Construction of a Prognostic Model for Hepatocellular Carcinoma Based on Immunoautophagy-Related Genes and Tumor Microenvironment

Zhen Sun¹,²,*
Zhenhua Lu¹,*
Rui Li¹,²
Weiwei Shao¹,²
Yangyang Zheng¹,²
Xiaolei Shi¹
Yao Li¹
Jinghai Song¹,²

¹Department of General Surgery, Department of Hepato-Bilio-Pancreatic Surgery, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, 100730, People’s Republic of China; ²Graduate School of Peking Union Medical College, Beijing, 100730, People’s Republic of China

*These authors contributed equally to this work

Background: The aim of this study was to screen and identify immunoautophagy-related genes (IARGs) in HCC patients and clarify their potential prognostic value in HCC patients.

Methods: Immune-related genes and autophagy-related gene were downloaded from public databases. Cox regression analysis was used to selected several immunoautophagy-related genes to establish a prognostic model, and patients were divided into high- and low-risk groups based on median risk score. We analyzed the overall survival and clinicopathological characteristics between two groups. Meanwhile, internal validation dataset and external ICGC dataset were used to verify robustness of the model. Associations between six immune cells infiltrates and risk score were analyzed.

Results: A prognostic model was established based on CANX and HDAC1. The prognoses of the high-risk group were worse than low-risk group in both TCGA and ICGC datasets. Multivariate Cox regression analysis showed that risk score was an independent prognostic factor for HCC patients. Results showed that the risk score in young group was higher than elderly group. Patients with poorly differentiated tumor may have high risk score and poor survival. The score was positively correlated with immune cells.

Conclusion: Our study shows that immunoautophagy-related genes have potential prognostic value for patients with HCC and may provide new information and direction for targeted therapy.

Keywords: hepatocellular carcinoma, immune-related genes, autophagy-related gene, overall survival

Introduction

Hepatocellular carcinoma (HCC) is the second deadliest cancer worldwide, due to its high incidence and poor prognosis. As an immune organ, liver is associated with a variety of immune cells and receives blood both the hepatic artery and portal vein. The innate and adaptive immune system play a key role in carcinogenesis of HCC by supporting tumor growth, survival, angiogenesis and motility.¹ The immune cells within the tumor microenvironment (TME) play important roles in tumorigenesis. Various immune cells might function as a tumor inhibitor or promoter in HCC.² The antigenicity of tumor contributes to the attraction of tumor-infiltrating lymphocytes, and immune cells also shape the antigenicity of tumors.³ Given its high vascularization and immunogenicity, immune therapies such as anti-CTLA-4...
(cytotoxic T lymphocyte antigen-4), anti-PD-1 (programmed cell death-1), and anti-PD-L1 (programmed death-ligand 1) strategies have shown efficacy in HCC.\textsuperscript{4} While some immune checkpoint inhibitor including pembrolizumab, nivolumab failed to show expected efficacy in clinical trial.\textsuperscript{5} Thus, HCC might have a complex immune status, and more studies are required to understand its underlying mechanisms.

Autophagy, or cellular “self-eating” is a highly conserved lysosomal pathway that involves in regulation of macromolecules’ recycling and maintains homeostasis and survival.\textsuperscript{6} Besides, it is also involved in preventing certain types of disease, including cancer, muscular disorders, and neurodegeneration, such as Huntington’s, Alzheimer’s, and Parkinson’s diseases.\textsuperscript{7} It serves as a double-edged sword in tumorigenesis and metastasis. Therefore, an optimal combination of autophagy inhibition and promotion, according to the properties of the cancer, is needed. Autophagy can be involved in innate and adaptive immune tolerance at multiple levels.\textsuperscript{8} Studies found autophagy could inhibit cancer development by orchestrating inflammation and immunity.\textsuperscript{9} And autophagy also plays an important role in the pathogenesis of HCC. Autophagy levels in HCC tumor tissues are noticeably higher adjacent normal tissues. What’s more, autophagy could promote HCC cell proliferation and invasion.\textsuperscript{10}

Thus, the combination of autophagy and immunotherapy may provide us with a promising strategy for tumor treatment, requires further study. However, few previous studies have established some prognosis model of HCC based on immune-related genes\textsuperscript{11,12} or autophagy-related genes\textsuperscript{13,14} but no studies have explored the relationship between immunoautophagy-related genes and investigate its prognosis of HCC. With the currently available autophagy-related genes (ARGs) and immune-related genes (IRGs), and the accumulative data deposited in public databases, it is hypothesized that a prognostic signature based on the immune response and autophagy might help to identify HCC patients’ survival profile. This study aims to establish a risk prognosis model based on immunoautophagy-related genes (IARGs) in HCC so as to provide a new target for future anti-cancer therapy.

**Materials and Methods**

**Data Collection**

The RNA-seq expression data and clinical data of HCC patient samples were downloaded from the TCGA data portal (TCGA-LIHC cohort). For validation, the gene expression data and the corresponding clinical data of LIRI-JP cohort were downloaded from the ICGC data portal. All databases are open-access and the present study followed the data access policy and publishing guidelines of these databases. There was no need for ethics approval. What’s more, we downloaded the immune-related genes (IRGs) from the ImmPort database (https://www.immport.org/shared/home)\textsuperscript{15} and autophagy-related genes from the Human Autophagy Database (HADb, http://www.autophagy.lu/index.html).\textsuperscript{16}

**Identification of Differentially Expressed Gene (DEGs)**

The “limma” package in R was used to normalize the data and Pearson correlation was used to calculate the correlation between the expression of DEGs in HCC patients.\textsuperscript{17,18} Differentially expressed genes of HCC and normal liver tissues with an absolute log2 fold change (FC) >1 and an adjusted p value < 0.05 were considered for subsequent analysis.\textsuperscript{19} And they were visualized by “pheatmap”, “ggplot2” and “ggpubr” package.

**Functional Enrichment Analyses**

DEGs were subjected to Gene Ontology (GO) term (biological processes, molecular function, and cellular component) enrichment and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis using the “clusterprofiler” package in R.\textsuperscript{20} Significance was defined as P < 0.05, and data were graphically visualized by “GOplot” package.

**Development and Validation of the Prognostic Signature**

The DEGs expression profile and clinical datum were incorporated into the complete dataset and further randomly divided into training dataset and test dataset, using “caret” package.\textsuperscript{21} The survival-related immunoautophagy related genes decided by univariate Cox analysis and Kaplan–Meier (K-M) analysis in training set were selected as candidate prognostic genes. Then multivariate Cox regression analysis was used to establish an optimal prognostic signature.\textsuperscript{19} The risk score was calculated for each tumor sample based on the relative expression of each IARG and its correlation coefficient, using the following formula: Risk Score=$\sum_{i=1}^{n}$ coef\textsuperscript{i} * Expr\textsuperscript{i}, where coef\textsuperscript{i} is the multivariate Cox coefficient of IARGs, and Expr\textsuperscript{i} is the relative expression of the IARGs\textsubscript{i}.\textsuperscript{19}
Patients in TCGA training set, test set and ICGC dataset were divided into low- and high-risk groups based on the median value of risk score in the TCGA training set. Kaplan–Meier survival analysis was carried out, and the prognostic value of the prediction model was evaluated based on the area under the time-dependent receiver operating characteristic (ROC) curve (AUC) using the “survivalROC” package and concordance index (c-index) by “Hmisc” package.

Independent of the Prognostic Signature
Univariate and multivariate Cox Regression analyses were applied to prove the independent nature of the signature with clinical characteristics such as age, sex, tumor grade and pathological TNM stage in training dataset. The 95% confidence intervals (CIs) and the hazard ratios (HRs) were calculated and forest plots were constructed to display the relation between risk score and different clinical factor. A p-value < 0.05 was considered statistically significant.

Correlation Between the Signature and Clinicopathological Characteristics
The correlation between clinicopathological characteristics and the prognostic signature were analyzed. T-test or Kruskal–Wallis test was used to analyze statistical significance. Statistical significance was defined as p < 0.05.

Immune Infiltrates Correlation Analysis
Associations between six immune cells infiltrates and risk score were analyzed and visualized using R. We downloaded the abundances of immune cells in HCC patients from the Tumor Immune Estimation Resource (TIMER) online database (timer.cistrome.org). The correlation index cor and the corresponding p-value were plotted.

Results
Identification of Differentially Expressed IARGs
Figure 1A showed our article structure. RNA-seq and clinical data of 374 HCC tissue samples and 50 non-tumor samples were downloaded from TCGA. We identified 7647 DEGs, including 11 IARGs (Figure 1B and C). In addition, the expression patterns of 11 differentially expressed IARG-genomes in HCC and non-tumor tissues were shown in the box diagram (Figure 1D). From the box diagram, 9 up-regulated genes (CANX, HSPA5, HSP90AB1, IKBKE, MAPK3, HDAC1, BIRC5, NRG2, CASP3) and 2 down-regulated genes (FOS, NRG1) could be directly observed. The IARGs were mostly enriched for GO terms related to positive regulation of protein kinase B signaling and ERBB2 signaling pathway. IL-17 signaling and Hepatitis B were the most frequently identified KEGG pathway (Figure 2).

Construction of Prognostic Signature with Two Genes
The expression data and clinical features of those IARGs were further integrated into a complete dataset and patient samples were randomly divided into training (75%, n=278) and test (25%, n=92) sets (4 patients has repetitive samples were excluded in the subsequent analysis).

Univariate Cox regression analysis and K-M analysis were performed on the data from the training set to investigate the correlation between differentially expressed IARGs and OS in patients with HCC. It was found that 7 genes were significantly correlated with OS in patients with HCC when p < 0.05. Then, multivariate Cox regression analysis was performed to construct an optimal prognostic signature. The final prognostic signature consisted of two genes. The predictive model was based on the linear combination of the expression levels of the two genes weighted by the relative coefficients in multiple Cox regression, as follows: risk score = (0.00325 × expression level of CANX) + (0.03955 × expression level of HDAC1). Both genes have high-risk characteristics since genes showed a positive coefficient in Cox regression analysis. Thus, their elevated expression is associated with shorter OS.

In the training set, we were divided into high expression group and low expression group by the median expression of each gene, and the K-M survival curve was plotted (Figure 3A and B). K-M analysis showed that the expression of CANX and HDAC1 was closely related to OS in HCC patients (p =0.014, p =0.011). In addition, we also searched the Oncomine database and found that the mRNA expression level of CANX in HCC tissues were significantly higher than those in normal tissues (Figure 3C and D), while the difference of HDAC1 expression level was not significant. But OS of patients with elevated expression of CANX and HDAC1 were significantly lower than that of patients with low expression.
Figure 1 (A) Study workflow of our analysis; (B) expression heatmap of differentially expressed IARGs in TCGA dataset. (C) Volcano plots of the differentially expressed IARGs (red and green nodes represent gene expression upregulation and downregulation). (D) Visualization of 11 differentially expressed IARGs in a box diagram.
Validation of Prognostic Role of Gene Signature

According to the signature we obtained, patients in the training set were divided into high- and low-risk groups according to the median value of risk score, and we visualized the number of patients, survival, and heatmap of the two gene expression profiles in different risk groups in the training set (Figure 4). The K-M curve we draw indicating significant differences ($p < 0.008$, Figure 5A). ROC curve analysis showed that the 1-year, 2-year, 3-year, and 5-year AUC of our signature were 0.696, 0.654, 0.639, and 0.642 (Figure 5D), indicating the good prognostic efficacy of the model. In the meanwhile, we used internal dataset (test set) and external dataset (ICGC dataset) to evaluate the predictive value of the prognosis signature (Figure 5B and C). The 1-year, 2-year, 3-year, and 5-year ROC in TCGA test set (Figure 5E), ICGC dataset (Figure 5F) were 0.728, 0.651, 0.685 and 0.612; 0.757, 0.681, 0.669 and 0.644. Corresponding with the results of the training set, high-risk group HCC patients in the test set and the ICGC dataset were related to poor OS ($p =0.048$, $p < 0.001$). The c-index of TCGA training set, test set and ICGC dataset were 0.602, 0.622 and 0.640, respectively.

Clinical Utility of the Prognostic Signature

Univariate and multivariate Cox regression analyses were performed on 203 HCC patients with complete clinical data in the training set to evaluate the independent predictive value of the relative clinical data and the prognostic signature. Univariate Cox regression analysis showed that age, clinical stage, tumor size and risk score had certain prognostic value. Multivariate Cox regression analysis showed that risk score was an independent prognostic
factor for OS (HR = 1.752, 95% CI, 1.358–2.260, \( p < 0.001 \)) (Figure 6).

We also discussed the relationship between clinicopathological characters and the prognostic model. After divided patients into the young group (≤65 years) and the elderly group (>65 years), the results showed that the risk score in young group was higher than those in elderly group (\( p = 0.010 \), Figure 7A). And the score of patients with higher pathological grade was higher than that of patients with lower pathological grade (\( p = 0.003 \), Figure 7B). This indicated that patients with poorly differentiated tumors have poor prognosis. However, no difference was found between risk score and gender (\( p = 0.104 \), Figure 7C), or TNM stage (\( p = 0.052 \), Figure 7D).

**Validation of the Immune Correlation**

We analyzed the relationship between the signature and immune cell infiltration to determine whether it can accurately reflect the immune microenvironment of HCC. The signature was positively correlated with B cells, CD4 T cells, neutrophils, macrophages, and dendritic cells (Figure 8A–F).
Discussion
The occurrence of liver cancer is related to genetic, environment, lifestyle and other factors, which makes diagnosis and prevention of liver cancer become difficult. For the treatment of liver cancer, the study of the pathogenesis has become more and more important. The occurrence and development of HCC involve many aspects such as proliferation, apoptosis, autophagy and invasion. Disorders of the immune system lead to the escape of the tumor from the immune and promote tumor pathogenesis. HCC is a highly inflammation-related tumor that develops along with inflammation. Immunotherapy has appeared as an attractive option for improving outcome for advanced hepatocellular carcinoma, including Anti-CTLA-4, Anti-PD-1, Anti-PD-L1, Cytotoxic agents or combinations of these therapies.4

Figure 4 Rank of prognostic index and distribution of groups (A), survival status of patients in different groups (B), and expression heatmap of the two genes included (C).
Currently, autophagy is thought to play a dual role in HCC. During tumor initiation, autophagy inhibits the transformation of normal cells to tumor cells by removing damaged proteins, DNA and necrotic organelles. While in the advanced stages of cancer, autophagy provides a large amount of nutrients and energy for tumor cells, which enables tumor cells to withstand the harsh tumor microenvironment, resist necrosis, inflammation and pressure of chemotherapy or radiotherapy, and promote tumor cells survive and grow. It may promote tumor metastasis and invasion. Thus, within the same cancer, both inhibition and promotion of autophagy may be beneficial. It also play tumor-suppression function by inhibiting the “inflammation-carcinogenesis” pathway in liver. Autophagy can regulate inflammation and inflammatory cytokines play a different role in autophagy as well. Since the intricate relationship between autophagy and immunity, investigating the mechanism of autophagy interfere with the function of inflammation in HCC may provide a new target for immunotherapy of liver cancer.

Wang et al built a nine-gene prognostic model that showed a good performance for HCC prognosis prediction. Another study developed an eight-gene model based on the immune expression profile. In the meanwhile, several studies also constructed prognostic signatures for HCC overall survival prediction based on

Figure 5 KM survival analysis of high- and low-risk groups in training set (A), test set (B), ICGC dataset (C); the 1, 2, 3, 5-year ROC in TCGA training set (D), test set (E), ICGC dataset (F).

Figure 6 Univariate (A) and multivariate (B) Cox regression analyses demonstrated that the prognostic model was independently associated with the OS of HCC patients.
autophagy-associated genes. These prognostic models could serve as predictors for survival and play key roles in tumor microenvironment. In order to identify potential biomarkers, we used the TCGA database to analyze differences in immunoautophagy-related gene expression between HCC and normal tissues in our study. We first screened out 11 differentially expressed IARGs. Considering that these genes may be closely related to the development and growth of HCC, we performed GO and KEGG enrichment analyses. The results showed that the 11 DEGs were mostly enriched for GO terms related to positive regulation of protein kinase B signaling and ERBB2 signaling pathway. KEGG enrichment showed that genes were particularly enriched in the IL-17 signaling and hepatitis B pathways. As the member of human epidermal growth factor receptor (EGFR) family, Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2) is subjected to an additional layer of regulation mediated by the molecular chaperone HSP90. The ERBB receptors could engage different downstream signaling modules, such as Ras/Raf/MAPK and phosphatidylinositol 3-kinase (PI3K)/AKT pathways. The mitogen-activated protein kinases (MAPK) signaling pathway was related to the autophagy pathway in cancer, which provided insights for a potential therapeutic strategy. Song et al found that activated the MAPK signaling pathway could block the development of liver fibrosis by enhancing cell apoptosis and reducing autophagy. Most HCCs (80–90%) develop on underlying chronic liver disease, the main causes include chronic hepatitis B virus (HBV) or hepatitis C virus (HCV) infections. In South-East Asia and Central Africa, the endemic prevalence of chronic HBV infections accounts for 70% of HCC. IL-17 signaling pathway as an inflammation-related pathway, the regulatory potential of it makes it a compelling target in cancer immunotherapy.

We identified 7 survival related IARGs through univariate cox regression analysis and KM analysis. Then, we used multivariate cox regression analysis to determine two key IARGs: CANX and HDAC1, and established...
a prognosis model. Overexpression of CANX and HDAC1 was also found in HCC samples and elevated expression of both genes was associated with poor prognosis in HCC patients. Based on the risk score of our model, HCC patients in different datasets were divided into high-risk and low-risk groups, respectively. High-risk group patients were associated with worse prognosis. Our study demonstrated that risk score of the prognostic model was an independent prognostic factor for HCC patients. Doctor can adjust the treatment plan of patients according to the model. For patients at high risk, more aggressive treatment strategies are recommended. Further analysis found that age and pathological grade were correlated with risk score. Patients with poorly differentiated tumor may have high risk score and poor survival according this model. Abnormal differentiation is the main characteristic of tumor cells. Poorer differentiation status indicates worse prognosis of HCC patients, which is corresponded with our result.

Calnexin (CANX), an essential endoplasmic reticulum (ER) chaperone protein that prevents the aggregation and export of incompletely folded proteins from the ER, is involved in the metastatic progression of tumors. CANX was revealed to inhibit the proliferation and activation of CD4+T and CD8+T cells, and may impair the anti-tumor immunity of CD4+ and CD8+ T Cells by upregulating the expression of PD-1 in oral squamous cancer. Okayama et al reported that CANX played an important role in tumor invasion and metastasis, indicating poor prognosis of lung adenocarcinoma patient. Kobayashi et al subsequently suggested CANX expression could detect early lung cancer as its level was significantly higher in lung cancer patients. Ryan et al found depletion of Calnexin gene reduced tumor cell survival and increased colorectal cancer cells susceptibility to 5-FU chemotherapy. Extracellular matrix (ECM) is a physical barrier to the growth of HCC. Tumor growth and invasiveness could be induced by ECM degradation. Ros et al revealed that

Figure 8 Relationships between the signature and immune cell infiltration. Correlations were determined by Pearson correlation analysis. (A) B cells; (B) CD4 T cells; (C) neutrophil; (D) dendritic; (E) macrophages; (F) CD8 T cells.
CANX complexes in cell surface may reduce extracellular disulfide bonds and are essential for ECM degradation. And they found anti-CANX antibody could prevent lung metastasis from endogenous liver tumours. This indicates that CANX is a potential target for HCC treatment.

Histone deacetylase 1 (HDAC1) can antagonize the acetylation of histones proteins. Studies shows that HDAC1 is over-expressed in many tumors including lung carcinoma, glioma, renal cell cancer, gastric cancer, prostate cancer and colorectal cancer, and is associated with prognostic outcomes of patient. HDAC inhibitors could induce cell-cycle arrest, promote differentiation, sensitize cancer cells to DNA damaging therapies and stimulate tumor cell death. Consisting with this discovery, Zhang found that downregulation of HDAC1 inhibited cell proliferation, prevented cell migration, decreased cell invasion, reduced tumor angiogenesis and induced cell apoptosis in non-small cell lung cancer cells. Yoo et al discovered a positive cross talk between HBx and the MTA1/HDAC 1/2 complex in stabilizing HIF-1 alpha, which may play a vital role in angiogenesis and metastasis of HBV-associated HCC. Ler et al found that mortality was also increased with high HDAC1 expression in hepatocellular carcinoma among a South East Asian population. While the roles and molecular mechanisms of HDAC1 in HCC is still limited. What’s more, further studies should be performed to reveal the relationship between CANX and HDAC1 in HCC.

Tumors are evolving under the pressure of TME, and interaction between tumor and the immune system plays crucial parts. To characterize the tumor immune microenvironment status, the relationships between immune-related prognostic signature and immune cell infiltration were investigated. Our study found that B-cells, CD4T-cells, dendritic cells, neutrophil and macrophage displayed positive correlation with IARGs prognostic model, revealing that the model may serve as predictor for increased immune cells infiltration. It may be a promising way to anti-HCC by better understanding of the role of immune cells. While the clinical relevance of HCC neoantigens and their interactions with immune microenvironment still remains unknown.

However, this study also has some limitations. This study is a retrospective study with data from the TCGA and ICGC databases, so there may be selection bias. Thus, a further well-designed prospective analysis is necessary to validate the value of the novel model. And researches are needed to reveal the function of the model gene and its exact molecular mechanism.

Conclusion
Based on bioinformatics analysis of large-scale data, we identified differentially expressed IARGs in HCC, and we developed a prognostic model. This model could be used for prognostic prediction and the selection of patients for immunotherapies and targeted therapies.

Disclosure
The authors declare no competing interests.

References


