Comprehensive Analysis of Autophagy-related Genes in Hepatocellular Carcinoma

Li Wang
Qingdao University

Jialin Qu
Qingdao University

Na Zhou
Qingdao University

Man Jiang
Qingdao University

Xiaochun Zhang (✉ zhangxiaochun9671@126.com)
Nanchang University https://orcid.org/0000-0003-3297-2093

Research article

Keywords: Survival model, Nomogram, Autophagy-related genes, HCC, TCGA database, ICGC database

DOI: https://doi.org/10.21203/rs.3.rs-139968/v1

License: ☝️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: Hepatocellular carcinoma (HCC) is the common type of cause of cancer-related death among human cancers. There are ample evidences to showing that autophagy-related genes (ARGs) may play a significant role in the biological process of HCC.

Methods: In this study, we aim to identify survival model and nomogram that could effectively predict the prognosis of HCC based on ARGs. First, we download the data of HCC patients from TCGA database. Second, we analysis the function of ARGs by utilized GO and the KEGG analysis. Finally, we screen 5 ARGs (SQSTM1, CAPN10, EIF2S1, ATIC, RHEB) for survival model by performed the Cox regression and Lasso regression analysis. We further built and verified a prognostic nomogram base on prognostic ARGs. Moreover, its efficacy was validated by the ICGC database. The expressions level of 5 ARGs was performed using Oncomine database, the Human Protein Atlas and Kaplan-Meier plotter.

Result: We found patients the survival of patients in the different groups was significantly different both in the TCGA cohort and ICGC cohort. The survival model showed good performance for predicting the prognosis of HCC patients by having a better area under the receiver operating characteristic curve than other clinicopathological parameters.

Conclusion: our survival models and prognostic ARGs nomogram can be independent risk factors for hepatocellular carcinoma patients.

Background

Hepatocellular carcinoma (HCC) is a common type of cause of cancer-related death among human cancers [1]. Cancer of the liver and intrahepatic bile ducts was responsible for an estimated 42,220 new cases and 30,200 deaths in the year 2018 alone in the United States[2]. And it’s dismal prognosis with overall 1- and 3- year survival rates of only 36% and 17%[3]. Although the great progress has been obtained in chemotherapy, surgical resection, radiofrequency ablation, systemic therapy, and liver transplantation, the prognosis of Hepatocellular carcinoma still poor due to most of patients at an advanced stage [4]. Hence, it is essential to explore alternative biomarker to predictive the diagnosis and outcomes of HCC.

Autophagy, a vital catabolic process of transporting damaged, denatured or senescent proteins and organelles into lysosomes for digestion and degradation[5]. In normal cells, autophagy is conducive to cells to maintain a stable state and inhibit cell canceration. However, autophagy plays a important role in tumor cell that digestion and degradation nonfunctional organelles to meet the demands of tumor development[6–8]. Previous study had showed that autophagy was important in many pathophysiological processes including neurodegenerative diseases, inflammation, immune responses, tumorigenesis and cancer development[9, 10]. Moreover, a study reported that autophagy plays an significant role in tumor suppression, and the loss of autophagy was essential to tumorigenesis[11]. Autophagy has been described as one of the main pathways for liver health and disease. Some studies
showed that autophagy is related to several liver diseases, including fatty liver disease and HCC [12, 13]. Liver cancer formation has been observed in a number of autophagy mice models [14].

In order to explore the relationship between autophagy and the prognosis of HCC and whether autophagy-related genes can be used as a potential biomarker for prognosis of HCC and guiding clinical medication. We used gene expression data that acquired from TCGA database to develop ARGs expression signature and establish a survival model and nomogram as independent index for HCC patients.

Materials And Methods

Data source and pre-processing

Total of 232 autophagy-related genes was download from The Human Autophagy Database. RNA microarray data that consisting of 374 HCC tissues and 50 normal tissues was obtained from TCGA database. The expression of ARGs comprising 374 HCC tissues and 50 normal tissue were processed by the mean function, and normalized the mean expression level by log2 transformation. We analysis the differential expression of ARGs between HCC and normal tissues by limma package, with thresholds of false discovery rate < 0.05 and |log2 fold change | > 1.

Functional and pathway enrichment analysis

The Gene ontology database mainly includes three parts biological process (BP), cellular component (CC), and molecular function (MF) [15]. The Kyoto Encyclopedia of Genes and Genomes database, an integrated database, resource for biological interpretation of genome sequences [16]. In this study, GO and KEGG enrichment analysis performed by the clusterProfiler package [17].

Construction of survival model

We obtain the significantly related to survival ARGs by using univariate Cox regression analysis. Next, we remove highly correlated survival-related ARGs by employed the Lasso regression analysis [18].

We identify the prognostic ARGs and their coefficients by multivariate Cox regression analysis, on which we constructed the survival model. We calculated the risk score of each HCC patients according to the survival model. The calculation of the risk score of survival model was conducted as follows:

\[
\text{Risk score} = \sum_{i=1}^{n} v_i \times c_i
\]

(the \(v_i\) is the means expression of gene, \(c_i\) means the regression coefficient of gene).
The Kaplan–Meier survival curve was used to evaluate the high- and low-risk groups according to survival model by the log-rank test. Moreover, we determine the accuracy of the survival model by receiver operating characteristic (ROC) curves.

We predict survival model can be used as an independent prognostic factor for HCC patients by independent prognostic analysis. In addition, we verify the relationship between ARGs and clinicopathological parameters by performed clinical correlation analysis.

**Evaluation of Tumor-infiltrating immune cells**

CIBERSORT, an analytical tool, that estimates infiltrating immune composition of a tumor biopsy using gene expression data. In this study, we quantify the fractions of immune cells in the HCC samples by CIBERSORT method. The results of the inferred fractions of immune cell populations produced by CIBERSORT were considered accurate with a threshold of P< 0.05[19].

**Survival model validation by independent cohort**

Our survival model external validation by the International Cancer Genome Consortium (ICGC) database by the same formula. HCC Patients data that consisting of 116 Japanese HCC patient and adjacent normal tissues was downloaded from ICGC database. The analyses process of ICGC cohort was same as the TCGA cohort.

**ARGs validation by Oncomine database and The Human Protein Atlas, and Kaplan-Meier plotter**

The Oncomine database, a cancer microarray database and web-based data-mining platform, for the purpose of mining cancer gene information [20].The Human Pathology Atlas allowed for generation of personalized genome-scale metabolic models for cancer patients to identify key genes involved in tumor growth. In this study, we utilized Oncomine database and The Human Protein Atlas to explore the protein expression of the ARGs in HCC tissues and normal tissues. The prognostic value of the RBPs in HCC was verified by the Kaplan-Meier plotter online tool.

**Establish ARGs nomogram**

The ARGs nomogram was build based on the TCGA cohort. The predictive accuracy and discriminative value of the nomogram were mainly including concordance index (C-index), AUC and calibration curve[21]. In addition, the predictive accuracy and discriminative value of the ARGs nomogram were validated using ICGC cohort.
Statistical analysis

In this work, statistical analyses were performed by Perl language and R language. Lasso regression analysis and Cox regression analyses were used to screen the ARGs that related to prognosis of HCC. And all comparisons statistical significance was defined as a P < 0.05.

Results

Differentially expressed ARGs

A total 4 downregulated ARGs (DIRAS3, FOS, FOXO1, NRG1) and 58 upregulated ARGs (ATG10, ATG16L2, ATG4B, ATG9B, ATIC, BAK1, BAX, BIRC5, CANX, CAPN10, CAPN2, CAPNS1, CASP3, CASP8, CD46, CDKN2A, CLN3, DAPK2, DDIT3, DRAM1, FKBP1A, GAPDH, HDAC1, HGS, HSP90AB1, HSPA5, HSPB8, IKBKE, IRGM, ITGA3, ITGA6, ITGB4, MAPK3, MLST8, NKX2−3, NPC1, NRG2, PARP1, PEA15, PELP1, PRKCD, RAB24, RB1CC1, RGS19, RHEB, RPTOR, RUBCN, SPHK1, SPNS1, SQSTM1, TMEM74, TP63, TP73, TSC1, TSC2, ULK3, VMP1, WDR45B) were obtained from data analysis (Figure 1 A). We visualize the expression pattern of the differentially expressed ARGs by volcano plot and box plot (Figure 1 B, C).

Functional and pathway enrichment analysis

The GO and KEGG enrichment analysis showed the differentially expressed ARGs were mainly associated with the BP terms autophagy, process utilizing autophagic mechanism, macro-autophagy, and regulation of autophagy. In addition, CC terms showed that the ARGs were associated with vacuolar membrane, membrane region, lysosomal membrane, and lytic vacuole membrane. Moreover, MF terms were mainly protein kinase regulator activity, kinase regulator activity, and ubiquitin−like protein ligase binding (Figure 2 A). We also performed KEGG pathway enrichment analysis to determine the pathways most enriched, which included Autophagy − animal, Human papillomavirus infection, and Apoptosis (Figure 2 B, C).

Identification of prognosis-related ARGs survival model

A total of 48 ARGs were identified closely related to HCC patient survival by the univariate Cox regression analysis. Then, we exclude genes that may be highly correlated with other genes by the Lasso regression analysis (Figure 3 A, B). Finally, seven survival-related ARGs were submitted to multivariate Cox proportional hazards model, generate 5 candidate ARGs (SQSTM1, ATIC, CAPN10, RHEB, EIF2S1) were identified to construct the survival model (Table 1). Our survival model was formed using the following formula: risk score = (the means expression of SQSTM1*0.002097) + (the means expression of ATIC * 0.021484) + (the means expression of CAPN10 * 0.229976) + (the means expression of RHEB * 0.032636) + (the means expression of EIF2S1 * 0.108674).
The HCC patients divided into low- and high-risk groups based on the median of the risk score. The low-risk group had a higher survival rate than high-risk group (HR = 1.202, 95% CI = 1.141–1.268, P < 0.001) according Kaplan–Meier survival curves (Figure 4A). In addition, we assess the accuracy of the OS-related survival model by constructed the ROC curve, and the AUC of risk score was significant than other clinicopathological parameters (Figure 4B). Finally, we ranked the HCC patients by survival model to analyses the survival distribution. We could identify the mortality rate of HCC patients with their risk scores. Moreover, with the increase of risk score, the mortality rate of patients rose (Figure 4 C, D). We described the expression of ARGs between low- and high-risk groups by heat maps (Figure 4 E).

### The relationships between prognosis-related survival model, Tumor-infiltrating immune cells and clinicopathological parameters

The univariate and multivariate Cox regression analysis showed that the survival model can be used as an independent prognostic factor for HCC patients (Figure 5 A, B). The survival model was significantly related to clinicopathological features mainly including survival state, stage, and tumor T status (Figure 5 C-E).

We next evaluated the association of prognosis-related survival model and tumor-infiltrating immune cells in HCC microenvironment using CIBERSORT algorithm. The CIBERSORT analysis showed that the composition of 22 immune cell types in High-risk group and low-risk group of HCC varied significantly (Figure 6 A, B). Moreover, Dendritic cell resting were more enriched in High-risk group (Figure 6 C).

### Validation of the survival model

We calculated the risk score of each HCC patients in the ICGC data portal project Liver Cancer - RIKEN, JP (LIRI-JP) as an independent external validation by the same formula. And HCC patients divided into high- and low-risk groups based on the median of risk score. The Kaplan–Meier survival curves showed the
prognostic value of our survival model (HR = 1.215, 95% CI = 1.160–1.271, P < 0.001) (Figure 7 A). In
addition, the ROC curve also showed a good ability of the OS-related survival model to predict prognosis
of HCC patients (Figure 7 B). And with the increase of risk score, the mortality rate of patients rose
(Figure 7 C, D). Heat maps described the expression level of ARGs with the different risk scores of
samples (Figure 7 E). Therefore, these validation results confirmed the predict prognosis ability of our
survival model.

The expression of ARGs in Oncomine database, Human
Protein Atlas, and Kaplan-Meier plotter

We analyze the SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1 in liver cancer by Oncomine database. The
expression level of SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1 in different hepatocellular carcinoma was
higher than the normal group in the Wurmbach Liver (201471_s_at), Roessler Liver 2 (208758_at),
Roessler Liver (219333_s_at), Roessler Liver 2 (201453_x_at), and Roessler Liver 2 (201144_s_at) studies
(Figure 8 A-E).

In addition, we verify the histological level of SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1 by the Human
Protein Atlas database, and the results showed that SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1 is
upregulated in HCC tissues and downregulated in normal tissue (Figure 8 F-J).

The prognostic significance of SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1 was identified using the
Kaplan-Meier plotter server. The results showed that the 5 RBP s were closely related to OS of HCC
patients (Figure 8 K-O).

ARGs nomogram

We tried to establish nomograms to connect survival model with 1,2,3-year survival. We analyze the
ARGs that affecting the prognosis of HCC patients and establish an ARGs nomogram using the Cox
multivariate analysis. Ultimately, the prognosis-related ARGs nomogram included 5 prognostic ARGs
(SQSTM1, ATIC, CAPN10, RHEB, EIF2S1) (Figure 9 A). Moreover, we validation the ARGs nomogram by
ICGC cohort. The C-index for OS prediction was 0.683 (95% confidence interval (CI): 0.672-0.713) in TCGA
cohort, 0.747 (95%CI: 0.71-0.77) in ICGC cohort. The calibration curve for predicting 3-years survival and
5-years survival in the TCGA cohort (Figure 9 B, C) and at 3-years survival in the ICGC cohort (Figure 9 D).
The 3-year and 5-year survival AUC of ARGs nomogram also explained that nomogram suitable for
clinical application (Figure 9 E, F). In summary, these results showed that our ARGs nomogram had great
value for application in clinical practice.

Discussion
The carcinogenesis of hepatocellular carcinoma is an intricate regulatory network. Compared to using a single clinicopathological parameter, gathering diverse biomarkers and establishing a survival model and nomogram is an effective way to predict tumor prognosis. Nowadays, autophagy has dual roles in many cancers, including hepatocellular carcinoma. Autophagy was ubiquitous in the occurrence and development of cancer.

In this study, we aim to identify survival model and nomogram that could effectively predict the prognosis of HCC based on ARG. We obtained the RNA expression profiles of ARGs from the TCGA database. Then, we screened 4 downregulated ARGs and 58 upregulated ARGs between tumor and normal tissues. The results of GO analysis showed that the ARGs were mainly enriched in autophagy, process utilizing autophagic mechanism, vacuolar membrane, lysosomal membrane, protein kinase regulator activity, and kinase regulator activity. It can be seen that ARGs revolve around autophagy, lysosomes, and enzyme regulation. The KEGG analysis results showed that the ARGs were primarily enriched autophagy–animal, and apoptosis. After the univariate Cox regression analysis, a total of 48 survival-related ARGs that were identified significantly related to HCC survival. Finally, we determine 5 genes (SQSTM1, ATIC, CAPN10, RHEB, EIF2S1) and construct the survival model by multivariate Cox regression analysis. The survival model could be an independent prognostic biomarker for HCC patients.

The CIBERSORT analysis showed that Dendritic cell resting were more enriched in High-risk group. Dendritic cells are the critical professional APC for T cell priming in spontaneous antitumor adaptive immunity[22]. Moreover, Dendritic cells have been shown to recruit CD8 + T cells into the tumor microenvironment[23]. Some studies showed that Dendritic cells are required for spontaneous T cell-mediated tumor rejection and response to immunotherapy in a variety of cancer mouse models[24–26]. A recent study showed that dendritic cell paucity can lead to dysfunctional immune surveillance against an engineered model neoantigen, accelerating neoplastic progression in cancer mouse model [27]. These results indicated that high-risk patients may be sensitive to immunotherapy.

Nomogram, a user-friendly graphical composite model, have been developed and shown to be more accurate than the conventional staging systems for predicting prognosis in many cancers[28]. Nomogram can predict the likelihood of an event based on the patient's personal data, such as survival, and recurrence. In order to make our survival models achieve a more credible and valuable prediction power in clinical application, the prognostic ARGs nomogram including SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1, was developed for assessing the individualized survival risk of patients and demonstrate satisfactory discrimination.

Sequestosome 1 (SQSTM1), a multifunctional protein, involved in multiple cellular processes including proliferation, drug sensitivity and autophagy-associated cancer cell growth[29]. Some studies showed that SQSTM1 are related to cancer occurrence and development including liver cancer, ovarian cancer and colorectal cancer[30–32]. SQSTM1 also plays a significant role in our survival model and ARGs nomogram. Inosine monophosphate cyclohydrolase (ATIC), a bifunctional protein enzyme, catalyzes the last two steps of the de novo purine biosynthetic pathway[33]. ATIC acts as an oncogenic gene that
promotes survival, proliferation and migration by targeting AMPK-mTOR-S6 K1 signaling in hepatocellular carcinoma[33]. However, there is still lack some research to explore the specific mechanism. Calpain-10 (CAPN10), the calpain family protease, identified as the first candidate susceptibility gene for type 2 diabetes mellitus[34]. Moreno-Luna et al found that CAPN10 UCSNP-63 genotype 12 seems to be related with a worse prognosis in laryngeal cancer[35]. Some studies also showed CAPN10 are related to ovarian cancer and laryngeal cancer[35, 36]. However, the mechanism between CAPN10 and HCC still unclear. Ras homolog enriched in brain (RHEB) is a conserved small GTPase that belongs to the Ras superfamily. RHEB mainly involved in activation of cell growth through stimulation of mTORC1 activity[37, 38]. Liu et al found that the expression of RHEB is a significant prognostic factor and may be important in HCC cell growth [39]. Eukaryotic translation initiation factor 2 subunit alpha (EIF2S1), which functions as a marker of endoplasmic reticulum stress. Some study has been reported that EIF2S1 widely associated with patient prognosis in various cancers including breast cancer, gastrointestinal carcinomas, malignant melanoma [40–42]. Previous studies have reported that the expression levels of EIF2S1 was high in tumor compared with in matched noncancerous tissues[40].

In our study, our survival model and nomogram that based on autophagy-related genes for hepatocellular carcinoma patient are validation by ICGC cohort. In addition, the expressions of 5 ARGs was performed using Oncomine database, the Human Protein Atlas and Kaplan - Meier plotter server. Autophagy is a dynamic and continuous process. It is still not clear if analyzing changes in ARGs are sufficient to reflect autophagic activity. Therefore, there are some limitations in our work. First, there is no experimental studies regarding the link between expression data and functional autophagy states to verify our results. Second, lack of clinical research to confirm our results.

Conclusion

Our study develop a survival model based on autophagy-related genes for hepatocellular carcinoma patients. Moreover, we further establish a prognostic ARGs nomogram for hepatocellular carcinoma patients. Our survival model and ARGs nomogram had great value for application in clinical practice. However, these ARGs need further research explore whether they may be helpful for molecular targeted therapy of liver hepatocellular carcinoma.

Declarations

Authors’ contributions

W.L and Z.X.C contributed to the conception and design of the study; All authors participated in data download, analysis and statistical analysis, writing and editing the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Consent for publication

Not applicable

Availability of data and materials

Not applicable

Declaration of Competing Interest

The authors report no conflicts of interest in this work.

Funding

This work was supported by the Taishan Scholar foundation of Shandong Province (No tshw201502061).

Acknowledgments

The authors thank my family for helpful comments, and Zhixuan Ren and Jian Tang for technical assistance.

References


Figures

Figure 1

Differentially expressed of ARGs between HCC and normal tissues. (A) Heatmap of differentially expressed ARGs. (B) The volcano plot for the 232 ARGs from the TCGA database. Red, higher expression; green, lower expression; Black, no difference. (C)
Figure 3

The Lasso regression analysis of ARGs. (A) Screening of optimal parameter (lambda) at which the vertical lines were drawn. (B) Lasso coefficient profiles of the seven ARGs with non-zero coefficients determined by the optimal lambda.
Figure 4

Survival model of HCC patients in TCGA cohort. (A) Kaplan–Meier curve of high-risk and low-risk HCC patients. (B) ROC curve of OS-related survival model. (C) Risk score distribution of HCC patients with different risks. (D) Scatterplots of HCC patients with different survival status. (E) Heatmap of expression of 5 ARGs in HCC patients.
Figure 5

Independent prognostic analysis. (A) Univariate factor independent prognostic analysis. (B) Multivariate factor independent prognostic analysis. (C) The relationships between survival state and prognosis-related survival model. (D) The relationships between Stage and prognosis-related survival model. (E) The relationships between Tumor T status and prognosis-related survival model.
Figure 6

The relationships between prognosis-related survival model and tumor-infiltrating immune cells. (A) 22 immune cells in High-risk group. (B) 22 immune cells in low-risk group. (C) Violin plot displays the differentially infiltrated immune cells between High-risk group and low-risk group.
Figure 8

The expression level of SQSTM1, ATIC, CAPN10, RHEB, and EIF2S1 in the Oncomine database and The Human Protein Atlas. (A) The expression level of SQSTM1 in HCC and Liver. (B) The expression level of ATIC in HCC and Liver. (C) The expression level of CAPN10 in HCC and Liver. (D) The expression level of RHEB in HCC and Liver. (E) The expression level of EIF2S1 in HCC and Liver. (F) Immunohistochemistry of SQSTM1 in HCC (Staining: high; Intensity: strong; Quantity: > 75%; Location: cytoplasmic/membranous, nuclear) and in normal tissue (Staining: low; Intensity: moderate; Quantity: < 25%; Location: nuclear); (G) Immunohistochemistry of ATIC in HCC (Staining: high; Intensity: strong; Quantity: > 75%; Location: cytoplasmic/membranous) and in normal tissue (Staining: medium; Intensity: moderate; Quantity: 75%-25%; Location: cytoplasmic/membranous); (H) Immunohistochemistry of CAPN10 in HCC (Staining: high; Intensity: strong; Quantity: > 75%; Location: cytoplasmic/membranous) and in normal tissue (Staining: medium; Intensity: moderate; Quantity: 75%-25%; Location: cytoplasmic/membranous); (I) Immunohistochemistry of RHEB in HCC (Staining: medium; Intensity: moderate; Quantity: > 75%; Location: cytoplasmic/membranous) and in normal tissue (Staining: medium; Intensity: moderate; Quantity: 75%-25%; Location: cytoplasmic/membranous); (J) Immunohistochemistry of EIF2S1 in HCC (Staining: high; Intensity: strong; Quantity: > 75%; Location: cytoplasmic/membranous) and in normal tissue (Staining: medium; Intensity: moderate; Quantity: 75%-25%; Location: cytoplasmic/membranous).
Location: cytoplasmic/ membranous) and in normal tissue (Staining: medium; Intensity: moderate; Quantity: 75%-25%; Location: cytoplasmic/ membranous); (J) Immunohistochemistry of EIF2S1 in HCC (Staining: high; Intensity: strong; Quantity: > 75%; Location: cytoplasmic/ membranous) and in normal tissue (Staining: low; Intensity: moderate; Quantity: < 25%; Location: cytoplasmic/ membranous). (K) prognostic value of SQSTM1. (L) prognostic value of ATIC. (M) prognostic value of CAPN10. (N) prognostic value of RHEB. (O) prognostic value of EIF2S1.

Figure 9
The ARGs nomogram for prediction on survival probability of HCC patients. (A) Development of ARGs nomogram for predicting 1-, 3-, and 5-years OS for HCC patients. (B) The calibration curve for predicting HCC patient 3-year survival in TCGA cohort. (C) The calibration curve for predicting HCC patient 5-year survival in TCGA cohort. (D) The calibration curve for predicting HCC patient 3-year survival in ICGC cohort. X-axis: Nomogram-predicted probability of overall survival; Y-axis: actual overall survival. (E)3-year survival ROC of ARGs nomogram, AUC=0.694. (F) 5-year survival ROC of ARGs nomogram, AUC=0.692.