

Immune-associated molecular occurrence and prognosis predictor of hepatocellular carcinoma: an integrated analysis of GEO datasets

Guanjun Liu^{a,b,c,*}, Dongmei Wu^{d,*}, Yiyang Wen^{a,b,c}, and Shundong Cang^{a,b,c,*}

^aDepartment of Oncology, Henan Provincial People's Hospital, Zhengzhou, Henan, P.R. China; ^bDepartment of Oncology, People's Hospital of Zhengzhou University, Zhengzhou, Henan, P.R. China; ^cDepartment of Oncology, People's Hospital of Henan University, Zhengzhou, Henan, P.R. China; ^dDepartment of Radiotherapy, Xinxiang Center Hospital, Xinxiang, Henan, P.R. China

ABSTRACT

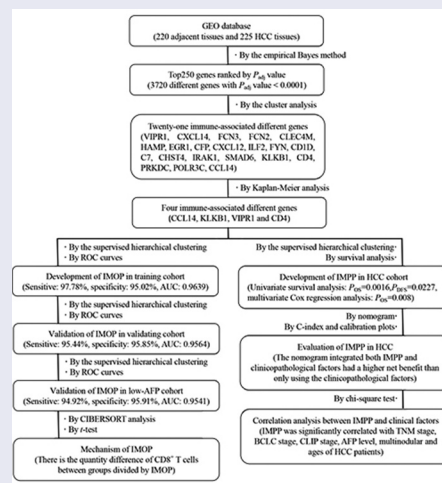
Hepatocellular carcinoma (HCC) is the fifth most common cancer and the second most common cause of cancer-related deaths worldwide. As immune response failure is the main factor in the occurrence and poor prognosis of HCC, our study aimed to develop an immune-associated molecular occurrence and prognosis predictor (IMOPP) of HCC. To that end, we discovered a 4-gene immune-associated gene signature: C–C motif chemokine ligand 14 (*CCL14*), kallikrein B1 (*KLKB1*), vasoactive intestinal peptide receptor 1 (*VIPR1*), and cluster of differentiation 4 (*CD4*). When tested on three cohorts as an immune-associated molecular occurrence predictor (IMOP), it had high sensitivity, specificity, and area under the receiver operating characteristics curve. When tested as an immune-associated molecular prognosis predictor (IMPP), it stratified the HCC prognosis for overall survival (Kaplan–Meier analysis, log rank $P = 0.0016$; Cox regression, HR = 1.832, 95% CI = 1.173–2.859, $P = 0.008$) and disease-free survival (Kaplan–Meier analysis, log rank $P = 0.0227$). IMPP also significantly correlated with the clinicopathological characteristics of HCC; integrating it with clinicopathological characteristics improved the accuracy of a nomogram for overall survival prediction (C-index: 0.7097 vs. 0.6631). In HCC tumor micro-environments, the proportion of CD8⁺ T cells significantly differed between IMOP-stratified groups. We conclude that IMOPP can potentially predict the occurrence of HCC in high-risk populations and improve prognostic accuracy by providing new biomarkers for risk stratification. In addition, we believe that the IMOP mechanism may be related to its effect on the proportion of CD8⁺ T cells in tumor-infiltrating lymphocytes.





ARTICLE HISTORY

Received 4 May 2021
Revised 29 June 2021
Accepted 30 June 2021


KEYWORDS

Molecular predictor;
immune; hepatocellular
carcinoma; occurrence;
prognosis; cd8⁺ t cells



CONTACT Dongmei Wu  mdong0822@126.com  Department of Radiotherapy, Xinxiang Center Hospital, No. 56 of Jinsui Road, Xinxiang, Henan 453000, P.R. China; Shundong Cang  cangshundong@163.com  Department of Oncology, Henan Provincial People's Hospital, No. 7 Of Weiwu Road, Zhengzhou, Henan, 450003, P.R. China

*These authors are contributed equally to this work.

 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer and the second most common cause of global cancer-related deaths [1]. Approximately 782,500 new HCC cases and 745,500 deaths occurred worldwide during 2012, with China alone accounting for roughly 50% of the total number of cases and deaths [2]. More worrisome is the incidence and mortality of HCC, which have been increasing internationally in recent decades [2,3]. However, molecular mechanisms involved in the development of HCC remain ambiguous, and therapeutic methods are limited largely to surgical resection and liver transplantation. Diagnostic delays, high recurrence rates, and high metastatic rates contribute to a 5-year survival rate of less than 30% in HCC patients who undergo surgical resection [4]. As early HCC diagnosis and the identification of HCC patients with poor prognosis are critical to improving survival, investigating the biomarkers for HCC occurrence and prognosis is of considerable importance.

Serum alpha fetal protein (AFP) has been used extensively for decades as a molecular predictor for the diagnosis of HCC. However, its suitability for HCC surveillance is reportedly low [5]. In recent years, several molecular predictors for the occurrence and prognosis of HCC have been proposed [6,7]. Unfortunately, these molecular predictors are not always suitable for evaluating the occurrence and prognosis of HCC [6,7].

Immune response, an important factor in maintaining internal homeostasis, reportedly plays a role in immune surveillance and prevention of infections. Immune-associated genetic dysfunction is the key cause of cancer development and occurrence [8,9]. Ahmad et al. found that Toll-like receptor 4 can suppress antitumor response and trigger the development of ultraviolet B-induced skin cancer [10]; Nicoud et al. [11] discovered that histamine H₄ receptor can upregulate the proportion of cluster of differentiation (CD)4⁺ CD25⁺ FoxP3⁺ regulatory T cells and promote an immunosuppressive milieu. Immune-associated predictors have been shown to accurately predict the occurrence or prognosis of ovarian cancer and colorectal cancer [12–14]. For example, a study on

Ewing sarcoma [15] reported that a gene signature consisting of 11 immune-associated genes correlates with patient prognosis and can be used as a reliable prognostic biomarker. However, there are few studies applying immune-associated signatures to the prediction of HCC occurrence and prognosis.

In the current study, we hypothesized that immune-associated genes play an important role in HCC development. We aimed to develop an immune-associated molecular occurrence and prognosis predictor (IMOPP) for HCC and evaluate its potential mechanisms. In order to accomplish our goals, we first analyzed data from the Gene Expression Omnibus (GEO) database on 445 patients with HCC to identify differentially expressed immune-associated genes for HCC. We then developed and validated an immune-associated molecular occurrence predictor (IMOP) by using three cohorts that had been established by two independent GEO datasets, and we evaluated the IMOP using areas under the receiver operating characteristic (ROC) curve (AUC). At the same time, we developed an immune-associated molecular prognosis predictor (IMPP), which we evaluated by using survival analysis. In addition, we constructed a nomogram by combining the IMPP with clinico-pathological characteristics to increase the accuracy of prognostic predictions. Finally, we used a CIBERSORT algorithm to determine the potential mechanism of IMOP in the development of HCC.

Methods and materials

Study design

In this retrospective study of data from two GEO databases, we aimed to develop an IMOPP that can identify patients at high risk of HCC occurrence and those with poor HCC prognosis. First, we reanalyzed the GSE14520 database utilizing the GPL3921 platform (GSE14520-GPL3921) and identified immune-associated genes by means of cluster analysis and Kaplan–Meier analysis. Second, an IMOP was developed and validated by analyzing AUC in a training cohort, validation cohort, and low-AFP cohort. Third, an IMPP was developed using an HCC cohort, and IMPP efficiency was evaluated by Kaplan–Meier analysis,

Cox regression, Harrell's concordance index (C-index), and a calibration plot. Finally, we used the CIBERSORT algorithm to quantify tumor-infiltrating immune cells in tumor microenvironments and discover the potential mechanism of the HCC IMOP.

Patients and cohorts

For the HCC IMOP, we used data collected from a total of 878 patients in two GEO databases. To develop the IMOP, we established a training cohort using all 445 tissues in the GSE14520-GPL3921 dataset (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14520>). To validate the IMOP, we established a validation cohort using all 433 tissues in the GSE36376 dataset (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36376>). To establish the low-AFP cohort, we chose 220 adjacent tissues and 118 HCC tissues with AFP levels lower than 300 ng/mL from the GSE14520-GPL3921 dataset. Detailed clinical information and informed consent had been separately provided to and reported by Roessler S et al. [16,17] and Lim HY [18].

Owing to insufficient survival information, we developed the IMPP by choosing 221 HCC tissues from the GSE14520-GPL3921 dataset to establish an HCC cohort (Supplementary Table 1). The endpoints were HCC overall survival (defined as the time from surgery to death) and HCC disease-free survival (defined as the time from surgery to any recurrence, distant metastasis, or death from any cause) as described by Roessler and colleagues [16,17].

Identification of differentially expressed immune-associated genes

In order to identify differentially expressed genes associated with the pathogenesis of HCC, we downloaded the GSE14520-GPL3921 microarray expression profile and reanalyzed it using 'limma' package in the R programming environment (v2.15.3) [19]. Then, the top 250 of those genes with P values < 0.001 were chosen for further Gene Ontology (GO) analysis using the DAVID website (<https://david.ncifcrf.gov/>) [20]; we then focused

on the pathways associated with immune response (P value < 0.05).

Development and validation of the IMOPP

Machine learning is a type of artificial intelligence that is widely used for the diagnosis and prognosis of disease [6]. In the present study, we used hierarchical clustering to develop and validate an IMOP. We first extracted immune-associated gene expression data from the training cohort, validation cohort, and low-AFP cohort. Then, hierarchical clustering (based on the Manhattan distance and the clustering methods of Ward. D) was conducted using 'pheatmap' package in the R programming environment (v3.2.2).

We used the HCC cohort to develop an IMPP, employing the same strategy as that used for IMOP. Generally, all patients were categorized into two groups: the normal immunity group (containing patients with relatively more adjacent tissues or better prognosis) and the immunodeficient group (containing patients with relatively fewer adjacent tissues or poorer prognosis).

Evaluation of tumor-infiltrating immune cells in tumor microenvironment

CIBERSORT algorithm is an analytical tool that uses gene expression data to evaluate specific cell types within a mixed cell population [21]. To estimate the composition of tumor-infiltrating immune cells in HCC and quantify the relative levels of distinct immune cell types, we utilized the CIBERSORT platform (<https://cibersort.stanford.edu/>). The analysis was performed using an arrangement of 100 default statistical parameters. At a threshold of $P < 0.05$, results of the inferred fractions of tumor-infiltrating immune cells produced by CIBERSORT were considered accurate. When the average relative percent of tumor-infiltrating immune cells exceeded 0.01 for any group, a comparison was performed between the normal immunity group and the immunodeficient group.

Statistical analysis

In the R programming environment, ROC curves were plotted with the 'pROC' package, and cluster

analysis was performed using the ‘GOplot’ package. The chi-square test was used for analyzing the correlation between two groups, Kaplan–Meier method was used to compare overall survival and disease-free survival between different groups, and the log-rank test was used to estimate the significance of differences in survival rates. Survival curves were plotted with Graphpad prism 5.0 software (GraphPad Software Inc., San Diego, CA, USA). Cox regression was utilized to analyze overall survival and disease-free survival, and results were reported as hazard ratios (HR) with 95% confidence intervals (CI). A nomogram combining the risk score model with clinicopathological characteristics was constructed using the ‘rms’ package [22]. Discrimination by the prognostic models was measured and compared using Harrell’s concordance index (C-index) and a calibration plot. Statistical analyses were performed using SPSS Statistics software (v19.0; SPSS Inc., IL, USA). All statistical tests were two-tailed, and $P \leq 0.05$ was considered statistically significant.

Results

Considering the harm caused by HCC and the important role that immune response plays in HCC development, our aims were to develop an IMOPP for HCC and to evaluate its potential mechanisms. To that end, this study identified an immune-associated gene signature consisting of four genes: C–C motif chemokine ligand 14 (*CCL14*), kallikrein B1 (*KLKB1*), vasoactive intestinal peptide receptor 1 (*VIPRI*) and *CD4*. We found that when tested as an IMOP on three cohorts, the predictor had high sensitivity, specificity, and AUC. When tested as an IMPP, the predictor could stratify the HCC prognosis for overall survival (based on Kaplan–Meier analysis and Cox regression) and disease-free survival (based on Kaplan–Meier analysis). A nomogram that integrated IMPP with clinicopathological characteristics showed improved accuracy in overall survival prediction compared to a nomogram that used only clinicopathological characteristics. In addition, IMPP results also significantly correlated with the clinicopathological characteristics of HCC. In HCC tumor microenvironments, the

difference in number of CD8⁺ T cells was significant among IMOP-stratified groups.

Identification of a differentially expressed immune-associated gene signature for HCC

To gain a global understanding of the pathogenesis of HCC, the GSE14520-GPL3921 microarray expression profile was used to discover genes that were differentially expressed between HCC tissue and adjacent tissues. The preliminary results showed that expression levels significantly differed among 3720 genes (adjusted P value < 0.0001 , Figure 1(a)). The top 250 of those genes were used for further Gene Ontology (GO) analysis. According to their attributes and biological functions (Figure 1(b)), these differentially expressed genes were classified as: i) cell cycle (e.g., GO: 0007049, GO: 0022402, GO: 0000278), ii) immune response (e.g., GO: 0050778, GO: 0048548, GO:002684), and iii) oxidation reduction (e.g., GO:0055114).

Owing to the importance of their function, special attention was paid to pathways associated with immune response. Twenty-one genes associated with immune response were involved with the immune activation response and acute inflammation response (Figure 1(c)).

Immune-associated genes can influence HCC overall and disease-free survival. Kaplan–Meier analysis revealed that *CCL14* was significantly associated with HCC disease-free survival (Figure 1(d)) and that *CCL14* and *KLKB1* were significantly associated with overall survival (Figure 1(e)). In order to maximize the predictive power of these immune-associated genes, we extended the survival analysis P value threshold to 0.1 and thereafter identified four genes for use as an IMOPP of HCC: *CCL14*, *KLKB1*, *VIPRI*, and *CD4*.

Development and validation of an IMOP for HCC

Clustering, which is widely used for data mining, feature extraction, disease diagnosis and prognostic evaluation, is a fundamental data analysis method in machine learning [23,24]. In order to develop a robust molecular occurrence predictor for HCC, we used a hierarchical clustering

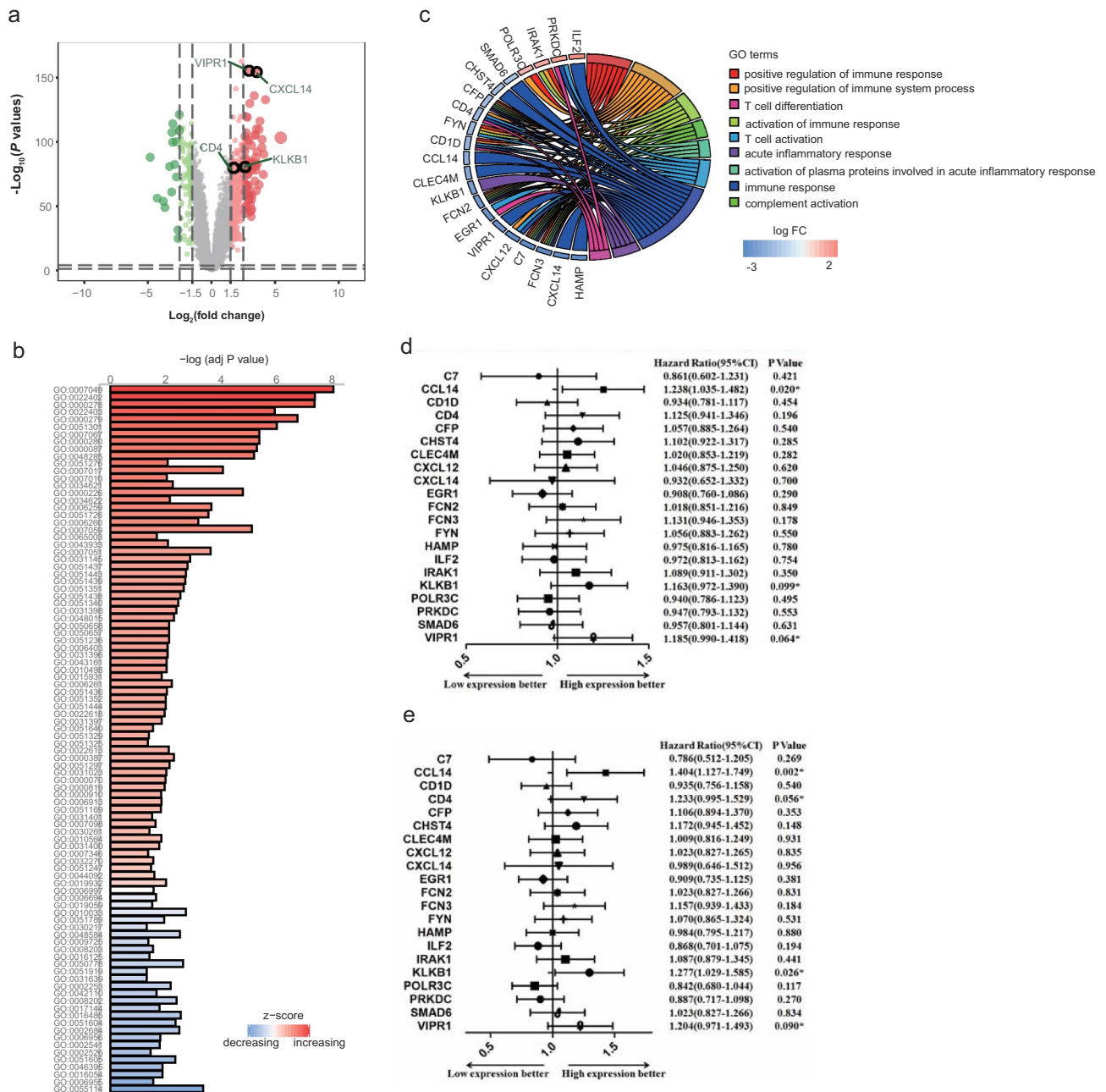


Figure 1. Identification of a differentially expressed immune-associated gene signature for HCC. (a) Volcano plot shows differentially expressed genes based on an absolute log_2 fold change (FC) >1.5 and adjusted P value < 0.05; black circles indicate four immune-associated genes. (b) GO analysis of the top 250 differentially expressed genes. (c) Circos plot illustrating the relationship between differentially expressed immune-associated genes and GO terms for HCC. (d) Univariate analysis of the relationship between differentially expressed immune-associated genes and HCC overall survival. (e) Univariate analysis of the relationship between differentially expressed immune-associated genes and HCC disease-free survival.

algorithm. Results showed that the sensitive and specificity of IMOP in the training cohort were 97.78% and 95.02%, respectively (Figure 2(a), Table 1); the AUC was 0.9639 (Figure 2(b)). For the validation cohort, the sensitive and specificity were 95.44% and 95.85%, respectively (Supplementary Figure 1a, Table 1); the AUC

was 0.9564 (Figure 2(b)). For the low-AFP cohort, the sensitive and specificity were 94.92% and 95.91%, respectively (Supplementary Figure 1b, Table 1); the AUC was 0.9541 (Figure 2(b)). High sensitivity, specificity, and AUC can indicate the superiority of a predictor. These results prove that the IMOP possessed significant capability for

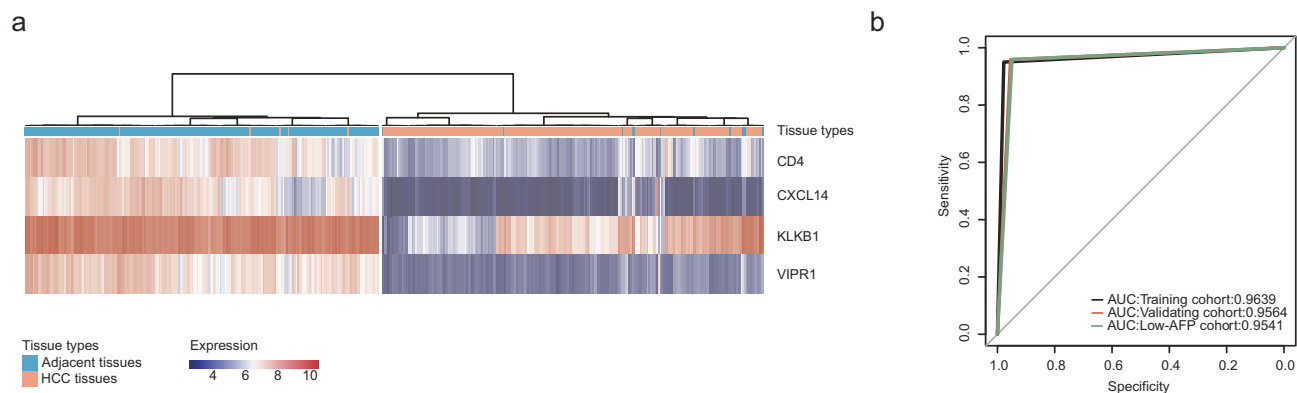


Figure 2. Development and validation of an IMOP for HCC. (a) Development of the IMOP via hierarchical clustering analysis of the training cohort. (b) ROC validation of IMOP in the training cohort, validation cohort, and low-AFP cohort.

Table 1. Sensitivity and specificity of the immune-associated molecular occurrence predictor in the training cohort, validation cohort, and low-AFP cohort.

Clinical diagnosis	Training cohort			Validating cohort			Low-AFP cohort		
	HCC tissues	adjacent normal tissues	Percent	HCC tissues	adjacent normal tissues	Percent	HCC tissues	adjacent normal tissues	Percent
HCC tissues	220	5	97.78%	230	11	95.44%	112	6	94.92%
adjacent tissues	11	210	95.02%	8	185	95.85%	9	211	95.91%

distinguishing between HCC tissue and adjacent tissues.

Development and evaluation of an IMPP for HCC

In the correlation analysis between IMPP and overall survival or disease-free survival in HCC patients, we first used hierarchical clustering to divide HCC patients from the GSE14520-GPL3921 dataset into two groups: the immunodeficient group and the normal immunity group (Figure 3(a)). For prognostic estimation, we found that the IMPP was significantly associated with HCC overall survival (log rank $P = 0.0016$; Figure 3(b)) and disease-free survival (log rank $P = 0.0227$; Figure 3(c)).

In order to eliminate the influence of clinicopathological characteristics on HCC overall survival and disease-free survival, we performed Cox regression analysis on the two IMPP-stratified groups. After eliminating the influence of clinicopathological characteristics, we found that IMPP was significantly associated with HCC overall survival (HR = 1.832, 95% CI = 1.173–2.859,

$P = 0.008$; Table 2) but not associated with disease-free survival ($P = 0.145$; Table 2).

To develop a quantitative method for predicting overall survival in HCC patients, we created a nomogram that integrated IMPP with clinicopathological characteristics (Figure 3(d)). Subsequently, we utilized C-index and calibration plots to evaluate the nomogram. C-index analysis revealed that the net benefit was higher for the nomogram that integrated IMPP with clinicopathological characteristics (C-index = 0.7097) than for the nomogram that was constructed using only clinicopathological characteristics (C-index = 0.6631). Similarly, the calibration plot (Figure 3(e)) showed that the nomogram that integrated IMPP with clinicopathological characteristics performed well when compared to the performance of an ideal model.

HCC prognosis is closely related to clinicopathological characteristics such as AFP level, tumor-node-metastasis (TNM) stages, and Barcelona Clinic Liver Cancer (BCLC) stages. For the correlation analysis between IMPP and the clinicopathological characteristics of HCC, we evaluated differences between the immunodeficient group and the normal immunity group

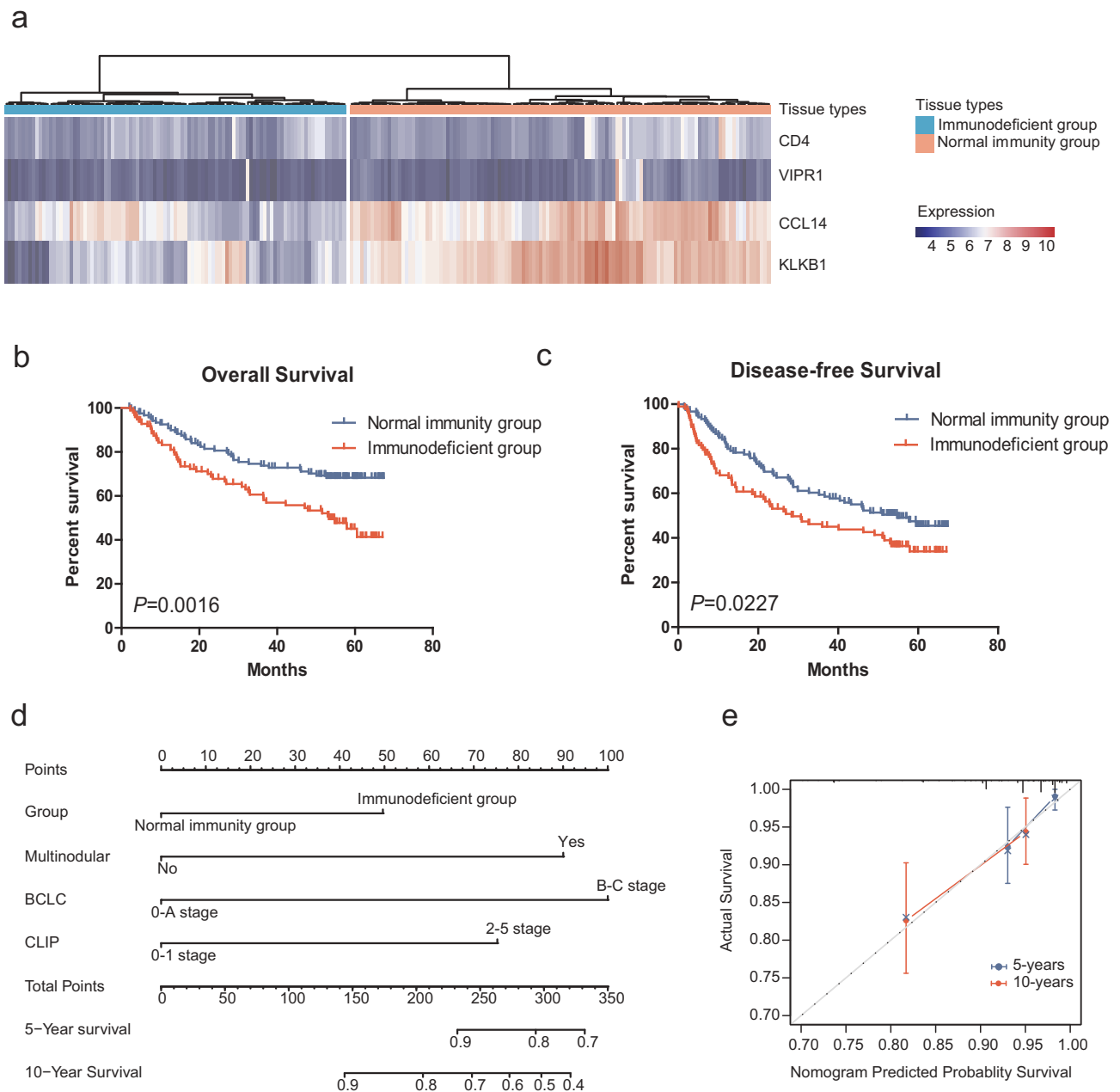


Figure 3. Development and validation of an IMPP for HCC. (a) Development of the IMPP via hierarchical clustering. (b) Overall survival was significantly lower in the immunodeficient group than in the normal immunity group. (c) Disease-free survival was also significantly lower in the immunodeficient group than in the normal immunity group. (d) Nomogram for predicting 5- and 10-year overall survival for HCC patients based on IMPP and clinicopathological characteristics (MN, Multinodular; BCLC, Barcelona Clinic liver cancer; CLIP, Cancer of the Liver Italian Program). (e) Calibration plot illustrating agreement between observed and nomogram-predicted 5- and 10-year outcomes. The 45° dashed line represents perfect prediction; the actual results from our nomogram are indicated by blue and red lines.

using the chi-square test. Our findings showed that IMPP significantly correlated with TNM stage, BCLC stage, Cancer of the Liver Italian Program (CLIP) stage, AFP level, multinodular presentation, and HCC patient age (Table 3).

Analysis of differences in proportion of CD8⁺ T cells between the HCC IMOP-stratified immunodeficient and normal immunity groups

We used the CIBERSORT algorithm to evaluate the relationship between HCC IMOP and tumor-

Table 2. Univariate and multivariate Cox regression analyses of disease-free survival and overall survival between HCC patient groups stratified by the immune-associated molecular prognosis predictor.

Characteristics	Univariates analysis		Multivariates analysis	
	Hazard ratio (95% CI)	P Value	Hazard ratio (95% CI)	P Value
Disease-free survival				
Group (immunodeficient group or normal immunity group)	1.537(1.074–2.200)	0.019	-	0.145
Gender (male or female)	1.480(1.071–2.047)	0.018	1.387(1.001–1.921)	0.049
MTS (≤ 5 cm or > 5 cm)	0.832(0.692–1.001)	0.051	-	0.880
Cirrhosis (no or yes)	-	0.053	-	0.122
TNM stage (I-II stage or III stage)	0.668(0.547–0.816)	0.000	-	0.746
BCLC stage (0-A stage or B-C stage)	0.609(0.500–0.741)	0.000	0.627(0.514–0.764)	0.000
CLIP stage (0–1 stage or 2–5 stage)	0.699(0.573–0.854)	0.000	-	0.754
Overall survival				
Group (immunodeficient group or normal immunity group)	2.029(1.316–3.129)	0.001	1.832(1.173–2.859)	0.008
MTS (≤ 5 cm or > 5 cm)	0.707(0.570–0.877)	0.002	-	0.588
MN (no or yes)	-	0.056	1.729(1.241–2.410)	0.001
Cirrhosis (no or yes)	0.464(0.230–0.935)	0.032	-	0.052
TNM stage (I-II stage or III stage)	0.535(0.428–0.671)	0.000	-	0.119
BCLC stage (0-A stage or B-C stage)	0.533(0.426–0.666)	0.000	0.540(0.380–0.767)	0.001
CLIP stage (0–1 stage or 2–5 stage)	0.561(0.449–0.701)	0.000	0.628(0.444–0.889)	0.009

MTS, Main tumor size; MN, Multinodular; BCLC, Barcelona Clinic liver cancer; CLIP, Cancer of the Liver Italian Program.

Table 3. Correlation analysis between immune-associated molecular prognosis predictor and the clinicopathological characteristics of HCC patients.

Characteristics		immunodeficient group	normal immunity group	X ²	P Value
Gender	Male	90(85.6)	101(105.4)	3.073	0.080
	Femal	9(13.4)	21(16.6)		
Age(years)	≤ 60	87(81.1)	94(99.9)	4.324	0.038
	> 60	12(17.9)	28(22.1)		
HBV vortal status	AVR-CC	27(24.8)	29(31.2)	0.463	0.496
	CC	67(69.2)	89(86.8)		
ALT	≤ 50 U/L	53(58.2)	77(71.8)	2.071	0.150
	> 50 U/L	46(40.8)	45(50.2)		
AFP	≤ 300 ng/ml	40(53.0)	78(65.0)	12.707	< 0.001
	> 300 ng/ml	58(45.0)	42(55.0)		
Main tumor size	≤ 5 cm	61(63.0)	79(77.0)	0.317	0.573
	> 5 cm	38(36.0)	42(44.0)		
Multinodular	no	73(78.8)	103(97.2)	3.851	0.050
	yes	26(20.2)	19(24.8)		
Cirrhosis	no	5(8.9)	13(11.1)	3.348	0.067
	yes	94(90.1)	109(102.9)		
TNM stage	I-II stage	68(76.1)	102(93.9)	6.930	0.008
	III stage	30(21.9)	19(27.1)		
BCLC stage	0-A stage	68(75.2)	100(92.8)	5.327	0.021
	B-C stage	30(22.8)	21(28.2)		
CLIP stage	0–1 stage	69(76.5)	102(94.5)	6.104	0.013
	2–5 stage	29(21.5)	19(26.5)		

NA, not available; HBV, Hepatitis B Virus; AVR-CC, active viral replication chronic carrier; CC, chronic carrier; ALT, alanine aminotransferase; AFP, Alpha Fetal Protein; BCLC, Barcelona Clinic liver cancer; CLIP, Cancer of the Liver Italian Program.

There are statistically significant between the two groups.

infiltrating immune cells [21]. As shown in Figure 4 (a), the composition of immunity cells (22 cell types) varied significantly among samples; furthermore, a large proportion of the cells were CD8⁺ T cells, CD4⁺ memory resting T cells, follicular helper T cells, gamma delta T cells, activated natural killer cells, M1 macrophages, M2 macrophages, and

resting dendritic cells. Mean proportional values for cell types that had a proportion greater than 0.01 in at least one group were compared between immunodeficient and normal immunity groups. The results showed that the memory B cell, CD8⁺ T cell, M1 macrophage, and M2 macrophage ratios were significantly higher in the normal immunity group

than in the immunodeficiency group ($P = 6.58E-4$, $P = 3.65E-10$, $P = 3.43E-4$, and $P = 3.03E-12$, respectively; **Figure 4(b)**). In contrast, the CD4⁺ naive T cell, regulatory T cell, activated natural killer cell, M0 macrophage, resting dendritic cell, activated dendritic cell, and resting mast cell ratios were significantly lower in the normal immunity group than in the immunodeficient group ($P = 2.21E-8$, $P = 3.42E-2$, $P = 6.91E-3$, $P = 5.18E-24$, $P = 1.80E-3$, $P = 3.27E-5$, and $P = 4.58E-4$, respectively; **Figure 4(b)**). These results indicated that the activation and

inhibition of various immune cells in tumor micro-environments occurred simultaneously.

CD8⁺ T cells can affect the development of HCC [25,26]. Thus, we compared the CD8⁺ T-cell markers between immunodeficient and normal immunity groups. The results (**Figure 4(c)**) showed that the levels of general T-cell markers (CD2 and CD3E) and specific CD8⁺ T-cells markers (CD8A and CD8B) were significantly lower in the immunodeficient group than in the normal immunity group.

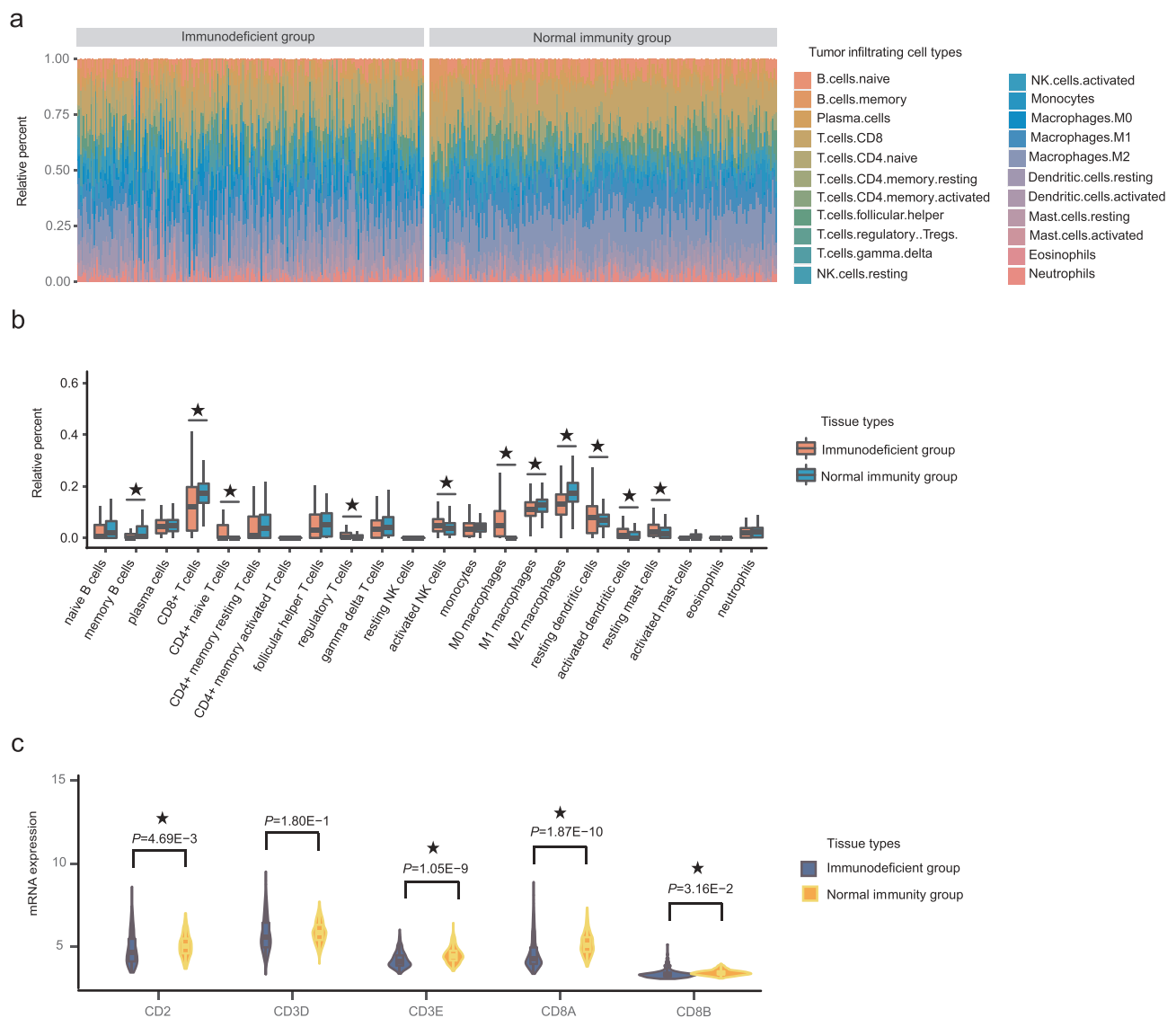


Figure 4. The proportion of CD8⁺ T cells significantly differed between the HCC IMOP-stratified immunodeficient and normal immunity groups. (a) Stacked bar chart showing the relative distribution immune cells (22 cell types) in each sample. (b) Box plot illustrates relative percent of tumor-infiltrating immune cells between the HCC IMOP-stratified immunodeficient and normal immunity groups. (c) Violin plot shows differential expression of CD8⁺ T-cell markers between the HCC IMOP-stratified immunodeficient and normal immunity groups.

Discussion

Effective methods are needed for predicting HCC occurrence and prognosis in patients. To address this issue, we identified a potential signature involving four immune-associated genes that were differentially expressed between HCC tissue and normal tissues and therefore could be valuable for predicting HCC occurrence and prognosis. Subsequently, we found that the IMOP possessed significant capability for distinguishing between HCC tissue and adjacent tissues in the training cohort, validation cohort, and low-AFP cohort. Meanwhile, the IMPP also showed excellent discriminatory ability in HCC patients with poor prognosis. To develop a quantitative method for predicting the overall survival of HCC patients, we combined clinicopathological characteristics with IMPP to construct a nomogram and then confirmed its performance using C-index and a calibration curve. With regard to the mechanism of IMOP, we found that the difference in proportion of CD8⁺ T cells was significant among IMOP-stratified groups.

Liver biopsy and pathological diagnosis continue to be the 'gold standard' for evaluating chronic hepatitis and fibrosis and for diagnosing HCC [27]. However, liver biopsy and pathological diagnosis have some disadvantages [27,28]. High-risk patients are usually diagnosed with HCC at the intermediate-advanced stage [29]. In tumors, molecular changes often precede morphological changes [30]. Thus, molecular predictors may be one of the most effective methods for monitoring HCC.

The role of immune response in the occurrence and development of HCC is well recognized [31–33]. Among HCC patients infected by hepatitis B virus (HBV) or hepatitis C virus, immune suppression owing to viral infection and viral gene integration are possible mechanisms associated with the occurrence of HCC [34–36]. The abnormal expression of various characteristic tumor surface antigens can inhibit antitumor immunity, thereby affecting HCC development and HCC treatment [37,38]. The present study found that in HCC patients, *CCL14* was significantly associated with overall survival and disease-free survival, and *KLKB1* was significantly associated with overall survival. Our results concur with those of

other studies. In one HCC study, researchers found that *CCL14* can serve as a novel HCC tumor suppressor by regulating cell cycles and promoting apoptosis [39]. Another study on epithelial ovarian cancer revealed that *CCL14* upregulation is associated with a favorable prognosis [40]. The differentially expressed immune-associated genes may play important roles in the development of tumor. In order to maximize the predictive power of differentially expressed immune-associated genes, we ensured that the four genes (*CCL14*, *KLKB1*, *VIPR1*, and *CD4*) chosen for development as an IMOPP for HCC had *P* values of less than 0.1 for overall or disease-free survival.

Serological tests performed alone or in combination (e.g., AFP, des-gamma-carboxy prothrombin, and AFP-L3) are the main methods for detecting HCC. However, AFP (at a cutoff of 20 ng/mL) has a sensitivity of 40–65% and 14–40% for clinically diagnosed HCC and preclinical HCC, respectively [41]. In addition, the sensitivity and specificity of des-gamma-carboxy prothrombin for clinically diagnosing HCC are only 28–89% and 87–96%, respectively – values which are similar to those for AFP-L3 [41]. Molecular predictors display superior stratification capability. For example, Lin et al. found that one miRNA classifier had high sensitivity (range 70.4–85.7%) and specificity (80.0–91.1%) [6]. In this study, our IMOP had 95% sensitivity and specificity in the training and validation cohorts. That is, our IMOP had higher sensitivity and specificity than do AFP [41,42] and other molecular predictors [6]. Moreover, the sensitivity and specificity of the IMOP reach approximately 95% in the low-AFP cohort. These results show that IMOP has robust capability for distinguishing HCC tissue from adjacent tissue.

Partial hepatectomy is regarded as a standard curative treatment for HCC, but the prognosis for HCC patients who undergo hepatectomy varies even among early-stage HCC patients [43,44]. Therefore, it is important to stratify HCC patients according to pathological parameters, protein biomarkers, mRNA expression levels, and genomic DNA abnormalities [45–47]. In the analysis of prognostic ability, we found that IMPP could effectively be used to divide HCC patients into

two groups – the immunodeficient group and the normal immunity group – the latter of which was significantly associated with HCC overall survival and disease-free survival. Even after eliminating the influence of clinicopathological characteristics, IMPP was significantly associated with overall survival. An analysis using C-index and a calibration plot with well-fitted calibration curves showed that a nomogram that integrated IMPP with clinicopathological characteristics had higher accuracy in overall survival prediction (C-index: 0.7097 vs. 0.6631) provided a greater net benefit than did a nomogram that used only clinicopathological characteristics. In the correlation analysis of IMPP and clinicopathological characteristics, results showed significant differences in TNM stage, BCLC stage, CLIP stage, AFP level, and ages between the two groups of HCC patients. To some extent, IMPP demonstrated its ability to stratify HCC patients with poor prognosis.

The genes associated with immune response are involved in positive regulation of the immune system process, such as T-cell activation and T-cell differentiation. We speculated that a strong immune-associated molecular predictor presence is significantly associated with tumor-infiltrating immune cells. CIBERSORT analysis revealed that the normal immunity group had a higher proportion of CD8⁺ T cells and higher expression levels of CD8⁺ T-cell markers than did the immunodeficient group. As the most predominant tumor-infiltrating lymphocyte, CD8⁺ T cell is ascribed the role of cytotoxic killer cell in cancer immunobiology, and infiltration of CD8⁺ T cell has been regarded as a marker of superior prognosis [48]. In a recent meta-analysis of 3509 patients from 21 observational studies, high levels of intratumoral CD8⁺ T cells were positively correlated with overall survival and disease-free survival [49]. Our experimental results confirmed these results. Regarding cause, most mechanisms of the molecular predictor are obscure in terms of number of CD8⁺ T cells affected. In future studies, we plan to explore the mechanism by which IMOP influences the occurrence of HCC.

There were some limitations in our study. First, we retrospectively studied genetic and clinical data from two independent cohorts in the GEO database; the patients were enrolled in East Asia and had

different ethnic and environmental backgrounds. Second, our IMOPP may be limited to HBV-related HCC because the majority of patients in our study were HBV-positive. Owing to these limitations, it will be necessary to validate our predictor in prospective cohorts with different etiological backgrounds and different ethnic characteristics.

Conclusions

In summary, we developed and validated an IMOPP that can potentially predict the occurrence of HCC in high-risk populations and provide new biomarkers for risk stratification and accurate prognostication. In addition, we found that the IMOP mechanism may involve inducing changes in the number of CD8⁺ T cells. Early diagnosis and accurate stratification of HCC patients help clinicians personalize clinical treatment. Thus, we believe that in spite of its limitations, our HCC IMOPP is important because it allows for individualized treatment recommendations.

Highlights

- (1) We identified a 4-gene signature for HCC: *CCL14*, *KLKB1*, *VIPRI* and *CD4*.
- (2) IMOPP potentially predicts HCC occurrence and improves prognostic accuracy.
- (3) IMOP may affect the proportion of CD8⁺ T cells in tumor-infiltrating lymphocytes.

Acknowledgements

We would like to thank Jie Ping and Zihao Chen for their technical support with data analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was funded by the Doctor Research Initial Foundation of Henan Provincial People's Hospital [ZC20170017].

Author's contributions

GJL analyzed the data, prepared figures and tables, authored drafts of the paper, and approved the final draft. DMW conceived and designed the experiments, analyzed the data, reviewed drafts of the paper, and approved the final draft. YYW analyzed the data and approved the final draft. SDC conceived and designed the experiments, reviewed drafts of the paper, and approved the final draft.

Data availability statement

Data analyzed during the present study are available at NCBI GEO: GSE14520 and GSE36376.

References

- [1] Wei L, Wang X, Lv L, et al. The emerging role of microRNAs and long noncoding RNAs in drug resistance of hepatocellular carcinoma. *Mol Cancer*. 2019;18:147.
- [2] Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87–108.
- [3] Yu X, Gao X, Mao X, et al. Knockdown of FOXO6 inhibits glycolysis and reduces cell resistance to paclitaxel in HCC cells via PI3K/Akt signaling pathway. *Onco Targets Ther*. 2020;13:1545–1556.
- [4] Singal AG, Lok AS, Feng Z, et al. Conceptual model for the hepatocellular carcinoma screening continuum: current status and research agenda. *Clin Gastroenterol Hepatol*. 2020;S1542: 3565:31297.
- [5] Liang KH, Lai MW, Lin YH, et al. Plasma interleukin-17 and alpha-fetoprotein combination effectively predicts imminent hepatocellular carcinoma occurrence in liver cirrhotic patients. *BMC Gastroenterol*. 2021;21:177.
- [6] Lin XJ, Chong Y, Guo ZW, et al. A serum microRNA classifier for early detection of hepatocellular carcinoma: a multicentre, retrospective, longitudinal biomarker identification study with a nested case-control study. *Lancet Oncol*. 2015;16:804–815.
- [7] Farinati F, Vitale A, Spolverato G, et al. Development and validation of a new prognostic system for patients with hepatocellular carcinoma. *PLOS Med*. 2016;13: e1002006.
- [8] Wang C, Cheng Y, Zhang X, et al. Vacuolar protein sorting 33B is a tumor suppressor in hepatocarcinogenesis. *Hepatology*. 2018;68:2239–2253.
- [9] Wang D, Li X, Li J, et al. APOBEC3B interaction with PRC2 modulates microenvironment to promote HCC progression. *Gut*. 2019;68:1846–1857.
- [10] Ahmad I, Nasti TH, Rihan HM, et al. Toll-like receptor-4 deficiency inhibits ultraviolet radiation-induced tumor development by modulation of immune and inflammatory responses. *Mol Carcinog*. 2021;60:60–70.
- [11] Nicoud MB, Táquez Delgado MA, Sarasola MP, et al. Impact of histamine H4 receptor deficiency on the modulation of T cells in a murine breast cancer model. *Cancer Immunol Immunother*. 2021;70:233–244.
- [12] Hao D, Liu J, Chen M, et al. Immunogenomic analyses of advanced serous ovarian cancer reveal immune score is a strong prognostic factor and an indicator of chemosensitivity. *Clin Cancer Res*. 2018;24:3560–3571.
- [13] Maker AV, Ito H, Mo Q, et al. Genetic evidence that intratumoral T-cell proliferation and activation are associated with recurrence and survival in patients with resected colorectal liver metastases. *Cancer Immunol Res*. 2015;3:380–388.
- [14] Poźniak J, Nsengimana J, Laye JP, et al. Genetic and environmental determinants of immune response to cutaneous melanoma. *Cancer Res*. 2019;79:2684–2696.
- [15] Ren EH, Deng YJ, Yuan WH, et al. An immune-related gene signature for determining Ewing sarcoma prognosis based on machine learning. *J Cancer Res Clin Oncol*. 2021;147:153–165.
- [16] Roessler S, Jia HL, Budhu A, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res*. 2010;70:10202–10212.
- [17] Roessler S, Long EL, Budhu A, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. *Gastroenterology*. 2012;142:957–966.e12.
- [18] Lim HY, Sohn I, Deng S, et al. Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. *Ann Surg Oncol*. 2013;20:3747–3753.
- [19] Zhang Y, Di X, Chen G, et al. An immune-related signature that to improve prognosis prediction of breast cancer. *Am J Cancer Res*. 2021;11:1267–1285.
- [20] Dennis G Jr., Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4:P3.
- [21] Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453–457.
- [22] Yin CD, Hou YL, Liu XR, et al. Development of an immune-related prognostic index associated with osteosarcoma. *Bioengineered*. 2021;12:172–182.
- [23] Tosado J, Zdilar L, Elhalawani H, et al. Clustering of largely right-censored oropharyngeal head and neck cancer patients for discriminative groupings to improve outcome prediction. *Sci Rep*. 2020;10(1):3811.
- [24] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res*. 2018;46:10546–10562.
- [25] Gabrielson A, Wu Y, Wang H, et al. Intratumoral CD3 and CD8 T-cell densities associated with relapse-free survival in HCC. *Cancer Immunol Res*. 2016;4:419–430.

- [26] Ma J, Zheng B, Goswami S, et al. PD1^{Hi} CD8⁺ T cells correlate with exhausted signature and poor clinical outcome in hepatocellular carcinoma. *J Immunother Cancer*. 2019;7:331.
- [27] Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2019;156:1264–1281.e4.
- [28] Zhou JH, Cai JJ, She ZG, et al. Noninvasive evaluation of nonalcoholic fatty liver disease: current evidence and practice. *World J Gastroenterol*. 2019;25:1307–1326.
- [29] Jun L, Yang G, Zhisu L. The utility of serum exosomal microRNAs in hepatocellular carcinoma. *Biomed Pharmacother*. 2019;111:1221–1227.
- [30] Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2014;15:178–196.
- [31] Makarova-Rusher OV, Medina-Echeverz J, Duffy AG, et al. The yin and yang of evasion and immune activation in HCC. *J Hepatol*. 2015;62:1420–1429.
- [32] Chen Y, Tian Z. HBV-induced immune imbalance in the development of HCC. *Front Immunol*. 2019;10:2048.
- [33] Itoh S, Yoshizumi T, Yugawa K, et al. Impact of immune response on outcomes in hepatocellular carcinoma: association with vascular formation. *Hepatology*. 2020;72:1987–1999.
- [34] Lafaro KJ, Demirjian AN, Pawlik TM. Epidemiology of hepatocellular carcinoma. *Surg Oncol Clin N Am*. 2015;24:1–17.
- [35] Song G, Shi Y, Zhang M, et al. Global immune characterization of HBV/HCV-related hepatocellular carcinoma identifies macrophage and T-cell subsets associated with disease progression. *Cell Discov*. 2020;6:90.
- [36] Lim CJ, Lee YH, Pan L, et al. Multidimensional analyses reveal distinct immune microenvironment in hepatitis B virus-related hepatocellular carcinoma. *Gut*. 2019;68:916–927.
- [37] Kondo Y, Kimura O, Tanaka Y, et al. Differential expression of CX3CL1 in hepatitis B virus-replicating hepatoma cells can affect the migration activity of CX3CR1⁺ immune cells. *J Virol*. 2015;89:7016–7027.
- [38] Dargel C, Bassani-Sternberg M, Hasreiter J, et al. T cells engineered to express a T-cell receptor specific for Glypican-3 to recognize and kill hepatoma cells in vitro and in mice. *Gastroenterology*. 2015;149:1042–1052.
- [39] Zhu M, Xu W, Wei C, et al. CCL14 serves as a novel prognostic factor and tumor suppressor of HCC by modulating cell cycle and promoting apoptosis. *Cell Death Dis*. 2019;10:796.
- [40] Cai Y, Ling Y, Huang L, et al. C-C motif chemokine 14 as a novel potential biomarker for predicting the prognosis of epithelial ovarian cancer. *Oncol Lett*. 2020;19:2875–2883.
- [41] Marrero JA, Lok AS. Newer markers for hepatocellular carcinoma. *Gastroenterology*. 2004;127(Supplement 1):S113–9.
- [42] Trevisani F, D’Intino PE, Morselli-Labate AM, et al. Serum alpha-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. *J Hepatol*. 2001;34:570–575.
- [43] Liu S, Miao C, Liu J, et al. Four differentially methylated gene pairs to predict the prognosis for early stage hepatocellular carcinoma patients. *J Cell Physiol*. 2018;233:6583–6590.
- [44] Nathan H, Schulick RD, Choti MA, et al. Predictors of survival after resection of early hepatocellular carcinoma. *Ann Surg*. 2009;249:799–805.
- [45] Bruix J, Reig M, Sherman M. Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. *Gastroenterology*. 2016;150:835–853.
- [46] Kim JH, Sohn BH, Lee HS, et al. Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLOS Med*. 2014;11:e1001770.
- [47] Wen L, Li J, Guo H, et al. Genome-scale detection of hypermethylated CpG islands in circulating cell-free DNA of hepatocellular carcinoma patients. *Cell Res*. 2015;25:1250–1264.
- [48] Liao H, Zhang Z, Chen J, et al. Preoperative radiomic approach to evaluate tumor-infiltrating CD8⁺ T cells in hepatocellular carcinoma patients using contrast-enhanced computed tomography. *Ann Surg Oncol*. 2019;26:4537–4547.
- [49] Yao W, He JC, Yang Y, et al. The prognostic value of tumor-infiltrating lymphocytes in hepatocellular carcinoma: a systematic review and meta-analysis. *Sci Rep*. 2017;7:7525.