Am J Transl Res 2021;13(8):8766-8776
www.ajtr.org /ISSN:1943-8141/AJTR0131590

Original Article

A novel recurrence associated immune gene signature and its clinical significance in liver cancer

Yubing Chen, Zhenya Guo, Yonglian Zeng, Huizhao Su, Fudi Zhong, Keqing Jiang, Guandou Yuan

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China

Received February 7, 2021; Accepted May 17, 2021; Epub August 15, 2021; Published August 30, 2021

Abstract: Hepatocellular carcinoma (HCC) is one of the most common malignancies across nations. Although the outcome of HCC has been improved significantly with the advances in comprehensive treatment, patients remain suffered from recurrence as well as metastasis. Therefore, it is urgent to identify reliable biomarkers for predicting the recurrence of HCC, by which the treating strategy can be made to restrain tumor progress. Increasing evidence has shown the association between immune signature and prognosis of HCC. Thus, we aimed to discover an immune-related gene signature that can estimate the recurrence rates of HCC. We collected gene expression profiles and clinical information of patients from GEO and TCGA dataset. Furthermore, we conducted a lasso regression analysis and established a recurrence-related model consisting of 36 immune-related gene pairs (IRGPs) with 54 genes. We validated the IRGPs in the validation cohort and observed that the immune-related signature robustly stratified patients with HCC into high- and low-risk groups in terms of recurrence ($P < 0.001$). Multivariant Cox regression analysis showed the relationship between the model and recurrence outcomes (Hazard Ratio: 3.81 95% Confidence Interval: 2.90-5.00). Gene Ontology and KEGG enrichment analyses revealed that those genes were enriched in important signaling pathways. In summary, we developed a robust model based on the signature of immune-related genes for forecasting the recurrence outcome of patients with HCC, which holds the potential to assist clinical practice.

Keywords: Hepatocellular carcinoma, immune gene, recurrence, survival

Introduction

Hepatocellular carcinoma (HCC) is a dismal health concern and the sixth common cancer in the world [1]. The incidence of HCC is continuously rising worldwide [2]. Despite the consistent advances in treatment and diagnosis, the five-year survival rate of liver cancer remains very low, approximately 30.5% based on the Surveillance, Epidemiology, and End Results (SEER) database [3]. Although the therapeutic approaches of liver cancer have been improved remarkably due to the development of chemotherapeutic regimes and immune-related therapies, the most promising option to cure patients with this disease still relies on surgical resection. However, most patients are diagnosed at an advanced stage that is not suitable for curative approaches [4]. Nearly 54% of patients experienced recurrence or distal metastasis [5]. Thus, we decided to find a signature to predict the recurrence outcomes of HCC, which has the potential to guide the treatment options and improve patient survival outcomes.

Immune related genes (IRGs) have been extensively studied in recent years. Immunotherapy has emerged as a new star in cancer treatment, which increasingly extended the survival of patients, at least partially, especially with the help of immune checkpoint inhibitors [6]. The prognosis-predicting value of those genes was investigated in various cancers, indicating the noteworthy of their roles in guiding the management of cancers. Predicting the recurrence of liver cancer plays a curial role in the treatment of HCC. Unfortunately, many kinds of biomarkers were used to estimate the prognosis of HCC

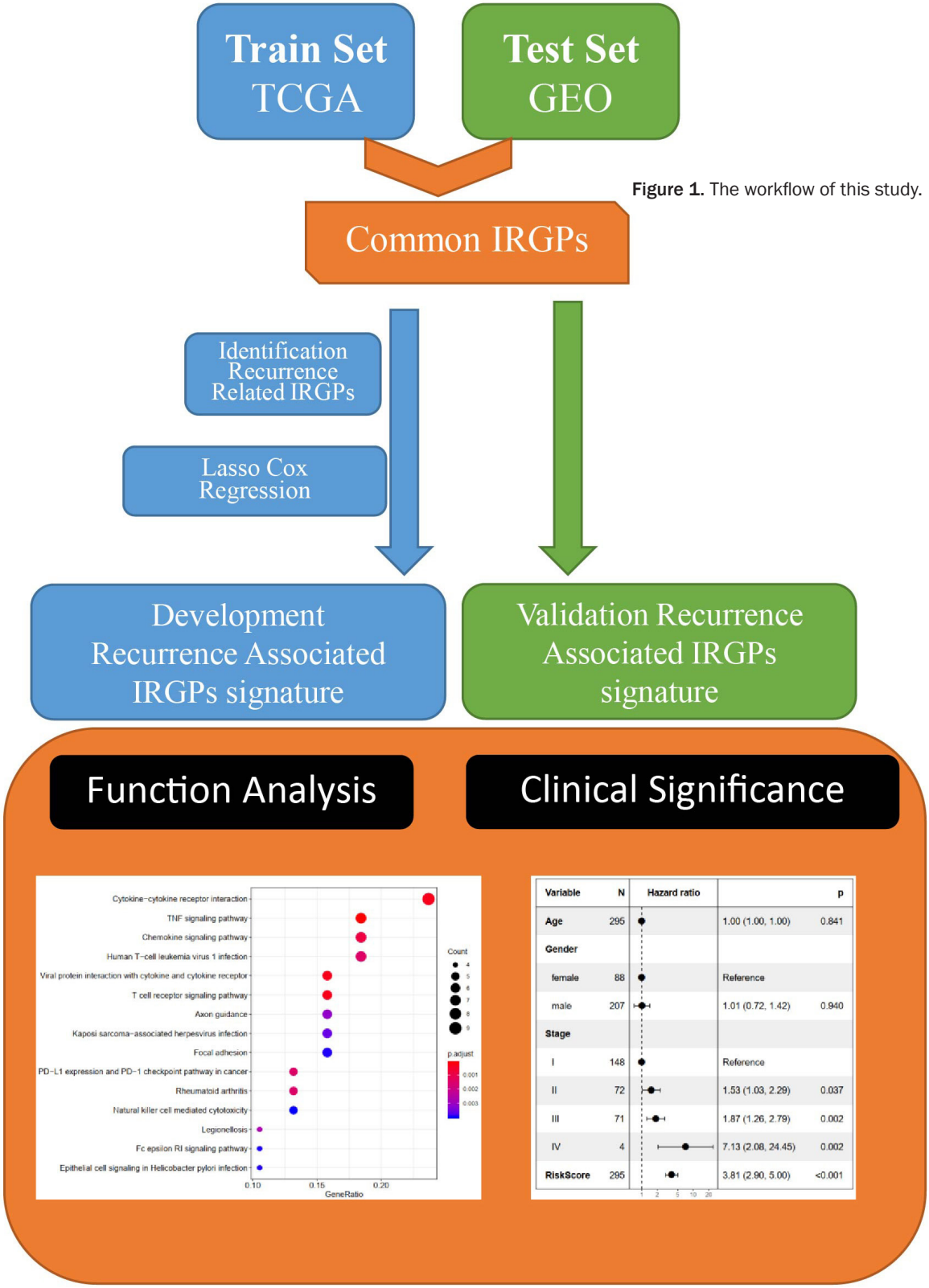

Figure 1. The workflow of this study.

rather than the recurrence. Thus, we attempted to discover an immune related gene pair (IRGPs) signature to predict the recurrence of HCC and validate in other datasets.

Table 1. The IRGPs and the coefficient of signature

Gene	Coef
ADM|CD3D	0.228989
ADM|CSF1R	0.141239
ADM|FYN	0.020781
ADM|LPA	0.163601
BPHL|ROBO1	-0.1252
C3AR1|SEMA5A	-0.16781
CACYBP|HLA.DMA	0.141385
CACYBP|TMPRSS6	0.02436
CCL16|SPP1	-0.01041
CD3D|DCK	-0.03016
CD4|RFX5	-0.15539
CKLF|NR5A2	0.432557
CKLF|TAPBPL	0.158347
CSRP1|NR1I3	0.117471
CXCL1|GCGR	0.03163
CXCL12|PLSCR1	-0.03197
CXCL2|NDRG1	-0.03
DCK|EDNRB	0.112755
DCK|HLA.DQA1	0.097435
DCK|IRF1	0.228557
DCK|KDR	0.015743
DCK|LYN	0.028205
DDX17|MASP2	0.057228
ENG|IL17RC	-0.03625
GCGR|SEMA5A	-0.04596
HLA.DPA1|HSPA1B	-0.22601
IL15RA|ITGB2	0.112393
ITGAV|NR1I2	0.086925
ITGAV|PIK3R1	0.162313
JAG1|NR4A1	0.20422
JUN|MASP2	0.018153
KDR|LTBP1	-0.0439
MASP2|PSMB8	-0.17593
NRAS|TNFRSF1B	0.055925
PRF1|SECTM1	-0.25685
ROBO1|TAPBPL	0.162024

Materials and methods

Data collection

The publicly available dataset was utilized to perform the comprehensive analysis. The workflow of the present study is shown in **Figure 1**. Gene expression profiles were collected from XENA (https://xenabrowser.net/) and clinical data were downloaded from cBioPortal (http://www.cbioportal.org/), which was utilized as a training cohort. GSE14520 database was downloaded from Gene Expression Omnibus and employed as a validation cohort [7]. Immune-related genes were gathered from the ImmPort database (https://immport.niaid.nih.gov) [8].

Identification of recurrence-related IRGPs

A total of 1811 unique genes, which were classified into 17 categories, including antigen processing and presentation antimicrobials, and BCR signaling pathway, were involved in this study. To improve the accuracy of the model, we filter out genes with a median absolute deviation below 0.5. The IRGPs were constructed as previously descripted [9]. Briefly, if the expression level of the first gene was higher than that of the second one in a specific IRGP, the score of this IRGP was defined as 1; otherwise, the score was defined as 0. The percent of a validated IRGP with a score of either 0 or 1 should account for 20-80% of all the samples.

Construction of IRGPs signature for predicting recurrence outcomes of HCC

Univariate Cox proportional hazards regression model was employed to identify recurrence-related IRGPs. The least absolute shrinkage and selection operator (Lasso) regression model was used to develop IRGPs signature for predicting recurrence rates of HCC using R package glmne (https://glmnet.stanford.edu/) in the train set. The most stable model was selected to construct the recurrence-related prognostic signature. All patients were classified into either a high or a low recurrence risk group according to a recurrence risk cutoff score. The time-dependent receiver operating characteristic curve was constructed using R package survival ROC [10]. The optimal cutoff value of the risk score was determined by the Youden index. This model was further validated in the GEO dataset. The clinical significance was examined by the multivariate Cox proportional hazards regression model. An R package of forest model was used to draw a forest plot.

Gene set enrichment analysis

Gene set enrichment of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were analyzed to explore the underlying

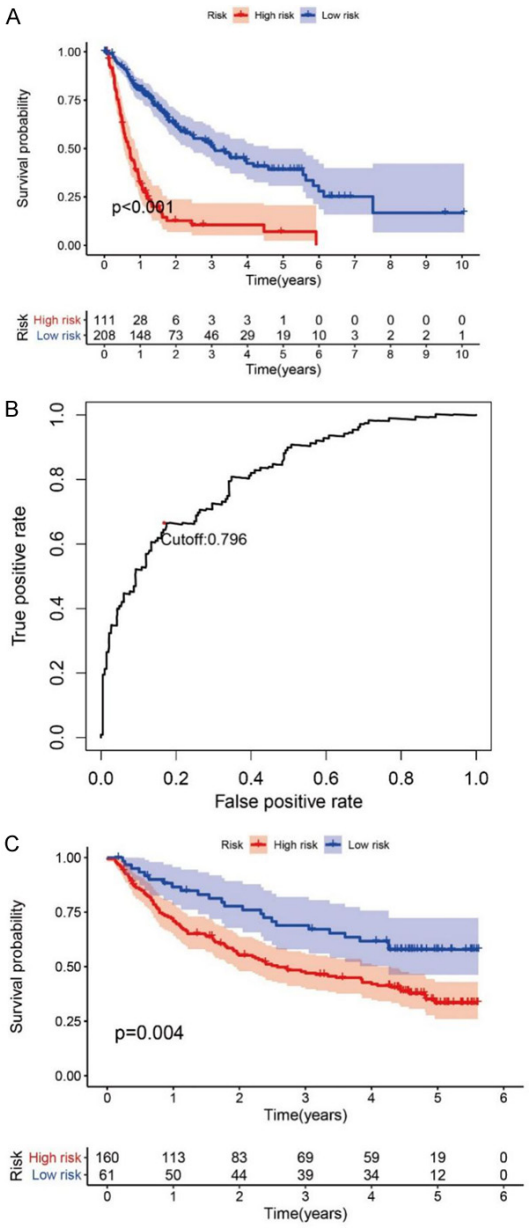

Figure 2. The IRGPs signature in training sets and test sets. A. The IRGPs signature was identified and patients with high risk score had poor survival. B. The ROC of the IRGPs signature. C. IRGPs signature was validated in independent dataset.

biological mechanisms of those immune-related genes which facilitate recurrence of HCC using R package clusterProfilter [11]. A false discovery rate below 0.2 and $P \leq 0.05$ were considered statistically significant.

Statistical analysis

Correlating IRGDs signature with immune infiltrates: We explored the relationship between IRGPs signature and immune infiltrates. TIMER (https://cistrome.shinyapps.io/timer/) [12], a web tool to analyze the immune infiltrate characteristics across TCGA samples, was used to explore the correlation of NDC1 expression and immune infiltrates, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells via gene modules [13].

Correlating the IRGDs signature with immune checkpoint: To interrogate the differences in immune checkpoints between the high and low-risk groups, we analyzed PD1 and PDL1 expression within the IRGPs signature.

Developing nomogram with the IRGDs signature

The pathological parameters and other clinical information were integrated into the multivariant COX regression analysis in the TCGA dataset and sequentially constructed a nomogram to predict the probability of recurrence of HCC patients. The calibration curve was drawn to evaluate the performance of the nomogram.

All statistical analyses were performed using R software (version 3.6.0, https://www.r-project.org/). The log-rank test was applied to evaluate the relationship between IRGPs signature and disease-free survival. The survival curves were plotted using the R package "survminer" [14].

Results

Construction of IRGPs signature for predicting recurrence of HCC

Figure 1 shows the overall workflow of this study. A total of 540 patients were included in this study, with a training cohort including 319 patients and a validation cohort including 221 patients. We observed that 258 IRGs were shared in the two datasets, resulting in a total of 6363 IRGPs that were identified after filtering with above mentioned approach. Eventually, we identified 119 disease-free-survival related IRGPs ($P < 0.05$) using univariate Cox regression analysis. The disease-free-survival related IRGPs were further employed to construct the recurrence-predicting model through lasso regression analysis and 36-IRGPs based model was developed (**Table 1**). The area of ROC of 1-year disease-free survival was 0.81 (**Figure 2A**). The best cutoff was 0.80 which was determined by the Youden index. The best cutoff sig-

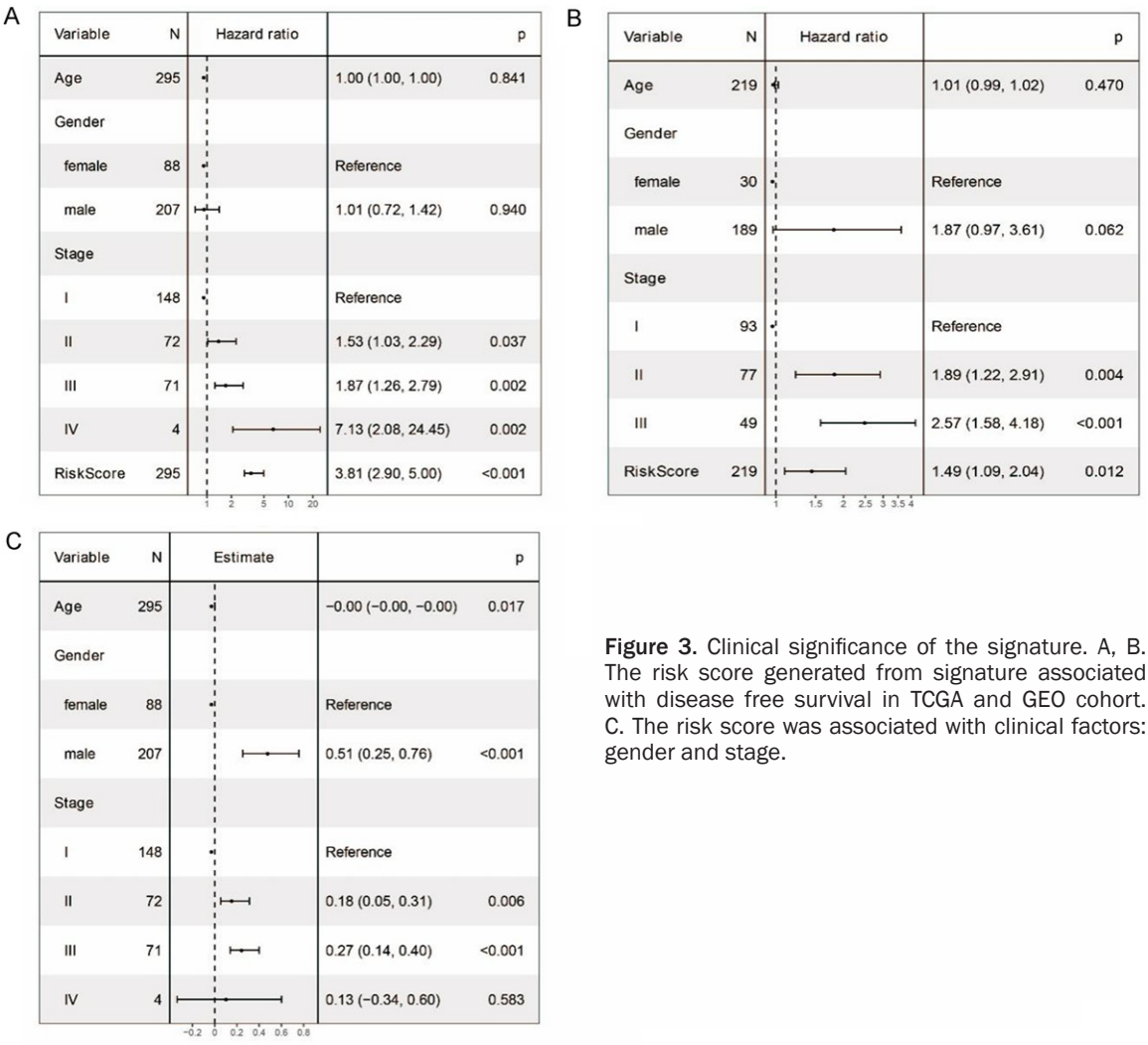

Figure 3. Clinical significance of the signature. A, B. The risk score generated from signature associated with disease free survival in TCGA and GEO cohort. C. The risk score was associated with clinical factors: gender and stage.

nificantly stratified patients into the high or low recurrence risk group (**Figure 2B**) in the training cohort. Patients in the high-risk group appeared to have poor disease-free survival. Furthermore, the patients in the validation cohort were divided into two groups based on the optimal cutoff point resulting from the training cohort (**Figure 2C**).

Clinical significance of the IRGPs signature

Each patient could get a risk score based on the model. We treat the risk model as a clinical factor to explore whether this factor could be used as an independent risk factor for predicting disease-free survival by using a multivariant cox regression model. **Figure 3** shows that the risk score of recurrence signature could work as an independent risk factor.

Next, we explored whether the recurrence signature was associated with other clinical factors. A logistic regression analysis was conducted, and results showed that risk score was associated with gender and stage (**Figure 3C**).

Gene function enrichment

The recurrence signature consisted of 54 IRGs which are enriched in many signaling pathways. The top 15 GO terms of molecular function, biological process, and cellular component are shown in **Figure 4**. Moreover, the top 15 KEGG pathways are listed in **Figure 4D**.

We compared the difference in the immune cells between the high and low-risk groups. As is shown in **Figure 5**, the two groups had statistical differences. Moreover, the high-risk group had a higher expression of PD1 with $P < 0.05$ (**Figure 6**).

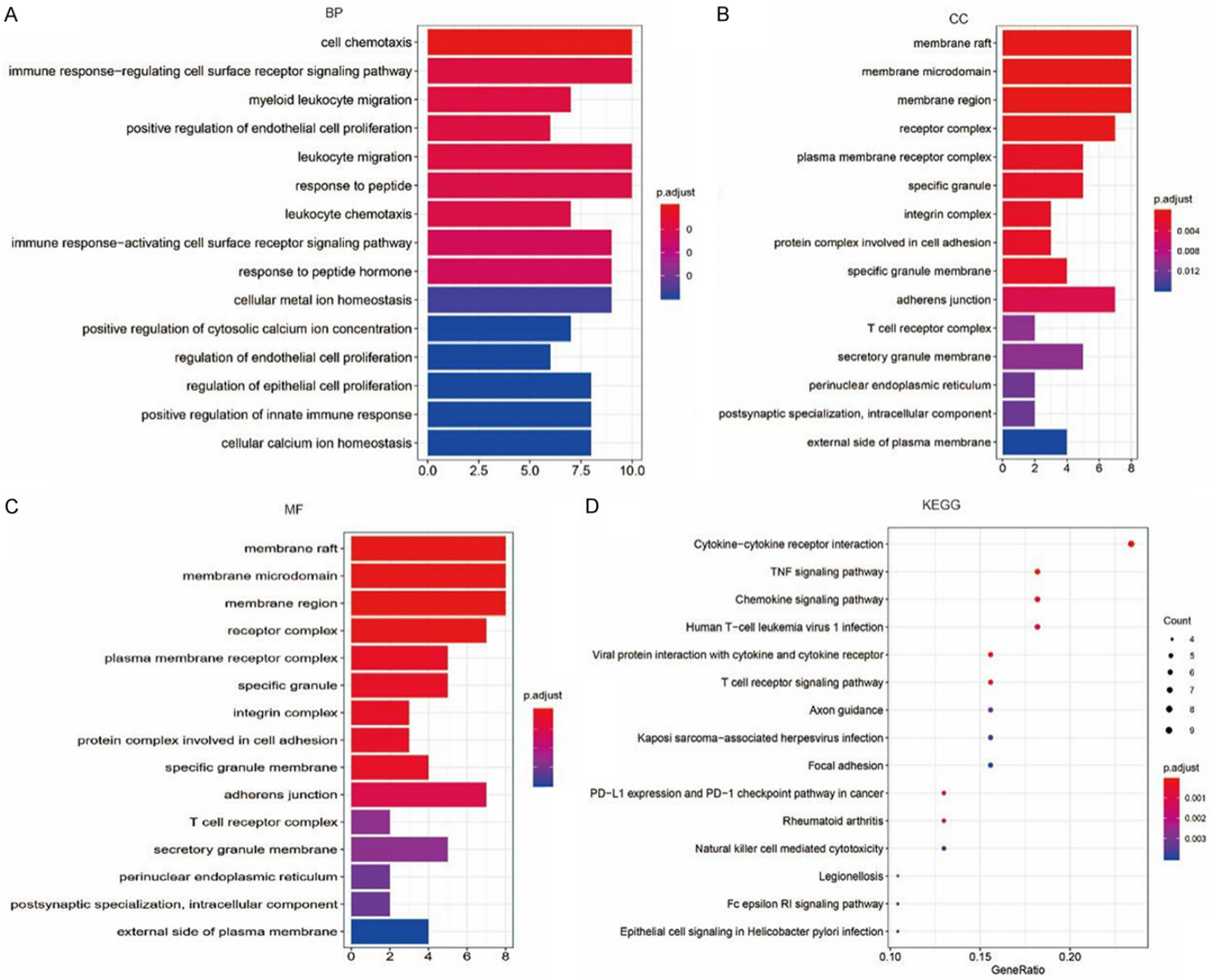
A
BP
cell chemotaxis
immune response−regulating cell surface receptor signaling pathway
myeloid leukocyte migration
positive regulation of endothelial cell proliferation
leukocyte migration
response to peptide
leukocyte chemotaxis
immune response−activating cell surface receptor signaling pathway
response to peptide hormone
cellular metal ion homeostasis
positive regulation of cytosolic calcium ion concentration
regulation of endothelial cell proliferation
regulation of epithelial cell proliferation
positive regulation of innate immune response
cellular calcium ion homeostasis
0.0 2.5 5.0 7.5 10.0
p.adjust
B
CC
membrane raft
membrane microdomain
membrane region
receptor complex
plasma membrane receptor complex
specific granule
integrin complex
protein complex involved in cell adhesion
specific granule membrane
adherens junction
T cell receptor complex
secretory granule membrane
perinuclear endoplasmic reticulum
postsynaptic specialization, intracellular component
external side of plasma membrane
0 2 4 6 8
p.adjust
0.004
0.008
0.012
C
MF
membrane raft
membrane microdomain
membrane region
receptor complex
plasma membrane receptor complex
specific granule
integrin complex
protein complex involved in cell adhesion
specific granule membrane
adherens junction
T cell receptor complex
secretory granule membrane
perinuclear endoplasmic reticulum
postsynaptic specialization, intracellular component
external side of plasma membrane
0 2 4 6 8
p.adjust
D
KEGG
Cytokine−cytokine receptor interaction
TNF signaling pathway
Chemokine signaling pathway
Human T−cell leukemia virus 1 infection
Viral protein interaction with cytokine and cytokine receptor
T cell receptor signaling pathway
Axon guidance
Kaposi sarcoma−associated herpesvirus infection
Focal adhesion
PD−L1 expression and PD−1 checkpoint pathway in cancer
Rheumatoid arthritis
Natural killer cell mediated cytotoxicity
Legionellosis
Fc epsilon RI signaling pathway
Epithelial cell signaling in Helicobacter pylori infection
0.10 0.15 0.20
GeneRatio
Count
4
5
6
7
8
9
p.adjust
0.001
0.002
0.003

Figure 4. Gene enrichment in recurrence signature. BP: biological process, CC: cellular component; MF: molecular function.

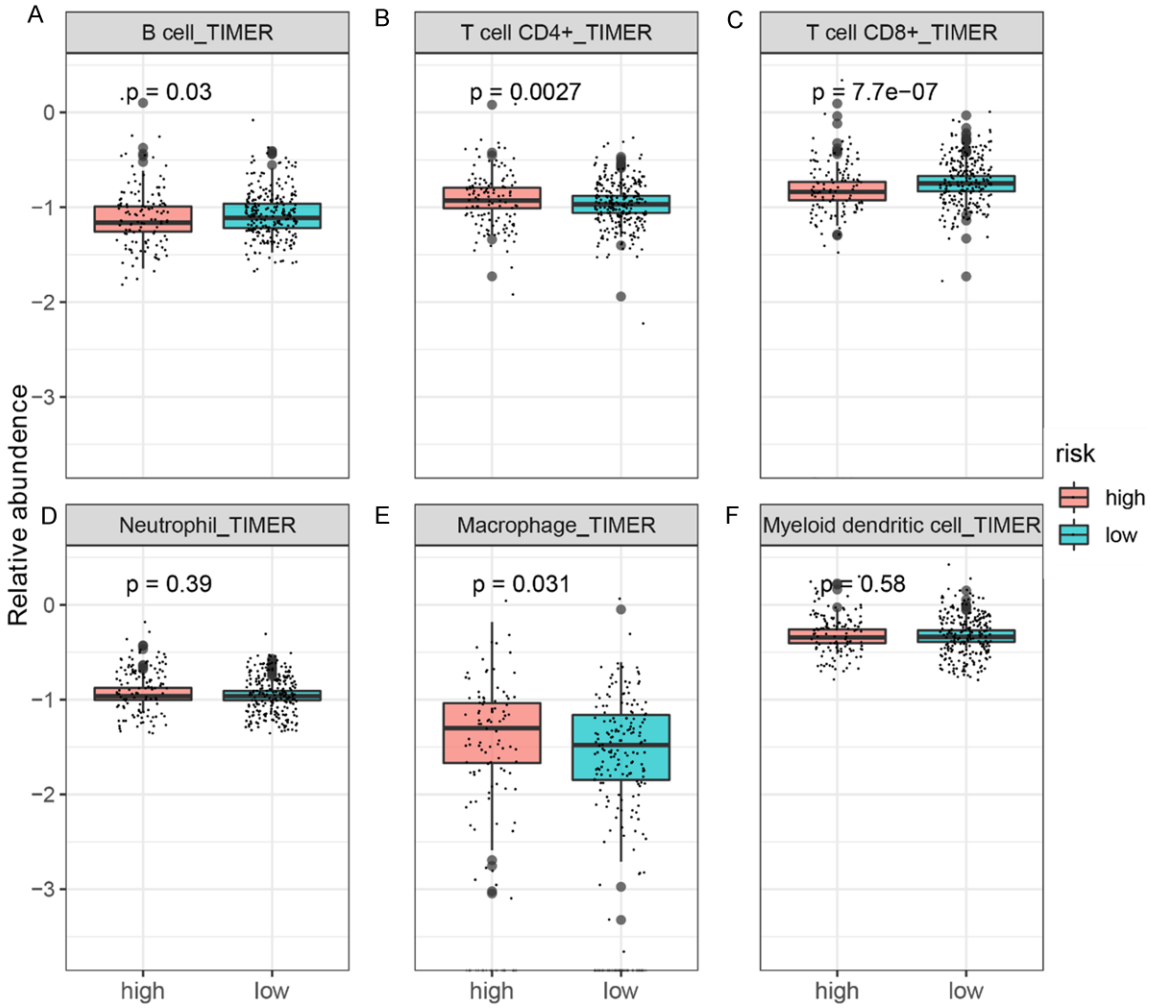

Figure 5. The different immune infiltrate in the two groups.

IRGPs related nomogram

To enable individualized prediction of the prognosis of HCC, we integrated all clinicopathological parameters with risk scores to construct a novel nomogram (**Figure 7A**). The calibration curves showed reliable consistency between the predicted value and the actually observed value of the nomogram (**Figure 7B**).

Discussion

HCC is the most common primary malignancy of the liver, accounting for nearly 90% of all the malignant cases [15]. It has been well documented that the development of HCC is a multistep process, and it is a multigene alteration-induced cancer with a high level of heterogeneity. Several etiologic factors have been identified, including hepatitis B, hepatitis C, alcohol abuse, steatohepatitis, and obesity [16]. Recent molecular studies have shown that some special gene alterations play a crucial role in the development of HCC [17, 18].

It is a promising way to treat liver cancer with immunotherapy. Several clinical trials of immunotherapy for HCC were already conducted [4]. Immune checkpoint inhibitor blockade therapy has been paid more attention, especially for Nivolumab, which is the first FDA-approved immune checkpoint inhibitor for HCC. Several

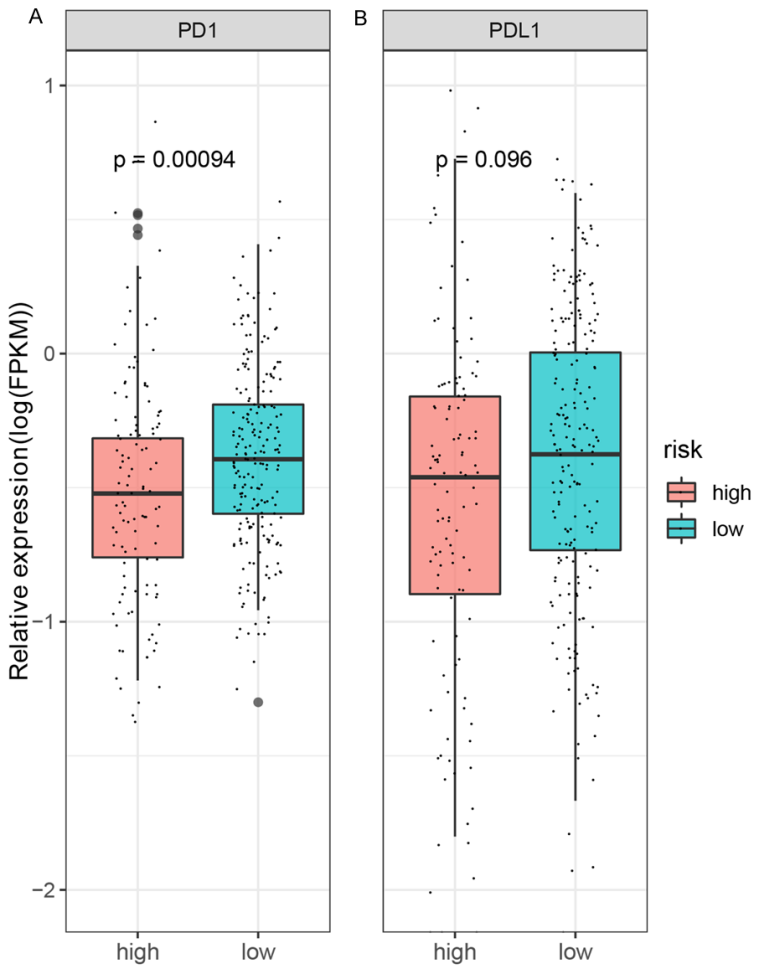

Figure 6. The immune check point genes of PD1 and PDL1 in the two groups.

studies have shown the encouraging results of immune checkpoint inhibitors on HCC. Integrating immune therapy with other therapy has been an important exploration. However, unmanageable relapse of HCC is still a thorny obstacle in the way to completely cure liver cancer. It is urgent to develop a recurrence-related model to predict the relapse of HCC. In our study, we used IRGs to construct IRGPs which consisted of the risk assessment model. IRGPs were produced by comparing the immune-related genes with each other only based on the level of expression within an individual patient. This method makes data more comparative and overcomes the batch effects resulting from different types of data. Several studies have reported the efficiency and practicability of this method [9, 19, 20]. We used this method to construct a recurrence-related model by lasso regression. A 36-IRGPs based recurrence-related signature was developed and validated in an independent dataset. The area of AUC of the model was up to 0.796, which is better than the recurrence signature [20]. The signature could separate patients into low and high recurrence risk groups in both training set and testing set. We treat the risk score as a clinical variant and further validated it by the multivariate Cox proportional hazard regression, showing that the signature was an independent risk factor for predicting recurrence. Logistic regression analysis demonstrated that the risk model was associated with gender and tumor stage. Combined with multivariate cox analysis, the model has important clinical significance in estimating HCC recurrence rates.

The 54 IRGs, which consisted of the model, were enriched in many important pathways. For instance, TNF signaling pathways play crucial roles in various cancers, including HCC [21]. PD-L1/PD-1 checkpoint pathway in cancer is also involved in HCC. In the context of HCC, approximately 82% of the patients express PDL1 and 19% are responsive to the anti-PDL1 reagent, in which 5% of patients could get complete remission from the anti-PDL1 treatment [22]. Those exciting results indicate that immune therapy is on the way to become a promising strategy for tackling the disease. Our model is enriched with the immune-related pathways, reflecting the significance of the IRGPs signature. However, there are some shortcomings in this study. Although the model was validated in an independent dataset, we did not test the model in the clinical and experimental samples.

Conclusion

We developed an IRGPs signature to predict the recurrence of HCC and validated it in an inde-

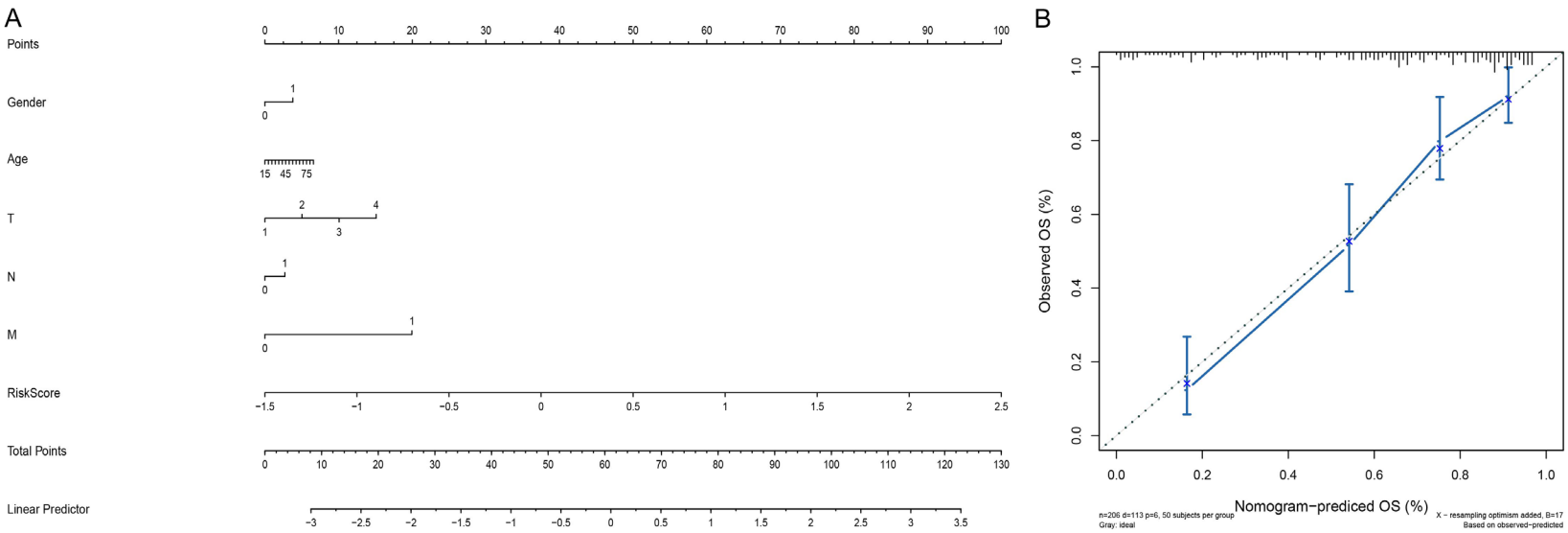

Figure 7. The nomogram of IRGPs for HCC (A). The calibration curves showed good consistency between the predicted value and the actual observed value of the nomogram (B).

pendent cohort as well. We found that the IRGPs signature was associated with gender and tumor stage. The genes which consisted of IRGPs signature enriches in many important pathways.

Acknowledgements

This work was supported by Fund of The 111 Project (D17011); Guangxi Key Research and Development Plan (2018AD03001); Guangxi BaGui Scholars; National Natural Science Foundation of China (81771674 and 9194-9122).

Disclosure of conflict of interest

None.

Address correspondence to: Guandou Yuan, Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, No. 6 Shuangyong Road, Nanning 530021, Guangxi, China. Tel: +86-0771-5359339; E-mail: dr_yuangd@gxmu.edu.cn

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu; European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. J Hepatol 2018; 69: 182-236.

[3] Oweira H, Petrausch U, Helbling D, Schmidt J, Mehrabi A, Schöb O, Giryes A and Abdel-Rahman O. Prognostic value of site-specific extrahepatic disease in hepatocellular carcinoma: a SEER database analysis. Expert Rev Gastroenterol Hepatol 2017; 11: 695-701.

[4] Greten TF, Lai CW, Li G and Staveley-O'Carroll KF. Targeted and immune-based therapies for hepatocellular carcinoma. Gastroenterology 2019; 156: 510-524.

[5] Tabrizian P, Jibara G, Shrager B, Schwartz M and Roayaie S. Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. Ann Surg 2015; 261: 947-955.

[6] Teng F, Meng X, Kong L and Yu J. Progress and challenges of predictive biomarkers of anti PD-1/PD-L1 immunotherapy: a systematic review. Cancer Lett 2018; 414: 166-173.

[7] Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX and Wang XW. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010; 70: 10202-10212.

[8] Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS, Thomson E, Wiser J and Butte AJ. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data 2018; 5: 180015.

[9] Li B, Cui Y, Diehn M and Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. JAMA Oncol 2017; 3: 1529-1537.

[10] Heagerty PJ, Lumley T and Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000; 56: 337-344.

[11] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 2012; 16: 284-287.

[12] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B and Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 2017; 77: e108-e110.

[13] Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS and Liu XS. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 2016; 17: 174.

[14] Kassambara A, Kosinski M, Biecek P and Fablian S. Survminer: drawing survival curves using 'ggplot2',URL:https://cran.r-project.org/web/packages/survminer/index.html.

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

[16] Fujiwara N, Friedman SL, Goossens N and Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. J Hepatol 2018; 68: 526-549.

[17] Gao Q, Zhu H, Dong L, Shi W, Chen R, Song Z, Huang C, Li J, Dong X, Zhou Y, Liu Q, Ma L, Wang X, Zhou J, Liu Y, Boja E, Robles AI, Ma W, Wang P, Li Y, Ding L, Wen B, Zhang B, Rodriguez H, Gao D, Zhou H and Fan J. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. Cell 2019; 179: 561-577, e522.

[18] Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu; Cancer Genome Atlas Research Network. Comprehen-

sive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017; 169: 1327-1341, e1323.
[19] Sun XY, Yu SZ, Zhang HP, Li J, Guo WZ and Zhang SJ. A signature of 33 immune-related gene pairs predicts clinical outcome in hepatocellular carcinoma. Cancer Med 2020; 9: 2868-2878.
[20] Liu H, Yan Y, Chen R, Zhu M, Lin J, He C, Shi B, Wen K, Mao K and Xiao Z. Integrated nomogram based on five stage-related genes and TNM stage to predict 1-year recurrence in hepatocellular carcinoma. Cancer Cell International 2020; 20: 140.
[21] Park EJ, Lee JH, Yu GY, He G, Ali SR, Holzer RG, Osterreicher CH, Takahashi H and Karin M. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. Cell 2010; 140: 197-208.
[22] Khemlina G, Ikeda S and Kurzrock R. The biology of Hepatocellular carcinoma: implications for genomic and immune therapies. Mol Cancer 2017; 16: 149.