



Research article

A novel immune checkpoint-related gene signature for hepatocellular carcinoma to predict clinical outcomes and therapeutic response

Siyuan Tian^{1,†}, Yinan Hu^{1,†}, Chunmei Yang^{1,†}, Jiahao Yu¹, Jingyi Liu², Guoyun Xuan¹, Yansheng Liu¹, Keshuai Sun³, Miao Zhang¹, Shuoyi Ma¹, Yulong Shang^{1,*}, Xia Zhou^{1,*} and Ying Han^{1,*}

¹ State Key Laboratory of Cancer Biology, Xijing Hospital of Digestive Diseases, The Fourth Military Medical University, Xi'an 710032, China

² Department of Radiation Oncology, Xijing Hospital, The Fourth Military Medical University, Xi'an 710032, China

³ Department of Gastroenterology, The Air Force Hospital from Eastern Theater of PLA, Nanjing 210002, China

* **Correspondence:** Email: hanying1@fmmu.edu.cn, zhouxialyg@163.com, shangyul870222@163.com; Tel: +862984771506; Fax: +862982539041.

† These authors contributed equally to this work.

Abstract: Immune checkpoint genes (ICGs) have recently been proven to perform instrumental functions in the maintenance of immune homeostasis and represent a promising therapeutic strategy; however, their expression patterns and prognostic values are not fully elucidated in hepatocellular carcinoma (HCC). In this investigation, we focused on establishing and validating a prognostic gene signature to facilitate decision-making in clinical practice. Clinical information, as well as transcriptome data, was obtained from the Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) database. Univariate Cox regression and least absolute shrinkage and selection operator (LASSO) Cox method were employed to build a multi-gene signature in the TCGA database, while the ICGC database was used for validation. Subsequently, utilizing the six-gene signature, we were able to categorize patients into high- and low-risk groups. In two cohorts, survival analysis findings revealed a dismal outlook for the high-risk group. The receiver operating characteristic curves were utilized to estimate the gene signature's prediction ability. Moreover, correlation analysis showed high-risk group was linked to advanced pathological stage,

infiltration of immune cells and therapeutic response. In summary, this unique gene profile might serve not only as a useful prognostic indicator but also as a marker of therapy responsiveness in HCC.

Keywords: HCC; immune checkpoint; gene signature; survival analysis; therapeutic response

1. Introduction

Hepatocellular carcinoma (HCC) is amongst the most prevalent and deadly malignancies globally and has been ranked 6th in occurrence and 3rd in fatality rate [1]. Even though there has been some advancement in prevention, surveillance, early-HCC diagnosis and therapy over the past few decades [2–5], HCC is still a malignant tumor with a dismal prognosis as evidenced by a limited five-year survival probability of as low as 18% [6]. Most HCC patients receive their diagnosis at intermediate or advanced stages and lose the opportunity to benefit from surgery intervention [7]. Moreover, within the first five years following surgery, around 70 percent of patients experience recurrence or distant metastasis [8]. The traditional TNM stage acts as a common prognostic marker and determines treatment choices for HCC patients. However, HCC is a highly heterogeneous disease both at the molecular and clinical levels [9]. And genetic heterogeneity is closely related to prognosis and therapeutic reactivity [10]. In this sense, the traditional TNM staging system remains to be further improved. Therefore, it is important to construct an effective prognostic prediction model to identify patients with poor prognosis to optimize individualized and rational adjuvant treatments.

Over the past few years, the advent of immune checkpoint blocking treatment has heralded in a new epoch of cancer treatment. So far, three immune checkpoint targeted agents have been approved by the Food and Drug Administration for treating HCC including nivolumab (anti-PD-1), pembrolizumab (anti-PD-1) and ipilimumab (anti-CTLA-4) [11]. Recently, the IMbrave150 trial, evaluating the efficacy of atezolizumab in combination with bevacizumab versus sorafenib in patients suffering from the advanced hepatocellular carcinoma, has demonstrated significant benefits of this combination on overall and progression-free survival [12]. Some studies showed that PD-L1 expression in HCC could regulate several pathways in hepatocellular carcinogenesis, such as inducing inactivation of CD8⁺ T cells adhered to liver sinusoid cells to promote immune tolerance, inducing the exhaustion of follicular helper T cells, impairing cytokine expression and B cell function [13–15]. Itoh et al. [16] reported that HCC patients having PD-L1 positive expression had a worse prognosis as opposed to the ones having negative PD-L1 expression. Still other studies have pointed out that immune checkpoint-associated receptors and ligands, such as TIM-3 [17], TIGIT [15], LAG-3 [18] and GAL-9 [19] are involved in HCC progress and may serve as prognostic indicators. However, there are few research have systematically studied the expression profile of genes associated with immune checkpoints and the prognostic significance of these genes in HCC patients. Further, it is noteworthy that the predictive value of a single immune checkpoint gene in prognosis or treatment response seems to be limited. In a meta-analysis conducted by Shen et al. [20] to assess the effectiveness of PD-1 or PD-L1 inhibitors in patients with PD-L1 positive and PD-L1 negative cancers, it was discovered that the expression of PD-L1 was not a sufficient predictor of responsiveness to immunotherapy.

In the present research, taking the advantage of genomic databases, we focused on identifying prognosis-related immune checkpoint genes and constructing an ICG-based gene signature, followed by

the validation of its robustness through an external patient cohort. Moreover, immune status and drug sensitivity were compared between the high- and low-risk cohorts. We expected this novel gene signature might help to refine risk assessment and promote the development of effective individualized therapy in clinical settings.

2. Materials and methods

2.1. Data sources and processing

The original RNA-sequence data along with matching clinical data were downloaded from the Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>) database, including 374 HCC tissues and 50 adjacent non-cancerous tissues (There were 370 HCC patients who had comprehensive follow-up information), which were utilized to create a prognostic model. Meanwhile, level 3 mRNA expression data together with clinical features of additional 231 HCC samples were downloaded from the International Cancer Genome Consortium (ICGC; <https://icgc.org/icgc>) database and were employed for validation. All of the data that was utilized in this investigation was accessible to the public. The gene symbol was first transformed from the probe IDs by referring to the annotation files. From the related literature, an aggregate of 79 ICGs was identified, with the majority of the ICGs being receptors, ligands, or other key components in immune checkpoint pathways [21–24] (Table S1). A subsequent analysis was conducted using the expression matrix of the ICGs that had been identified.

2.2. DE-ICG identification and enrichment analysis

The “limma” R package was employed to search for differential expression of immune checkpoint-related genes (DE-ICGs) across tumor and surrounding tissues in TCGA database. The $|\log_2 \text{fold change}| > 1$ and $P\text{-value} < 0.05$ served as the screening threshold. The “ggplot2” and “pheatmap” packages were employed to create heatmap and volcano plots. Then, the DE-ICGs were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses by means of the R package “clusterProfiler”. The GO terms included 3 biological categories (MF: molecular functions, CC: cell components and BP: biological process). Utilizing the R package “ggplot2”, we built a representation of the 10 leading highly enriched GO terms and KEGG pathways.

2.3. Creation and verification of a prognostic gene signature associated with ICG

Firstly, univariate Cox analysis was undertaken to filter out ICGs substantially linked to overall survival (OS). To further narrow the candidate genes range, the LASSO-Cox regression analysis was applied to determine the independent OS predictors using the “glmnet” R package. The genes with independent prognostic values were incorporated into the gene signature. These hub genes were also validated in the Kaplan-Meier (K-M) Plotter database (www.kmplot.com). Next, these genes were employed to construct a prognostic gene signature. Each patient’s risk score was computed utilizing the equation below: Risk score = $\sum_{i=1}^n \text{Coef}_i * X_i$ (where Coef_i was coefficient calculated by multivariate Cox proportional hazard regression, while X_i represented the expression level for each

ICG). Following that, subjects were categorized into low- and high-risk cohorts depending on their median risk rating. Using the “survival” R package, a survival analysis was carried out to examine the variation in OS between the two cohorts. After generating the ROC curves, we computed the area under the curve (AUC) to measure the prediction ability. For external validation, the same methods were adopted to estimate the gene’s signature precision in the ICGC cohort. In addition, considering that disease-specific survival (DSS) and disease-free survival (DFS) could more specifically indicate the clinical benefit to patients, we additionally explored the prognostic significance of the established gene signature using DFS and DSS as clinical outcomes in the TCGA-LIHC cohort.

2.4. Correlation analysis between gene signature, ICGs and clinical traits

To thoroughly investigate the gene signature’s prognostic significance, we conducted subgroup analyses to clarify the clinicopathological features, gene expression patterns and OS of patients in distinct risk cohorts. In addition, survival analyses were also conducted according to patients’ clinical information like age, gender, tumor grade and clinical stage.

2.5. Establishment and assessment of a predictive nomogram

The gene signature, in conjunction with the other clinical parameters listed above, was subjected to univariate and multivariate Cox regression analysis in TCGA cohort in an attempt to discover independent prognostic markers. The “rms” R platform was utilized to create the nomogram from all independent predictors with a P -value < 0.05 . In order to visualize the precision of the prediction nomogram, we generated calibration curves plus time-dependent ROC curves. Finally, when it comes to clinical applications, decision curve analysis (DCA) was utilized to indicate the nomogram’s potential value.

2.6. Immune status and therapeutic response analysis

To additionally examine the link between risk scores and immune status of patients, a single-sample genomic enrichment analysis (ssGSEA) was carried out to contrast the infiltration scores of 16 distinct immune cell subpopulations and 13 immune functional pathways among both risk cohorts utilizing “GSVA” package in the R software. To forecast the possible responsiveness to an immune checkpoint inhibitor, we applied an additional tool called Immunophenoscore (IPS). IPS is an extensively utilized machine learning-based algorithm, reflecting the tumor immunogenicity of patients [25]. It was computed on a scale ranging from 0 to 10, with higher scores being related to greater immunogenicity and improved immunotherapeutic responsiveness. The Cancer Immunome Atlas (TCIA, <https://tcia.at/home>) database was utilized to retrieve the IPS of patients, which were subsequently subjected to a comparison between the two risk cohorts. In addition, we also compared the sensitivity of five common chemotherapy drugs (sorafenib, doxorubicin, cisplatin, vinblastine and mitomycin. C) between the two groups, which were recommended by AJCC guidelines for the adjuvant therapy of liver cancer. The half-maximal inhibitory concentration (IC50) values were computed for drug resistance measurements. Results of high- and low-risk cohorts were visualized utilizing box plots created by using the R package “pRRophetic” and “ggplot2”.

3. Results

3.1. Patient demographics and clinical characteristics

A total of 374 HCC patients from the TCGA-LIHC cohort and 231 HCC patients from the ICGC (LIRI-JP) cohort were eventually included in the present research. Among the patients in the TCGA cohort, four patients were excluded from survival analysis due to incomplete follow-up information. Table 1 summarizes the comprehensive clinical parameters of the patients in more depth. The study's workflow is outlined in Figure 1.

Table 1. The clinical characteristics of the patients in the TCGA and ICGC cohorts.

Characteristics	TCGA cohort (N = 370)	ICGC cohort (N = 231)
Age at diagnosis (years)		
>65	138 (37.3%)	142 (61.5%)
≤65	232 (62.7%)	89 (38.5%)
Sex		
Male	249 (67.3%)	170 (72.6%)
Female	121 (32.7%)	61 (26.4%)
Stage		
I	171 (46.2%)	36 (15.6%)
II	85 (23.0%)	105 (45.5%)
III	85 (23.0%)	71 (30.7%)
IV	5 (1.3%)	19 (8.2%)
Unknown	24 (6.5%)	0 (0.0%)
T classification		
T1	181 (48.9%)	NA
T2	93 (25.2%)	NA
T3	80 (21.6%)	NA
T4	13 (3.5%)	NA
Unknown	3 (0.8%)	NA
N classification		
N0	252 (68.1%)	NA
N1	4 (1.1%)	NA
Unknown	114 (30.8%)	NA
M classification		
M0	266 (71.9%)	NA
M1	4 (1.1%)	NA
Unknown	100 (27.0%)	NA
Tumor grade		
Grade 1	55 (14.9%)	NA
Grade 2	177 (47.8%)	NA
Grade 3	121 (32.7%)	NA
Grade 4	12 (3.2%)	NA
Unknown	5 (1.4%)	NA

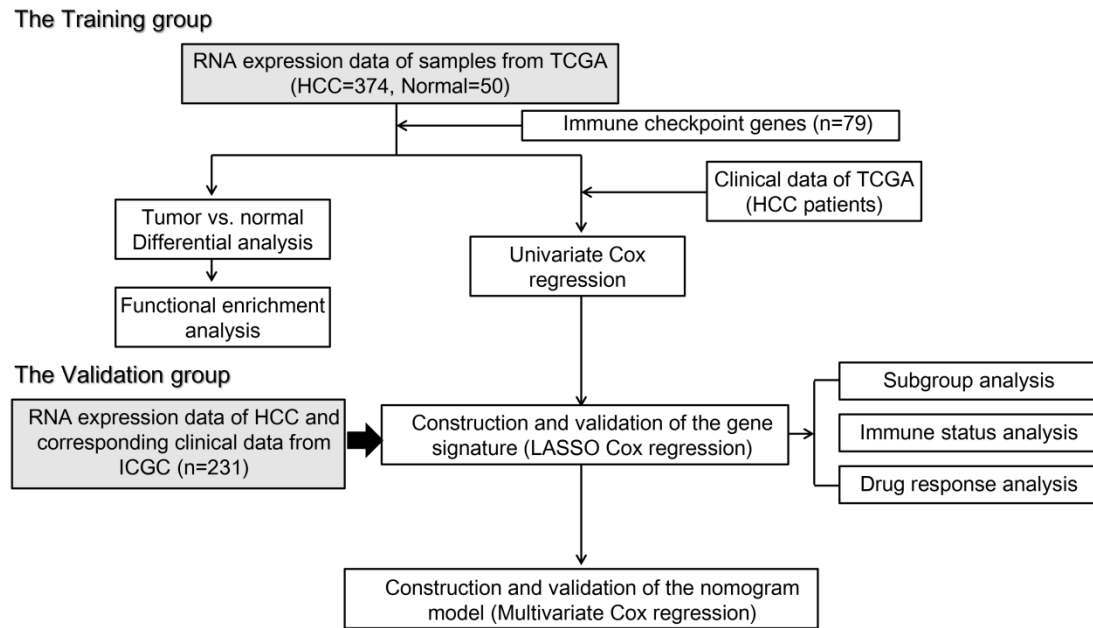


Figure 1. The schematic illustration of the study's overall design.

3.2. Differentially expressed genes and functional enrichment analysis

Among the 79 ICGs, 68 genes were identified from the TCGA dataset. From these genes, we extracted the expression profiles and collated them into the expression matrix. Based on the screening criteria, a total of 21 genes showed significant alteration in the levels of expression in HCC as opposed to the adjoining non-cancerous tissues, including 18 up-regulated and 3 down-regulated. Heatmap (Figure 2A) and volcano plot (Figure 2B) showed the distribution of DE-ICGs. Furthermore, in order to better understand the biological role of DE-ICGs, a functional enrichment analysis was performed. As shown in Figure 2C, the bar plot showed the top 10 enriched GO terms, including: “MHC protein complex binding”, “external side of plasma membrane” and “T-cell activation”. Premised on the findings of the KEGG pathway analysis, the identified DE-ICGs were related to “Cytokine-cytokine receptor interaction” and “Antigen processing and presentation” (Figure 2D,E).

3.3. Construction of an ICG-based prognostic gene signature

We firstly conducted a univariate Cox regression analysis and discovered eight genes of significant correlation with OS, among which six genes were protective factors and two genes were risk factors (Figure 3A). Of note, not all OS-related ICGs were DE-ICGs. Figure 3B depicts the relationship between these genes and their expression. Considering these genes may interact with each other, we employed Lasso Cox regression analysis to lower the dimensionality of data in order to find the real OS-affecting factors (Figure 3C,D). Six genes were selected as hub genes including LGALS9, CD209, CD40LG, SIRPA, BTNL9 and TNFSF4. The coefficients of these hub genes were depicted in Figure 3E. Detailed information was summarized in Table 2. Besides, the prognostic values of these six hub genes were also confirmed by the Kaplan-Meier plotter database (Figure S1).

Subsequently, the following equation was established for this six-gene signature: Risk score = 0.002 * Exp_{LGALS9} + 0.089 * Exp_{CD209} - 0.381 * Exp_{CD40LG} + 0.019 * Exp_{SIRPA} - 0.07 * Exp_{BTNL9} - 0.046 * Exp_{TNFSF4}. Finally, we were able to determine each patient's risk score utilizing the equation provided above.

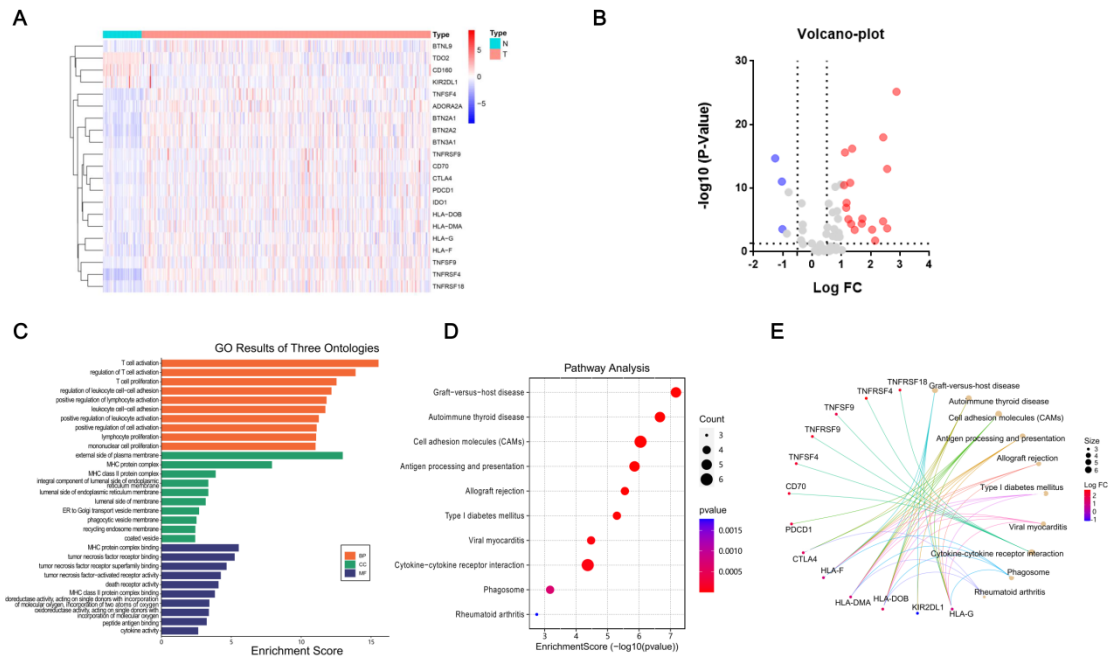


Figure 2. Results of enrichment analysis of differentially expressed immune checkpoint genes (DE-ICGs) (A) The heatmap of DE-ICGs in the TCGA cohort; N normal T tumor. (B) The volcano-plot of DE-ICGs; red denotes up-modulated genes and blue denotes down-modulated genes. (C) GO enrichment analysis of DE-ICGs; BP biological processes, CC cellular components, MF molecular functions. (D) KEGG pathway analysis of DE-ICGs. (E) The chord plot showing the top 10 significant KEGG signaling pathways.

Table 2. Hub immune checkpoint genes included in the prognostic gene signature.

Gene symbol	Description	Risk coefficient
LGALS9	Galectin 9	0.002263861
CD209	CD209 Molecule	0.089413494
CD40LG	CD40 Ligand	-0.381133418
SIRPA	Signal Regulatory Protein Alpha	0.019462250
BTNL9	Butyrophilin Like 9	-0.069846262
TNFSF4	TNF Superfamily Member 4	0.046198837

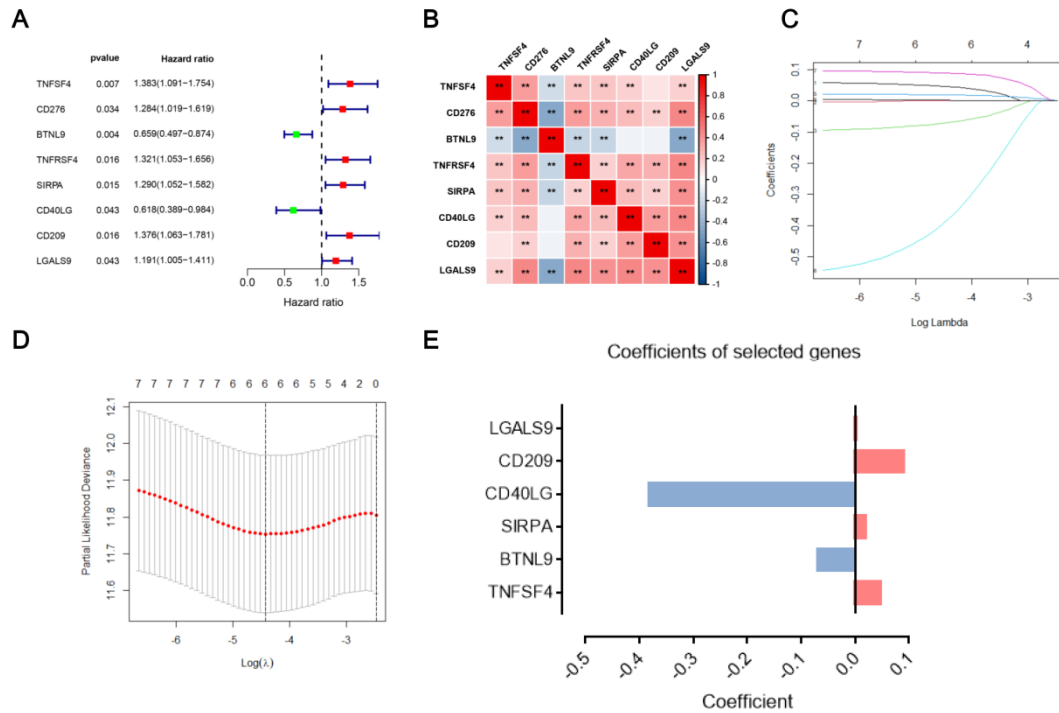


Figure 3. Identification of the candidate immune checkpoint genes (ICGs) associated with the prognosis of HCC patients in the TCGA database. (A) The univariate Cox regression analysis to identify ICGs related to overall survival (OS). (B) The correlation of candidate ICGs; correlation coefficients are represented by different colors. $*P < 0.05$, $**P < 0.01$. (C) The Least Absolute Shrinkage and Selector Operation (LASSO) coefficient profiles. (D) Selecting the best parameters (λ) in the LASSO model. (E) The coefficient of selected hub genes in the gene signature.

3.4. The performance of gene signature

In the TCGA cohort, the median risk value of all HCC patients was utilized to separate them into high- and low-risk cohorts, with a cutoff value of 0.1 to distinguish between the two. First, we charted the survival curves to probe into whether there was a substantial difference in survival between the two cohorts. As illustrated in Figure 4A, patients at high risk demonstrated a considerably reduced OS as opposed to the ones at low risk. Figure 4B illustrated the risk score distribution as well as the survival status of patients. Next, in order to assess the gene signature's predictive ability even further, ROC curves for survival over one, two and three years were created and the AUC was determined for each of these outcomes. The AUC values as depicted in Figure 4C were 0.679, 0.649 and 0.662, correspondingly. Moreover, an external dataset retrieved from the ICGC database was employed to validate this gene signature. In the validation cohort, high-risk patients exhibited a substantially unfavorable outcome as opposed to the ones at low risk (Figure 4D,E). Similar findings were achieved in the ICGC cohort, whereby, high predictive values of AUC were observed in the time-dependent ROC curves (Figure 4F).

Compared with OS, DFS and DSS are more precise indicators reflecting clinical benefits of patients. This study employed DSS and DFS as clinical endpoints in order to determine if the

prognostic model was also associated with these 2 prognostic markers. As illustrated in Figure S2, this prognostic model still could be used to distinguish patients with different survival statuses and held good stability. The results described above suggested that the novel gene signature was a robust index, which might be applied to anticipate clinical outcomes of HCC patients.

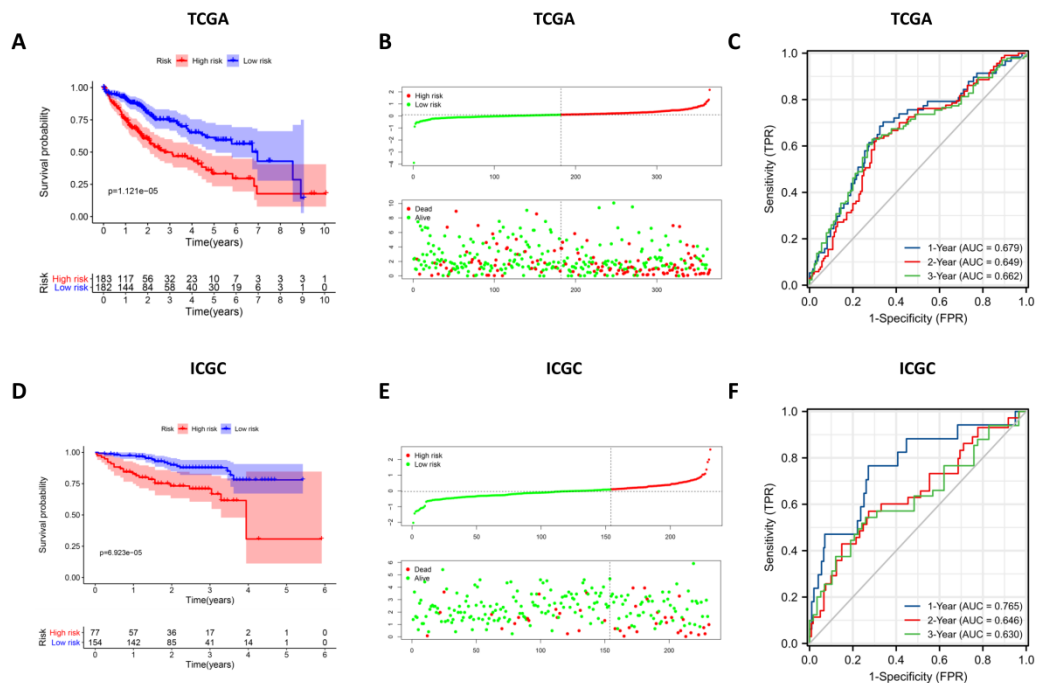


Figure 4. Development and validation of an immune checkpoint-related gene signature. Survival curves for the high- and low-risk groups in the TCGA cohort (A) and ICGC cohort (D). Patients' survival status and risk score in the TCGA cohort (B) and ICGC cohort (E). ROC curve analysis for predicting the overall survival of HCC patients according to the risk score in the TCGA cohort (C) and ICGC cohort (F).

3.5. Clinical relevance investigation and subgroup analysis

To study the link between risk cohorts and clinical parameters such as grade, TNM stage, gender and age, we plotted a heat map to display the correlation among these indexes. Figure 5A depicts the distribution of expression patterns of six genes, clinical parameters and survival status. On the basis of these data, it is evident that higher risk scores had a high likelihood of being detected in those patients with advanced TNM stages or higher pathological grade (Figure 5B,C), hinting at the risk signature's precision. Furthermore, survival analysis among distinct clinical subgroups was carried out according to the above different parameters. The results showed that this gene signature could effectively distinguish samples into low- or high-risk cohorts between age, gender, stage as well as grade (Figure 5D–K).

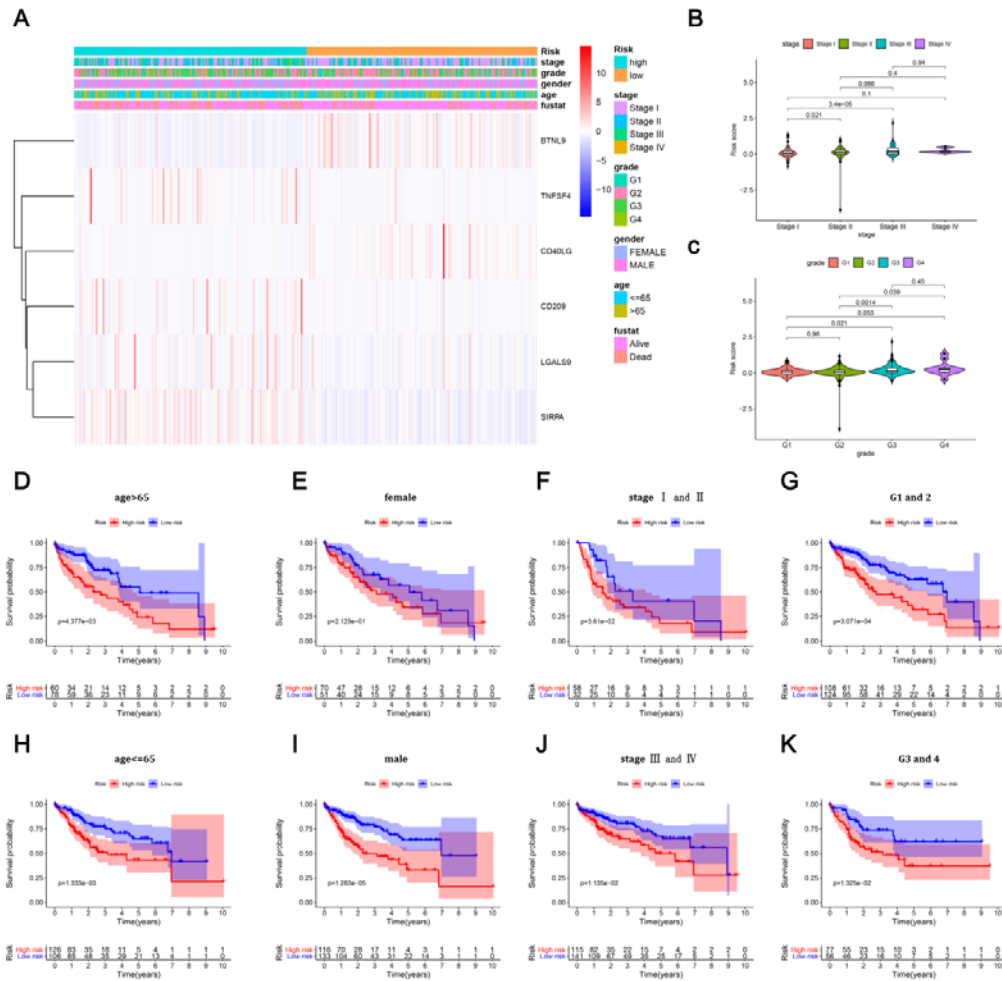


Figure 5. Clinical relevance assessment and subgroup survival analysis. (A) The heatmap showing the correlations between clinical traits, gene expression and risk groups. (B,C) Correlation analysis between the risk groups and pathological parameters. (D–K) Survival curves for different subgroups categorized by age, sex, stage and grade.

3.6. Development of a nomogram for clinical application

The findings from a univariate Cox regression analysis indicated that risk score ($P < 0.001$) and stage ($P < 0.001$) were substantially linked to unfavorable clinical outcomes (Figure 6A). Subsequently, we examined whether the risk score independently served as a reliable prognostic indicator. As shown in Figure 6B, risk score ($P < 0.001$) and stage ($P < 0.001$) independently acted as prognostic markers revealed by the multivariate analysis. We included these variables and constructed a nomogram (Figure 6C), aiming at providing individualized survival estimation for HCC patients. The calibration curves (Figure 6D–F) illustrated that the established model has an extremely high consistency with the ideal model. Besides, compared with the pathologic stage, higher prediction accuracy was found in the integration nomogram, exhibiting the largest AUC value (Figure 6G–I). More importantly, the DCA curves demonstrated that the nomogram exhibited larger net benefits, suggesting that it had more therapeutic value for clinical application (Figure 6J–L).

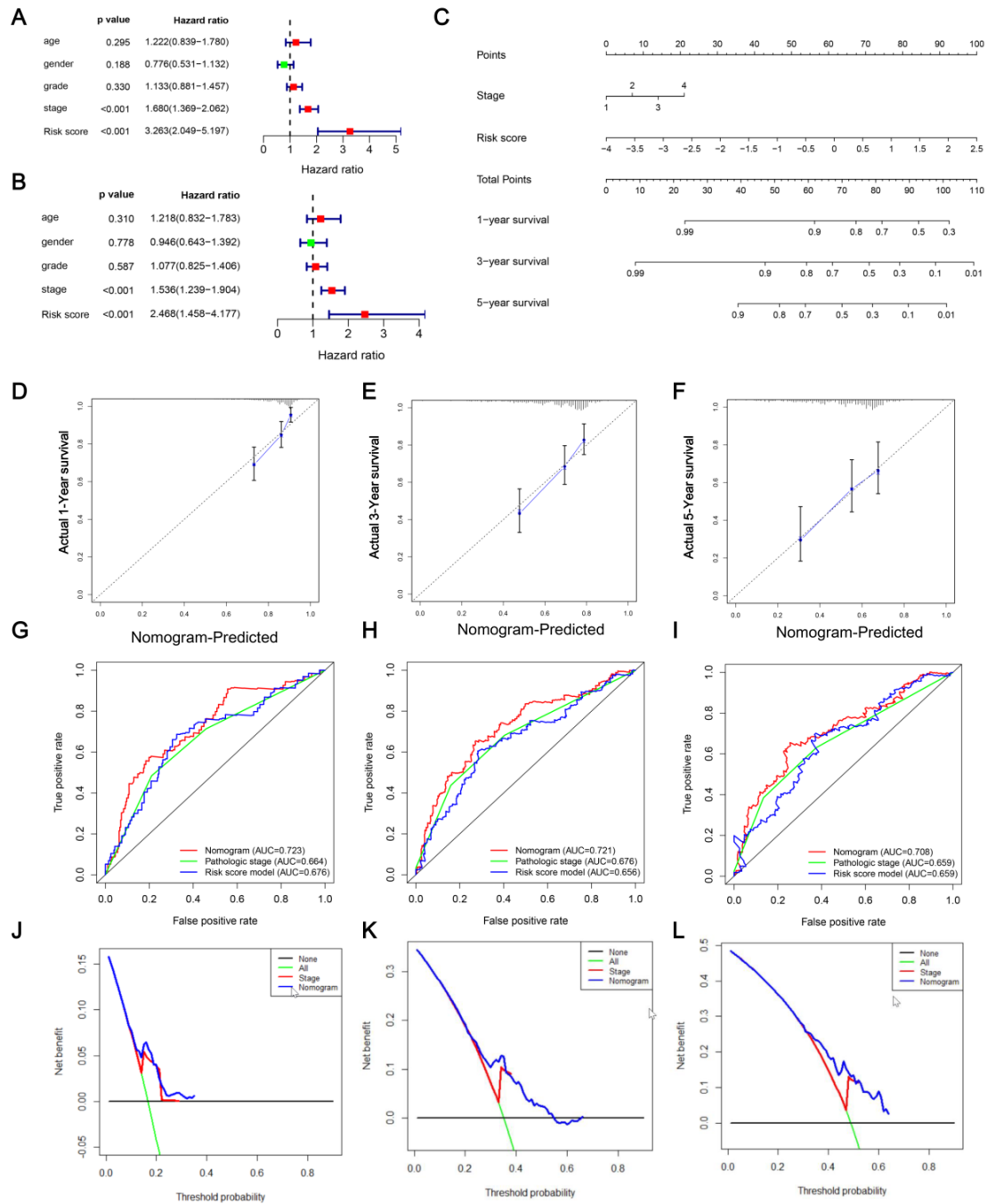


Figure 6. Development of a nomogram for overall survival (OS) prediction for HCC patients. (A,B) Univariate and multivariate Cox regression analyses revealed that risk score was an independent prognostic predictor in the TCGA datasets. (C) The nomogram to predict 1-year, 3-year and 5-year OS of HCC patients. (D–F) Calibrations curves for predicting OS at 1, 3 and 5-year; diagonal line: ideal model, vertical bars: 95% confidence interval. (G–I) Comparison of the time-dependent ROC curves for nomogram, pathologic stage and risk score. (J–L) Decision curves analysis of the nomogram and stage for OS prediction at 1, 3 and 5-year.

3.7. Assessment of tumor immune microenvironment

To examine the indicative roles of the gene signature on the tumor immune microenvironment, we applied the ssGSEA algorithm to measure the enrichment scores of a variety of immune cell subpopulations, associated pathways, or functions. As anticipated, highly infiltrated immune cells, including NK cells, CD8⁺ T cells and B cells, were found in the low-risk cohort (Figure 7A). Meanwhile, immune signaling pathways such as cytolytic activity, IFN response had an elevated score in the low-risk cohort (Figure 7B). Lollipop plot (Figure 7C) further showed that risk scores were negatively related to immune effectors cells including NK cells, CD8⁺ T cells, B cells and tumor-infiltrating lymphocytes. This result was in accordance with the conclusion that patients at high risk might exhibit unfavorable responsiveness to immunotherapy. Additionally, we examined the prognostic significance of these immune effector cells, among which higher infiltration levels of CD8⁺ T cells, B cells, NK cells and TIL were correlated with the better OS (Figure 7D–G).

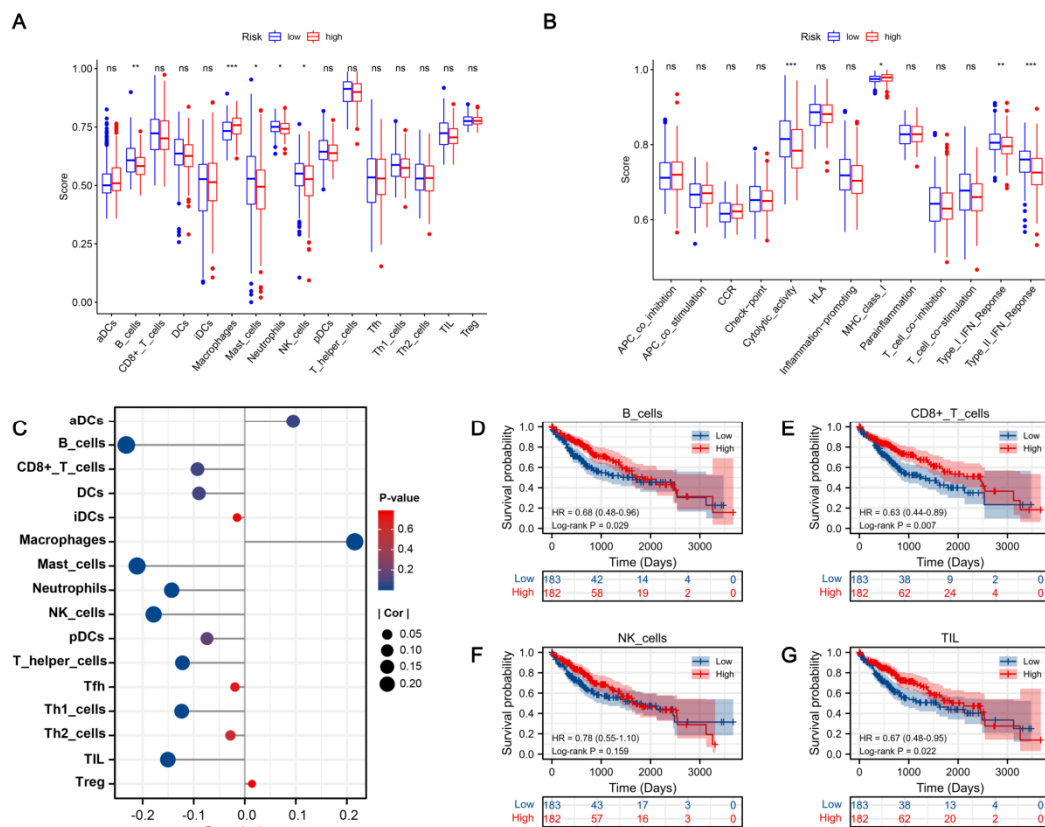


Figure 7. Evaluation of tumor immune microenvironment. (A, B) Comparison of the ssGSEA score of immune infiltrating cells and immune-related functions in different risk groups. (C) The lollipop plot depicting the link between immune infiltrating cells and risk scores; node size represents correlation coefficient; node color represents P-value. (D–G) Kaplan-Meier curves for HCC patients of different immune infiltration subgroups.

3.8. Potential of the risk score as an indicator reflecting therapeutic response

In previous studies, IPS has been proved to be reliable for predicting immunotherapy response [26,27]. We found that the scores of IPS, IPS-CTLA4, IPS-PD1/PD-L1/PD-L2, IPS-PD1/PDL1/PD-L2 + CTLA4 were elevated among low-risk patients (Figure 8A–D), suggesting these patients had a high likelihood of responding to inhibitors of immune checkpoint. In addition to the immunotherapy, we next sought to explore the link between risk scores and the effectiveness of chemotherapy. The IC₅₀ of five common chemotherapeutic drugs was computed via the pRRophetic algorithm. Figure 8E–I illustrated that higher risk scores were linked to a greater IC₅₀ for doxorubicin ($P = 0.015$) and mitomycin c ($P < 0.001$).

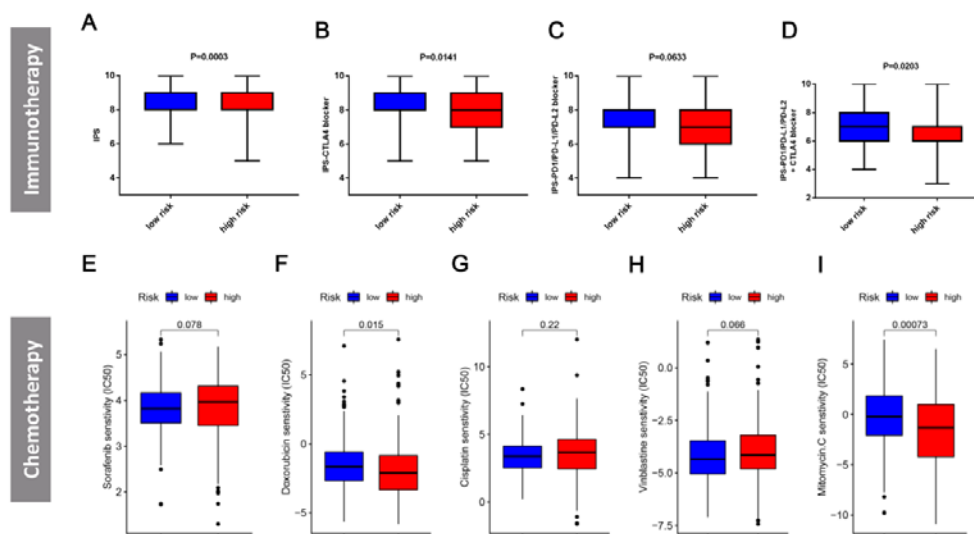


Figure 8. Analysis of therapeutic response. (A–D) The relationship between the immunophenoscore (IPS) and risk groups. (E–I) Comparison of the IC₅₀ of different chemotherapeutics in different risk groups.

4. Discussion

HCC is troublesome for having a high risk of relapse and a dismal prognosis. Thus, biomarkers correlated with HCC incidence, progression and prognosis must be unraveled to improve the clinical outcomes of patients. Although a variety of tumor markers have been widely used in clinical practice, only one biomarker might not be reliable enough to anticipate long-term outcomes or treatment responses. In contrast, the gene signatures could achieve better predictive performance, facilitating clinicians to judge the prognosis of patients accurately and distinguish subgroups to select appropriate treatments.

In recent years, immunotherapies that target programmed cell-death-1 receptor (PD-1) or its ligand (PD-L1) have gained substantial interest in the HCC treatment. The CheckMate-040 [28] and Keynote-224 [29] studies showed that OS of advanced HCC patients receiving nivolumab and pembrolizumab as second-line treatment was 15.6 months and 12.9 months, respectively. Compared with the regorafenib treatment (OS, 10.6 months), there was a significant increase. Growing evidence suggested that immune checkpoint blockade is one of the most promising approaches in HCC treatment.

However, it was also worth noting that the overall response rate of ICIs was only 10%–20% [28]. Hence, there is an unmet need to identify biomarkers guiding personalized treatment to effectively utilize medical resources and prevent adverse events caused by overtreatment. Several other immune checkpoint molecules in addition to PD-L1, which encompass TIGIT, TIM-3, LAG-3 and CTLA-4, have been proposed to participate in the immunomodulation of tumor microenvironment [30]. However, their exact prognostic value is controversial and remains to be further explored. Meanwhile, ICG-based prognostic models are still lacking in clinical practice.

So, in the current study, we conducted a literature review from which we collected 79 immune checkpoint genes and extracted the gene expression profiles in the TCGA database. OS-related ICGs were firstly identified in the univariate Cox regression analysis. Subsequently, we utilized Lasso Cox regression to minimize overfitting and constructed an optimal six-gene prognostic model. The prognostic model enabled us to classify all patients into two prognostic cohorts based on their risk ratings. The high-risk cohort exhibited a significant worse prognosis. Furthermore, the model's predictive capacity was demonstrated by ROC curve analysis, which was satisfactory. Meanwhile, the above results were also externally validated in the ICGC cohort. Interestingly, we found that Zhao et al. [31]. developed a similar predictive model based on PD-L1 and CTLA-4 signaling pathways genes. We plotted a time-dependent ROC curve to contrast the precision of the two models and the findings illustrated that our model exhibited a higher AUC index, reflecting its better prediction efficiency (Figure S3). Besides, our model is suitable for different clinical endpoints, like DSS and DFS, facilitating clinicians to make more comprehensive evaluations of patients' prognosis. Eventually, we created a nomogram incorporating risk signature and TNM stage to enhance clinical application values.

Our risk signature is composed of six genes, including LGALS9, CD209, CD40LG, SIRPA, BTNL9 and TNFSF4. The results showed that CD40LG and BTNL9 were protective factors, while the others were risk factors for HCC. Galectin 9, encoded by LGALS9 gene, is a ligand for TIM-3 and participates in immune tolerance [32]. The galectin-9 expression level was correlated with cancer progression and aggressiveness in diverse types of cancers [33,34], such as HCC [35]. Similar to our findings, Sideras et al. [19] found that the galectin-9 was associated with poor HCC-specific survival. CD40LG affiliated to the tumor necrosis factor gene superfamily, performs a pivotal function in T cell-dependent humoral responses through an interface with its classical receptor CD40 [36]. It has been suggested that CD40LG-expressing dendritic cells could suppress HCC progression by activating immune system in vivo [37]. SIRPA was related to the phenotype switch of tumor-polarized macrophages and served important functions in HCC progression [38]. Regrettably, the remaining three genes were identified and reported to be associated with the progression as well as the prognosis of the other tumors [39–41], but have never been reported in HCC. This remains to be explored in further studies.

Previous studies have confirmed that immune infiltration is a key factor affecting the clinical benefits of immunotherapy [42]. Our findings suggested that the immune cells infiltration levels were strongly associated with the gene signature as well as the prognosis of patients. Highly infiltrated CD8⁺ T cells, NK cells and B cells were found in the low-risk cohort and linked to improved prognosis. Kesh et al. [43] pointed out that elevated infiltration of CD8⁺ T cells may sensitive tumors to respond to immunotherapy. It was also reported that NK cells could cooperate with CD8⁺ T cells to improve therapeutic response in patients receiving CTLA-4 blockade therapy [44]. Furthermore, based on infiltrated immune subsets and immunomodulatory molecules, Charoentong et al. [25] applied the machine learning algorithm to develop the Immunophenoscore. It could be used to predict the therapeutic response to anti-CTLA-4 and anti-PD-1 therapy. In our study, low risk patients had a higher

IPS, suggesting that they were more likely to have higher immunogenicity. Thus, the current findings may have important implications for the identification of patients with a high likelihood of benefiting from immunotherapy.

Nonetheless, our study had some limitations. Firstly, some important clinical parameters, determining the prognosis, are not available in the TCGA database, such as the history of hepatitis or liver cirrhosis. Secondly, although we performed an external validation in the ICGC cohort, the conclusion would be more convincing with validation in a prospective, large-sample cohort. Furthermore, the biological activities of these hub genes in the prognostic signature remain to be further explored in experiments.

5. Conclusions

In conclusion, we established and verified a new ICG-related gene signature for HCC based on six hub genes, which could possibly function as a prognostic marker for clinical outcomes and therapeutic response.

Acknowledgements

This study was funded by National Key Research and Development Program of China (2020YFA0710803) and National Natural Science Foundation of China (No. 81770569, 81870421 and 81773072).

Conflict of interest

The authors declare that they have no competing interest.

Data availability statement

The data that support the findings of this study could be downloaded from TCGA (<https://portal.gdc.cancer.gov/>) databases and ICGC (<https://dcc.icgc.org/repositories>) database.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. Siegel, L. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.*, **68** (2018), 394–424. <https://doi.org/10.3322/caac.21492>
2. T. Greten, C. Lai, G. Li, K. Staveley-O' Carroll, targeted and immune-based therapies for hepatocellular carcinoma, *Gastroenterology*, **156** (2019), 510–524. <https://doi.org/10.1053/j.gastro.2018.09.051>
3. J. Llovet, R. Montal, D. Sia, R. Finn, Molecular therapies and precision medicine for hepatocellular carcinoma, *Nat. Rev. Clin. Oncol.*, **15** (2018), 599–616. <https://doi.org/10.1038/s41571-018-0073-4>
4. Y. Jiang, A. Sun, Y. Zhao, W. Ying, H. Sun, X. Yang, et al., Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma, *Nature*, **567** (2019), 257–261. <https://doi.org/10.1038/s41586-019-0987-8>

5. C. Qu, Y. Wang, P. Wang, K. Chen, M. Wang, H. Zeng, et al., Detection of early-stage hepatocellular carcinoma in asymptomatic HBsAg-seropositive individuals by liquid biopsy, *Proc. Natl. Acad. Sci. U. S. A.*, **116** (2019), 6308–6312. <https://doi.org/10.1073/pnas.1819799116>
6. S. Torrecilla, D. Sia, A. Harrington, Z. Zhang, L. Cabellos, H. Cornella, et al., Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma, *J. Hepatol.*, **67** (2017), 1222–1231. <https://doi.org/10.1016/j.jhep.2017.08.013>
7. C. Cainap, S. Qin, W. Huang, I. Chung, H. Pan, Y. Cheng, et al., Linifanib versus Sorafenib in patients with advanced hepatocellular carcinoma: results of a randomized phase III trial, *J. Clin. Oncol.: Off. J. Am. Soc. Clin. Oncol.*, **33** (2015), 172–179. <https://doi.org/10.1200/jco.2013.54.3298>
8. Y. Sun, Y. Xu, X. Yang, W. Guo, X. Zhang, S. Qiu, et al., Circulating stem cell-like epithelial cell adhesion molecule-positive tumor cells indicate poor prognosis of hepatocellular carcinoma after curative resection, *Hepatology*, **57** (2013), 1458–1468. <https://doi.org/10.1002/hep.26151>
9. J. Nault, A. Villanueva, Intratumor molecular and phenotypic diversity in hepatocellular carcinoma, *Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res.*, **21** (2015), 1786–1788. <https://doi.org/10.1158/1078-0432.ccr-14-2602>
10. L. Dong, L. Peng, L. Ma, D. Liu, S. Zhang, S. Luo, et al., Heterogeneous immunogenomic features and distinct escape mechanisms in multifocal hepatocellular carcinoma, *J. Hepatol.*, **72** (2020), 896–908. <https://doi.org/10.1016/j.jhep.2019.12.014>
11. S. Tella, A. Mahipal, A. Kommalapati, Z. Jin, Evaluating the safety and efficacy of nivolumab in patients with advanced hepatocellular carcinoma: evidence to Date, *Oncol. Targets Ther.*, **12** (2019), 10335–10342. <https://doi.org/10.2147/ott.s214870>
12. P. Galle, R. Finn, S. Qin, M. Ikeda, A. Zhu, T. Kim, et al., Patient-reported outcomes with atezolizumab plus bevacizumab versus sorafenib in patients with unresectable hepatocellular carcinoma (IMbrave150): an open-label, randomised, phase 3 trial, *Lancet Oncol.*, **22** (2021), 991–1001. [https://doi.org/10.1016/s1470-2045\(21\)00151-0](https://doi.org/10.1016/s1470-2045(21)00151-0)
13. H. Tan, N. Wang, C. Zhang, Y. Chan, M. Yuen, Y. Feng, Lysyl oxidase-like 4 fosters an immunosuppressive microenvironment during hepatocarcinogenesis, *Hepatology*, **73** (2021), 2326–2341. <https://doi.org/10.1002/hep.31600>
14. M. Wu, X. Xia, J. Hu, N. Fowlkes, S. Li, WSX1 act as a tumor suppressor in hepatocellular carcinoma by downregulating neoplastic PD-L1 expression, *Nat. Commun.*, **12** (2021), 3500. <https://doi.org/10.1038/s41467-021-23864-9>
15. X. Liu, M. Li, X. Wang, Z. Dang, Y. Jiang, X. Wang, et al., PD-1 TIGIT CD8 T cells are associated with pathogenesis and progression of patients with hepatitis B virus-related hepatocellular carcinoma, *Cancer Immunol. Immunother.*, **68** (2019), 2041–2054. <https://doi.org/10.1007/s00262-019-02426-5>
16. S. Itoh, T. Yoshizumi, K. Yugawa, D. Imai, S. Yoshiya, K. Takeishi, et al., Impact of immune response on outcomes in hepatocellular carcinoma: association with vascular formation, *Hepatology*, **72** (2020), 1987–1999. <https://doi.org/10.1002/hep.31206>
17. F. Liu, Y. Liu, Z. Chen, Tim-3 expression and its role in hepatocellular carcinoma, *J. Hematol. Oncol.*, **11** (2018), 126. <https://doi.org/10.1186/s13045-018-0667-4>
18. M. Guo, F. Yuan, F. Qi, J. Sun, Q. Rao, Z. Zhao, et al., Expression and clinical significance of LAG-3, FGL1, PD-L1 and CD8T cells in hepatocellular carcinoma using multiplex quantitative analysis, *J. Transl. Med.*, **18** (2020), 306. <https://doi.org/10.1186/s12967-020-02469-8>

19. K. Sideras, K. Biermann, J. Verheij, B. Takkenberg, S. Mancham, B. Hansen, et al., PD-L1, Galectin-9 and CD8 tumor-infiltrating lymphocytes are associated with survival in hepatocellular carcinoma, *Oncoimmunology*, **6** (2017), e1273309. <https://doi.org/10.1080/2162402x.2016.1273309>
20. X. Shen, B. Zhao, Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: meta-analysis, *BMJ*, **362** (2018), k3529. <https://doi.org/10.1136/bmj.k3529>
21. F. Hu, C. Liu, L. Liu, Q. Zhang, A. Guo, Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response, *Briefings Bioinf.*, **22** (2021), 176. <https://doi.org/10.1093/bib/bbaa176>
22. D. Pardoll, The blockade of immune checkpoints in cancer immunotherapy, *Nat. Rev. Cancer*, **12** (2012), 252–264. <https://doi.org/10.1038/nrc3239>
23. K. Campbell, A. Purdy, Structure/function of human killer cell immunoglobulin-like receptors: lessons from polymorphisms, evolution, crystal structures and mutations, *Immunology*, **132** (2011), 315–325. <https://doi.org/10.1111/j.1365-2567.2010.03398.x>
24. H. Afrache, P. Gouret, S. Ainouche, P. Pontarotti, D. Olive, The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response, *Immunogenetics*, **64** (2012), 781–794. <https://doi.org/10.1007/s00251-012-0619-z>
25. P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, *Cell Rep.*, **18** (2017), 248–262. <https://doi.org/10.1016/j.celrep.2016.12.019>
26. Y. Dai, W. Qiang, K. Lin, Y. Gui, X. Lan, D. Wang, An immune-related gene signature for predicting survival and immunotherapy efficacy in hepatocellular carcinoma, *Cancer Immunol. Immunother.*, **70** (2021), 967–979. <https://doi.org/10.1007/s00262-020-02743-0>
27. X. Hua, S. Ge, J. Zhang, H. Xiao, S. Tai, C. Yang, et al., A costimulatory molecule-related signature in regard to evaluation of prognosis and immune features for clear cell renal cell carcinoma, *Cell Death Discovery*, **7** (2021), 252. <https://doi.org/10.1038/s41420-021-00646-2>
28. A. El-Khoueiry, B. Sangro, T. Yau, T. Crocenzi, M. Kudo, C. Hsu, 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*, **389** (2017), 2492–2502. [https://doi.org/10.1016/s0140-6736\(17\)31046-2](https://doi.org/10.1016/s0140-6736(17)31046-2)
29. A. Zhu, R. Finn, J. Edeline, S. Cattan, S. Ogasawara, D. Palmer, 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.*, **19** (2018), 940–952. [https://doi.org/10.1016/s1470-2045\(18\)30351-6](https://doi.org/10.1016/s1470-2045(18)30351-6)
30. Z. Liu, X. Liu, J. Liang, Y. Liu, X. Hou, M. Zhang, et al., Immunotherapy for hepatocellular carcinoma: current status and future prospects, *Front. Immunol.*, **12** (2021), 765101. <https://doi.org/10.3389/fimmu.2021.765101>
31. E. Zhao, S. Chen, Y. Dang, Development and external validation of a novel immune checkpoint-related gene signature for prediction of overall survival in hepatocellular carcinoma, *Front. Mol. Biosci.*, **7** (2020), 620765. <https://doi.org/10.3389/fmolb.2020.620765>
32. S. Kim, H. Lee, M. Jeon, T. Yi, S. Song, Galectin-9 is involved in immunosuppression mediated by human bone marrow-derived clonal mesenchymal stem cells, *Immune Network*, **15** (2015), 241–251. <https://doi.org/10.4110/in.2015.15.5.241>

33. T. Kageshita, Y. Kashio, A. Yamauchi, M. Seki, M. Abedin, N. Nishi, et al., Possible role of galectin-9 in cell aggregation and apoptosis of human melanoma cell lines and its clinical significance, *Int. J. Cancer*, **99** (2002), 809–816. <https://doi.org/10.1002/ijc.10436>
34. A. Irie, A. Yamauchi, K. Kontani, M. Kihara, D. Liu, Y. Shirato, et al., Galectin-9 as a prognostic factor with antimetastatic potential in breast cancer, *Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res.*, **11** (2005), 2962–2968. <https://doi.org/10.1158/1078-0432.ccr-04-0861>
35. S. Chen, J. Pu, J. Bai, Y. Yin, K. Wu, J. Wang, et al., EZH2 promotes hepatocellular carcinoma progression through modulating miR-22/galectin-9 axis, *J. Exper. Clin. Cancer Res.*, **37** (2018), 3. <https://doi.org/10.1186/s13046-017-0670-6>
36. S. Schoenberger, R. Toes, E. van der Voort, R. Offringa, C. Melief, T-cell help for cytotoxic T lymphocytes is mediated by CD40-CD40L interactions, *Nature*, **393** (1998), 480–483. <https://doi.org/10.1038/31002>
37. M. Gonzalez-Carmona, V. Lukacs-Kornek, A. Timmerman, S. Shabani, M. Kornek, A. Vogt, et al., CD40ligand-expressing dendritic cells induce regression of hepatocellular carcinoma by activating innate and acquired immunity in vivo, *Hepatology*, **48** (2008), 157–168. <https://doi.org/10.1002/hep.22296>
38. Y. Pan, Y. Tan, M. Wang, J. Zhang, B. Zhang, C. Yang, et al., Signal regulatory protein α is associated with tumor-polarized macrophages phenotype switch and plays a pivotal role in tumor progression, *Hepatology*, **58** (2013), 680–691. <https://doi.org/10.1002/hep.26391>
39. M. Yuan, X. Zhang, J. Zhang, K. Wang, Y. Zhang, W. Shang, et al., DC-SIGN-LEF1/TCF1-miR-185 feedback loop promotes colorectal cancer invasion and metastasis, *Cell Death Differ.*, **27** (2020), 379–395. <https://doi.org/10.1038/s41418-019-0361-2>
40. W. Ma, J. Liang, J. Mo, S. Zhang, N. Hu, D. Tian, et al., Butyrophilin-like 9 expression is associated with outcome in lung adenocarcinoma, *BMC Cancer*, **21** (2021), 1096. <https://doi.org/10.1186/s12885-021-08790-9>
41. Y. Li, Y. Chen, L. Miao, Y. Wang, M. Yu, X. Yan, et al., Stress-induced upregulation of TNFSF4 in cancer-associated fibroblast facilitates chemoresistance of lung adenocarcinoma through inhibiting apoptosis of tumor cells, *Cancer Lett.*, **497** (2021), 212–220. <https://doi.org/10.1016/j.canlet.2020.10.032>
42. W. Fridman, F. Pagès, C. Sautès-Fridman, J. Galon, The immune contexture in human tumours: impact on clinical outcome, *Nat. Rev. Cancer*, **12** (2012), 298–306. <https://doi.org/10.1038/nrc3245>
43. K. Kesh, V. Garrido, A. Dosch, B. Durden, V. Gupta, N. Sharma, et al., Stroma secreted IL6 selects for “stem-like” population and alters pancreatic tumor microenvironment by reprogramming metabolic pathways, *Cell Death Dis.*, **11** (2020), 967. <https://doi.org/10.1038/s41419-020-03168-4>
44. F. Kohlhapp, J. Broucek, T. Hughes, E. Huelsmann, J. Lusciks, J. Zayas, et al., NK cells and CD8+ T cells cooperate to improve therapeutic responses in melanoma treated with interleukin-2 (IL-2) and CTLA-4 blockade, *J. Immunother. Cancer*, **3** (2015), 18. <https://doi.org/10.1186/s40425-015-0063-3>

