Am J Transl Res 2020;12(9):5108-5130
www.ajtr.org /ISSN:1943-8141/AJTR0108446

Original Article

Construction of a prognostic model based on nine immune-related genes and identification of small molecule drugs for hepatocellular carcinoma (HCC)

Jiaxin Zhang[1,2*], Guang Chen[1,2*], Jiaying Zhang[3*], Peng Zhang[1,2], Yong'an Ye[1,2]

[1]Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China; [2]Institute of Liver Diseases, Beijing University of Chinese Medicine, Beijing, China; [3]Ministry of Education Key Laboratory of Bioinformatics, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. [*]Equal contributors.

Received January 29, 2020; Accepted July 3, 2020; Epub September 15, 2020; Published September 30, 2020

Abstract: This study aimed to develop a prognostic model for hepatocellular carcinoma (HCC) based on immune-related genes and to identify new potential small-molecule drugs. A differential gene expression analysis of high-throughput microarray data from The Cancer Genome Atlas (TCGA) was performed to identify immune-related genes. By comparison with an immune-related genome, nine genes with important prognostic value for HCC were identified. The prognostic characteristics were established based on univariate and multivariate COX and Lasso regression analyzes. Subsequently, immune-related HCC risk signatures were constructed based on these identified nine immune-related genes and patients were classified as being at high or low risk according to these signatures. The overall survival (OS) time of high-risk patients was significantly shorter than that of low-risk patients. When studied as an independent prognostic factor of HCC, the significant prognostic value of this feature can be seen in the stratified cohorts. For clinical application, it was developed a nomogram that includes nine clinical risk factors and the prognostic model built based on the identified immune-related genes. Internal and external verification on 243 HCC tissues through International Cancer Genome Consortium (ICGC) database were performed to make this model more accurate and reliable. In addition, it was observed a positive regulation between the identified immune-related genes and their transcription factors found in HCC patients. Moreover, physiological function and signaling pathway of identified immune-related genes were studied by GO and KEGG enrichment analysis. Finally, several new small molecular drugs with potential for the treatment of HCC have been identified in the CMap database.

Keywords: Immune gene, hepatocellular carcinoma, prognostic signature, small molecular drugs

Introduction

Hepatocellular carcinoma (HCC), also known as malignant liver cancer, is the most common cause of cancer-related death in the third world [1-3]. It is a multifactorial disease whose main risk factors worldwide are viral hepatitis and excessive alcohol consumption [4]. Non-alcoholic fatty liver, diabetes, aflatoxin, and immune-related diseases, such as autoimmune hepatitis and primary biliary cirrhosis, are other common risk factors for HCC [5]. Chronic liver disease and cirrhosis are the main limitations of liver cancer treatments [6, 7]. Surgical resection is a viable option for early HCC, but there are high mortality and morbidity risks when using this method due to underlying condition of the cirrhotic liver [8]. Liver transplantation is also a viable option for treatment of HCC and unresectable primary cirrhosis. However, its adequacy is limited by tumor stage, the health status of the patients, the lack of a suitable liver and the possible problems with graft rejection [9]. In addition, HCC is usually diagnosed at a later stage, which results in less efficient surgical resection or liver transplantation and a lower survival rate [10]. Sorafenib is the only drug currently approved for the treatment of advanced HCC patients, but it can only be used in patients with good liver function and provides only a small increase in mean survival time [11, 12]. Although many studies on the clinical and molecular types of tumors have been performed, the mortality associated with HCC is

still comparable to its morbidity. Therefore, there is an urgent need for the development of new treatments.

Currently, it is believed that tumor-promoting immune disease is the cause of the development of liver cancer. HCC cells can stimulate an obvious immune response and provide an appropriate microenvironment for their development [13]. Immunotherapy as a supplementary treatment for HCC has been increasingly studied due to the poor prognosis after standard treatments [14]. In addition, several immune-related parameters have been reported to predict the prognosis of HCC patients, further emphasizing the importance of immunological status in determining the HCC prognosis [13, 15].

High-throughput gene microarray can analyze cancer and non-cancer samples on a large scale. This technique can identify tumor-related genes at multiple levels; from genes related to molecular diagnosis and pathological classification to genes related to therapeutic evaluation and prognosis prediction. In addition, it can indicate genes that can be markers of drug sensitivity and neoplasm recurrence [16-18]. In order to identify immune genes related to prognosis in HCC, we carried out a study of immune-related genes. Here, a differential gene expression analysis of high-throughput microarray data of normal liver and HCC tissues was performed. Nine immune-related genes were identified and they can be used as biomarkers genes for HCC. A predictive model for HCC was built based on these genes and including different clinical factors. This model showed a good accuracy in the overall survival prediction.

Materials and methods

Data acquisition and preparation

Data from the transcriptome analysis and related clinical information from the HCC were obtained from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/). The ImmPort database (https://www.immport.org/shared/home) is a dedicated database that contains human immune-related genes. The RNA-seq data obtained from TCGA were compared with the ImmPort database and standardized using the R programming language.

Identification and enrichment analysis of differently expressed immune-related genes

The "Lima" package in the R statistical software was used to evaluate the differential expression of immune-related genes between HCC and non-tumour samples. The threshold for recognizing differentially expressed immune-related genes was set to the adjusted *P*-value (FDR) less than 0.05 and at least 2 times the change of (FCs). Gene functional enrichment analyses were performed, including gene ontology (GO) functional annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, as well as analyzes of the biological function, cell localization and signaling pathway of the target gene. "Clusterprofiler" R package was used to analyze the enrichment of the differentially expressed immune-related genes provided by GO and KEGG.

Identification of differently expressed TFs and construction of a network between TFs and differentially expressed immune-related genes

The Cistrome database (http://www.cistroes.org/) is a comprehensive resource for predicting the profile of transcription factors (TFs) targets and enhancers in cancer. The prediction results come from the comprehensive analysis of the TCGA expression profile and ChIP-seq open spectrum. The difference of TFs expression between HCC and non-tumor samples was analyzed by "Lima" package in the R statistical software. The correlation between the differential expression of TFs and differentially expressed immune-related genes was tested by the language R. Correlations were considered significant if they presented a correlation coefficient of at least 0.4 and a *P*-value less than 0.01.

Establishment of an independent prognostic index (risk score) based on differentially expressed immune-related genes

In order to identify the key differentially expressed immune-related genes, we performed univariate and multivariate COX and Lasso regression analyzes, excluding some differentially expressed immune-related genes with little prognostic value. The correlation coefficient in the model formula to predict the prognosis of HCC patients was obtained according to the weight of each gene in multivariate COX regres-

sion analysis. These data were combined with the expression of various prognosis-related genes to establish an independent prognostic model (Risk score; RS). Thus, RS was calculated using the following formula $\beta_1 \times gene_1$ expression + $\beta_2 \times gene_2$ expression + ····· + $\beta_n \times gene_n$ expression, where β is the corresponding correlation coefficient.

Prognostic index evaluation

Each HCC patient received a RS according to the constructed prognostic model. The median RS was used as the cutoff value to divide HCC patients into high and low-risk groups. The survival curve was plotted by Kaplan-Meier (K-M) method and the difference of survival rate between the low and high-risk groups was evaluated by the log-rank test. The working characteristic curve (receiver operating characteristic curves, ROC), was established by "survival ROC" package to calculate the area under the curve (AUC) value and to evaluate the specificity and sensitivity of the model. The working characteristic curve also showed the risk score distribution, the patient examined number and the heatmap of the prognosis-related differentially expressed immune-related genes in different risk groups. The "rms" package of R software was used based on the Cox proportional hazard regression model, to visualize the prognostic nomogram of HCC patients and to show the relationship between individual predictors and survival rates.

To further evaluate whether the constructed model can be used as an independent prognostic factor, age, gender, race, grade, stage, T, M, N, and RS was used as independent variables. Then, changes of survival time and survival outcome were analyzed by univariate and multivariate cox regression analysis, respectively. In addition, a clinical correlation analysis of the correlation between RS and clinical characteristics, such as age, gender, race, grade, stage, T, M, N was performed. Finally, the correlation between these clinical characteristics and each prognosis-related differentially expressed immune-related genes was also evaluated.

Prognostic index verification

After evaluating the constructed prognostic model, its progress was verified. For this, internal and external verification it was performed on 243 HCC tissues through International Cancer Genome Consortium (ICGC) database to make the model more accurate and reliable. Each HCC patient in ICGC received a RS based on previous established prognostic index. The median RS was used as a dividing line between the high and the low-risk groups of HCC patients. The survival curve was drawn by K-M method and the difference of survival rate between high and low-risk groups was verified by log-rank test. The working characteristic curve (receiver operating characteristic curves, ROC) was established by the "survival ROC" program to calculate the area under the curve (AUC) value and to verify the specificity and sensitivity of the established model. Moreover, a heatmap was drawn to show the RS distribution, the number of patients examined and the differential expression of immune-related genes related to prognosis in different risk groups. The "rms" package of R software was used based on the Cox proportional hazard regression model to visualize the prognostic nomogram of HCC patients and to verify the relationship between individual predictors and survival rates.

Identification of small bioactive molecules candidates for HCC treatment

Connectivity Map (CMap) was used to identify potential small-molecule drugs (SMDs) for HCC treatment [19]. CMap is a database of gene expression profiles from different cultured human cells treated with small bioactive molecules. CMap is a useful tool for analyzing connections between SMDs and disease-related genes expression, in order to provide rapid identification of potential disease-related SMDs and their structure and possible mechanisms of action.

Statistical analysis

Univariate Cox regression analysis was utilized to evaluate gene expression values and overall survival time associations and the Lasso regression analysis was used to prevent the occurrence of overfitting. Subsequently, multivariate Cox regression analysis was utilized to construct the RS based on differently expressed immune-related genes. The nine-gene risk signature was finally constructed by weighting the estimated Cox regression coefficients. Diffe-

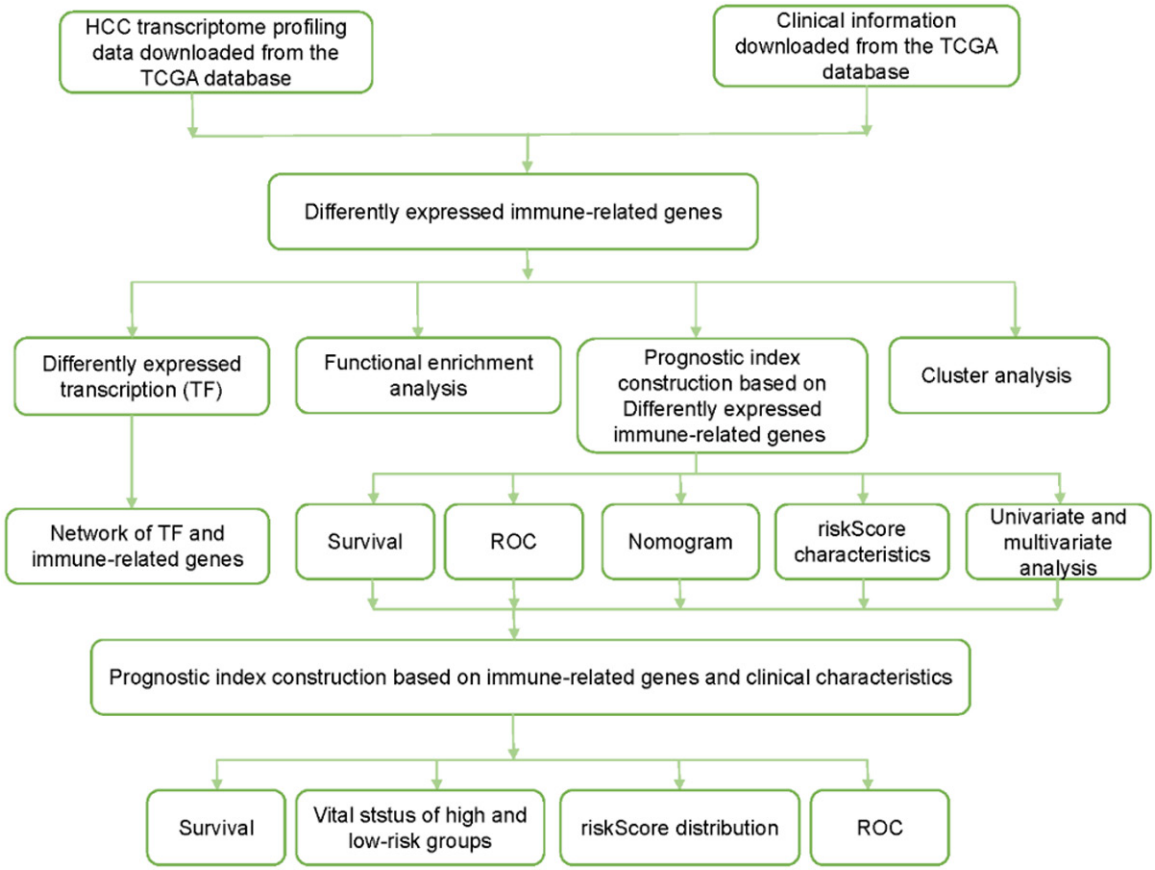


Figure 1. The flow diagram of the analysis procedure. Abbreviations: ROC, receiver operator characteristic.

rently expressed immune-related genes were considered statistically significant, when the joint satisfaction satisfied FDR < 0.05 and the fold changes ≥ 2. Univariate and multivariate COX regression analyses were used to verify whether the established model here can indeed be an independent prognostic factor. The nomogram was created using the "rms" package of the R software. The operating characteristic curve of the receiver and the AUC value were generated using the R "survival ROC" package. The differences in clinicopathological parameters between high and low-risk groups were tested by Students t-test or the X^2 test. All tests were bilateral and *P* values < 0.05 were considered statistically significant. All data used in this work were analyzed statistically by software R (version 3.4.1, https://www.r-project.org/).

Results

Differentially expressed immune genes set

The flowchart of this study is shown in **Figure 1**. Through the online TCGA database, the RNA sequences and clinical information of 374 HCC samples were obtained and compared with 50 normal liver tissues. The gene sequence data were compared with the differentially expressed immune-related genes from the ImmPort database and thus, the expression of 2,498 immune-related genes was obtained. In order to further screen the valuable differentially expressed immune genes, we set the joint satisfaction as FDR < 0.05 and $\log_2$FC > 1. Then, it was obtained 9 differentially expressed immune-related genes and constructed a prediction model and a prognostic index (risk score; RS) based on immune-related genes and clinical characteristics.

The heatmaps and volcano plots that demonstrated significant differential distribution between each data set (normal and HCC tissues) are shown in **Figure 2A** and **2B**. As shown in **Figure 2A**, N stands for Normal, T stands for Tumor; higher and lower expression are represented in red and green, respectively. Principal component analysis of HCC and normal liver tissues based on immune-related genes show-

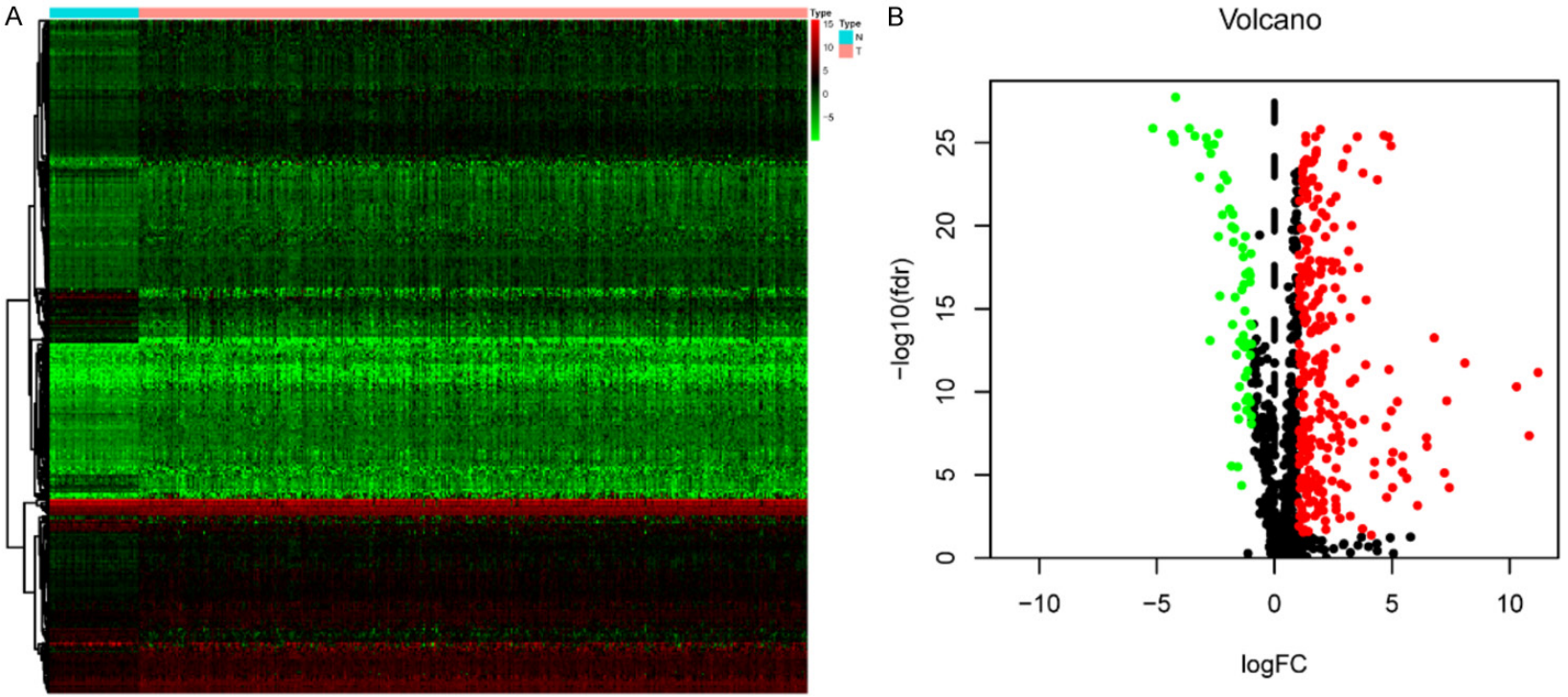


Figure 2. Differentially expressed immune-related genes; A. Hierarchical clustering of HCC tissues and normal liver tissues expressing the immune-related genes. Red indicates higher expression and green indicates lower expression. B. Volcano map of differentially expressed immune-related genes, red dots represent upregulated and green dots represent downregulated.

ed two significantly different distribution patterns. The normal liver and HCC samples were distributed on the left and right sides of heatmaps, respectively (**Figure 2B**). This suggests that there is a significant difference in the regulation of immune-related genes between normal liver and HCC.

Differentially expressed immune-related genes enrichment analysis

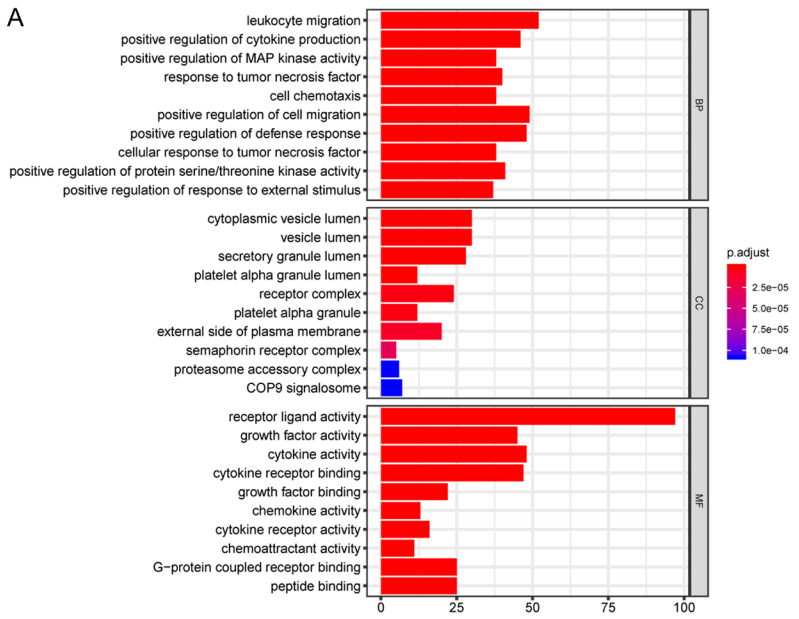
To explore the biological functional implication of differentially expressed immune-related genes, the top 10 GO enrichment analysis of up-regulated differentially expressed immune-related genes was performed (**Figure 3A**). The up-regulated genes were mostly associated with the Biological Process (BP) terms response to leukocyte migration, positive regulation of cytokine production, positive regulation of MAP kinase activity process and tumor necrosis factor. The bubble plot of enriched GO terms is shown in **Figure 3B**. The green, red and blue circles correspond to the BP, cellular component (CC) and the molecular function category (MF) terms, respectively. In addition, CC analysis showed that the up-regulation genes are related to cytoplasmic vesicle lumen, vesicle lumen, secretory granule lumen and platelet alpha granule lumen. Additionally, for MF terms, the up-regulated genes were enriched in receptor-ligand activity, growth factor activity, cytokine activity and cytokine receptor binding. Furthermore, KEGG pathway enrichment analysis of differentially expressed immune-related genes using DAVID (**Figure 3C**) shows the most significant KEGG pathways of the up-regulated differentially expressed immune-related genes, including cytokine-cytokine receptor interaction, antigen processing and presentation, axon guidance and MAPK signaling pathway. **Figure 3D** shows a bubble diagram of the relationship between immune genes and pathways. **Table 1** shows the up-regulated of differentially expressed immune-related genes in the top 10 sites both of GO enrichment analysis and the KEGG pathway enrichment analysis by DAVID methods, such as Terms, *P*-value and related genes.

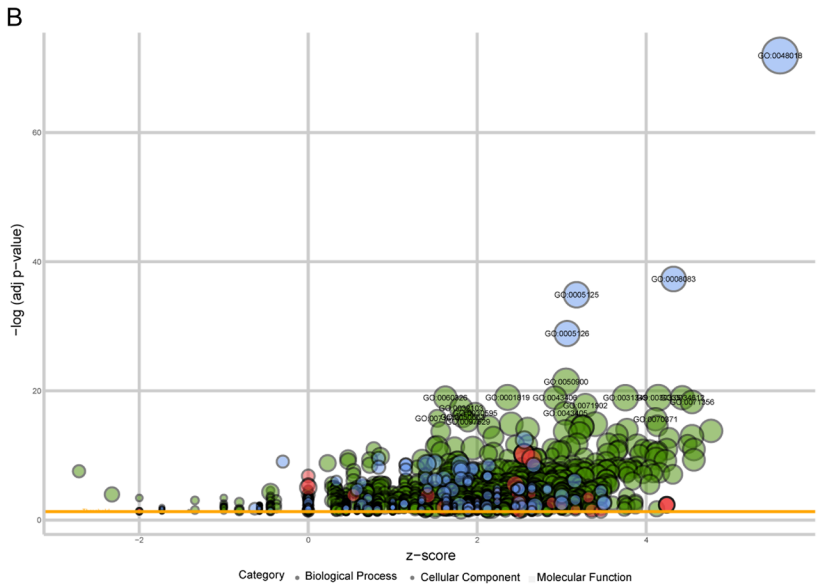
Differentially expressed TFs

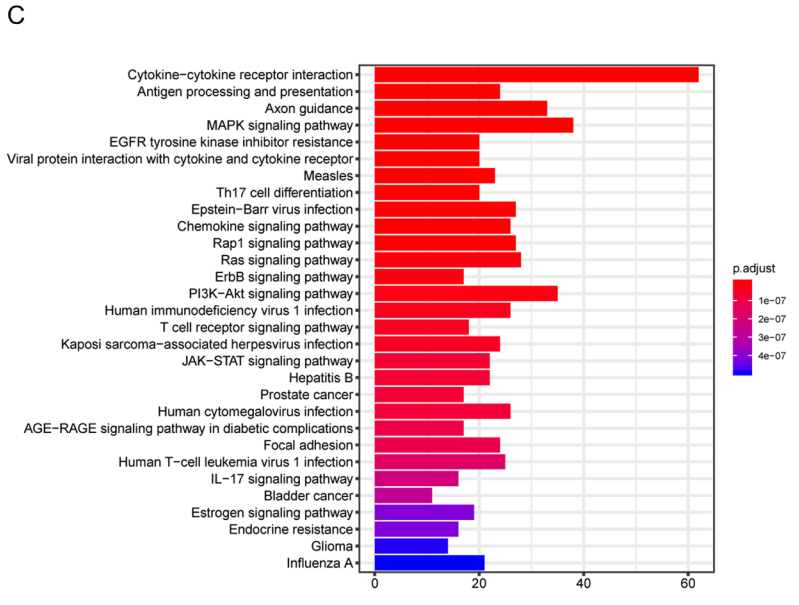
The upstream pathway of regulating HCC has been studied to confirm the identified HCC immune-related genes and to identify related TFs, since the immune genes are regulated by them. A total of 117 differentially expressed TFs related to HCC were found, as shown in a heatmap (**Figure 4A**), where the higher and lower expression between normal and HCC tissues are represented in red and green, respectively. It was found that normal liver and HCC samples were distributed on the left and right sides of the heatmaps, respectively (**Figure 4B**). This indicates that there is a significant difference in the regulation of transcription factors between normal liver and HCC. In addition, the regulatory relationship between immune-related genes and transcription factors has also been described (**Figure 4C**). The red circles, green triangles and red lines represent the immune-related genes, the related transcription factors and the positive regulation, respectively. A positive regulatory relationship was found between immune-related genes and transcription factors. Taken together, these data suggest that transcription factors can predict the survival and prognosis of patients with HCC.

Establishment of a prognostic index (risk score) based on immune related genes

The expression profiles of immune-related genes downloaded from the TCGA were standardized by [log2 (count+1)] transformation. Univariate Cox regression analysis was used to select the immune-related genes that were significantly correlated with overall survival (OS) in HCC patients (**Figure 5A**). Subsequently, these survival-related genes were analyzed by Lasso and multivariate Cox regression and genes that may not be independent indicators of prognosis monitoring were excluded (**Figure 5B** and **5C**). Then, 73 prognostic immune-related genes were obtained (**Figure 5D**), for which multiple stepwise Cox regression was performed to explore the effect of these genes on the survival time and outcome of the patients. Finally, nine gene markers were identified as independent predictors in HCC patients (**Table 2**) and selected to build a predictive model. A multivariate Cox regression model was used to develop the following immune-related risk signature associated with the survival of HCC patients [20]. The RS formula is based on the linear combination of gene relative expression levels multiplied by regression coefficients and represents the relative weight of genes in multiple Cox analysis.


A
leukocyte migration
positive regulation of cytokine production
positive regulation of MAP kinase activity
response to tumor necrosis factor
cell chemotaxis
positive regulation of cell migration
positive regulation of defense response
cellular response to tumor necrosis factor
positive regulation of protein serine/threonine kinase activity
positive regulation of response to external stimulus
cytoplasmic vesicle lumen
vesicle lumen
secretory granule lumen
platelet alpha granule lumen
receptor complex
platelet alpha granule
external side of plasma membrane
semaphorin receptor complex
proteasome accessory complex
COP9 signalosome
receptor ligand activity
growth factor activity
cytokine activity
cytokine receptor binding
growth factor binding
chemokine activity
cytokine receptor activity
chemoattractant activity
G−protein coupled receptor binding
peptide binding
BP
CC
MF
0
25
50
75
100
p.adjust
2.5e−05
5.0e−05
7.5e−05
1.0e−04


B
−log (adj p−value)
z−score
Category
Biological Process
Cellular Component
Molecular Function


C
Cytokine−cytokine receptor interaction
Antigen processing and presentation
Axon guidance
MAPK signaling pathway
EGFR tyrosine kinase inhibitor resistance
Viral protein interaction with cytokine and cytokine receptor
Measles
Th17 cell differentiation
Epstein−Barr virus infection
Chemokine signaling pathway
Rap1 signaling pathway
Ras signaling pathway
ErbB signaling pathway
PI3K−Akt signaling pathway
Human immunodeficiency virus 1 infection
T cell receptor signaling pathway
Kaposi sarcoma−associated herpesvirus infection
JAK−STAT signaling pathway
Hepatitis B
Prostate cancer
Human cytomegalovirus infection
AGE−RAGE signaling pathway in diabetic complications
Focal adhesion
Human T−cell leukemia virus 1 infection
IL−17 signaling pathway
Bladder cancer
Estrogen signaling pathway
Endocrine resistance
Glioma
Influenza A
0
20
40
60
p.adjust
1e−07
2e−07
3e−07
4e−07

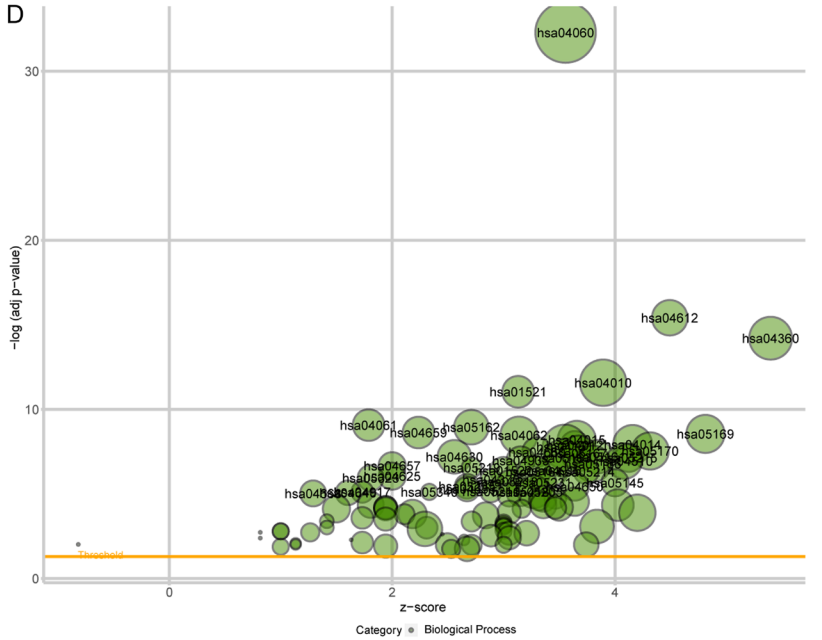

D
hsa04060
hsa04612
hsa04360
hsa04010
hsa01521
hsa04061
hsa04659
hsa05162
hsa04062
hsa05169
hsa05170
hsa04657
hsa04630
hsa05145
−log (adj p−value)
z−score
Category
Biological Process

Figure 3. Functional enrichment analysis of differentially expressed immune-related genes; A. The bar plot of enriched GO terms. Red bars display up-regulation and blue ones down-regulation; B. The bubble plot of enriched GO terms. The z-score is assigned to the x-axis, and the negative logarithm of the *P*-value to the y-axis, as in the bar plot (the higher the more significant). The size of the displayed circles is proportional to the number of genes assigned to the term. Greed circles correspond to the biological process, red indicates the cellular component, and blue shows the molecular function category; C. The bar plot of KEGG pathways. Red bars display up-regulation and blue ones down-regulation; D. The bubble plot of KEGG pathways. Greed circles correspond to the biological process, red indicates the cellular component, and blue shows the molecular function category.

Table 1. GO and KEGG analysis of differentially expressed immune-related genes

Category	ID	Term	*P*-value	GeneRatio
Biological Process	GO:0050900	leukocyte migration	9.20E-26	52/326
Biological Process	GO:0001819	positive regulation of cytokine production	4.82E-23	46/326
Biological Process	GO:0043406	positive regulation of MAP kinase activity	8.09E-23	38/326
Biological Process	GO:0034612	response to tumor necrosis factor	1.10E-22	40/326
Biological Process	GO:0060326	cell chemotaxis	1.59E-22	38/326
Biological Process	GO:0030335	positive regulation of cell migration	1.67E-22	49/326
Biological Process	GO:0031349	positive regulation of defense response	1.88E-22	48/326
Biological Process	GO:0071356	cellular response to tumor necrosis factor	9.71E-22	38/326
Biological Process	GO:0071902	positive regulation of protein serine/threonine kinase activity	4.04E-21	41/326
Biological Process	GO:0032103	positive regulation of response to external stimulus	1.12E-20	37/326
Cellular Component	GO:0060205	cytoplasmic vesicle lumen	3.26E-13	30/328
Cellular Component	GO:0031983	vesicle lumen	3.52E-13	30/328
Cellular Component	GO:0034774	secretory granule lumen	3.17E-12	28/328
Cellular Component	GO:0031093	platelet alpha granule lumen	1.76E-09	12/328
Cellular Component	GO:0043235	receptor complex	5.19E-08	24/328
Cellular Component	GO:0031091	platelet alpha granule	5.59E-08	12/328
Cellular Component	GO:0009897	external side of plasma membrane	1.86E-07	20/328
Cellular Component	GO:0002116	semaphorin receptor complex	6.83E-07	5/328
Cellular Component	GO:0022624	proteasome accessory complex	2.87E-06	6/328
Cellular Component	GO:0008180	COP9 signalosome	3.14E-06	7/328
Molecular Function	GO:0048018	receptor ligand activity	2.14E-75	97/322
Molecular Function	GO:0008083	growth factor activity	1.73E-40	45/322
Molecular Function	GO:0005125	cytokine activity	7.17E-38	48/322
Molecular Function	GO:0005126	cytokine receptor binding	9.66E-32	47/322
Molecular Function	GO:0019838	growth factor binding	3.68E-15	22/322
Molecular Function	GO:0008009	chemokine activity	3.05E-12	13/322
Molecular Function	GO:0004896	cytokine receptor activity	1.21E-11	16/322
Molecular Function	GO:0042056	chemoattractant activity	1.32E-11	11/322
Molecular Function	GO:0001664	G-protein coupled receptor binding	3.62E-11	25/322
Molecular Function	GO:0042277	peptide binding	4.25E-11	25/322
KEGG PATHWAY	hsa04060	Cytokine-cytokine receptor interaction	2.25E-35	294/7914
KEGG PATHWAY	hsa04612	Antigen processing and presentation	3.33E-18	77/7914
KEGG PATHWAY	hsa04360	Axon guidance	8.09E-17	181/7914
KEGG PATHWAY	hsa04010	MAPK signaling pathway	4.63E-14	295/7914
KEGG PATHWAY	hsa01521	EGFR tyrosine kinase inhibitor resistance	1.98E-13	79/7914
KEGG PATHWAY	hsa04061	Viral protein interaction with cytokine and cytokine receptor	2.23E-11	100/7914
KEGG PATHWAY	hsa05162	Measles	3.45E-11	138/7914
KEGG PATHWAY	hsa04659	Th17 cell differentiation	8.17E-11	107/7914
KEGG PATHWAY	hsa05169	Epstein-Barr virus infection	1.11E-10	201/7914
KEGG PATHWAY	hsa04062	Chemokine signaling pathway	1.47E-10	189/7914

This formula is as follows: Risk score = (0.0250 × HSPA4 expression) + (0.0213 × RBP2 expression) + (0.2424 × MAPT expression) + (0.1347 × TRAF3 expression) + (0.0057 × NDRG1

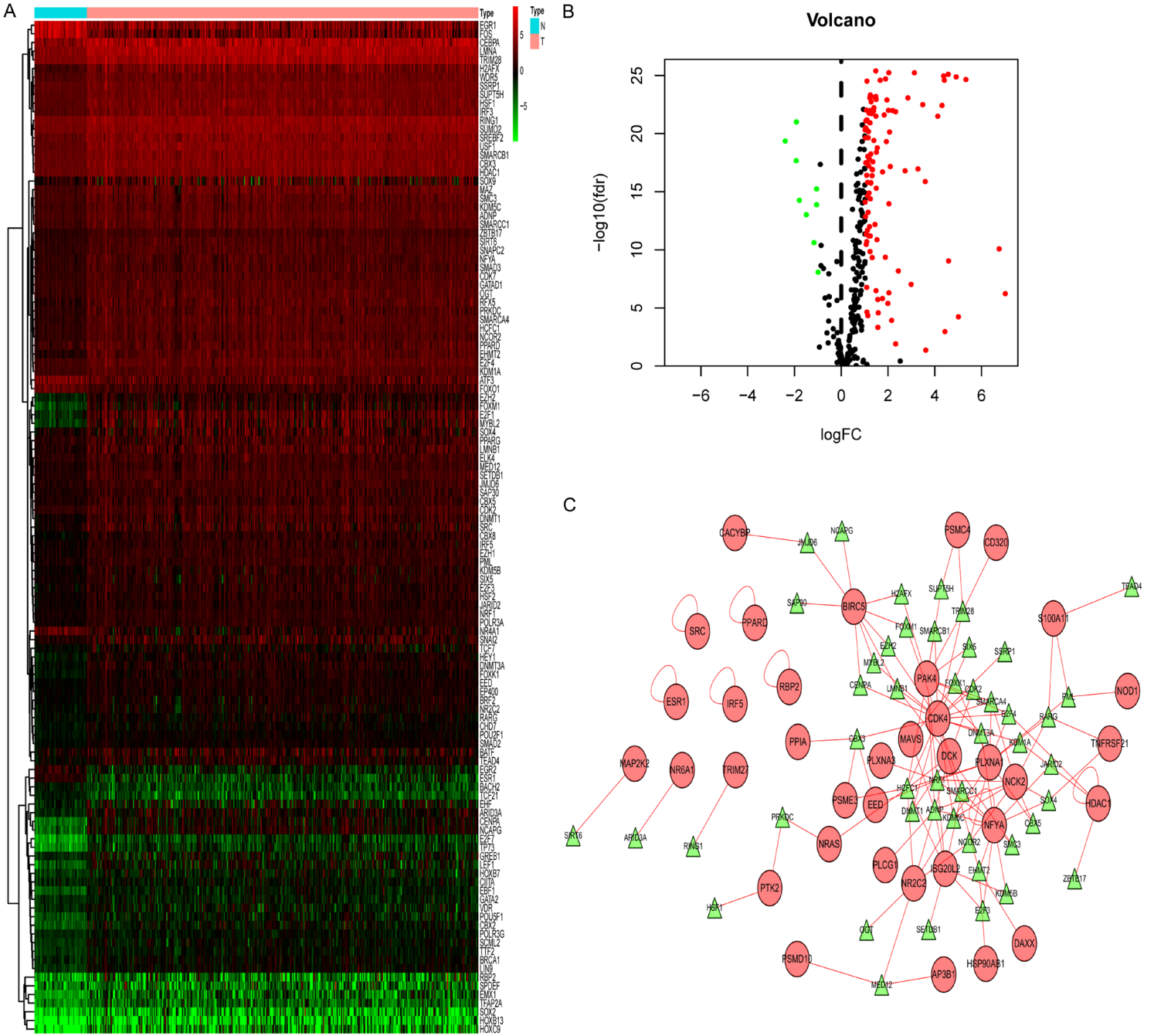

A
Type
N
T
B
Volcano
−log10(fdr)
logFC
C

Figure 4. 117 differentially expressed transcription factors (TFs) associated with HCC were demonstrated; A. Heatmap of differentially expressed TFs. The high expression was red and the low one was green. B. Volcano map of differentially expressed TFs, red dots represent upregulated and green dots represent downregulated, indicating that there was a significant difference in the regulation of transcription factors between normal liver and HCC. C. A network shows the relationship between TFs and immune-related genes. The red circle represents the immune-related genes, the green triangle represents the related transcription factors, and the red lines represent positive regulation. Shows a positive regulatory relationship between immune-related genes and transcription factors.

expression) + (0.0330 × NRAS expression) + (0.1501 × GAL expression) + (0.0799 × IL17D expression) + (0.0001 × SPP1 expression). It is possible to note that the coefficients of all immune-related genes were positive. This indicates that the expression of 9 immune-related genes is positively correlated with the survival time of HCC patients.

Evaluation of established prognostic model index (risk score)

The immune-related gene prognosis model was evaluated after its construction (**Figure 6**). The estimated survival for HCC patients in the high and low-risk groups was significantly different, with an increased risk of death in the high-risk group. The RS, patients survival status distribution, and expression heatmap of the 9 prognostic genes in the two groups are shown in **Figure 6A-C**. The survival rate was significantly different between the high and low-risk groups (**Figure 6D**). The prognostic capacity of the nine-gene signature was evaluated by using the AUC of a time-dependent ROC curve. The AUC of prognostic model based on biomarker genes was 0.805 (**Figure 6E**), indicating that this predictive model had a high sensitivity and specificity. Then, a nomogram was developed to predict 1, 3 and 5-year overall survival (OS) probabilities. Predictors of the nomogram included nine independent prognostic factors, namely HSPA4, RBP2, MAPT, TRAF3, NDRG1, NRAS, GAL, IL17D, and SPP1 (**Figure 6F**).

Verification of established prognostic model index (risk score)

The immune-related gene prognosis model was verified after its evaluation (**Figure 7**). The estimated survival for HCC patients in the high and low-risk groups was significantly different, with an increased risk of death in the high-risk group. The RS, patients survival status distribution, and expression heat map of the 9 prognostic genes in two groups are shown in **Figure 7A-C**. Survival rate was significantly different between the high and low-risk groups (**Figure 7D**). The AUC of a time-dependent ROC curve obtained was 0.582 (**Figure 7E**), indicating the predictive model had a certain sensitivity and specificity. Then, a nomogram was developed to predict 1 and 3-year OS probabilities. Predictors of the nomogram included the same nine independent prognostic factors presented before (HSPA4, RBP2, MAPT, TRAF3, NDRG1, NRAS, GAL, IL17D, and SPP1) (**Figure 7F**).

Independent prognostic factor evaluation and correlation with clinical characteristics

Univariate and multivariate Cox regression analysis were performed on the TCGA obtained dataset to study whether immune-related genes were independent factors associated with OS (**Table 3**). They showed that the RS was less than 0.05, suggesting that it can be used as an independent prognostic factor (**Figure 8A** and **8B**). In addition, the AUC values of the ROC curve for RS is the highest compared to the other prognostic characteristics such as age, gender, race, grade, stage, T, M and N (**Figure 8C**). These results show that the RS-based model constructed here was better than the other eight factors in assessing the survival and prognosis of HCC patients. Therefore, these results suggest that immune-related genes may be an independent prognostic factor. Moreover, to provide clinicians with a quantitative method for predicting the survival of HCC, a nomogram was collected that combines RS and various clinicopathological risk factors. This nomogram was constructed to estimate the probability of 1, 3 and 5-year survival and the results showed that the RS was the most basic factor among the different variables analyzed (**Figure 8D**).

The correlation analysis between 9 prognostic immune-related genes, RS and clinical characteristics is shown in **Table 4**. In addition, the relationship between clinicopathological parameters, immune-related genes and RS is shown in the Supplementary Figure 1. A significant difference in the expression of the immune genes

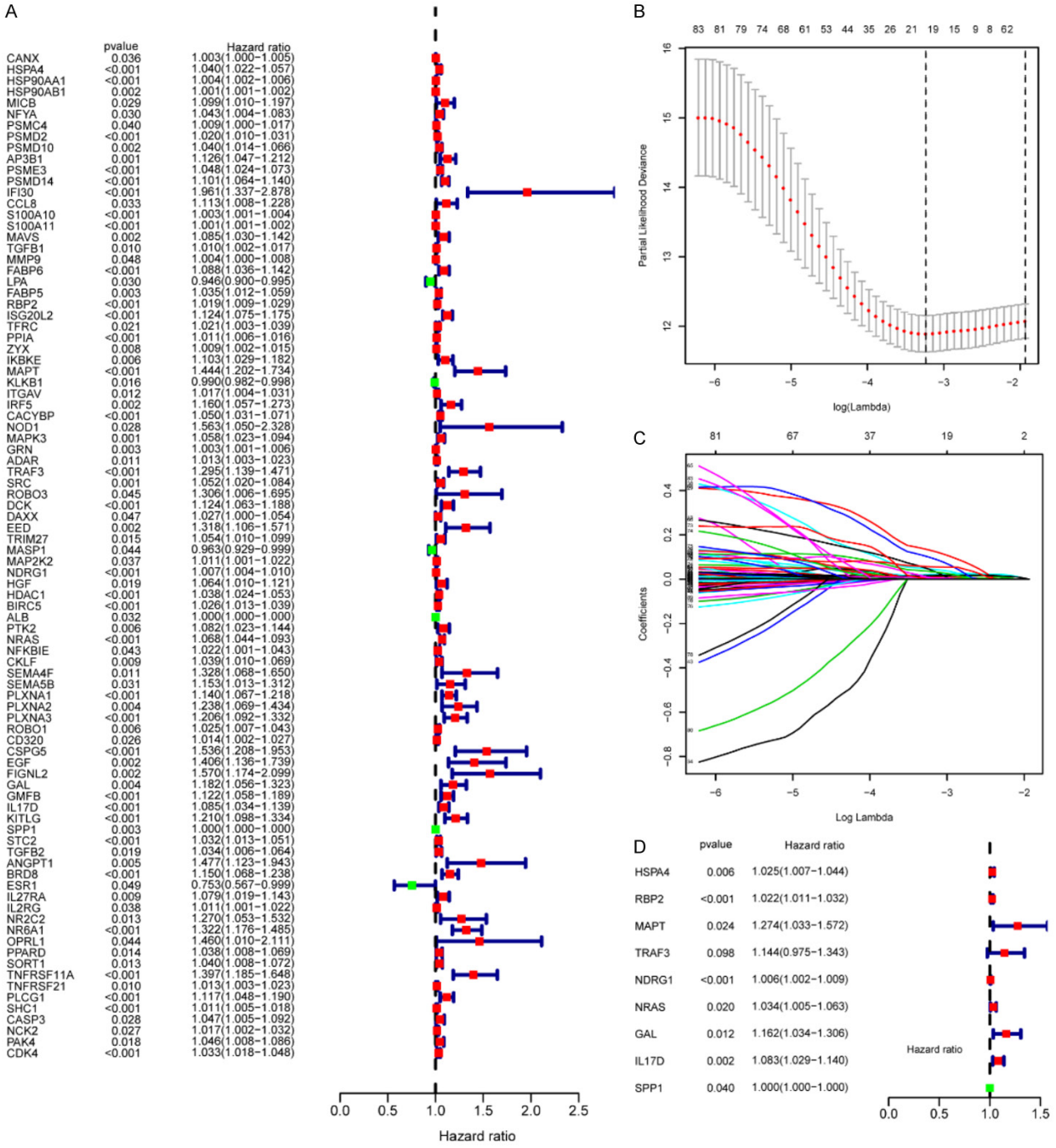

A
pvalue
Hazard ratio
CANX 0.036 1.003(1.000−1.005)
HSPA4 <0.001 1.040(1.022−1.057)
HSP90AA1 <0.001 1.004(1.002−1.006)
HSP90AB1 0.002 1.001(1.001−1.002)
MICB 0.029 1.099(1.010−1.197)
NFYA 0.030 1.043(1.004−1.083)
PSMC4 0.040 1.009(1.000−1.017)
PSMD2 <0.001 1.020(1.010−1.031)
PSMD10 0.002 1.040(1.014−1.066)
AP3B1 0.001 1.126(1.047−1.212)
PSME3 <0.001 1.048(1.024−1.073)
PSMD14 <0.001 1.101(1.064−1.140)
IFI30 <0.001 1.961(1.337−2.878)
CCL8 0.033 1.113(1.008−1.228)
S100A10 <0.001 1.003(1.001−1.004)
S100A11 <0.001 1.001(1.001−1.002)
MAVS 0.002 1.085(1.030−1.142)
TGFB1 0.010 1.010(1.002−1.017)
MMP9 0.048 1.004(1.000−1.008)
FABP6 <0.001 1.088(1.036−1.142)
LPA 0.030 0.946(0.900−0.995)
FABP5 0.003 1.035(1.012−1.059)
RBP2 <0.001 1.019(1.009−1.029)
ISG20L2 <0.001 1.124(1.075−1.175)
TFRC 0.021 1.021(1.003−1.039)
PPIA <0.001 1.011(1.006−1.016)
ZYX 0.008 1.009(1.002−1.015)
IKBKE 0.006 1.103(1.029−1.182)
MAPT <0.001 1.444(1.202−1.734)
KLKB1 0.016 0.990(0.982−0.998)
ITGAV 0.012 1.017(1.004−1.031)
IRF5 0.002 1.160(1.057−1.273)
CACYBP <0.001 1.050(1.031−1.071)
NOD1 0.028 1.563(1.050−2.328)
MAPK3 0.001 1.058(1.023−1.094)
GRN 0.003 1.003(1.001−1.006)
ADAR 0.011 1.013(1.003−1.023)
TRAF3 <0.001 1.295(1.139−1.471)
SRC 0.001 1.052(1.020−1.084)
ROBO3 0.045 1.306(1.006−1.695)
DCK <0.001 1.124(1.063−1.188)
DAXX 0.047 1.027(1.000−1.054)
EED 0.002 1.318(1.106−1.571)
TRIM27 0.015 1.054(1.010−1.099)
MASP1 0.044 0.963(0.929−0.999)
MAP2K2 0.037 1.011(1.001−1.022)
NDRG1 <0.001 1.007(1.004−1.010)
HGF 0.019 1.064(1.010−1.121)
HDAC1 <0.001 1.038(1.024−1.053)
BIRC5 <0.001 1.026(1.013−1.039)
ALB 0.032 1.000(1.000−1.000)
PTK2 0.006 1.082(1.023−1.144)
NRAS <0.001 1.068(1.044−1.093)
NFKBIE 0.043 1.022(1.001−1.043)
CKLF 0.009 1.039(1.010−1.069)
SEMA4F 0.011 1.328(1.068−1.650)
SEMA5B 0.031 1.153(1.013−1.312)
PLXNA1 <0.001 1.140(1.067−1.218)
PLXNA2 0.004 1.238(1.069−1.434)
PLXNA3 <0.001 1.206(1.092−1.332)
ROBO1 0.006 1.025(1.007−1.043)
CD320 0.026 1.014(1.002−1.027)
CSPG5 <0.001 1.536(1.208−1.953)
EGF 0.002 1.406(1.136−1.739)
FIGNL2 0.002 1.570(1.174−2.099)
GAL 0.004 1.182(1.056−1.323)
GMFB <0.001 1.122(1.058−1.189)
IL17D <0.001 1.085(1.034−1.139)
KITLG <0.001 1.210(1.098−1.334)
SPP1 0.003 1.000(1.000−1.000)
STC2 <0.001 1.032(1.013−1.051)
TGFB2 0.019 1.034(1.006−1.064)
ANGPT1 0.005 1.477(1.123−1.943)
BRD8 <0.001 1.150(1.068−1.238)
ESR1 0.049 0.753(0.567−0.999)
IL27RA 0.009 1.079(1.019−1.143)
IL2RG 0.038 1.011(1.001−1.022)
NR2C2 0.013 1.270(1.053−1.532)
NR6A1 <0.001 1.322(1.176−1.485)
OPRL1 0.044 1.460(1.010−2.111)
PPARD 0.014 1.038(1.008−1.069)
SORT1 0.013 1.040(1.008−1.072)
TNFRSF11A <0.001 1.397(1.185−1.648)
TNFRSF21 0.010 1.013(1.003−1.023)
PLCG1 <0.001 1.117(1.048−1.190)
SHC1 <0.001 1.011(1.005−1.018)
CASP3 0.028 1.047(1.005−1.092)
NCK2 0.027 1.017(1.002−1.032)
PAK4 0.018 1.046(1.008−1.086)
CDK4 <0.001 1.033(1.018−1.048)
0.0 0.5 1.0 1.5 2.0 2.5
Hazard ratio
B
83 81 79 74 68 61 53 44 35 26 21 19 15 9 8 62
Partial Likelihood Deviance
log(Lambda)
C
81 67 37 19 2
Coefficients
Log Lambda
D
pvalue
Hazard ratio
HSPA4 0.006 1.025(1.007−1.044)
RBP2 <0.001 1.022(1.011−1.032)
MAPT 0.024 1.274(1.033−1.572)
TRAF3 0.098 1.144(0.975−1.343)
NDRG1 <0.001 1.006(1.002−1.009)
NRAS 0.020 1.034(1.005−1.063)
GAL 0.012 1.162(1.034−1.306)
IL17D 0.002 1.083(1.029−1.140)
SPP1 0.040 1.000(1.000−1.000)
Hazard ratio
0.0 0.5 1.0 1.5

Figure 5. Establishment of a prognostic index (*Risk score*) based on immune related genes for HCC patients in TCGA; A. Univariate forest map shows the immune-related genes which were significantly correlated with overall survival (OS) in patients with HCC; B. CvFit Line shows the nine prognostic immune-related genes; C. Lambda expression to exclude genes that may not be independent indicators of prognostic monitoring; D. Multivariable forest map. Abbreviations: TCGA, The Cancer Genome Atlas; OS, overall survival.

Table 2. Univariate and multivariate cox regression analysis to establish a prognostic index (*Risks core*) based on immune genes for HCC patients in TCGA

Univariate Cox			Multivariate Cox			
Gene name	HR	*P*-value	Gene name	coef	HR	*P*-value
CANX	1.002841	0.03554	HSPA4	0.025021	1.025337	0.006281
HSPA4	1.039627	5.8E-06	RBP2	0.021319	1.021548	0.0000877
HSP90AA1	1.0039	0.000382	MAPT	0.242423	1.274333	0.023789
HSP90AB1	1.001366	0.001853	TRAF3	0.134779	1.144284	0.098462
MICB	1.099311	0.029065	NDRG1	0.005741	1.005758	0.000916
NFYA	1.042658	0.030266	NRAS	0.033029	1.033581	0.019685
PSMC4	1.008587	0.039789	GAL	0.15015	1.162008	0.011858
PSMD2	1.020417	0.000109	IL17D	0.079997	1.083284	0.00214
PSMD10	1.039754	0.001894	SPP1	0.00017	1.00017	0.040037
AP3B1	1.126435	0.001399				
PSME3	1.04786	8.18E-05				
PSMD14	1.101054	5.32E-08				
IFI30	1.961409	0.000578				
CCL8	1.112773	0.033354				
S100A10	1.002922	0.000237				
S100A11	1.001362	0.000316				
MAVS	1.084641	0.001955				
TGFB1	1.009537	0.009947				
MMP9	1.003831	0.04845				
FABP6	1.087902	0.000729				
LPA	0.945956	0.030312				
FABP5	1.034978	0.00317				
RBP2	1.018819	0.000114				
ISG20L2	1.124014	2.89E-07				
TFRC	1.02095	0.020738				
PPIA	1.010821	6.41E-05				
ZYX	1.008847	0.007793				
IKBKE	1.10279	0.005921				
MAPT	1.443894	8.51E-05				
KLKB1	0.990051	0.015847				
ITGAV	1.017426	0.011844				
IRF5	1.160037	0.001733				
CACYBP	1.050463	4.23E-07				
NOD1	1.563151	0.02793				
MAPK3	1.058037	0.001034				
GRN	1.003403	0.00293				
ADAR	1.01298	0.01058				
TRAF3	1.294759	7.39E-05				
SRC	1.051565	0.001274				
ROBO3	1.306036	0.044743				
DCK	1.12381	3.48E-05				

DAXX	1.026971	0.047352
EED	1.318217	0.002053
TRIM27	1.053734	0.014683
MASP1	0.963126	0.043581
MAP2K2	1.011149	0.036798
NDRG1	1.007283	4.55E-07
HGF	1.064074	0.018886
HDAC1	1.038413	2.88E-07
BIRC5	1.025993	0.000119
ALB	0.999979	0.032075
PTK2	1.082116	0.005506
NRAS	1.068198	2.6E-08
NFKBIE	1.021607	0.043308
CKLF	1.038973	0.008775
SEMA4F	1.327702	0.010545
SEMA5B	1.153094	0.030711
PLXNA1	1.14023	9.98E-05
PLXNA2	1.237901	0.004392
PLXNA3	1.205989	0.000222
ROBO1	1.024783	0.006341
CD320	1.014069	0.026135
CSPG5	1.535673	0.000469
EGF	1.405617	0.001722
FIGNL2	1.569922	0.002354
GAL	1.181853	0.003586
GMFB	1.121682	0.000116
IL17D	1.085121	0.000965
KITLG	1.20988	0.000126
SPP1	1.000185	0.003112
STC2	1.031944	0.000701
TGFB2	1.034382	0.019003
ANGPT1	1.477401	0.005228
BRD8	1.150199	0.0002
ESR1	0.752963	0.049388
IL27RA	1.079227	0.008894
IL2RG	1.011266	0.037656
NR2C2	1.270032	0.012538
NR6A1	1.321713	2.76E-06
OPRL1	1.46006	0.044173
PPARD	1.037898	0.014132
SORT1	1.039857	0.012782
TNFRSF11A	1.39747	7.12E-05
TNFRSF21	1.013304	0.009625
PLCG1	1.116812	0.000677
SHC1	1.011393	0.000911
CASP3	1.047491	0.028095
NCK2	1.016819	0.026724
PAK4	1.04617	0.018176
CDK4	1.032565	1.49E-05

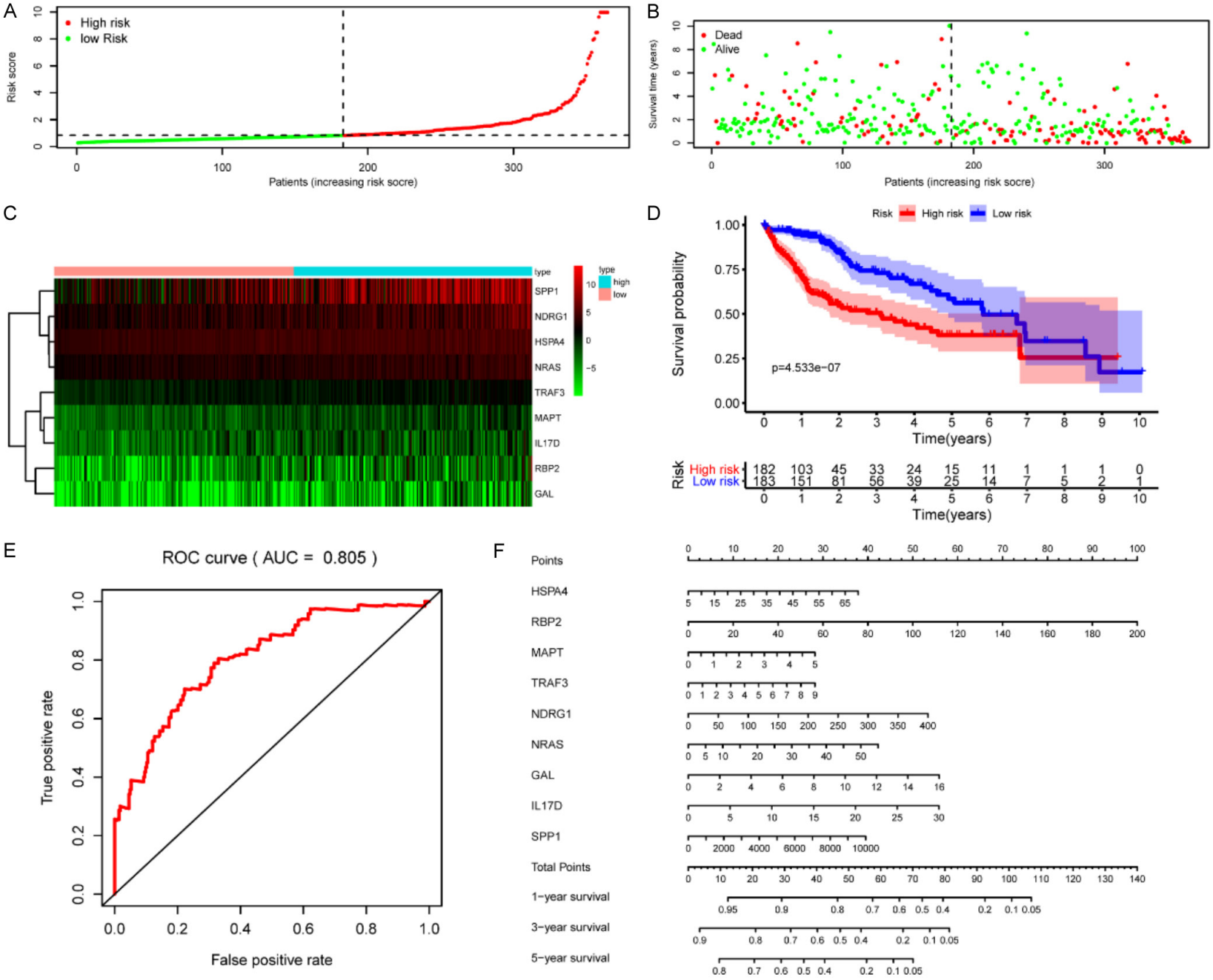

A
High risk
low Risk
Risk score
Patients (increasing risk socre)
B
Dead
Alive
Survival time (years)
Patients (increasing risk socre)
C
type
high
low
SPP1
NDRG1
HSPA4
NRAS
TRAF3
MAPT
IL17D
RBP2
GAL
D
Risk
High risk
Low risk
Survival probability
p=4.533e−07
Time(years)
High risk 182 103 45 33 24 15 11 1 1 1 0
Low risk 183 151 81 56 39 25 14 7 5 2 1
E
ROC curve (AUC = 0.805)
True positive rate
False positive rate
F
Points
HSPA4
RBP2
MAPT
TRAF3
NDRG1
NRAS
GAL
IL17D
SPP1
Total Points
1−year survival
3−year survival
5−year survival

Figure 6. Evaluation of prognostic index (*Risk score*) based on immune-related genes for HCC patients; A. Clinical characteristics in TCGA database (in order from top to bottom): The risk score distribution of HCC patients in high and low risk groups; B. The overall survival status distribution of HCC patients with increasing risk score; C. The heatmap of the 9 key genes expression profiles in the TCGA dataset; D. The Kaplan Meier curves for low-risk and high-risk groups; E. The ROC curves for predicting OS by the risk score; F. Nomogram for predicting 1-, 3-, and 5-year survival rate in HCC patients. By adding up the points identified on the point scale for each variable, the total score on the bottom scale shows the probability of survival. Abbreviations: HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas; ROC, receiver operator characteristic.

HSPA4 and NDRG1 was observed between Africans, Asians and Caucasians ($P < 0.05$; Supplementary Figure 1F and 1G). The expression of TRAF3, NDRG1, NRAS, IL17D immune genes, and the RS were significantly different between G 1-2 and G 3-4 ($P < 0.05$; Supplementary Figure 1A-E). The expression of TRAF3, NDRG1, NRAS immune genes ,and the RS were different between I-II and III-IV stages ($P < 0.05$; Supplementary Figure 1H-K), as well as the expression ofTRAF3, NDRG1 and the RS were different between T1-2 and T3-4 ($P < 0.05$; Supplementary Figure 1L-N). Finally, the HSPA4 immune gene expression was different between N0 and N1-X or M0 and M1-X stages ($P < 0.05$; Supplementary Figure 1O and 1P).

Relationship between prognostic model of immune-related genes and immune cells

The correlation between the immune-related gene model constructed here and immune cells was assessed after model construction and the related clinical evaluation (Supplementary Figure 2). *P-values* below than 0.05 indicating that the prognostic model is correlated with dendritic, macrophages and neutrophils immune cells (Supplementary Figure 2D-F), but not with B, CD4+ T, and CD8+ T cells (Supplementary Figure 2A-C).

Related small molecule drugs screening

The CMap database was used to screen potential SMDs for HCC. According to the probe group in which the expression of the HCC samples was consistent with that of adjacent normal samples, SMDs related to high correlation were screened. Nine SMDs were identified according to the absolute value (mean > 0.4) and *P*-value ($P < 0.01$) of the mean value (**Table 5**). Among these SMDs, orlistat, lycorine, cycloheximide, NS-398, and biotin showed a high negative correlation and therapeutic potential for HCC, while the others showed a positive correlation, suggesting that HCC can be treated by targeted inhibitors.

Discussion

HCC is a progressive disease and there is an urgent need for the establishment of reliable prognostic markers that are able to assist in diagnosis and treatment. Computational models have recently been used to explore possible mRNA and non-coding RNA biomarkers for HCC. In addition, a large number of studies have focused on the role of immunity in tumorigenesis and the results of cancer treatment. Although most studies of immune-related genes use cell lines or animal models, current studies from our group use the high-throughput expression profile of immune-related genes to explore the progress and prognosis of HCC patients.

HCC immunological pathogenesis has received increasing attention in recent years. At the same time, immunotherapy is gradually becoming a potential powerful treatment for advanced HCC due to three main reasons: (i) HCC is an inflammation-related cancer so immunotherapy is more likely to be effective [21]; (ii) the liver is an immune-privileged organ, therefore immunotherapeutic drugs can not be metabolized in the liver and have predictable pharmacokinetic characteristics in patients with liver cirrhosis [21] and; (iii) the immune response of the liver to antigens is susceptible to infection, which is balanced by the activation of immature T cells and several immunosuppressive mechanisms, such as cytokine scretion, antigen and immune checkpoint expression imbalance and local immune microenvironment changes [10, 22, 23]. In addition, the immune system has been shown to be a decisive factor in the occurrence and development of cancer [24, 25]. The clinical success of immune checkpoint inhibitor (ICI) in treating a variety of malignant tumors, including advanced melanoma, opens up the prospect of ICIs as a potential immunotherapy strategy for HCC [26, 27].

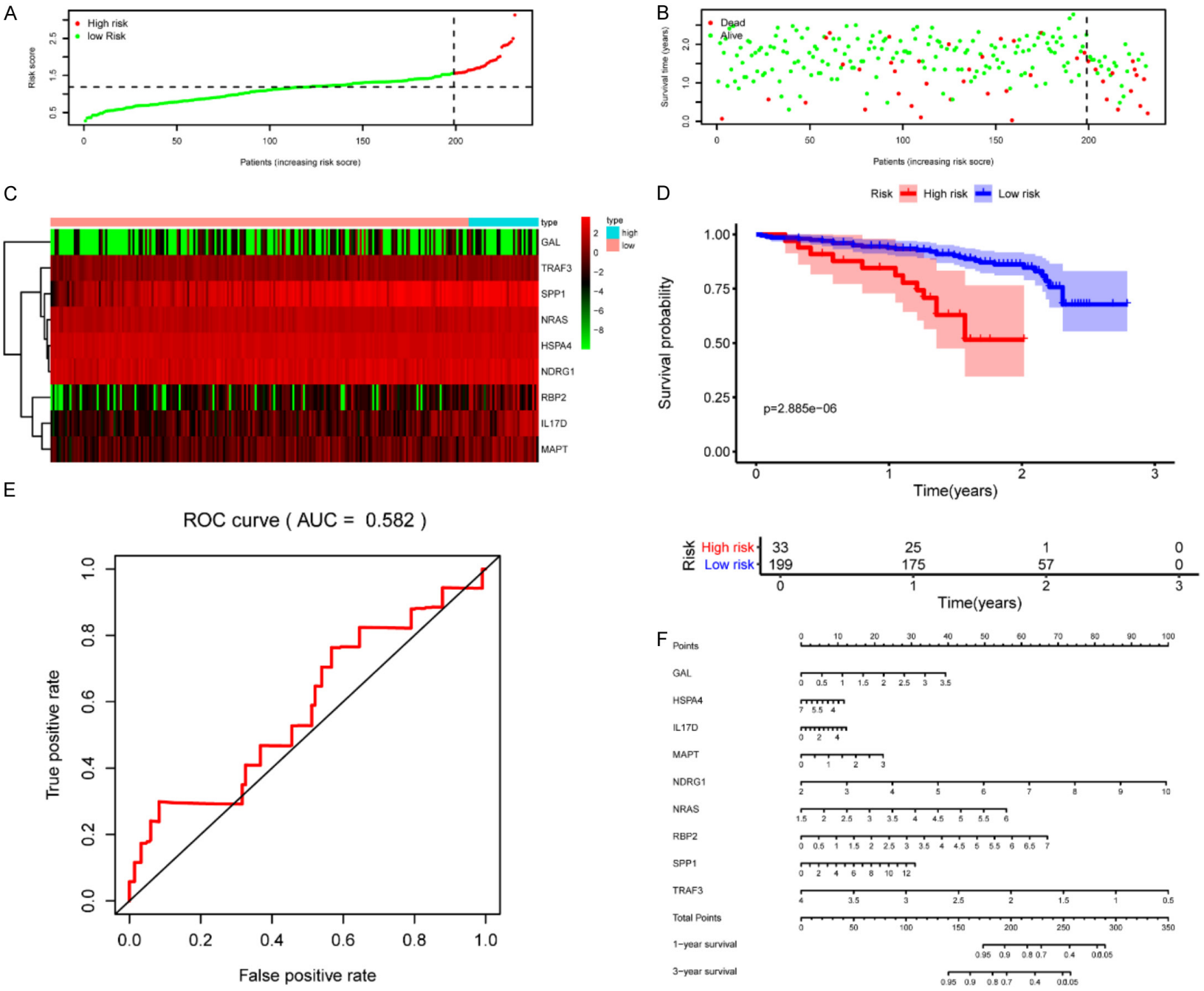

A
High risk
low Risk
Risk score
Patients (increasing risk socre)
B
Dead
Alive
Survival time (years)
Patients (increasing risk socre)
C
type
high
low
GAL
TRAF3
SPP1
NRAS
HSPA4
NDRG1
RBP2
IL17D
MAPT
D
Risk
High risk
Low risk
Survival probability
p=2.885e−06
Time(years)
E
ROC curve (AUC = 0.582)
True positive rate
False positive rate
Risk
High risk 33 25 1 0
Low risk 199 175 57 0
Time(years)
F
Points
GAL
HSPA4
IL17D
MAPT
NDRG1
NRAS
RBP2
SPP1
TRAF3
Total Points
1−year survival
3−year survival

Figure 7. Verification of prognostic index (*Risk score*) based on immune-related genes for HCC patients in ICGC database; A. Clinical characteristics in ICGC database (in order from top to bottom): The risk score distribution of HCC patients in high and low risk groups; B. The overall survival status distribution of HCC patients with increasing risk score; C. The heatmap of the 9 key genes expression profiles in the ICGC dataset; D. The Kaplan Meier curves for low-risk and high-risk groups; E. The ROC curves for predicting OS by the risk score; F. Nomogram for predicting 1-, 3-year survival rate in HCC patients. By adding up the points identified on the point scale for each variable, the total score on the bottom scale shows the probability of survival. Abbreviations: HCC, hepatocellular carcinoma; ICGC, International Cancer Genome Consortium; ROC, receiver operator characteristic.

Table 3. Univariate and multivariate analyses of OS for HCC patients based on TCGA

	Univariate Cox		Multivariate Cox	
Gene name	HR	*P*-value	HR	*P*-value
age	1.010346	0.182343	1.015047	0.073856
gender	0.795041	0.254908	0.967971	0.887404
race	1.082731	0.430024	0.818392	0.145855
grade	1.101676	0.465355	1.10461	0.510164
stage	1.642365	4.09E-06	1.098775	0.843857
T	1.62586	2.12E-06	1.524711	0.357548
M	1.207148	0.090729	1.544997	0.005152
N	1.053726	0.652257	0.925932	0.617729
risk Score	1.143628	2.53E-16	1.137082	3.90E-12

The effects of immunotherapy as a new cancer treatment method have been discussed and evaluated for many years. In this study, differentially expressed immune-related genes in normal and HCC liver tissues were identified. Although the role of immunization in cancer development is still controversial, the development of a significant risk marker of immune-related genes is conducive to the science of disease prognosis. In addition, it may also provide a theoretical basis for clinicians to use gene-targeting immune therapy in the treatment of HCC patients. In this context, this study established, evaluated and verified a prognostic HCC model based on nine differentially expressed immune-related genes.

Previous studies have shown that clinical characteristics such as age, gender, race, grade, stage, T, M, N are an independent prognostic factor for HCC, suggesting that immune-related genes may be an accurate prognostic indicator. In this study, it was observed that the higher the RS, the worse the prognosis. The nomogram results show that the RS associated with the immunological signature of each patient is the most important variable, indicating the RS importance in predicting the prognosis of HCC patients. It is known that the prognosis of elderly patients with high FIGO stage and high tumor grade is poor, which is consistent with the results of this study. Therefore, this nomogram provides a complete and individual survival assessment, rather than using specific covariables. Finally, the nomogram results highlight that the selected nine gene markers can accurately predict the prognosis of HCC patients.

The clinical value of immune-related genes and clinical characteristics were compared and it was found that immune-related genes can more accurately estimate the OS of HCC patients. Furthermore, the upstream transcription factors that regulate immune-related genes were examined. A positive regulatory relationship between transcription factors and immune genes has been identified, indicating that transcription factors can also be potential targets for HCC therapy. This comprehensive study using multiple databases can contribute to a new understanding of HCC biology and map out potential therapeutic interventions. In addition, it can provide new treatment ideas for clinicians and provide confidence and hope for HCC patients.

In recent years, the development of high-throughput sequencing technology has given rise to large databases, such as TCGA and GEO, which provide an effective means of selecting gene markers. In the present study, the expression profile of immune-related genes in TCGA was explored in order to find molecular markers for detecting the prognosis of HCC patients. The up-regulated differentially expressed immune-related genes were subjected to gene functional enrichment analyses, including GO functional annotations and KEGG pathway enrichment analysis. These analyses showed that these genes were mainly enriched in tumor suppression pathways. These results suggest that immune-related genes may play an anti-cancer role in the process of tumorigenesis. In

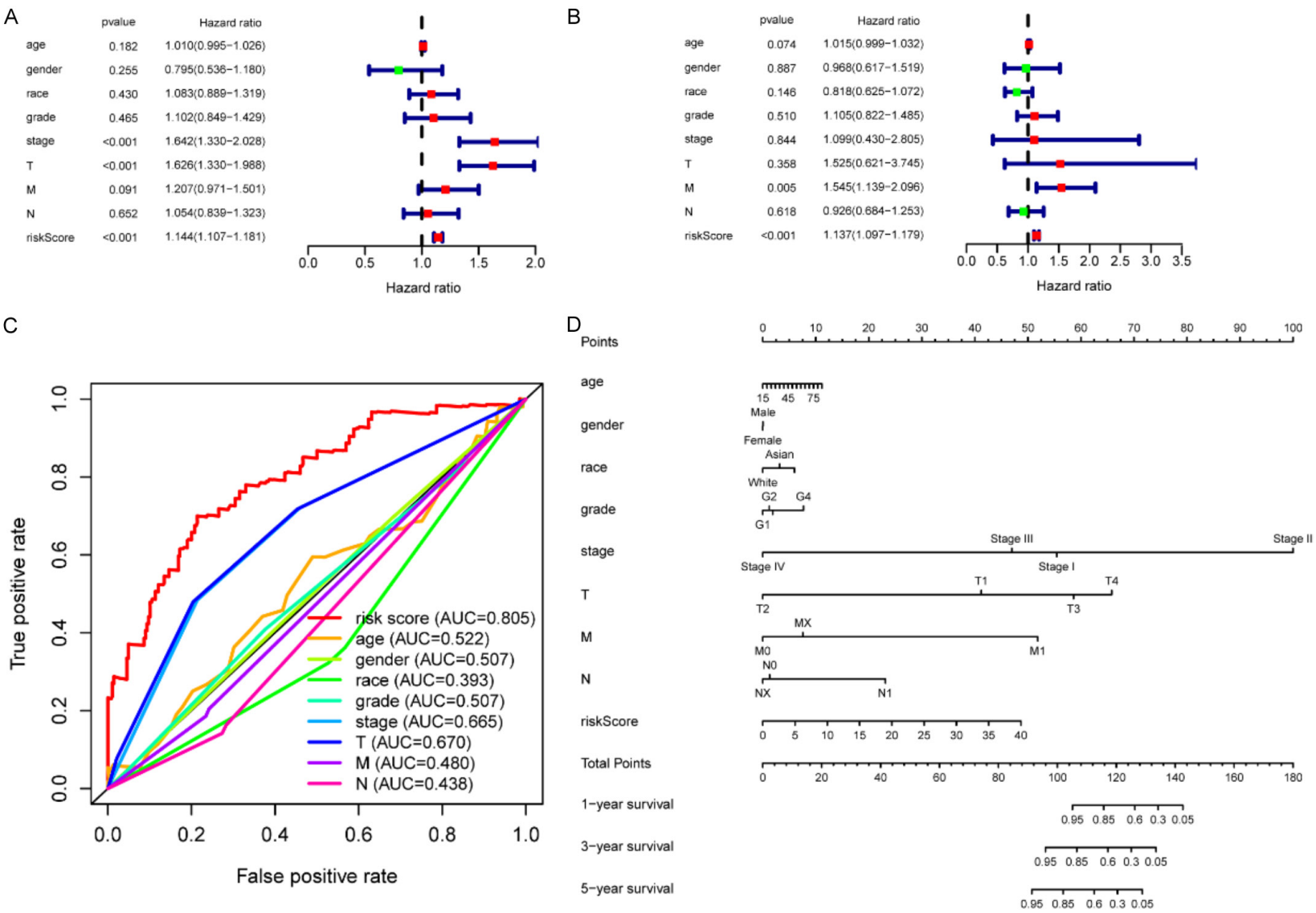


Figure 8. Independent prognostic factor evaluation; A. Univariate cox regression analysis (score < 0.05, suggesting the value of risk can be used as an independent prognostic factor); B. Multivariate cox regression analysis (risk score < 0.05, suggesting the value of risk can be used as an independent prognostic factor); C. ROC analysis of overall survival for the nine-gene signature and classical clinicopathologic parameters (age, gender, race, grade, stage, T, M, N) in the TCGA cohort; D. Prognostic nomogram for HCC patients. Abbreviations: ROC, receiver operator characteristic; TCGA, The Cancer Genome Atlas; HCC, hepatocellular carcinoma.

Table 4. Correlation analysis between 9 prognostic immune-related genes, *Risk score* and clinical characteristics

Gene name	age	gender	race	grade	stage	T	M	N
HSPA4	0.408 (0.684)	-1.174 (0.242)	21.216 (2.472e-05)	-4.904 (1.899e-06)	0.151 (0.880)	0.112 (0.911)	4.094 (6.227e-05)	3.412 (8.086e-04)
RBP2	1.311 (0.191)	1.268 (0.208)	1.348 (0.510)	-1.414 (0.160)	-1.282 (0.203)	-1.279 (0.205)	0.347 (0.729)	0.335 (0.738)
MAPT	-1.101 (0.272)	1.07 (0.286)	3.953 (0.139)	0.267 (0.789)	-0.87 (0.386)	-1.051 (0.295)	1.74 (0.083)	1.698 (0.091)
TRAF3	-0.403 (0.688)	1.478 (0.141)	0.655 (0.721)	-2.123 (0.035)	-2.203 (0.030)	-2.116 (0.037)	-0.543 (0.588)	-0.044 (0.965)
NDRG1	-0.444 (0.657)	-0.549 (0.583)	8.258 (0.016)	-1.705 (0.090)	-2.247 (0.027)	-2.216 (0.029)	1.148 (0.252)	-1.174 (0.242)
NRAS	0.827 (0.409)	0.97 (0.333)	4.828 (0.089)	-2.759 (0.006)	-2.228 (0.028)	-1.852 (0.067)	-0.208 (0.836)	0.014 (0.989)
GAL	0.371 (0.711)	-0.413 (0.680)	0.408 (0.815)	-1.385 (0.167)	-1.46 (0.147)	-1.542 (0.126)	-0.108 (0.914)	1.339 (0.182)
IL17D	1.851 (0.065)	1.9 (0.060)	0.217 (0.897)	-2.314 (0.022)	-0.808 (0.421)	-0.761 (0.448)	0.576 (0.565)	0.938 (0.349)
SPP1	-1.334 (0.184)	0.049 (0.961)	0.18 (0.914)	0.158 (0.874)	-0.336 (0.737)	-0.465 (0.643)	-0.758 (0.450)	-0.627 (0.532)
risk Score	0.285 (0.776)	1.292 (0.199)	5.587 (0.061)	-2.614 (0.010)	-2.533 (0.013)	-2.526 (0.013)	1.646 (0.101)	0.665 (0.507)

Table 5. Results of CMap analysis

rank	cmap name	mean	n	enrichment	P	specificity	percent non-null
1	ikarugamycin	0.569	3	0.916	0.00118	0.0111	100
2	metoprolol	0.474	4	0.827	0.00135	0	75
3	orlistat	-0.439	5	-0.762	0.00142	0	80
4	lycorine	-0.543	5	-0.73	0.00298	0.1333	80
5	cicloheximide	-0.484	4	-0.801	0.00302	0.0519	75
6	NS-398	-0.619	3	-0.876	0.00385	0.0065	100
7	biotin	-0.495	3	-0.868	0.00457	0	100
8	5182598	0.598	2	0.947	0.00499	0.0719	100
9	16-phenyltetranorprostaglandin E2	0.414	4	0.75	0.00754	0	75

recent years, immune-related genes have attracted great attention due to their multifaceted nature in cancer and their antagonistic effects in different types of tumors [28, 29].

Subsequently, the CMap database was used to identify SMDs that may have therapeutic effects on HCC. Some molecules have been identified to have anticancer effects, such as orlistat, lycorine, cycloheximide, NS-398, and biotin. Previous studies have found that orlistat can inhibit the proliferation of colorectal, pancreatic, ovarian and other cancer cells in addition to improving the paclitaxel effect in the liver cancer treatment [30-34]. Lycorine can inhibit the proliferation of many types of cancers, such as gastric, liver, small cell lung cancer among others [35-40]. NS-398 can promote the invasion of CD147 and MMP-2 to pancreatic cancer cells by activating P38 [41]. Biotin is a targeted anticancer precursor drug that can enhance the uptake of cancer cells and reverse cisplatin resistance [42, 43]. The identification of these molecules in CMap database suggest that they may be potential drugs for the treatment of HCC. In addition, it also provides some potential biomarkers and molecular targets for HCC treatment.

To date, some prognostic features of cancer based on expression profiles have been proposed with the help of large public databases. For example, Bao *et al.* [44] analyzed TCGA RNA-Seq data from 234 patients with bladder urothelial carcinoma and successfully obtained four-lncRNA signatures, which have prognostic value. Zhong *et al.* [45] also proposed a prognostic marker with six genes as a potential predictor of survival in er-positive breast cancer patients. However, these studies focused only on molecular markers and ignored traditional clinical parameters. On the other hand, the present study integrated the molecular markers and mechanisms identified with and clinical characteristics, thus increasing its prognostic potential for clinical application.

The advance of this study in relation to the previous ones lies in the accomplishment of a systematic analysis of the spectra of HCC disease and of the immune genes. This provided a reliable statistical method for exploring the role of these genes in HCC. Although the markers of these nine genes provide the possibility for an actual independent prognosis of HCC, this study does have some limitations. This study includes only immune-related genes and risk markers identified do not represent the entire gene transcription spectrum associated with HCC. In addition, other potential prognostic variables related to HCC, such as body mass index (BMI), tumor size among others, have not been studied and should be considered.

Conclusion

A prognostic model based on nine differentially expressed immune-related genes and including nine clinical risk factors has been established. This model presents a great ability to independently predict the overall survival time of HCC patients. In addition, several potential SMDs have been identified that can be explored as future therapeutic sources for the HCC treatment.

Acknowledgements

This work was supported by China National Science and Technology major projects 13th 5-year plan (No.2018ZX10725505), and Inno-

vation team project of Beijing University of Chinese Medicine (2019-JYB-TD-009).

Disclosure of conflict of interest

None.

Address correspondence to: Yong'an Ye, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China. E-mail: yeyongan@vip.163.com

References

[1] El-Serag HB and Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 2007; 132: 2557-2576.

[2] Altekruse SF, McGlynn KA and Reichman ME. Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. J Clin Oncol 2009; 27: 1485-1491.

[3] El-Serag HB. Hepatocellular carcinoma. N Engl J Med 2011; 365: 1118-1127.

[4] Balogh J, Victor D 3rd, Asham EH, Burroughs SG, Boktour M, Saharia A, Li X, Ghobrial RM and Monsour HP Jr. Hepatocellular carcinoma: a review. J Hepatocell Carcinoma 2016; 3: 41-53.

[5] Parikh S and Hyman D. Hepatocellular cancer: a guide for the internist. Am J Med 2007; 120: 194-202.

[6] Tsochatzis E, Meyer T, O'Beirne J and Burroughs AK. Transarterial chemoembolisation is not superior to embolisation alone: the recent European Association for the Study of the Liver (EASL) - European Organisation for Research and Treatment of Cancer (EORTC) guidelines. Eur J Cancer 2013; 49: 1509-1510.

[7] Dhungel B, Jayachandran A, Layton CJ and Steel JC. Seek and destroy: targeted adeno-associated viruses for gene delivery to hepatocellular carcinoma. Drug Deliv 2017; 24: 289-299.

[8] Llovet JM, Burroughs A and Bruix J. Hepatocellular carcinoma. Lancet 2003; 362: 1907-1917.

[9] Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, Montalto F, Ammatuna M, Morabito A and Gennari L. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 1996; 334: 693-699.

[10] Chen Z, Shen S, Peng B and Tao J. Intratumoural GM-CSF microspheres and CTLA-4 blockade enhance the antitumour immunity induced by thermal ablation in a subcutaneous murine hepatoma model. Int J Hyperthermia 2009; 25: 374-382.

[11] Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, de Oliveira AC, Santoro A, Raoul JL, Forner A, Schwartz M, Porta C, Zeuzem S, Bolondi L, Greten TF, Galle PR, Seitz JF, Borbath I, Haussinger D, Giannaris T, Shan M, Moscovici M, Voliotis D and Bruix J. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008; 359: 378-390.

[12] Mir N, Jayachandran A, Dhungel B, Shrestha R and Steel JC. Epithelial-to-mesenchymal transition: a mediator of sorafenib resistance in advanced hepatocellular carcinoma. Curr Cancer Drug Targets 2017; 17: 698-706.

[13] Huang Y, Wang FM, Wang T, Wang YJ, Zhu ZY, Gao YT and Du Z. Tumor-infiltrating FoxP3+ Tregs and CD8+ T cells affect the prognosis of hepatocellular carcinoma patients. Digestion 2012; 86: 329-337.

[14] Long J, Lin J, Wang A, Wu L, Zheng Y, Yang X, Wan X, Xu H, Chen S and Zhao H. PD-1/PD-L blockade in gastrointestinal cancers: lessons learned and the road toward precision immunotherapy. J Hematol Oncol 2017; 10: 146.

[15] Huang Y, Wang F, Wang Y, Zhu Z, Gao Y, Ma Z, Xu R and Du Z. Intrahepatic interleukin-17+ T cells and FoxP3+ regulatory T cells cooperate to promote development and affect the prognosis of hepatocellular carcinoma. J Gastroenterol Hepatol 2014; 29: 851-859.

[16] Liu M, Xu Z, Du Z, Wu B, Jin T, Xu K, Xu L, Li E and Xu H. The identification of key genes and pathways in glioma by bioinformatics analysis. J Immunol Res 2017; 2017: 1278081.

[17] Wang J, Xie Y, Bai X, Wang N, Yu H, Deng Z, Lian M, Yu S, Liu H, Xie W and Wang M. Targeting dual specificity protein kinase TTK attenuates tumorigenesis of glioblastoma. Oncotarget 2018; 9: 3081-3088.

[18] Kasemeier-Kulesa JC, Schnell S, Woolley T, Spengler JA, Morrison JA, McKinney MC, Pushel I, Wolfe LA and Kulesa PM. Predicting neuroblastoma using developmental signals and a logic-based model. Biophys Chem 2018; 238: 30-38.

[19] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES and Golub TR. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006; 313: 1929-1935.

[20] Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D and Levy R. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004; 350: 1828-1837.

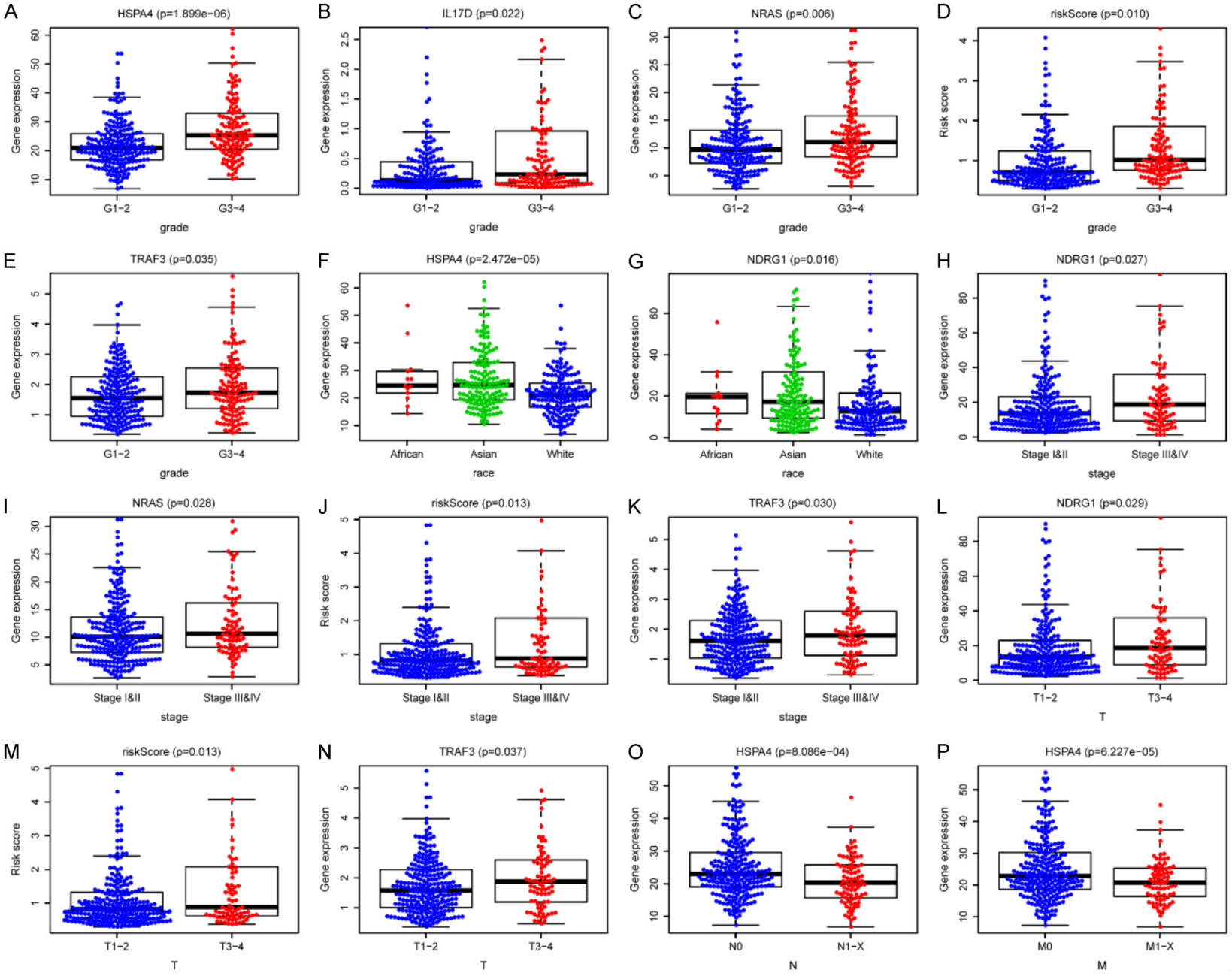
[21] Hato T, Goyal L, Greten TF, Duda DG and Zhu AX. Immune checkpoint blockade in hepatocel-

lular carcinoma: current progress and future directions. Hepatology 2014; 60: 1776-1782.
[22] Nemeth E, Baird AW and O'Farrelly C. Microanatomy of the liver immune system. Semin Immunopathol 2009; 31: 333-343.
[23] Thomson AW and Knolle PA. Antigen-presenting cell function in the tolerogenic liver environment. Nat Rev Immunol 2010; 10: 753-766.
[24] Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK and Alizadeh AA. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015; 21: 938-945.
[25] Angell H and Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. Curr Opin Immunol 2013; 25: 261-267.
[26] Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, Gonzalez R, Robert C, Schadendorf D, Hassel JC, Akerley W, van den Eertwegh AJ, Lutzky J, Lorigan P, Vaubel JM, Linette GP, Hogg D, Ottensmeier CH, Lebbe C, Peschel C, Quirt I, Clark JI, Wolchok JD, Weber JS, Tian J, Yellin MJ, Nichol GM, Hoos A and Urba WJ. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010; 363: 711-723.
[27] Topalian SL, Drake CG and Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. Cancer Cell 2015; 27: 450-461.
[28] Bono S, Lulli M, D'Agostino VG, Di Gesualdo F, Loffredo R, Cipolleschi MG, Provenzani A, Rovida E and Dello Sbarba P. Different BCR/Abl protein suppression patterns as a converging trait of chronic myeloid leukemia cell adaptation to energy restriction. Oncotarget 2016; 7: 84810-84825.
[29] Simioni C, Ultimo S, Martelli AM, Zauli G, Milani D, McCubrey JA, Capitani S and Neri LM. Synergistic effects of selective inhibitors targeting the PI3K/AKT/mTOR pathway or NUP214-ABL1 fusion protein in human acute lymphoblastic leukemia. Oncotarget 2016; 7: 79842-79853.
[30] Czumaj A, Zabielska J, Pakiet A, Mika A, Rostkowska O, Makarewicz W, Kobiela J, Sledzinski T and Stelmanska E. In vivo effectiveness of orlistat in the suppression of human colorectal cancer cell proliferation. Anticancer Res 2019; 39: 3815-3822.
[31] Peng H, Wang Q, Qi X, Wang X and Zhao X. Orlistat induces apoptosis and protective autophagy in ovarian cancer cells: involvement of Akt-mTOR-mediated signaling pathway. Arch Gynecol Obstet 2018; 298: 597-605.
[32] Sokolowska E, Presler M, Goyke E, Milczarek R, Swierczynski J and Sledzinski T. Orlistat reduces proliferation and enhances apoptosis in human pancreatic cancer cells (PANC-1). Anticancer Res 2017; 37: 6321-6327.
[33] Wright C, Iyer AKV, Kaushik V and Azad N. Antitumorigenic potential of a novel orlistat-AICAR combination in prostate cancer cells. J Cell Biochem 2017; 118: 3834-3845.
[34] You BJ, Chen LY, Hsu PH, Sung PH, Hung YC and Lee HZ. Orlistat displays antitumor activity and enhances the efficacy of paclitaxel in human hepatoma Hep3B cells. Chem Res Toxicol 2019; 32: 255-264.
[35] Chen X, Long X, Liao Y and Zheng J. Lycorine inhibits the proliferation and promotes apoptosis of SGC-7901 gastric cancer cells by downregulating expression of cyclin D1. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi 2018; 34: 354-358.
[36] Li L, Zhang Z, Yang Q and Ning M. Lycorine inhibited the cell growth of non-small cell lung cancer by modulating the miR-186/CDK1 axis. Life Sci 2019; 231: 116528.
[37] Wang C, Wang Q, Li X, Jin Z, Xu P, Xu N, Xu A, Xu Y, Zheng S, Zheng J, Liu C and Huang P. Lycorine induces apoptosis of bladder cancer T24 cells by inhibiting phospho-Akt and activating the intrinsic apoptotic cascade. Biochem Biophys Res Commun 2017; 483: 197-202.
[38] Wang J, Xu J and Xing G. Lycorine inhibits the growth and metastasis of breast cancer through the blockage of STAT3 signaling pathway. Acta Biochim Biophys Sin (Shanghai) 2017; 49: 771-779.
[39] Wu S, Qiu Y, Shao Y, Yin S, Wang R, Pang X, Ma J, Zhang C, Wu B, Koo S, Han L, Zhang Y, Gao X, Wang T and Yu H. Lycorine displays potent antitumor efficacy in colon carcinoma by targeting STAT3. Front Pharmacol 2018; 9: 881.
[40] Yu H, Qiu Y, Pang X, Li J, Wu S, Yin S, Han L, Zhang Y, Jin C, Gao X, Hu W and Wang T. Lycorine promotes autophagy and apoptosis via TCRP1/Akt/mTOR axis inactivation in human hepatocellular carcinoma. Mol Cancer Ther 2017; 16: 2711-2723.
[41] Liu H, Xu XF, Zhao Y, Tang MC, Zhou YQ and Gao FH. NS-398 promotes pancreatic cancer cell invasion by CD147 and MMP-2 via the activation of P38. Mol Med Rep 2016; 13: 2208-2214.
[42] Hu W, Fang L, Hua W and Gou S. Biotin-Pt (IV)-indomethacin hybrid: a targeting anticancer prodrug providing enhanced cancer cellular uptake and reversing cisplatin resistance. J Inorg Biochem 2017; 175: 47-57.
[43] Muhammad N, Sadia N, Zhu C, Luo C, Guo Z and Wang X. Biotin-tagged platinum(iv) com-

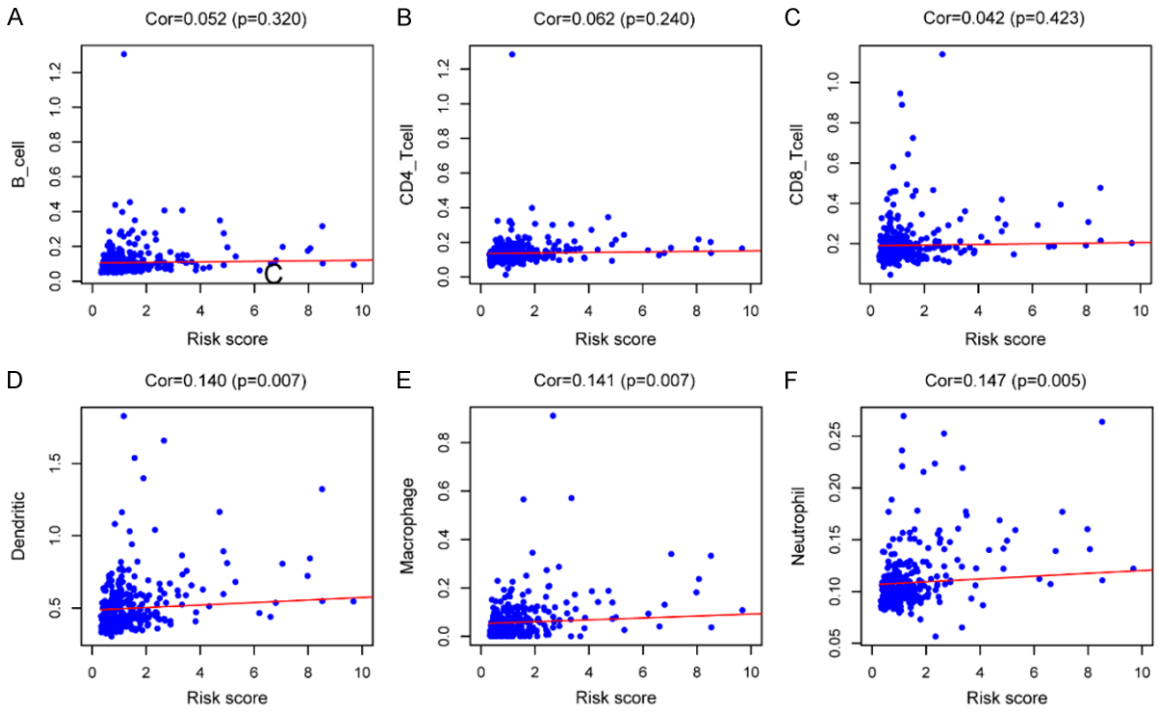
plexes as targeted cytostatic agents against breast cancer cells. Chem Commun (Camb) 2017; 53: 9971-9974.

[44] Bao Z, Zhang W and Dong D. A potential prognostic lncRNA signature for predicting survival in patients with bladder urothelial carcinoma. Oncotarget 2017; 8: 10485-10497.

[45] Zhong L, Lou G, Zhou X, Qin Y, Liu L and Jiang W. A six-long non-coding RNAs signature as a potential prognostic marker for survival prediction of ER-positive breast cancer patients. Oncotarget 2017; 8: 67861-67870.


A
HSPA4 (p=1.899e−06)
Gene expression
G1−2
G3−4
grade
B
IL17D (p=0.022)
Gene expression
G1−2
G3−4
grade
C
NRAS (p=0.006)
Gene expression
G1−2
G3−4
grade
D
riskScore (p=0.010)
Risk score
G1−2
G3−4
grade
E
TRAF3 (p=0.035)
Gene expression
G1−2
G3−4
grade
F
HSPA4 (p=2.472e−05)
Gene expression
African
Asian
White
race
G
NDRG1 (p=0.016)
Gene expression
African
Asian
White
race
H
NDRG1 (p=0.027)
Gene expression
Stage I&II
Stage III&IV
stage
I
NRAS (p=0.028)
Gene expression
Stage I&II
Stage III&IV
stage
J
riskScore (p=0.013)
Risk score
Stage I&II
Stage III&IV
stage
K
TRAF3 (p=0.030)
Gene expression
Stage I&II
Stage III&IV
stage
L
NDRG1 (p=0.029)
Gene expression
T1−2
T3−4
T
M
riskScore (p=0.013)
Risk score
T1−2
T3−4
T
N
TRAF3 (p=0.037)
Gene expression
T1−2
T3−4
T
O
HSPA4 (p=8.086e−04)
Gene expression
N0
N1−X
N
P
HSPA4 (p=6.227e−05)
Gene expression
M0
M1−X
M

Supplementary Figure 1. Correlation between 9 immune-related genes, *Risk score* and clinical characteristics for HCC patients; A-E. TRAF3, NDRG1, NRAS, IL17D, and *Risk score* were significantly related to G 1-2 and G 3-4 ($P < 0.05$); F, G. HSPA4 and NDRG1 were significantly related to race ($P < 0.05$); H-K. TRAF3, NDRG1, NRAS, and *Risk score* were significantly related to stage ($P < 0.05$); L-N. TRAF3, NDRG1 and *Risk score* were significantly related to T stage ($P < 0.05$); O, P. HSPA4 was significantly related to N stage or M stage ($P < 0.05$). Abbreviations: HCC, hepatocellular carcinoma.



Supplementary Figure 2. Correlations between the *Risk score* and immune cells; (A) B cell; (B) CD4 T cell; (C) CD8 T cell; (D) Dendritic; (E) Macrophage; (F) Neutrophil. Shows the immune-related gene model is correlated with Dendritic, Macrophage, and Neutrophil immune cells (D-F, $P < 0.05$), but not with B cells, CD4 T cells, and CD8 T cells (D-F, $P > 0.05$).