Construction of an immune-related prognostic model to predict prognosis and immunotherapy in liver cancer patients with hepatitis B virus-infected

Nan Dong1,*, Chen Fu1**

1 School Health Center, Shenyang Center for Disease Control and Prevention, Shenyang 110122, China. 2 Department of Pharmacology, School of Pharmacy, China Medical University, Shenyang 110122, China.

* These authors contributed equally to this work and are co-first authors for this paper.

Author contributions
Nan Dong and Chen Fu conceived the project. Nan Dong and Chen Fu conducted the research. Nan Dong and Chen Fu analyzed and interpreted the data. Nan Dong and Chen Fu were responsible for the provided materials. Chen Fu provided research guidance. Nan Dong and Chen Fu was a major contributor in writing the manuscript. All authors read and approved the final manuscript. Competing interests
The authors declare no conflicts of interest.

Acknowledgments
This study was supported by the Shenyang City-School Joint Funding Project (No. 2400022093).

Peer review information
Medical Data Mining thanks all anonymous reviewers for their contribution to the peer review of this paper.

Abstract
Background: Hepatocellular carcinoma (HCC) appears to be strongly associated with immune-related genes. However, immune-related genes are not well understood as a prognostic marker in HCC caused by the hepatitis B virus (HBV). The purpose of this study was to investigate the prognostic significance of immune-related genes in HBV-infected HCC.

Methods: Gene expression data from 114 HBV-infected HCC and 50 normal tissues were integrated into The Cancer Genome Atlas. Differentially expressed immune-associated genes were analyzed to identify immune-associated differential genes associated with overall survival. Least Absolute Shrinkage and Selection Operator and multivariate Cox regressions were used to constructing immunoprognostic models. An independent prognostic factor analysis using multiple Cox regressions was also performed for HBV-infected HCCs. Immunocorrelation analysis markers and immune cell infiltration were also investigated.

Results: We found 113 differentially expressed immune-associated genes. Immune-related differential genes were significantly correlated with the overall survival of HCC patients. We constructed an immune-based prognostic model using multivariate Cox regression analysis including seven immune-related genes. According to further analysis, immune-related prognostic factors may serve as independent prognostic indicators in the clinical setting. There is also evidence that the 7-gene prognostic model reflects the tumor immune microenvironment as a result of the risk score model and immune cell infiltration.

Conclusions: As a result of our study, we screened immune-related genes for prognosis in HBV-infected HCC and developed a novel immune-based prognostic model. The research not only provides new prognostic biomarkers but also offers insight into the tumor immune microenvironment and lays the theoretical groundwork for immunotherapy.

Keywords: hepatocellular carcinoma; hepatitis B virus; immune-related genes; prognosis
Background

Globally, the most common form of liver cancer is hepatocellular carcinoma (HCC) [1]. It is estimated that millions of people will die from liver cancer by the end of this century [2]. Generally, patients with HCC are diagnosed in the middle to late stages of the disease as there are no obvious clinical findings in the early stages of the disease [3]. In addition, liver cancer is most effectively treated by surgery. Despite considerable improvements in surgical techniques, surgery remains the most effective method of treatment. It is estimated that 5% of all HCC patients will survive beyond five years [4]. To improve prognosis, it is imperative to develop new therapies that address these problems.

Under normal physiological conditions, the liver is exposed to antigens from the gut, including those derived from diet and microbial products. This exposure makes the liver inherently immunogenic and capable of suppressing inappropriate inflammatory responses. The relationship between inflammation and the development of hepatocellular carcinoma and cholangiocarcinoma has led to a growing interest in immune-based approaches for treating these cancers. The liver serves as an organ that regulates the immune system, allowing it to respond to harmful stimuli while maintaining immune tolerance. However, this carefully regulated immune tolerance network becomes disrupted in chronic inflammatory liver diseases, which in turn promotes the progression of liver tumors. The tumor microenvironment in hepatocellular carcinoma is complex, consisting of various innate and adaptive immune cells that influence the evasion of the immune system by the tumor, the response to immunotherapy, and the overall survival of the patient.

The hepatitis B virus (HBV) increases the risk of liver cancer [5]. Current research suggests that, in addition to insertional mutations and activation and trans-activation of host genes by viral proteins and some other factors, chronic liver injury and tissue regeneration are important triggers for hepatocyte malignancy and the development of HCC [6]. However, HBV is a non-cytotoxic virus and does not cause hepatocellular damage per se; liver damage during HBV infection is mainly caused by the body’s anti-HBV immune response [7]. Thus, pathological mechanisms related to the immune response are critical to the development of HBV-associated HCC. However, the role of the natural immune response in this process is not clear.

Current research suggests that even when chronic HBV is clinically and functionally cured, the HBV is still not completely cleared. When patients use chemotherapy drugs, immunosuppressive agents, and biological agents for other diseases, the related immune response is suppressed, leading to HBV reactivation, which may be manifested as asymptomatic alanine transaminase elevation in mild cases and spontaneous remission in some patients; in severe cases, signs of liver failure such as xanthogranuloma, ascites, coagulation abnormalities, and encephalopathy may occur [8]. It is characterized by rapid onset, not easily controlled, high mortality, and a poor prognosis, and requires special clinical attention in these patients. Generally, this is due to genetic heterogeneity and the immune system [9, 10]. It is characterized by rapid onset, not easily controlled, high mortality, and a poor prognosis, and requires special clinical attention in these patients [11, 12].

Therefore, the construction of a reliable immune-based prognostic model for predicting HCC. The study started with a detailed analysis of differential and immune-associated genes associated with HBV infection. Secondly, the relationship between further detection of immune-associated significant differential genes and prognosis was elaborated. Then, we integrated the immune-related genes of HBV-infected HCC to construct an immune-related prognostic model. Furthermore, building on the immune-associated independent prognostic value model to further look at its relationship with immune infiltrating cells. In this study, we aim to provide new biomarkers that can be evaluated by the tumor immune microenvironment to predict the prognosis of HBV-infected HCC patients.

Materials and methods

Data collection

HCC gene expression data and sample clinical profile were downloaded from the The Cancer Genome Atlas (TCGA) database (as of 16 July 2022) [13]. Gene expression data included RNA-Seq FPKM data from a total of 375 HCC and 50 normal cases, after careful searching and examination of 114 HCC cases with co-infection with HBV. The Immunology Database and Analysis Portal (ImmPort, https://www.immport.org/shared/genel) provided us with 1811 immune-related genes, based on various molecular functions, the immune system is divided into 17 categories [14]. There are no ethical or moral requirements for the publication of the data since it is directly obtained from public databases and strictly adheres to the publication guidelines.

Differential expression analysis of immune genes

Differentially expressed immune-related genes (IRGs) were detected in hepatocellular carcinoma and normal tissue by the Wilcoxon test, with \( \log 2 FC > 2 \) and \( P < 0.05 \) defined as significant. Heat and volcano maps were mapped using the phentmap package in R software [15, 16]. Assessment of potential differentially expressed biologically functional of IRGs, Gene Ontology [17], and Kyoto Encyclopedia of Genes and Genomes analysis [18]. \( P < 0.05 \) was considered a significant screening condition.

Construction of immune-related markers for hepatocellular carcinoma

The expression of prognostic IRGs and overall survival (OS) was assessed using Least Absolute Shrinkage and Selection Operator (LASSO) and multivariate Cox regression analyses. The survival status and clinicl features were used to perform LASSO analyses and Cox multivariate analyses to avoid overfitting and deletion of highly correlated genes. Genes were multivariate Cox regression analyses by stepwise assessment of the LASSO algorithm assays. Risk scores were obtained as a linear product of combined regression coefficients based on gene expression. To compare the overall survival of high- and low-risk patients, Kaplan-Meier (K-M) analysis was performed [19]. Subgroup characteristics curve analysis was performed by survival status using R software. The characteristics of risk scores on OS and multiple clinical effects were also assessed using univariate and multivariate analyses of variance.

Construction of prognostic features and column line graphs

In order to simplify the COX model, we analyzed K-M curves and receiver operating characteristic (ROC) curves within the LASSO model. Genes were selected to have good prognostic performance and diagnostic ability to construct a predictive signal. A risk score formula based on prognostic characteristics was used to determine the risk level for each patient [8].

\[ \text{Risk Score} = \sum_{i=1}^{n} \text{Coef}_i \times \text{Exp}_i \]

The \( n \) denotes the significance of the number of genes included in the prognosis. Coef_i denotes the LASSO coefficient for gene i. Exp_i denotes the expression value of gene i.

Depending on the expression of the gene characterized, each patient has a unique risk score. All patients were divided into median groups of high and low risk. A ROC analysis of the prognostic significance and diagnostic power of K-M curves for risk scores. Subsequently, we further analyzed the immune microenvironment tumor-infiltrating immune cells in heterogeneous different risk groups in a stepwise analysis of immune infiltration data. Immunological evaluation using the Wilcoxon test showed statistically significant differences in cell infiltration between the high-risk and low-risk groups (\( P < 0.05 \)). A multivariate COX regression for multiple clinical characteristics was used to determine whether risk scores under interference have independent prognostic value. Independent predictive features were then included in the construction of column line plots. An index of agreement between actual and predicted survival is also given for assessing nomogram survival predictive power [19]. Visualization of
prediction results and observation curves during the calibration process to measure prediction performance in line graphs.

**Correlation analysis of immune-related features and infiltration of immune cells**
The Tumor Immune Estimation Resource (TIMER) program (Comprehensive Analysis of Immune Cells for Tumour Infiltration) is used to explore prognostic genetic models and immune cell infiltration resources in tumor immunity estimation. The TIMER estimates correlations between 6 types of component infiltrating immune cells. Using TIMER data on immune infiltration in hepatocellular carcinoma, prognostic models were plotted against six immune cells.

**Results**

**mRNA expression profiles between HBV-infected liver cancer tissues and paracancerous tissues**

HCC and paracancerous liver tissues infected with HBV were differentially compared \( (P < 0.05, \log \text{FC} > 2) \) and 1416 differential genes were identified (Figure 1A, 1B). A total of 1316 mRNAs were significantly up-regulated, and 100 mRNAs were significantly down-regulated. IRGs were further screened in hepatocellular carcinoma tissues infected with HBV normal liver tissues (Figure 1C, 1D). The study identified 113 differentially expressed genes related to immunity, of which 83 were upregulated and 20 were downregulated.

**Construction of an immune-related risk profile model**

Immune-related risk features were constructed using the training set in the TCGA database and validated using the test set and the entire test set. The results of our univariate Cox regression analysis indicated that 33 immune-related genes had prognostic predictive power out of 113 IRGs investigated. In the LASSO-Cox regression analysis, 33 immune-related genes were included. After 100 rounds of 10-fold cross-validation, when lambda took a minimum value of 0.059, 7 best IRGs were identified. The immune-related risk profiles of HCC patients were established using these 7 IRGs (Figure 2A, 2B). And Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses were performed on them (Figure 2C, 2D). The risk score equations were as follows.

\[
\text{Risk score} = (0.0514) \times \text{FABP6} + (0.0314) \times \text{RP2P2} + (0.1439) \times \text{MAPT} + (0.1252) \times \text{BIRC5} + (0.0756) \times \text{CSPG5} + (0.0456) \times \text{SP1} + (0.1257) \times \text{STC2}
\]

**Immunity-related risk characteristics**

To validate the predictive power of these characteristics, we recruited 114 samples from the TCGA database that had been infected with HBV. The risk score formula described above was used to calculate risk scores for each patient, training scores were set as critical scores for each group based on patients’ risk scores (wilcorank-sum test, \( P < 0.001 \)), as shown in Figure 3A–3C. The K-M survival curve analysis combined with log-rank tests indicated a significantly shorter survival time for patients in the high-risk group. According to the K-M curves, patients with high-risk conditions had a significantly lower overall survival than those with low-risk conditions (K-M Log-Rank \( P = 8.29e-7 \)) (Figure 3D). Area under curve values for immune-related risk signals at 1, 3, and 5 years were 0.794, 0.845, and 0.831, respectively (Figure 3E). Overall, our risk factor model was able to accurately predict the overall survival of HBV-infected HCC patients.

![Volcano and heat maps of differential and immune-related differential genes in the TCGA dataset comparing HBV-infected liver cancer tissue with normal tissue.](https://doi.org/10.53388/MDM202407005)

Submit a manuscript: https://www.tmrjournals.com/mdm
Figure 2 Prognostic IRG identification and expression levels. (A & B) Prognostic IRGs identified by LASSO regression analysis. (C) Immune-related gene biological processes. (D) Immune-related gene Kyoto Encyclopedia of Genes and Genomes pathway. IRG, immune-related gene; LASSO, Least Absolute Shrinkage and Selection Operator; GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function.

Figure 3 An analysis of immune-related traits in HCC using time-dependent ROC analysis, risk score analysis, and K-M analysis. (A) Evaluation of the risk score model for the construction of characteristic immune genes in the TCGA database. (B) Scatterplot of the risk score model for the construction of signature immune genes in the TCGA database. (C) Heat map of mRNA expression. (D) K-M curves. (E) The TCGA cohort test set was analyzed using time-dependent ROC analysis. HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas; K-M, Kaplan-Meier; ROC, receiver operating characteristic; CI, confidence interval; AUC, area under curve.

Submit a manuscript: https://www.tmrjournals.com/mdm
Clinical relevance of immune-related risk profiles

The purpose of this study was to investigate whether our characteristics could predict the development of HCC infected with HBV. The relationship between our characteristics and clinical parameters and risk factors (risk score and five risk genes) was analyzed (Figure 4A, 4B). We then created a new column line plot with three independent prognostic factors to predict overall survival at 1, 3, and 5 years for the 114 HBV-infected HCC patients in the TCGA cohort (Figure 4C). A column line graph was used to divide patients into high- and low-point groups. The median value is used as the cut-off value. The Area under curves for the 1-year, 3-year, and 5-year overall survival predictions for the column line graphs were 0.75, 0.793, and 0.833, respectively (Figure 4D). According to the column line graph, the C-index was 0.75 (95% confidence interval: 0.664-1). In patients with HCC, the column line plot performed well in predicting OS, and greatly improved the clinical evaluation of patients HCC relative to traditional tumor-node-metastasis staging and immune-related markers. Columnar maps also reduce the cost of misdiagnosis compared to tumor-node-metastasis staging. Therefore, the use of our line chart has proven to be of survival prediction adds further benefit, which may enhance the application of clinical management of HCC patients.

Immunne analysis

We found a strong correlation between most immune cells, especially Treg, and neutrophils, suggesting that immune cells play an important role in the tumor microenvironment of HBV-infected HCC (Figure 5A, 5B). A risk score distribution was also examined for individual immune genes and immune cells (Figure 5C). HBV-associated liver cancer tissues showed a significant correlation between immune cells and risk scores, as a result, high-risk individuals were significantly more likely to have these immune cells infiltrated, showing a correlation between immune cell infiltration. As a result, our features were well able to predict the immune status of liver cancer in both high- and low-risk groups.

Discussion

Tumor initiation and progression are largely influenced by immune responses in the tumor microenvironment (TME) [20, 21]. Inhibiting cancer growth is possible through up-or down-regulation of certain insulin-like growth factors by the immune system [22, 23]. However, due to the immune evasion properties of tumors, some tumor cells can mimic the IRG expression pattern of normal cells to evade the immune system [24, 25]. Thus, IRG expression may be an important predictor of liver cancer progression and prognosis. In this study, an immune-related risk signal consisting of 7 IRGs was established and validated using the TCGA database, which predicted OS and correlated significantly with clinicopathological data. In addition, we created column line plots based on clinicopathological factors to improve accuracy of HBV-related HCC prognosis prediction. The abnormal expression of immune-related risk genes was associated with the development of risk scores based on the seven risk genes model studied in this study. Additionally, we found that enhanced immune phenotypes were associated with the signature as well as mutation status and immune status. Based on these results, we can infer that our signature may be valuable for predicting the outcome of patients infected with HBV who develop HCC, as well as identifying potential new immunotherapy targets.

Figure 4 Univariate and multivariate Cox regression analysis with comparative analysis of time-dependent column line graphs between characteristics and all clinical factors. (A & B) Univariate and multivariate Cox regression analysis of immune-related genes in HBV-infected HCC. (C) Prognostic line graphs predicting 1-, 3- and 5-year overall survival in patients with HCC. (D) 1-, 3- and 5-year calibration plots for internal validation. HCC, hepatocellular carcinoma; HBV, hepatitis B virus; CI, confidence interval.
Figure 5 Analysis of immunogenetic risk factors and tumor immune microenvironment. (A) Heatmap of immune cell scores, where different colors represent expression trends in different samples. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$), and the asterisks represent the level of significance ($P$). Significance of two samples by Wilcoxon test, three and more samples by Kruskal-Wallis test. (B) The number of tumor-infiltrating immune cells in each sample, represented by different colors, horizontal coordinates are samples, and vertical coordinates are tumors. (C) Heat map of correlations between multiple genes or models and immune scores, both horizontal and vertical coordinates represent genes, where different colors represent correlation coefficients, with darker colors representing stronger correlations between the two, *$P < 0.05$, **$P < 0.01$, and (P) represent significance.
Tumor-associated or tumor-induced inflammation can be promoted by chemokines and cytokines released by tumor cells and other cellular components of the TME. An important component of the tumor microenvironment that drives tumor progression [26, 27]. Through their receptors on tumor cells, chemokines and cytokines activate the NF-kB family of potentially oncogenic transcription factors. In addition to inducing survival factors, they also promote tumor growth [28]. Accordingly, HCC patients with high-risk HBV infections may have experienced increased inflammation in the tumor microenvironment, which would promote the progression of the cancer process and lead to poor OS [29]. According to the univariate Cox regression analysis of the seven IRGs composing the signature, RBP2 was the most significant gene; in the cell cycle regulatory network, RBP2 is a core protein whose expression activity directly affects the process; in addition, overexpression of RBP2 is closely associated with cancer [30].

The column line chart was developed by combining risk factors with conventional clinical factors, such as TMN staging and neoplasm development. Line charts such as this may be valuable diagnostic and prognostic tools for patients with HBV-infected HCC. Our goal was to determine whether risk factors have any clinical utility, so we assessed the correlation between risk factors and clinical factors. There was a significant association between risk scores and seven risk genes and the progression of liver cancer caused by HBV. Moreover, we found that different IRG expression can lead to different states of TME, which further impacts prognoses of different clinical subgroups. In HBV infection, our risk factors demonstrate excellent clinical applicability in predicting the progression of HCC.

We also examined whether our risk factors were associated with tumor mutational burden. Tumor mutational burden was higher among high-risk individuals than low-risk individuals in our risk score, which indicates that our risk score is both accurate and predictive. Additionally, the number of immune cells infiltrating liver cancer tissue increased with an increasing risk score. A positive correlation was found between risk scores and neutrophils. Our phenotypes are also strongly immune-related, further indicating our close relationship to the immune system.

The first study of its kind to systematically establish and validate a polygenic immune-related risk profile for HBV-infected HCC patients that can be used independently for prognosis. Based on the TCGA database cohort, we identified key prognostic IRGs and constructed features using multiple algorithms. Our study focused on immune cell infiltration as well as immune-related functions that reflect the strength of the immune response in the microenvironment of HBV-infected HCC. We may be able to use our scores for clinical management and customized immunotherapy. Our study, however, has several limitations. First, we did not conduct a prospective clinical trial to validate the profile. Furthermore, further in vivo and in vitro experimental studies should be conducted.

Conclusion

Our study identified immune-related risk profiles and developed combined features and clinical parameters to predict OS. There is a strong association between the risk score and tumor invasion, progression, and prognosis, as well as local immune status. Clinical management and medical decision-making might be improved with the help of that predicts OS in cancer.

References


Submit a manuscript: https://www.tmrjournals.com/mdm