European Review for Medical and Pharmacological Sciences 2020; 24: 1134-1141

Identification of prognostic immune-related signature predicting the overall survival for colorectal cancer

J. BAI[1,3,4], X. ZHANG[2,3,4], Z.-X. XIANG[1,3,4], P.-Y. ZHONG[1,3,4], B. XIONG[1,3,4]

[1]Department of Gastrointestinal Surgery & Department of Gastric and Colorectal Surgical Oncology, Zhongnan Hospital of Wuhan University, Wuhan, China
[2]Department of Radiation Oncology and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan, China
[3]Hubei Key Laboratory of Tumor Biological Behaviors, Wuhan, China
[4]Hubei Cancer Clinical Study Center, Wuhan, China

Jian Bai and *Xue Zhang* contributed equally to this work

Abstract. – OBJECTIVE: The morbidity and mortality of patients with colorectal cancer, one of the most common malignant tumors worldwide, is steadily increasing. The aim of this study was to investigate the association between prognostic immune-related gene profile and the outcome of colorectal cancer in patients by analyzing datasets from The Cancer Genome Atlas (TCGA).

MATERIALS AND METHODS: Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) further demonstrated that these genes were enriched in many immune-related biological processes. Univariate Cox regression analysis was applied to examine the association of immune-related genes with the prognosis in patients with colorectal cancer. The least absolute shrinkage and selection operation (LASSO) Cox regression model was then used to establish the immune-related signature for the prognostic evaluation of colorectal cancer in patients. Survival differences were assessed by the Kaplan-Meier method along with the log-rank test.

RESULTS: A total of 133 prognostic immune-related signatures were identified by using the univariate Cox proportional hazards regression analysis. A 14-gene signature-based risk score was constructed using the LASSO Cox regression. According to the cut-off of the risk-score, patients were assigned to the low-risk and high-risk groups. The log-rank test suggested that the survival time of the low-risk group was significantly higher than that of the high-risk group. In the time-dependent ROC curve analysis, the AUC for 1-year, 3-year, and 5-year overall survival (OS) were 0.781, 0.742, and 0.791, respectively. GO and KEGG analysis further revealed that the gene sets were actively involved in immune and inflammatory response, as well as the cytokine-cytokine receptor interaction pathway.

CONCLUSIONS: To summarize, we identified a novel 14-gene immune-related signature that may potentially serve as a prognostic predictor for colorectal cancer, thereby contributing to patient personalized treatment decisions. Further research needs to be conducted to validate the prognostic value of the selected genes.

Key Words:
Colorectal cancer, Immune-related signature, Prognosis, Lasso, Overall survival.

Abbreviations

TCGA, The Cancer Genome Atlas; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, False discovery rate; AUC, area under the curve; AJCC, American Joint Committee on Cancer; LASSO, least absolute shrinkage and selection operation; OS, overall survival.

Introduction

Colorectal cancer is the third most frequently diagnosed malignancy and the second leading cause of cancer-related death globally; it is also the third most common cancer in men and women[1,2]. Colorectal adenocarcinoma accounts for 90% of all colon cancer cases and is the most histological type. Early symptoms of colorectal cancer are not evident; the most common symptoms in patients with advanced disease include changes in bowel habits and fecal traits[3,4]. Despite significant devel-

Corresponding Author: Bin Xiong, MD, Ph.D; e-mail: binxiongwhu1961@163.com

opments in the treatment of colorectal cancer in recent years, its prognosis in patients remains poor due to a lack of early diagnostic and predictive biomarkers[2,3]. Therefore, identifying effective potential diagnostic markers and therapeutic targets to combat colorectal cancer is an urgent need.

It is well-known that various components of the immune system are related to the occurrence and development of cancer[5]. Various investigations[6-8] have verified that colorectal cancer is an immunogenic tumor and immunotherapy is strongly pursued *via* targeting the immune checkpoints. Additionally, the normalization of immune-microenvironment has an effect on improving the other anti-tumor treatments, including targeted therapy, radiotherapy, as well as chemotherapy[9]. A variety of immune-relevant gene signatures have been reported to be associated with the sensitivity of various chemotherapeutic drugs[10]. Notwithstanding, there has been no immune-relevant gene signature that can systematically assess and predict the prognosis of colorectal cancer in patients.

In this study, the transcriptome data and corresponding clinical follow-up information were applied to identify some key immune-related genes with significant prognostic value. Subsequently, we constructed a survival model to predict the prognosis of colorectal cancer in patients using these key immune-related genes.

Materials and Methods

Data Collection

The expression data of mRNA (514 samples, Workflow Type: HTSeq-Counts) and the corresponding clinical follow-up information were downloaded from The Cancer Genome Atlas (TCGA; https://cancergenome.nih.gov). The samples for which the gene expression was "zero" were excluded from the analysis. The list of immune-relevant genes consisting of 2498 genes was downloaded from the Molecular Signature Database V7.0 (MSigDB; http://software.broadinstitute.org/gsea/msigdb/)[11].

Data Processing

Colorectal cancer samples were randomly assigned into training and validation cohorts at a 7:3 ratio. Meanwhile, univariate Cox regression analysis was applied to examine the association between immune-related genes in relation to the prognosis in patients with colorectal cancer. In univariate Cox regression analysis, $p < 0.05$ was

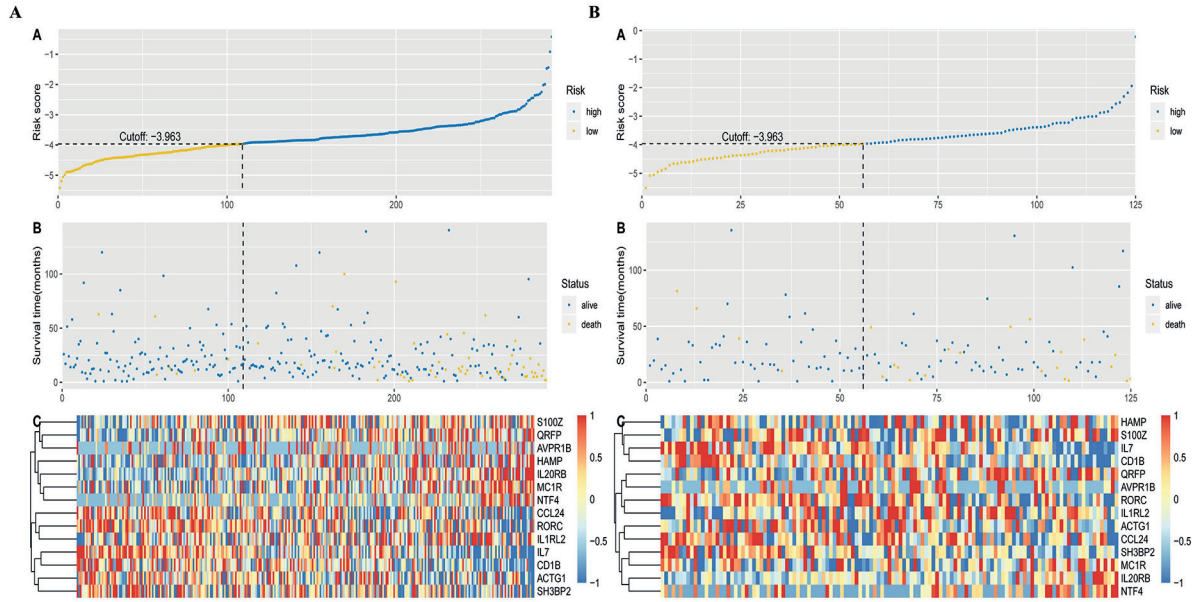

Figure 1. The prognostic signature in colorectal cancer. **A,** Training cohort: a. risk score of each colorectal cancer: the risk score increased from yellow to blue; b. survival time of each colorectal cancer: blue and yellow scatter represent alive and dead, respectively; c. heatmap of the 14-gene immune-related signature. **B,** Validation cohort: a. risk score of each colorectal cancer: the risk score increased from yellow to blue; b. survival time of each colorectal cancer: blue and yellow scatter represent alive and dead, respectively; c. heatmap of the 14-gene immune-related signature.

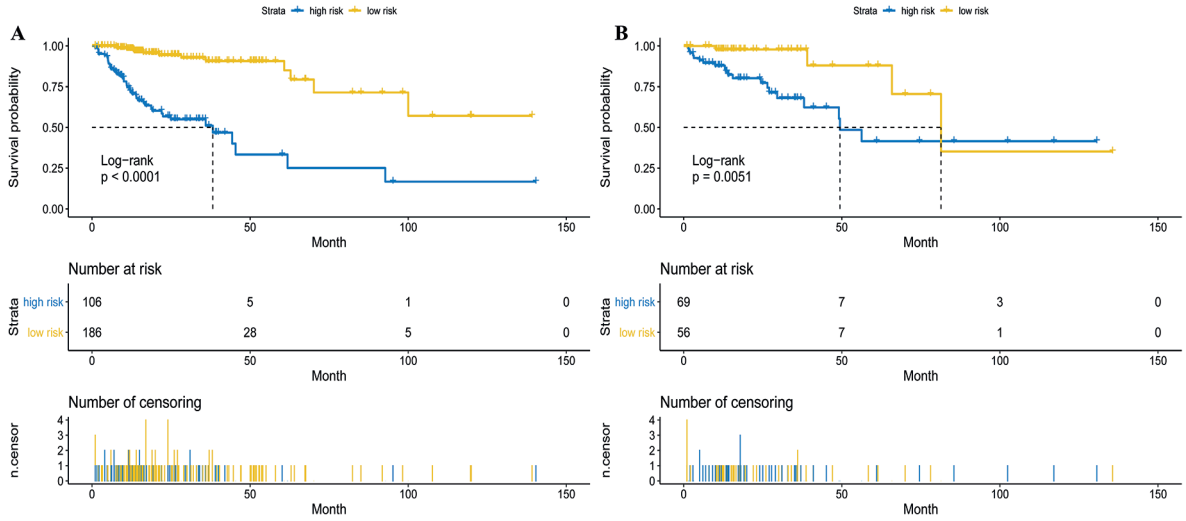

Figure 2. Kaplan-Meier curves of OS stratified by the 14-gene immune-related signature score in the high- and low-risk patients in (**A**) Training cohort and (**B**) Validation cohort.

considered to be statistically significant. After that, the LASSO Cox selection method was used to construct the survival-predicting model[12].

Enrichment Analyses

To better understand the potential function of immune-related genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed using the "clusterProfiler" R package[13]. False discovery rate (FDR) of less than 0.05 was set as the threshold.

Construction of the Prognostic Immune-Related Gene Signature

The immune-related prognostic risk score was constructed using the LASSO Cox selection method at a 10-fold cross-validation[14], by using the "glmnet" R package[15]. The risk score for each patient was calculated according to the signature gene expression weighted by its associated Cox regression coefficient. The prognostic immune-related gene signatures were shown as risk score = ($expr_{gene1}$ * $coefficient_{gene1}$) + ($expr_{gene2}$ * $coefficient_{gene2}$) +…+ ($expr_{genen}$ * $coefficient_{genen}$). The "surv_cutpoint" function of the "survminer" R package was applied to generate the optimal cut-off value of the risk-score. Based on the cut-off value of the risk-score, colorectal cancer patients were assigned into the low-risk group and high-risk group. Subsequently, the area under the curve (AUC) was calculated to validate the predictive ability of the immune-related risk signature, by using the "survivalROC" R package[16]. To assess the significance of the survival difference between the low-risk and high-risk group, the "survdiff" function of the "survival" R packages was applied[17]. Additionally, the differences in the other clinicopathological features between these two groups of patients were evaluated by the Chi-square test. All the analyses were conducted using R version 3.6.1. Significance was defined as $p < 0.05$.

Results

Construction and Assessment of the Immune-Related Gene Prognostic Signature

Colorectal cancer samples were randomly assigned into training and validation cohorts at a 7:3 ratio. Subsequently, the 2498 immune-related genes were analyzed using the univariate Cox regression model in the training set. A total of 133 probable prognostic immune-relevant genes were selected for further evaluation based on $p < 0.05$. We then utilized the LASSO Cox regression model with a 10-fold cross-validation for selecting genes with the best prognostic value. A total of 14 immune-related genes were identified and the risk score was calculated based on their expression level and associated Cox regression coefficient. The risk score = ($expr_{ACTG1}$ * -0.116) + ($expr_{SH3BP2}$ * -0.429) + ($expr_{CCL24}$ * -0.088) + ($expr_{RORC}$ * -0.132) + ($expr_{IL7}$ * -0.253) + ($expr_{MC1R}$ * 0.050) + ($expr_{IL1RL2}$ * 0.0975) + ($expr_{IL20RB}$ * 0.385) + ($expr_{QRFP}$ * 0.261)+ ($expr_{HAMP}$ * 0.153) + ($expr_{CD1B}$ * -0.364) + ($expr_{S100Z}$

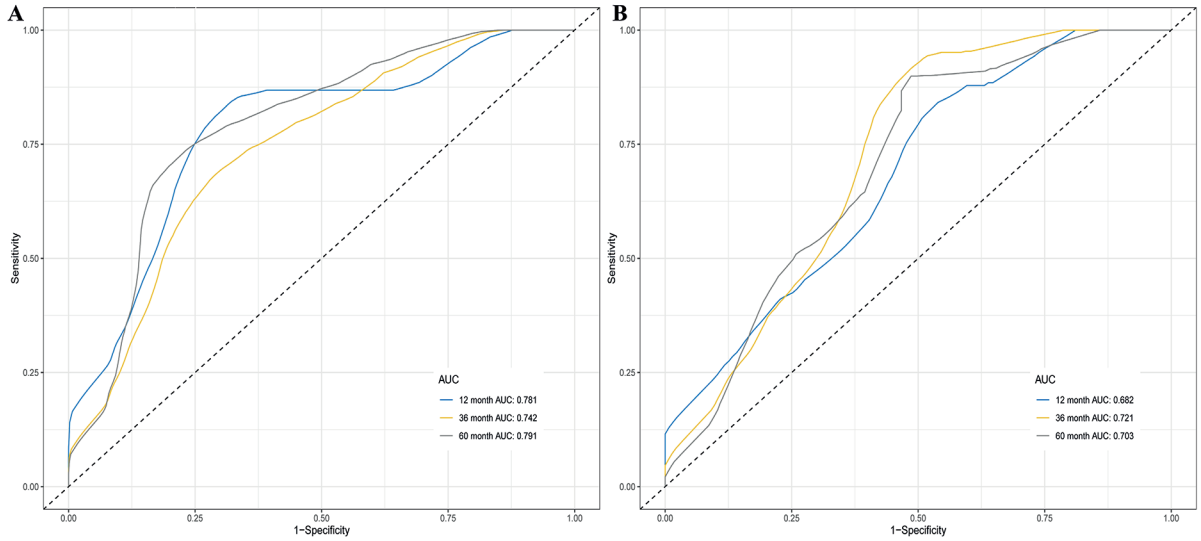

Figure 3. Time-dependent ROC curves of OS for the 14-gene immune-related signature score in the (**A**) Training cohort and (**B**) Validation cohort, at 1-, 3-, and 5-year period.

* 0.363) + ($expr_{NTF4}$ * 0.273) + ($expr_{AVPR1B}$ * 0.081). Based on the optimal cut-off value of -3.963 for the risk-score, patients were further assigned into the low-risk and high-risk group (Figure 1A). The Kaplan-Meier log-rank test demonstrated that the high-risk patients had a worse overall survival (OS) rate compared to that of low-risk patients in the training set (Figure 2A, $p < 0.001$). In the time-dependent ROC curve analysis, the AUC for 1-year, 3-year, and 5-year OS period were 0.781, 0.742, and 0.791, respectively (Figure 3A).

We further verified the prognostic ability of this 14-immune-related gene signature in the validation cohort. Based on the optimal cut-off value of -3.963 for the risk-score, patients were assigned into the low-risk and high-risk group (Figure 1B). The Kaplan-Meier log-rank test demonstrated that the low-risk patient group had significant survival advantages compared to those in the high-risk group in the validation set (Figure 2B, $p < 0.001$). The AUC for 1-year, 3-year, and 5-year OS period was 0.682, 0.721, and 0.703, respectively (Figure 3B).

These results demonstrated great applicability and stability of the immune-relevant gene signature for predicting prognosis in patients with colorectal cancer.

Function Enrichment Analysis

50 significant KEGG pathways and 1698 GO terms were identified *via* the enrichment analysis of 676 immune-relevant genes. The results showed that the top six KEGG pathways were hsa04060: Cytokine-cytokine receptor interaction, hsa04061: Viral protein interaction with cytokine and cytokine receptor, hsa04062: Chemokine signaling pathway, hsa04010: MAPK signaling pathway, hsa04640: Hematopoietic cell lineage, and hsa04080: Neuroactive ligand-receptor interaction. (Figure 4A). Further, pathway analysis according to GO identified various significant terms, the top six of them being 0050900: leukocyte migration, 0006959: humoral immune response, 0002526: acute inflammatory response, 0006958: complement activation, classical pathway, 0002920: regulation of humoral immune response, and 0002673: regulation of acute inflammatory response (Figure 4B).

Association Between Immune-Related Gene Signature and Clinical Parameters

To further confirm the clinical value of the survival-predicting model, the Chi-square test was applied to assess the association between the signature and clinical parameters, including age, gender, and American Joint Committee on Cancer (AJCC) staging system. In the training cohort, a higher-risk score was found to be associated significantly with age ($p = 0.001$), AJCC staging system ($p < 0.001$), T ($p = 0.001$), N ($p < 0.001$), and M ($p = 0.001$). However, no significant difference was found in gender ($p = 0.668$). Similar results were found in the validation cohort of colorectal cancer (Table I and Figure 5).

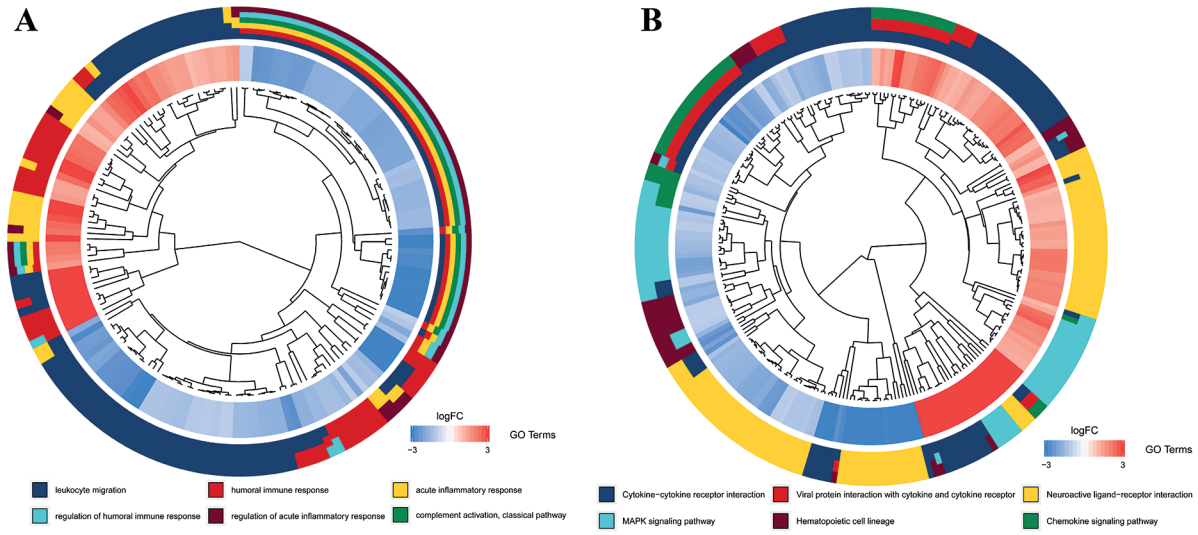

Figure 4. Gene functional enrichment of immune-related signature. **A,** The top six most significant GO terms. **B,** The top six most significant KEGG pathways. Function analysis of the 14-gene immune-related signature.

Discussion

Despite significant developments in the treatment of colorectal cancer in the past ten years, its prognosis in patients remains poor. In most of the patients diagnosed at an advanced stage, simple surgical resection treatment could not achieve satisfactory results, and had to be supplemented by radiotherapy or chemotherapy at the same time[18]. Considering the importance of immune-envi-

Table I. Correlation between the clinical features of colon carcinoma and 14 immune-relevant genes signature.

Parameters	Training cohort (n = 280)			Validation cohort (n = 123)		
	High risk	Low risk	*p*-value	High risk	Low risk	*p*-value
Age (years)			**0.001**			0.459
>65	57	103		44	32	
≤65	43	77		24	23	
Gender			0.668			0.250
Female	44	84		30	30	
Male	56	96		38	25	
AJCC stage			**<0.001**			**0.004**
I+II	39	118		32	40	
III+IV	61	62		36	15	
T			**0.001**			**0.005**
T0-2	9	46		10	14	
T3-4	91	134		58	41	
N			**<0.001**			**0.005**
N0	40	120		35	42	
N1-3	60	60		33	13	
M			**0.001**			**0.001**
M0	66	151		44	47	
M1	34	29		24	8	

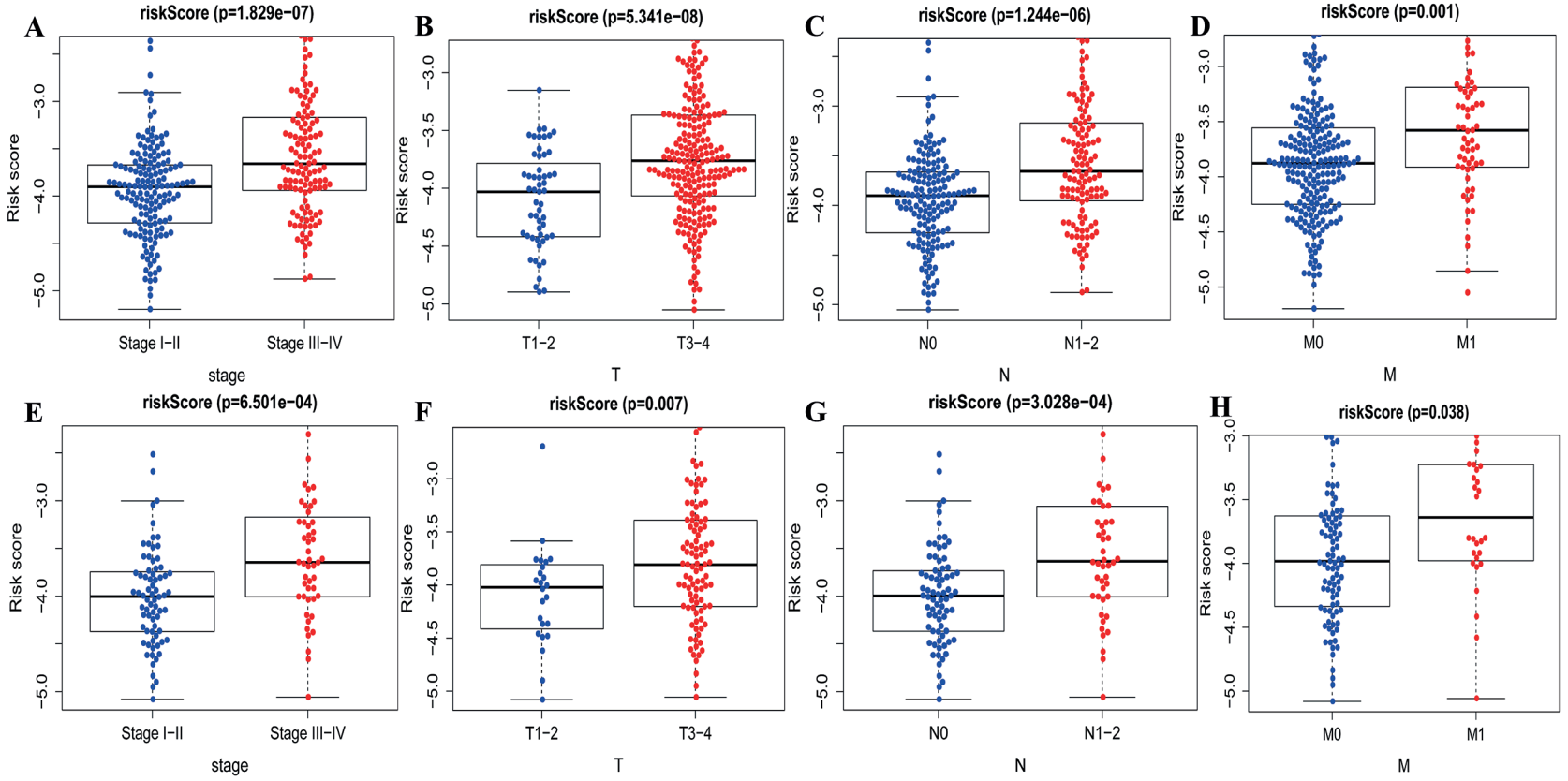

Figure 5. Association between the immune-related signature and clinical parameters in the Training cohort (**A-D**) and Validation cohort (**E-H**). Distribution of the risk score is stratified by age, AJCC staging system, T, N, and M.

ronment in the progression of cancer, the identification of immune-relevant biomarkers is very important for the prognosis of colorectal cancer in patients[19]. We therefore established a robust prognostic signature according to immune-relevant genes using the TCGA datasets to predict patients' survival outcomes.

The survival-predicting model revealed a total of 14 immune-related genes with a prognostic ability. We showed that the signature was associated significantly with the OS of patients with colorectal cancer in the training and validation cohort. These results demonstrated great applicability and stability of the immune-relevant gene signature for predicting prognosis in patients with colorectal cancer. We further evaluated the association between the signature and clinical parameters to figure out the clinical value of this immune-related signature. It was found that high-risk patients were associated significantly with age, AJCC staging system, T, N, and M. These results further highlight the prognostic ability of the immune-related signature.

There is growing evidence[20] that innate and adaptive immune systems make a crucial contribution to the occurrence and development of cancer. In this study, we utilized GO and KEGG analysis to better understand the potential function of immune-related genes. The results showed that these immune-related genes were actively involved in cytokine-cytokine receptor interaction, functioning as significant contributors in the inflammatory process of tumor occurrence and development[21]. These cytokines and cytokine receptors can directly or indirectly affect tumor cells in the tumor-microenvironment *via* chronic inflammatory reactions, free radicals, and signal pathways[22]. They can function to inhibit the development and progression of tumor and are also verified to be effective in the treatment of cancer[23,24]. Future researches might uncover their therapeutic potential in tumor immunotherapy by elucidating the mechanisms of cytokines and immune response.

In addition, the pathway enrichment analyses increased the evidence of its association with cancer and the clinical applicative potential. PI3K-Akt signal pathway is one of the key regulatory pathways in multiple cancers, including colorectal cancer. PI3K-Akt signaling pathway inhibitors can inhibit the growth of colorectal cancer xenografts and enhance the cytotoxic effect of cetuximab and panitumumab[25]. Furthermore, other pathways such as MAPK signaling pathway, Ras signaling pathway, Chemokine signaling pathway, and RAS signaling pathway were also involved in the growth, invasion, proliferation, and metastasis of cancer cells and played a key role in colorectal cancer[26-29].

To the best of our knowledge, this is the first study to focus on the association between prognostic immune-related genes and the outcome of colorectal cancer in patients. Nonetheless, this investigation has some limitations. First, the signature was established using retrospective data. Thus, it is necessary to carry out clinical verification using a sufficient number of colorectal cancer samples to confirm the clinical value of this survival prediction model. Second, due to a lack of patients treated with immune-checkpoint inhibitors, we could not verify the association between the prognostic immune-related signature and responses to tumor immunotherapy.

Conclusions

To summarize, we identified a novel 14-gene immune-related signature that may potentially serve as a prognostic predictor for colorectal cancer, thereby contributing to patient personalized treatment decisions. Future prospective studies are needed to verify the clinical value of this survival-predicting model in patients with colorectal cancer.

Funding

The present study was supported by the National Natural Science Foundation of China (81572874, 81872376).

Author Contributions

B.J and Z.X conceived, designed, analyzed the data, and wrote the manuscript; X.Z.X and Z.P.Y acquired the data and performed the statistical analysis; X.B provided administrative support and manuscript review. All authors read and approved this article.

Data Availability Statement

All data generated or analysed during this study are included in this paper. This article is based on data generated by the TCGA Research Network: https://cancergenome.nih.gov, and the data are also available in the TCGA.

Conflicts of interest

The authors declare no conflicts of interest.

References

1) Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019; 69: 7-34.
2) Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, Jemal A. Colorectal cancer statistics, 2017. CA Cancer J Clin 2017; 67: 177-193.
3) Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol 2019; 27: 1-20.
4) Cong N, Huang W, Yuan JP, Li GZ, Zhai GS, Li BS. Peroxiredoxin1 promotes cell proliferation, migration and invasion of colorectal cancer via p38MAPK signaling. Eur Rev Med Pharmacol Sci 2018; 22: 1922-1928.
5) Dougan M, Dranoff G. Immune therapy for cancer. Annu Rev Immunol 2009; 27: 83-117.
6) Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat Rev Immunol 2017; 17: 559-572.
7) Zaravinos A, Roufas C, Nagara M, de LMB, Oblovatskaya M, Efstathiades C, Dimopoulos C, Ayiomamitis GD. Cytolytic activity correlates with the mutational burden and deregulated expression of immune checkpoints in colorectal cancer. J Exp Clin Cancer Res 2019; 38: 364.
8) Bashir B, Snook AE. Immunotherapy regimens for metastatic colorectal carcinomas. Hum Vaccin Immunother 2018; 14: 250-254.
9) Palucka AK, Coussens LM. The basis of oncoimmunology. Cell 2016; 164: 1233-1247.
10) Pedrosa L, Esposito F, Thomson TM, Maurel J. The tumor microenvironment in colorectal cancer therapy. Cancers (Basel) 2019; 11. pii: E1172.
11) Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.
12) Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997; 16: 385-395.
13) Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-287.
14) Tibshirani R. Regression shrinkage and selection via the Lasso. J R Statist Soc B 1996; 58: 267-288.
15) Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw 2011; 39: 1-13.
16) Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000; 56: 337-344.
17) Lin H, Zelterman DJT. Modeling survival data: extending the Cox model. Technometrics 2000; 44: 85-86.
18) Moriarity A, O'Sullivan J, Kennedy J, Mehigan B, McCormick P. Current targeted therapies in the treatment of advanced colorectal cancer: a review. Ther Adv Med Oncol 2016; 8: 276-293.
19) Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature 2017; 541: 321-330.
20) Ogino S, Galon J, Fuchs CS, Dranoff G. Cancer immunology--analysis of host and tumor factors for personalized medicine. Nat Rev Clin Oncol 2011; 8: 711-719.
21) Lippitz BE. Cytokine patterns in patients with cancer: a systematic review. Lancet Oncol 2013; 14: e218-228.
22) Lin WW, Karin M. A cytokine-mediated link between innate immunity, inflammation, and cancer. J Clin Invest 2007; 117: 1175-1183.
23) Lee S, Margolin K. Cytokines in cancer immunotherapy. Cancers (Basel) 2011; 3: 3856-3893.
24) Santarpia L, Lippman SM, El-Naggar AK. Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. Expert Opin Ther Targets 2012; 16: 103-119.
25) Bahrami A, Khazaei M, Hasanzadeh M, ShahidSales S, Joudi MM, Farazestanian M, Sadeghnia HR, Rezayi M, Maftouh M, Hassanian SM, Avan A. Therapeutic potential of targeting PI3K/AKT pathway in treatment of colorectal cancer: rational and progress. J Cell Biochem 2018; 119: 2460-2469.
26) Tolaymat M, Larabee SM, Hu S, Xie G, Raufman JP. The role of M3 muscarinic receptor ligand-induced kinase signaling in colon cancer progression. Cancers (Basel) 2019; 11. pii: E308.
27) Gunawardene A, Dennett E, Larsen P. Prognostic value of multiple cytokine analysis in colorectal cancer: a systematic review. J Gastrointest Oncol 2019; 10: 134-143.
28) Cabrero-de LHS, Martinez-Balibrea E. CXC family of chemokines as prognostic or predictive biomarkers and possible drug targets in colorectal cancer. World J Gastroenterol 2018; 24: 4738-4749.
29) Wang Y, Yan XL, Tian SK. Downregulating long non-coding RNA CCAT5 inhibits tumor growth, invasion and metastasis in colorectal cancer through suppressing STAT3. Eur Rev Med Pharmacol Sci 2019; 23: 7899-7904.