

## Original Article

# Development and validation of a survival model based on autophagy-associated genes for predicting prognosis of hepatocellular carcinoma

Wanli Yang<sup>1\*</sup>, Liaoran Niu<sup>1\*</sup>, Xinhui Zhao<sup>2\*</sup>, Lili Duan<sup>1\*</sup>, Yiding Li<sup>1</sup>, Xiaoqian Wang<sup>1</sup>, Yujie Zhang<sup>1</sup>, Wei Zhou<sup>1</sup>, Jinqiang Liu<sup>1</sup>, Qingchuan Zhao<sup>1</sup>, Yu Han<sup>3</sup>, Daiming Fan<sup>1</sup>, Liu Hong<sup>1</sup>

<sup>1</sup>State Key Laboratory of Cancer Biology and National Clinical Research Center for Digestive Diseases, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an 710032, Shaanxi Province, China; <sup>2</sup>Department of Thyroid and Breast Surgery, Xi'an No. 3 Hospital, The Affiliated Hospital of Northwest University, Northwest University, Xi'an 710018, Shaanxi Province, China; <sup>3</sup>Department of Otolaryngology, Xijing Hospital, Fourth Military Medical University, Xi'an 710032, Shaanxi Province, China. \*Equal contributors.

Received June 1, 2020; Accepted September 24, 2020; Epub October 15, 2020; Published October 30, 2020

**Abstract:** Objective: This study aimed to identify the novel prognostic gene signature based on autophagy-associated genes (ARGs) in hepatocellular carcinoma (HCC). Methods: The RNA sequencing data and clinical information of HCC and normal tissues were obtained from The Cancer Genome Atlas (TCGA) database. The differentially expressed ARGs were screened by the Wilcoxon signed-rank test. Cox regression analysis and Lasso regression analysis were performed to screen the ARGs and establish the prognostic prediction model. Kaplan-Meier and receiver operating characteristic (ROC) curves were both used to evaluate the accuracy of the model. GSE14520 dataset (testing cohort) was used to validate the prognostic risk model in TCGA. A clinical nomogram was established to predict the survival rate of HCC patients. Results: Totally 27 differentially expressed ARGs were identified. Three OS-related ARGs (SQSTM1, HSPB8, and BIRC5) were identified via the Cox regression and Lasso regression analyses. Based on these three ARGs, a prognostic prediction model was constructed. HCC patients with high risk score present poorer prognosis than those with low risk score both in TCGA cohort ( $P=4.478e-04$ ) and testing cohort ( $P=1.274e-03$ ). Moreover, the risk score curve shows a well feasibility in predicting the patients' survival both in TCGA and GEO cohort with the area under the ROC curve (AUC) of 0.756 and 0.672, respectively. Besides, the calibration curves and C-index indicated that the clinical nomogram performs well to predict survival rate in HCC patients. Conclusions: The survival model based on the ARGs may be a promising tool to predict the prognosis in HCC patients.

**Keywords:** Hepatocellular carcinoma, autophagy-associated genes, the cancer genome atlas, prognosis, nomogram

## Introduction

Hepatocellular carcinoma (HCC) ranks as the third leading cause for cancer deaths in the world, and is the most frequent liver malignancy [1]. Studies have demonstrated that hepatitis B virus, hepatitis C virus, and alcoholism are the most common risk factors for HCC [2, 3]. Despite advances made in diagnosing and treating HCC, this disease remains a formidable threat to human health. Moreover, the frequency of tumor recurrence, metastasis, and drug resistance results in a 5-year survival rate for HCC patients, which is not satisfactory [4].

Therefore, identifying novel, specific biomarkers and targets may be useful for the diagnosis, prognostic analysis, and the development of targeted therapy approaches in HCC.

Autophagy is an important process that allows lysosomes to degrade damaged and non-functional proteins or organelles [5]. A growing number of recent studies have demonstrated that abnormal autophagy is involved in many types of cancers, including esophageal, gastric, and breast cancer [5-7]. Furthermore, other reports have demonstrated that some autophagy genes have the potential to serve as bio-

# Prognostic model based on autophagy-associated genes in HCC

**Table 1.** Primer sets used in this study

Genes	Primer sets
SQSTM1	Forward Primer: AGATTCGCCGCTTCAGCTT Reverse Primer: AACCAAGTCCCCGTCCTCAT
HSBP8	Forward Primer: CTTCACATGCCTGGCCTAA Reverse Primer: AGCGTCCTTAGGGAGTGCTA
BIRC5	Forward Primer: GAGGTATCTCGGCTGTCC Reverse Primer: CAAACAGGTCTGGGGTTCGT
GAPDH	Forward Primer: GCACCGTCAAGGCTGAGAAC Reverse Primer: TGGTGAAGACGCCAGTGA

markers or therapeutic targets for cancer management [5, 6]. Recent studies have also revealed an association between the pathophysiological processes of HCC and autophagy. For example, Fang et al. found that suppression of autophagy can inhibit hepatitis C virus replication in human hematoma cells [8]. Pan et al. reported that up-regulation of p62/SQSTM1 can decrease the sensitivity of HCC cells to sorafenib [9]. LC3, a vital marker for autophagy, was also demonstrated to be a promising indicator for predicting the prognosis of HCC patients [10]. In addition, studies have shown that small molecules involved in autophagy regulatory mechanisms may provide new clues for targeted therapy in advanced HCC [11]. For instance, ATG5 siRNA could suppress autophagy and enhance norcantharidin-induced apoptosis in HCC [12]. Xue et al. found that ULK1 may act as a novel target for HCC treatment [13]. The aim of our study was to demonstrate the use of prognostic models based on autophagy-associated genes (ARGs) to predict HCC prognosis, thus helping to improve the accuracy of prognosis and enable the administration of a targeted therapy.

In this study, we identified three ARGs from the TCGA database that are closely related to the overall survival (OS) of HCC patients. A prognosis prediction model was constructed based on these three ARGs and was demonstrated to perform well for HCC patients in both training and testing cohorts. Moreover, we established a clinical nomogram combining the OS-related ARGs and clinicopathological factors (age, gender, grade, stage, T (primary tumor), N (lymph nodes), M (metastasis), and risk score), and demonstrated its effectiveness in predicting the 3- and 5-year survival rates of HCC patients.

## Materials and methods

### *Data collection and processing*

The Human Autophagy Database (<http://www.autophagy.lu/index.html>) was used to download the 232 known ARGs [14]. RNA sequencing data and the corresponding clinical data for 374 HCC and 50 non-tumor tissues were acquired from the TCGA database (<https://tcga-data.nci.nih.gov/tcga/>) [15]. Data were processed as previously described [16]. The cBio Cancer Genomics Portal (<http://cbioportal.org>) was utilized to explore the genetic alterations and clinical information associated with select ARGs in HCC [17].

### *Sample collection, total RNA extraction, and qPCR*

Twenty paired HCC and adjacent non-tumor tissues were collected in Xijing Hospital from 2018-2019. Written informed consent was obtained from the patients. All these HCC patients had not received chemotherapy or radiotherapy prior to surgery. All tissue samples were snap-frozen and stored in liquid nitrogen (-80°C) until RNA extraction. The Ethics Committees of Xijing Hospital approved this study. Total RNA was extracted from tissues with TRIzol Reagent (Invitrogen, Life Technologies Corporation, Carlsbad, CA, USA) and then subjected to reverse transcription followed by quantitative real-time PCR (qPCR) reactions using PrimeScript™ RT-PCR Kit (TaKaRa, Otsu, Shiga, Japan). The primer sequences used are presented in **Table 1**.

### *Differentially expressed ARGs in HCC*

The Wilcoxon signed-rank test in package “lrimma” in R software (version 3.6.3) was applied to compare the differentially expressed ARGs between HCC and normal tissues, with the criteria of  $|\log_2 \text{fold-change (FC)}| > 1.5$  and an adjusted  $p$ -value  $< 0.05$  [16, 18]. Then, we combined the expression data of the differentially expressed ARGs with the corresponding clinical data. Univariate and multivariate Cox regression analyses were applied to identify the ARGs that are closely associated with OS in HCC [18]. We also performed Lasso regression to remove ARGs that might be closely correlated with others [14, 16].

# Prognostic model based on autophagy-associated genes in HCC

## *Functional enrichment analyses*

To explore the biological implications and potential mechanisms of ARGs in HCC, GO annotation and KEGG pathway analyses were conducted for the ARGs using the R software with the “GO plot”, “ggplot2”, “Cluster Profiler”, and “DOSE” packages [16]. Top enriched terms with  $p$ -value < 0.05 and  $q$ -value < 0.05 were regarded as noteworthy.

## *Establishment of the prognostic model for HCC*

A linear combination of ARGs expression and regression coefficient, based on the multivariate Cox regression, was used to construct the risk signature [14]. The OS-related predictive formula was calculated as follows: risk score = (Expression gene 1 × Coefficient gene 1) + (Expression gene 2 × Coefficient gene 2) + ... + (Expression gene n × Coefficient gene n) [14]. Then, each HCC patient was assigned a risk score according to the formula.

## *Assessment of the prognostic model*

Kaplan-Meier survival curves and time-dependent ROC curves were developed to evaluate the efficiency of the prognostic model [19].

## *External validation of the prognostic gene signature*

A GSE14520 dataset consisting of 221 HCC patients was downloaded from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) [20, 21]. Each patient received a risk score which was calculated using the same prognostic gene signature-based risk model as the TCGA dataset. Then, Kaplan-Meier and ROC curves were developed to evaluate the predictive performance of the prognostic gene signature.

## *Development of the clinical nomogram*

We downloaded clinical data of HCC patients from the TCGA database. Then, a clinical nomogram based on several factors (age, gender, grade, stage, primary tumor (T), lymph nodes (N), metastasis (M), and risk score) was built to predict the 3- and 5-year survival rates of HCC patients using the “survival” and “rms” packages in the R software package [19]. Moreover, the concordance index (C-index) and calibra-

tion curves were both applied to evaluate the accuracy of the nomogram.

## *Validation of ARG expression at the protein and mRNA levels*

The Human Protein Atlas database (<https://www.proteinatlas.org/>) contains more than 11,200 unique proteins [22]. Therefore, we utilized it to evaluate the protein level of the prognostic ARGs in HCC and normal tissues. The TIMER database (<https://cistrome.shinyapps.io/timer/>) was used to validate the mRNA level of ARGs in HCC tissues and normal samples [23]. qPCR was performed to detect the expression of prognostic ARGs in HCC and adjacent non-tumor tissues.

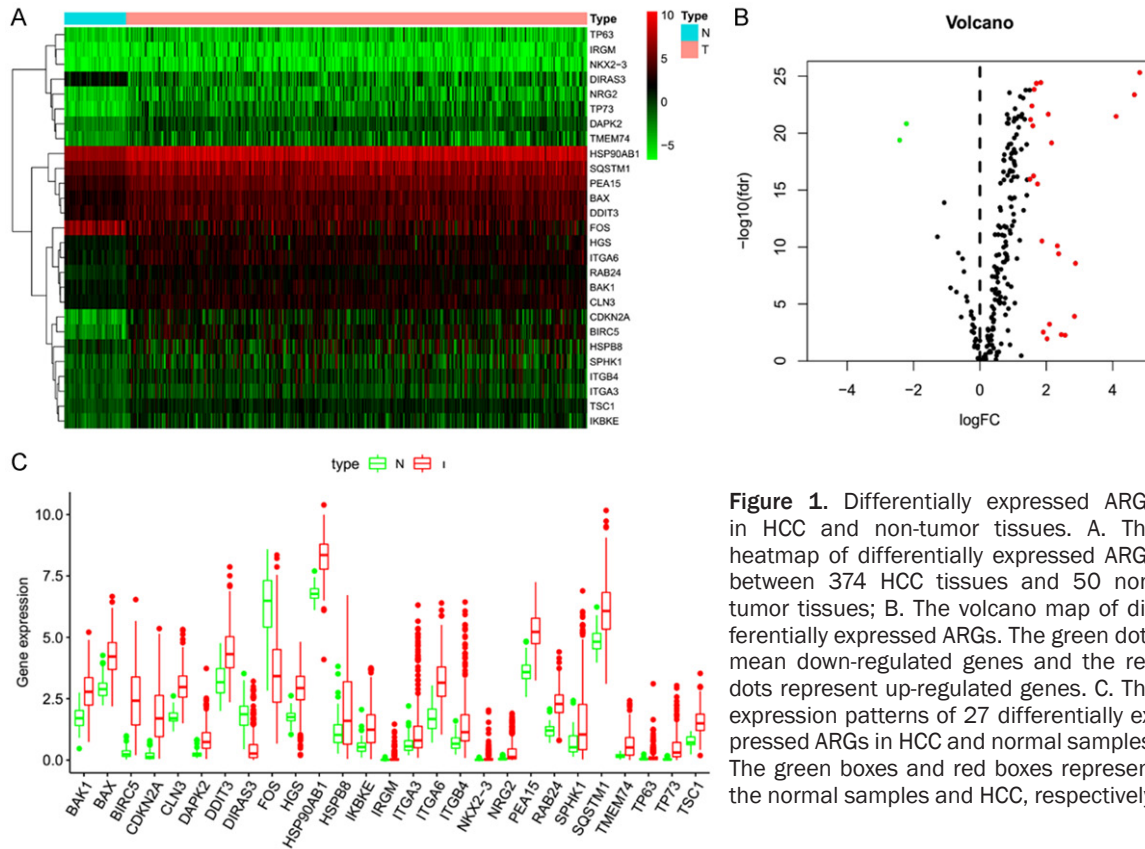
## *Development of transcription factor-gene networks and miRNA-gene networks*

The NetworkAnalyst database (<http://www.networkanalyst.ca>) was utilized to predict the transcription factors (TFs) and miRNAs of SQSTM1, HSPB8, and BIRC5 [24]. TF prediction was based on the ENCODE database with ChIP-seq data [24]. Only results exhibiting a peak intensity signal value greater than 500 and a potential score value less than 1 were identified for further study [24]. The miRNA-genes network was predicted and constructed using TarBase and miRTarBase in the NetworkAnalyst database.

## *Statistical analysis*

The Perl language and R software (version 3.6.3) were used to conduct all the statistical tests and produce the resulting graphics. The Wilcoxon signed-rank test was used to screen the differentially expressed ARGs between HCC and normal tissues. Cox regression analyses were used to identify the ARGs associated with OS in HCC patients. A Lasso regression analysis was utilized to remove ARGs that might be closely correlated with others. Chi-square test was utilized to exam the relationships between risk scores and clinical features. The Kaplan-Meier curve was plotted and the log-rank test was utilized to determine the differences of OS between the two groups. The ROC curve was used to evaluate the prognostic accuracy of the model. A  $p$ -value of less than 0.05 was considered as statistically significant.

## Prognostic model based on autophagy-associated genes in HCC



**Figure 1.** Differentially expressed ARGs in HCC and non-tumor tissues. A. The heatmap of differentially expressed ARGs between 374 HCC tissues and 50 non-tumor tissues; B. The volcano map of differentially expressed ARGs. The green dots mean down-regulated genes and the red dots represent up-regulated genes. C. The expression patterns of 27 differentially expressed ARGs in HCC and normal samples. The green boxes and red boxes represent the normal samples and HCC, respectively.

## Results

### Differential expression of ARGs between HCC and adjacent non-tumor tissues

In this study, RNA sequencing data from 374 HCC and 50 non-tumor tissues were analyzed to assess differential gene expression. Our analysis identified 27 differentially expressed ARGs, including twenty-five up-regulated genes (DDIT3, BAX, TSC1, HGS, BAK1, HSP90AB1, RAB24, CLN3, SQSTM1, PEA15, IKBKE, TP63, HSPB8, ITGA6, ITGA3, DAPK2, TMEM74, ITGB4, IRGM, NKX2-3, SPHK1, NRG2, TP73, CDKN2A, and BIRC5) and two down-regulated genes (FOS and DIRAS3), which exhibited  $|\log_2\text{FC}| > 1.5$  (Figure 1A and 1B). Figure 1C presents these 27 differentially expressed ARGs in HCC and non-tumor tissues.

### Bioinformatics analysis of differentially expressed ARGs

To investigate the biological functions and molecular mechanisms of the 27 ARGs in HCC, gene ontology (GO) enrichment and KEGG pa-

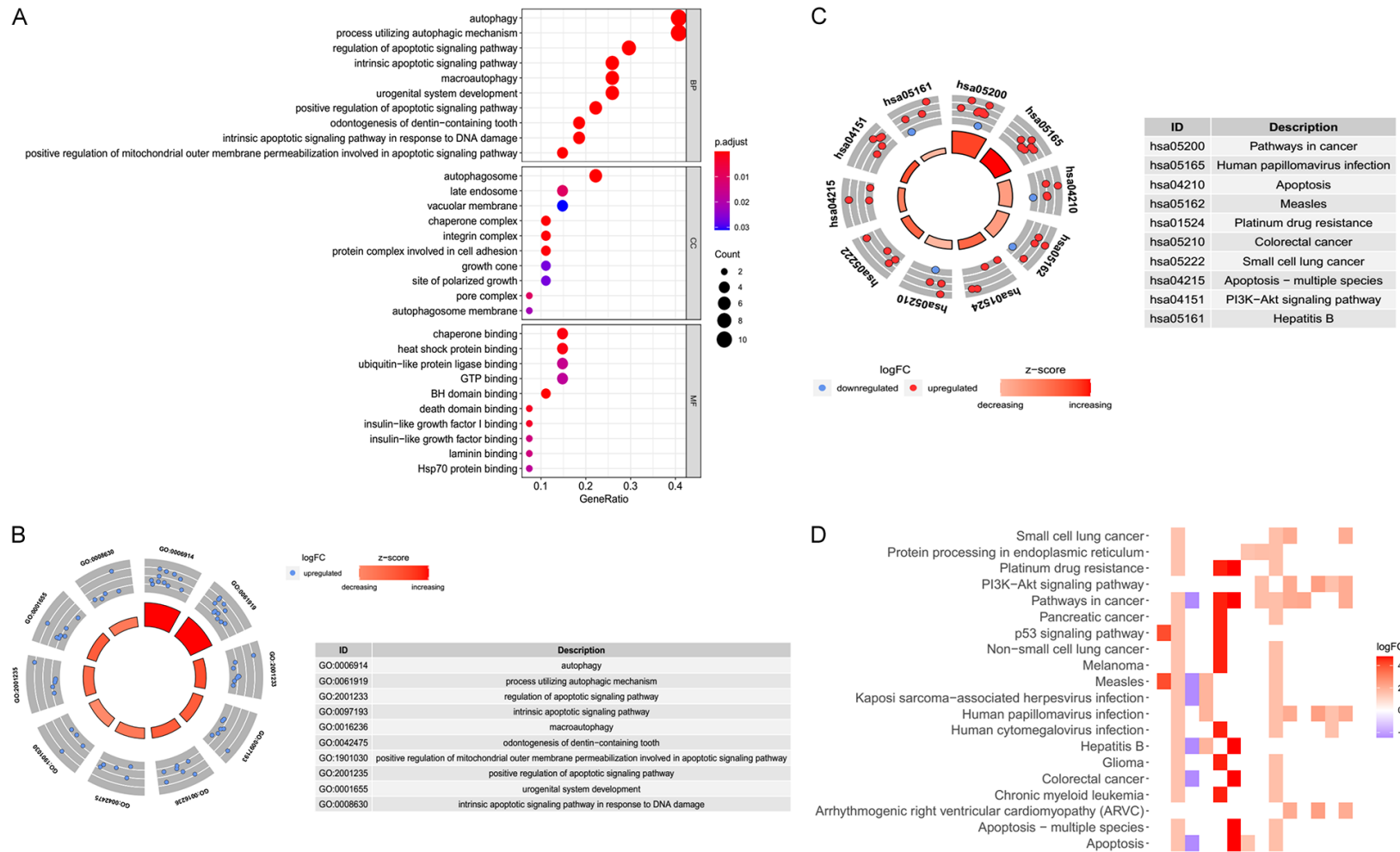
thway analyses were conducted (Figure 2). The GO enrichment analysis can be divided into three parts: biological processes (BP), cellular components (CC), and molecular function (MF). As shown in Figure 2A, the top enriched terms for BP included autophagy, process utilizing autophagic mechanism, and regulation of apoptotic signaling pathways. As for CC, the most significant terms were involved in autophagosome, chaperone complex, and integrin complex. Differentially expressed ARGs in the MF group were mainly associated with BH domain binding, chaperone binding, and heat shock protein binding. KEGG enrichment analysis showed that the identified differentially expressed ARGs are mainly associated with pathways involved in cancer, human papillomavirus infection, apoptosis, measles, and platinum drug resistance (Figure 2B).

### Prognostic gene signature for HCC cohorts

Using a univariate Cox regression analysis, a total of 13 differentially expressed ARGs (BAX, SQSTM1, PEA15, CDKN2A, HSPB8, HGS, IKBKE, HSP90AB1, RAB24, BIRC5, DDIT3, BAK1,

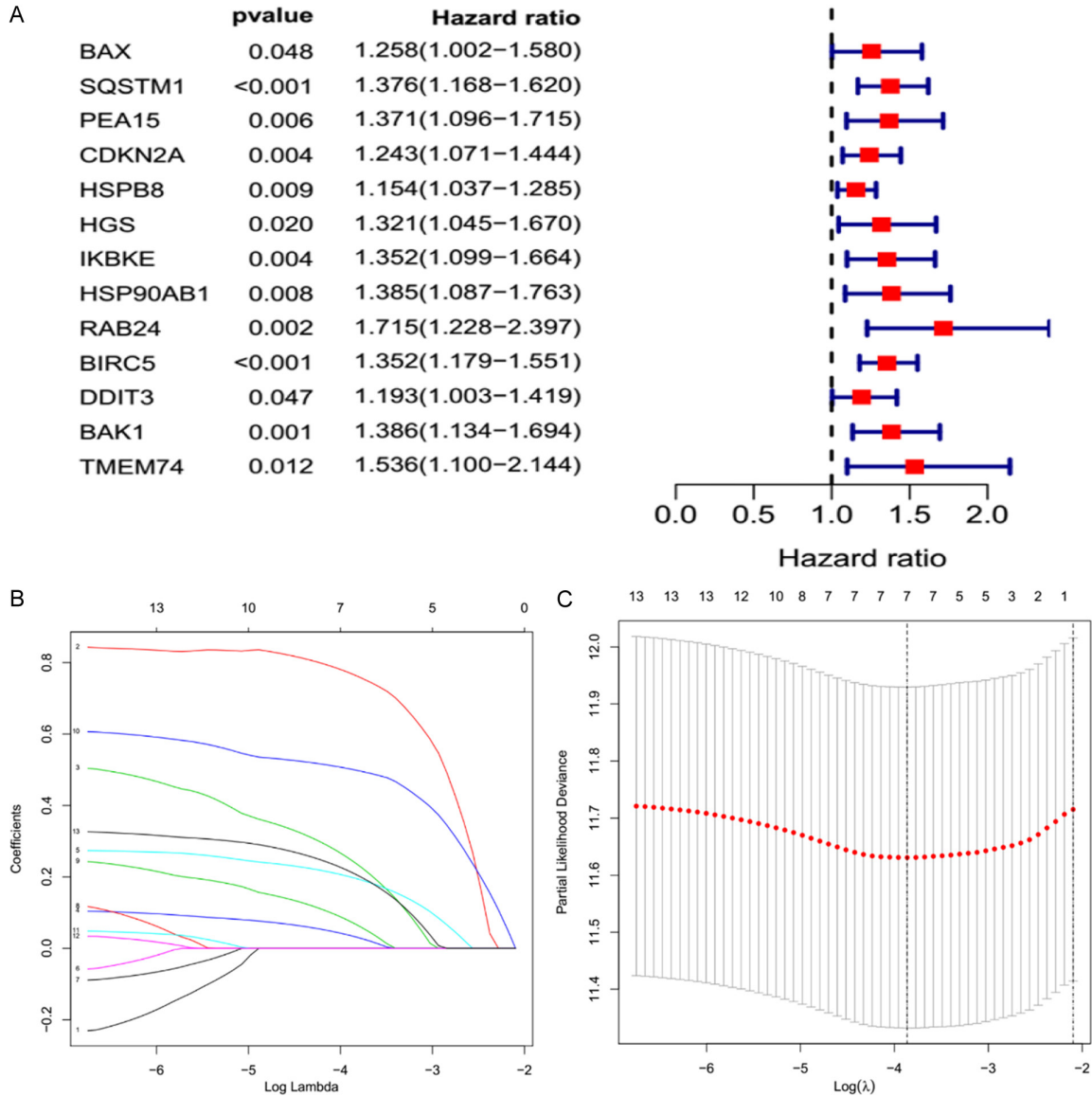


# Prognostic model based on autophagy-associated genes in HCC



**Figure 2.** GO enrichment analysis and KEGG pathway of differentially expressed ARGs. A. Bubble plot of significant GO terms. The change in color from red to blue indicates the decrease in the adjusted *P*-value, and the size of the circles represents the number of gene enriched in the GO terms. B. GOCircle plot of significant GO terms. C. The significant KEGG terms of differentially expressed ARGs. D. The heatmap illustrates the relationships between the differentially expressed ARGs and KEGG pathways.

Prognostic model based on autophagy-associated genes in HCC



**Figure 3.** Predictive gene signature constructed by Lasso regression. A. Forest plot of ARGs associated with OS in HCC. B. Lasso coefficient profiles of the 13 ARGs in HCC. C. The optimal lambda value in Lasso model for HCC.

and TMEM74) were found to be markedly related to OS in HCC patients (**Figure 3A**). All 13 survival-related ARGs were regarded as risk factors (HRs, 1.154-1.715;  $P < 0.05$ ) and their overexpression may worsen prognosis. When we validated the prognostic roles of these 13 survival-related ARGs using the Kaplan-Meier plotter website, we found that the results are consistent with those in the TCGA dataset (**Supplementary Figure 1**; The OS curve of BAX presented a similar trend to that of other genes, but was not statistically significant). Finally, we performed a Lasso regression analysis on these 13 differentially expressed ARGs.

**Figure 3B** presents the regression coefficient of the 13 ARGs in HCC. Although only seven ARGs were included in the model (SQSTM1, PEA15, CDKN2A, HSPB8, RAB24, BIRC5, and TMEM74), it still performed optimally (**Figure 3C**). **Table 2** lists the biological functions and risk coefficients of these seven ARGs, which are mainly associated with the formation of autophagosomes, regulation of autophagy, and the regulation of apoptosis.

To explore how these seven ARGs contribute to hepatocellular carcinogenesis, we explored the impact of genetic alteration of these genes

## Prognostic model based on autophagy-associated genes in HCC

**Table 2.** Biological functions and coefficient of seven ARGs

No	Gene symbol	Full name	Function	Risk coefficient
1	SQSTM1	Sequestosome 1	Functions as a bridge between polyubiquitinated cargo and autophagosomes	0.76707221
2	PEA15	Proliferation And Apoptosis Adaptor Protein 15	Functions as a regulator of apoptosis and autophagy	0.24377181
3	CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	Involved in regulation of autophagy and caspase-independent cell death	0.03399741
4	HSPB8	Heat Shock Protein Family B (Small) Member 8	Functions as a regulator of macroautophagy	0.19802769
5	RAB24	Ras-Related Protein Rab-24	Involved in autophagy-related processes	0.07147873
6	BIRC5	Baculoviral IAP Repeat Containing 5	Functions as a regulator of apoptosis and autophagy	0.50048513
7	TMEM74	Transmembrane Protein 74	Plays an essential role in autophagy	0.21098334

## Prognostic model based on autophagy-associated genes in HCC

**Table 3.** Univariate and multivariate Cox regression analyses of OS in HCC patients

Genes	Univariate analysis		Multivariate analysis		Coef
	HR (95% CI)	P	HR (95% CI)	P	
SQSTM1	1.3759 (1.1682-1.6204)	0.000132	1.2940 (1.1007-1.5213)	0.001792	0.257771
PEA15	1.3709 (1.0956-1.7153)	0.005811	---	---	---
CDKN2A	1.2434 (1.0706-1.4441)	0.004323	---	---	---
HSPB8	1.1542 (1.0371-1.2846)	0.008595	1.1264 (1.0079-1.2588)	0.035779	0.119042
RAB24	1.7155 (1.2278-2.3970)	0.001566	---	---	---
BIRC5	1.3523 (1.1790-1.5510)	1.60E-05	1.3565 (1.1810-1.5581)	1.61E-05	0.304895
TMEM74	1.5356 (1.0999-2.1439)	0.011769	---	---	---

using the cBio Cancer Genomics database. The PanCancer Atlas dataset (353 samples) and Firehose Legacy dataset (366 samples) of HCC were both included in this analysis. We found that the ARGs of interest are altered in 149 out of the 366 (41%) sequenced patients (TCGA, Firehose Legacy dataset) (Supplementary Figure 2). For comparison, altered ARGs were detected in 110 out of the 353 (31%) sequenced patients (TCGA, PanCancer Atlas dataset) (Supplementary Figure 3A). Moreover, HCC patients with altered genes exhibited poorer progression-free survival (PFS) (Supplementary Figure 3B) and disease-free survival (DFS) (Supplementary Figure 3C), than those with unaltered genes. The OS curves presented a similar trend to that of PFS, but were not statistically significant (Supplementary Figure 3D). These results suggest that the seven ARGs play a crucial role in HCC.

Finally, multivariate Cox regression analysis on the seven ARGs identified three candidate genes (SQSTM1, HSPB8, and BIRC5) as prognostic markers for HCC patients (Table 3). Each patient received a risk score that was calculated as follows: risk score = (0.2578 × expression value of SQSTM1) + (0.1190 × expression value of HSPB8) + (0.3049 × expression value of BIRC5).

### Identification of independent risk factors of OS for HCC patients

In order to screen the independent risk factors of OS for HCC patients, univariate and multivariate Cox regression analyses were conducted. According to Figure 4A, tumor stage, primary tumor (T), and risk score were closely related to OS (HR=1.669 (95% CI: 1.357-2.053), P < 0.001; HR=1.649 (95% CI: 1.354-2.009), P <

0.001; HR=1.755 (95% CI: 1.511-2.039), P < 0.001, respectively). As shown in Figure 4B, the results of the multivariate Cox regression indicated that metastasis (M) (HR=1.394 (95% CI: 1.065-1.824), P=0.016) and a risk score (HR=1.769 (95% CI: 1.478-2.116), P < 0.001) should be considered as independent risk factors of OS.

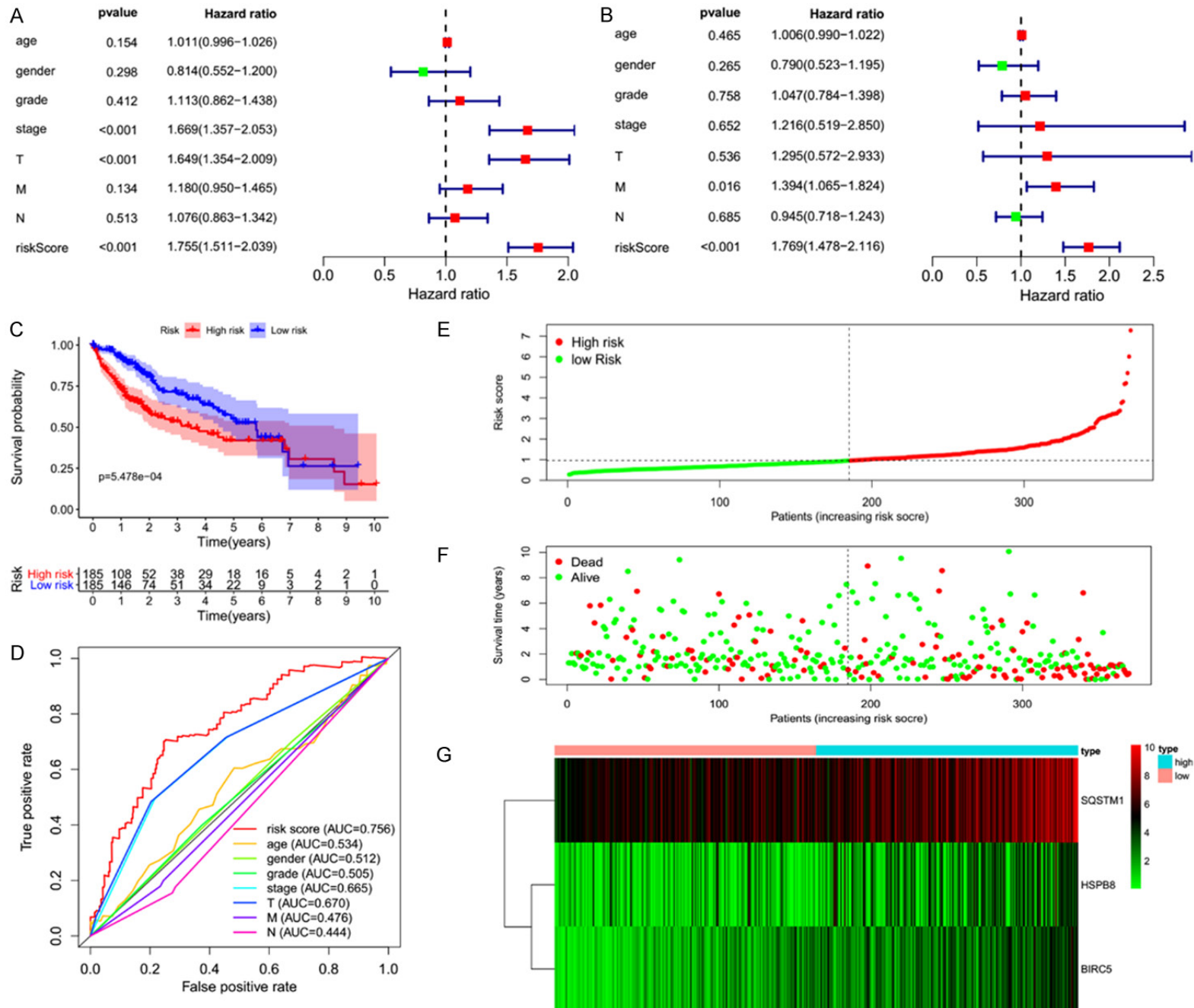
### Validation of the risk model

We divided the 374 HCC cases into high-risk and low-risk groups based on the median risk score. To validate the performance of the risk model, we plotted Kaplan-Meier curves to compare HCC survival in the two groups. Our results demonstrated that HCC patients in the low-risk group had a better prognosis than those in the high-risk group (3-year OS, 70.2% vs 53.7%; 5-year OS, 55.2% vs 42.0%; P=4.478e-04) (Figure 4C). ROC curves were also plotted using the risk factors related to OS (age, gender, grade, stage, primary tumor (T), metastasis (M), lymph nodes (N), and risk score). Evaluation of the area under the curve (AUC) values for each risk factor revealed that the risk score curve was a better predictor of survival when the AUC equaled 0.756 (Figure 4D). In addition, our results demonstrated that the survival time of HCC patients decreased with an increasing risk score (Figure 4E-G).

The prognostic effectiveness of the three-gene risk model was then validated using the GEO validation cohort (GSE14520; n=221). Similar to above, the patients were divided into the high-risk and low-risk groups based on the median risk score. Kaplan-Meier survival analysis showed that HCC patients in the high-risk group had a significantly worse OS compared to cases in the low-risk group in the validation



# Prognostic model based on autophagy-associated genes in HCC



## Prognostic model based on autophagy-associated genes in HCC

**Figure 4.** Prognostic prediction model of HCC patients. A. Assessment of the contribution of each factor to HCC survival by univariate Cox regression analysis. B. Assessment of the contribution of each factors to HCC survival by multivariate Cox regression analysis. C. Kaplan-Meier curves show that HCC patients in low-risk group have better OS than those in high-risk group. D. ROC curves of different variable in the prognostic model. E. The distribution of risk scores of HCC patients in prognostic model. F. The distribution of HCC patients with different survival status. G. The heatmap of three risk genes (SQSTM1, HSPB8, and BIRC5) in HCC patients with different risk score.

dataset (3-year OS, 57.7% vs 73.5%; 5-year OS, 43.2% vs 63.0%;  $P=1.274e-03$ ) (**Figure 5A**). Moreover, we found that the risk score was associated with a favorable predictive ability when the AUC equaled 0.672 (**Figure 5B**). Patients in the high-risk group exhibited a shorter survival time than those with a low-risk score in the validation cohort (**Figure 5C-E**). Taken together, our results show that the three-gene signature consisting of the ARGs SQSTM1, HSPB8, and BIRC5 can effectively predict prognosis in HCC.

### *Development of a nomogram for predicting the survival rate of HCC*

To quantitatively assess patient survival, a clinical nomogram was constructed by combining several risk factors (age, gender, grade, stage, primary tumor (T), lymph nodes (N), metastasis (M), and risk score). As shown in **Figure 6A**, the total points of the risk factors were utilized to evaluate an individual's 3- and 5-year survival rates. The concordance index (C-index) was 0.68 (95% CI: 0.63-0.73). In addition, calibration curves demonstrated good concordance between actual survival and nomogram-predicted survival (**Figure 6B and 6C**), especially for the 3-year survival rate.

### *Relationships between ARGs and clinical factors*

The Student's t-test was applied to investigate the relationships between the expression of the three ARGs (SQSTM1, HSPB8, and BIRC5) and clinical factors. Our results show that the risk scores increased along with the primary tumor (T) and tumor grade scores (**Supplementary Figure 4A and 4B**). In addition, the expression of SQSTM1 was higher in patients older than 65 years, male patients and in patients with a higher tumor grade (**Supplementary Figure 4C-E**). We also found that the levels of BIRC5 were related to the grade, tumor stage, and primary tumor (T) in HCC patients (**Supplementary Figure 4F-H**).

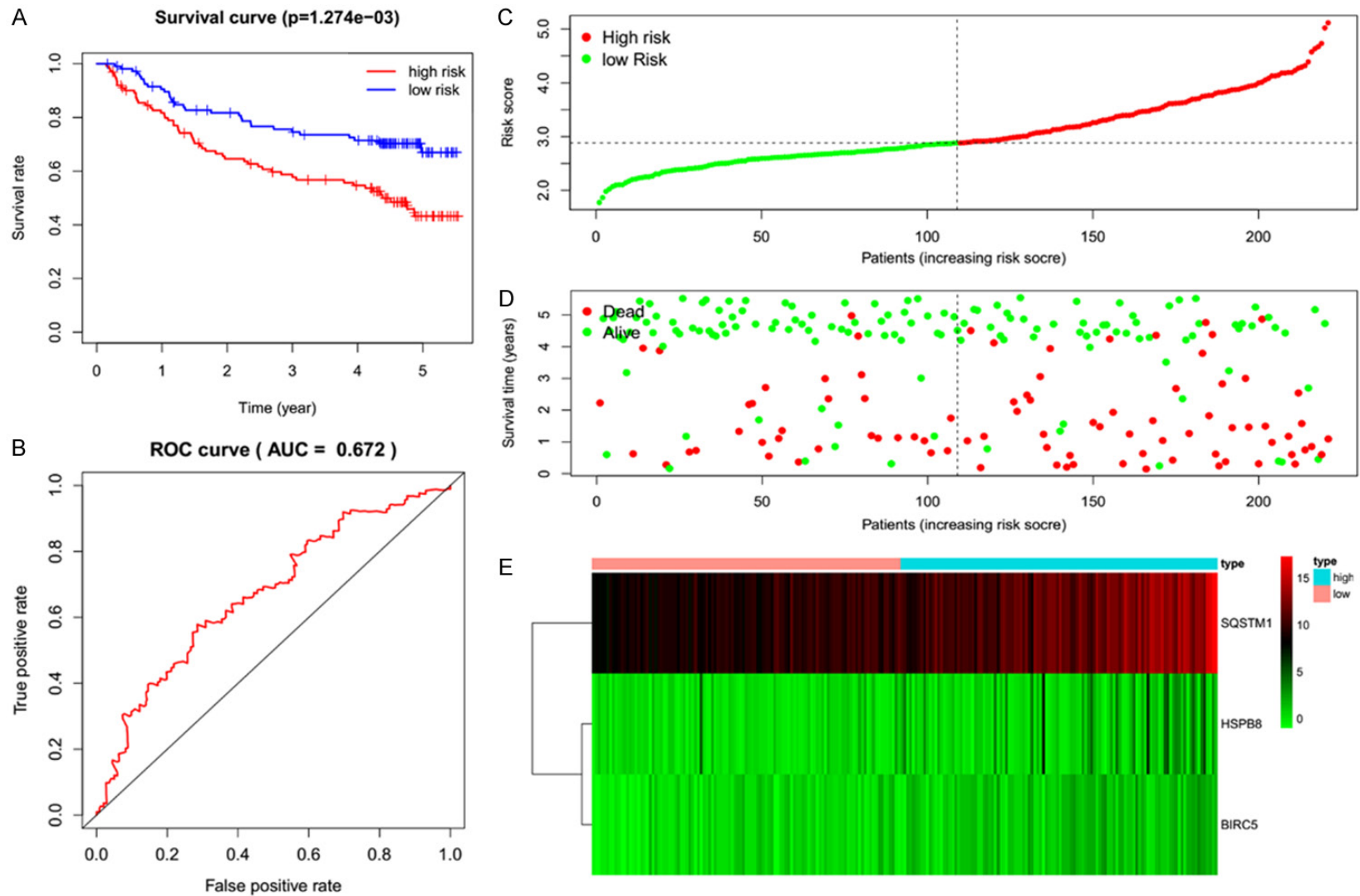
### *Validation of ARG expression at the protein and mRNA levels*

The Human Protein Atlas database (<https://www.proteinatlas.org/>) was used to evaluate the protein levels of SQSTM1, HSPB8, and BIRC5 in HCC tissues compared with their expression in normal tissues. As expected, the protein levels of SQSTM1, HSPB8, and BIRC5 were markedly higher in HCC tissues compared with normal samples (**Figure 7A-C**). qPCR analysis demonstrated that SQSTM1, HSPB8, and BIRC5 displayed higher expression in HCC tissues than in adjacent non-tumor tissues (**Figure 7D-F**). In addition, the TIMER database (<https://cistrome.shinyapps.io/timer/>) showed that the mRNA levels of SQSTM1, HSPB8, and BIRC5 were dramatically higher in HCC compared with the normal controls (**Supplementary Figure 5**).

### *Development of transcription factor (TF)-gene networks and miRNA-gene networks for SQSTM1, HSPB8, and BIRC5*

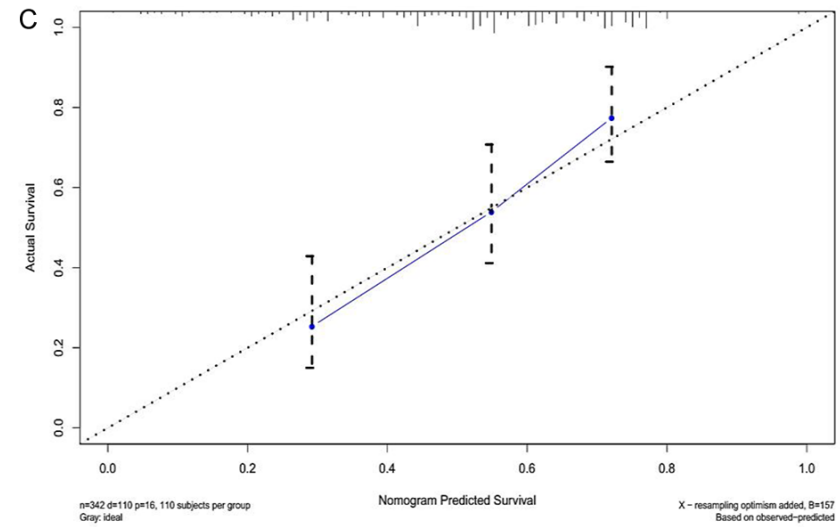
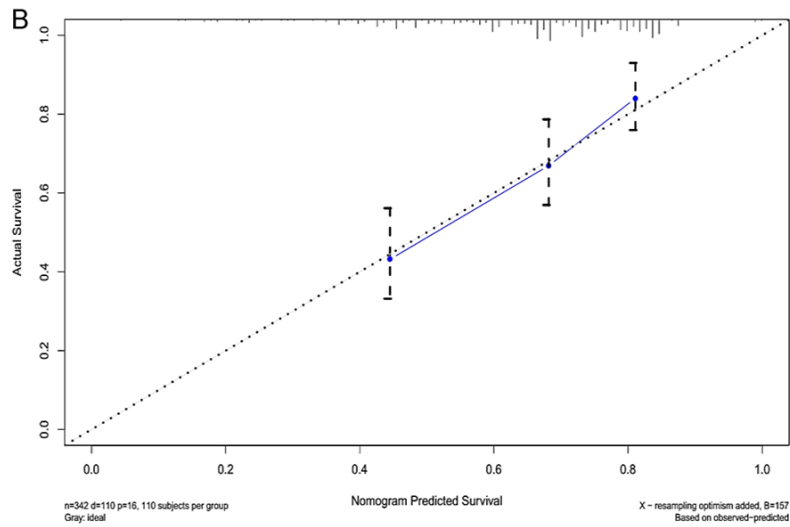
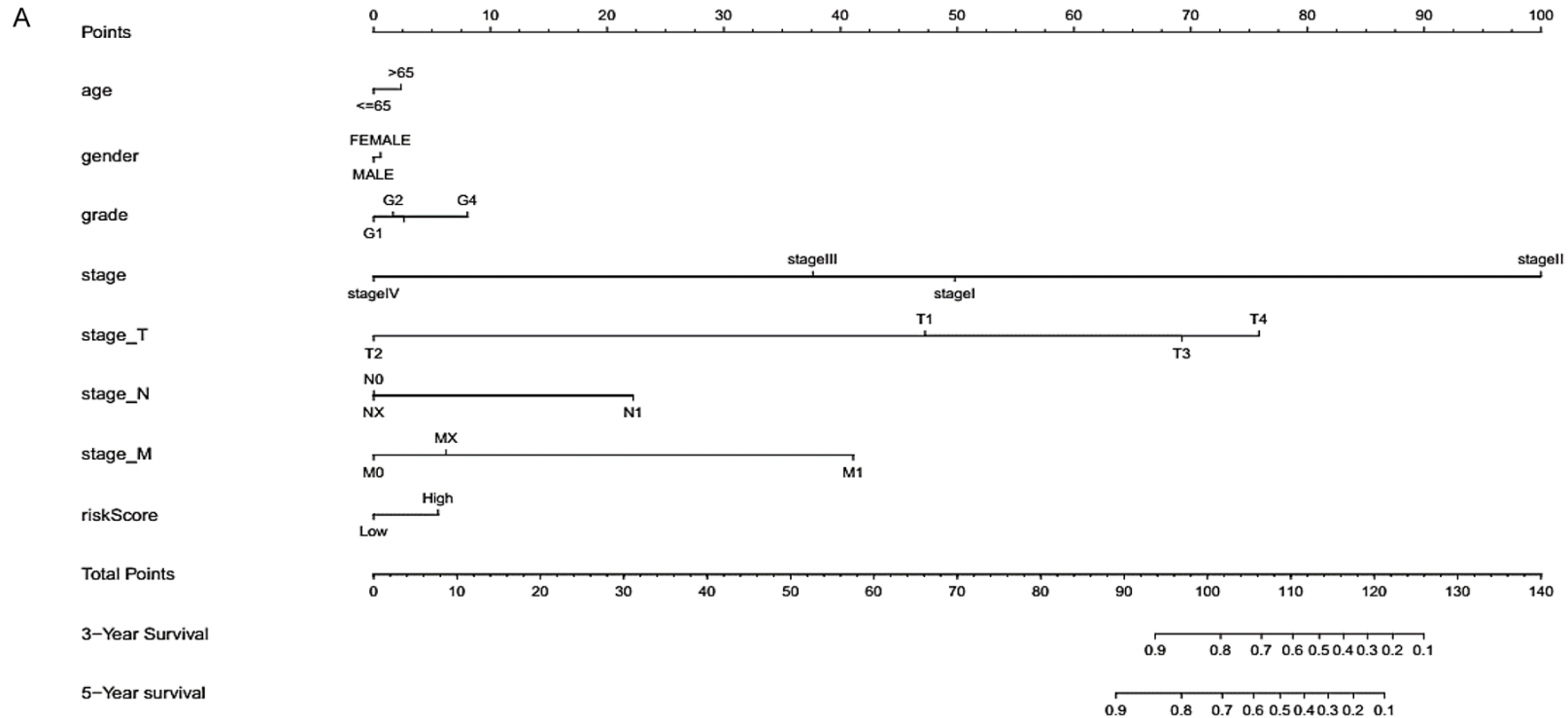
To better understand the contributions of SQSTM1, HSPB8, and BIRC5 in the development and progression of HCC, we constructed the TF-gene network and the miRNA-gene network for these ARGs (**Figure 8A and 8B**). The number of TFs and miRNAs in the networks were 100 and 117, respectively. In the TF-gene network (**Figure 8A**), ZNF394 was identified as the hub TF for these three target genes. Moreover, SQSTM1 and BIRC5 shared ten TFs (ZNF394, ZBTB7A, SCRT1, ZNF644, SSRP1, MLX, PPARG, CTCF, MDX3, and ZNF501). In the miRNA-gene networks (**Figure 8B**), miR-218-5p, miR-646, miR-93-5p, miR-16-5p, miR-484, miR-335-5p, and miR-1252-3p could regulate both SQSTM1 and BIRC5. In addition, SQSTM1 and HSPB8 were predicted as the target genes of miR-1226-3p. Taken together, the TF-gene and the miRNA-gene networks provide new clues to better understand the underlying molecular mechanisms of HCC in future studies.

Prognostic model based on autophagy-associated genes in HCC

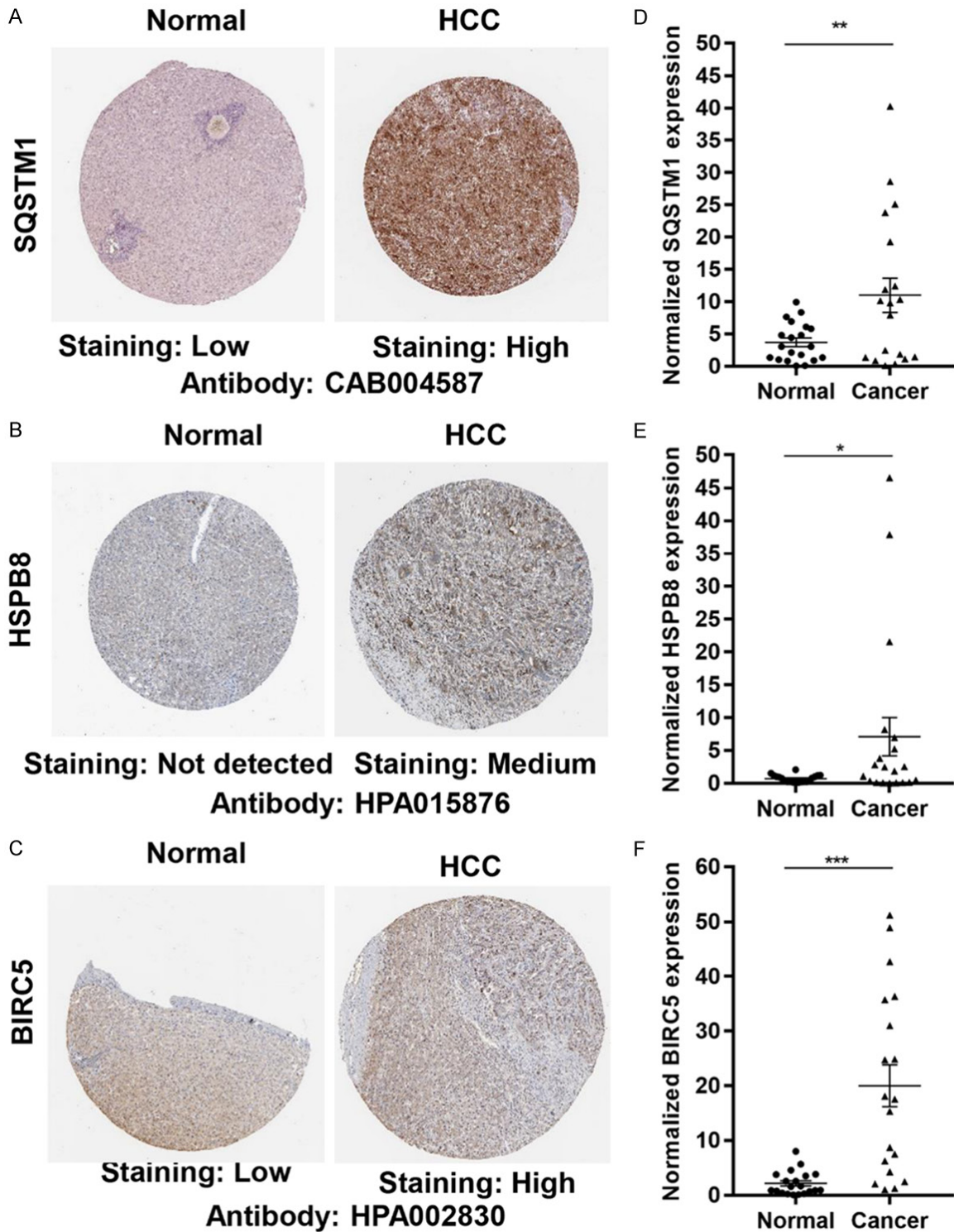


**Figure 5.** Validation of the risk signature in the testing cohort. A. Kaplan-Meier analysis shows that HCC patients with high risk score have poorer OS than those with low risk score; B. The ROC curve for assessing the prediction performance of the risk gene signature; C. The distribution of risk scores of HCC patients in prognostic model. D. The distribution of HCC patients with different survival status and survival time. E. The distributions gene expression profiles of SQSTM1, HSPB8, and BIRC5.

# Prognostic model based on autophagy-associated genes in HCC



**Figure 6.** The nomogram for predicting survival rate of HCC. A. The clinical nomogram for predicting the 3- and 5-year survival rate in HCC patients. B. The calibration curves present the concordance of 3-year survival between the observation and the prediction. C. The calibration curves present the concordance of 5-year survival between the observation and the prediction.



**Figure 7.** Protein and mRNA levels of SQSTM1, HSPB8, and BIRC5 in HCC and normal tissues. A. Immunohistochemistry results of SQSTM1 in normal tissue (staining: low; intensity: moderate; quantity: < 25%; location: nuclear) and in HCC (staining: high; intensity: strong; quantity: >75%; location: cytoplasmic/membranous/nuclear).



## Prognostic model based on autophagy-associated genes in HCC

B. Immunohistochemistry results of HSPB8 in normal tissue (staining: not detected; intensity: negative; quantity: none; location: none) and in HCC (staining: medium; intensity: moderate; quantity: >75%; location: cytoplasmic/membranous). C. Immunohistochemistry results of BIRC5 in normal tissue (staining: low; intensity: weak; quantity: >75%; location: cytoplasmic/membranous) and in HCC (staining: high; intensity: strong; quantity: 75-25%; location: cytoplasmic/membranous/nuclear). D. qPCR shows an increased expression of SQSTM1 in 20 paired HCC relative to adjacent non-tumor tissues. E. qPCR shows an increased expression of HSPB8 in 20 paired HCC relative to adjacent non-tumor tissues. F. qPCR shows an increased expression of BIRC5 in 20 paired HCC relative to adjacent non-tumor tissues. \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001.

### Discussion

HCC is a common and frequently occurring malignancy worldwide. Due to the lack of effective prognostic biomarkers, HCC patients usually cannot receive a reasonable treatment immediately [25]. Traditional prognostic risk factors (such as tumor size, histological type, stage, and grade) can only be adopted and evaluated post-surgery. Some scholars even believe that the current TNM stage system in use should be revised due to its ineffectiveness to accurately predict the prognosis of cancer patients [26-28]. Moreover, different patients can respond differently to treatments. Thus, more specific and effective markers need to be identified to evaluate prognosis and to screen for potentially high-risk HCC patients.

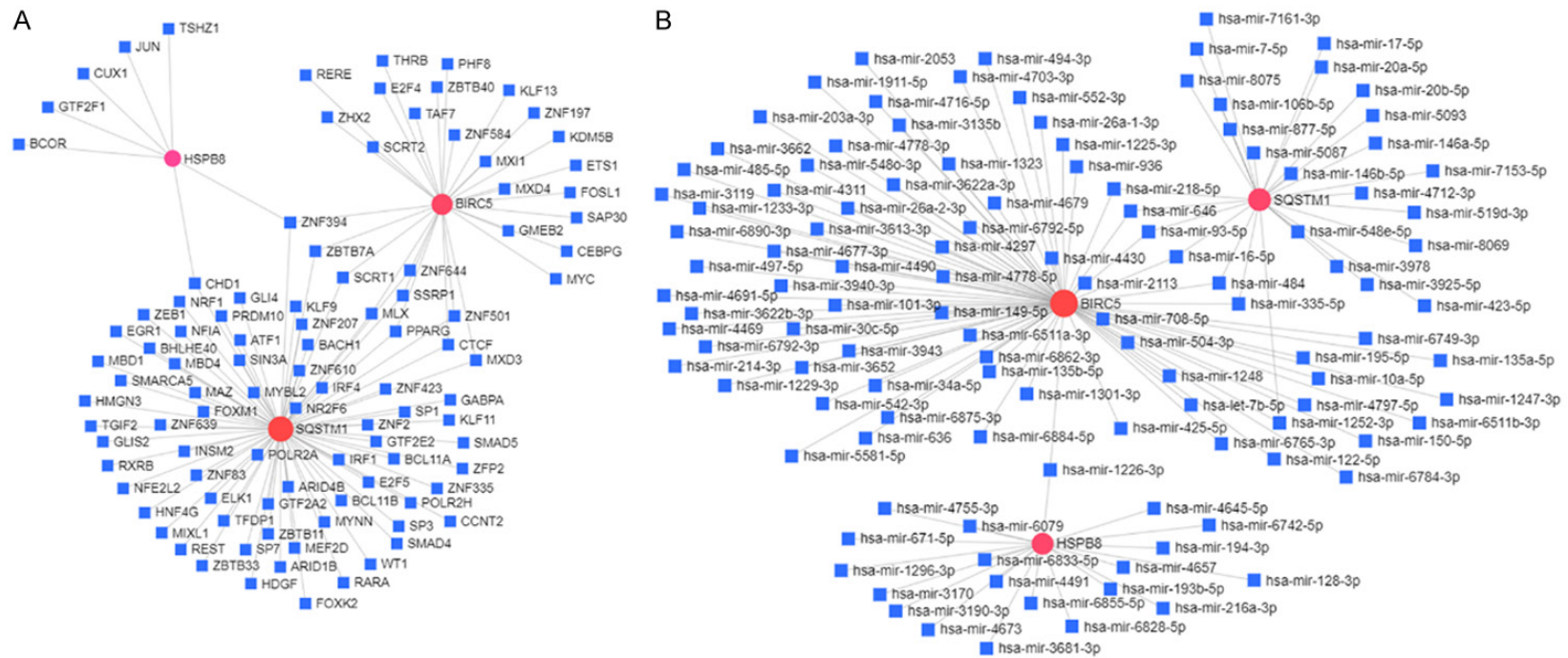
To date, an increasing number of biomarkers have been identified for use in predicting the prognosis of HCC [29, 30]. For example, Dai et al. confirmed that high expression of HIF-1 $\alpha$  is an independent prognostic factor for OS and DFS in HCC patients [31]. Similarly, our lab previously found that high levels of SOX12 is an independent and important risk factor for HCC patients [32]. However, further investigation is required to translate these biomarkers into clinical application. A better understanding of the molecular mechanisms behind the dysregulation of these markers in HCC is required. The markers must also be evaluated together, in studies consisting of larger sample sizes. Single-gene studies may lead to inaccurate conclusions given that the expression level of a single gene can be affected by many factors. A prognostic gene signature, consisting of multiple genes, may resolve these challenges and provide more accurate results because it is based on a statistical model comprised of multiple parameters.

Numerous publications have reported that dysfunctional autophagy is associated with several pathophysiological processes, including inflam-

mation, metabolic disorder, neurodegeneration, and cancer [33, 34]. Autophagy can function as a tumor suppressor or exert an oncogenic role in tumorigenesis, depending on the tumor microenvironment and tumor heterogeneity [6, 35]. For example, Fan et al. reported that autophagy can promote the metastasis and glycolysis of HCC cells through the Wnt/ $\beta$ -catenin pathway [36]. Conversely, exenatide-induced autophagy can inhibit the proliferation of HCC cells [37]. Considering its emerging role in cancers, studies focused on autophagy may provide us with a better understanding of the pathogenesis and prognosis of HCC. Importantly, developing a risk gene signature using the entire set of ARGs may provide a superior advantage to accurately predict survival.

Thus far, the rapid development of gene chip assays and second-generation gene sequencing have greatly facilitated the development of both personalized and precision medicine. Increasing numbers of biomarkers have been identified by integrating the analysis of genomic data from individual specimens. It is indisputable that these methods can be consolidated to improve current cancer management practices. To our knowledge, this study is the first to explore the prognostic roles of all reported ARGs in HCC. Here, we identified 27 differentially expressed ARGs from 374 HCC and 50 non-tumor tissues. Then, functional enrichment analyses were conducted to explore the roles and mechanisms of the differentially expressed ARGs in HCC. Through Cox and Lasso regression analyses, we established a risk model based on three OS-related ARGs, namely SQSTM1, HSPB8, and BIRC5. HCC patients were then divided into high-risk and low-risk groups according to the risk score derived from this model. Kaplan-Meier and ROC curves suggest that the risk model performed well in both the training and testing cohorts. A clinical nomogram combining clinicopathological features and a risk score were

## Prognostic model based on autophagy-associated genes in HCC



**Figure 8.** TF-gene networks and miRNA-gene networks of SQSTM1, HSPB8, and BIRC5. A. The predicted networks of TFs and three ARGs (SQSTM1, HSPB8, and BIRC5); The red circles represent SQSTM1, HSPB8, and BIRC5; The blue squares represent the predicted TFs; B. The predicted networks of miRNAs and three ARGs (SQSTM1, HSPB8, and BIRC5). The red circles represent SQSTM1, HSPB8, and BIRC5; The blue squares represent the predicted miRNAs.

# Prognostic model based on autophagy-associated genes in HCC

applied to predict the 3- and 5-year survival rates of HCC patients. Moreover, the calibration plots and C-index supported the effectiveness of the nomogram in predicting patient survival.

This study possesses a few deficiencies, which can be addressed in future investigations. First, our study focused mainly on the prognostic role of selected ARGs, but we did not examine the roles of the other ARGs. Second, we did not validate the expression levels of prognosis-related ARGs via *in vitro* assays. Third, we did not evaluate the prognostic model on a large clinical sample size. Finally, an in-depth investigation of the three prognosis-related ARGs is needed to better understand their potential as therapeutic targets for HCC.

## Conclusion

In conclusion, we identified three OS-related ARGs, namely SQSTM1, HSPB8, and BIRC5, as prognostic markers for HCC. Based on these genes, we developed a model to predict the survival rate in HCC, which exhibited good efficacy in guiding personalized therapy for HCC patients. Taken together, our results suggest that an ARG signature can act as an effective and promising prognostic indicator for HCC patients. Further studies on the identified ARGs may provide insight into their potential as therapeutic targets for HCC, in addition to improving current cancer management practices.

## Acknowledgements

This study was supported in part by grant from the Scientific Foundation of Shaanxi Province (2019ZDLSF01-02-01; 2018SF-240) and grant from the National Clinical Research Center for Digestive Diseases (2015BAI13B07) and Boost Program of Xijing Hospital (XJZT-18ML06).

## Disclosure of conflict of interest

None.

## Abbreviations

HCC, Hepatocellular carcinoma; FC, Fold change; FDR, False discovery rate; OS, Overall survival; TCGA, The Cancer Genome Atlas; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ROC, Receiver operating characteristic; AUC, The area under the ROC curve; qPCR, quantitative real-time PCR.

**Address correspondence to:** Dr. Yu Han, Department of Otolaryngology, Xijing Hospital, Fourth Military Medical University, No. 127, Changle West Road, Xi'an 710032, Shaanxi Province, China. E-mail: hanyufmmu@126.com; Drs. Daiming Fan and Liu Hong, State Key Laboratory of Cancer Biology and National Clinical Research Center for Digestive Diseases, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, No. 127, Changle West Road, Xi'an 710032, Shaanxi Province, China. E-mail: hlhyhj@126.com (DMF); hongliufmmu@163.com (LH)

## References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394-424.
- [2] Llovet JM, Montal R, Sia D and Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* 2018; 15: 599-616.
- [3] Fujiwara N, Friedman SL, Goossens N and Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J Hepatol* 2018; 68: 526-549.
- [4] Chang Lee R and Tebbutt N. Systemic treatment of advanced hepatocellular cancer: new hope on the horizon. *Expert Rev Anticancer Ther* 2019; 19: 343-353.
- [5] Onorati AV, Dyczynski M, Ojha R and Amaravadi RK. Targeting autophagy in cancer. *Cancer* 2018; 124: 3307-3318.
- [6] Huang T, Song X, Yang Y, Wan X, Alvarez AA, Sastry N, Feng H, Hu B and Cheng SY. Autophagy and hallmarks of cancer. *Crit Rev Oncog* 2018; 23: 247-267.
- [7] Amaravadi R, Kimmelman AC and White E. Recent insights into the function of autophagy in cancer. *Genes Dev* 2016; 30: 1913-1930.
- [8] Fang S, Su J, Liang B, Li X, Li Y, Jiang J, Huang J, Zhou B, Ning C, Li J, Ho W, Li Y, Chen H, Liang H and Ye L. Suppression of autophagy by mycophenolic acid contributes to inhibition of HCV replication in human hepatoma cells. *Sci Rep* 2017; 7: 44039.
- [9] Pan J, Lu C, Jun W, Wu Y, Shi X and Ding Y. The up-regulation of P62 levels is associated with resistance of sorafenib in hepatocarcinoma cells. *Int J Clin Exp Pathol* 2019; 12: 2622-2630.
- [10] Meng YC, Lou XL, Yang LY, Li D and Hou YQ. Role of the autophagy-related marker LC3 expression in hepatocellular carcinoma: a meta-analysis. *J Cancer Res Clin Oncol* 2020; 146: 1103-1113.

## Prognostic model based on autophagy-associated genes in HCC

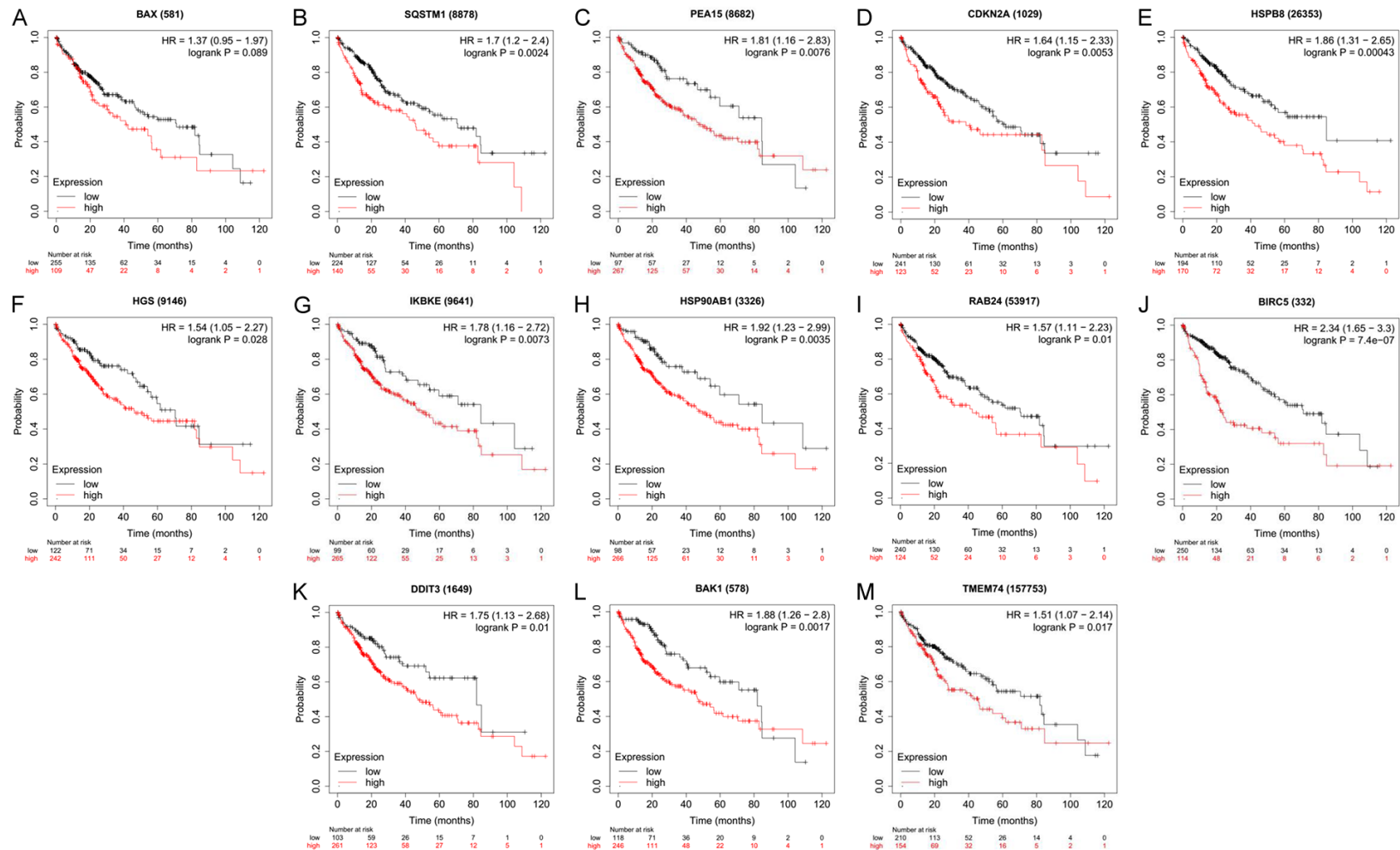
- [11] Lee YG and Jeon TI. Modulation of the autophagy-lysosomal pathway in hepatocellular carcinoma using small molecules. *Molecules* 2020; 25: 1580.
- [12] Xiong X, Wu M, Zhang H, Li J, Lu B, Guo Y, Zhou T, Guo H, Peng R, Li X, Tian Q and Wang Y. Atg5 siRNA inhibits autophagy and enhances norcantharidin-induced apoptosis in hepatocellular carcinoma. *Int J Oncol* 2015; 47: 1321-1328.
- [13] Xue ST, Li K, Gao Y, Zhao LY, Gao Y, Yi H, Jiang JD and Li ZR. The role of the key autophagy kinase ULK1 in hepatocellular carcinoma and its validation as a treatment target. *Autophagy* 2020; 16: 1823-1837.
- [14] Wang Y, Zhao W, Xiao Z, Guan G, Liu X and Zhuang M. A risk signature with four autophagy-related genes for predicting survival of glioblastoma multiforme. *J Cell Mol Med* 2020; 24: 3807-3821.
- [15] Wang Z, Jensen MA and Zenklusen JC. A practical guide to the cancer genome atlas (TCGA). *Methods Mol Biol* 2016; 1418: 111-141.
- [16] Wang X, Yao S, Xiao Z, Gong J, Liu Z, Han B and Zhang Z. Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes. *J Transl Med* 2020; 18: 149.
- [17] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; 2: 401-404.
- [18] Hu D, Jiang L, Luo S, Zhao X, Hu H, Zhao G and Tang W. Development of an autophagy-related gene expression signature for prognosis prediction in prostate cancer patients. *J Transl Med* 2020; 18: 160.
- [19] Liu Y, Wu L, Ao H, Zhao M, Leng X, Liu M, Ma J and Zhu J. Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. *Aging (Albany NY)* 2019; 11: 11440-11462.
- [20] Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX and Wang XW. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res* 2010; 70: 10202-10212.
- [21] Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, Walker R, Jia HL, Ye QH, Qin LX, Tang ZY, He P, Hunter KW, Thorgeirsson SS, Meltzer PS and Wang XW. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. *Gastroenterology* 2012; 142: 957-966, e912.
- [22] Ponten F, Schwenk JM, Asplund A and Edqvist PH. The Human Protein Atlas as a proteomic resource for biomarker discovery. *J Intern Med* 2011; 270: 428-446.
- [23] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B and Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 2017; 77: e108-e110.
- [24] Xia J, Gill EE and Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015; 10: 823-844.
- [25] Hartke J, Johnson M and Ghabril M. The diagnosis and treatment of hepatocellular carcinoma. *Semin Diagn Pathol* 2017; 34: 153-159.
- [26] Tellapuri S, Sutphin PD, Beg MS, Singal AG and Kalva SP. Staging systems of hepatocellular carcinoma: a review. *Indian J Gastroenterol* 2018; 37: 481-491.
- [27] Cattoni M, Vallières E, Brown LM, Sarkeshik AA, Margaritora S, Siciliani A, Filosso PL, Guertera F, Imperatori A, Rotolo N, Farjah F, Wandell G, Costas K, Mann C, Hubka M, Kaplan S, Fari-var AS, Aye RW and Louie BE. Improvement in TNM staging of pulmonary neuroendocrine tumors requires histology and regrouping of tumor size. *J Thorac Cardiovasc Surg* 2018; 155: 405-413.
- [28] Ilhan E, Ureyen O and Meral UM. Ongoing problems concerning 7(th) TNM staging system and proposals for 8(th) TNM staging system of gastric cancer. *Prz Gastroenterol* 2016; 11: 223-225.
- [29] De Stefano F, Chacon E, Turcios L, Marti F and Gedaly R. Novel biomarkers in hepatocellular carcinoma. *Dig Liver Dis* 2018; 50: 1115-1123.
- [30] Yim SY, Hae NJ, Shin JH, Jeong YS, Kang SH, Park YN, Um SH and Lee JS. Identification of prognostic biomarker in predicting hepatocarcinogenesis from cirrhotic liver using protein and gene signatures. *Exp Mol Pathol* 2019; 111: 104319.
- [31] Dai CX, Gao Q, Qiu SJ, Ju MJ, Cai MY, Xu YF, Zhou J, Zhang BH and Fan J. Hypoxia-inducible factor-1 alpha, in association with inflammation, angiogenesis and MYC, is a critical prognostic factor in patients with HCC after surgery. *BMC Cancer* 2009; 9: 418.
- [32] Huang W, Chen Z, Shang X, Tian D, Wang D, Wu K, Fan D and Xia L. Sox12, a direct target of FoxQ1, promotes hepatocellular carcinoma metastasis through up-regulating Twist1 and FGF19. *Hepatology* 2015; 61: 1920-1933.
- [33] Saha S, Panigrahi DP, Patil S and Bhutia SK. Autophagy in health and disease: a comprehensive review. *Biomed Pharmacother* 2018; 104: 485-495.
- [34] Mizushima N. A brief history of autophagy from cell biology to physiology and disease. *Nat Cell Biol* 2018; 20: 521-527.

## Prognostic model based on autophagy-associated genes in HCC

- [35] Mowers EE, Sharifi MN and Macleod KF. Functions of autophagy in the tumor microenvironment and cancer metastasis. *FEBS J* 2018; 285: 1751-1766.
- [36] Fan Q, Yang L, Zhang X, Ma Y, Li Y, Dong L, Zong Z, Hua X, Su D, Li H and Liu J. Autophagy promotes metastasis and glycolysis by upregulating MCT1 expression and Wnt/ $\beta$ -catenin signaling pathway activation in hepatocellular carcinoma cells. *J Exp Clin Cancer Res* 2018; 37: 9.
- [37] Krause GC, Lima KG, Levorse V, Haute GV, Gasen RB, Garcia MC, Pedrazza L, Donadio MVF, Luft C and de Oliveira JR. Exenatide induces autophagy and prevents the cell regrowth in HepG2 cells. *EXCLI J* 2019; 18: 540-548.

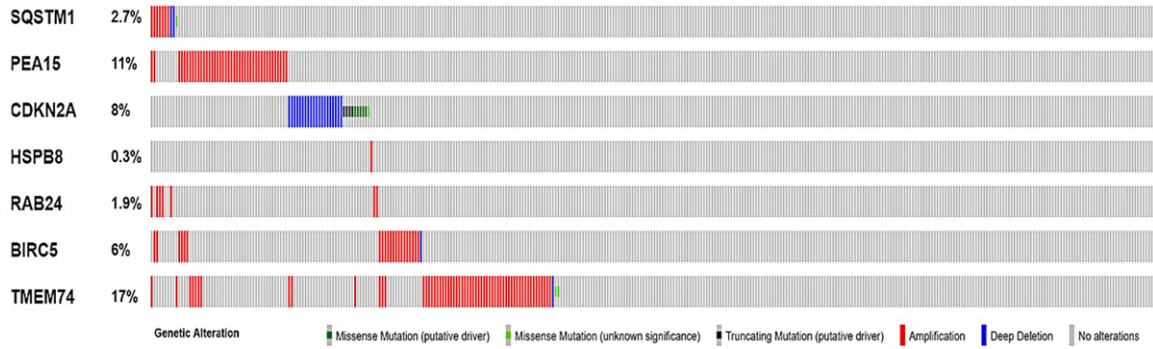


## Prognostic model based on autophagy-associated genes in HCC



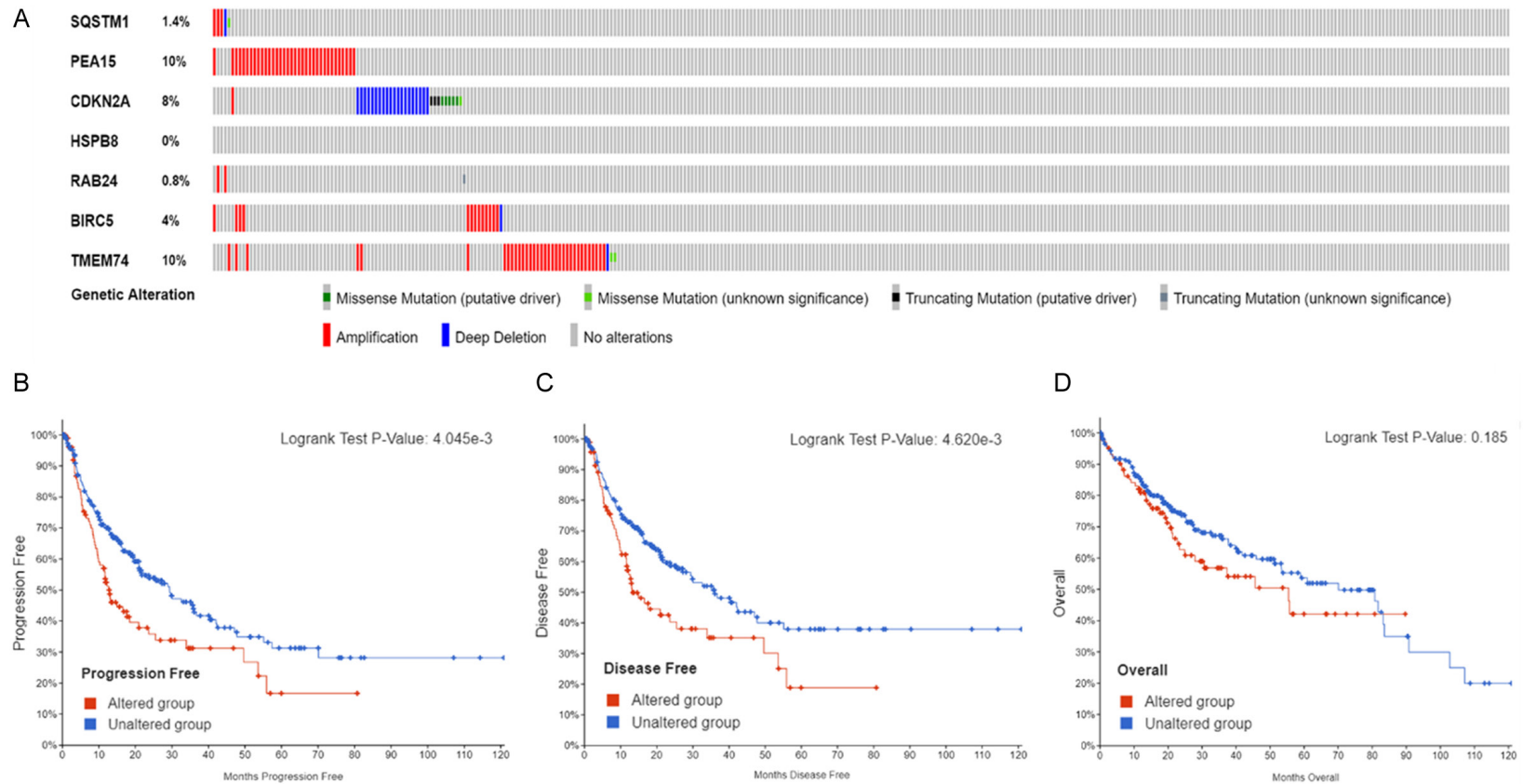
**Supplementary Figure 1.** Validation of prognostic role of 13 ARGs in Kaplan Meier-plotter website. The Kaplan Meier-plotter website (<http://kmplot.com/analysis/index.php>) was used to validate the prognostic role of 13 ARGs (BAX (A), SQSTM1 (B), PEA15 (C), CDKN2A (D), HSPB8 (E), HGS (F), IKBKE (G), HSP90AB1 (H), RAB24 (I), BIRC5 (J), DDIT3 (K), BAK1 (L), and TMEM74 (M)). High expression of SQSTM1, PEA15, CDKN2A, HSPB8, HGS, IKBKE, HSP90AB1, RAB24, BIRC5, DDIT3, BAK1, and TMEM74 was correlated with poorer OS of HCC patients. The OS curve of BAX showed a similar trend to that of other genes, but was not statistically significant.

## Prognostic model based on autophagy-associated genes in HCC



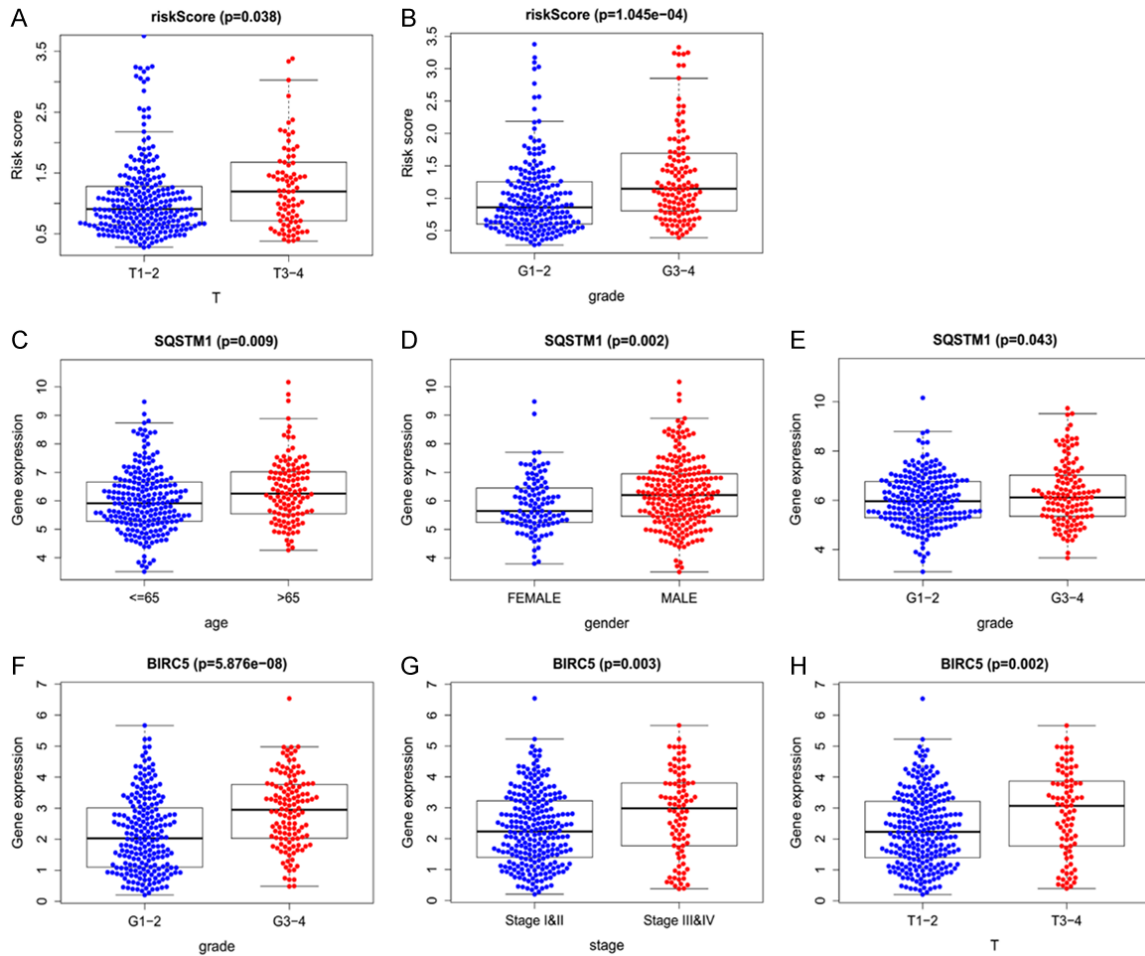
**Supplementary Figure 2.** Genetic alteration of seven ARGs in the HCC cohort (TCGA, Firehose Legacy dataset). Genetic alteration of SQSTM1, PEA15, CDKN2A, HSPB8, RAB24, BIRC5, and TMEM74 in HCC patients.

## Prognostic model based on autophagy-associated genes in HCC



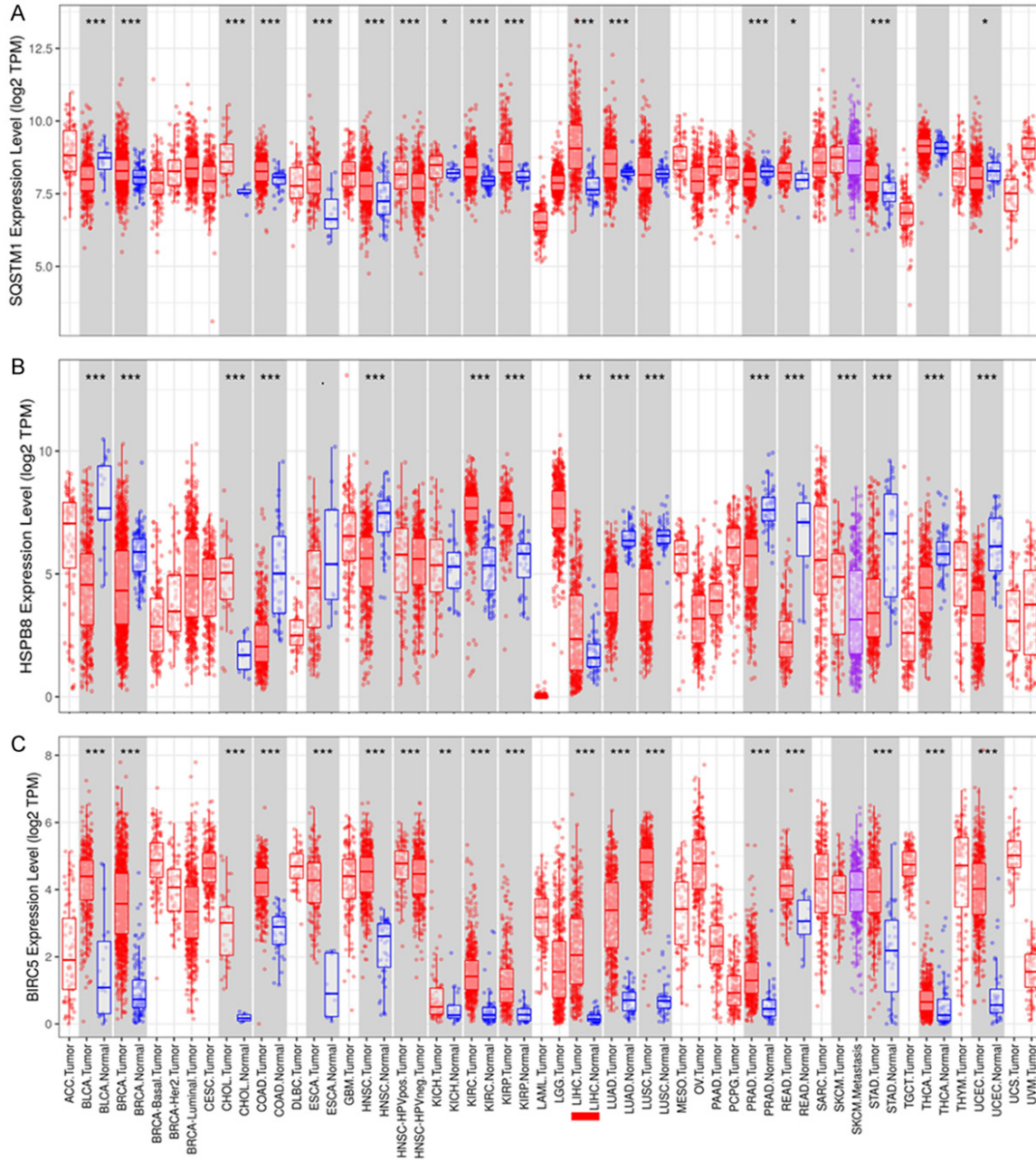
**Supplementary Figure 3.** Genetic alteration and prognostic role of seven ARGs in HCC cohort (TCGA, PanCancer Atlas dataset). A. Genetic alteration of seven ARGs (SQSTM1, PEA15, CDKN2A, HSPB8, RAB24, BIRC5, and TMEM74) in HCC patients. B. HCC patients with altered genes exhibits poorer progression-free survival (PFS) than these with unaltered genes. C. HCC patients with altered genes have poorer disease-free survival (DFS) than these with unaltered genes. D. The OS curves present similar trend to PFS curves and DFS curves, but not statistically significant.

## Prognostic model based on autophagy-associated genes in HCC



**Supplementary Figure 4.** Relationships between ARGs and clinical factors. A. The risk scores increase along with the T (primary tumor). B. The risk scores increase along with tumor grade. C. The expression of SQSTM1 is higher in patients older than 65 years. D. The expression of SQSTM1 is higher in the groups of male. E. The expression of SQSTM1 is higher in patients with a higher tumor grade. F-H. The levels of BIRC5 are associated with the grade, tumor stage and T (primary tumor) in HCC patients.

Prognostic model based on autophagy-associated genes in HCC



**Supplementary Figure 5.** Validation of ARGs expression in the TIMER database. The TIMER database (<https://cistrome.shinyapps.io/timer/>) was used to validate the mRNA expression levels of SQSTM1 (A), HSPB8 (B), and BIRC5 (C) in HCC and normal tissues. The mRNA levels of SQSTM1 (A), HSPB8 (B), and BIRC5 (C) are dramatically higher in HCC compared with the normal controls. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .