data.* 2018; 5:180015.
<https://doi.org/10.1038/sdata.2018.15> PMID:29485622
54. 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
55. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017; 45:D353–61.
<https://doi.org/10.1093/nar/gkw1092> PMID:27899662
56. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019; 47:D419–26.
<https://doi.org/10.1093/nar/gky1038> PMID:30407594
57. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–87.

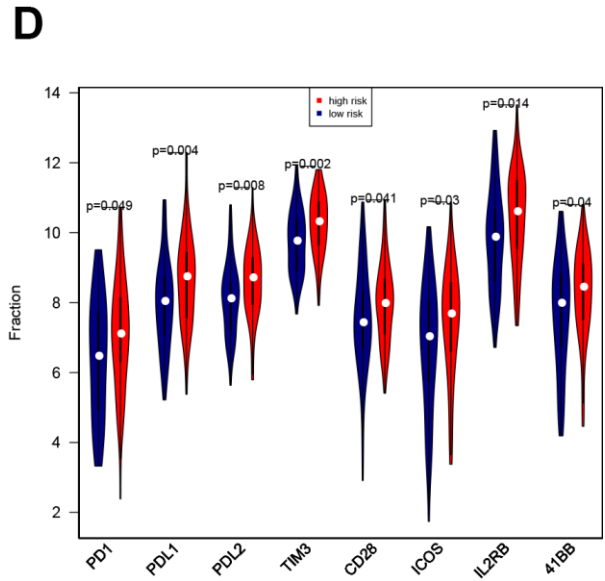
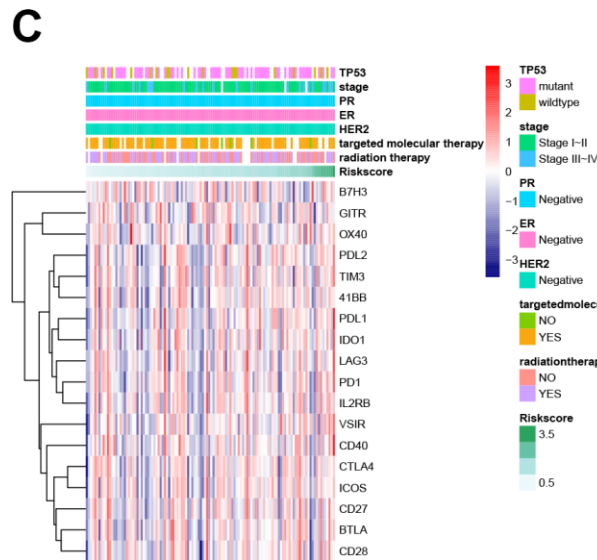
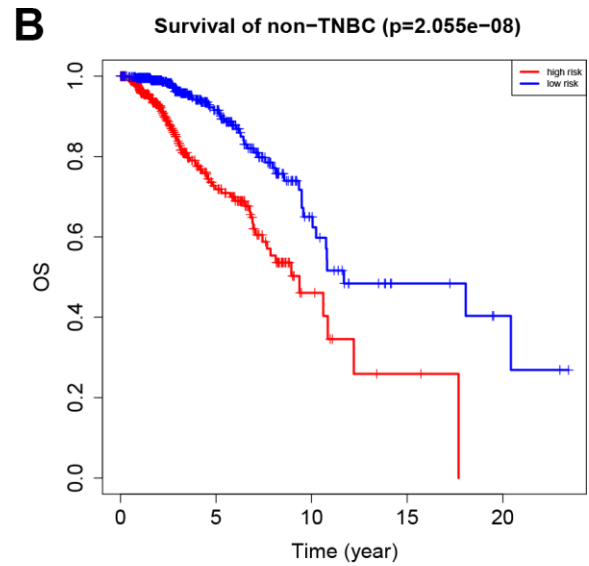
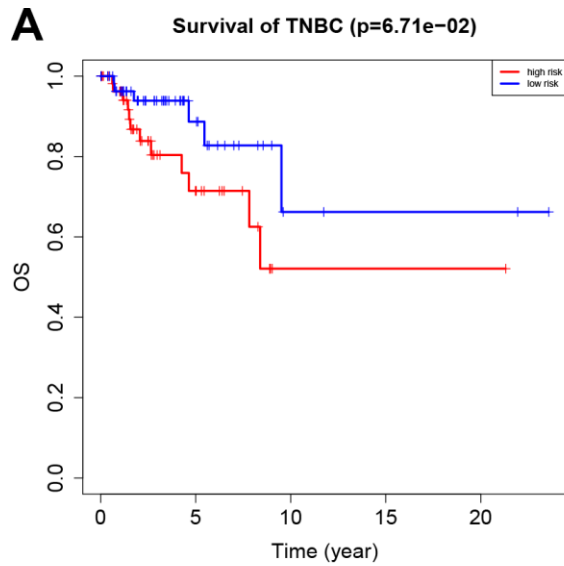
- <https://doi.org/10.1089/omi.2011.0118>
PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
58. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CG. Time-varying covariates and coefficients in cox regression models. *Ann Transl Med.* 2018; 6:121.
<https://doi.org/10.21037/atm.2018.02.12>
PMID:[29955581](https://pubmed.ncbi.nlm.nih.gov/29955581/)
59. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010; 33:1–22.
PMID:[20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
60. Goel MK, Khanna P, Kishore J. Understanding survival analysis: kaplan-meier estimate. *Int J Ayurveda Res.* 2010; 1:274–78.
<https://doi.org/10.4103/0974-7788.76794>
PMID:[21455458](https://pubmed.ncbi.nlm.nih.gov/21455458/)
61. 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/)
62. 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/)
63. 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:D991–95.
<https://doi.org/10.1093/nar/gks1193> PMID:[23193258](https://pubmed.ncbi.nlm.nih.gov/23193258/)
64. Kao KJ, Chang KM, Hsu HC, Huang AT. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer.* 2011; 11:143.
<https://doi.org/10.1186/1471-2407-11-143>
PMID:[21501481](https://pubmed.ncbi.nlm.nih.gov/21501481/)
65. 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/)
66. 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/)
67. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell.* 2017; 65:631–43.e4.
<https://doi.org/10.1016/j.molcel.2017.01.023>
PMID:[28212749](https://pubmed.ncbi.nlm.nih.gov/28212749/)
68. 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/)
69. Aran D. Cell-type enrichment analysis of bulk transcriptomes using xCell. *Methods Mol Biol.* 2020; 2120:263–76.
https://doi.org/10.1007/978-1-0716-0327-7_19
PMID:[32124326](https://pubmed.ncbi.nlm.nih.gov/32124326/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Validation of the eight-gene signature using survival analysis of GEO datasets and in different subtypes and stages of breast cancer. (A, B) Survival analysis of the GSE20685 and GSE21653 GEO datasets. **(C, D)** Survival analysis of breast cancer patients depending on TP53 mutation status. **(E, F)** Survival analysis of stage I-II vs. stage III-IV breast cancer patients. **(G–L)** Survival analysis in breast cancer patients with the following subtypes: ER positive/negative, PR positive/negative, HER2 positive/negative.



Supplementary Figure 2. Predictive value of the eight-gene signature for survival outcomes and immune checkpoints in TNBC subgroup. (A) Survival analysis of high- and low-risk TNBC patients. (B) Survival analysis of high- and low-risk non-TNBC patients. (C) Heatmap showing associations between risk score, clinical variables, and 18 immune checkpoints. (D) Predictive value of the eight-gene signature for PD1, PDL1, PDL2, TIM3, CD28, ICOS, IL2RB, and 41BB.