

Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of colorectal cancer

Wei Li^{1,*}, Xiaojing Jin^{2,*}, Shang Guo¹, Fei Xu¹, Xingkai Su¹, Xia Jiang¹, Guiqi Wang¹

¹Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

²Department of Emergency Medicine, The First Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

*Equal contribution

Correspondence to: Xia Jiang, Guiqi Wang; **email:** jiangxia0925@hebmh.edu.cn; wangguiqi2011@126.com, <https://orcid.org/0000-0002-7253-11569>

Keywords: colorectal cancer, TCGA, GEO, prognosis, nomogram

Received: July 22, 2020

Accepted: November 20, 2020

Published: February 1, 2021

Copyright: © 2021 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

In this study, we used the ESTIMATE algorithm to analyze clinical data and transcriptome profiles of 1635 colorectal cancer (CRC) samples from the Gene Expression Omnibus and The Cancer Genome Atlas databases and identify prognostic immune-related genes (IRGs). We identified 941 differentially expressed (4 downregulated and 937 upregulated) genes by comparing samples with high and low immune, stromal scores and tumor purity. LASSO Cox regression analyses showed that the risk score based on a ten-IRG signature was an independent prognostic factor in CRC. The nomogram with pathological stages (TNM) and the ten-IRG signature showed a C-index of 0.769 (95% CI, 0.717-0.821), and area under ROC curve values of 0.788, 0.782 and 0.789 for 1-, 3-, and 5-year OS, respectively. TIMER database analysis showed positive correlation between the ten prognostic IRGs and the levels of tumor-infiltrated immune cells, including CD4⁺ and CD8⁺ T cells, macrophages, neutrophils, and dendritic cells. These findings demonstrate that this novel ten-IRG signature correlates with the pathological stages and the levels of multiple tumor-infiltrated immune cell types. This makes the ten-IRG signature a potential prognostic factor for CRC patients.

INTRODUCTION

Colorectal cancer (CRC) is the third most commonly diagnosed cancer among men and women, and the second leading cause of cancer deaths worldwide [1, 2]. The tumor microenvironment (TME) plays a crucial role in tumor development and progression and consists of cancer cells, stromal cells, and several types of immune cells [3, 4]. Colorectal cancer is highly immunogenic and may show positive response to immunotherapy [5–7]. In several cancers, normalization of the tumor microenvironment improves the efficacy of targeted therapies, radiotherapy, and chemotherapy [8–10]. This suggests that the proportion of different immune cell types and their functional status in the TME determines the efficacy of immunotherapy in CRC.

Several studies have demonstrated the association between TME and survival outcomes as well as tumor recurrence in CRC patients. Xiong et al. showed that reduced levels of M1 macrophages and elevated levels of M2 macrophages, neutrophils, and eosinophils in the TME were associated with poor prognosis of CRC patients [11]. Furthermore, Ye et al. reported significant correlation between the levels of Tregs, neutrophils, and macrophages in the TME and the prognosis of CRC patients [12]. Recent advances in high-throughput sequencing technologies have helped define the genomic landscape of CRC and identify several TME-related molecular signatures for predicting the prognosis of CRC patients [13–15]. However, these molecular signatures require further analysis and validation to explore their utility in clinical applications.

Large-scale transcriptome, genomic, mutational, and clinical data of cancer patients is now available from public databases like Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). Yoshihara et al. developed a novel algorithm to called ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumors using Expression data) to compute the fraction of stromal and immune cells in tumor tissues [16]. ESTIMATE algorithm generates three scores-stromal, immune, and estimate. Stromal and immune scores are a measure of the proportion of stromal cells and immune cells in the tumor tissues, respectively. Estimate score indicates tumor purity, which is defined as the percentage of tumor cells in the TME and is closely related to the prognosis of cancer. Recent studies have shown that ESTIMATE scores show the extent of infiltration of non-tumor cells such as the immune cell types into the TME of cutaneous melanoma, glioblastoma, and adrenocortical carcinoma [17–19].

In this study, we analyzed the transcriptome data of CRC patients from the TCGA and GEO databases using the ESTIMATE algorithm to identify and characterize immune-related genes (IRGs) that can be used to accurately predict the prognosis of CRC patients. Furthermore, we analyzed the potential prognostic IRGs using the LASSO (Least Absolute Shrinkage and

Selection Operator) Cox regression model and We constructed a nomogram with ten prognostic IRGs after analysis using the LASSO Cox regression model and further evaluated the correlation between clinicopathological features and the expression status of the prognostic IRGs in CRC samples.

RESULTS

ESTIMATE analysis shows relationship between tumor purity, stromal scores, and immune scores with CRC prognosis

We downloaded OS and clinicopathological data for 1635 eligible CRC patients from the TCGA and GEO databases with a mean age of 68 years at diagnosis. Among these, 53.9% of the CRC patients were males. ESTIMATE results showed that immune scores ranged from -899.57 and 3202.84, stromal scores ranged from -2232.54 to 2193.08, and tumor purity ranged from 0.27 to 0.98. We then classified the CRC patients into low- and high-score groups based on their immune, stromal, and tumor purity scores, and investigated their correlation with OS rates. We observed positive correlation between stromal scores and OS of CRC patients ($p=0.035$; Figure 1A). However, immune scores did not show any significant correlation with OS of CRC

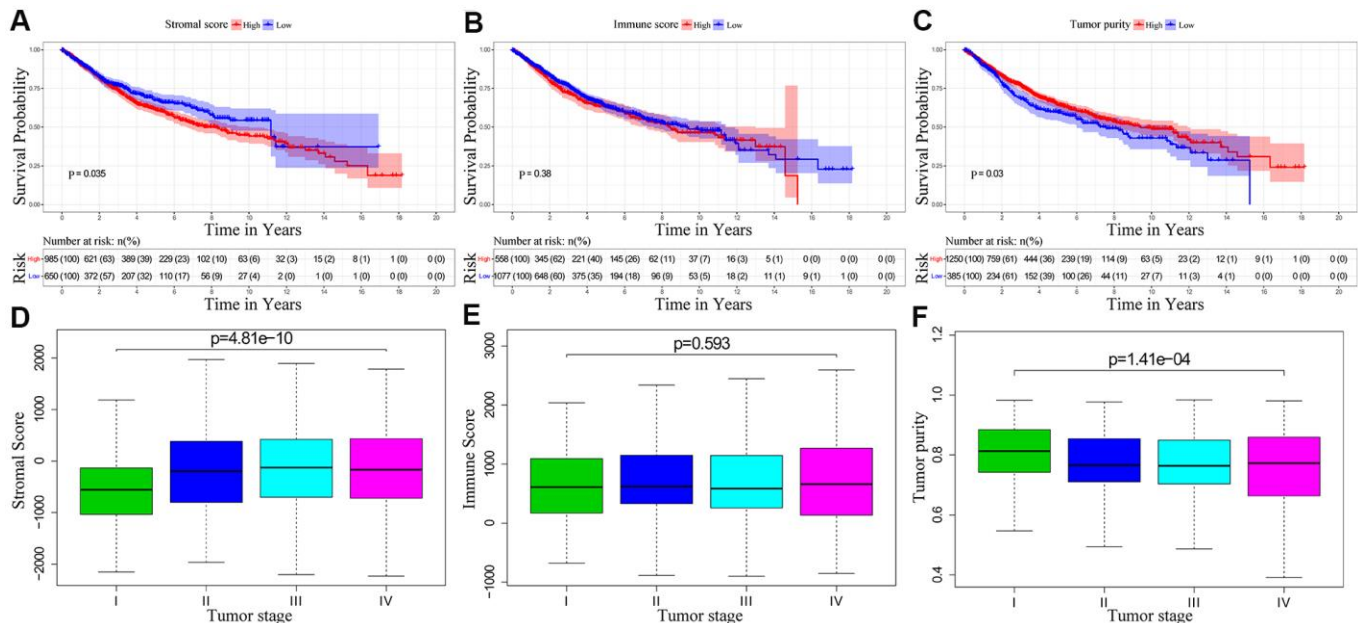


Figure 1. The relationship between stromal score, immune score and tumor purity of CRC patient samples with pathological stages and overall survival. (A) Kaplan-Meier survival curve analysis shows overall survival of CRC patients with high and low stromal scores. (B) Kaplan-Meier survival curve analysis shows overall survival of CRC patients with high and low immune scores. (C) Kaplan-Meier survival curve analysis shows overall survival of CRC patients with high and low tumor purity. (D) Correlation analysis between stromal scores and pathological stages of CRC patients. (E) Correlation analysis between immune scores and pathological stages of CRC patients. (F) Correlation analysis between tumor purity and pathological stages of CRC patients.

patients ($p=0.381$; Figure 1B). We also observed negative correlation between tumor purity and OS ($p=0.03$; Figure 1C). We then assessed the relationship between immune scores, stromal scores and tumor purity with the pathological stages of CRC. We observed positive correlation between stromal scores and pathologic stage ($p = 4.815e-10$; Figure 1D). Immune scores were not associated with the pathological stage of CRC patients ($p = 0.593$; Figure 1E). Tumor purity showed inverse correlation with the pathological stages of CRC patients ($p = 1.411e-04$; Figure 1F).

Identification of differentially expressed immune-related genes in CRC tissues

As shown by the volcano plots, we identified several differentially expressed genes (upregulated and downregulated) between the low- and high-score groups in the TCGA-CRC cohort ($n=611$; Figure 2A–2C). We

identified 9 downregulated and 1493 upregulated genes in the high stromal score group compared to the low stromal score group (Figure 2A), 61 downregulated and 1235 upregulated genes in the high immune score group compared to the low immune score group (Figure 2B), and 1830 upregulated genes and 69 downregulated genes in the low tumor purity group compared to the high tumor purity score group (Figure 2C). Overall, we identified 4 downregulated and 937 upregulated IRGs that were common among all the three groups (Figure 2D–2E).

Functional enrichment analysis of the differentially expressed IRGs

We performed functional enrichment analysis using the clusterProfiler R package and identified 1454 significantly enriched GO terms and 54 significantly enriched KEGG pathways ($FDR<0.05$) associated with

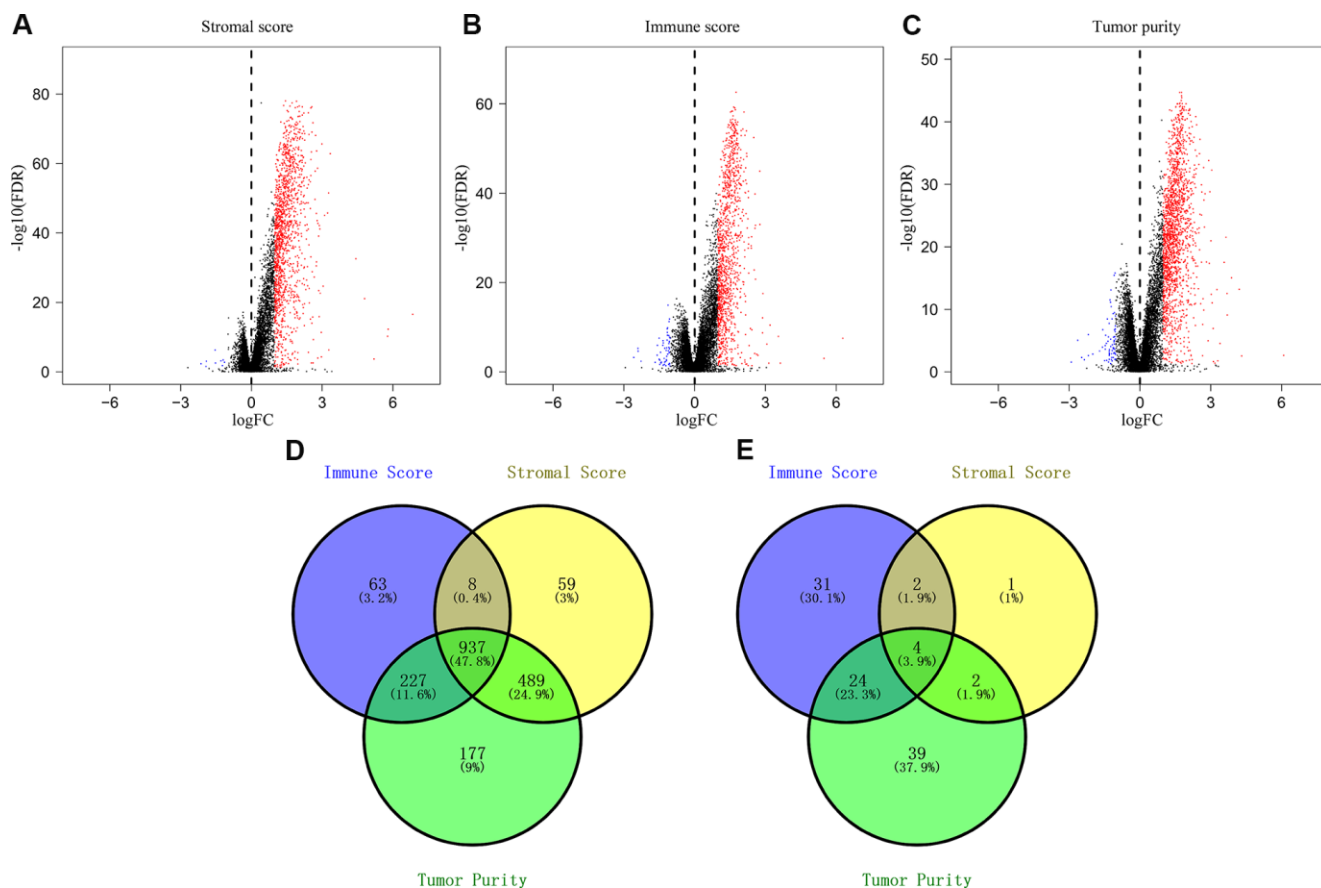


Figure 2. Identification of differentially expressed genes based on the immune scores, stromal scores, and tumor purity of CRC patients. (A) Volcano plots of the DEGs based on stromal scores. (B) Volcano plots of the DEGs based on immune scores. (C) Volcano plots of the DEGs based on tumor purity. (D) Venn diagram shows the numbers of upregulated genes in the immune score, stromal score and tumor purity groups as well as the upregulated genes that are common among the three groups. (E) Venn diagram shows the numbers of downregulated genes in immune score, stromal score and tumor purity groups as well as the downregulated genes that are common among the three groups.

the 941 differentially expressed IRGs. The top 20 GO terms and KEGG pathways are shown in Figure 3 and Supplementary Table 1. The top GO terms related to biological processes (BF) included regulation of leukocyte activation (GO:0002694), leukocyte migration (GO:0050900), and T cell activation (GO:0042110). The top GO terms related to cellular components (CC) were membrane (GO:0098552) and extracellular matrix (GO:0031012). The top GO terms related to molecular functions (MF) included receptor regulator activity (GO:0030545) and receptor ligand activity (GO:0048018). The top enriched KEGG pathways were cytokine-cytokine receptor interactions (hsa04060), chemokine signaling pathway (hsa04062), PI3K-Akt signaling pathway (hsa04151), and phagosomes (hsa04145).

Construction and validation of the IRG-based prognostic signature

We used univariate COX regression and Lasso regression analysis to identify IRGs that predict the prognosis of CRC patients. We identified 172 IRGs based on the LASSO regression analysis results that were significantly correlated with the OS of CRC patients (Figure 4A, 4B). Subsequently, based on the co-efficient values, we generated an IRG-based prognostic risk signature with ten IRGs. The risk score was calculated with the following formula: $(0.1235 \times \text{expression of MAP2}) + (0.0873 \times \text{expression of NKAIN4}) + (0.2936 \times \text{expression of VAX2 status}) + (0.0321 \times \text{expression of CALB2}) + (0.3958 \times \text{expression of FCRL2}) + (0.0471 \times \text{expression of HAND1}) + (0.0986 \times \text{expression of A2ML1}) + (0.002 \times \text{expression of IDO1}) + (0.0134 \times$

expression of COL22A1) + $(-0.4143 \times \text{expression of CD1B})$. CRC patients from the TCGA database (n=611) were then classified into low- and high-risk groups according to the cutoff risk score of 1.099 (Figure 4C, 4D). The high-risk group CRC patients showed significantly lower OS than those in the low-risk group ($p < 0.0001$; Figure 4E). The area under the ROC curve (AUC) values for 1-, 3-, and 5-year OS were 0.708, 0.716, and 0.680, respectively (Figure 4F). In the internal validation cohort, patients from the high-risk group demonstrated significantly worse survival outcomes compared to the low-risk group ($P < 0.001$; Supplementary Figure 1). The AUC values for the internal validation cohort were 0.722, 0.712, and 0.738, for 1-, 3- and 5-year OS, respectively (Supplementary Figure 1). Next, we verified the prognostic model in the GSE39582 cohort (n=531) by first classifying the CRC patients into low- and high-risk groups according to the cutoff risk score of 0.989 (Figure 4G, 4H). The high-risk CRC patients from the GSE39582 cohort showed significantly lower OS than the corresponding low-risk patients ($p = 0.006$; Figure 4I). The AUC values for 1-, 3-, and 5-year OS for CRC patients from the GSE39582 cohort were 0.711, 0.632, and 0.613, respectively (Figure 4J).

The ten-IRG signature is an independent prognostic predictor for CRC patients

We performed univariate and multivariate Cox regression analyses using the TCGA and GEO CRC patient datasets after adjusting for clinicopathological parameters such as M stage, N stage, T stage, tumor stage, age and gender to determine if the risk score based on the ten-IRG signature accurately predicted

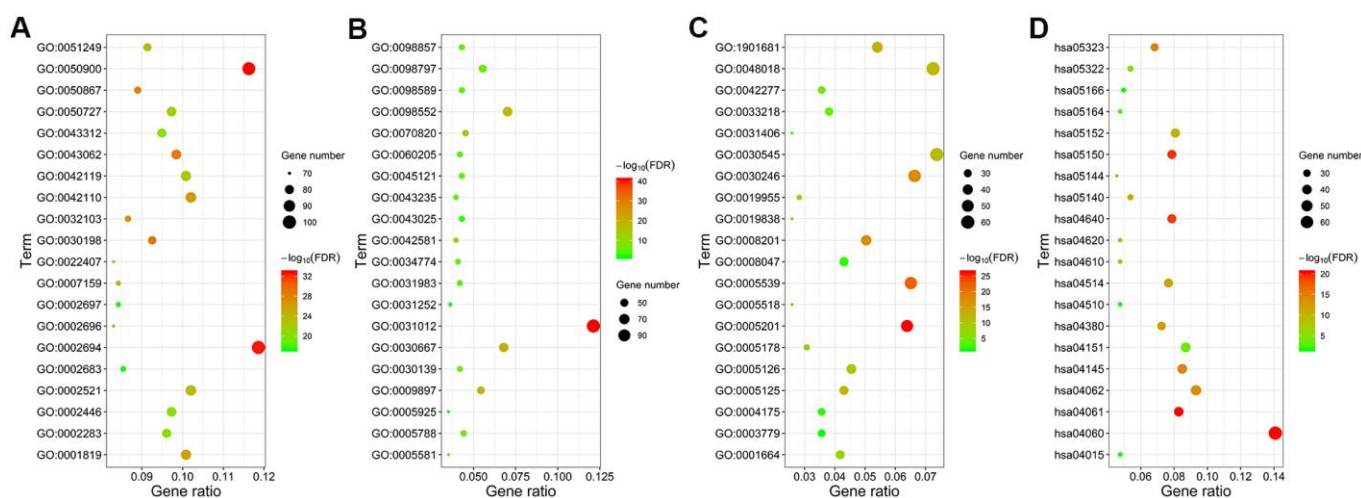


Figure 3. Functional enrichment analysis of the differentially expressed IRGs. GO and KEGG pathway analyses results show the most enriched GO terms related to (A) biological functions (BF), (B) cellular component (CC), and (C) molecular functions (MF), and (D) KEGG pathways related to differentially expressed IRGs.

prognosis of CRC patients. The results showed that the ten-IRG signature-based risk score was an independent prognosis factor for determining the OS of CRC patients from the two datasets (Figure 5A, 5B, 5D, 5E and Supplementary Figure 1). Next, we analyzed the correlation between the ten-IRG signature and other clinicopathological parameters using the Chi-square test. In the TCGA cohort, the higher risk score based on the ten-IRG signature was associated with tumor stage ($p < 0.001$), M stage ($p < 0.05$), N stage ($p < 0.001$), and T stage ($p < 0.001$) (Figure 5C), but was not related to gender and age of CRC patients. We obtained similar results with the GSE39582 validation cohort of CRC patients (Figure 5F).

Nomogram construction and validation

Next, we analyzed if the ten-IRG prognostic signature would enhance the accuracy of predicting OS of CRC patients. Towards this, we constructed a nomogram with the ten-IRG signature-based risk score, M stage, N stage, T stage, and age to determine the 1-, 3-, and 5-year OS of CRC patients (Figure 6A). The calibration plots showed that the nomogram accurately predicted the 3-year OS of CRC patients (Figure 6B). Overall, the nomogram significantly improved the prediction of the 1-, 3-, and 5-year OS of CRC patients (Figure 6C). The OS prediction of the nomogram was significantly higher compared to the OS prediction based on TNM stages

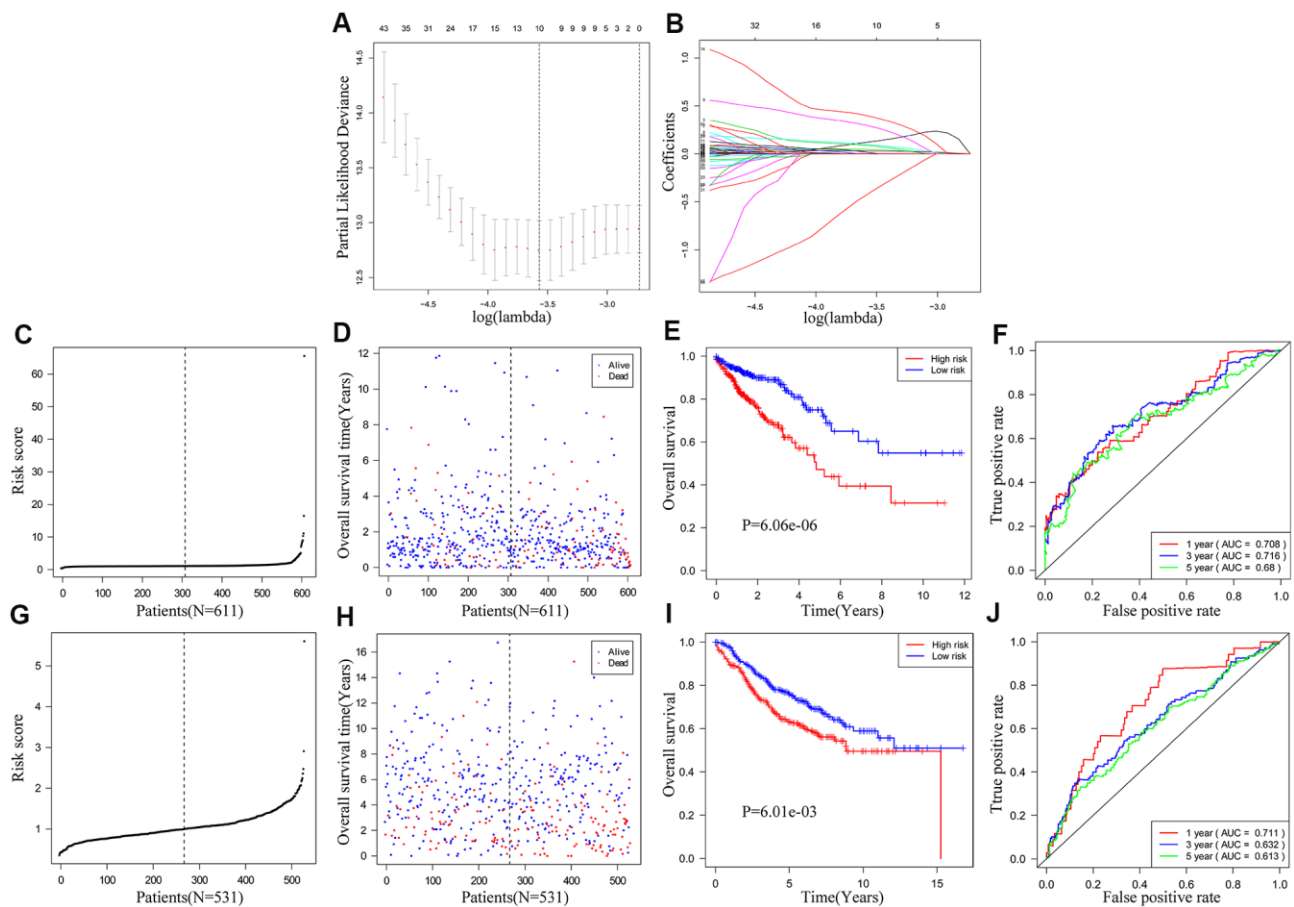


Figure 4. Construction and validation of the ten-IRG prognostic signature. (A) A plot of partial likelihood deviance of the LASSO coefficient profiles of differentially expressed IRGs. (B) A plot of the LASSO coefficient profiles for the differentially expressed IRGs associated with overall survival of CRC. (C) Distribution of risk scores for the CRC patients in the TCGA cohort based on the ten-IRG prognostic signature. (D) The survival status of 611 CRC patients in the TCGA cohort belonging to the high- and low-risk groups based on the ten-IRG prognostic signature. (E) Kaplan–Meier survival curves show overall survival of high and low-risk CRC patients in the TCGA cohort based on the ten-IRG prognostic signature. (F) Time-dependent ROC curves show the accuracy of overall survival prediction for the TCGA-CRC cohort based on the ten-IRG prognostic signature. (G) Distribution of risk scores for the GSE39582 cohort based on the ten-IRG prognostic signature. (H) The survival status of 531 CRC patients in the GSE3958 cohort belonging to the high- and low-risk groups based on the ten-IRG prognostic signature. (I) Kaplan–Meier survival curve analysis shows the overall survival of high and low-risk CRC patients in the GSE3958 cohort based on the ten-IRG prognostic signature. (J) Time-dependent ROC curves show the accuracy of overall survival prediction for the GSE3958 cohort based on the ten-IRG prognostic signature.

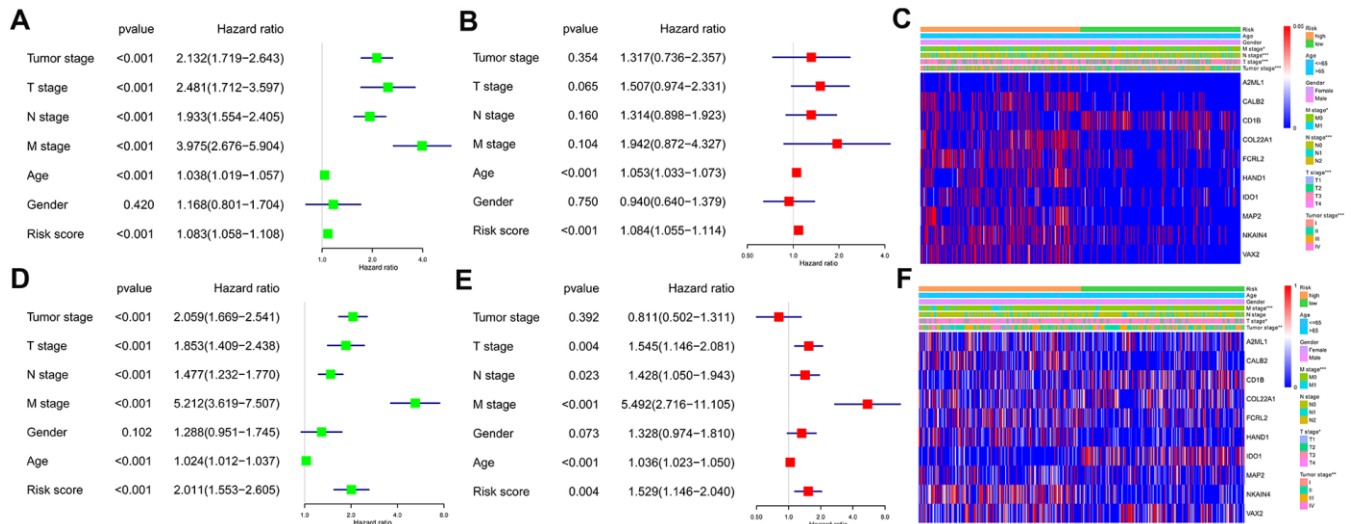


Figure 5. Association of the ten-IRG prognostic signature with overall survival of CRC patients. (A) Univariate COX regression analysis shows the clinicopathological parameters associated with the overall survival of CRC patients in the TCGA cohort. (B) Multivariate COX regression analysis shows clinicopathological parameters associated with the overall survival of CRC patients in the TCGA cohort. (C) Correlation analysis results show the relationship between the ten-IRG prognostic signature and the clinicopathological parameters in the TCGA-CRC cohort. (D) Univariate COX regression analysis shows the clinicopathological parameters associated with the overall survival of CRC patients in the GSE39582 cohort. (E) Multivariate COX regression analysis shows clinicopathological parameters associated with the overall survival of CRC patients in the GSE39582 cohort. (F) Correlation analysis results show the relationship between the ten-IRG prognostic signature and the clinicopathological parameters in the GSE39582-CRC cohort.

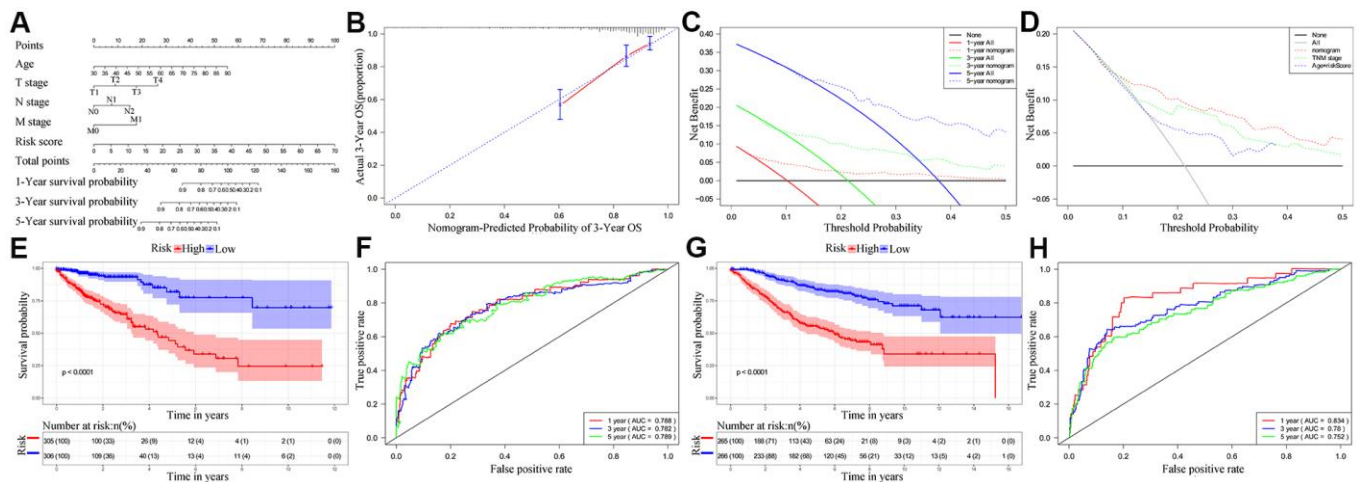


Figure 6. Establishment and validation of the nomogram for predicting overall survival of CRC patients in the TCGA and GSE39582 cohorts. (A) Nomogram with the ten-IRG prognostic risk score, TNM stages and age for predicting the 1-year, 3-year, and 5-year OS of CRC patients. In the nomogram, each variable is assigned a score. The sum of scores for all variables is used to predict the probability of survival of the CRC patients. (B) Calibration plot shows the comparison between nomogram predicted and actual 3-year OS of the TCGA cohort. (C) Decision curve analysis shows the predicted 1-year, 3-year and 5-year overall survival of CRC patients based on the nomogram. (D) Decision curve analysis shows the predicted 1-year, 3-year, and 5-year OS of CRC patients based on the nomogram, TNM stage only, and age plus ten-IRG signature. (E) Kaplan–Meier survival curves show the overall survival of CRC patients in the TCGA cohort based on the nomogram. (F) Time-dependent ROC curves show the accuracy of overall survival prediction in the TCGA cohort based on the nomogram. (G) Kaplan–Meier survival curves show the overall survival of CRC patients in the GSE39582 cohort based on the nomogram. (H) Time-dependent ROC curves show the accuracy of overall survival prediction in the GSE39582 cohort based on the nomogram.

alone or age plus the ten-IRG signature-based risk score (Figure 6D). The C-index value for the nomogram was 0.769 (95% CI, 0.717-0.821). The AUC values for the 1-, 3-, and 5-year OS based on the nomogram were 0.788, 0.782 and 0.789, respectively (Figure 6F). We then verified the predictive value of the nomogram in the TCGA and GSE39582 validation cohorts. The C-index value for the internal validation cohort (TCGA) was 0.781 (95% CI, 0.709-0.853), and the AUC values for 1-, 3-, and 5-year OS were 0.834, 0.780, and 0.752, respectively (Supplementary Figure 2). The C-index value for the GSE39582 cohort was 0.732 (95% CI, 0.691-0.773) and the AUC values for 1-, 3-, and 5-year OS were 0.834, 0.780, and 0.752, respectively (Figure 6H). Furthermore, we stratified the CRC patients into high- and low-risk groups based on the median risk score cut off value of the nomogram. Kaplan-Meier survival curves and time-dependent ROC curves showed significantly worse survival outcomes for patients from the high-risk group compared to those from the low-risk group in both the TCGA cohort ($P < 0.001$; Figure 6E, 6F) and GSE39582 cohort ($P < 0.001$; Figure 6G, 6H and Supplementary Figure 2).

Survival and immune cells infiltration analysis of ten IRGs in signature

We performed Kaplan-Meier survival curve analysis of the TCGA-CRC patients based on the expression of the ten prognostic IRGs to determine their independent prognostic values. High expression of A2ML1 (Figure 7A), CALB2 (Figure 7B), COL22A1 (Figure 7D),

FCRL2 (Figure 7E), HAND1 (Figure 7F), IDO1 (cut off value: 34, Figure 7G), MAP2 (Figure 7H), NKAIN4 (Figure 7I) and VAX2 (Figure 7J) correlated with shorter OS in CRC patients, whereas, high expression of CD1B (Figure 7C) significantly correlated with longer OS. We observed similar results in the GSE39582 cohort (Supplementary Figure 3A–3J). Interestingly, IDO1 may also have another tumor suppressor gene identity in the GSE39582 cohort (Supplementary Figure 3G) and in TCGA database when expression was between 1.06 and 34 (Supplementary Figure 4).

We performed TIMER database analysis to determine the correlation between the ten-IRG prognostic signature and tumor infiltration of six types of immune cells, namely B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells in the TME. The ten-IRG signature-related risk score showed positive correlation with the levels of CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells in the TME (all $P < 0.05$), but was not significantly associated with the levels of B cells ($P > 0.05$; Figure 8). The levels of CD1B expression correlated with the infiltration levels of neutrophils ($r = 0.468$, $p = 3.31E-23$) and dendritic cells ($r = 0.505$, $P = 2.05E-27$) in colon adenocarcinoma (COAD) samples (Figure 9A and Table 1). Moreover, FCRL2 expression levels showed significant correlation with the infiltration levels of B cells ($r = 0.484$, $p = 3.74E-25$), CD4⁺ T Cells ($r = 0.447$, $p = 3.63E-21$), and dendritic cells ($r = 0.434$, $P = 6.73E-20$) (Figure 9B and Table 1). IDO1 expression levels were significantly associated with the infiltration levels

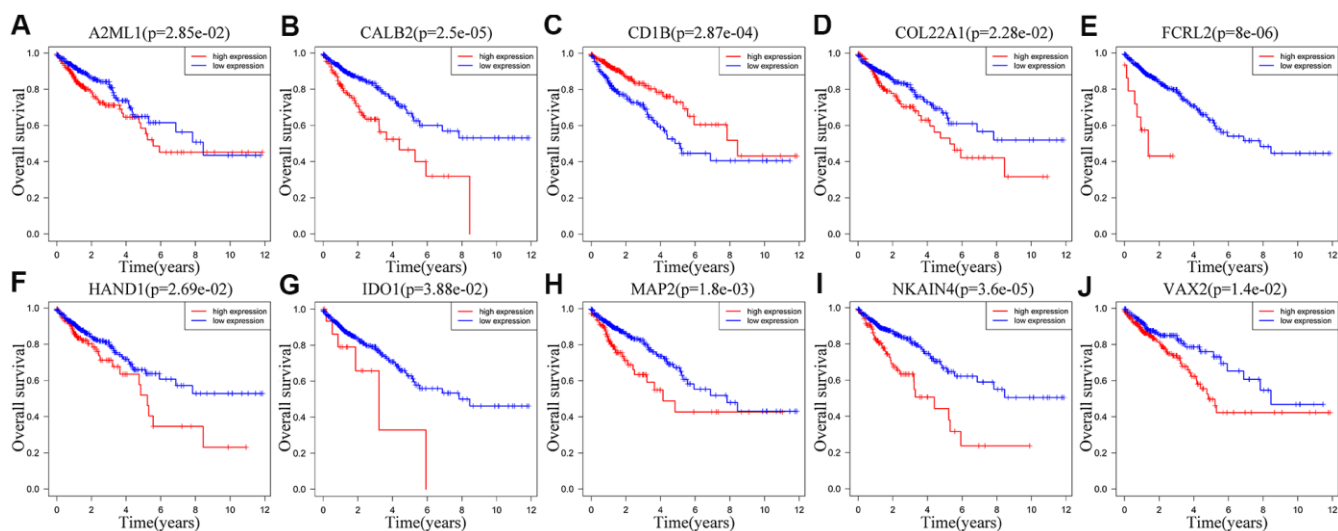


Figure 7. Correlation between the expression levels of individual IRGs from the ten-IRG prognostic signature and OS of CRC patients in the TCGA database. Kaplan-Meier survival plots show OS of CRC patients with high (red line) and low (blue line) expression of the ten individual IRGs. Kaplan-Meier survival curves for (A) A2ML1, (B) CALB2, (C) CD1B, (D) COL22A1, (E) FCRL2, (F) HAND1, (G) IDO1, (H) MAP2, (I) NKAIN4, (J) VAX2. Note: $p < 0.05$ by the log-rank test.

of CD8⁺ T cells ($r = 0.4$, $P = 4.66E-17$), neutrophils ($r = 0.638$, $p = 3.97E-47$), and dendritic cells ($r = 0.564$, $P = 3.47E-35$) (Table 1). MAP2 expression levels were significantly associated with the infiltration levels of CD4⁺ T Cells ($r = 0.441$, $p = 1.38E-20$), macrophages ($r = 0.535$, $p = 3.03E-31$), and dendritic cells ($r = 0.465$, $P = 5.68E-23$) (Figure 9C and Table 1).

DISCUSSION

In this study, we used the ESTIMATE algorithm to analyze the transcriptome and clinical data of CRC patients from the GEO and TCGA databases to determine immune-related gene (IRG) signatures that predict survival outcomes. The ESTIMATE algorithm calculates stromal and immune scores and predicts tumor purity based on single-sample gene set enrichment analysis (ssGSEA) of specific gene signatures that are related to the proportions of stromal and immune cells in the TME. The accuracy of the ESTIMATE algorithm has been validated in several cancer types and shown promising results. However, the ESTIMATE algorithm has not been used to evaluate prostate or pancreatic cancers because of the atypical nature of the tumor cells and lack of sufficient data [16]. We screened differentially expressed IRGs by comparing CRC

patients from the TCGA database with high and low immune, stromal or tumor purity scores and built a risk signature including ten IRG genes based on the LASSO Cox regression analyses. The ten-IRG risk signature was significantly associated with the OS of CRC patients from the TCGA and GSE39582 cohorts. Furthermore, the risk signature was significantly associated with the TNM stages. These results demonstrate that the ten-IRG signature predicts the prognosis of patients with CRC.

We then established a nomogram that included the risk score based on the ten-IRG signature and clinicopathological parameters including TNM stages and age to accurately predict OS of CRC patients. The predictive performance of the nomogram was satisfactory. The calibration plots demonstrated that the actual OS and predicted OS values were comparable. The C-index values for the nomogram were 0.769 and 0.732 for the TCGA and GSE39582 cohorts, respectively. The prognostic accuracy was higher when the clinicopathological parameters were combined with the ten-IRG prognostic signature compared to clinicopathological parameters alone or the ten-IRG prognostic signature alone. This demonstrated that the ten-IRG signature improved the prognostic prediction of the clinicopathological parameters.

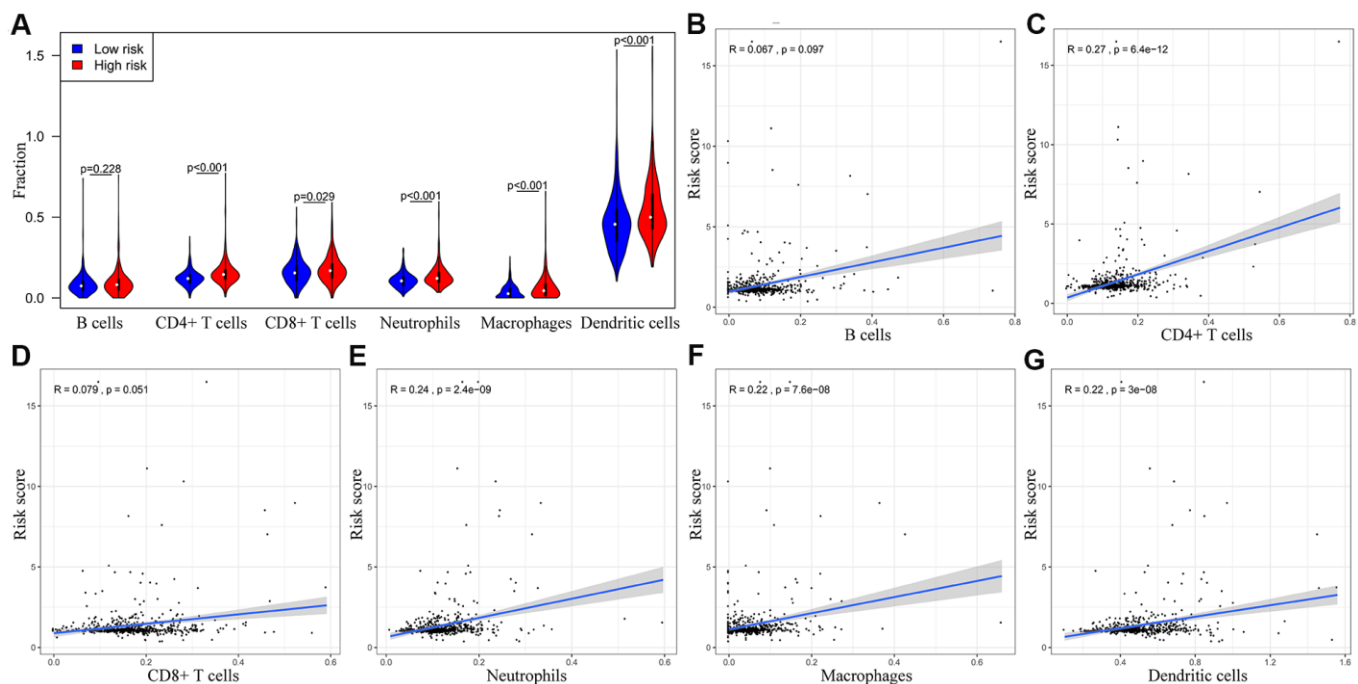


Figure 8. The abundance of six immune cell types in the CRC tissues correlates with the risk scores according to the ten-IRG prognostic signature. (A) The abundance of six immune cell types (B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells) in the high and low risk groups based on the ten-IRG prognostic signature. **(B–G)** The correlation between the ten-IRG prognostic signature and the abundance of B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells in the CRC tissues.

In general, the status and function of tumor-infiltrating immune cells regulates the biological behavior of cancers. CD4⁺ and CD8⁺ T cells identify cancer cell-related antigens and play a significant role in cancer immunotherapy [20, 21]. Conversely, tumor-associated macrophages (TAMs) provide trophic and nutritional support to the malignant cancer cells and mediate therapeutic resistance and disease progression [22]. TAMs also represent potential targets for human

anticancer therapies [23]. Neutrophils play a significant role in tumor initiation, growth and progression, and have been identified as therapeutic targets and clinical biomarkers in several cancers [24]. TIMER database analysis demonstrated significant positive correlation between the ten-IRG signature and the infiltration of CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells. This suggests that higher tumor infiltration of immune cell types indicates CRC patients

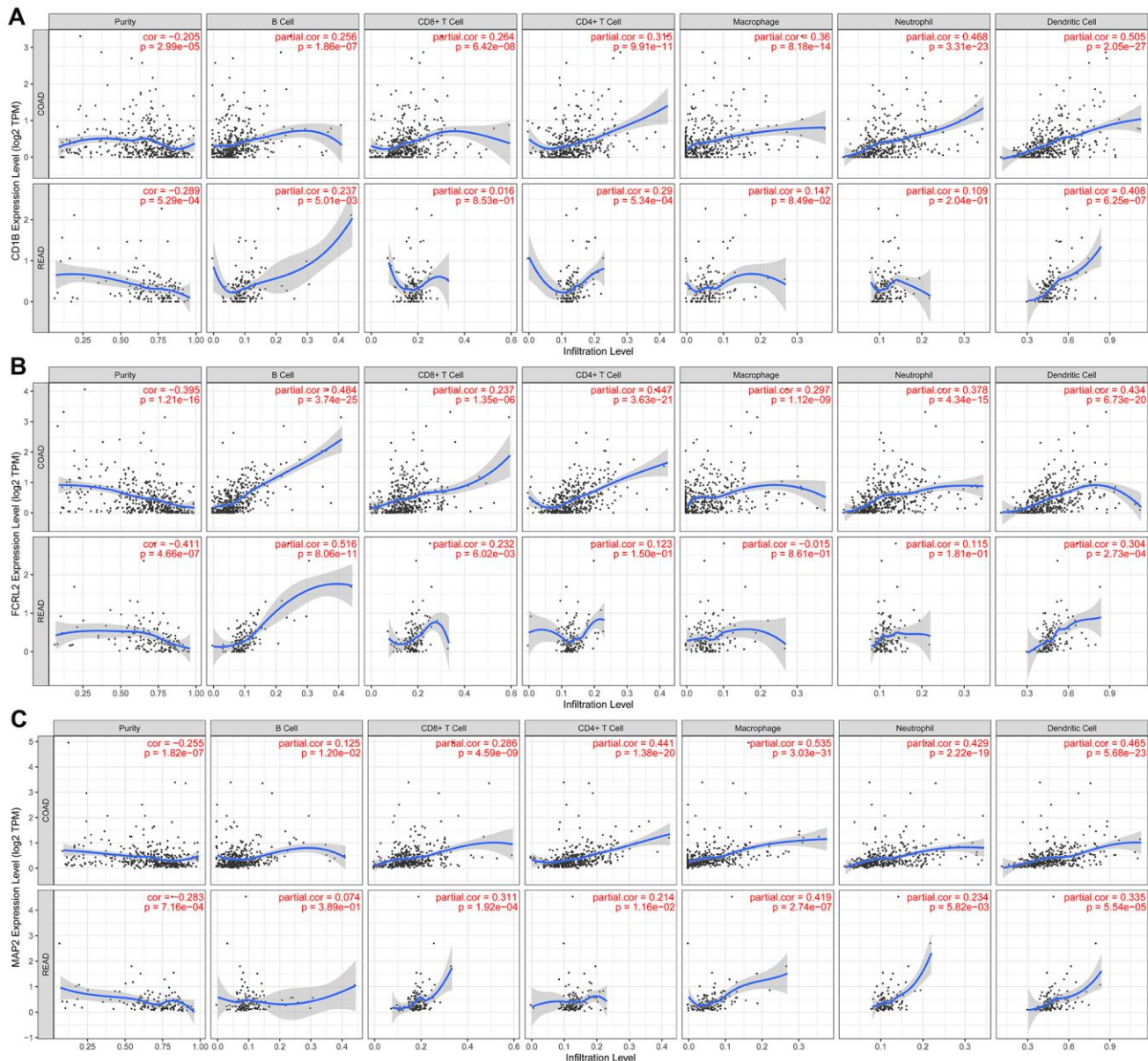


Figure 9. The expression of specific prognostic IRGs correlates with the abundance of tumor-infiltrated immune cell types in the CRC tissues. (A) The correlation analysis between *CD1B* gene expression and the proportions of six types of immune cells (B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells) in the CRC tissues. **(B)** The correlation analysis between *FCRL2* gene expression and the proportions of six types of immune cells (B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells) in the CRC tissues. **(C)** The correlation analysis between *MAP2* gene expression and the proportions of six types of immune cells (B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells) in the CRC tissues.

Table 1. Correlation analysis between the ten prognostic IRGs and immune cell infiltration in the CRC tissues.

IRGs	CRC TYPES	Tumor purity		B Cells		CD8 ⁺ T Cells		CD4 ⁺ T Cells		Macrophages		Neutrophils		Dendritic cells	
		Cor	P	Cor	P	Cor	P	Cor	P	Cor	P	Cor	P	Cor	P
A2ML1	COAD	-0.156	0.002	-0.103	0.039	-0.016	0.743	0.136	0.006	0.048	0.339	0.019	0.705	0.025	0.624
	READ	-0.029	0.73	-0.04	0.639	-0.119	0.162	-0.101	0.239	0.014	0.866	-0.005	0.957	0.115	0.179
CALB2	COAD	-0.291	2.26E-09	0.014	0.769	0.109	0.028	0.263	9.11E-08	0.331	8.51E-12	0.284	7.19E-09	0.299	8.76E-10
	READ	-0.334	5.49E-05	-0.081	0.345	-0.002	0.984	0.286	6.37E-04	0.321	1.16E-04	0.204	0.02	0.291	5.13E-04
CD1B	COAD	-0.205	2.99E-05	0.256	1.86E-07	0.264	6.42E-08	0.315	9.91E-11	0.36	8.18E-14	0.468	3.31E-23	0.505	2.05E-27
	READ	-0.289	5.29E-04	0.237	0.005	0.016	0.853	0.29	5.34E-04	0.147	0.085	0.109	0.204	0.408	6.25E-07
COL22A1	COAD	-0.265	5.40E-08	0.022	0.664	0.131	0.008	0.396	1.62E-16	0.473	6.29E-24	0.398	1.06E-16	0.395	1.91E-16
	READ	-0.296	3.84E-04	0.099	0.242	0.0006	0.994	0.411	5.07E-07	0.399	1.16E-06	0.254	0.002	0.299	3.50E-04
FCRL2	COAD	-0.395	1.21E-16	0.484	3.74E-25	0.237	1.35E-06	0.447	3.63E-21	0.297	1.12E-09	0.378	4.34E-15	0.434	6.73E-20
	READ	-0.411	4.66E-07	0.516	8.06E-11	0.232	0.006	0.123	0.15	-0.015	0.861	0.115	0.181	0.304	2.73E-04
HAND1	COAD	-0.176	3.49E-04	-0.103	0.038	0.038	0.446	0.256	2.06E-07	0.377	4.04E-15	0.155	0.002	0.211	1.93E-05
	READ	-0.067	0.433	-0.016	0.856	-0.06	0.479	0.219	0.01	0.317	1.46E-04	-0.019	0.828	0.111	0.191
IDO1	COAD	-0.353	2.17E-13	0.248	4.42E-07	0.4	4.66E-17	0.286	5.44E-09	0.303	5.33E-10	0.638	3.97E-47	0.564	3.47E-35
	READ	-0.377	4.46E-06	0.25	0.003	0.342	3.82E-05	0.184	0.031	0.046	0.594	0.364	1.16E-05	0.567	3.38E-13
MAP2	COAD	-0.255	1.82E-07	0.125	0.012	0.286	4.59E-09	0.441	1.38E-20	0.535	3.03E-31	0.429	2.22E-19	0.465	5.68E-23
	READ	-0.283	7.15E-04	0.074	0.389	0.311	1.92E-04	0.214	0.012	0.419	2.74E-07	0.234	0.006	0.335	5.54E-05
NKAIN4	COAD	-0.274	1.92E-08	-0.07	0.161	-0.007	0.887	0.301	7.29E-10	0.341	1.90E-12	0.229	3.78E-06	0.273	2.79E-08
	READ	-0.325	9.17E-05	-0.08	0.327	-0.123	0.149	0.389	2.20E-06	0.454	1.93E-08	0.113	0.188	0.197	0.02
VAX2	COAD	-0.239	1.12E-06	-0.016	0.742	0.118	0.017	0.239	1.21E-06	0.352	3.26E-13	0.284	7.00E-09	0.285	5.61E-09
	READ	-0.108	0.203	-0.157	0.07	-0.119	0.163	0.206	0.015	0.139	0.104	0.012	0.89	0.096	0.261

with advanced stage cancer. Furthermore, these data confirm the regulatory role of tumor-infiltrated immune cells in CRC progression.

We observed significant correlation between the expression of several genes in the ten-IRG signature and tumor progression. Bagchi et al. demonstrated the anti-tumor potential of CD1b-autoreactive T cells and their potential in adoptive immunotherapy [25]. Guo et al. demonstrated that miR-582/CD1B regulates the function of dendritic cells and is associated with immunotherapeutic outcomes in patients with lung adenocarcinoma [26]. The role of CD1B in CRC is unclear. Chronic inflammation is a well-established risk factor for CRC [27]. IDO1 plays an important role in limiting adaptive immune responses in several inflammatory and malignant gut diseases including gastric, pancreatic, esophageal, and stromal tumors [28]. High IDO1 expression is associated with tumor progression and poor clinical outcomes in CRC patients [29]. IDO1 contributes to immune tolerance in colon cancer by suppressing CD8⁺ T cell responses [30]. Reversing IDO1-mediated immunosuppression improved responses to immunogenic chemotherapy in a subcutaneous colorectal tumor model by promoting dendritic cell maturation, increasing tumor infiltrating T lymphocytes, and decreasing the numbers of regulatory T cells [31]. Our study showed IDO1 may be a gene with Janus face. On one face, IDO1 expression was higher than 34 that will be as an oncogene (Figure 7). The other face, IDO1 will be a cancer suppressor gene when expression was between 1.06 and 34 (Supplementary Figure 4). And the significant correlation was confirmed between IDO1

expression levels and the numbers of tumor-infiltrating dendritic cells and CD8⁺ T cells.

There are several limitations in our study. Firstly, we performed this study by analyzing data of CRC patients from public databases with genetic algorithms. However, the findings of our study require further clinical validation. Secondly, we did not validate the selected IRGs independently. Therefore, further studies are necessary to establish the individual role of these IRGs in tumorigenesis and immunotherapy response in CRC.

In conclusion, we used the ESTIMATE algorithm to analyze clinical and transcriptome data of CRC patients from the TCGA and GEO databases and identified a ten-IRG signature that is significantly associated with the status of tumor-infiltrating immune cells and prognosis of CRC patients. Further investigations are necessary to determine the clinical utility of this ten-IRG prognostic signature.

MATERIALS AND METHODS

Data collection

We searched the GSE and TGCA databases to identify CRC datasets with sample sizes exceeding 50 subjects. The selection criteria included: (1) RNA-Seq or microarray data from transcriptome studies; and (2) clinical data including data regarding clinicopathological parameters such as TNM stages and OS data. Based on these parameters, we selected six datasets, namely GSE103479, GSE72970, GSE41258, GSE39582, GSE12945, and TCGA [32–37] for this study. This

included 1635 primary CRC patients that were staged according to the American Joint Committee on Cancer (AJCC) staging system. The clinicopathological data was downloaded from the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) and TCGA (<https://tcga-data.nci.nih.gov/tcga/>) databases. The ESTIMATE algorithm was used to evaluate tumor purity as well as stromal and immune scores [16]. The entire TCGA dataset was used for genomic analysis and 60% of the patients from this dataset were randomly selected as an internal validation cohort. The results were also validated using the largest CRC dataset GSE39582 as the external validation cohort. This study complied with the approved publication guidelines for the TCGA and GEO databases. Since we used the data from these public databases, we did not require the approval of the ethics committee from our university or the consent of the patients included in this study.

Identification of immune-related genes (IRGs)

The ESTIMATE algorithm was used to calculate tumor purity, stromal scores, and immune scores of the CRC patient samples, which were then classified into low and high groups using X-tile [38]. X-tile is a novel tool for assessing the biological relationships between potential biomarker genes and prognosis, and identifies optimal cut-points based on the gene expression values [38]. Then, we used the limma R package [39] to identify differentially expressed genes using the cutoff values set at false discovery rate (FDR) < 0.05 and \log_2 | fold change | > 1. Finally, we identified the prognostic immune-related genes (IRGs) by comparing their stromal scores, immune scores, and tumor purity.

Functional enrichment analysis of IRGs

We used the clusterProfiler R package [40] to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and identified the most enriched biological pathways and functions related to the IRGs. We used the false discovery rate (FDR) to determine the adjusted P values and set the statistical significance as FDR < 0.05 to identify the most enriched GO terms for the biological processes (BP), molecular functions (MF), and cellular components (CC) as well as the KEGG pathways related to the IRGs.

Establishment of the IRG prognostic signature for CRC

We investigated the association between IRGs and prognosis of CRC patients by performing univariate Cox regression analysis. $P < 0.05$ was considered

statistically significant. Based on this analysis, we selected a panel of IRGs, which were then evaluated using the LASSO Cox regression analysis with the R package “glmnet”. We then set up a multi-gene signature to evaluate the prognosis of the CRC patients after cross-validating 1000 times with standard error within one standard deviation of the minimum. We then established the most simplified (smallest perimeter) immune gene expression signature model and calculated the risk scores for all patients based on the sum of the corresponding coefficients and expression value of each gene in the prognostic IRG signature model.

TIMER database analysis

We used the deconvolution algorithm from the TIMER database (<https://cistrome.shinyapps.io/timer/>) to analyze the association between the prognostic risk-related IRGs and the tumor-infiltration levels of immune cells, such as B cells, CD4⁺ T cells, CD8⁺ T cells, macrophages, neutrophils, and dendritic cells [41].

Statistical analysis

The median risk score was used as the cutoff value to classify CRC patients from the GSE39582 and TCGA datasets into low- and high-risk groups. Kaplan-Meier survival curves and log-rank test were used to evaluate the differences in OS rates between the low- and high-risk groups. Univariate and multivariate analysis of clinicopathological parameters and the ten-IRG prognostic signature was performed to determine the prognostic factors for the CRC patients from the TCGA dataset. A nomogram was constructed with the prognostic clinicopathological factors to determine the OS of the CRC patients in the TCGA cohort. The performance of the nomogram was evaluated using C-index and calibration plots with the “rms” R package. The efficiency of the nomogram was validated with the internal validation TCGA cohort and the GSE39582 cohort. Two-tailed t-tests were used to determine statistical significance, which was set at $P < 0.05$. R software version 3.6.0 was used to perform all statistical analyses.

AUTHOR CONTRIBUTIONS

Wei Li, Guiqi Wang and Xia Jiang planned this study; Wei Li and Xiaojing Jin performed data analysis; Wei Li, Shang Guo, Xingkai Su, and Fei Xu drafted and revised the manuscript; Wei Li, Xiaojing Jin, and Guiqi Wang prepared all the figures and tables; Xia Jiang interpreted the data and reviewed the manuscript. All the authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

This study was performed using the data available at The Cancer Genome Atlas (TCGA; <https://tcga-data.nci.nih.gov/>) and Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) databases.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

FUNDING

This study was financially supported by the Young Top-notch Talent Foundation of Hebei Province Office of Education (Grant No. BJ2018042); The National Natural Science Foundation of Hebei (Grant No. H202006374); International Science and Technology Cooperation Program of Hebei (Grant Nos. YZ201802 and 2019YX006A); other foundation from Hebei province (Grant Nos. LNB201911, zh2018002, 20180222, 20190445); and the Spark Program of the First Hospital of Hebei Medical University (Grant No. XH201701).

REFERENCES

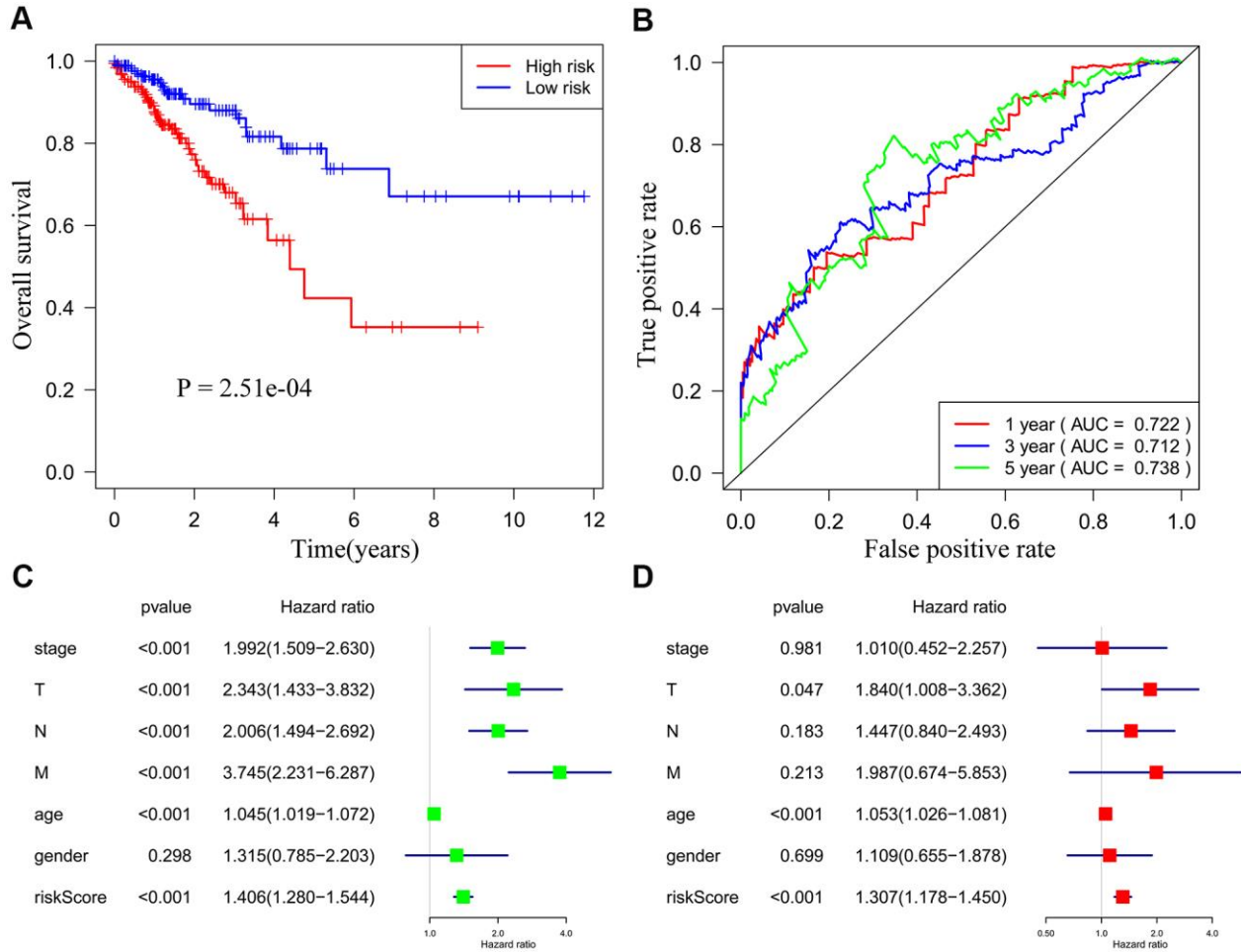
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:31912902
2. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:145–64. <https://doi.org/10.3322/caac.21601> PMID:32133645
3. 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
4. Hui L, Chen Y. Tumor microenvironment: sanctuary of the devil. *Cancer Lett.* 2015; 368:7–13. <https://doi.org/10.1016/j.canlet.2015.07.039> PMID:26276713
5. Bashir B, Snook AE. Immunotherapy regimens for metastatic colorectal carcinomas. *Hum Vaccin Immunother.* 2018; 14:250–54. <https://doi.org/10.1080/21645515.2017.1397244> PMID:29083978
6. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature.* 2013; 501:346–54. <https://doi.org/10.1038/nature12626> PMID:24048067
7. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol.* 2019; 16:361–75. <https://doi.org/10.1038/s41575-019-0126-x> PMID:30886395
8. Palucka AK, Coussens LM. The basis of oncoimmunology. *Cell.* 2016; 164:1233–47. <https://doi.org/10.1016/j.cell.2016.01.049> PMID:26967289
9. Kou Y, Koag MC, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. *J Am Chem Soc.* 2015; 137:14067–70. <https://doi.org/10.1021/jacs.5b10172> PMID:26517568
10. Wen R, Umeano AC, Kou Y, Xu J, Farooqi AA. Nanoparticle systems for cancer vaccine. *Nanomedicine (Lond).* 2019; 14:627–48. <https://doi.org/10.2217/nnm-2018-0147> PMID:30806568
11. Xiong Y, Wang K, Zhou H, Peng L, You W, Fu Z. Profiles of immune infiltration in colorectal cancer and their clinical significant: a gene expression-based study. *Cancer Med.* 2018; 7:4496–508. <https://doi.org/10.1002/cam4.1745> PMID:30117315
12. Ye L, Zhang T, Kang Z, Guo G, Sun Y, Lin K, Huang Q, Shi X, Ni Z, Ding N, Zhao KN, Chang W, Wang J, et al. Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer. *Front Immunol.* 2019; 10:2368. <https://doi.org/10.3389/fimmu.2019.02368> PMID:31681276
13. Lin K, Huang J, Luo H, Luo C, Zhu X, Bu F, Xiao H, Xiao L, Zhu Z. Development of a prognostic index and screening of potential biomarkers based on immunogenomic landscape analysis of colorectal cancer. *Aging (Albany NY).* 2020; 12:5832–57. <https://doi.org/10.18632/aging.102979> PMID:32235004
14. Bai J, Zhang X, Xiang ZX, Zhong PY, Xiong B. Identification of prognostic immune-related signature predicting the overall survival for colorectal cancer. *Eur Rev Med Pharmacol Sci.* 2020; 24:1134–41. https://doi.org/10.26355/eurev_202002_20164 PMID:32096169
15. Zhao X, Liu J, Liu S, Yang F, Chen E. Construction and validation of an immune-related prognostic model based on TP53 status in colorectal cancer. *Cancers (Basel).* 2019; 11:1722. <https://doi.org/10.3390/cancers11111722> PMID:31689990
16. 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/)
17. Li X, Gao Y, Xu Z, Zhang Z, Zheng Y, Qi F. Identification of prognostic genes in adrenocortical carcinoma microenvironment based on bioinformatic methods. *Cancer Med.* 2020; 9:1161–72.
<https://doi.org/10.1002/cam4.2774> PMID:[31856409](https://pubmed.ncbi.nlm.nih.gov/31856409/)
18. 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/)
19. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S, Sui J, Liang G. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J Cell Physiol.* 2020; 235:1025–35.
<https://doi.org/10.1002/jcp.29018> PMID:[31240705](https://pubmed.ncbi.nlm.nih.gov/31240705/)
20. Borst J, Ahrends T, Bąbała N, Melief CJ, Kastenmüller W. CD4⁺ T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol.* 2018; 18:635–47.
<https://doi.org/10.1038/s41577-018-0044-0> PMID:[30057419](https://pubmed.ncbi.nlm.nih.gov/30057419/)
21. Farhood B, Najafi M, Mortezaee K. CD8⁺ cytotoxic T lymphocytes in cancer immunotherapy: a review. *J Cell Physiol.* 2019; 234:8509–21.
<https://doi.org/10.1002/jcp.27782> PMID:[30520029](https://pubmed.ncbi.nlm.nih.gov/30520029/)
22. Netea-Maier RT, Smit JW, Netea MG. Metabolic changes in tumor cells and tumor-associated macrophages: a mutual relationship. *Cancer Lett.* 2018; 413:102–09.
<https://doi.org/10.1016/j.canlet.2017.10.037> PMID:[29111350](https://pubmed.ncbi.nlm.nih.gov/29111350/)
23. Komohara Y, Fujiwara Y, Ohnishi K, Takeya M. Tumor-associated macrophages: potential therapeutic targets for anti-cancer therapy. *Adv Drug Deliv Rev.* 2016; 99:180–85.
<https://doi.org/10.1016/j.addr.2015.11.009> PMID:[26621196](https://pubmed.ncbi.nlm.nih.gov/26621196/)
24. Ocana A, Nieto-Jiménez C, Pandiella A, Templeton AJ. Neutrophils in cancer: prognostic role and therapeutic strategies. *Mol Cancer.* 2017; 16:137.
<https://doi.org/10.1186/s12943-017-0707-7> PMID:[28810877](https://pubmed.ncbi.nlm.nih.gov/28810877/)
25. Bagchi S, Li S, Wang CR. CD1b-autoreactive T cells recognize phospholipid antigens and contribute to antitumor immunity against a CD1b⁺ T cell lymphoma. *Oncoimmunology.* 2016; 5:e1213932.
<https://doi.org/10.1080/2162402X.2016.1213932> PMID:[27757307](https://pubmed.ncbi.nlm.nih.gov/27757307/)
26. Guo J, Jin H, Xi Y, Guo J, Jin Y, Jiang D. The *miR-582/CD1B* axis is involved in regulation of dendritic cells and is associated with clinical outcomes in advanced lung adenocarcinoma. *Biomed Res Int.* 2020; 2020:4360930.
<https://doi.org/10.1155/2020/4360930> PMID:[32258122](https://pubmed.ncbi.nlm.nih.gov/32258122/)
27. Kou Y, Koag MC, Lee S. Promutagenicity of 8-chloroguanine, a major inflammation-induced halogenated DNA lesion. *Molecules.* 2019; 24:3507.
<https://doi.org/10.3390/molecules24193507> PMID:[31569643](https://pubmed.ncbi.nlm.nih.gov/31569643/)
28. Santhanam S, Alvarado DM, Ciorba MA. Therapeutic targeting of inflammation and tryptophan metabolism in colon and gastrointestinal cancer. *Transl Res.* 2016; 167:67–79.
<https://doi.org/10.1016/j.trsl.2015.07.003> PMID:[26297050](https://pubmed.ncbi.nlm.nih.gov/26297050/)
29. Ferdinande L, Decaestecker C, Verset L, Mathieu A, Moles Lopez X, Negulescu AM, Van Maerken T, Salmon I, Cuvelier CA, Demetter P. Clinicopathological significance of indoleamine 2,3-dioxygenase 1 expression in colorectal cancer. *Br J Cancer.* 2012; 106:141–47.
<https://doi.org/10.1038/bjc.2011.513> PMID:[22108515](https://pubmed.ncbi.nlm.nih.gov/22108515/)
30. Lou Q, Liu R, Yang X, Li W, Huang L, Wei L, Tan H, Xiang N, Chan K, Chen J, Liu H. miR-448 targets IDO1 and regulates CD8⁺ T cell response in human colon cancer. *J Immunother Cancer.* 2019; 7:210.
<https://doi.org/10.1186/s40425-019-0691-0> PMID:[31391111](https://pubmed.ncbi.nlm.nih.gov/31391111/)
31. Huang H, Jiang CT, Shen S, Liu A, Gan YJ, Tong QS, Chen SB, Gao ZX, Du JZ, Cao J, Wang J. Nanoenabled reversal of IDO1-mediated immunosuppression synergizes with immunogenic chemotherapy for improved cancer therapy. *Nano Lett.* 2019; 19:5356–65.
<https://doi.org/10.1021/acs.nanolett.9b01807> PMID:[31286779](https://pubmed.ncbi.nlm.nih.gov/31286779/)
32. Cancer Genome Atlas Network. 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/)
33. Staub E, Groene J, Heinze M, Mennerich D, Roepcke S, Klamann I, Hinzmann B, Castanos-Velez E, Pilarsky C, Mann B, Brümmendorf T, Weber B, Buhr HJ, Rosenthal A. An expression module of WIPF1-coexpressed genes identifies patients with favorable prognosis in three tumor types. *J Mol Med (Berl).* 2009; 87:633–44.
<https://doi.org/10.1007/s00109-009-0467-y> PMID:[19399471](https://pubmed.ncbi.nlm.nih.gov/19399471/)

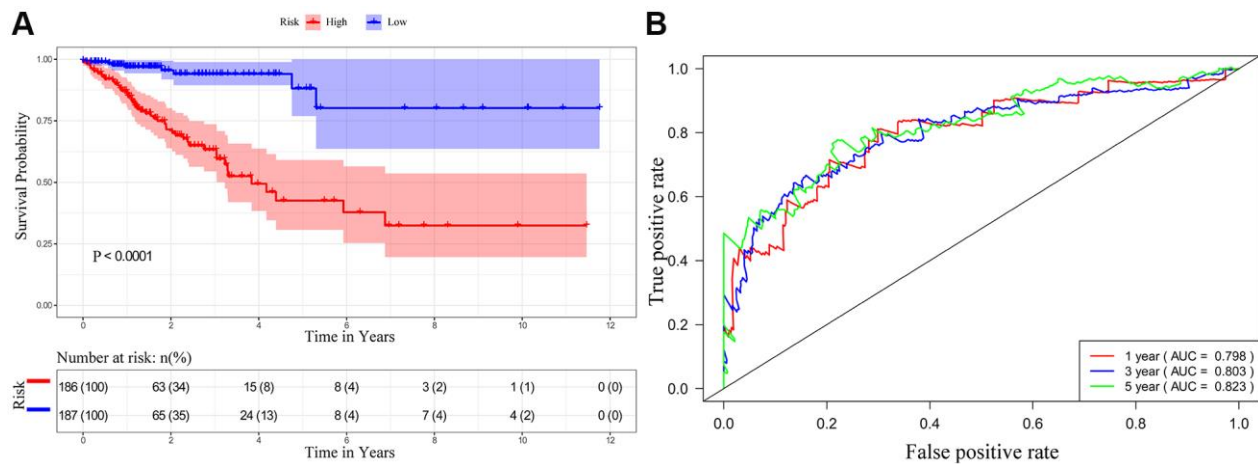
34. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Fléjou JF, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013; 10:e1001453.
<https://doi.org/10.1371/journal.pmed.1001453>
PMID:[23700391](https://pubmed.ncbi.nlm.nih.gov/23700391/)
35. Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, Paty PB, Gerald WL, Notterman DA, Domany E. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci USA.* 2009; 106:7131–36.
<https://doi.org/10.1073/pnas.0902232106>
PMID:[19359472](https://pubmed.ncbi.nlm.nih.gov/19359472/)
36. Del Rio M, Mollevi C, Bibeau F, Vie N, Selves J, Emile JF, Roger P, Gongora C, Robert J, Tubiana-Mathieu N, Ychou M, Martineau P. Molecular subtypes of metastatic colorectal cancer are associated with patient response to irinotecan-based therapies. *Eur J Cancer.* 2017; 76:68–75.
<https://doi.org/10.1016/j.ejca.2017.02.003>
PMID:[28284171](https://pubmed.ncbi.nlm.nih.gov/28284171/)
37. Allen WL, Dunne PD, McDade S, Scanlon E, Loughrey M, Coleman H, McCann C, McLaughlin K, Nemeth Z, Syed N, Jithesh P, Arthur K, Wilson R, et al. Transcriptional subtyping and CD8 immunohistochemistry identifies poor prognosis stage II/III colorectal cancer patients who benefit from adjuvant chemotherapy. *JCO Precis Oncol.* 2018; 2018:PO.17.00241.
<https://doi.org/10.1200/PO.17.00241>
PMID:[30088816](https://pubmed.ncbi.nlm.nih.gov/30088816/)
38. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bioinformatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res.* 2004; 10:7252–59.
<https://doi.org/10.1158/1078-0432.CCR-04-0713>
PMID:[15534099](https://pubmed.ncbi.nlm.nih.gov/15534099/)
39. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-seq 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/)
40. 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/)
41. 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

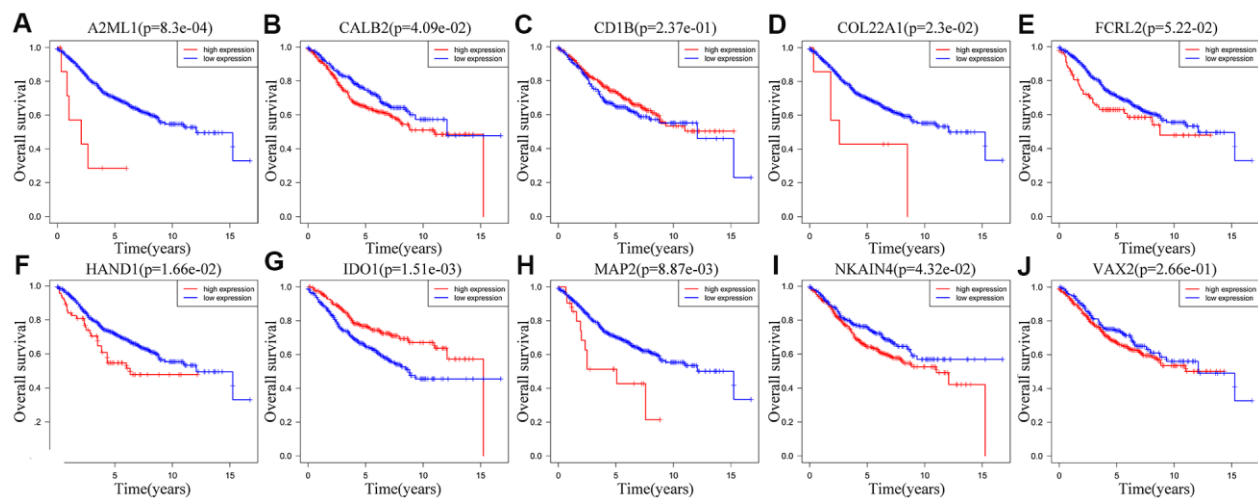
Supplementary Figures



Supplementary Figure 1. Validation of the ten-IRG signature in the internal validation cohort. (A) Kaplan–Meier survival curves show the overall survival of high- and low-risk CRC patients in the internal validation cohort based on the ten-IRG prognostic signature. (B) Time-dependent ROC curves show accuracy of the overall survival prediction based on the ten-IRG prognostic signature in the internal validation cohort. (C) Univariate Cox regression analysis shows the clinicopathological parameters associated with overall survival of CRC patients in the internal validation cohort. (D) Multivariate Cox regression analysis shows the clinicopathological parameters associated with overall survival of CRC patients in the internal validation cohort.

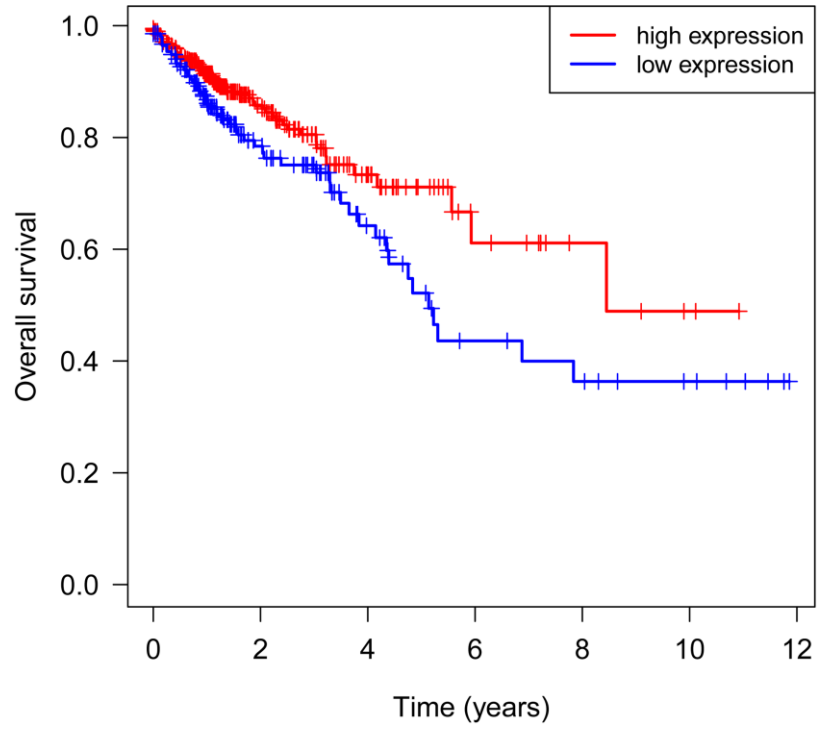


Supplementary Figure 2. Validation of the nomogram in accurately predicting overall survival of CRC patients in the internal validation cohort. (A) Kaplan–Meier survival curves show the overall survival of high-and low-risk CRC patients in the internal validation cohort based on the nomogram. (B) Time-dependent ROC curves show the accuracy of overall survival prediction based on the nomogram in CRC patients from the internal validation cohort.



Supplementary Figure 3. Correlation analysis between expression levels of individual prognostic IRGs and OS of CRC patients in the GSE39582 cohort. Kaplan-Meier survival plots show the OS of CRC patients in the GSE39582 cohort with high (red line) and low (blue line) expression of the ten individual IRGs. (A), Kaplan-Meier survival curves for A2ML1. (B), Kaplan-Meier survival curves for CALB2. (C), Kaplan-Meier survival curves for CD1B. (D), Kaplan-Meier survival curves for COL22A1. (E), Kaplan-Meier survival curves for FCRL2. (F), Kaplan-Meier survival curves for HAND1. (G), Kaplan-Meier survival curves for IDO1. (H), Kaplan-Meier survival curves for MAP2. (I), Kaplan-Meier survival curves for NKAIN4. (J), Kaplan-Meier survival curves for VAX2. Note: $p < 0.05$ according to the log-rank test.

IDO1($p=2.44e-02$)



Supplementary Figure 4. Kaplan-Meier survival plot of IDO1 (cut-off value=1.06).

Supplementary Table

Supplementary Table 1. Top 20 KEGG pathways and GO terms enriched by the IRGs.

Categories	Term	Description	Count	FDR
KEGG pathways	hsa04060	Cytokine-cytokine receptor interaction	68	4.11E-21
	hsa04062	Chemokine signaling pathway	45	1.51E-14
	hsa04151	PI3K-Akt signaling pathway	42	0.000122
	hsa04145	Phagosome	41	3.35E-15
	hsa04061	Viral protein interaction with cytokine and cytokine receptor	40	3.85E-21
	hsa05152	Tuberculosis	39	2.31E-11
	hsa05150	Staphylococcus aureus infection	38	2.70E-20
	hsa04640	Hematopoietic cell lineage	38	7.47E-20
	hsa04514	Cell adhesion molecules (CAMs)	37	9.65E-13
	hsa04380	Osteoclast differentiation	35	3.21E-13
	hsa05323	Rheumatoid arthritis	33	5.13E-16
	hsa05140	Leishmaniasis	26	2.94E-12
	hsa05322	Systemic lupus erythematosus	26	1.23E-06
	hsa05166	Human T-cell leukemia virus 1 infection	24	0.01797
	hsa04610	Complement and coagulation cascades	23	1.06E-08
	hsa04620	Toll-like receptor signaling pathway	23	5.17E-07
	hsa05164	Influenza A	23	0.001474
	hsa04510	Focal adhesion	23	0.012233
	hsa04015	Rap1 signaling pathway	23	0.020275
	hsa05144	Malaria	22	4.66E-13
GO BP	GO:0002694	regulation of leukocyte activation	100	2.45E-33
	GO:0050900	leukocyte migration	98	1.54E-33
	GO:0042110	T cell activation	86	3.07E-27
	GO:0002521	leukocyte differentiation	86	1.34E-24
	GO:0001819	positive regulation of cytokine production	85	1.89E-26
	GO:0042119	neutrophil activation	85	3.42E-23
	GO:0043062	extracellular structure organization	83	3.31E-29
	GO:0050727	regulation of inflammatory response	82	1.68E-22
	GO:0002446	neutrophil mediated immunity	82	2.31E-21
	GO:0002283	neutrophil activation involved in immune response	81	2.31E-21
	GO:0043312	neutrophil degranulation	80	5.85E-21
	GO:0030198	extracellular matrix organization	78	7.27E-30
	GO:0051249	regulation of lymphocyte activation	77	3.24E-24
	GO:0050867	positive regulation of cell activation	75	4.78E-29
	GO:0032103	positive regulation of response to external stimulus	73	7.40E-28
	GO:0002683	negative regulation of immune system process	72	4.76E-18
	GO:0007159	leukocyte cell-cell adhesion	71	2.26E-25
	GO:0002697	regulation of immune effector process	71	5.28E-18
	GO:0002696	positive regulation of leukocyte activation	70	1.60E-26
	GO:0022407	regulation of cell-cell adhesion	70	2.12E-20
GO CC	GO:0031012	extracellular matrix	107	1.76E-41
	GO:0098552	side of membrane	62	3.05E-17
	GO:0030667	secretory granule membrane	60	3.28E-20
	GO:0098797	plasma membrane protein complex	49	1.32E-06
	GO:0009897	external side of plasma membrane	48	2.95E-19
	GO:0070820	tertiary granule	40	2.38E-16

	GO:0005788	endoplasmic reticulum lumen	39	5.17E-07
	GO:0045121	membrane raft	38	1.69E-06
	GO:0098857	membrane microdomain	38	1.74E-06
	GO:0098589	membrane region	38	3.53E-06
	GO:0043025	neuronal cell body	38	0.004156
	GO:0030139	endocytic vesicle	37	9.93E-07
	GO:0060205	cytoplasmic vesicle lumen	37	2.55E-05
	GO:0031983	vesicle lumen	37	2.65E-05
	GO:0034774	secretory granule lumen	36	2.13E-05
	GO:0042581	specific granule	35	8.24E-13
	GO:0043235	receptor complex	35	7.29E-06
	GO:0031252	cell leading edge	32	0.009968
	GO:0005581	collagen trimer	31	7.80E-18
	GO:0005925	focal adhesion	31	0.027118
GO MF	GO:0030545	receptor regulator activity	60	3.55E-12
	GO:0048018	receptor ligand activity	59	7.02E-13
	GO:0030246	carbohydrate binding	54	1.28E-18
	GO:0005539	glycosaminoglycan binding	53	7.93E-23
	GO:0005201	extracellular matrix structural constituent	52	5.50E-27
	GO:1901681	sulfur compound binding	44	6.34E-14
	GO:0008201	heparin binding	41	2.33E-18
	GO:0005126	cytokine receptor binding	37	3.10E-10
	GO:0005125	cytokine activity	35	7.02E-13
	GO:0008047	enzyme activator activity	35	0.010079
	GO:0001664	G-protein coupled receptor binding	34	9.06E-09
	GO:0033218	amide binding	31	8.09E-05
	GO:0042277	peptide binding	29	1.07E-05
	GO:0004175	endopeptidase activity	29	0.008894
	GO:0003779	actin binding	29	0.037972
	GO:0005178	integrin binding	25	5.00E-09
	GO:0019955	cytokine binding	23	9.95E-09
	GO:0005518	collagen binding	21	2.15E-11
	GO:0019838	growth factor binding	21	6.77E-06
	GO:0031406	carboxylic acid binding	21	0.00145