Original Article

Identification and Validation of an Immune-related Prognostic Signature for Hepatocellular Carcinoma

Xinxin Xia[1] , Ping Tang[2,3] , Hui Liu[4]* and Yuejun Li[2,3]*

[1]Department of Traditional Chinese Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China;
[2]Department of Oncology, The Third Affiliated Hospital of Hunan University of Chinese Medicine, Zhuzhou, Hunan, China;
[3]Department of Oncology, The First Affiliated Hospital of Hunan College of Traditional Chinese Medicine, Zhuzhou, Hunan, China; [4]Department of Breast Surgery, The First Hospital of Hunan University of Chinese Medicine, Changsha, Hunan, China

Abstract

Background and Aims: The immune system plays vital roles in hepatocellular carcinoma (HCC) initiation and progression. The present study aimed to construct an immune-gene related prognostic signature (IRPS) for predicting the prognosis of HCC patients. ***Methods:*** Gene expression data were retrieved from The Cancer Genome Atlas database. The IRPS was established via least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analysis. The prognostic values of the IRPS were further validated using the International Cancer Genome Consortium (ICGC) dataset. ***Results:*** A total of 62 genes were identified as candidate immune-related prognostic genes. According to the results of Lasso and multivariate Cox regression analysis, we established an IRPS and confirmed its stability and reliability in the ICGC dataset. The IRPS was significantly associated with advanced clinicopathological characteristics. Both Cox regression analyses revealed that the IRPS could be independent risk factors influencing prognosis of HCC patients. The relationships between the IRPS and infiltration of immune cells demonstrated that the IRPS was associated with immune cell infiltration. Furthermore, a nomogram was constructed to estimate the survival probability of HCC patients. ***Conclusions:*** The IRPS was effective for predicting prognosis of HCC patients, which might serve as novel prognostic and therapeutic biomarkers for HCC.

Introduction

Globally, hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related mortality. According to the World Health Organization's estimation, about one million patients will die from HCC in 2030.[1,2] Sorafenib and lenvatinib are the standard of therapy for patients with advanced stage HCC, however, the median overall survival is only 13 months.[3,4] Although new treatment algorithms and drugs have been constantly developed and tested during the past decades, the 5-year overall survival (OS) rate of HCC patients is as low as 18%.[2] Thus, it is urgently required to explore novel targets for HCC treatment.

The immune system can play vital roles in tumor initiation and progression, including for HCC.[5] Various studies have demonstrated that aberrantly expressed immune-related genes (IRGs) are related to a high risk of HCC development and a poor clinical outcome.[6–8] Abnormal expression of the immune gene programmed death-ligand 1 (PD-L1) was found to be associated with vascular formation in HCC patients.[7] An IRG expression pattern in cirrhosis patients has been reported to be correlated with the risk of HCC occurrence. Mice with chronic liver inflammation administrated with nintedanib or aspirin and clopidogrel lost IRG expression pattern and exhibited an attenuation in tumor size.[6] Hence, systematic analysis and characterization of the IRGs is of great importance.

Immunotherapy has shed new light on HCC treatment. In a phase 2 trial, nivolumab (PD-1 inhibitor) achieved a median survival of 15.6 months in HCC patients who had been treated with sorafenib.[9] Pembrolizumab, another PD-1 inhibitor, also showed similar results in a phase 2 trial.[10] However, phrase 3 trials of pembrolizumab as second-line treatment did not demonstrate a longer OS or progression-free survival compared to placebo, indicating the urgent need for accurate biomarkers to predicate the treatment

response and prognosis of HCC.[1]

In the present study, based on the IRGs obtained from the Immunology Database and Analysis Portal database (ImmPort; https://www.immport.org/shared/genelists),[11] we aimed to construct an immune-related prognosis signature (IRPS) with data retrieved from The Cancer Genome Atlas (TCGA) dataset (https://portal.gdc.cancer.gov/). The relationship of the prognosis signature with clinicopathological characteristics and immune cell infiltration were then explored. Further, to assess its predictive accuracy and effectiveness, the prognosis signature was validated with the data acquired from the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/releases/current/Projects).

Methods

Data preprocessing

The messenger RNA (mRNA) expression profiles and clinical data of HCC patients were retrieved from TCGA. The processed RNA-Sep FPKM data of a total of 374 HCC tissues and 50 normal samples were obtained. Another RNA-seq dataset, with 240 HCC tissues and 202 normal samples, was retrieved from ICGC and used as validation cohort for the prognosis signature. Only patients with complete survival information and OS time more than 30 days were selected for further analysis.

In total 2,498 IRGs, containing 17 immune categories according to different molecular functions, were derived from the ImmPort database, while 318 transcription factors (TFs) were obtained from the Cistrome Cancer database (http://cistrome.org/CistromeCancer/CancerTarget/).[12]

All of our data were directly acquired from open-access public databases, and this study was carried out strictly in accordance with the publishing guidelines provided by TCGA and ICGC. Thus, ethical approval was not necessary.

Construction of a TF-immunogene regulatory network

To detect differentially expressed IRGs (DE-IRGs) and TFs in normal and HCC samples, Wilcoxon test method was used in R software (version 3.6.1, https://www.r-project.org/). The significant cut-off values were set as log2-foldchange>1 and false discovery rate (FDR)<0.05. Heatmaps and volcano plots were generated with pheatmap package in R. To identify the prognostic value of DE-IRGs, univariate Cox analysis was performed with survival package in R. Only those with *p* value <0.01 were considered as prognostic IRGs. To assess the regulation of TFs on IRGs, the prognosis-associated TFs (*p*-value <0.01) were screened out using univariate Cox analysis. The correlation analysis between prognosis-related TFs and prognostic IRGs were performed using Pearson's test, and the calculations are done using cor.test in R. The significant cut-off values were set as *p*-value <0.001 and correlation coefficient >0.6. Cytoscape software was employed to build and visualize the TF-IRG regulatory network.[13]

Establishment of an IRPS for HCC

To establish a prognostic signature, the relationship between prognostic IRG expression and OS was evaluated by Least Absolute Shrinkage and Selection Operator (LASSO) and multivariate Cox regression analyses using the survival and glmnet packages in R. The prognostic signature was exhibited as risk score = (coefficient$_{mRNA1}$×expression of $_{mRNA1}$)+

(coefficient$_{mRNA2}$×expression of $_{mRNA2}$)+(coefficient$_{mRNA3}$×expression of $_{mRNA3}$)+…+ (coefficient$_{mRNAn}$×expression of $_{mRNAn}$). The eight-gene based risk scores for HCC patients were calculated via using the survminer R package to find median risk-score as cut-off value, and then the patients were allocated to high- and low-risk groups. The Kaplan-Meier (K-M) survival analysis was implemented to compare the differences in OS rates between the high- and low-risk groups by using the survival package in R. A scatter plot and interactive distribution-based scatter plot for evaluating the performance of the signature were generated to illustrate the gene expression heatmap and K-M survival curves. The receiver operating characteristic (ROC) curve was generated from the survival ROC package in R.

Independence evaluation of the IRPS

To evaluate the independent predictive power of the IRPS based on risk scores, HCC patients with complete demographic and clinical information, such as age, gender, tumor stage, and tumor grade, were included in subsequent analyses. Both univariate and multivariate Cox regression analyses were carried out by a forward stepwise method. A *p*-value <0.05 was regarded as statistically significant. The results are presented as forest_plots, which were generated via the survival package in R.

Correlation analysis between the IRPS and clinicopathological characteristics

To assess the correlation of prognostic signature and clinicopathological characteristics such as age at diagnosis, TNM stage, and tumor grade, the Chi-squared test was conducted via the beeswarm package in R.

Association analysis between the IRPS and immune cell infiltration

The Tumor Immune Estimation Resource (TIMER; http://timer.cistrome.org/), an important database for the systemic analysis of tumor-infiltrating immune cells (e.g., B cells, neutrophils, CD4+ T cells, CD8+ T cells, macrophages, and dendritic cells), was employed to evaluate the association of prognostic signature with immune cells infiltration. The immune cell infiltration data of HCC patients were obtained from TIMER database, and the association between the IRPS and immune cell infiltration was determined in R.

Validation of the IRPS

The validation cohort was retrieved from the ICGC database to validate the predictive values of the IRPS. The same formula was employed to determine the risk scores, and patients were assigned to high- and low-risk groups according to the same cut-off point of the TCGA cohort. Both K-M and ROC curves were generated by the methods described above. Independence evaluation and correlation analysis between IRPS and clinicopathological characteristics were also performed as described above.

External validation of the expression of the IRGs involved in IRPS

The mRNA expression data of the IRGs involved in IRPS

were retrieved from both TCGA and ICGC datasets, and then validated using the Oncomine database (https://www.oncomine.org/resource/main.html). The protein expression levels of the IRPS were also confirmed through the Human Protein Atlas (HPA) database (http://www.proteinatlas.org).

Gene set enrichment analyses (GSEA)

To disclose the biological information of the IRPS, GSEA were conducted to analyze the enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on the C2 in TCGA cohort.[14] A *p*-value <0.05 and FDR <0.05 were deemed as statistically significant.

Establishment of a predictive nomogram

A predictive nomogram was established by including all independent prognostic factors, in order to visualize the personal risk score and survival rate. The graph was generated with the rms package in R according to the multivariate Cox regression data of each independent assessment.

Statistical analysis

All statistical analyses were carried out with R. The relationship between the IRPS and OS were evaluated via log-rank test with survival package in R. The survival curves were generated via the survminer package, while the ROC curves were constructed with the survialROC package. The association analysis between the IRPS and clinicopathological characteristics of HCC was conducted by the beeswarm package in R.

Results

Identification of IRGs with prognostic value in HCC

In total, 329 IRGs (including 62 downregulated and 267 upregulated genes) were identified as differentially expressed genes in HCC samples compared with normal samples. Both heatmap and volcano plots showeds the distribution of DE-IRGs between HCC and normal tissues (Supplementary Fig. 1A, C). Then, the 329 IRGs were evaluated with univariate Cox regression analysis to determine prognostic characteristics. Finally, 62 genes were identified as candidate prognostic IRGs, which are shown in Figure 1.

Construction of TF–immunogene regulatory network

To discover underlying molecular mechanisms of these candidate prognostic IRGs, we explored the interaction between each gene. A total of 118 TFs were differentially expressed between HCC and normal tissues (Supplementary Fig. 1B, D). A regulatory network was built upon these 118 TFs and 62 prognostic IRGs. A correlation score of >0.6 was set as the cut-off value. The constructed TF-immunogene regulatory network revealed the interaction among these IRGs (Fig. 2).

Establishment of the IRPS for HCC prognosis

According to the univariate Cox regression results of the

TCGA cohort, LASSO-penalized Cox analysis identified eight genes to establish the IRPS, including retinol binding protein 2 (RBP2), microtubule-associated protein tau (MAPT), baculoviral IAP repeat-containing 5 (BIRC5), roundabout guidance receptor 1 (ROBO1), fidgetin like 2 (FIGNL2), interleukin 17D (IL17D), secreted phosphoprotein 1 (SPP1), and stanniocalcin 2 (STC2). The risk scores were calculated as follows: 0.0132×expression of RBP2+0.1397×expression of MAPT+0.0202×expression of BIRC5+0.0512×expression of ROBO1+0.2357×expression of FIGNL2+0.0560×expression of IL17D+0.0001×expression of SPP1+0.2095×expression of STC2. Based on the median cut-off risk score (0.387082), 343 HCC samples were assigned to the high- (*n*=171) and low-risk (*n*=172) groups. K-M survival analysis demonstrated that HCC patients with high-risk scores exhibited remarkably worse OS compared to the low-risk group (*p*-value=7.731e−08; Fig. 3A). The area under the curve (AUC) for OS was determined to be 0.825 (Fig. 3C), suggesting that this prognostic signature exhibits outstanding sensitivity and specificity. The gene expression heatmap and survival overview are illustrated in Figure 3E.

The prognostic signature was validated in the ICGC cohort. A total of 232 HCC patients were assigned to the high- (*n*=197) and low-risk groups (*n*=35) according to the same cut-off value of 0.387082 in the TCGA cohort. Notably, the OS of HCC patients was longer in the low-risk group than in the high-risk group (*p*-value=3.3632e−02; Fig. 3B). The AUC for OS was determined to be 0.668 (Fig. 3D), indicating moderate sensitivity and specificity of this prognostic signature. Moreover, the gene expression heatmap and survival overview are illustrated in Figure 3F.

Independence evaluation of the IRPS

For the TCGA cohort, univariate analysis demonstrated that tumor stage, metastasis stage, and risk score were independent prognostic factors for OS (Fig. 4). Furthermore, multivariate Cox regression analysis showed that only risk score remained an independent prognostic factor of the survival outcome of HCC patients after adjustment by age, gender, tumor grade and TNM stage (Fig. 4B). For the ICGC cohort, both univariate and multivariate Cox regression analyses revealed that tumor stage and IRPS were the independent prognostic factors for OS (Fig. 4C, D), which were supported by the findings of the TCGA cohort. Taken together, these results confirmed that IRPS could serve as a biomarker for predicting survival outcome of HCC patients.

Relationships between the IRPS and clinical-related factors

To further determine the predictive values of the IRPS, we analyzed the correlation between the IRPS and clinicopathological characteristics. In the TCGA cohort, this prognostic signature was significantly correlated with pathologically poor differentiation (*p*-value=0.003; Fig. 5A), more advanced tumor stage (*p*-value=0.004; Fig. 5B), and advanced T stage (Fig. 5C). Similarly, higher risk scores were associated with more advanced tumor stage in the ICGC cohort (Fig. 5D). The clinical significance of each identified IRG is shown in Supplementary Tables 1 and 2.

To analyze whether the IRPS is related to tumor-infiltrating immune cells, the association between the signature and immune cell infiltration were investigated in the TCGA cohort. The results indicated that CD4+ T cells and CD8+ T cells were negatively correlated with the signature, and no correlations were observed between the signature and B cells, dendritic cells, macrophage cells and neutrophil cells.
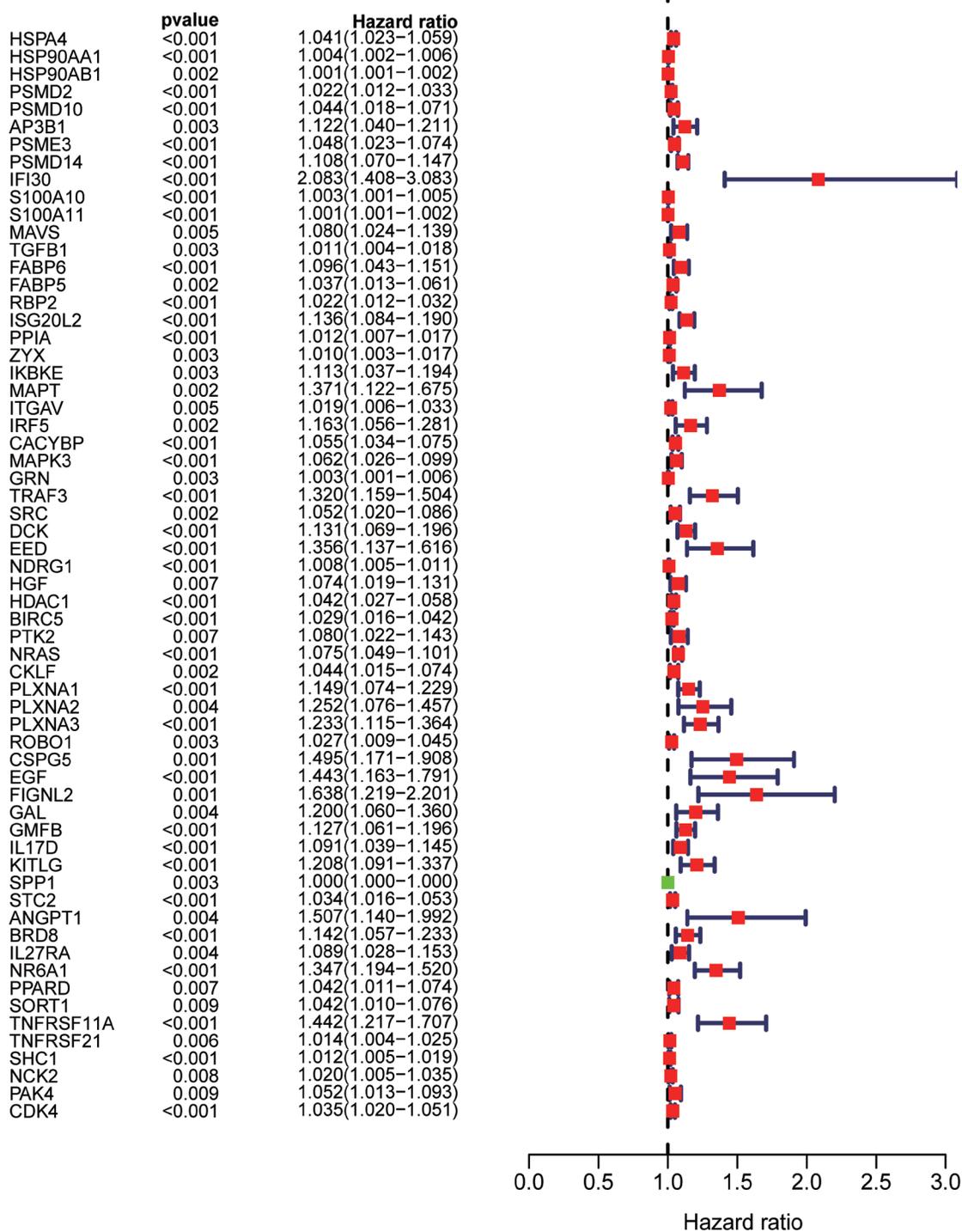
	pvalue	Hazard ratio
HSPA4	<0.001	1.041(1.023–1.059)
HSP90AA1	<0.001	1.004(1.002–1.006)
HSP90AB1	0.002	1.001(1.001–1.002)
PSMD2	<0.001	1.022(1.012–1.033)
PSMD10	<0.001	1.044(1.018–1.071)
AP3B1	0.003	1.122(1.040–1.211)
PSME3	<0.001	1.048(1.023–1.074)
PSMD14	<0.001	1.108(1.070–1.147)
IFI30	<0.001	2.083(1.408–3.083)
S100A10	<0.001	1.003(1.001–1.005)
S100A11	<0.001	1.001(1.001–1.002)
MAVS	0.005	1.080(1.024–1.139)
TGFB1	0.003	1.011(1.004–1.018)
FABP6	<0.001	1.096(1.043–1.151)
FABP5	0.002	1.037(1.013–1.061)
RBP2	<0.001	1.022(1.012–1.032)
ISG20L2	<0.001	1.136(1.084–1.190)
PPIA	<0.001	1.012(1.007–1.017)
ZYX	0.003	1.010(1.003–1.017)
IKBKE	0.003	1.113(1.037–1.194)
MAPT	0.002	1.371(1.122–1.675)
ITGAV	0.005	1.019(1.006–1.033)
IRF5	0.002	1.163(1.056–1.281)
CACYBP	<0.001	1.055(1.034–1.075)
MAPK3	<0.001	1.062(1.026–1.099)
GRN	0.003	1.003(1.001–1.006)
TRAF3	<0.001	1.320(1.159–1.504)
SRC	0.002	1.052(1.020–1.086)
DCK	<0.001	1.131(1.069–1.196)
EED	<0.001	1.356(1.137–1.616)
NDRG1	<0.001	1.008(1.005–1.011)
HGF	0.007	1.074(1.019–1.131)
HDAC1	<0.001	1.042(1.027–1.058)
BIRC5	<0.001	1.029(1.016–1.042)
PTK2	0.007	1.080(1.022–1.143)
NRAS	<0.001	1.075(1.049–1.101)
CKLF	0.002	1.044(1.015–1.074)
PLXNA1	<0.001	1.149(1.074–1.229)
PLXNA2	0.004	1.252(1.076–1.457)
PLXNA3	<0.001	1.233(1.115–1.364)
ROBO1	0.003	1.027(1.009–1.045)
CSPG5	0.001	1.495(1.171–1.908)
EGF	<0.001	1.443(1.163–1.791)
FIGNL2	0.001	1.638(1.219–2.201)
GAL	0.004	1.200(1.060–1.360)
GMFB	<0.001	1.127(1.061–1.196)
IL17D	<0.001	1.091(1.039–1.145)
KITLG	<0.001	1.208(1.091–1.337)
SPP1	0.003	1.000(1.000–1.000)
STC2	<0.001	1.034(1.016–1.053)
ANGPT1	0.004	1.507(1.140–1.992)
BRD8	<0.001	1.142(1.057–1.233)
IL27RA	0.004	1.089(1.028–1.153)
NR6A1	<0.001	1.347(1.194–1.520)
PPARD	0.007	1.042(1.011–1.074)
SORT1	0.009	1.042(1.010–1.076)
TNFRSF11A	<0.001	1.442(1.217–1.707)
TNFRSF21	0.006	1.014(1.004–1.025)
SHC1	<0.001	1.012(1.005–1.019)
NCK2	0.008	1.020(1.005–1.035)
PAK4	0.009	1.052(1.013–1.093)
CDK4	<0.001	1.035(1.020–1.051)

Fig. 1. Univariate Cox data for the OS of 62 candidate prognostic IRGs in patients with HCC.

These results are shown in Figure 6.

External validation of the expression of the IRGs involved in IRPS

Compared to the adjacent non-tumor liver tissues, the ex-
pression of the eight IRGs involved in IRPS were significantly increased in HCC in TCGA cohort. These results were validated in the ICGC cohort (Supplementary Table 3). Moreover, we confirmed the expression patterns of these eight genes in HCC samples through the Oncomine database. Except for FIGNL2, the mRNA expression levels of RBP2, MAPT, BIRC5, ROBO1, IL17D, SPP1 and STC2 were
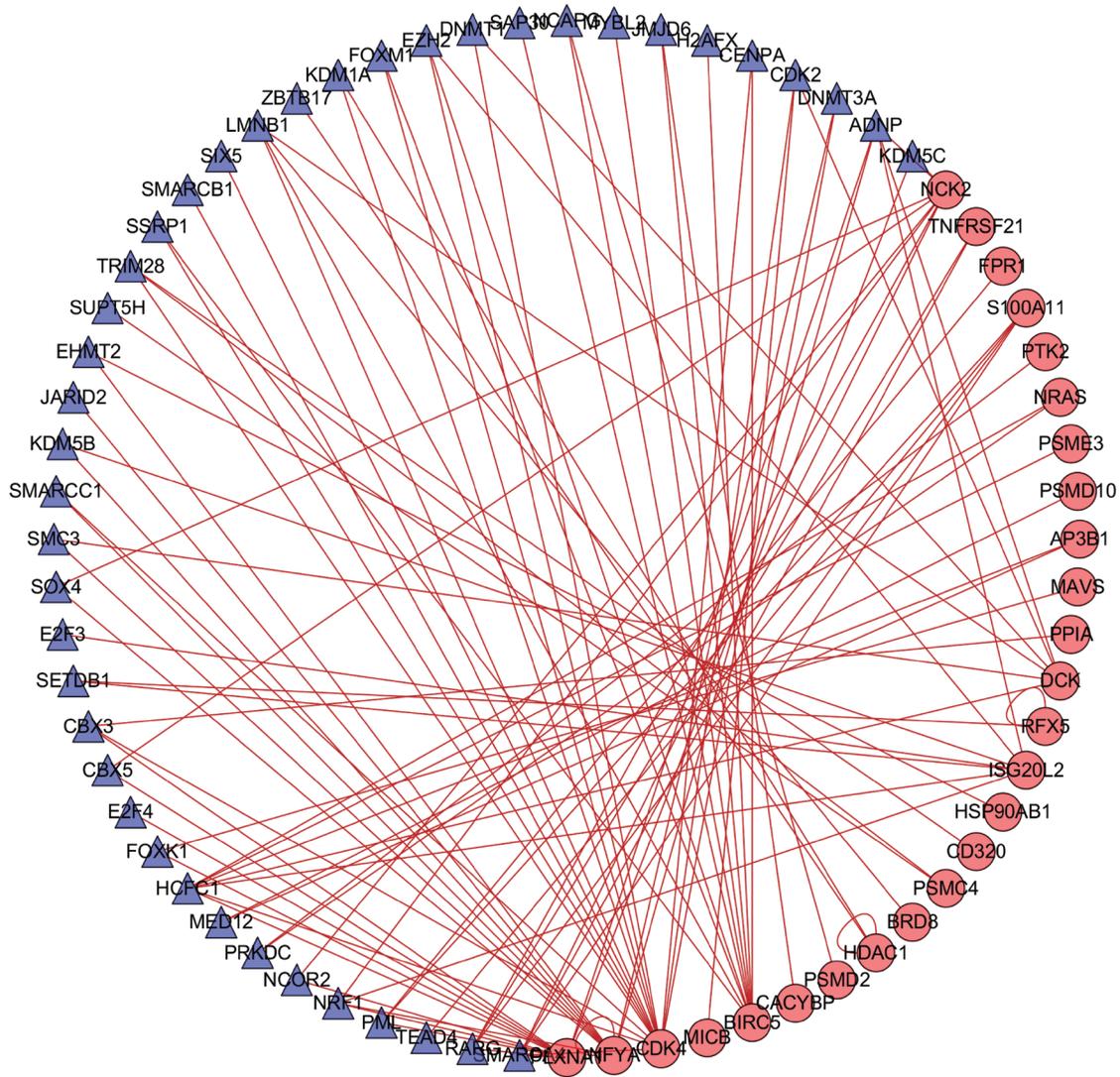
Fig. 2. The regulatory network built upon the prognosis-related TFs and candidate prognostic IRGs in patients with HCC. The blue triangle represents prognosis-associated TFs, while the red circle represents the candidate prognostic IRGs.

also increased in HCC tissues by using the Wurmbach liver cohort[15] (Fig. 7A). The protein expression of these eight genes was confirmed in the HPA database. Compared to their expression in normal liver tissues, MAPT and ROBO1 were strongly expressed, BIRC5 and SPP1were moderately expressed, and STC2 were weakly expressed in HCC tissues (Fig. 7B, C). However, the expression of RBP2 was not detected in both HCC and normal liver tissues (Fig. 7B, C), the expression of FIGNL2 and IL17D were not found on the website.

GSEA

GSEA were preformed and identified 65 significantly enriched KEGG pathways in the TCGA cohort. The top five KEGG pathways in the high-risk group were: "KEGG_PY-RIMIDINE_METABOLISM", "KEGG_PURINE_METABOLISM", "KEGG_RNA_DEGRADATION", "KEGG_BASE_EXCISION_RE-PAIR", and "KEGG_ CELL_CYCLE". The top five KEGG pathways in the low-risk group were: "KEGG_COMPLEMENT_

AND_COAGULATION_CASCADES", "KEGG_FATTY _ACID_ METABOLISM", "KEGG_DRUG_METABOLISM_CYTO-CHROME_ P450", "KEGG_PRIMARY_BILE_ACID_BIOSYN-THESIS", and "KEGG_RETINOL _METABOLISM" (Fig. 8A).

Establishment of a predictive nomogram

A predictive nomogram was established to systematically validate the predictive value of the prognostic signature in HCC patients (Fig. 8B). All independent clinical risk factors, including age, gender, grade, stage, and tumor, node, and metastasis stages combined with the prognostic signature, were involved. Each factor was assigned a score point according to its risk for survival. In the nomogram plot, the influence of different factors on survival were represented by the length of the line. The prognostic signature had the longest line, indicating that it had the greatest influence on the predication of survival probability. Compared to other clinical characteristics, our prognostic signature contributed the highest number of risk points (from 0 to 100) in the
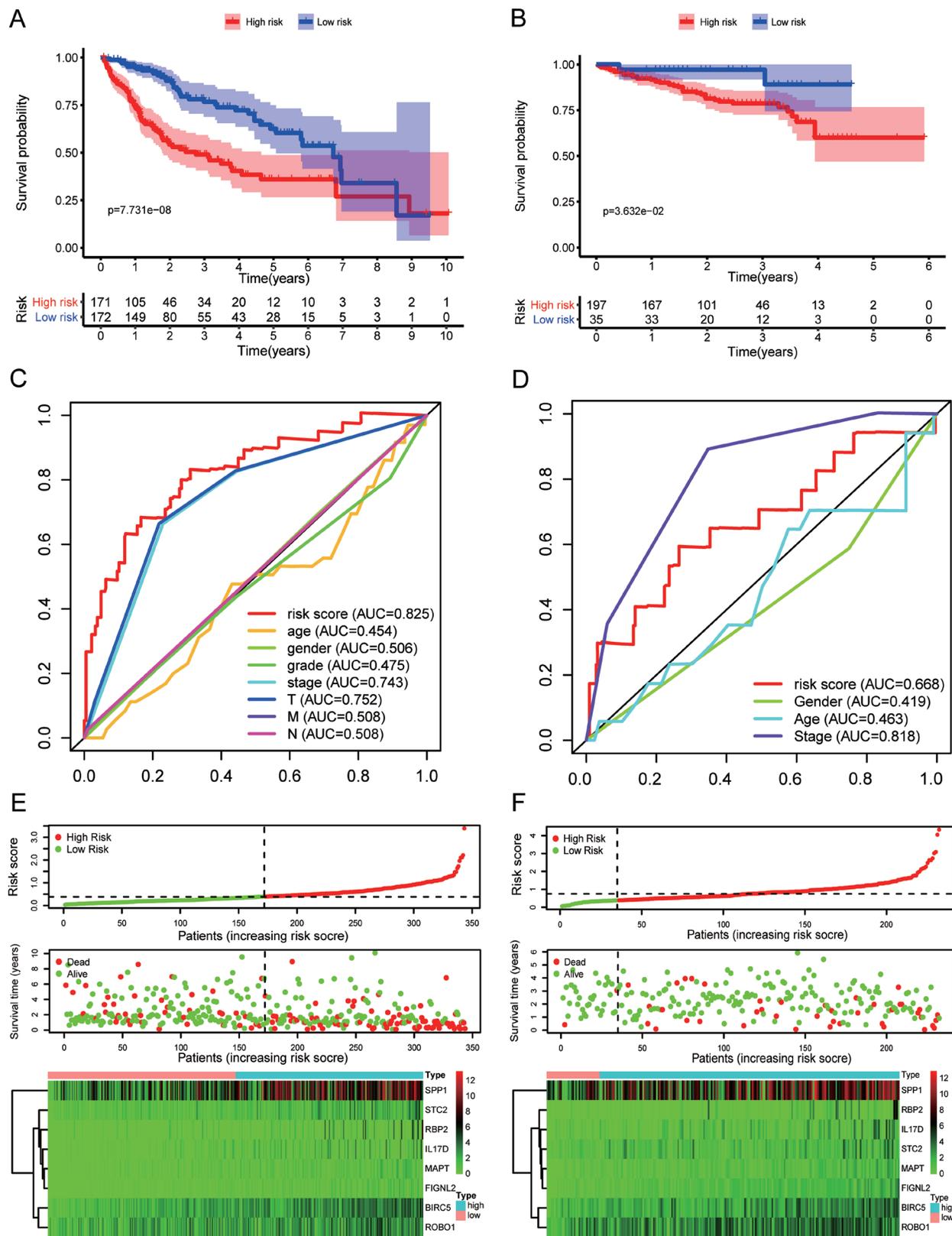
Fig. 3. Establishment of the IRPS for HCC patients. K-M curves of the IRPS in the TCGA (A) and ICGC (B) cohorts. ROC curves of the IRPS in the TCGA (C) and ICGC (D) cohorts. (E) Risk score distribution with survival status for the IRPS of HCC patients in the TCGA cohort. (F) Risk score distribution with survival status for the IRPS of HCC patients in the ICGC cohort.
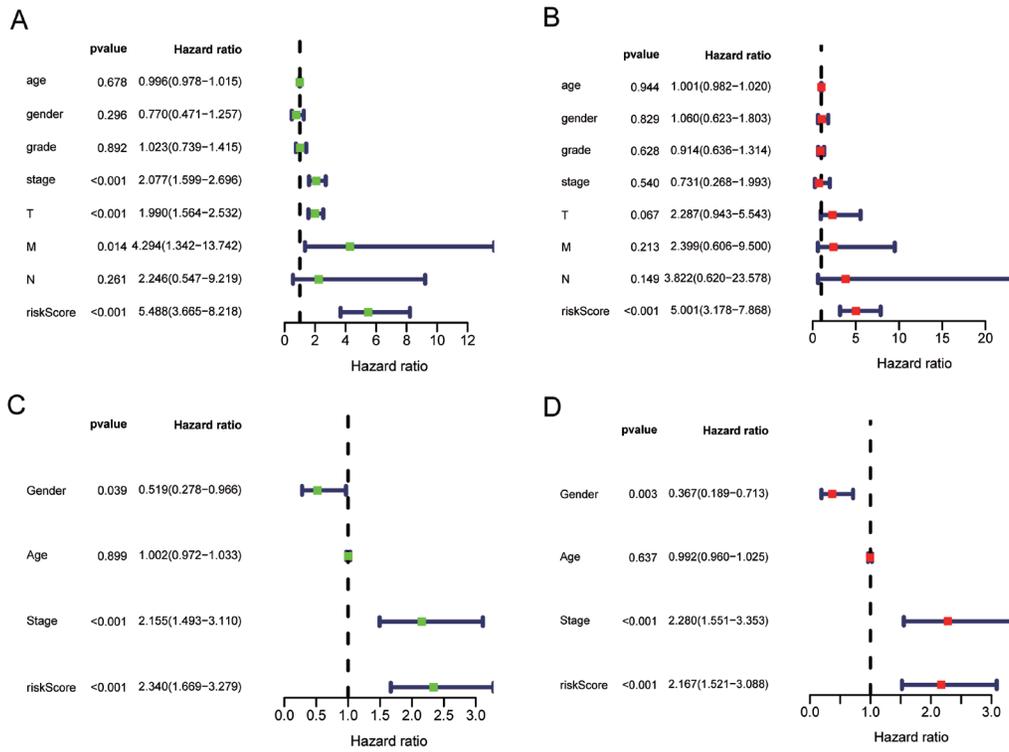
Fig. 4. Independence evaluation of the IRPS in patients with HCC. Univariate Cox regression analysis for the independence evaluation of the IRPS in the TCGA (A) and ICGC (C) cohorts. Multivariate Cox regression analysis for the independence evaluation of the IRPS in the TCGA (B) and ICGC (D) cohorts.
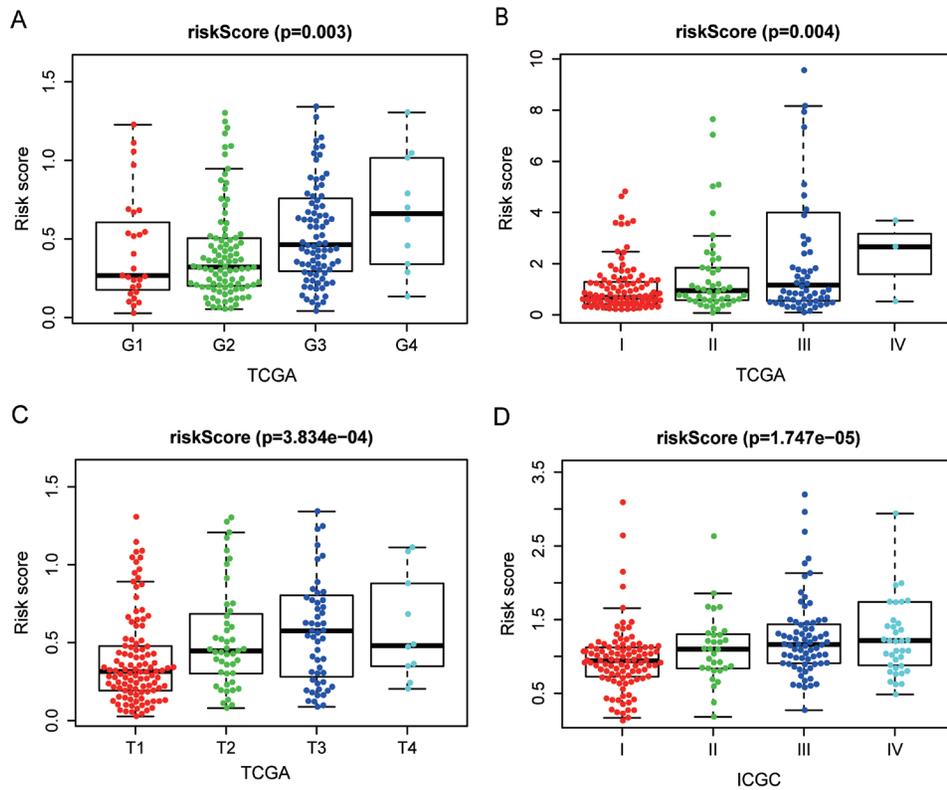

Fig. 5. Association analyses between the IRPS and clinicopathological characteristics. The IRPS was associated with more advanced tumor stage (A), tumor stage (B), and poorly differentiated carcinoma (C) in the TCGA cohort. (D) The IRPS was associated with more advanced tumor stage in the ICGC cohort.
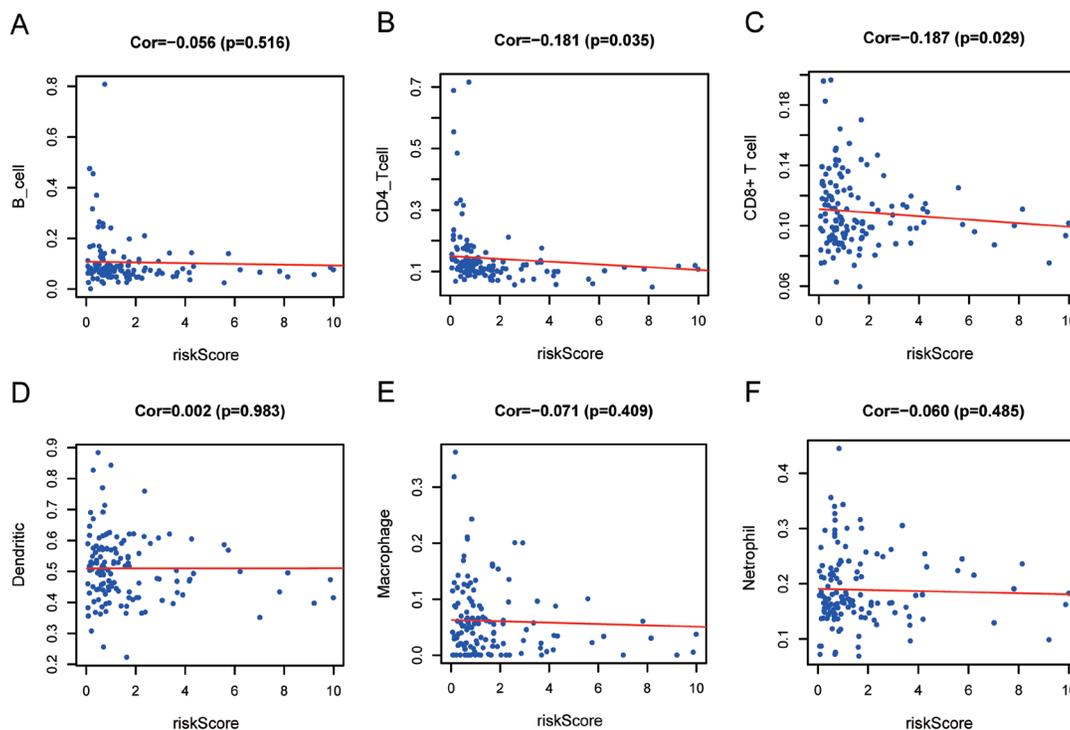
Fig. 6. Relationships between IRPS and immune cell infiltration. (A) B cells; (B) CD4 T cells; (C) CD8 T cells; (D) Dendritic cells; (E) Macrophage cells; (F) Neutrophil cells.
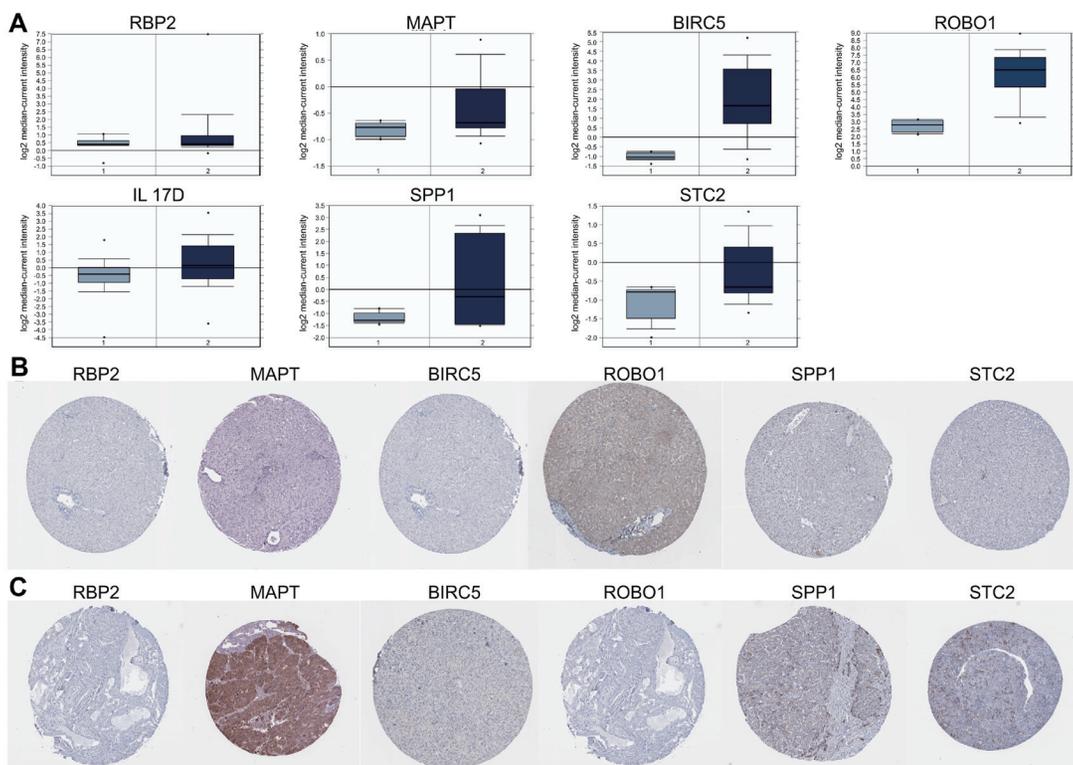

Fig. 7. Validation of the mRNA and protein expression of the IRGs involved in IRPS. (A) The mRNA expression of the eight IRGs in HCC samples according to the Oncomine database (https://www.oncomine.org/resource/main.html). Data of FIGNL2 was not found on the dataset. (B) The representative immunohistochemical images of the eight IRGs in HCC and normal liver samples. Data were obtained from the HPA database (http://www.proteinatlas.org). Expression data of FIGNL2 and IL17D was not found on the website.
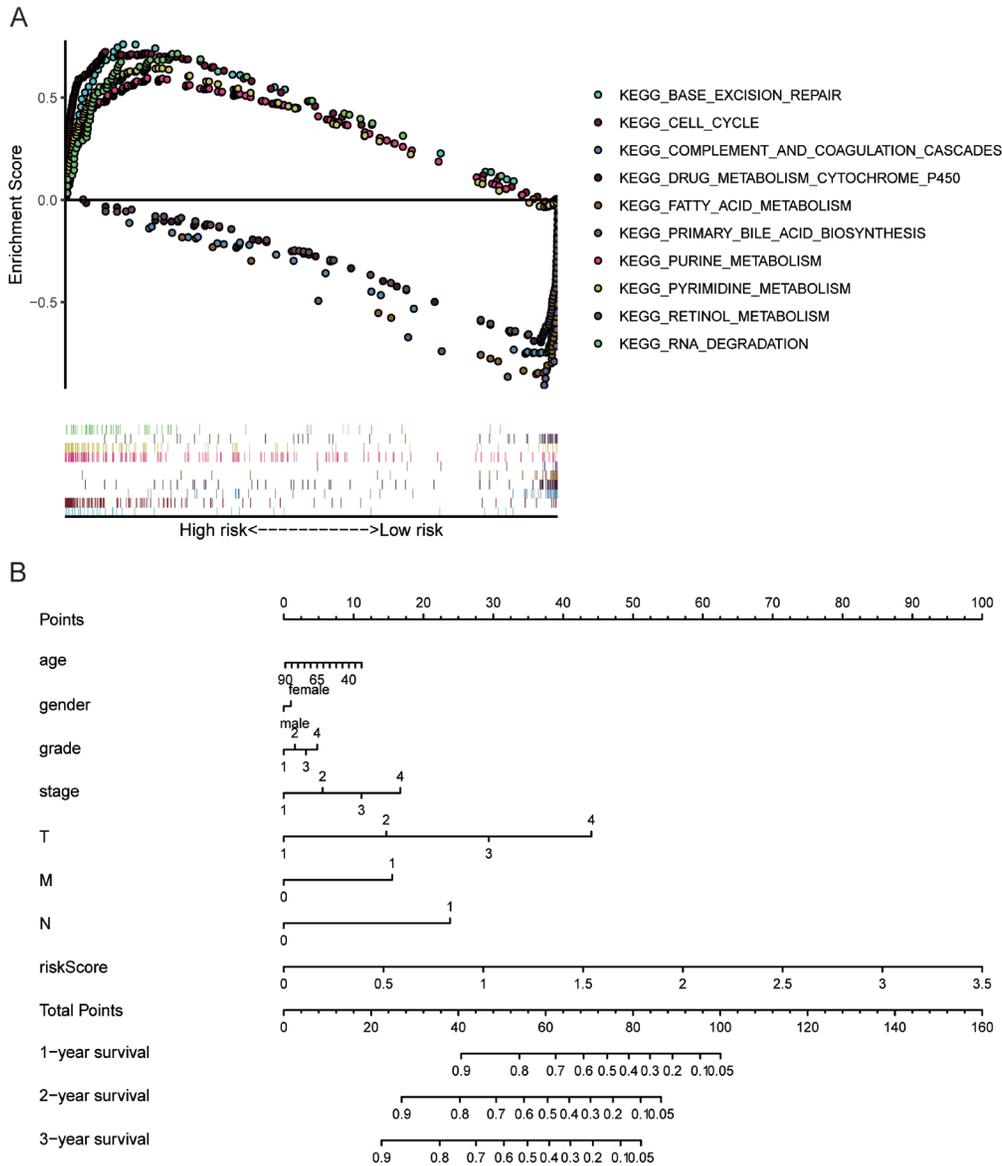
Fig. 8. **GSEA of the significantly enriched KEGG pathways and the nomogram for predicting the survival probability of HCC patients.** (A) GSEA of the significantly enriched KEGG pathways in TCGA cohort. (B) The nomogram for predicting the survival probability of HCC patients in TCGA cohort.

nomogram, which was consistent with the results of independence evaluation analyses.

Discussion

HCC is one of the most prevalent and life-threatening cancers around the world, with rapid progression and difficulty in treatment.[16] Although radiotherapy, molecular targeted therapy, and other treatment regimens can modestly prolong the survival time of HCC patients, the clinical outcomes of these patients are still unsatisfactory.[1,17,18] It has been reported that the immune system is involved in tumor development and progression. However, immune PD-1/PD-L1 checkpoint therapy, such as nivolumab and pembrolizumab, remains to be further investigated due to the unsatisfactory OS.[5,18]

Thus far, there is no established personal treatment strategy for HCC patients. Individual risk assessment might be an essential step for the successful implementation of personal treatment strategies.[19–21] Construction of a gene signature has been proposed to estimate the treatment outcomes of patients with HCC.[22–26] Thus, the prognostic signatures based on IRGs might serve as effective targets for HCC treatment.[27,28]

In the present study, we screened out 62 candidate prognostic IRGs based on the transcriptional data of the TCGA cohort. A TF regulatory network was established to elucidate the interactions among these candidate prognostic IRGs. According to the results of LASSO-penalized Cox analysis, a novel IRPS was constructed and validated with the ICGC cohort. This prognostic signature could efficiently classify the OS of HCC patients into two different cohorts. The predictive performance of the IRPS was excellent in the TCGA and ICGC cohorts, with ROC=0.825 and 0.668, respectively. Furthermore, this IRPS was an independent prognostic factor of HCC and positively correlated with clinicopathological

characteristics in the two cohorts, indicating a high effectiveness of the IRPS. The mRNA and protein expression of these genes involved in IRPS were confirmed through the Oncomine and HPA databases, respectively. More importantly, the cut-off value of our signature in both the TCGA and ICGC cohorts are the same, indicating that our prognostic signature is of high clinical practicability. In addition, a nomogram based on our prognostic signature combined with other clinical information was established to predicate the survival probability of patients with HCC, which could provide a convenient and efficient means for clinical use.

Besides, our IRG-based prognostic signature demonstrated a negative correlation with CD4+ T cells and CD8+ T cells. These results implied that HCC patients in the high-risk group might exhibit decreased infiltration levels of CD4+ T and CD8+ T cells. Tumor-infiltrating T cells have been shown to control the progression of HCC, thus contributing to better prognosis.[29] Ma and co-workers[30] founded that CD4+ T cells loss might be the mechanism underlying hepatocarcinogenesis induced by non-alcoholic fatty liver disease. Fu and co-workers[31] reported that the abnormal levels CD4+ cytotoxic T cells was associated with high recurrence and poor survival rates in HCC patients. In addition, CD8+ T cells were linked to prolonged survival in patients with HCC.[7,32] Nonetheless, the relationship between IRGs and immune cells in HCC is still indistinct. Our results demonstrated that IRPS could accurately reflect the status of HCC tumor immune microenvironments, which provides new insights into the molecular basis of HCC. However, more research is needed in the future.

To disclose the underlying biological mechanisms of the IRPS, GSEA were performed on the high- and low-risk group. In the high-risk group, the top five enriched KEGG pathways were primarily focusing on metabolic processes and cell proliferation, indicating that these IRGs can play an essential role in HCC prognosis.

Most of the eight IRGs have been identified as candidate biomarkers in various types of cancers. RBP2 is highly expressed in HCC tissues compared to normal tissues, and has been found to be correlated with alpha-fetoprotein, tumor differentiation and TNM stage.[33,34] Knocking down of RBP2 significantly suppressed HCC proliferation, and vice versa, indicating that RBP2 may be involved in the pathogenesis of HCC.[33] MAPT has been reported to be associated with poor prognosis and decreased sensitivity to taxane-based therapies in several tumor types.[35–37] Both BIRC5 and SPP1 have been found to promote HCC cell proliferation.[38,39] ROBO1 is overexpressed in HCC samples.[40] Overexpression of ROBO1 could promote HCC cell growth and metastasis both *in vitro* and *in vivo*.[41] STC2 is upregulated in HCC, thereby facilitating HCC proliferation and survival as well as mediating chemotherapeutic resistance.[42–44] The potential molecular mechanisms of genes involved in IRPS to regulate HCC progression is illustrated in Supplementary Figure 2. However, further studies are needed to elucidate the regulatory roles of candidate IRGs in HCC.

Nevertheless, some limitations to this study should be highlighted. First, this signature was constructed based on retrospective data, and thus a prospective study is required to validate our findings. Second, the IRGs in this signature do not represent the same biological processes. Further experiments should be carried out to reveal the exact mechanisms both *in vitro* and *in vivo*. Finally, the relationship between the IRPS and immunotherapy response has not been studied due to lack of patients treated with immunotherapy.

Conclusions

In summary, we constructed a prognostic signature based on IRGs in HCC patients. This signature may provide novel prognostic and therapeutic biomarkers for HCC.

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

Conception and design (HL, YL), administrative support (YL), provision of study materials or patients (XX, HL, PT), Collection and assembly of data (XX, PT), and data analysis and interpretation, manuscript writing, final approval of the manuscript (HL, YL, XX, PT).

Data sharing statement

All data are available upon request.

References

[1] Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019;380(15):1450–1462. doi:10.1056/NEJMra1713263.
[2] Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, *et al*. Annual report to the nation on the status of cancer, 1975-2014, featuring survival. J Natl Cancer Inst 2017;109(9):djx030. doi:10.1093/jnci/djx030.
[3] Kudo M, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, *et al*. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. Lancet 2018;391(10126):1163–1173. doi:10.1016/S0140-6736(18)30207-1.
[4] Chen LT, Martinelli E, Cheng AL, Pentheroudakis G, Qin S, Bhattacharyya GS, *et al*. Pan-Asian adapted ESMO Clinical Practice Guidelines for the management of patients with intermediate and advanced/relapsed hepatocellular carcinoma: a TOS-ESMO initiative endorsed by CSCO, ISMPO, JSMO, KSMO, MOS and SSO. Ann Oncol 2020;31(3):334–351. doi:10.1016/j.annonc.2019.12.001.
[5] Zongyi Y, Xiaowu L. Immunotherapy for hepatocellular carcinoma. Cancer Lett 2020;470:8–17. doi:10.1016/j.canlet.2019.12.002.
[6] Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, *et al*. An immune gene expression signature associated with development of human hepatocellular carcinoma identifies mice that respond to chemopreventive agents. Gastroenterology 2019;157(5):1383–1397.e11. doi:10.1053/j.gastro.2019.07.028.
[7] Itoh S, Yoshizumi T, Yugawa K, Imai D, Yoshiya S, Takeishi K, *et al*. Impact of immune response on outcomes in hepatocellular carcinoma: Association with vascular formation. Hepatology 2020;72(6):1987–1999. doi:10.1002/hep.31206.
[8] Hou J, Zhang H, Sun B, Karin M. The immunobiology of hepatocellular carcinoma in humans and mice: Basic concepts and therapeutic implications. J Hepatol 2020;72(1):167–182. doi:10.1016/j.jhep.2019.08.014.
[9] El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, *et al*. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. Lancet 2017;389(10088):2492–2502. doi:10.1016/S0140-6736(17)31046-2.
[10] Zhu AX, Finn RS, Edeline J, Cattan S, Ogasawara S, Palmer D, *et al*. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase

2 trial. Lancet Oncol 2018;19(7):940–952. doi:10.1016/S1470-2045(18)30351-6.

[11] Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, *et al*. ImmPort: disseminating data to the public for the future of immunology. Immunol Res 2014;58(2-3):234–239. doi:10.1007/s12026-014-8516-1.

[12] Mei S, Meyer CA, Zheng R, Qin Q, Wu Q, Jiang P, *et al*. Cistrome cancer: A web resource for integrative gene regulation modeling in cancer. Cancer Res 2017;77(21):e19–e22. doi:10.1158/0008-5472.CAN-17-0327.

[13] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–2504. doi:10.1101/gr.1239303.

[14] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545–15550. doi:10.1073/pnas.0506580102.

[15] Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, *et al*. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology 2007;45(4):938–947. doi:10.1002/hep.21622.

[16] Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol 2019;16(10):589–604. doi:10.1038/s41575-019-0186-y.

[17] Faivre S, Rimassa L, Finn RS. Molecular therapies for HCC: Looking outside the box. J Hepatol 2020;72(2):342–352. doi:10.1016/j.jhep.2019.09.010.

[18] Liu Z, Lin Y, Zhang J, Zhang Y, Li Y, Liu Z, *et al*. Molecular targeted and immune checkpoint therapy for advanced hepatocellular carcinoma. J Exp Clin Cancer Res 2019;38(1):447. doi:10.1186/s13046-019-1412-8.

[19] Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, *et al*. Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. Gut 2019;68(11):2019–2031. doi:10.1136/gutjnl-2019-318912.

[20] Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 2018;15(10):599–616. doi:10.1038/s41571-018-0073-4.

[21] Huo J, Wu L, Zang Y. A prognostic model of 15 immune-related gene pairs associated with tumor mutation burden for hepatocellular carcinoma. Front Mol Biosci 2020;7:581354. doi:10.3389/fmolb.2020.581354.

[22] Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, *et al*. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 2008;359(19):1995–2004. doi:10.1056/NEJMoa0804525.

[23] Hoshida Y, Villanueva A, Sangiovanni A, Sole M, Hur C, Andersson KL, *et al*. Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. Gastroenterology 2013;144(5):1024–1030. doi:10.1053/j.gastro.2013.01.021.

[24] Ji J, Eggert T, Budhu A, Forgues M, Takai A, Dang H, *et al*. Hepatic stellate cell and monocyte interaction contributes to poor prognosis in hepatocellular carcinoma. Hepatology 2015;62(2):481–495. doi:10.1002/hep.27822.

[25] Kim JH, Sohn BH, Lee HS, Kim SB, Yoo JE, Park YY, *et al*. Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. PLoS Med 2014;11(12):e1001770. doi:10.1371/journal.pmed.1001770.

[26] Zhang DY, Goossens N, Guo J, Tsai MC, Chou HI, Altunkaynak C, *et al*. A hepatic stellate cell gene expression signature associated with outcomes in hepatitis C cirrhosis and hepatocellular carcinoma after curative resection. Gut 2016;65(10):1754–1764. doi:10.1136/gutjnl-2015-309655.

[27] Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, *et al*. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21(8):938–945. doi:10.1038/nm.3909.

[28] Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, *et al*. Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. Gastroenterology 2017;153(3):812–826. doi:10.1053/j.gastro.2017.06.007.

[29] Garnelo M, Tan A, Her Z, Yeong J, Lim CJ, Chen J, *et al*. Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. Gut 2017;66(2):342–351. doi:10.1136/gutjnl-2015-310814.

[30] Ma C, Kesarwala AH, Eggert T, Medina-Echeverz J, Kleiner DE, Jin P, *et al*. NAFLD causes selective CD4(+) T lymphocyte loss and promotes hepatocarcinogenesis. Nature 2016;531(7593):253–257. doi:10.1038/nature16969.

[31] Fu J, Zhang Z, Zhou L, Qi Z, Xing S, Lv J, *et al*. Impairment of CD4+ cytotoxic T cells predicts poor survival and high recurrence rates in patients with hepatocellular carcinoma. Hepatology 2013;58(1):139–149. doi:10.1002/hep.26054.

[32] Brunner SM, Rubner C, Kesselring R, Martin M, Griesshammer E, Ruemmele P, *et al*. Tumor-infiltrating, interleukin-33-producing effector-memory CD8(+) T cells in resected hepatocellular carcinoma prolong patient survival. Hepatology 2015;61(6):1957–1967. doi:10.1002/hep.27728.

[33] Liang X, Zeng J, Wang L, Fang M, Wang Q, Zhao M, *et al*. Histone demethylase retinoblastoma binding protein 2 is overexpressed in hepatocellular carcinoma and negatively regulated by hsa-miR-212. PLoS One 2013;8(7):e69784. doi:10.1371/journal.pone.0069784.

[34] Wang ZY, Yang J, Liu CK, Shen SQ. High expression of retinoblastoma-binding protein 2 (RBP2) in patients with hepatocellular carcinoma and its prognostic significance. Med Sci Monit 2017;23:2736–2744. doi:10.12659/msm.905262.

[35] He W, Zhang D, Jiang J, Liu P, Wu C. The relationships between the chemosensitivity of human gastric cancer to paclitaxel and the expressions of class III β-tubulin, MAPT, and survivin. Med Oncol 2014;31(5):950. doi:10.1007/s12032-014-0950-3.

[36] Schroeder C, Grell J, Hube-Magg C, Kluth M, Lang D, Simon R, *et al*. Aberrant expression of the microtubule-associated protein tau is an independent prognostic feature in prostate cancer. BMC Cancer 2019;19(1):193. doi:10.1186/s12885-019-5390-1.

[37] Wu H, Huang M, Lu M, Zhu W, Shu Y, Cao P, *et al*. Regulation of microtubule-associated protein tau (MAPT) by miR-34c-5p determines the chemosensitivity of gastric cancer to paclitaxel. Cancer Chemother Pharmacol 2013;71(5):1159–1171. doi:10.1007/s00280-013-2108-y.

[38] Wei Y, Chen X, Liang C, Ling Y, Yang X, Ye X, *et al*. A noncoding regulatory RNAs network driven by Circ-CDYL acts specifically in the early stages hepatocellular carcinoma. Hepatology 2020;71(1):130–147. doi:10.1002/hep.30795.

[39] Cai C, Yang L, Tang Y, Wang H, He Y, Jiang H, *et al*. Prediction of overall survival in gastric cancer using a nine-lncRNA. DNA Cell Biol 2019;38(9):1005–1012. doi:10.1089/dna.2019.4832.

[40] Ito H, Funahashi S, Yamauchi N, Shibahara J, Midorikawa Y, Kawai S, *et al*. Identification of ROBO1 as a novel hepatocellular carcinoma antigen and a potential therapeutic and diagnostic target. Clin Cancer Res 2006;12(11 Pt 1):3257–3264. doi:10.1158/1078-0432.CCR-05-2787.

[41] Yuan M, Guo H, Li J, Sui C, Qin Y, Wang J, *et al*. Slit2 and Robo1 induce opposing effects on metastasis of hepatocellular carcinoma Sk-hep-1 cells. Int J Oncol 2016;49(1):305–315. doi:10.3892/ijo.2016.3506.

[42] Cheng H, Wu Z, Wu C, Wang X, Liow SS, Li Z, *et al*. Overcoming STC2 mediated drug resistance through drug and gene co-delivery by PHB-PDMAEMA cationic polyester in liver cancer cells. Mater Sci Eng C Mater Biol Appl 2018;83:210–217. doi:10.1016/j.msec.2017.08.075.

[43] Wu F, Li TY, Su SC, Yu JS, Zhang HL, Tan GQ, *et al*. STC2 as a novel mediator for Mus81-dependent proliferation and survival in hepatocellular carcinoma. Cancer Lett 2017;388:177–186. doi:10.1016/j.canlet.2016.11.039.

[44] Zhang ZH, Wu YG, Qin CK, Rong ZH, Su ZX, Xian GZ. Stanniocalcin 2 expression predicts poor prognosis of hepatocellular carcinoma. Oncol Lett 2014;8(5):2160–2164. doi:10.3892/ol.2014.2520.