

# Identification and validation of an immune-related prognostic signature for hepatocellular carcinoma

Xinxin Xia

Xi'an Jiaotong University Medical College First Affiliated Hospital

Hui Liu

Hunan University of Chinese Medicine

Yuejun Li (✉ [liyuejun97@foxmail.com](mailto:liyuejun97@foxmail.com))

The Third Affiliated Hospital of Hunan University of Chinese Medicine <https://orcid.org/0000-0001-5714-2323>

---

## Primary research

**Keywords:** hepatocellular carcinoma, immune-related gene, prognostic signature

**Posted Date:** July 1st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-25879/v3>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related mortality. The immune system plays vital roles in 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. Univariate Cox regression analysis was carried out to identify differentially expressed genes that associated with overall survival. The IRPS was established via Lasso and multivariate Cox regression analysis. Both Cox regression analyses were conducted to determine the independent prognostic factors for HCC. Next, the association between the IRPS and clinical-related factors were evaluated. The prognostic values of the IRPS were further validated using the International Cancer Genome Consortium (ICGC) dataset. Gene set enrichment analyses (GSEA) were conducted to understand the biological mechanisms of the IRPS.

**Results:** A total of 62 genes were identified to be candidate immune-related prognostic genes. Transcription factors-immunogenes network was generated to explore the interactions among these candidate genes. According to the results of Lasso and multivariate Cox regression analysis, we established an IRPS and confirmed its stability and reliability in ICGC dataset. The IRPS was significantly associated with advanced clinicopathological characteristics. Both Cox regression analyses revealed that the IRPS could be an independent risk factor influencing the prognosis of HCC patients. The relationships between the IRPS and infiltration immune cells demonstrated that the IRPS was associated with immune cell infiltration. GSEA identified significantly enriched pathways, which might assist in elucidating the biological mechanisms of the IRPS. 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.

## Background

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 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 administered 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 immune-related genes 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, phase 3 trials of pembrolizumab as the second-line treatment didn't demonstrate 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

### 2.1 Data preprocessing

The messenger RNA (mRNA) expression profiles and clinical data of HCC patients were retrieved from TCGA. The processed RNA-Seq 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 2498 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 to be obtained.

### 2.2 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  $\log_2\text{-foldchange} \geq 1$  and false discovery rate (FDR)  $\leq 0.05$ . 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  $\leq 0.01$  were considered as prognostic IRGs. To assess the regulation of TFs on IRGs, the prognosis-associated TFs (p-value  $\leq 0.01$ ) were screened out using univariate Cox analysis. The correlation analysis between prognosis-related TFs and prognostic IRGs were performed using Pearson test, and the calculations are done using cor.test in R. The significant cut-off values were set as p-value  $\leq 0.001$  and correlation coefficient  $\geq 0.6$ . Cytoscape software was employed to build and visualize the TF-IRG regulatory network [13].

### 2.3 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 survival and glmnet package in R. The prognostic signature was exhibited as risk score =  $(\text{Coefficient}_{\text{mRNA1}} \times \text{expression of mRNA1}) + (\text{Coefficient}_{\text{mRNA2}} \times \text{expression of mRNA2}) + (\text{Coefficient}_{\text{mRNA3}} \times \text{expression of mRNA3}) + \dots + (\text{Coefficient}_{\text{mRNA}_n} \times \text{expression of mRNA}_n)$ . 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 survivalROC package in R.

### 2.4 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, tumor grade, were included in subsequent analyses. Both univariate and multivariate Cox regression analyses were carried out by forward stepwise method. P-value  $\leq 0.05$  was regarded as statistically significant. The results are presented as forest\_plots, which were generated via the survival package in R.

### 2.5 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 beeswarm package in R.

### 2.6 Association analysis between the IRPS and immune cells infiltration

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.

## 2.7 Validation of the IRPS

The validation cohort was retrieved from 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.

## 2.8 External validation of the expression of the immune-related genes involved in IRPS

Furtherly, the mRNA level expression of the immune-related genes involved in IRPS was retrieved from both TCGA and ICGC datasets, and then validated using the Oncomine database (<https://www.oncomine.org/resource/main.html>). The protein expression of the genes involved in IRPS was also confirmed through the Human Protein Atlas (HPA) database (<http://www.proteinatlas.org>).

## 2.9 Gene set enrichment analyses

To disclose the biological information of the IRPS, Gene Set Enrichment Analyses (GSEA) were conducted to analyze the enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on the C2 in TCGA cohort [14]. P-value $\leq$ 0.05 and FDR $\leq$ 0.05 were deemed as statistically significant.

## 2.10 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.

## 3.0 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 survminer package, while the ROC curves were constructed with survialROC package. The association analysis between the IRPS and clinicopathological characteristics of HCC was conducted by beeswarm package in R.

# Results

## 3.1 Identify immune-related genes 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 shows the distribution of differentially expressed immune-related genes between HCC and

normal tissues (supplementary Fig. a, 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 were shown in Fig. 1.

### 3.2 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. b, d). A regulatory network was built upon these 118 TFs and 62 prognostic IRGs. Correlation score more of  $\geq 0.6$  was set as the cut-off values. The constructed TF-immunogene regulatory network revealed the interaction among these IRGs (Fig.2).

### 3.3 Establishment of the IRPS for HCC prognosis

According to the univariate Cox regression results of 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), stanniocalcin 2 (STC2). The risk scores were calculated as follows:  $0.0132 \times \text{expression of RBP2} + 0.1397 \times \text{expression of MAPT} + 0.0202 \times \text{expression of BIRC5} + 0.0512 \times \text{expression of ROBO1} + 0.2357 \times \text{expression of FIGNL2} + 0.0560 \times \text{expression of IL17D} + 0.0001 \times \text{expression of SPP1} + 0.2095 \times \text{expression of STC2}$ . 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 low-risk group (p-value=7.731e-08; Fig. 3a). The area under the ROC 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 Fig. 3e.

The prognostic signature was validated in 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 TCGA cohort. Notably, the OS of HCC patients was longer in low-risk group than in high-risk group (p-value=3.3632e-02; Fig.3b). The AUC for OS was determined to be 0.668 (Fig. 3d), indicating the high sensitivity and specificity of this prognostic signature. Moreover, the gene expression heatmap and survival overview are illustrated in Fig. 3f.

### 3.4 Independence evaluation of the IRPS

For TCGA cohort, univariate analysis demonstrated that tumor stage, T stage, M stage, and risk score were independent prognostic factors for OS (Fig. 4a.). Furthermore, multivariate Cox regression analysis showed that only risk score remained an independent prognostic factor of the survival outcome of HCC patients after adjusted by age, gender, tumor grade and TNM stage (Fig. 4b). For 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 TCGA cohort. Taken together, these results confirmed that IRPS could serve as a biomarker for predicting survival outcome of HCC patients.

### 3.5 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 TCGA cohort, this prognostic signature was significantly correlated with pathological 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 was associated with more advanced tumor stage in ICGC cohort (Fig. 5d). The clinical significance of each identified IRG is shown in 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 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. These results are shown in Fig. 6.

### 3.6 External validation of the expression of the immune-related genes involved in IRPS

Compared to the adjacent non-tumor liver tissues, the expression of the eight immune-related genes involved in IRPS was significantly increased in HCC in TCGA cohort. These results were validated in ICGC cohort (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 also increased in HCC tissues by using Wurmbach liver cohort [15] (Fig.7a). The protein expression of these eight genes was confirmed in HPA database. Compared to their expression in normal liver tissues, MAPT and ROBO1 were strongly positive, while BIRC5 and SPP1 were moderately positively, and STC2 were weakly positively in HCC tissues (Fig. 7b, c). However, the expression of RBP2 were not detected in both HCC and normal liver tissues (Fig. 7b, c), the expression of FIGNL2 and IL17D were not found on the website.

### 3.7 GSEA

GSEA was preformed and identified 65 significantly enriched KEGG pathways in TCGA cohort. The top 5 KEGG pathways in high-risk group were "KEGG\_PYRIMIDINE\_METABOLISM", "KEGG\_PURINE\_METABOLISM", "KEGG\_RNA\_DEGRADATION", "KEGG\_BASE\_EXCISION\_REPAIR", "KEGG\_CELL\_CYCLE". The top 5 KEGG pathways in low-risk group were "KEGG\_COMPLEMENT\_AND\_COAGULATION\_CASCADES", "KEGG\_FATTY\_ACID\_METABOLISM", "KEGG\_DRUG\_METABOLISM\_CYTOCHROME\_P450", "KEGG\_PRIMARY\_BILE\_ACID\_BIOSYNTHESIS", "KEGG\_RETINOL\_METABOLISM" (Fig.8a).

### 3.8 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, T, N, and M combined with the prognostic signature, were involved. Each factor was assigned a score point according to its risk to survival. In the nomogram plot, the influence of different factors on the survival were presented in the length of the line. The prognostic signature had the longest line indicated 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 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, 20]. Gene signature has been proposed to estimate the treatment outcomes of patients with HCC [21-25]. Thus, the prognostic signatures based on IRGs might serve as effective targets for HCC treatment [26, 27].

In the present study, we screened out 62 candidate prognostic IRGs based on the transcriptional data of 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 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 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 Oncomine and HPA databases, respectively. More importantly, the cut-off value of our signature in both TCGA and ICGC cohort 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 negatively 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 [28]. Ma and co-workers [29] founded that CD4+ T cells loss might be the reason of hepatocarcinogenesis induced by non-alcoholic fatty liver disease. Fu and co-workers [30] 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, 31]. Nonetheless, the relationship between IRGs and immune cells in HCC is still indistinct. Our results may yield new insights into the molecular basis, and further research is required in the future.

To disclose the underlying biological mechanisms of the IRPS, GSEA was performed on the high- and low-risk group. In the high-risk group, the top 5 enriched KEGG pathways were primarily focusing on metabolic process 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 cancer. RBP2 is highly expressed in HCC tissues compared to normal tissues, and has been founded to be correlated with AFP, tumor differentiation and TNM stage [32, 33]. Knocking down of RBP2 significantly suppressed HCC proliferation, and vice versa, indicating that RBP2 may be involved in the pathogenesis of HCC [32]. MAPT has been reported to be associated with poor prognosis and decreased sensitivity to taxane-based therapies in several tumor types [34-36]. Both BIRC5 and SPP1 have been found to promote HCC cell proliferation [37, 38]. ROBO1 is overexpressed in HCC samples [39]. Overexpression of ROBO1 could promote HCC cell growth and metastasis both *in vitro* and *in vivo* [40]. STC2 is upregulated in HCC, thereby facilitating HCC proliferation and survival as well as mediating chemotherapeutic resistance [41-43]. However, further studies are needed to uncover the exact functions of candidate IRGs in HCC.

Nevertheless, some limitations to this study should be highlighted. Firstly, this signature was constructed based on retrospective data, a prospective study is required to validate our findings. Secondly, the IRGs in this signature does 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 IRSs in HCC patients. This signature may provide novel prognostic and therapeutic biomarkers for HCC.

## Abbreviations

HCC, hepatocellular carcinoma; OS, overall survival; IRGs, immune-related genes; PD-L1, immune gene programmed death-ligand 1; ImmPort, the Immunology Database and Analysis Portal database; IRPS, immune-related prognosis signature; TCGA, The Cancer Genome Atlas; ICGC, the International Cancer Genome Consortium; mRNA, messenger RNA; TFs, transcription factors; DE-IRGs, differentially expressed

IRGs; LASSO, Least Absolute Shrinkage and Selection Operator; K-M survival, Kaplan-Meier survival; ROC, receiver operating characteristic; TIMER, Tumor Immune Estimation Resource; GSEA, Gene Set Enrichment Analyses; KEGG, Kyoto Encyclopedia of Genes and Genomes; RBP2, retinol binding protein 2; MAPT, microtubule associated protein tau; BIRC5, baculoviral IAP repeat containing 5; ROBO1, roundabout guidance receptor 1; FIGNL2, fidgetin like 2; IL17D, interleukin 17D; SPP1, secreted phosphoprotein 1; STC2, stanniocalcin 2.

## Declarations

### Ethical approval and consent to participate

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was consistent with the publication guidelines provided by the TCGA and ICGC; therefore, further approval was not required from the ethics committee.

### Consent for publication

All the data were obtained from open access database, consent for publication is not needed.

### Availability of data and materials

The data used in the present study can be found at TCGA (<https://portal.gdc.cancer.gov/>) and ICGC database (<https://dcc.icgc.org/releases/current/Projects>).

### Competing interests

The authors have no competing interests to declare.

### Funding

This study was supported by the National Natural Science Foundation of China (81703916), Natural Science Foundation of Hunan Province (2018JJ6042).

### Author contributions

(✉) Conception and design: Hui LIU, Yuejun LI; (✉) Administrative support: Yuejun LI; (✉) Provision of study materials or patients: Xinxin XIA, Hui LIU; (✉) Collection and assembly of data: Xinxin XIA; (✉) Data analysis and interpretation: All authors; (✉) Manuscript writing: All authors; (✉) Final approval of manuscript: All authors.

### Acknowledgements

The authors would like to express their gratitude to EditSprings (<https://www.editsprings.com/>) for the expert linguistic services provided.

# References

1. Villanueva A: Hepatocellular Carcinoma. *N Engl J Med.* 2019; 380(15):1450-1462.
2. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, Mariotto A, Lake AJ, Wilson R, Sherman RL *et al*: Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. *J Natl Cancer Inst.* 2017; 109(9).
3. Kudo M, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, Baron A, Park JW, Han G, Jassem J *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.
4. Chen LT, Martinelli E, Cheng AL, Pentheroudakis G, Qin S, Bhattacharyya GS, Ikeda M, Lim HY, Ho GF, Choo SP *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.
5. Zongyi Y, Xiaowu L: Immunotherapy for hepatocellular carcinoma. *Cancer Lett.* 2020; 470:8-17.
6. Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, Higuera M, Pfister D, Ramadori P, Pinyol R *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 e1311.
7. Itoh S, Yoshizumi T, Yugawa K, Imai D, Yoshiya S, Takeishi K, Toshima T, Harada N, Ikegami T, Soejima Y *et al*: Impact of Immune Response on Outcomes in Hepatocellular Carcinoma: Association with Vascular Formation. *Hepatology.* 2020.
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.
9. El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, Kim TY, Choo SP, Trojan J, Welling THR *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.
10. Zhu AX, Finn RS, Edeline J, Cattani S, Ogasawara S, Palmer D, Verslype C, Zagonel V, Fartoux L, Vogel A *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.
11. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J *et al*: ImmPort: disseminating data to the public for the future of immunology. *Immunol Res.* 2014; 58(2-3):234-239.
12. Mei S, Meyer CA, Zheng R, Qin Q, Wu Q, Jiang P, Li B, Shi X, Wang B, Fan J *et al*: Cistrome Cancer: A Web Resource for Integrative Gene Regulation Modeling in Cancer. *Cancer Res.* 2017; 77(21):e19-e22.
13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks.

Genome Res. 2003; 13(11):2498-2504.

14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *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.
15. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J *et al*: Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology*. 2007; 45(4):938-947.
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.
17. Faivre S, Rimassa L, Finn RS: Molecular therapies for HCC: Looking outside the box. *J Hepatol*. 2020; 72(2):342-352.
18. Liu Z, Lin Y, Zhang J, Zhang Y, Li Y, Liu Z, Li Q, Luo M, Liang R, Ye J: Molecular targeted and immune checkpoint therapy for advanced hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2019; 38(1):447.
19. Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, Wang L, Huang X, Fu Q, Ye M *et al*: Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut*. 2019; 68(11):2019-2031.
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.
21. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, Moore J, Wrobel MJ, Lerner J *et al*: Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med*. 2008; 359(19):1995-2004.
22. Hoshida Y, Villanueva A, Sangiovanni A, Sole M, Hur C, Andersson KL, Chung RT, Gould J, Kojima K, Gupta S *et al*: Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. *Gastroenterology*. 2013; 144(5):1024-1030.
23. Ji J, Eggert T, Budhu A, Forgues M, Takai A, Dang H, Ye Q, Lee JS, Kim JH, Greten TF *et al*: Hepatic stellate cell and monocyte interaction contributes to poor prognosis in hepatocellular carcinoma. *Hepatology*. 2015; 62(2):481-495.
24. Kim JH, Sohn BH, Lee HS, Kim SB, Yoo JE, Park YY, Jeong W, Lee SS, Park ES, Kaseb A *et al*: Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med*. 2014; 11(12):e1001770.
25. Zhang DY, Goossens N, Guo J, Tsai MC, Chou HI, Altunkaynak C, Sangiovanni A, Iavarone M, Colombo M, Kobayashi M *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.
26. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD *et al*: The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. 2015; 21(8):938-945.

27. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, Putra J, Camprecios G, Bassaganyas L, Akers N *et al*: Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology*. 2017; 153(3):812-826.
28. Garnelo M, Tan A, Her Z, Yeong J, Lim CJ, Chen J, Lim KH, Weber A, Chow P, Chung A *et al*: Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. *Gut*. 2017; 66(2):342-351.
29. Ma C, Kesarwala AH, Eggert T, Medina-Echeverez J, Kleiner DE, Jin P, Stroncek DF, Terabe M, Kapoor V, ElGindi M *et al*: NAFLD causes selective CD4(+) T lymphocyte loss and promotes hepatocarcinogenesis. *Nature*. 2016; 531(7593):253-257.
30. Fu J, Zhang Z, Zhou L, Qi Z, Xing S, Lv J, Shi J, Fu B, Liu Z, Zhang JY *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.
31. Brunner SM, Rubner C, Kesselring R, Martin M, Griesshammer E, Ruemmele P, Stempfl T, Teufel A, Schlitt HJ, Fichtner-Feigl S: Tumor-infiltrating, interleukin-33-producing effector-memory CD8(+) T cells in resected hepatocellular carcinoma prolong patient survival. *Hepatology*. 2015; 61(6):1957-1967.
32. Liang X, Zeng J, Wang L, Fang M, Wang Q, Zhao M, Xu X, Liu Z, Li W, Liu S *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.
33. 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.
34. 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 beta-tubulin, MAPT, and survivin. *Med Oncol*. 2014; 31(5):950.
35. Schroeder C, Grell J, Hube-Magg C, Kluth M, Lang D, Simon R, Hoflmayer D, Minner S, Burandt E, Clauditz TS *et al*: Aberrant expression of the microtubule-associated protein tau is an independent prognostic feature in prostate cancer. *BMC Cancer*. 2019; 19(1):193.
36. Wu H, Huang M, Lu M, Zhu W, Shu Y, Cao P, Liu P: 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.
37. Wei Y, Chen X, Liang C, Ling Y, Yang X, Ye X, Zhang H, Yang P, Cui X, Ren Y *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.
38. Cai C, Yang L, Tang Y, Wang H, He Y, Jiang H, Zhou K: Prediction of Overall Survival in Gastric Cancer Using a Nine-lncRNA. *DNA Cell Biol*. 2019; 38(9):1005-1012.
39. Ito H, Funahashi S, Yamauchi N, Shibahara J, Midorikawa Y, Kawai S, Kinoshita Y, Watanabe A, Hippo Y, Ohtomo T *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.

40. Yuan M, Guo H, Li J, Sui C, Qin Y, Wang J, Khan YH, Ye L, Xie F, Wang H *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.
41. Cheng H, Wu Z, Wu C, Wang X, Liow SS, Li Z, Wu YL: 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.
42. Wu F, Li TY, Su SC, Yu JS, Zhang HL, Tan GQ, Liu JW, Wang BL: STC2 as a novel mediator for Mus81-dependent proliferation and survival in hepatocellular carcinoma. Cancer Lett. 2017; 388:177-186.
43. 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.

## Tables

Table 1. Association of clinicopathologic characteristics with the IRGs and prognostic signature in TCGA cohort.

Gene	Age		Gender		Tumor grade		Tumor stage		T stage		N stage		M stage	
	t	P	t	p	t	P	t	P	t	P	t	P	t	P
RBP2	0.929	0.354	1.131	0.262	8.36	0.039	5.934	0.115	2.504	0.474	0.724	0.486	2.034	0.043
MAPT	-0.696	0.487	1.184	0.239	1.023	0.796	10.709	0.013	6.241	0.1	9.634	1.07E-17	1.769	0.183
BIRC5	2.644	0.009	-0.855	0.393	11.509	0.009	6.491	0.09	15.6	0.001	-0.444	0.698	4.946	7.73E-04
ROBO1	-0.957	0.34	-0.349	0.728	11.443	0.01	7.895	0.048	4.796	0.187	-1.039	0.407	-0.551	0.637
FIGNL2	1.66	0.098	1.463	0.147	7.339	0.062	9.955	0.019	2.065	0.559	-0.732	0.54	4.081	0.01
IL17D	0.833	0.406	1.463	0.148	12.565	0.006	4.585	0.205	1.001	0.801	3.86	1.64E-04	1.209	0.297
SPP1	-1.884	0.061	-1.317	0.19	5.642	0.13	7.103	0.069	17.649	5.20E-04	1.32	0.28	-0.204	0.855
STC2	-1.175	0.243	-0.455	0.65	0.832	0.842	5.01	0.171	11.262	0.01	0.473	0.67	-0.565	0.626
riskScore	1.331	0.185	1.378	0.171	13.863	0.003	-1.467	0.184	18.289	3.83E-04	-0.166	0.883	2.052	0.144

Note: t, t value of student's test; P, p-value of student's test; IRGs, immune-related genes.

Table 2. Associations of clinicopathologic characteristics with the IRGs and prognostic signature in ICGC cohort

Gene	Gender		Age		Tumor stage	
	t	P	t	P	t	P
RBP2	0.186	0.853	34.049	0.883	10.997	0.012
MAPT	1.597	0.113	39.44	0.706	3.982	0.263
BIRC5	-0.449	0.654	38.353	0.748	18.838	2.953e-04
ROBO1	0.429	0.669	38.092	0.757	7.156	0.067
FIGNL2	0.723	0.471	49.532	0.297	15.895	0.001
IL17D	0.843	0.401	40.309	0.671	19.834	1.837e-04
SPP1	0.663	0.509	36.239	0.821	17.43	5.764e-04
STC2	-0.71	0.479	44.377	0.498	8.256	0.041
riskScore	1.15	0.253	44.312	0.501	24.743	1.747e-05

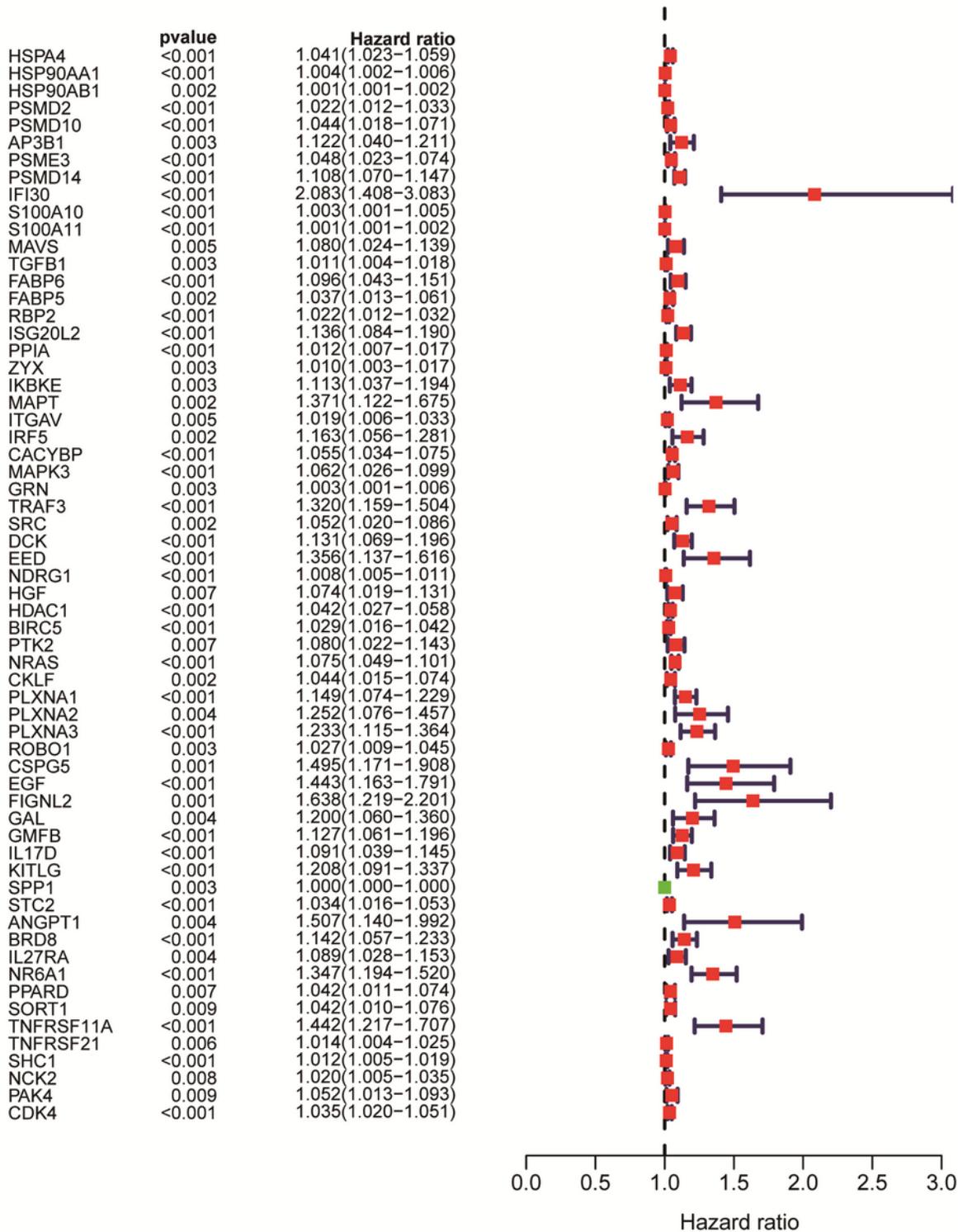
Note: t, t value of student's test; P, p-value of student's test; IRGs, immune-related genes.

Table 3. Differential expression of the immune-related genes involved in IRPS in hepatocellular carcinoma.

Gene	TCGA			ICGC		
	logFC	p-Value	FDR	logFC	p-Value	FDR
RBP2	4.984714	4.18E-05	5.87E-05	2.140651	6.82E-07	9.32E-07
MAPT	3.728369	3.34E-25	6.84E-24	3.156435	2.58E-59	4.32E-57
BIRC5	4.814625	2.35E-28	4.67E-26	3.352928	2.71E-60	6.43E-58
ROBO1	3.534811	7.16E-19	3.48E-18	2.994799	9.71E-50	2.07E-48
FIGNL2	2.311355	1.07E-10	2.27E-10	2.297298	1.46E-39	1.04E-38
IL17D	3.841742	9.47E-13	2.39E-12	2.64293	3.25E-16	6.38E-16
SPP1	5.006114	2.72E-07	4.49E-07	3.120633	9.71E-20	2.20E-19
STC2	2.8251	1.14E-18	5.37E-18	2.248714	4.55E-37	2.66E-36

Note: IRPS, immune-gene related prognostic signature; logFC, log2 of fold change; FDR, false discovery rate.

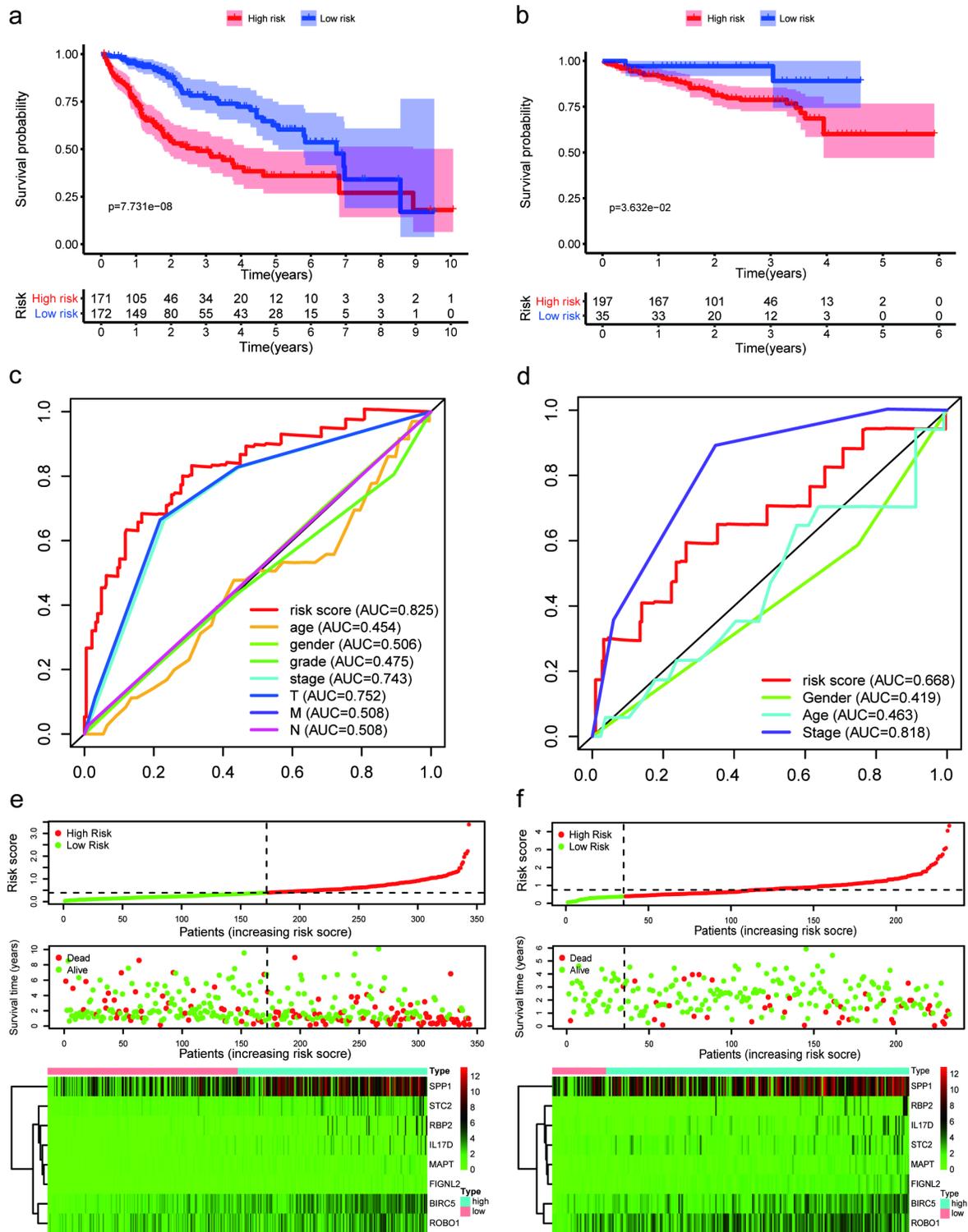
## Figures



**Figure 1**

Univariate Cox data for the overall survival of 62 candidate prognostic IRGs in patients with HCC. IRGs, immune-related genes; HCC, hepatocellular carcinoma.

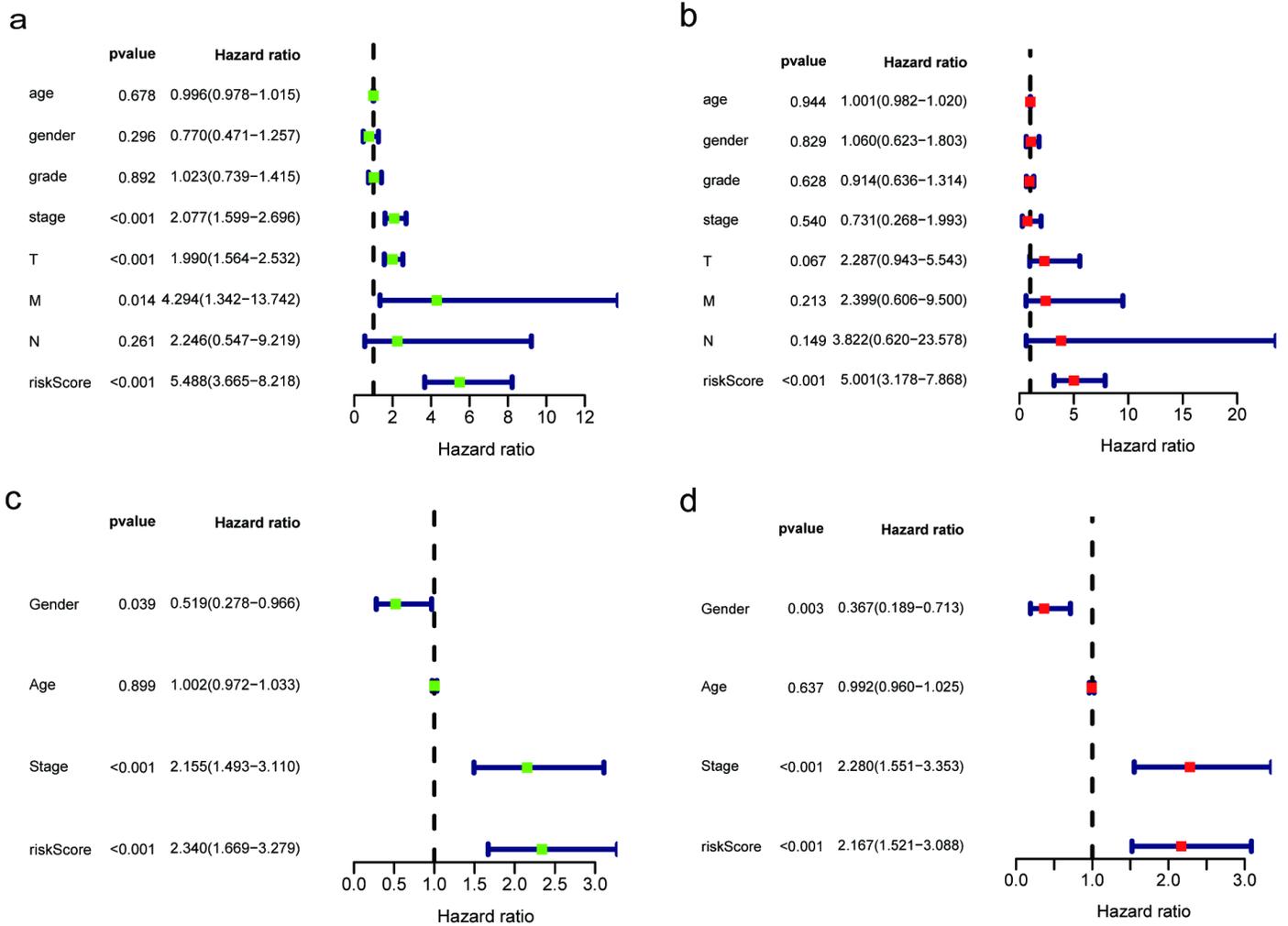




**Figure 3**

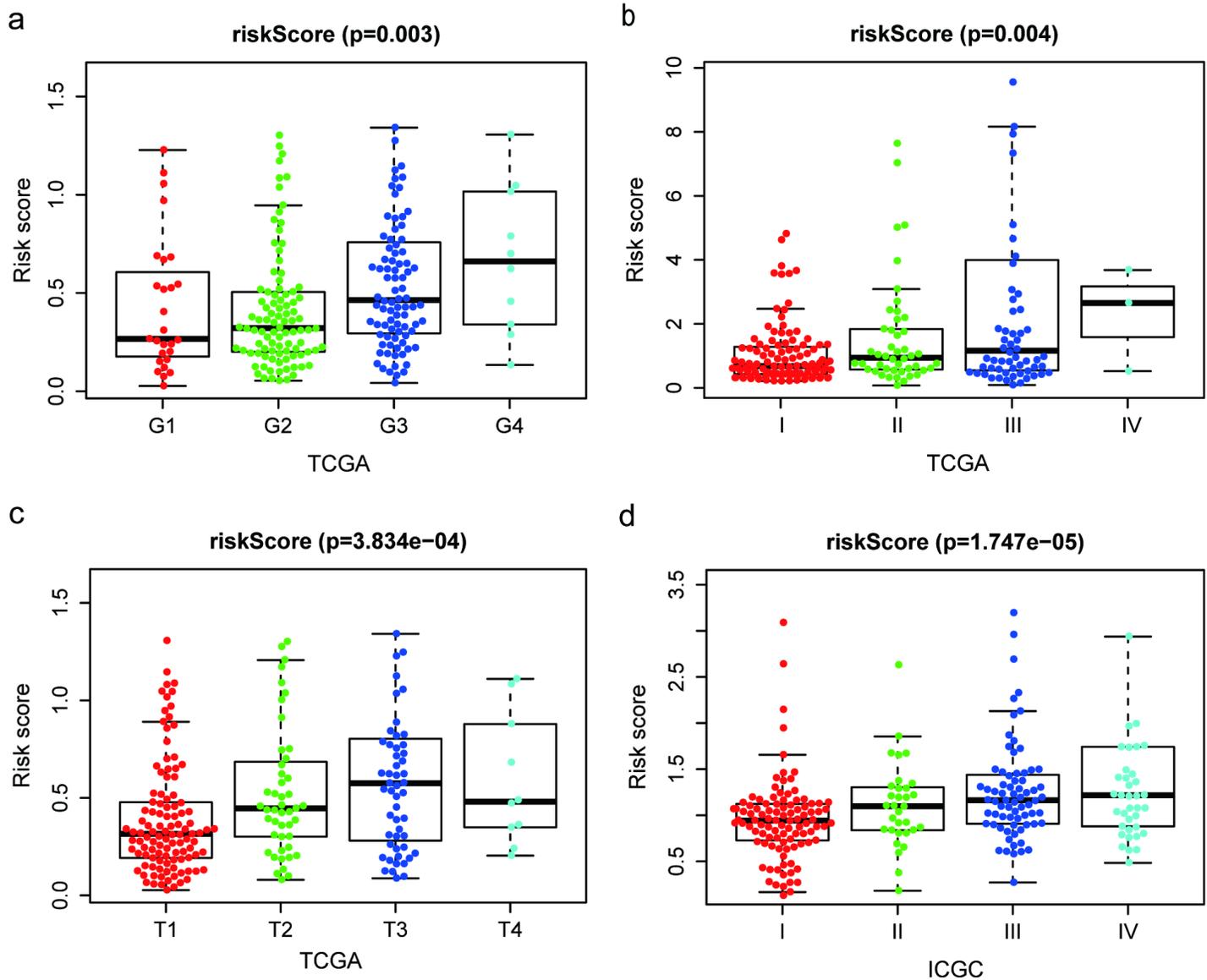
Establishment of the IRPS for HCC patients. Kaplan-Meier Curves of the IRPS in TCGA (a) and ICGC (b) cohort. ROC curves of the IRPS in TCGA (c) and ICGC (d) cohort. (e) Risk score distribution with survival status for the IRPS of HCC patients in TCGA cohort. (f) Risk score distribution with survival status for the IRPS of HCC patients in ICGC cohort. IRPS, immune-related prognostic signature; HCC, hepatocellular

carcinoma; TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium; ROC, receiver operating characteristic.



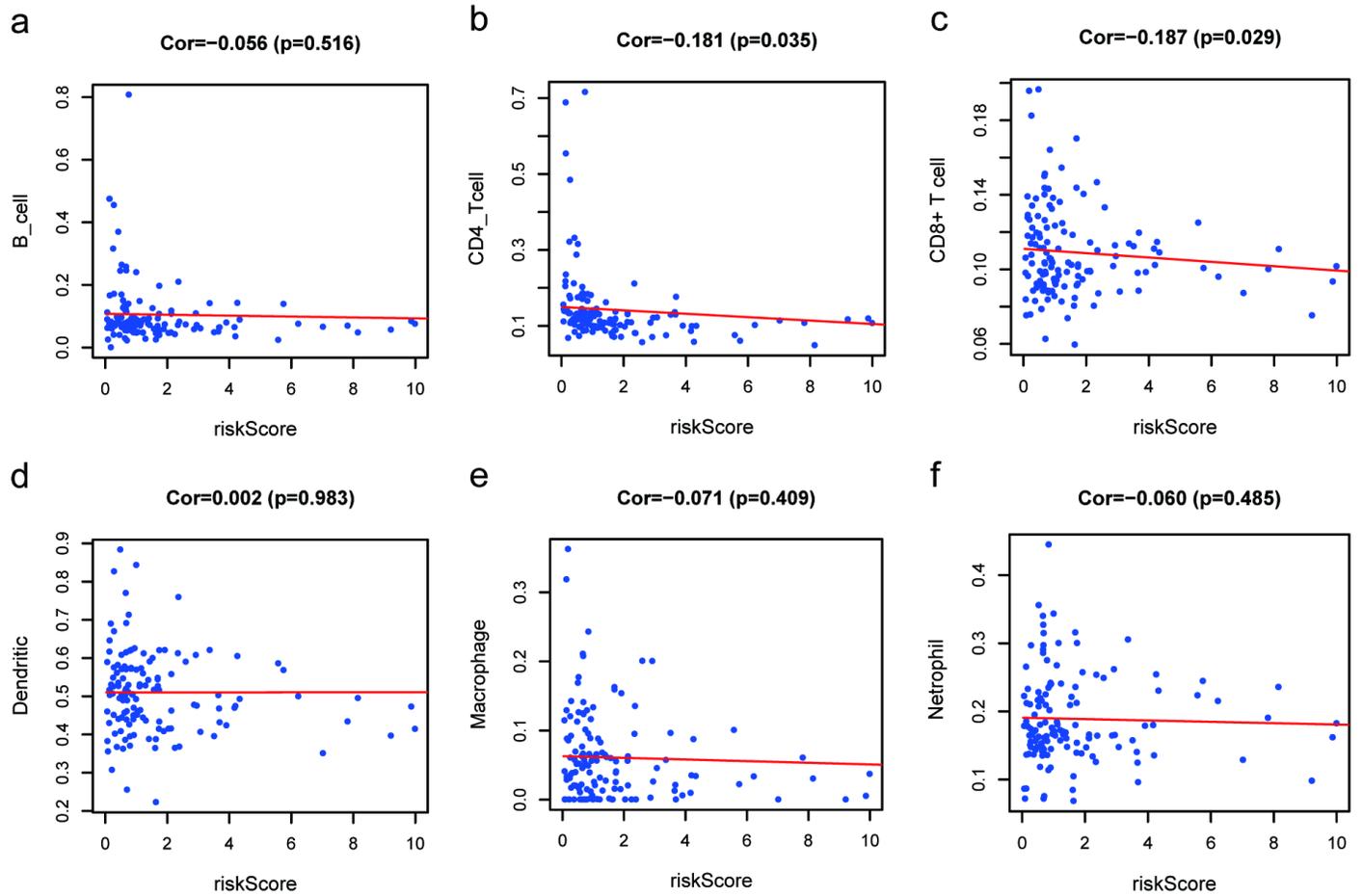
**Figure 4**

Independence evaluation of the IRPS in patients with HCC. Univariate Cox regression analysis for the independence evaluation of the IRPS in TCGA (a) and ICGC cohort (c). Multivariate Cox regression analysis for the independence evaluation of the IRPS in TCGA (b) and ICGC cohort (d). IRPS, immune-related prognostic signature; HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium.



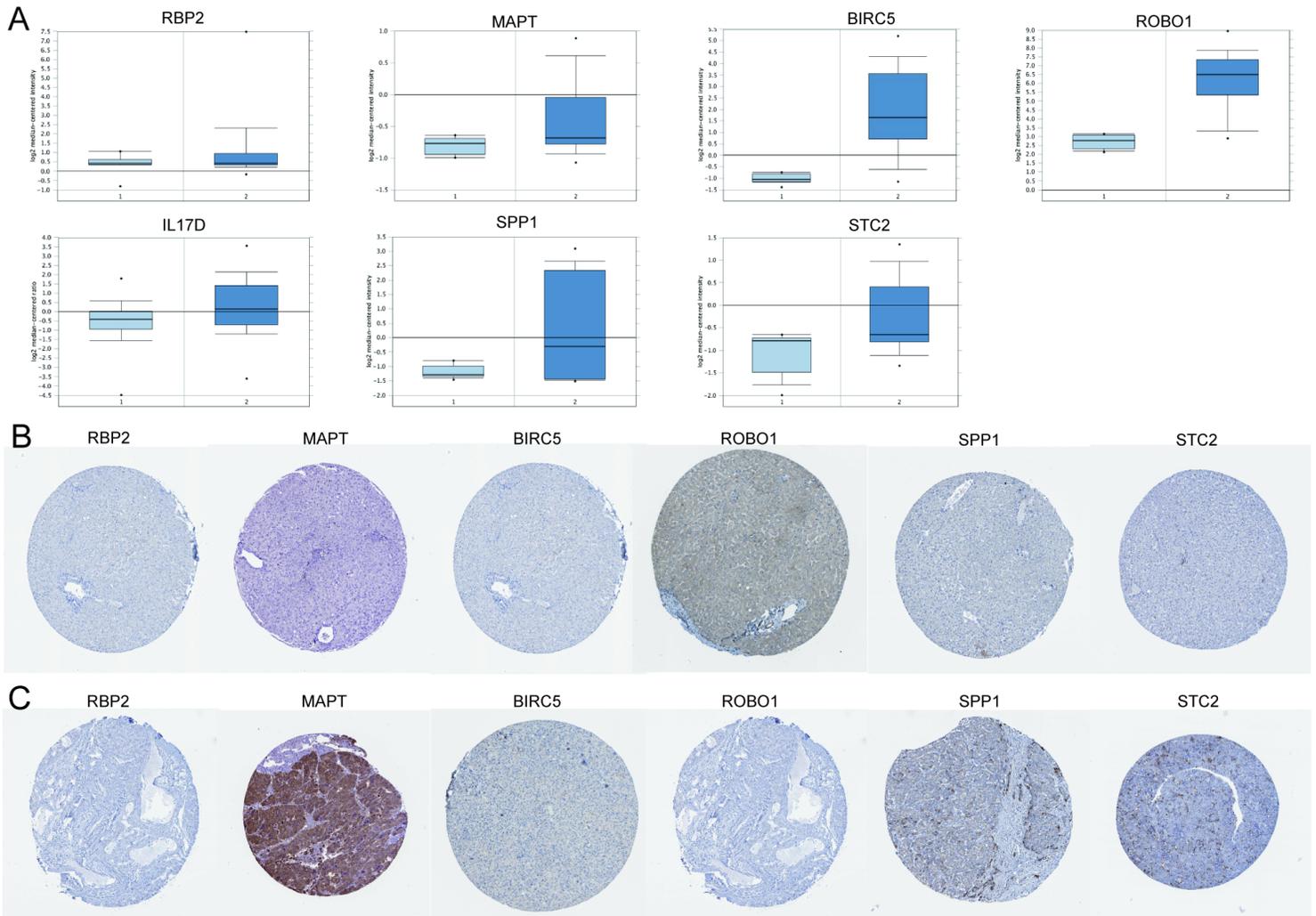
**Figure 5**

Association analyses between the IRPS and clinicopathological characteristics. The IRPS was associated with more advanced tumor stage (a), T stage (b), and poorly differentiated carcinoma (c) in TCGA cohort. (d) The IRPS was associated with more advanced tumor stage in ICGC cohort. IRPS, immune-related prognostic signature; TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium.



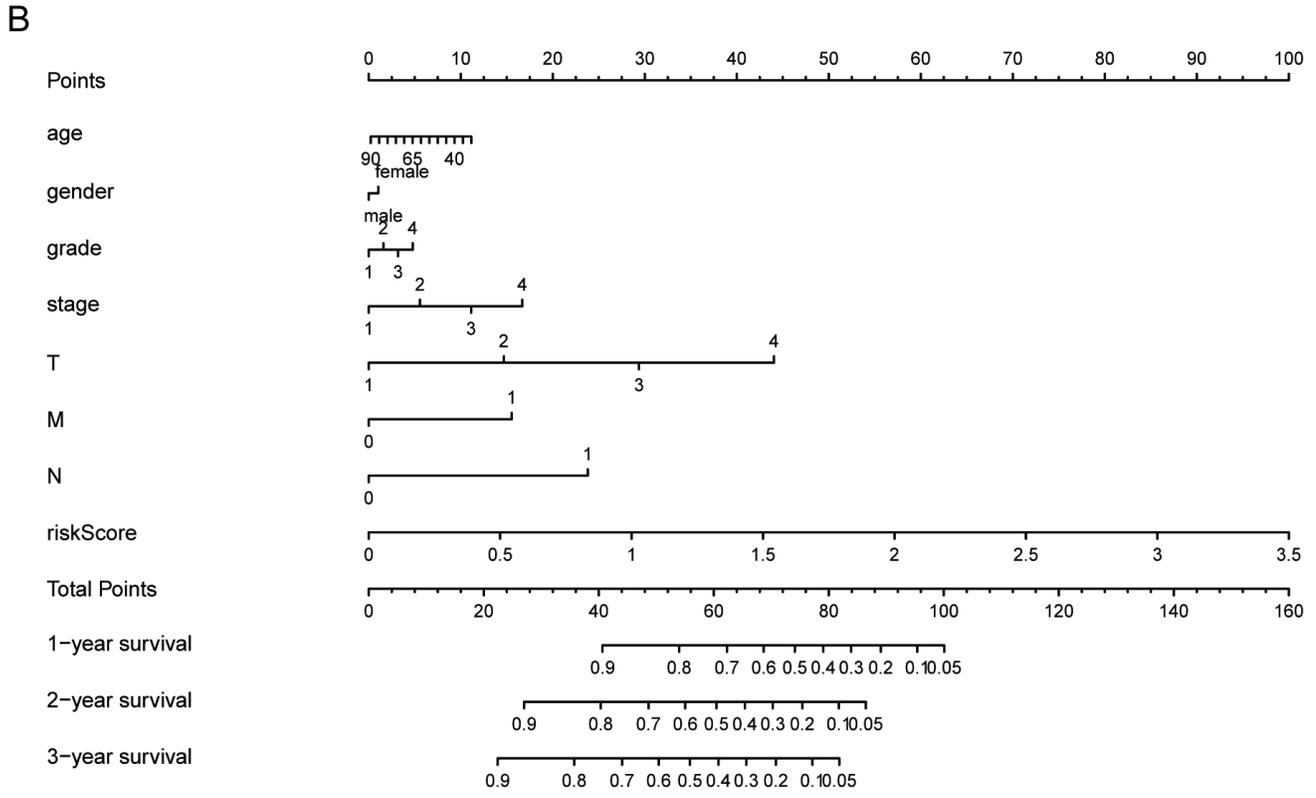
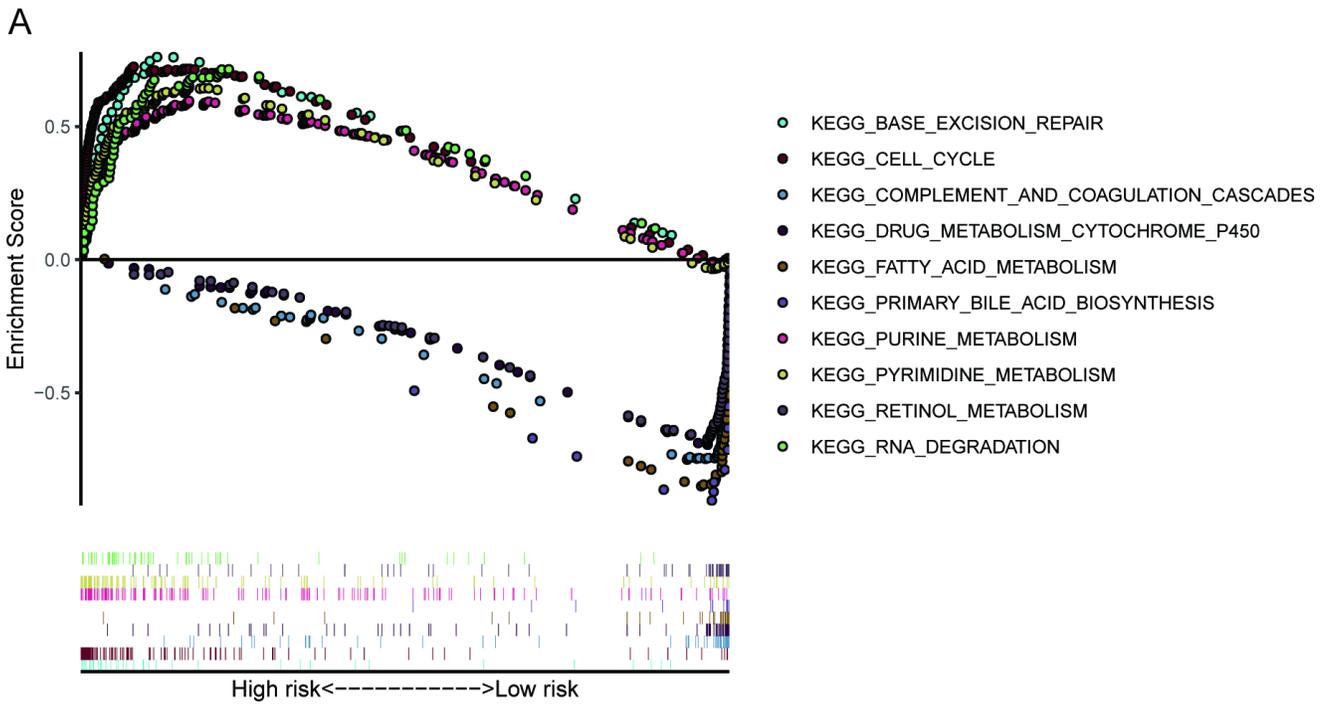
**Figure 6**

Relationships between IRPS and immune cells infiltration. (a) B cells; (b) CD4 T cells; (c) CD8 T cells; (d) Dendritic cells; (e) Macrophage cells; (f) Neutrophil cells. IRPS, immune-related prognostic signature.



**Figure 7**

Validation of the mRNA and protein expression of the eight immune-related genes involved in IRPS. (a) The mRNA expression of the eight immune-related genes 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 immune-related genes in HCC and normal liver samples. Data were obtained from the Human Protein Atlas database (<http://www.proteinatlas.org>). Expression data of FIGNL2 and IL17D was not found on the website. HCC, hepatocellular carcinoma.



**Figure 8**

(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. GSEA, gene set enrichment analyses; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The Cancer Genome Atlas; HCC, hepatocellular carcinoma.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigureLegend.docx](#)
- [supplementaryFigure.tif](#)