Development of a Prognostic Gene Signature For Hepatocellular Carcinoma

Cuiyun Wu

Shunde Hospital of Southern Medical University

Yaosheng Luo

First Peoples Hospital of Shunde: Shunde Hospital of Southern Medical University

Yinghui Chen

First Peoples Hospital of Shunde: Shunde Hospital of Southern Medical University

Hongling Qu

First Peoples Hospital of Shunde: Shunde Hospital of Southern Medical University

Lin Zheng

First Peoples Hospital of Shunde: Shunde Hospital of Southern Medical University

Jie Yao (✉ jie.yao413@yahoo.com)

Shunde hospital

Abstract

Background: Accurate prediction of overall survival is important for prognosis and the assignment of appropriate personalized clinical treatment in hepatocellular carcinoma (HCC) patients. The aim of the present study was to establish an optimal gene model for the independent prediction of prognosis associated with common clinical patterns.

Methods: Gene expression profiles and the corresponding clinical information of the LIHC cohort were obtained from The Cancer Genome Atlas. Differentially expressed genes were found using the R package "limma". Subsequently, a prognostic gene signature was developed using the LASSO Cox regression model. Kaplan–Meier, log-rank, and receiver operating characteristic (ROC) analyses were performed to verify the predictive accuracy of the prognostic model. Finally, a nomogram and calibration plot were created using the "rms" package.

Results: Differentially expressed genes were screened with threshold criteria (FDR < 0.01 and |log FC|>3) and 563 differentially expressed genes were obtained, including 448 downregulated and 115 upregulated genes. Using the LASSO Cox regression model, a prognostic gene signature was developed based on nine genes,*IQGAP3, BIRC5, PTTG1, STC2, CDKN3, PBK, EXO1, NEIL3, and HOXD9*, the expression levels of which were quantitated using RT-qPCR. According to the risk scores, patients were separated into high-risk and low-risk groups. Patients with lower risk scores generally had a better survival rate than those with higher risk scores. The mortality rate in the high-risk group was 42.02%, while that in the low-risk group was 12.50%. Results of the log-rank test showed significant differences in mortality between the two groups (HR: 4.86; 95% CI: 2.72−8.69; P = 1.01E-08). Subsequently, we assessed the prognostic accuracy of the gene signature using an ROC curve and the results show good sensitivity and specificity, with an average area under the curve (AUC) of 0.81 at 5 years (P < 0.01). Following multivariate adjustment for conventional clinical patterns, the prognostic gene signature remained a powerful and independent factor (HR: 4.70; 95% CI: 2.61−8.38; P = 2.06E-07), confirming its robust predictive ability of overall survival in HCC patients. Finally, a nomogram was established based on the gene signature and four clinicopathological features, which demonstrated an advantageous discriminating ability with the potential to facilitate clinical decision-making in HCC.

Conclusion: Our prognostic gene signature can be used as a combined biomarker for the independent prediction of overall survival in HCC patients. Moreover, we created a nomogram that can be used to infer prognosis and aid individualized decisions regarding treatment and surveillance.

Introduction

Hepatocellular carcinoma (HCC) is the second most lethal type of cancer, possessing a 5-year survival rate of 18%and a multi-step development process that includesmultiple biological factors(1, 2). The major risk factors for HCC development include gender(3), age(4, 5), chronic infection with the hepatitis B virus (HBV) or hepatitis C virus (HCV)(6, 7), cirrhosis(8), alpha-fetoprotein (AFP) levels(9, 10), and various

pathophysiological characteristics of the tumor(9, 11). Despite several clinicopathological factors being associated with survival, accurate prognostic biomarkers for patients with HCC remain elusive.

Recently, large-scale cancer genomics projects, such as The Cancer Genome Atlas (TCGA), have provided profiles of many cancer types according to the majority of patients(12). The latest state-of-the-art high-throughput techniques can assess the expression levels of thousands of genes in a single assay, serving as a strong analytical tool for the capture of prognostic characteristics of cancers. Additionally, owing to its ability to detect key features among complex datasets, machine learning technology is becoming increasingly popular for modeling the progression and treatment of cancer(13). Machine learning can be employed to extract important features from a large number of genes with a view to providing potentially useful prognostic information. In most human cancers, the prognosis is closely related to pathological criteria, particularly the histological grade. In addition to these high-risk etiological factors, the latest insights into HCC biology also indicate that genetic and epigenetic abnormalities play a vital role in the occurrence of liver carcinogenesis, such as molecular changes at different stages of HCC progression, which are induced by specific etiological factors in most cases(14–17). The majority of cancer research is focused on the elucidation of driver genes since this helps to uncover predictive information and identify prognostic biomarkers that can serve as therapeutic targets. In comparison with individual analyses, the integration of a panel of biomarkers substantially improves the prognostic value and greatly benefits the clinical management of cancer(18). For example, Liu et al. constructed a risk score model consisting of five stage-related genes, including NUP62, EHMT2, RANBP1, MSH6, and FHL2, which can successfully predict the recurrence of HCC within 1 year(19). Moreover, Li et al. validated a CIMP-associated prognostic model with an optimal ability to independentlypredict common clinical patterns related to survival (C-index:0.68;HR: 1.61; 95% CI:1.08–2.40; P = 0.019)(20). Furthermore, Yang et al. developed a four-long non-coding RNA signature as a novel candidate biomarker for prognosis in HCC patients (AUC: 0.689; 95% CI: 0.617–0.761; P < 0.01)(21). Nevertheless, due to the varying quality of prediction models, the construction of an optimal prognostic model remains a hot topic.

Owing to the development of various testing methods, an increasing number of biological markers have been reported. Due to the longer time period and higher cost involved in RNA sequencing or microarray, a faster, more economical, and feasible detection method is preferred for HCC patients. However, to the best of our knowledge, the current genome-wide models that focus on predicting the prognosis in HCC patients display varying success. Therefore, in the present study, we aimed to develop a multi-gene signature using the least absolute shrinkage and selection operator (LASSO) Cox regression model to predict the prognosis in a group of HCC patients obtained from TCGA. Through analysis of the areas classified as low-risk and high-risk by our model, we were able to obtain insight into the most relevant genes used for HCC patient stratification. In particular, we sought to establish a nomogram based on a prognostic gene signature and baseline clinicopathological risk factors with a view to assessing the efficacy of our prediction model. We show that the robust prognostic mode based on LASSO Cox regression possesses an optimal prediction of prognosis independent of common clinical patterns and may act as a promising therapeutic biomarker for HCC.

Materials And Methods

Data collection and processing. Gene expression profiles and the corresponding clinical information were obtained from TCGA. The original mRNA expression count values were normalized using the Trimmed Mean of M (TMM) method by package "edgeR".Main exclusion criteria were shown as follows, without follow-up or survival status, no mRNA expression data and the expression level of mRNA is greater than zero and exists more than three databases.As a result, data derived from239 HCC samples and 23 adjacent noncancerous tissue samples were included in the present study.

Identification of differentially expressed genes.Differentially expressed genes(DEGs) were screened among the data using limma package. Adjusted P-values of each gene was calculated using false-discovery rate (FDR). FDR < 0.01 and |log FC|>3 was set as the cut-off threshold.

Development of the prognostic model.A total of 115 up-regulated genes were matched and extracted for the subsequent analysis. Survival analysis was then performed through COX regression and log-rank tests, and P values 0.01 was defined as statistically significant.Importantly, the LASSO Cox regression model is optimal selection for high-dimensional microarray data analysis, which shrinks the respective regression coefficients by a penalty proportional to the size of the variable, thereby screening the variables and finally providing a relatively small number of variables with a weights less than zero. Based on the prognostic related DEGs, which were significant in the univariate Cox regression analysis, further screening was extracted using the LASSO algorithm. Consequently, multivariate Cox regression analysis was performed,and nine mRNAs were ultimately included to construct a Prognosis model. Hazard ratios (HRs) were calculated for each gene. Moreover, we constructed a prognostic risk score model which was based on a linear combination of expression levels weighted by multivariate Cox regression analysis coefficients according the following formula:

risk score = $\text{Exp}_{gene1} \times \beta_{gene1} + \text{Exp}_{gene2} \times \beta_{gene2} + .. . + \text{Exp}_{genen} \times \beta_{genen;}$

where β representsthe HR of each gene. Next,239 patients were classified into two groups(high-risk group and low-risk group)by median risk score.Kaplan-Meier survival analysis and log-rank test wereperformed to compare the overall survival (OS) time.Survival curves were draws using Surviminer package. Furthermore, Receiver operating characteristic (ROC) curve were conducted to analyze the prognostic model. P < 0.05 was considered statistically significant. Finally, univariate and multivariate Cox regression analysis were performed to further investigate the independence of the prognostic gene signature to predict the outcomes of HCC patients.

Create a nomogram for the overall survival prediction in HCC.The prognostic gene signature and the risk factors that we analyzed were included to generate a nomogram to investigate the probability of 1-, 3-, 5- and 7-year overall survival of patients with HCC. Additionally, a calibration curve was plotted to evaluate whether the actual outcomes approximate to predicted outcomes.

Subjects and samples.6 human HCC sampleswere obtainedconsecutively from patients who diagnosed with HCC needed segmental hepatectomy in ShundeHospitalof Southern Medical University. This study was approved by the Ethics Institutional Review Boardof Shunde Hospital of Southern Medical University.

Isolation of total RNA and quantitative real-time polymerase chain reaction.Total RNA was extracted from HCC patients' liver sample with the TRIzol (15596026, Invitrogen, USA). Then total RNA was reversely transcribed with a HiScript II Q RT SuperMix kit (R222-01,Vazyme, China) and PCR amplification was performed by ChamQ SYBR qPCR Master Mix kit (Q311-02,Vazyme), GAPDH was regarded as internal controls. Primers sequences are listed in Table 1, which were synthesized by BGI technology (Guangdong, China). These reactions were conducted according to the manufacturer's instructions with CFX96 Real Time System (Bio-Rad, USA).

Table 1
Primer sequences for qPCR

Gene name		Primer sequences(5'-3')
IQGAP3	Forward:	GTGCAGCGGATCAACAAAGC
	Reverse:	ACGATGCAACAGGGTACACTG
BIRC5	Forward:	GAGGCTGGCTTCATCCACTG
	Reverse:	CCTTTGCATGGGGTCGTCAT
PTTG1	Forward:	ACCCGTGTGGTTGCTAAGG
	Reverse:	ACGTGGTGTTGAAACTTGAGAT
STC2	Forward:	GGGTGTGGCGTGTTTGAATG
	Reverse:	TTTCCAGCGTTGTGCAGAAAA
CDKN3	Forward:	GGACTCCTGACATAGCCAGC
	Reverse:	CTGTATTGCCCCGGATCCTC
PBK	Forward:	TACTATGCAGCGTTGGGAACT
	Reverse:	GCAGAAGGACGATCTTTAGGGTC
EXO1	Forward:	GCAGCCACGGCCTATTAAG
	Reverse:	CCACCACGTACTTGCCACT
NEIL3	Forward:	CAAGCGTCCTAATTGTGGTCA
	Reverse:	CCCTGCTAGATGTCCAACTGATT
HOXD9	Forward:	ACTACGGGATTAAGCCTGAAACC
	Reverse:	CAGGAACGAGTTGCACGAGAA
GAPDH	Forward:	GGAGCGAGATCCCTCCAAAAT
	Reverse:	GGCTGTTGTCATACTTCTCATGG

Statistical analysis.We compared two groups using the student's *t* test for continuous variables and χ2 test for categorical variables. For survival analyses, we used the Kaplan-Meier method to analyze the correlation between variables and disease free survival, and the log-rank test to compare survival curves. Package "survival", "glmnet", "rms", "rms", "pROC" were used for Cox regression analysis, LASSO Cox regression, the nomogram and calibration plot and ROC curve respectively, carrying out by R software (ver3.6.4)

Study approval. Human Investigations Committee of Shundehospital(Fo Shan) approve to use thediscard tissue of liver cancer patients for qPCR. Informed consent for the procurement and analysis of these

sampleswas also obtained.

Results

Data preprocessing

Firstly, the gene expression profile data and the corresponding clinical information were downloaded from TCGA. The gene expression profile data were merged with clinical information into a data frame for the next step. Each of the molecular profiles consisted of RNA-Seq expression data of 60,488 genes. The imported dataset consisted of 'expression,' representing the gene expression profiles of patients in terms of counts value (described in the Methods), 'clinical data,' representing clinical information related to the patients, and 'merged data'.

Identification of DEGs

After data preprocessing and normalization, the limma package in R language analysis was used to identify the DEGs. According to the threshold criteria (FDR <0.01 and |log FC|>3), we obtained 563 DEGs, including 448 downregulated (Table S1) and 115 upregulated (Table S2) genes.

Identification of nine genes associated with prognosis in HCC

To identify hub DEGs with potential prognostic value, univariate Cox regression analysis was applied, revealing that 41 genes were significantly correlated with overall survival (OS) ($P < 0.01$) (Table S3). Subsequently, the LASSO Cox regression model was used to build a prognostic classifier, and a total of nine genes, IQGAP3, BIRC5, PTTG1, STC2, CDKN3, PBK, EXO1, NEIL3, and HOXD9, were confirmed to be significantly associated with survival (Figure 1). The hazard ratio (HR) of each gene was calculated: IQGAP3, PTTG1, STC2, PBK, NEIL3, and HOXD9 had an HR > 0 and were considered high-risk factors associated with a short survival time; and BIRC5, CDKN3, and EXO1 had an HR < 0 and were considered protective factors associated with a long survival time (Figure 1). The expression levels of these nine genes in liver samples from HCC patients were measured by RT-qPCR (Figure 2), seven of which trended in the same direction. In addition, three (NEIL3, STC2, IQGAP3) were statistically significant.Next,Kaplan–Meier survival curves were generated to explore the potential roles of individual genes in OS. Among the nine-gene prognostic signature, IQGAP3, STC2, PBK, NEIL3, and HOXD9 were shown to significantly predict poor OS in the log-rank test ($P < 0.05$, Figure 3).

Development of a prognostic gene signature for survival prediction

Nine screened genes were selected to formulate a prognostic gene signature. According to the risk score, patients were separated into high-risk and low-risk groups. The heat map shows that the genes were upregulated with increasing risk scores (Figure 4A). The risk score distribution and survival status of each patient are shown in Figure 4B and 4C, respectively. The high-risk group had a significantly higher mortality rate (42.02%) than the low-risk group (12.50%); the median survival time in the high-risk group was 4.27 years and that in the low-risk group was 8.56 years. The log-rank test shows a significant

difference in mortality between the two groups (HR: 4.86; 95% CI: 2.72−8.69; P = 1.01E-08), and the Kaplan−Meier curve reveals that patients in the high-risk group had a poorer prognosis (P < 0.0001; Figure 5A). The AUC of the ROC curves was used to determine the predictive power of the prognostic gene signature; the AUC at 1, 3, 5, and 7 years was 0.84, 0.711, 0.812, and 0.829, respectively (Figure 5B). These results show that the prognostic gene signature can effectively predict OS in patients with HCC. Furthermore, univariate and multivariate Cox regression analyses were performed to investigate whether the gene signature was an independent predictive factor for prognosis in HCC patients. Following multivariate adjustment for conventional clinical patterns, including gender, age at diagnosis, pathological TNM, pathological grade and stage, vascular invasion, and serum AFP level, the gene signature remained a powerful and independent factor, confirming its robust ability to predict the OS in HCC patients (HR: 4.70; 95% CI: 2.61−8.38; P = 2.06E-07; Figure 6). In addition, stratified analysis was employed for further data mining, which found that the prognostic gene signature was a predictive marker for male patients with HCC (Figure 7A). Following stratification of the clinicopathological risk factors, the risk score based on the prognostic signature was an independent predictive indicator of pathological TNM, pathological grade and stage, vascular invasion, and serum AFP level, and patients with a high risk score had a poorer prognosis (Figure 7B−H). Thus, the prognostic gene signature can add OS predictive value to clinicopathological features.

Construction of a nomogram to evaluate the predictive accuracy of the prognostic gene signature

To provide the clinician with a quantitative method to predict an HCC patient's probability of OS, we constructed a nomogram by integrating both the prognostic gene signature and clinicopathological risk factors (Figure 8A). The results of our nomogram show that the prognostic gene signature had the greatest weight among the total points, consistent with previous multivariate regression analysis. The 30-sample bootstrapped calibration plots for the prediction of OS at 1, 3, 5, and 7 years are shown in Figure 8B−E, respectively, the results of which indicate good consistency between the predicted and observed outcomes. Consequently, our results suggest that the nomogram is an optimal model for the prediction of HCC prognosis as compared with individual risk factors.

Discussion

HCC involves a complex molecular landscape and has a 5-year OS of only 10−15%, mostly due to late diagnosis(22-24). Studies have shown that the current clinicopathological parameters, such as age at diagnosis, gender, viral infection, TNM stage, and serum AFP levels, cannot accurately predict prognosis in HCC patients(25-30). Thus, it is imperative to elucidate novel biomarkers to predict survival and monitor treatment response, which may also provide a link to specific therapeutic options for combating HCC in individual patients.

In the present study, we evaluated the DEGs using mRNA expression profiles, and 115 significantly upregulated genes were selected for further analysis. Subsequently, LASSO Cox regression analysis was used to develop a prognostic model for HCC, the results of which show that the prognostic gene signature

was significantly associated with OS in HCC patients. Moreover, an ROC was used as a metric for assessing the predictive performance of our model; the AUC at 1, 5, and 7 years was greater than 0.8. To date, the TNM staging system is the globally recognized Gold Standard for classifying the extent of cancer metastasis and is widely accepted as a tool for predicting prognosis and guiding treatment options for HCC. Univariate and multivariate Cox regression analyses were performed to explore the predictive power of our prognostic gene signature in combination with the TNM staging system and certain important clinical features. The results indicate that the prognostic gene signature can act as an independent predictive factor of OS following adjustment for conventional clinical characteristics. Our prognostic gene signature appears to potentially possess greater predictive power than that of traditional prognostic patterns. Furthermore, a nomogram combining our model with other important clinical patterns was created. Indeed, our model outperformed all other common clinicopathological features in predicting survival. Based on the calibration plot, there was good consistency between the actual and predicted values for the 1-, 3-, and 5-year OS; therefore, our prognostic gene signature predicts survival in HCC and provides a personalized score for individual patients. Accordingly, our nomogram is a valuable novel prognostic method for clinicians in the future.

Previous studies have identified the biological function of multiple genes involved in the development of HCC, including those used in our model. STC2 overexpression can promote proliferation of HCC, ovarian cancer, and stem cells(31-33). Similarly, PBK promotes metastasis of HCC by activating the ETV4-uPAR signaling pathway(34). PTTG1 has been reported to be associated with poor prognosis and angiogenesis in HCC(35). IQGAP3 functions as a potential oncogene by activating metastasis and EMT in HCC(36). CDKN3 has been correlated with poor tumor differentiation and advanced tumor stage by promoting cell proliferation in HCC(37). HOXD9 acts as a tumor promoter to drive metastasis by regulating ZEB1 in HCC(38), and EXO1 overexpression is associated with poor prognosis in HCC patients(39); however, there have been few reports regarding the effect of NEIL3 on HCC. Two selected genes show different trends in comparison with TCGA data, likely due to race and sample size, which needs further confirmation using larger datasets. While these studies have profoundly improved our understanding of the progression of HCC, less is known regarding how such biomarkers impact patient outcome. It appears that such individual biomarkers are limited because they can be influenced by many factors, rendering their predictive ability unstable; thus, other recent studies have shown that combined analysis of a panel of biomarkers is the most promising approach to change clinical management(40, 41). With respect to HCC, since the accuracy of various prognostic models still varies, the field currently warrants further research(19-21). In the present study, the LASSO Cox regression model was used to select the most useful prognostic markers from many genes related to HCC. Furthermore, the LASSO Cox regression model allowed us to integrate multiple genes into one signature, with the AUC at 1, 5, and 7 years exceeding 80%, displaying significantly greater prognostic accuracy than previous studies and indicating that it can be potentially applied as an adjuvant therapy. Indeed, we present an optimal prognostic model for the first time; however, despite the fact that our research highlights novel ideas related to the prognostic assessment of HCC, it still has some limitations. Firstly, our research is based on a single-omics technique (RNA-Seq); therefore, it may have different heterogeneity due to other omics data

platforms. Secondly, our model does not have external data verification. Thirdly, the biological functions and molecular mechanisms of the nine indicator genes have been further evaluated with RNA expression to accelerate their clinical verification in HCC.

Conclusion

In summary, our findings indicate that the nine-gene prognostic signature can classify patients with HCC into low-risk and high-risk groups with respect to mortality. In comparison with other clinicopathological features, our prognostic gene signature proposes a more accurate assessment of prognosis. Moreover, we created a nomogram that numerically predicts the OS of individual HCC patients based on the prognostic gene signature and certain important clinicopathological features. This information can be used to predict patient prognosis and make individualized decisions regarding treatment and surveillance.

Declarations

Ethical Approval and Consent to participate

Human Investigations Committee of Shundehospital (Fo Shan) approve to use thediscard tissue of liver cancer patients for qPCR. Informed consent for the procurement and analysis of these sampleswas also obtained.

Author contributions

 Conception and design: Cuiyun Wu, Yinghui Chen, Hongling Qu, Lin Zheng, Jie Yao

 Data analysis and interpretation: Cuiyun Wu, Yinghui Chen, Hongling Qu, Lin Zheng Manuscript writing: Jie Yao

 RT-PCR: Yaosheng Luo

Consent for publication

Informed consent for publication was obtained from all participants.

Availability of supporting data

All the data used in the present study could be downloaded from TCGA (https://cancergenome.nih.gov/).

Competing interests

The authors have declared that no conflict of interest exists.

References

1. Llovet JM, Zucman-Rossi J, Pikarsky E, et al.: Hepatocellular carcinoma. Nature reviews. Disease primers 2: 16018, 2016.

2. Villanueva A: Hepatocellular Carcinoma. N Engl J Med 380: 1450-1462, 2019.

3. Villa E, Moles A, Ferretti I, et al.: Natural history of inoperable hepatocellular carcinoma: estrogen receptors' status in the tumor is the strongest prognostic factor for survival. Hepatology 32: 233-238, 2000.

4. Kao WY, Chiou YY, Hung HH, et al.: Younger hepatocellular carcinoma patients have better prognosis after percutaneous radiofrequency ablation therapy. Journal of clinical gastroenterology 46: 62-70, 2012.

5. Su CW, Lei HJ, Chau GY, et al.: The effect of age on the long-term prognosis of patients with hepatocellular carcinoma after resection surgery: a propensity score matching analysis. Archives of surgery (Chicago, Ill. : 1960) 147: 137-144, 2012.

6. Li T, Wang SK, Zhou J, et al.: Positive HBcAb is associated with higher risk of early recurrence and poorer survival after curative resection of HBV-related HCC. Liver Int 36: 284-292, 2016.

7. Yu SJ and Kim YJ: Hepatitis B viral load affects prognosis of hepatocellular carcinoma. World journal of gastroenterology 20: 12039-12044, 2014.

8. Sobin LH and Compton CC: TNM seventh edition: what's new, what's changed: communication from the International Union Against Cancer and the American Joint Committee on Cancer. Cancer 116: 5336-5339, 2010.

9. Zhu WJ, Huang CY, Li C, et al.: Risk factors for early recurrence of HBV-related hepatocellular carcinoma meeting milan criteria after curative resection. Asian Pacific journal of cancer prevention : APJCP 14: 7101-7106, 2013.

10. Peng SY, Chen WJ, Lai PL, Jeng YM, Sheu JC and Hsu HC: High alpha-fetoprotein level correlates with high stage, early recurrence and poor prognosis of hepatocellular carcinoma: significance of hepatitis virus infection, age, p53 and beta-catenin mutations. Int J Cancer 112: 44-50, 2004.

11. Arii S, Yamaoka Y, Futagawa S, et al.: Results of surgical and nonsurgical treatment for small-sized hepatocellular carcinomas: a retrospective and nationwide survey in Japan. The Liver Cancer Study Group of Japan. Hepatology 32: 1224-1229, 2000.

12. Giordano TJ: The cancer genome atlas research network: a sight to behold. Endocrine pathology 25: 362-365, 2014.

13. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV and Fotiadis DI: Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal 13: 8-17,

2015.

14. Huang W, Skanderup AJ and Lee CG: Advances in genomic hepatocellular carcinoma research. GigaScience 72018.

15. Rebouissou S and Nault JC: Advances in molecular classification and precision oncology in hepatocellular carcinoma. Journal of hepatology 72: 215-229, 2020.

16. Calderaro J, Ziol M, Paradis V and Zucman-Rossi J: Molecular and histological correlations in liver cancer. Journal of hepatology 71: 616-630, 2019.

17. Toh TB, Lim JJ and Chow EK: Epigenetics of hepatocellular carcinoma. Clinical and translational medicine 8: 13, 2019.

18. Venook AP, Niedzwiecki D, Lopatin M, et al.: Biologic determinants of tumor recurrence in stage II colon cancer: validation study of the 12-gene recurrence score in cancer and leukemia group B (CALGB) 9581. J Clin Oncol 31: 1775-1781, 2013.

19. Liu H, Yan Y, Chen R, et al.: Integrated nomogram based on five stage-related genes and TNM stage to predict 1-year recurrence in hepatocellular carcinoma. Cancer cell international 20: 140, 2020.

20. Li G, Xu W, Zhang L, et al.: Development and validation of a CIMP-associated prognostic model for hepatocellular carcinoma. EBioMedicine 47: 128-141, 2019.

21. Yang Z, Yang Y, Zhou G, et al.: The Prediction of Survival in Hepatocellular Carcinoma Based on A Four Long Non-coding RNAs Expression Signature. J Cancer 11: 4132-4144, 2020.

22. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. Journal of hepatology 56: 908-943, 2012.

23. Marrero JA, Kulik LM, Sirlin CB, et al.: Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. 68: 723-750, 2018.

24. Omata M, Cheng AL, Kokudo N, et al.: Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. Hepatology international 11: 317-370, 2017.

25. Trevisani F, D'Intino PE, Morselli-Labate AM, et al.: Serum alpha-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. Journal of hepatology 34: 570-575, 2001.

26. Zhou L, Liu J and Luo F: Serum tumor markers for detection of hepatocellular carcinoma. World journal of gastroenterology 12: 1175-1181, 2006.

27. Zhou L, Rui JA, Wang SB, Chen SG and Qu Q: LCSGJ-T classification, 6th or 5th edition TNM staging did not independently predict the long-term prognosis of HBV-related hepatocellular carcinoma after radical hepatectomy. The Journal of surgical research 159: 538-544, 2010.

28. Pawlik TM, Poon RT, Abdalla EK, et al.: Hepatitis serology predicts tumor and liver-disease characteristics but not prognosis after resection of hepatocellular carcinoma. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract 8: 794-804; discussion 804-795, 2004.

29. Tan JT, Zhao C, Peng NF, et al.: Association between age and overall survival of patients with hepatocellular carcinoma after hepatic resection. Journal of surgical oncology 114: 966-970, 2016.

30. Tangkijvanich P, Mahachai V, Suwangool P and Poovorawan Y: Gender difference in clinicopathologic features and survival of patients with hepatocellular carcinoma. World journal of gastroenterology 10: 1547-1550, 2004.

31. Wang H, Wu K, Sun Y, et al.: STC2 is upregulated in hepatocellular carcinoma and promotes cell proliferation and migration in vitro. BMB reports 45: 629-634, 2012.

32. Wang Y, Gao Y, Cheng H, Yang G and Tan W: Stanniocalcin 2 promotes cell proliferation and cisplatin resistance in cervical cancer. Biochem Biophys Res Commun 466: 362-368, 2015.

33. Kim PH, Na SS, Lee B, Kim JH and Cho JY: Stanniocalcin 2 enhances mesenchymal stem cell survival by suppressing oxidative stress. BMB reports 48: 702-707, 2015.

34. Yang QX, Zhong S, He L, et al.: PBK overexpression promotes metastasis of hepatocellular carcinoma via activating ETV4-uPAR signaling pathway. Cancer Lett 452: 90-102, 2019.

35. Fujii T, Nomoto S, Koshikawa K, et al.: Overexpression of pituitary tumor transforming gene 1 in HCC is associated with angiogenesis and poor prognosis. Hepatology 43: 1267-1275, 2006.

36. Shi Y, Qin N, Zhou Q, et al.: Role of IQGAP3 in metastasis and epithelial-mesenchymal transition in human hepatocellular carcinoma. J Transl Med 15: 176, 2017.

37. Xing C, Xie H, Zhou L, et al.: Cyclin-dependent kinase inhibitor 3 is overexpressed in hepatocellular carcinoma and promotes tumor cell proliferation. Biochem Biophys Res Commun 420: 29-35, 2012.

38. Lv X, Li L, Lv L, et al.: HOXD9 promotes epithelial-mesenchymal transition and cancer metastasis by ZEB1 regulation in hepatocellular carcinoma. J Exp Clin Cancer Res 34: 133, 2015.

39. Dai Y, Tang Z and Yang Z: EXO1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients. 17: 2386-2397, 2018.

40. Birkhahn M, Mitra AP and Cote RJ: Molecular markers for bladder cancer: the road to a multimarker approach. Expert review of anticancer therapy 7: 1717-1727, 2007.

41. Croner RS, Förtsch T, Brückl WM, et al.: Molecular signature for lymphatic metastasis in colorectal carcinomas. Annals of surgery 247: 803-810, 2008.
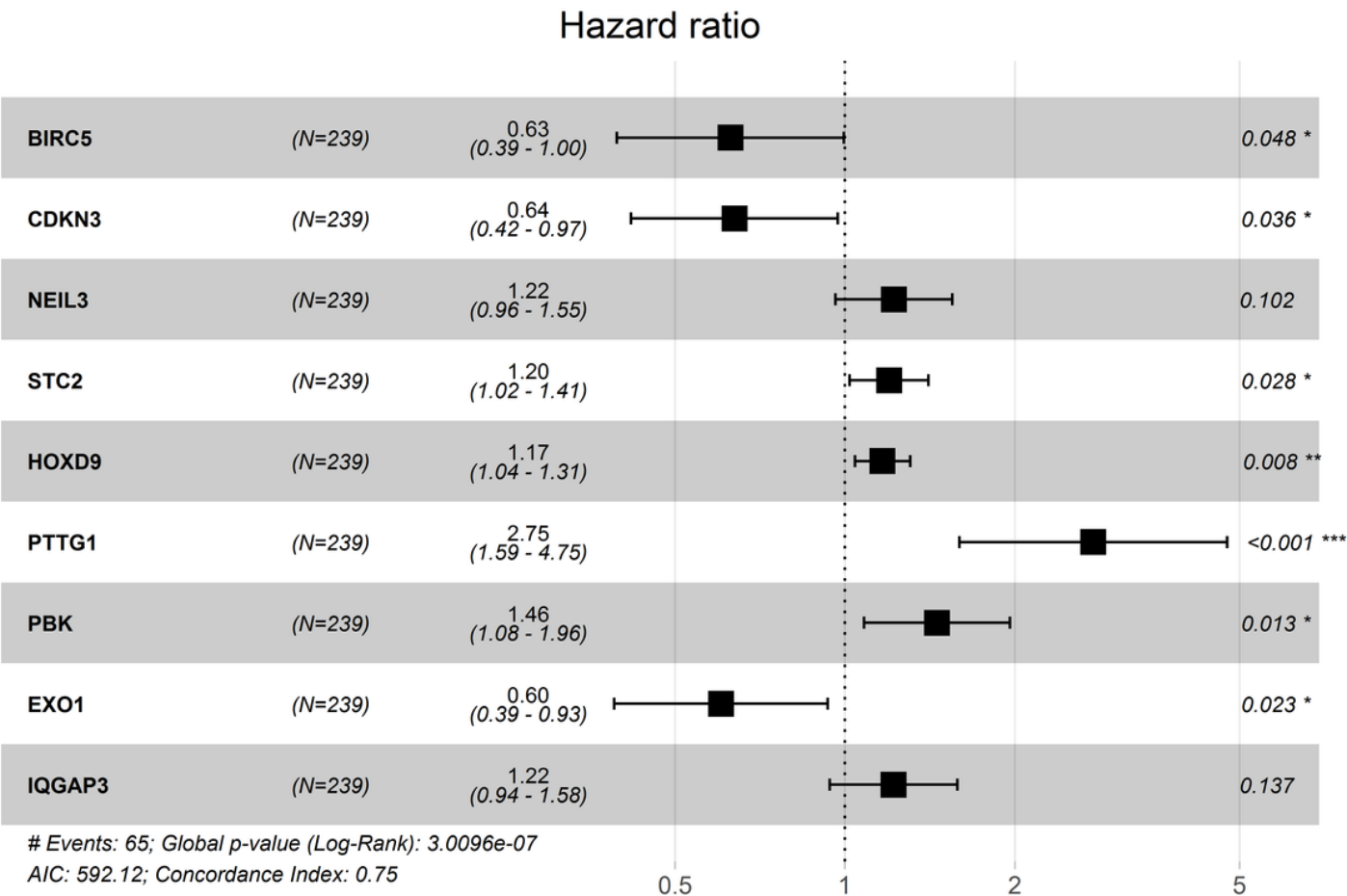
Figures

		Hazard ratio		
BIRC5	(N=239)	0.63 (0.39 - 1.00)		0.048 *
CDKN3	(N=239)	0.64 (0.42 - 0.97)		0.036 *
NEIL3	(N=239)	1.22 (0.96 - 1.55)		0.102
STC2	(N=239)	1.20 (1.02 - 1.41)		0.028 *
HOXD9	(N=239)	1.17 (1.04 - 1.31)		0.008 **
PTTG1	(N=239)	2.75 (1.59 - 4.75)		<0.001 ***
PBK	(N=239)	1.46 (1.08 - 1.96)		0.013 *
EXO1	(N=239)	0.60 (0.39 - 0.93)		0.023 *
IQGAP3	(N=239)	1.22 (0.94 - 1.58)		0.137

Events: 65; Global p-value (Log-Rank): 3.0096e-07
AIC: 592.12; Concordance Index: 0.75

Figure 1

Multivariate cox Regression analysis of the ninegenes.

Hazard ratio

Gene		HR (95% CI)	p-value
BIRC5	(N=239)	0.63 (0.39 - 1.00)	0.048 *
CDKN3	(N=239)	0.64 (0.42 - 0.97)	0.036 *
NEIL3	(N=239)	1.22 (0.96 - 1.55)	0.102
STC2	(N=239)	1.20 (1.02 - 1.41)	0.028 *
HOXD9	(N=239)	1.17 (1.04 - 1.31)	0.008 **
PTTG1	(N=239)	2.75 (1.59 - 4.75)	<0.001 ***
PBK	(N=239)	1.46 (1.08 - 1.96)	0.013 *
EXO1	(N=239)	0.60 (0.39 - 0.93)	0.023 *
IQGAP3	(N=239)	1.22 (0.94 - 1.58)	0.137

\# Events: 65; Global p-value (Log-Rank): 3.0096e-07
AIC: 592.12; Concordance Index: 0.75

Figure 1

Multivariate cox Regression analysis of the ninegenes.

Hazard ratio

Gene	N	HR (95% CI)		p-value
BIRC5	(N=239)	0.63 (0.39 - 1.00)		0.048 *
CDKN3	(N=239)	0.64 (0.42 - 0.97)		0.036 *
NEIL3	(N=239)	1.22 (0.96 - 1.55)		0.102
STC2	(N=239)	1.20 (1.02 - 1.41)		0.028 *
HOXD9	(N=239)	1.17 (1.04 - 1.31)		0.008 **
PTTG1	(N=239)	2.75 (1.59 - 4.75)		<0.001 ***
PBK	(N=239)	1.46 (1.08 - 1.96)		0.013 *
EXO1	(N=239)	0.60 (0.39 - 0.93)		0.023 *
IQGAP3	(N=239)	1.22 (0.94 - 1.58)		0.137

Events: 65; Global p-value (Log-Rank): 3.0096e-07
AIC: 592.12; Concordance Index: 0.75

Figure 1

Multivariate cox Regression analysis of the ninegenes.

Hazard ratio

Gene		HR (95% CI)		p-value
BIRC5	(N=239)	0.63 (0.39 - 1.00)		0.048 *
CDKN3	(N=239)	0.64 (0.42 - 0.97)		0.036 *
NEIL3	(N=239)	1.22 (0.96 - 1.55)		0.102
STC2	(N=239)	1.20 (1.02 - 1.41)		0.028 *
HOXD9	(N=239)	1.17 (1.04 - 1.31)		0.008 **
PTTG1	(N=239)	2.75 (1.59 - 4.75)		<0.001 ***
PBK	(N=239)	1.46 (1.08 - 1.96)		0.013 *
EXO1	(N=239)	0.60 (0.39 - 0.93)		0.023 *
IQGAP3	(N=239)	1.22 (0.94 - 1.58)		0.137

 # Events: 65; Global p-value (Log-Rank): 3.0096e-07
AIC: 592.12; Concordance Index: 0.75

0.5   1   2   5

Figure 1

Multivariate cox Regression analysis of the ninegenes.

Figure 2

Expression levels of the genes.(A-I) The mRNA expression in HCC patients were detected by RT-qPCR. *, p<0.05, ns no significance; CA, carcinoma tissue; PC, para-noncancerous tissue.
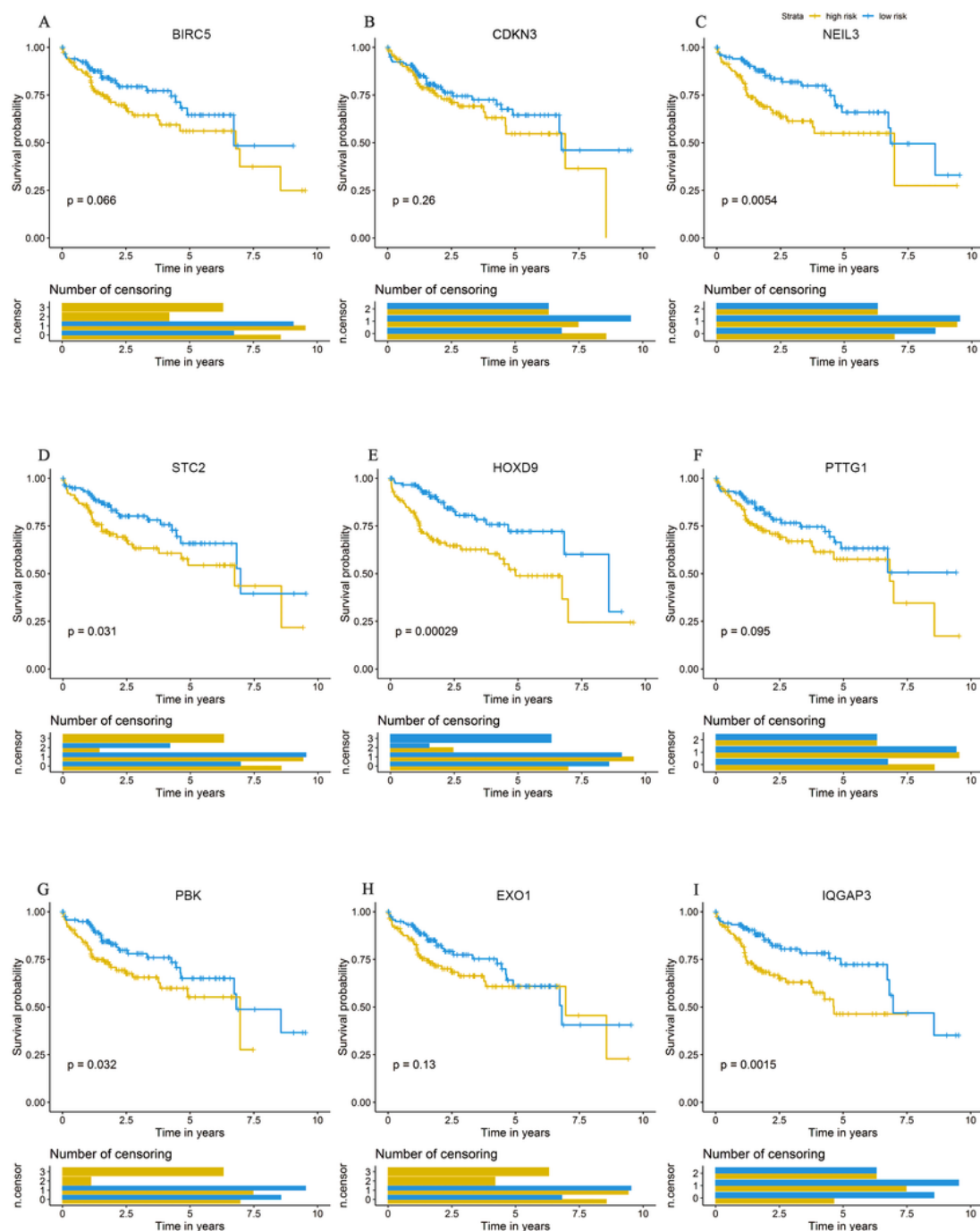
Figure 2

Expression levels of the genes.(A-I) The mRNA expression in HCC patients were detected by RT-qPCR. *, p<0.05, ns no significance; CA, carcinoma tissue; PC, para-noncancerous tissue.

Figure 2

Expression levels of the genes.(A-I) The mRNA expression in HCC patients were detected by RT-qPCR. *, p<0.05, ns no significance; CA, carcinoma tissue; PC, para-noncancerous tissue.

Figure 2

Expression levels of the genes.(A-I) The mRNA expression in HCC patients were detected by RT-qPCR. *, p<0.05, ns no significance; CA, carcinoma tissue; PC, para-noncancerous tissue.

Figure 3

Kaplan-Meier survival curves of genes.The yellow line represents the high-risk group and the blue represents the low risk-group.p<0.05 in Log-rank test.
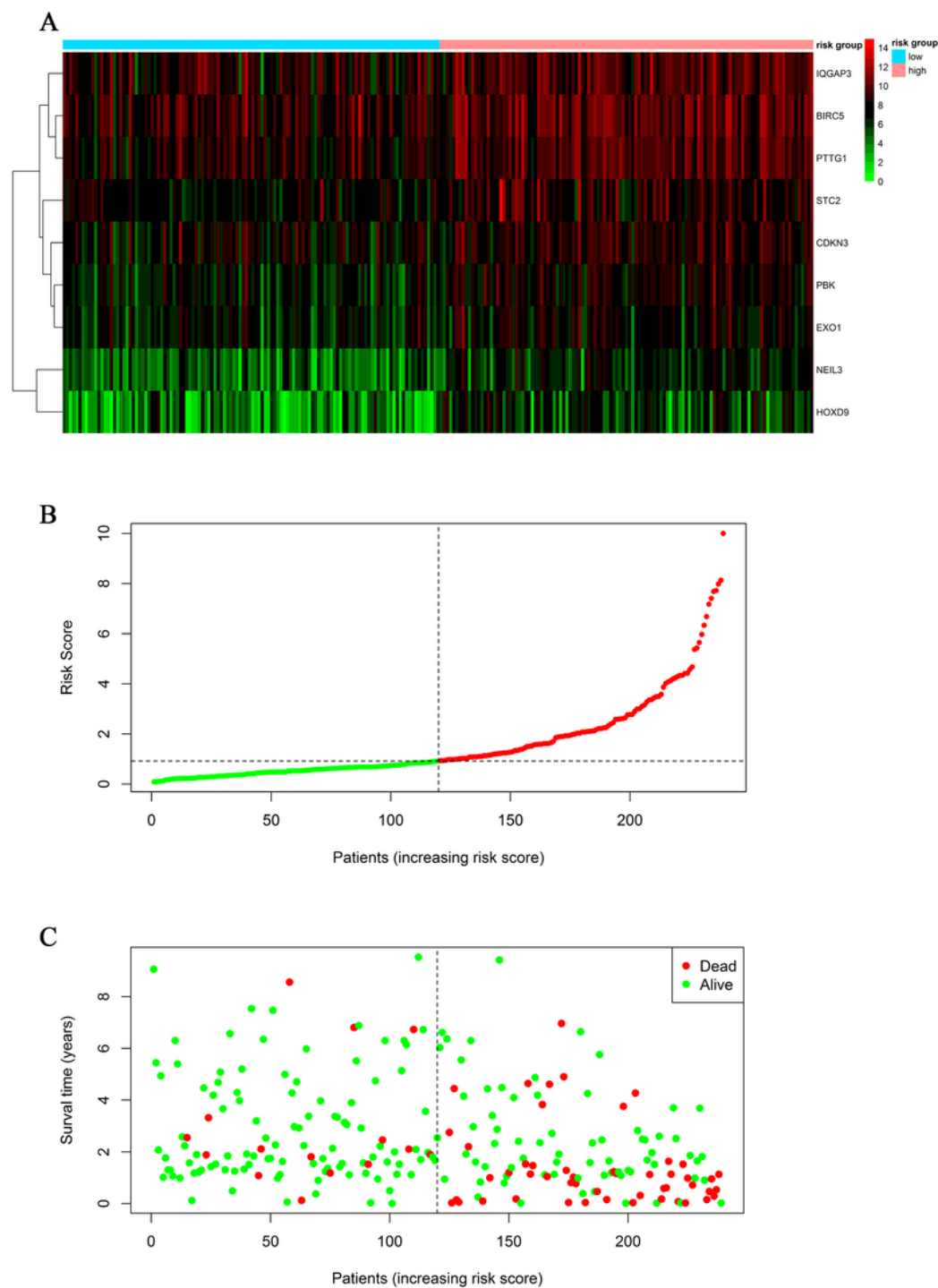
Figure 3

Kaplan-Meier survival curves of genes.The yellow line represents the high-risk group and the blue represents the low risk-group.p<0.05 in Log-rank test.

Figure 3

Kaplan-Meier survival curves of genes.The yellow line represents the high-risk group and the blue represents the low risk-group.p<0.05 in Log-rank test.

Figure 3

Kaplan-Meier survival curves of genes.The yellow line represents the high-risk group and the blue represents the low risk-group.p<0.05 in Log-rank test.
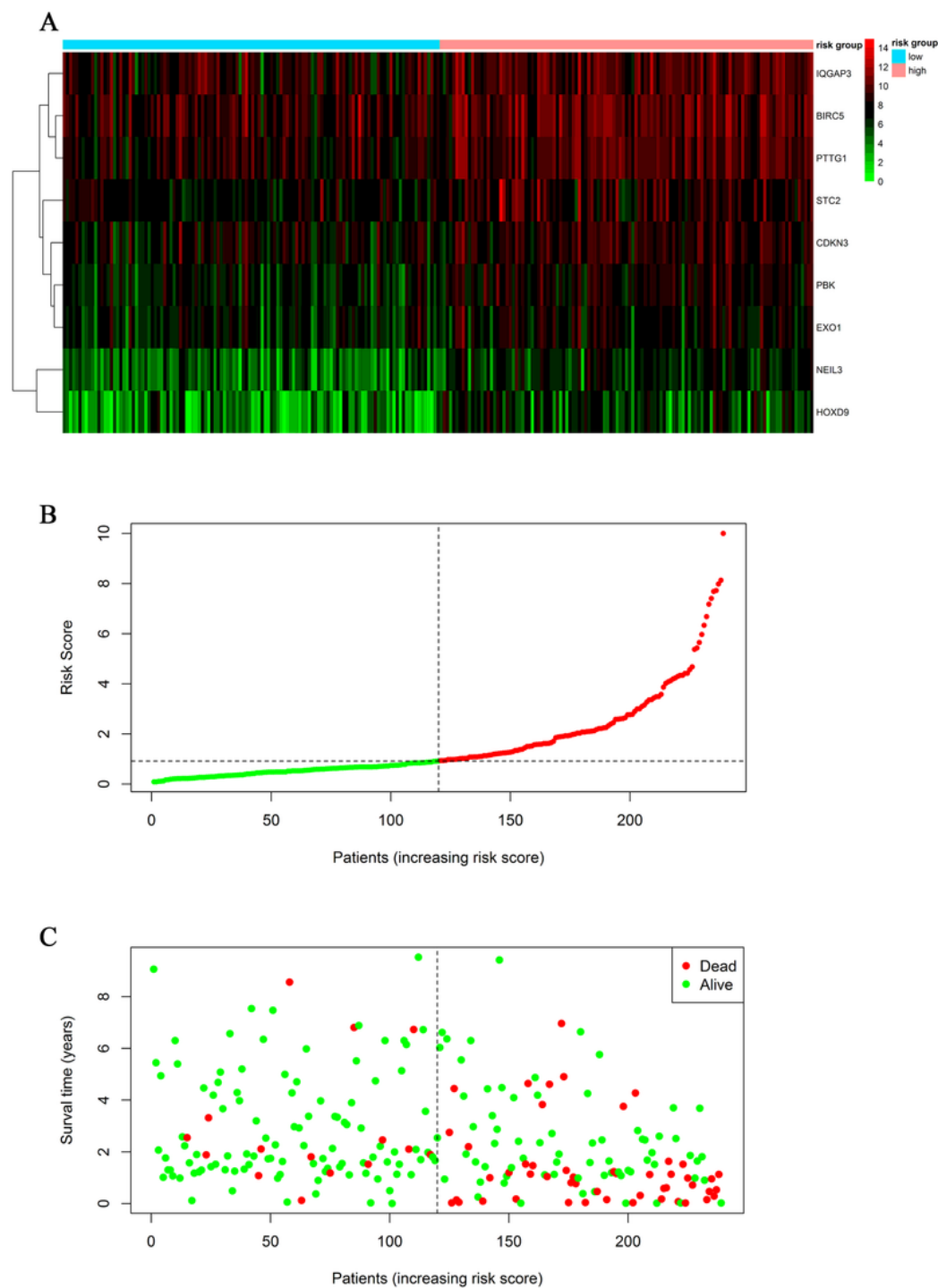

Figure 4

Risk score analysis of the prognostic gene signature of HCC. (A) The heat map of nine genes in HCC patients. Eachcolumn represents a patient and each rowrepresents a gene. The expression levels of genes are displayed in different colors. From green to red, the expression is gradually increasing.(B)The

distribution of high and low risk scores of nine genes. (C) Survival time and status of patients based on the high and low risk scores of ninegenes.
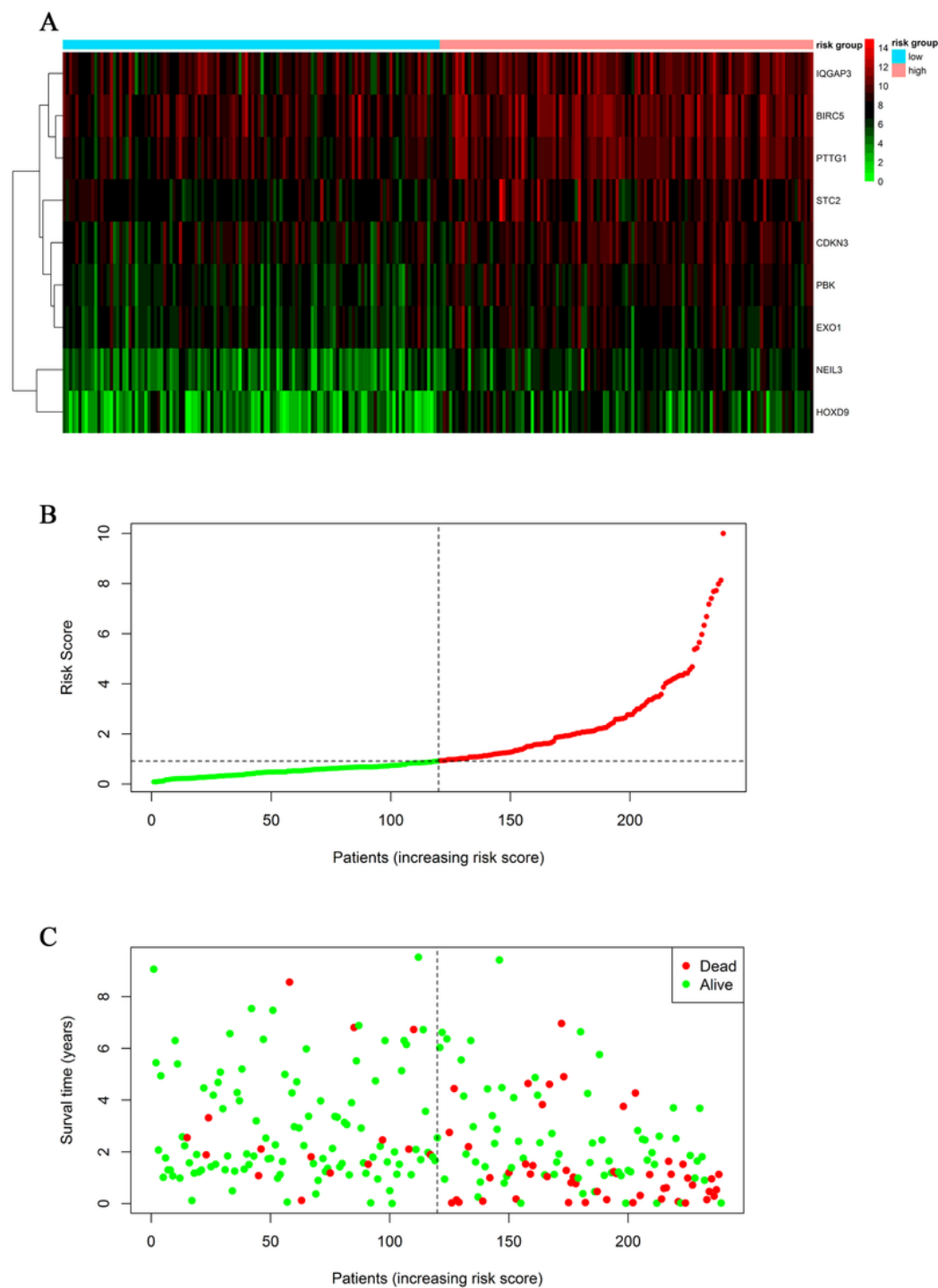

Figure 4

Risk score analysis of the prognostic gene signature of HCC. (A) The heat map of nine genes in HCC patients. Eachcolumn represents a patient and each rowrepresents a gene. The expression levels of genes are displayed in different colors. From green to red, the expression is gradually increasing.(B)The

distribution of high and low risk scores of nine genes. (C) Survival time and status of patients based on the high and low risk scores of ninegenes.
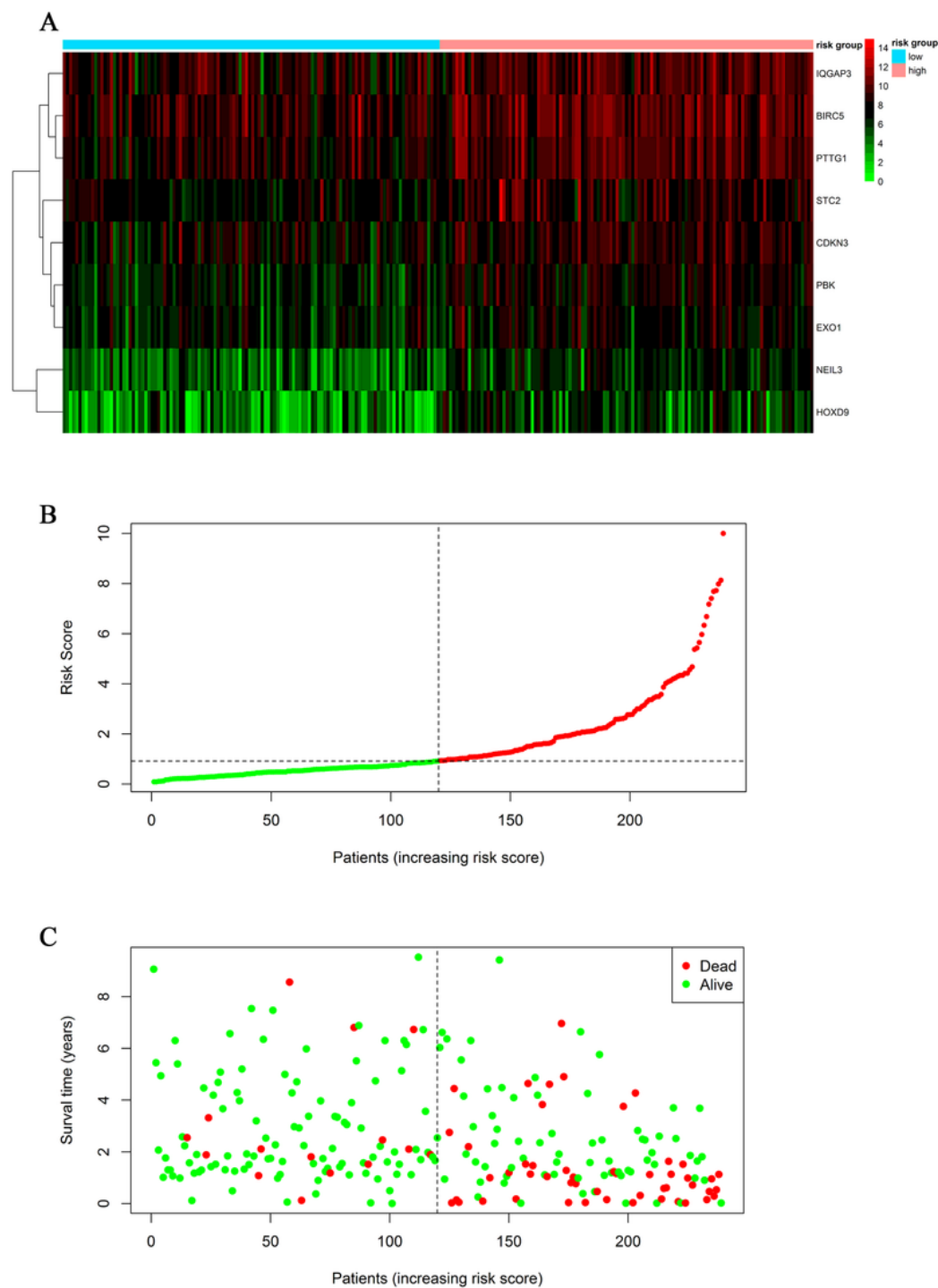

Figure 4

Risk score analysis of the prognostic gene signature of HCC. (A) The heat map of nine genes in HCC patients. Eachcolumn represents a patient and each rowrepresents a gene. The expression levels of genes are displayed in different colors. From green to red, the expression is gradually increasing.(B)The

distribution of high and low risk scores of nine genes. (C) Survival time and status of patients based on the high and low risk scores of ninegenes.

Figure 4

Risk score analysis of the prognostic gene signature of HCC. (A) The heat map of nine genes in HCC patients. Eachcolumn represents a patient and each rowrepresents a gene. The expression levels of genes are displayed in different colors. From green to red, the expression is gradually increasing.(B)The

distribution of high and low risk scores of nine genes. (C) Survival time and status of patients based on the high and low risk scores of ninegenes.
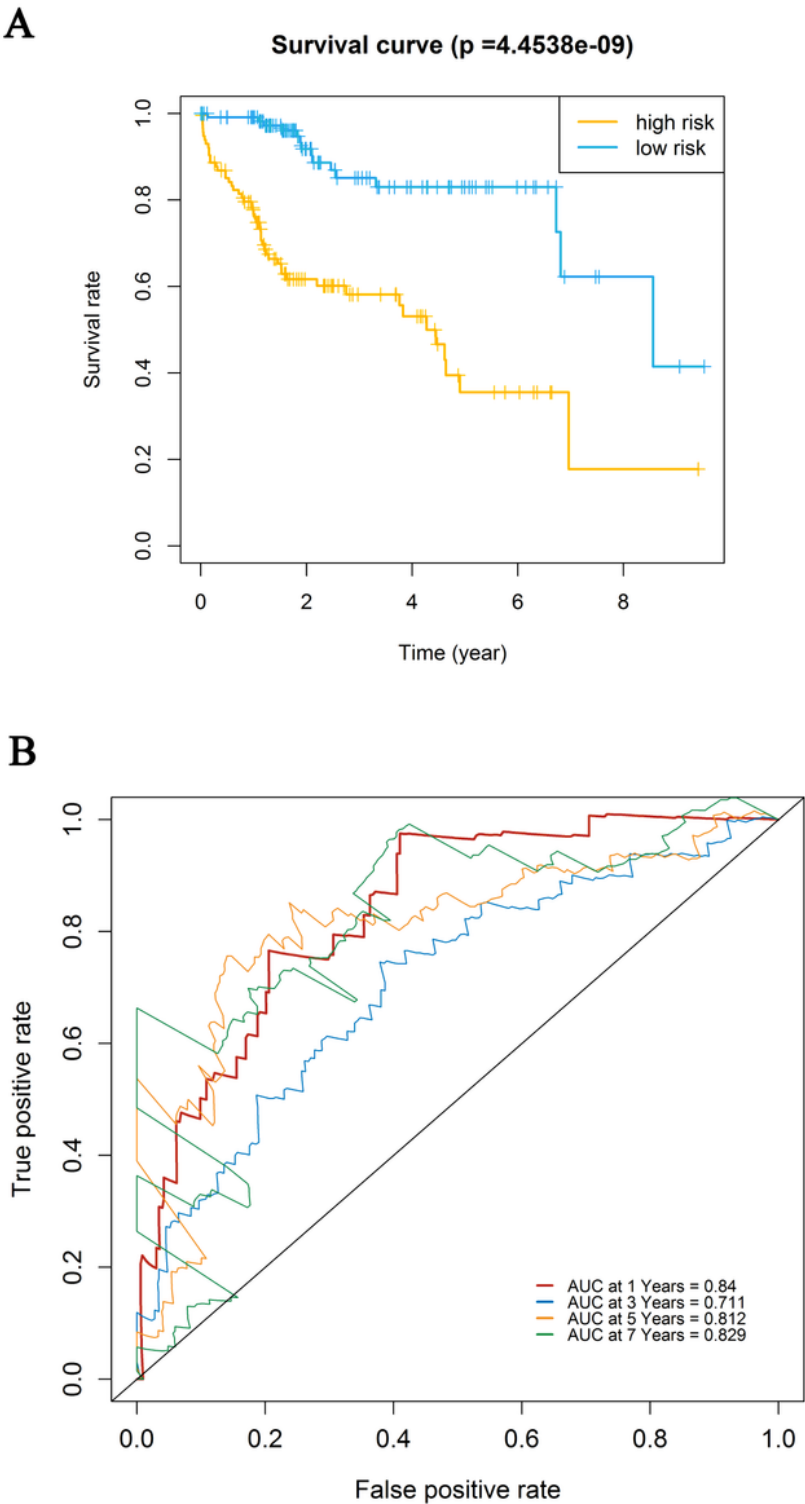

Figure 5

(A)Kaplan-Meier survival and(B) time-dependent ROC curves by theprognosis gene signature.

Figure 5

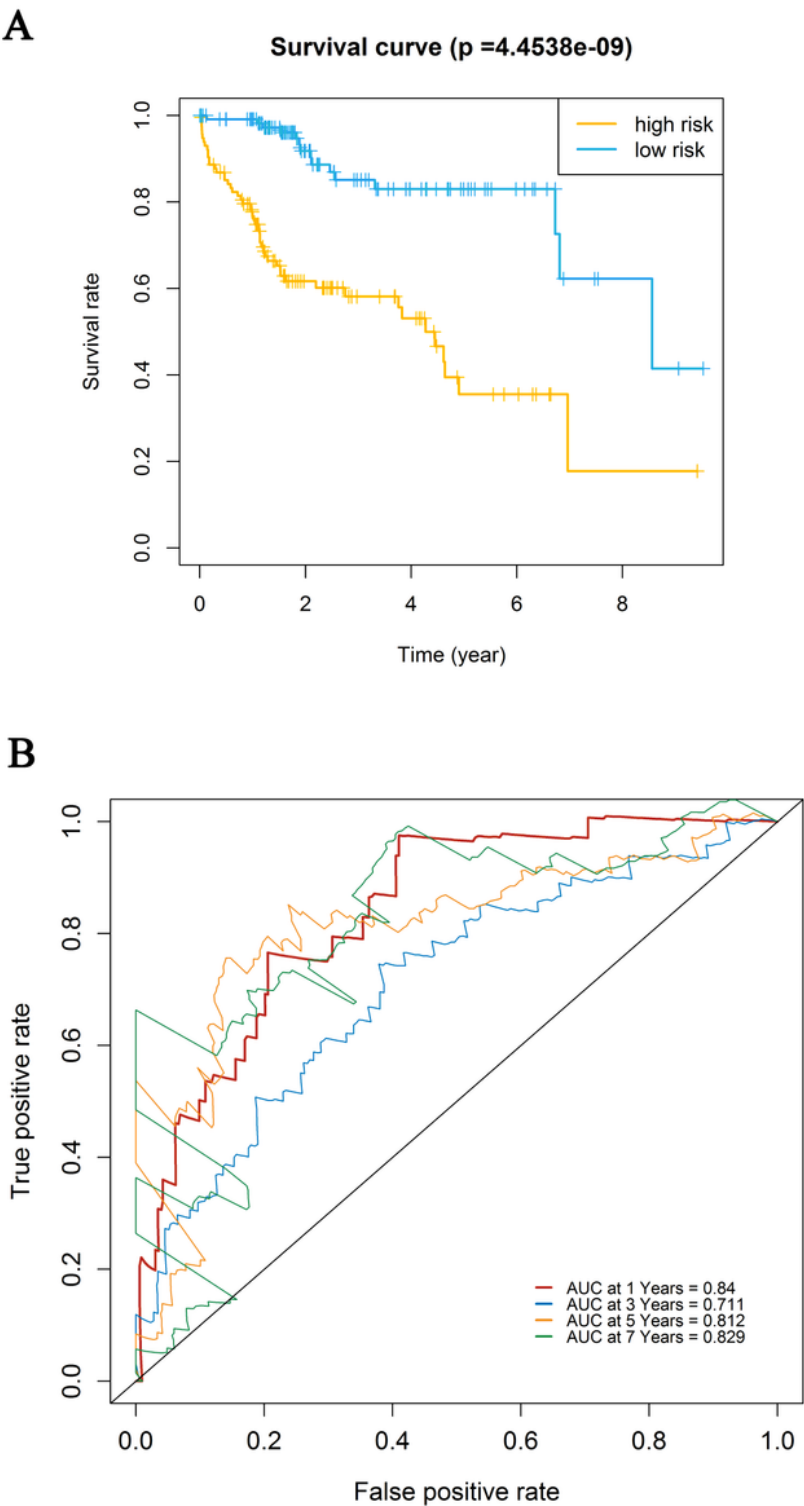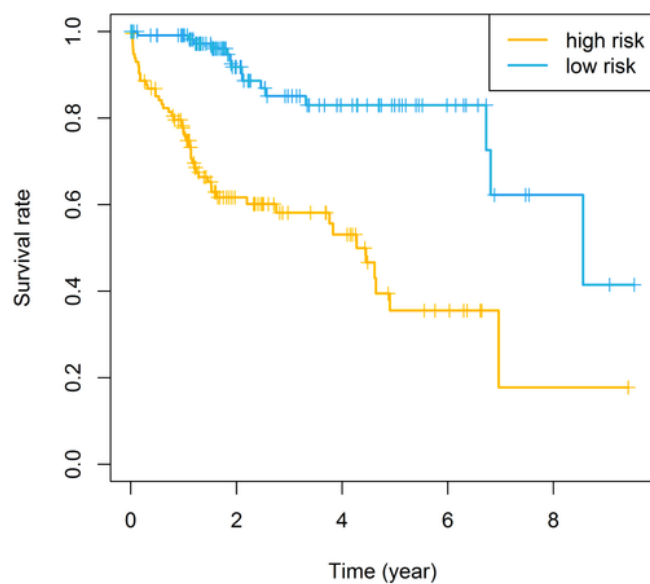(A)Kaplan-Meier survival and(B) time-dependent ROC curves by theprognosis gene signature.
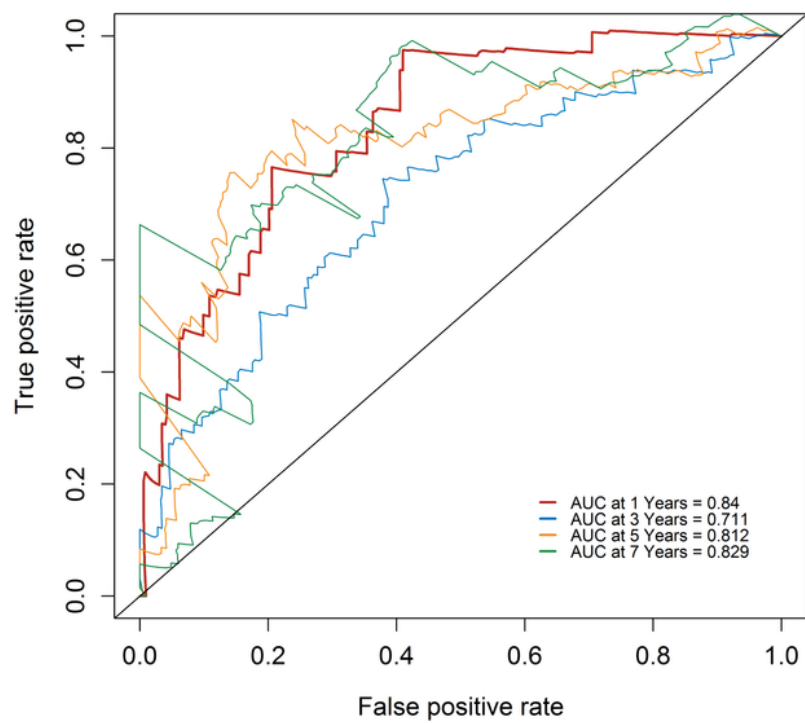
A

Survival curve (p =4.4538e-09)

B

Figure 5

(A)Kaplan-Meier survival and(B) time-dependent ROC curves by theprognosis gene signature.

A

B

Figure 5

(A)Kaplan-Meier survival and(B) time-dependent ROC curves by theprognosis gene signature.

A

Hazard ratio

age	(N=239)	1.49 (0.91 - 2.44)		0.11
apt	(N=239)	1.51 (0.92 - 2.47)		0.099
grade	(N=239)	1.35 (0.83 - 2.21)		0.226
M	(N=239)	2.54 (1.47 - 4.36)		<0.001 ***
N	(N=239)	1.75 (1.00 - 3.07)		0.051
T	(N=239)	1.97 (1.16 - 3.34)		0.012 *
vascular	(N=239)	1.54 (0.92 - 2.57)		0.099
gender	(N=239)	0.58 (0.35 - 0.95)		0.029 *
stage	(N=239)	2.05 (1.22 - 3.45)		0.007 **
risk	(N=239)	4.86 (2.72 - 8.69)		<0.001 ***

\# Events: 65

B

Hazard ratio

pathologic_M	(N=239)	2.3 (1.3 - 4.0)		0.003 **
Stage	(N=239)	1.7 (1.0 - 3.0)		0.039 *
Prognostic_model	(N=239)	4.7 (2.6 - 8.4)		<0.001 ***

\# Events: 65; Global p-value (Log-Rank): 1.9107e-10
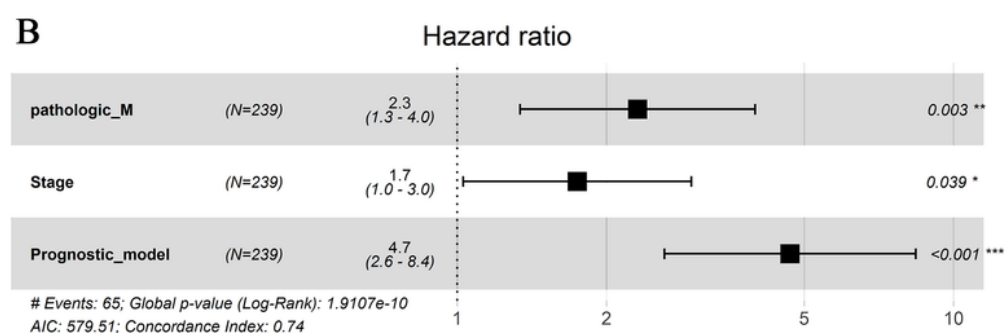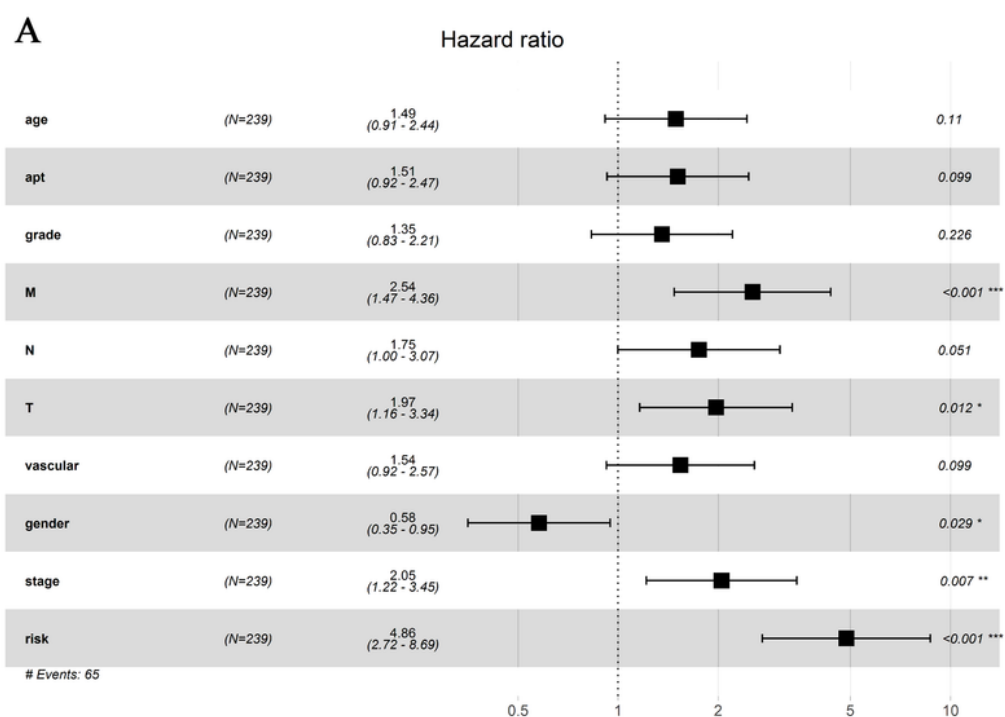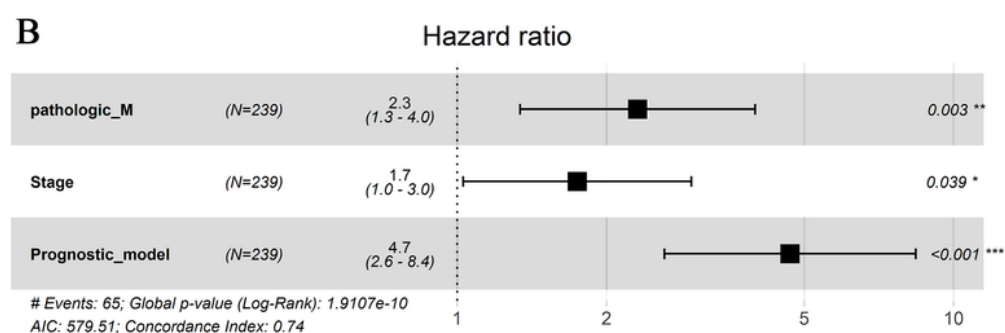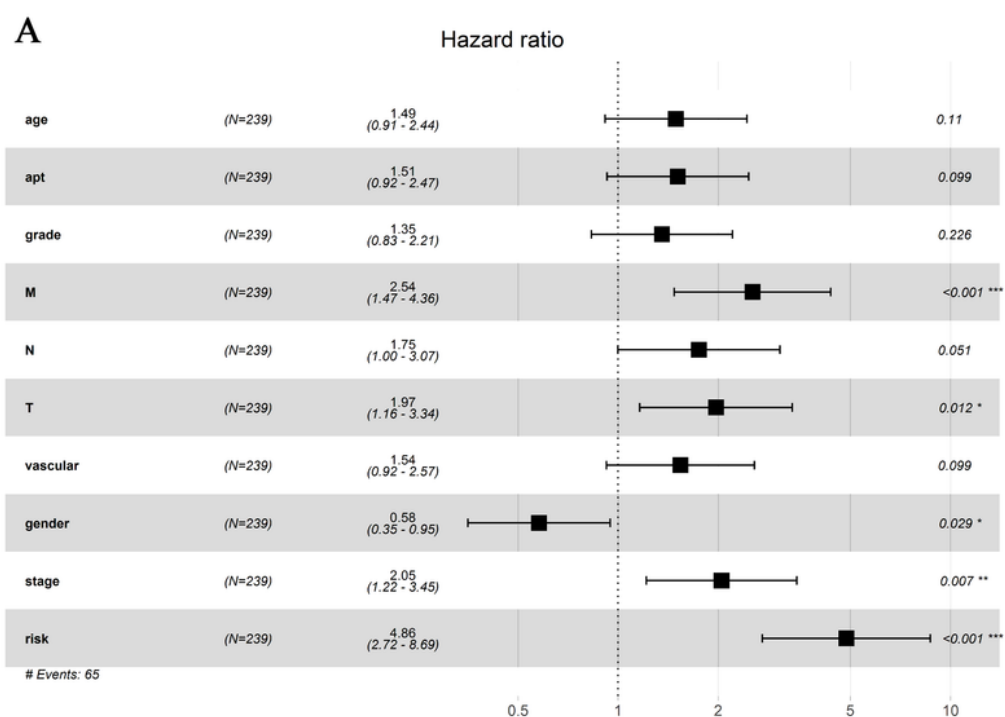AIC: 579.51; Concordance Index: 0.74

Figure 6

Univariate and multivariate Cox regression analysis of OS. Clinical patterns including gender, diagnostic age, pathologic TNM, pathologic grade, pathologic stage, vascular invasion, and serum AFP level in HCC patients. (A) univariate Cox regression analysis of OS. (B) multivariate Cox regression analysis of OS.
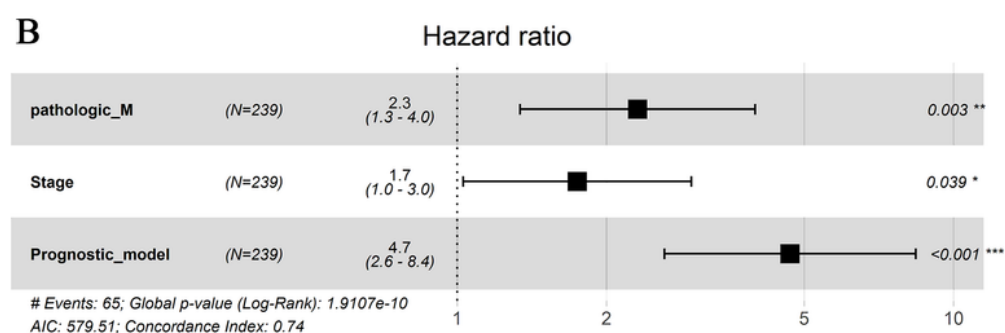
A

Hazard ratio

age	(N=239)	1.49 (0.91 - 2.44)		0.11
apt	(N=239)	1.51 (0.92 - 2.47)		0.099
grade	(N=239)	1.35 (0.83 - 2.21)		0.226
M	(N=239)	2.54 (1.47 - 4.36)		<0.001 ***
N	(N=239)	1.75 (1.00 - 3.07)		0.051
T	(N=239)	1.97 (1.16 - 3.34)		0.012 *
vascular	(N=239)	1.54 (0.92 - 2.57)		0.099
gender	(N=239)	0.58 (0.35 - 0.95)		0.029 *
stage	(N=239)	2.05 (1.22 - 3.45)		0.007 **
risk	(N=239)	4.86 (2.72 - 8.69)		<0.001 ***

\# Events: 65

0.5   1   2   5   10

B

Hazard ratio

pathologic_M	(N=239)	2.3 (1.3 - 4.0)		0.003 **
Stage	(N=239)	1.7 (1.0 - 3.0)		0.039 *
Prognostic_model	(N=239)	4.7 (2.6 - 8.4)		<0.001 ***

\# Events: 65; Global p-value (Log-Rank): 1.9107e-10
AIC: 579.51; Concordance Index: 0.74

1   2   5   10

Figure 6

Univariate and multivariate Cox regression analysis of OS. Clinical patterns including gender, diagnostic age, pathologic TNM, pathologic grade, pathologic stage, vascular invasion, and serum AFP level in HCC patients. (A) univariate Cox regression analysis of OS. (B) multivariate Cox regression analysis of OS.
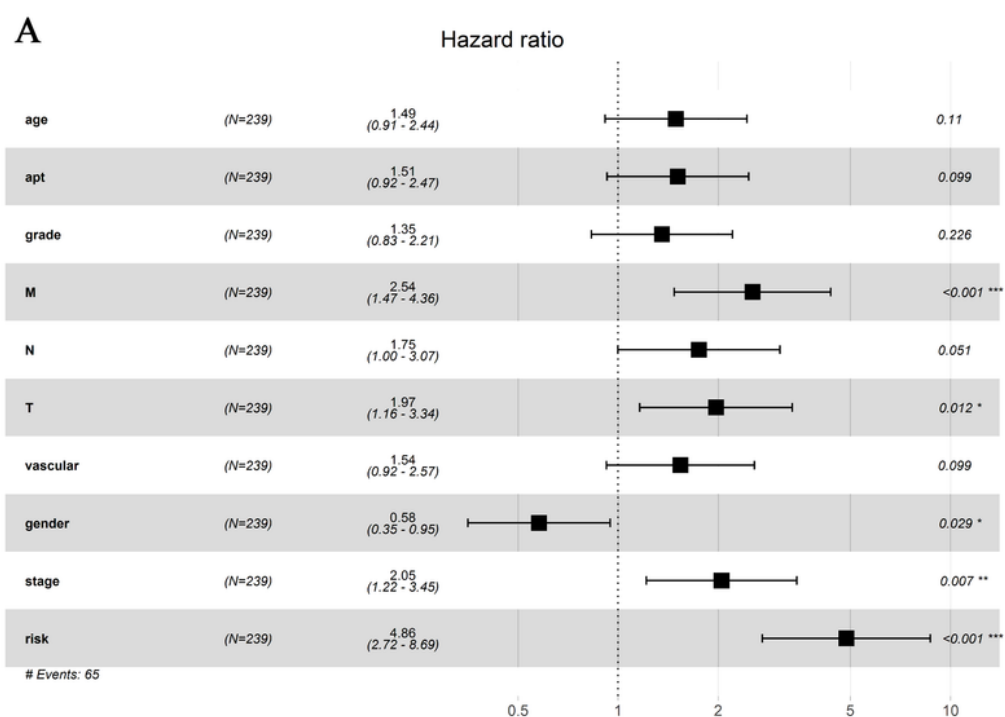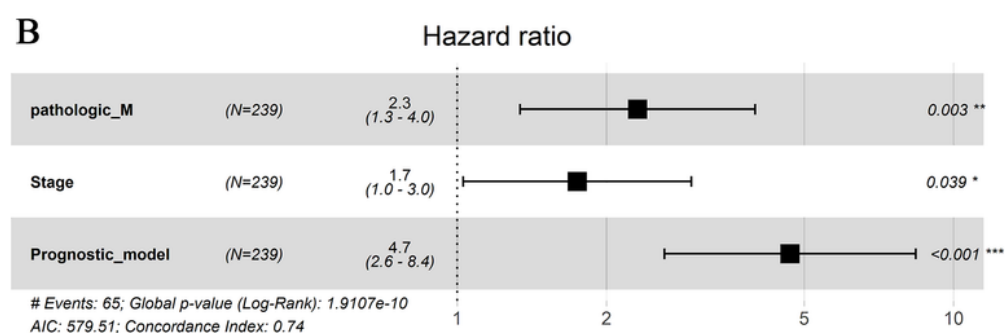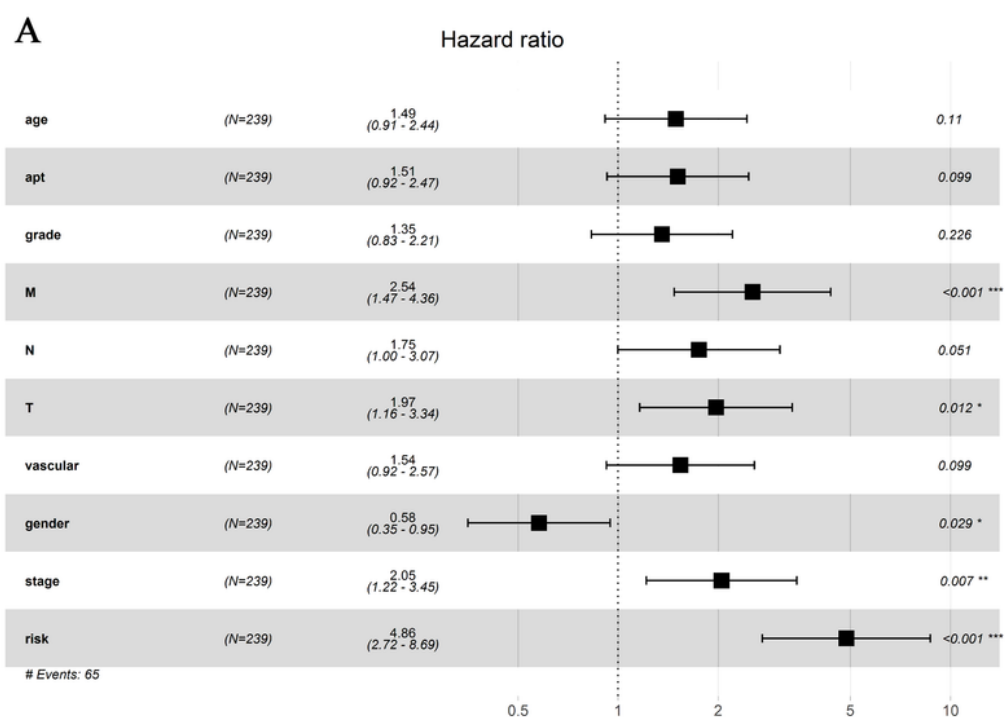
Figure 6

Univariate and multivariate Cox regression analysis of OS. Clinical patterns including gender, diagnostic age, pathologic TNM, pathologic grade, pathologic stage, vascular invasion, and serum AFP level in HCC patients. (A) univariate Cox regression analysis of OS. (B) multivariate Cox regression analysis of OS.

A

Hazard ratio

age	(N=239)	1.49 (0.91 - 2.44)		0.11
apt	(N=239)	1.51 (0.92 - 2.47)		0.099
grade	(N=239)	1.35 (0.83 - 2.21)		0.226
M	(N=239)	2.54 (1.47 - 4.36)		<0.001 ***
N	(N=239)	1.75 (1.00 - 3.07)		0.051
T	(N=239)	1.97 (1.16 - 3.34)		0.012 *
vascular	(N=239)	1.54 (0.92 - 2.57)		0.099
gender	(N=239)	0.58 (0.35 - 0.95)		0.029 *
stage	(N=239)	2.05 (1.22 - 3.45)		0.007 **
risk	(N=239)	4.86 (2.72 - 8.69)		<0.001 ***

Events: 65

0.5 1 2 5 10

B

Hazard ratio

pathologic_M	(N=239)	2.3 (1.3 - 4.0)		0.003 **
Stage	(N=239)	1.7 (1.0 - 3.0)		0.039 *
Prognostic_model	(N=239)	4.7 (2.6 - 8.4)		<0.001 ***

Events: 65; Global p-value (Log-Rank): 1.9107e-10
AIC: 579.51; Concordance Index: 0.74

1 2 5 10

Figure 6

Univariate and multivariate Cox regression analysis of OS. Clinical patterns including gender, diagnostic age, pathologic TNM, pathologic grade, pathologic stage, vascular invasion, and serum AFP level in HCC patients. (A) univariate Cox regression analysis of OS. (B) multivariate Cox regression analysis of OS.

Figure 7

Kaplan-Meier survival analysis for all HCC 239patientsdata according to theprognostic gene signature based on stratification of different clinical features. (A) Gender. (B) AFP. (C)Tumor grade. (D) Stage (E) Pathologic_T. (F) Pathologic_M. (G) Pathologic_N. (H)Vascular. We calculated p values using the log-rank test. Thestatistical significance level was 0.05.

Image not available with this version

Figure 8

Integration of gene and clinical characteristics (A) Nomogram constructed to predict the 1-, 3-,5- and7-year OS in the TCGA cohort. (B-E) Calibration curve of the nomogram forpredicting the probability of OS at 1-, 3-,5- and7-year.

Image not available with this version

Figure 8

Integration of gene and clinical characteristics (A) Nomogram constructed to predict the 1-, 3-,5- and7-year OS in the TCGA cohort. (B-E) Calibration curve of the nomogram forpredicting the probability of OS at 1-, 3-,5- and7-year.

Image not available with this version

Figure 8

Integration of gene and clinical characteristics (A) Nomogram constructed to predict the 1-, 3-,5- and7-year OS in the TCGA cohort. (B-E) Calibration curve of the nomogram forpredicting the probability of OS at 1-, 3-,5- and7-year.

Figure 8

Integration of gene and clinical characteristics (A) Nomogram constructed to predict the 1-, 3-,5- and7-year OS in the TCGA cohort. (B-E) Calibration curve of the nomogram forpredicting the probability of OS at 1-, 3-,5- and7-year.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1DownregualatedDEGsofHCCinTCGA.xls
- TableS1DownregualatedDEGsofHCCinTCGA.xls
- TableS1DownregualatedDEGsofHCCinTCGA.xls
- TableS1DownregualatedDEGsofHCCinTCGA.xls
- TableS2UpregulatedDEGsofHCCinTCGA.xls
- TableS2UpregulatedDEGsofHCCinTCGA.xls
- TableS2UpregulatedDEGsofHCCinTCGA.xls
- TableS2UpregulatedDEGsofHCCinTCGA.xls
- TableS3DEGscorrelatedwithOSinTCGA.xls
- TableS3DEGscorrelatedwithOSinTCGA.xls
- TableS3DEGscorrelatedwithOSinTCGA.xls
- TableS3DEGscorrelatedwithOSinTCGA.xls