


Identification of a Liver Progenitor Cell-Related Genes Signature Predicting Overall Survival for Hepatocellular Carcinoma

Technology in Cancer Research & Treatment
 Volume 20: 1-10
 © The Author(s) 2021
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/15330338211041425
journals.sagepub.com/home/tct


Xiaoyong Li, MD¹ , Jiaqong Lin, PhD², Yuguo pan, MD¹, Peng Cui, MD¹, and Jintang Xia, PhD¹

Abstract

Background: Liver progenitor cells (LPCs) play significant roles in the development and progression of hepatocellular carcinoma (HCC). However, no studies on the value of LPC-related genes for evaluating HCC prognosis exist. We developed a gene signature of LPC-related genes for prognostication in HCC. **Methods:** To identify LPC-related genes, we analyzed mRNA expression arrays from a dataset (GSE57812 & GSE 37071) containing LPCs, mature hepatocytes, and embryonic stem cell samples. HCC RNA-Seq data from The Cancer Genome Atlas (TCGA) were used to explore the differentially expressed genes (DEGs) related to prognosis through DEG analysis and univariate Cox regression analysis. Lasso and multivariate Cox regression analyses were performed to construct the LPC-related gene prognostic model in the TCGA training dataset. This model was validated in the TCGA testing set and an external dataset (International Cancer Genome Consortium [ICGC] dataset). Finally, we investigated the relationship between this prognostic model with tumor-node-metastasis stage, tumor grade, and vascular invasion of HCC. **Results:** Overall, 1770 genes were identified as LPC-related genes, of which 92 genes were identified as DEGs in HCC tissues compared with normal tissues. Furthermore, we randomly assigned patients from the TCGA dataset to the training and testing cohorts. Twenty-six DEGs correlated with overall survival (OS) in the univariate Cox regression analysis. Lasso and multivariate Cox regression analyses were performed in the TCGA training set, and a 3-gene signature was constructed to stratify patients into 2 risk groups: high-risk and low-risk. Patients in the high-risk group had significantly lower OS than those in the low-risk group. Receiver operating characteristic curve analysis confirmed the signature's predictive capacity. Moreover, the risk score was confirmed to be an independent predictor for patients with HCC. **Conclusion:** We demonstrated that the LPC-related gene signature can be used for prognostication in HCC. Thus, targeting LPCs may serve as a therapeutic alternative for HCC.

Keywords

hepatocellular carcinoma, LPC-related genes, prognostic signature, bioinformation, mRNAs

Abbreviations

LPCs, Liver progenitor cells; HCC, Hepatocellular carcinoma; DEGs, differentially expressed genes; TCGA, The cancer genome atlas; ICGC, International Cancer Genome Consortium; GO, Gene Ontology; CSCs, Cancer Stem Cells; EpCAM, epithelial cell adhesion molecule; NCAM, neural cell adhesion molecule; DKL1, delta-like 1 homolog; GEO, Gene Expression Omnibus database; HTSeq, Harmonized RNA sequencing; KEGG, Kyoto Encyclopedia of Genes and Genomes; TPM, Transcripts per million; PCA, The Principal Component Analysis; EZH2, zeste 2 polycomb repressive complex 2 subunit; PRKAA2, protein kinase AMP-activated catalytic subunit alpha 2; SPP1, secreted phosphoprotein

Received: March 1, 2021; Revised: July 31, 2021; Accepted: August 3, 2021.

Introduction

Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related deaths worldwide.¹ HCC occurs mainly in the context of cirrhosis, hepatitis B or C virus infection, alcohol addiction, nonalcoholic steatohepatitis, or exposure to dietary toxins such as aflatoxins and aristolochic acid. Preventing

¹ Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
² School of Basic Medical Sciences, Southern Medical University, Guangzhou, China

Corresponding Author:

Jintang Xia, Department of General Surgery, Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.
 Email: xiajintanggy@sina.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

these risk factors and clinical surveillance contribute to early diagnosis and potentially curative treatment. However, because of the rapid progression observed in the disease, intra-hepatic recurrence, and early metastasis, the prognosis of HCC patients remains poor. Generally, conventional models use clinicopathological parameters such as tumor-node-metastasis (TNM) stage, vascular invasion, and other parameters to prognosticate HCC.² However, considering genetic alterations and tumor heterogeneity, the predictive efficacy of conventional models is unsatisfactory. Therefore, it is important to consider molecular markers when establishing a novel prognostic model for patients with HCC. In a previous study, gene signatures were constructed to prognosticate HCC patients. For example, Long et al³ developed an immune prognostic model based on immune-related genes that were differentially expressed between TP53WT and TP53MUT HCC samples. Zhang et al⁴ constructed a risk score prognosis model by computing the immune/stromal scores of HCC patients obtained from The Cancer Genome Atlas (TCGA) based on the ESTIMATE algorithm. Moreover, Ao et al⁵ identified a prognostic signature consisting of 20 gene pairs, which could robustly predict the disease-free survival and overall survival (OS) of HCC patients. All these molecular prognostic models not only predict patient prognosis but also suggest their significant role in the development and progression of HCC. Despite this, no research has been published on the liver progenitor cell (LPC)-related gene signature.

LPCs, which have the capacity to differentiate into a lineage of hepatocytes and cholangiocytes, are bipotential cells residing in the canal of Hering.⁶ They are only activated under conditions when the proliferative capacity of hepatocytes is compromised, and they were considered to participate in the regeneration of liver mass. Cancer stem cells (CSCs) play a significant role in the development and progression of cancer. By comparing CSCs with adult stem cells, they show similar characteristics, such as self-renewal, differentiation, and longevity. As cancer is essentially a genetic disease, the longevity of stem cells means that they could have enough time to accumulate genetic alterations that result in carcinogenesis through the activation of proto-oncogenes or inactivation of tumor suppressor genes.^{7,8} Therefore, LPCs were not only considered to be involved in liver regeneration, but also in the origin of liver cancer. The controversy surrounding the specific biomarkers of LPCs remains, but previous studies have reported that LPC-related biomarkers such as epithelial cell adhesion molecule, neural cell adhesion molecule, delta-like 1 homolog (DLK1), and cytokeratin 19⁹ are shared by some HCC patients with poor prognosis.⁹ Li et al¹⁴ found that LPCs exhibited CSC properties when provoked by chronic inflammation, suggesting the malignant transformation of LPCs under these conditions. Furthermore, Liu et al¹⁵ found that autocrine osteopontin (SPP1) promoted LPC expansion and migration as well as subsequent tumorigenicity in a choline-deficient treated murine model. The expression of LPC-related genes was considered to contribute to the heterogeneity of HCC, which was verified to be closely associated with the prognosis of patients. Therefore, analysis of LPC-related genes may help to predict HCC outcomes

Table 1. Clinical Information of HCC patients in this study.

	TCGA cohort (n = 365)	ICGC cohort (n = 231)
Age (median, range)	61 (16-90)	69 (31-89)
Gender		
Female	119	61
Male	246	170
Grade		
G1	55	NA
G2	175	NA
G3	118	NA
G4	12	NA
Unknown	5	NA
Stage		
I	170	36
II	84	105
III	83	71
IV	4	19
Unknown	24	0
Vascular invasion		
Yes	106	NA
No	205	NA
Unknown	54	NA

Abbreviations: HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas.

more accurately and identify potential novel candidates for specific targeted therapies.

Methods

Data Collection

Genomic profiling datasets GSE57812 and GSE37031 were downloaded from the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>). GSE57812 contains 9 independent samples of second trimester LPCs, 3 independent mature hepatocyte samples, and 6 human embryonic stem cell (ESC) samples while GSE37031 contains 7 normal liver samples. Detailed information on the treatment protocol of each cell type is accessible through GSE accession numbers GSE57812 and GSE37031. Harmonized RNA sequencing data (HTSeq-counts) and associated clinical information for HCC were downloaded from TCGA (<https://portal.gdc.cancer.gov/>). To assess the prognostic performance of the gene signature in the external dataset, the International Cancer Genome Consortium (ICGC) dataset was obtained from <https://dcc.icgc.org/projects/LIRI-JP>. This dataset contains 231 RNA-seq data points and clinical information of tumor samples. The clinical and pathological characteristics of the patients are shown in Table 1.

Identification and Enrichment Analysis of LPC-Related Genes

To identify LPC-related genes, we merged the GSE57812 and GSE37031 datasets into a single file. Then, we removed the known batch effects using the sva package. Next, we calculated

DEGs between LPCs and mature hepatocytes and between ESCs and mature hepatocytes in the merged data using the LIMMA package in the R software. Based on the value of \log_2 (fold change) of DEGs, we divided the DEG list into 2 parts: upregulated DEGs and downregulated DEGs. By overlapping these 4 DEG lists, we obtained reliable LPC-related genes, which indicated the characteristics of the gene expression profile of LPCs. To investigate the potential biological functions of LPC-related genes, gene ontology (GO) biological process terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using the cluster profiler package.¹⁶

Prognostic Model Development

The differentially expressed LPC-related genes between HCC and normal tissues were detected in the TCGA dataset using the edgeR package in the R software.¹⁷ Volcano plots were also conducted in the R software. Then, we normalized the raw RNA-Seq read counts using the Transcripts per million (TPM) method. The formula is as follows:
$$TPMi = \frac{Ni/Li \cdot 10^6}{\sum(N1/L1 + N2/L2 + \dots + Nn/Ln)}$$
, where Ni is the number of read counts mapped to a particular gene region i and Li is the gene length of gene i . Furthermore, we took the normalized read counts for the $\log_2(x+1)$ transformation. The TCGA dataset was randomly divided into a training dataset and a testing dataset (70% vs 30%, respectively), and the initial seed size was 32. The transformed expression levels of the differentially expressed LPC-related genes were then analyzed in the TCGA training set using univariate Cox analysis. We screened prognosis-related genes with P values $<.01$ for further analysis. A prognostic gene model was then constructed in the training dataset. Lasso-penalized Cox regression analysis was performed to further reduce the number of prognosis-related genes in the training dataset based on the glmnet package. Furthermore, a stepwise Cox proportional hazards model was used to screen prognostic gene signatures. The risk scores of the patients were calculated according to the normalized expression level of each gene and its corresponding regression coefficients (Risk score = expression of gene 1 $\cdot \beta_1$ + expression of gene 2 $\cdot \beta_2$ + \dots + expression of gene $n \cdot \beta_n$). Patients were then divided into 2 groups based on the median risk score. Kaplan–Meier analysis was performed to compare OS between the high-risk and low-risk groups using the survival package. Additionally, receiver operating characteristic (ROC) curve analysis was also performed using the survival ROC package. The TCGA and ICGC cohorts were used to verify the prognostic value of the prognostic model.

Identification of Independent Prognostic Parameters for HCC

To identify independent prognostic parameters for HCC and to validate the independent prognostic value of the prognostic model, multivariate Cox regression analyses were performed

based on the risk score and clinical parameters, including age, gender, TNM stage, neoplasm histologic grade, and vascular invasion in the entire TCGA dataset. Statistical significance was set at $P < .05$. AS nomograms are widely used to predict the prognosis of cancer, the risk score and clinical parameters were included to build a nomogram to investigate the probability of 1-, 3-, and 5-year OS of HCC patients.

Results

Identification and GO, KEGG Analysis of LPC-Related Genes

To identify LPC-related genes, we first merged the GSE57812 and GSE37031 datasets into a single file. The GSE57812 dataset contains 9 LPC, 6 ESC, and 3 mature hepatocyte samples, while GSE37031 contains 8 nonalcoholic steatohepatitis liver samples and 7 normal liver samples. Three mature hepatocyte samples from GSE57812 and 7 normal liver samples from GSE37031 were chosen as the control group. Principal component analysis exhibited an excellent separation tendency of these 3 different samples, indicating the distinctly different characteristics of these 3 cell types (Figure 1A). DEG analysis was performed using the LIMMA package in R. Results show that $|\text{LogFC}| \geq 2$ and $\text{adj.P.Val} \leq .05$, which was considered statistically significant. A total of 1259 upregulated genes and 1930 downregulated genes were selected as LPCs and mature hepatocytes, and 2106 upregulated genes and 1043 downregulated genes were selected by comparing ESCs and mature hepatocytes (Figure 1B). Intersecting the sets was performed using the UpSet plot. The red and blue bars in the UpSet plot represent the upregulated and downregulated DEGs in the LPCs, respectively (Figure 1C). In total, 1770 genes were identified as LPC-related genes. To better understand the biological consequences of the LPC-related genes, we performed GO and KEGG enrichment analyses using the cluster profiler package. The GO enrichment results showed that LPC-related genes were significantly enriched in biological processes related to immune regulation (Figure 1D). On the other hand, KEGG pathway analysis revealed that the LPC-related genes participated in cell adhesion molecules, DNA replication, and adherens junctions (Figure 1D). These signaling pathways have been reported to play significant roles in cancer development. As such, the results above indicate that abnormal expression of LPC-related genes might influence the immune status of the tumor microenvironment and the development of HCC.

Construction of an LPC-Related Gene Prognostic Model in TCGA Dataset

To construct an LPC-related gene prognostic model, we first analyzed the expression levels of LPC-related genes in HCC and in normal tissues from the TCGA dataset. This dataset contained 371 HCC tissues and 50 normal tissues. The edgeR package was used to analyze the differential expression of LPC-related genes between the 2, where $|\text{LogFC}| \geq 2$ and

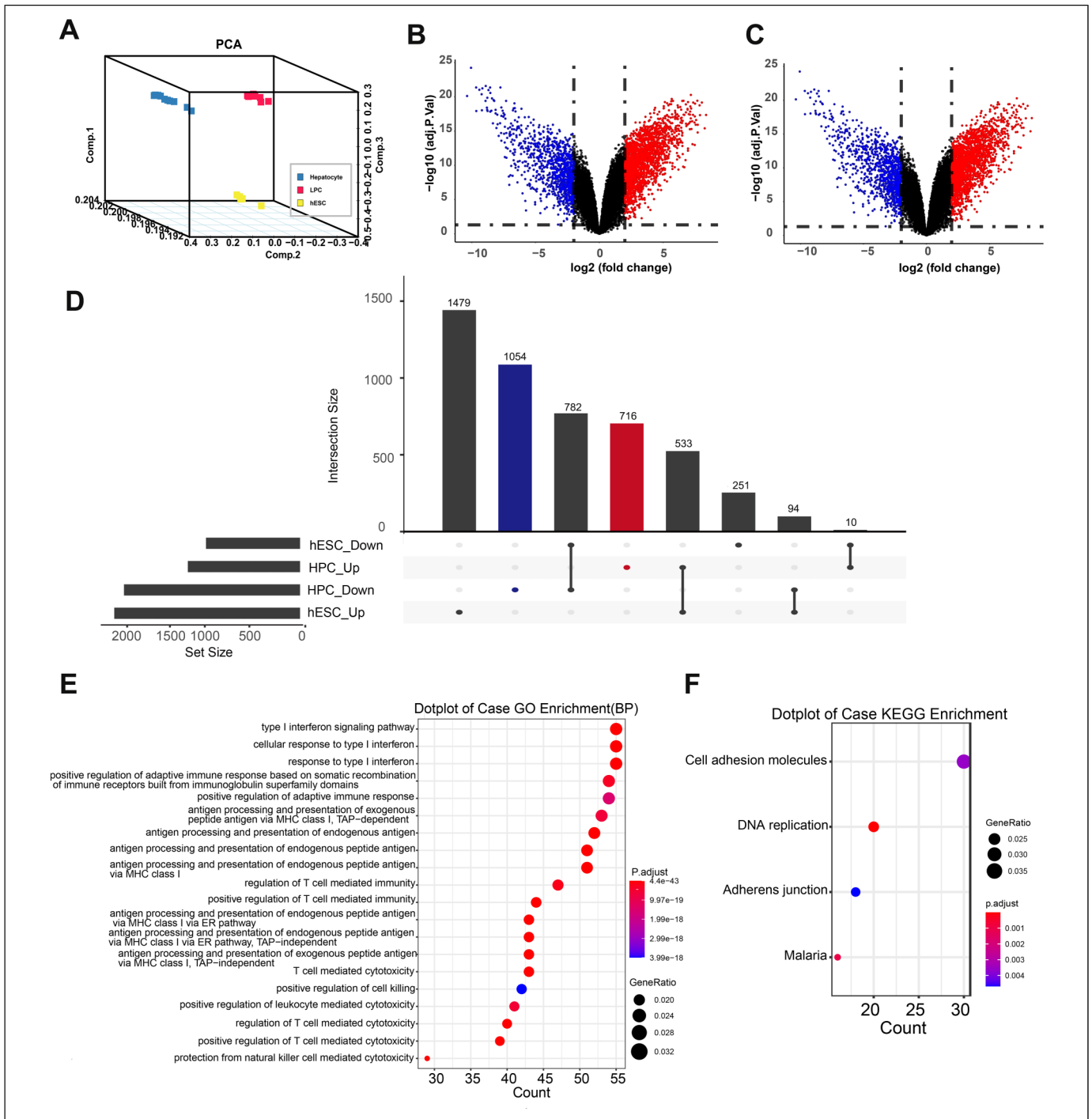


Figure 1. Identification and GO, KEGG analysis of LPC-related genes. (A) PCA of GSE57812 and GSE37031 (7 normal liver samples). (B) Volcano Plot on the left is the result of DEGs comparing between LPCs and mature hepatocytes and the volcano plot on the right represents the result of DEGs comparing between ESCs and mature hepatocytes. Threshold for the gene selection: $|\text{LogFC}| \geq 2$ and $\text{adj.P.Val} \leq .05$. (C) UpSet plot of the number of genes that were common or exclusive between comparison groups. (D) GO enrichment analysis of LPC-related genes. (E) KEGG pathway enrichment analysis of LPC-related genes.

Abbreviations: PCA, principal component analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LPC, liver progenitor cell; DEG, differentially expressed gene; ESC, embryonic stem cell.

$\text{adj.P.Val} \leq .01$ was considered statistically significant, and 92 DEGs, including 69 upregulated and 23 downregulated genes, were identified (Figure 2A and B). Next, we normalized the number of raw read counts from the TCGA dataset using the

TPM method. Furthermore, we took the normalized read counts for $\log_2(x+1)$ transformation. A total of 365 HCC patients with complete prognosis information and RNA-seq data were included in the survival analysis. Furthermore, we

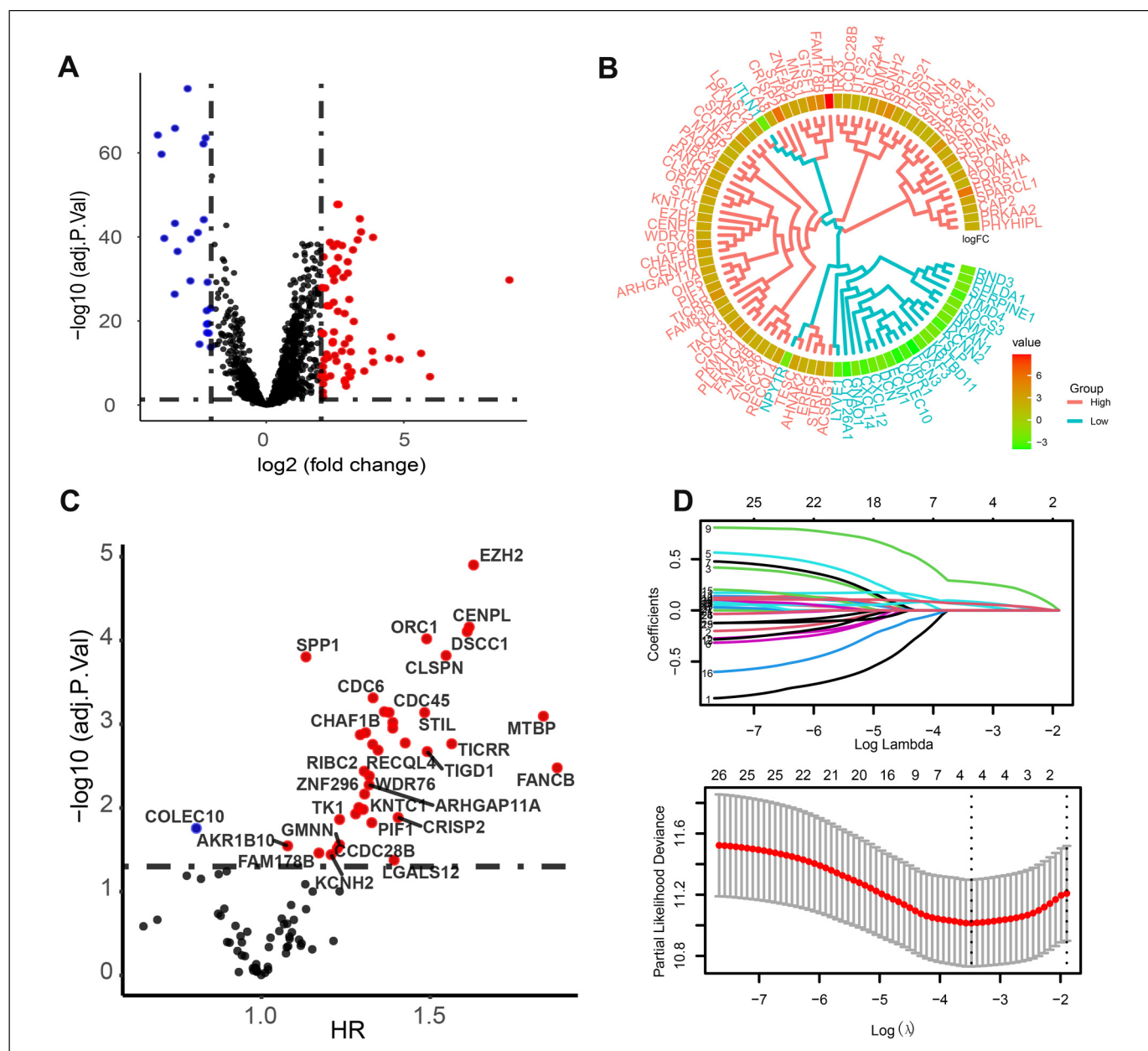


Figure 2. Construction of an LPC-related prognostic model in TCGA dataset. (A) Volcano plot represents the result of DEGs of LPC-related genes between tumor tissues and normal tissues. (B) Cluster dendrogram of the 92 DEGs of LPC-related genes. (C) Volcano plot represents the result of univariate Cox regression analysis. P.Val $\leq .01$ was considered statistically significant. (D) Lasso regression coefficient profiles of survival-associated mRNA and 'Leave-one-out' cross-validation for parameter selection during Lasso regression. Abbreviations: LPC, liver progenitor cell; TCGA, The Cancer Genome Atlas; DEG, differentially expressed gene.

divided the TCGA dataset into 2 cohorts: the TCGA training cohort (70%) and the testing dataset (30%). Univariate Cox regression analysis was performed to identify DEGs that were correlated with OS in the TCGA training cohort, and Figure 2C shows that a total of 26 DEGs were identified to be significantly associated with OS ($P < .01$). Using the Lasso-penalized Cox regression analysis in the training dataset, 4 genes were selected based on the glmnet package in R (Figure 2D). Multivariate Cox regression was then performed to narrow the gene signature and construct an LPC-related prognostic signature. Finally, 3 genes, including enhancer of zeste 2

polycomb repressive complex 2 subunit (EZH2), protein kinase AMP-activated catalytic subunit alpha 2 (PRKAA2), and secreted phosphoprotein 1 (SPP1), were identified in the prognostic model. Using the equation, (risk score = $0.39335 * \text{Expression}_{\text{EZH2}} + 0.14361 * \text{Expression}_{\text{PRKAA2}} + 0.08324 * \text{Expression}_{\text{SPP1}}$), we calculated the risk score for each patient in the training dataset. According to the risk score, we ranked them and then classified the patients into high- and low-risk groups based on the median value (Figure 3A). The Kaplan-Meier survival curve revealed significantly worse prognosis in the high-risk group ($P < .0001$) (Figure 3B). Next, the

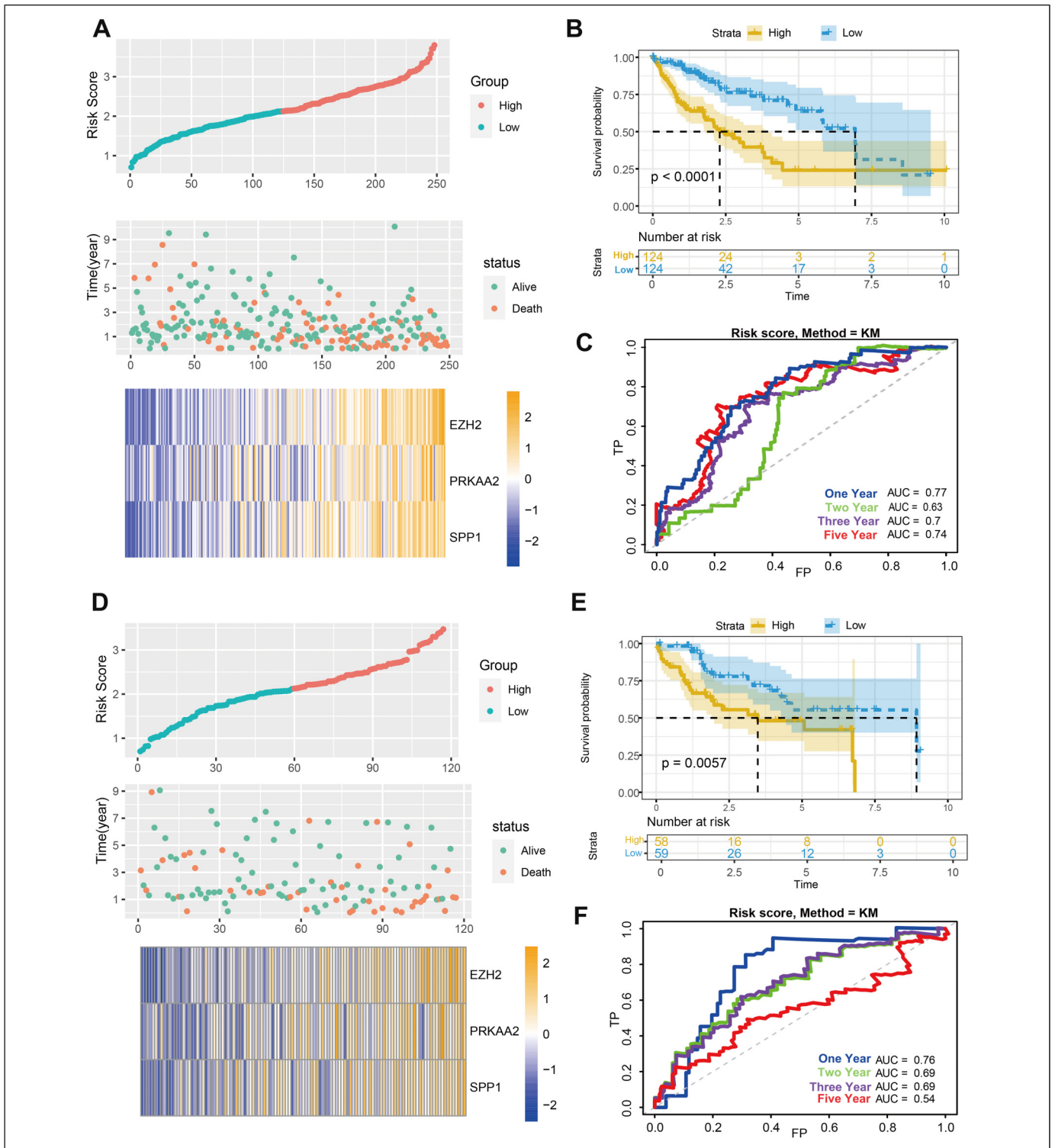


Figure 3. Evaluation of the performance of LPC-related prognostic model in TCGA training and testing dataset. (A) Relationship among the risk score (upper), survival status of patients in different groups (middle), and the expression profiles of the 3 prognostic genes (bottom) in the training dataset. Patients from the training dataset were stratified into 2 groups according to the median value of risk scores. (B) The Kaplan–Meier survival curves of the 3 prognostic genes in the training dataset. (C) The time-dependent ROC for 1-, 2-, 3-, and 5-year predictions for the 3 prognostic genes in the training dataset. (D) Relationship among the risk score (upper), survival status of patients in different groups (middle), and the expression profiles of the 3 prognostic genes (bottom) in the testing dataset. Patients from the training dataset were stratified into 2 groups according to the median value of risk scores. (E) The Kaplan–Meier survival curves of the 3 prognostic genes in the testing dataset. (F) The time-dependent ROC for 1-, 2-, 3-, and 5-year predictions for the 3 prognostic genes in the testing dataset. Abbreviations: LPC, liver progenitor cell; TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic curve.

prognostic value of the LPC-related prognostic model was evaluated based on the time-dependent ROC in the training dataset. The AUCs for 1-, 2-, 3-, and 5-year OS were 0.77, 0.63, 0.7, and 0.74, respectively (Figure 3C). Furthermore, we validated the performance of the LPC-related prognostic signature using the testing dataset, and it was observed that patients in the high-risk groups had a significantly poorer prognosis than those in the low-risk groups using Kaplan–Meier analysis ($P = .0057$) (Figure 3E). Moreover, AUCs for 1-, 2-, 3-, and 5-year OS were 0.76, 0.69, 0.69 and 0.54, respectively (Figure 3F).

External Validation of the LPC-Related Prognostic Model

To assess the performance of the LPC-related prognostic model in the external dataset, we downloaded the ICGC dataset from <https://dcc.icgc.org/projects/LIRI-JP>. The ICGC dataset contained RNA-seq data and clinical information of 231 tumor samples. The LPC-related prognostic model was then validated in the ICGC cohort. Next, the patients were divided into high- and low-risk groups by the median value, and it was observed that the OS was significantly poorer in the high-risk group than in the low-risk group ($P < .00001$) (Figure 4B). Moreover, AUCs for 1-, 2-, 3-, and 5-year OS were 0.7, 0.63, 0.66, and 0.72, respectively (Figure 4C). Taken together,

these results indicate that the LPC-related prognostic model is capable of predicting OS in HCC.

Investigation on the Clinical Relevance of the LPC-Related Prognostic Signature

The relationship between risk score and clinical features of HCC, such as TNM stage, neoplasm histologic grade, and vascular invasion, was analyzed in the TCGA cohort using the ggstatsplot package in R. The results showed significant differences in risk score in the different TNM stage samples (risk score for stages I, II, III, and III & IV were 1.97, 2.25, 2.37, respectively; all $P_{FDR\ corrected} < .05$) (Figure 5A). In addition, a higher risk score was associated with a higher incidence of vascular invasion (The mean of the risk score for vascular invasion: without vascular invasion = 2.21:2.04; $P < .05$) (Figure 5B). Moreover, the levels of risk score showed significant differences between different histopathological grades (mean of risk score for histopathological grades G1, G2, G3, G4 were 1.79, 2.07, 2.38, 2.47, respectively; all $P_{FDR\ corrected} < .05$) (Figure 5C). As this study aimed to determine whether risk score was an independent prognostic factor for OS, univariate and multivariate Cox regression analyses were performed in the TCGA cohort. In the univariate Cox regression analyses, the risk score and TNM stage were

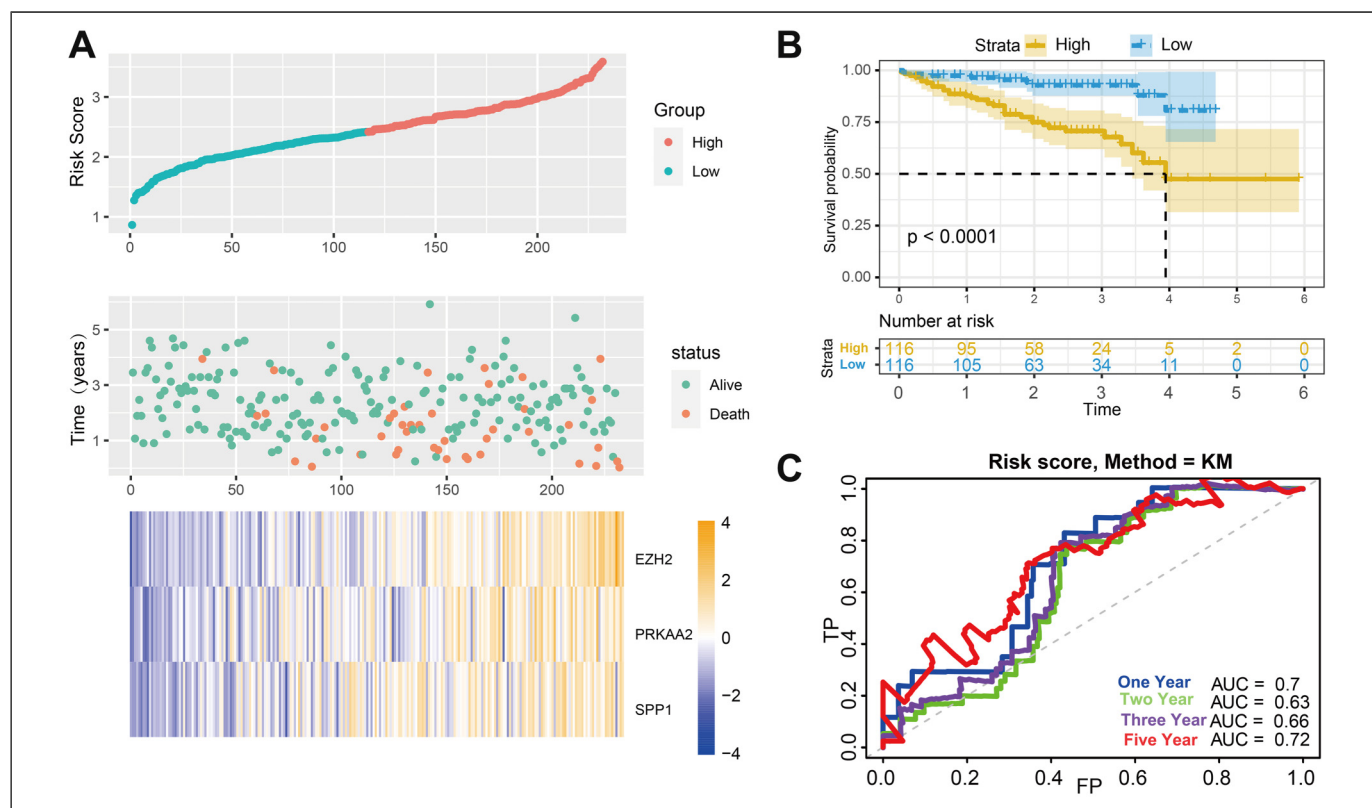


Figure 4. External validation of the LPC-related prognostic model in ICGC dataset. (A) Relationship among the risk score (upper), survival status of patients in different groups (middle), and the expression profiles of the 3 prognostic genes (bottom). Patients from the training dataset were stratified into 2 groups according to the median value of risk scores. (B) The Kaplan–Meier survival curves of the 3 prognostic genes. (C) The time-dependent ROC for 1-, 2-, 3-, and 5-year predictions for the 3 prognostic genes. Abbreviations: LPC, liver progenitor cell; ROC, receiver operating characteristic curve.

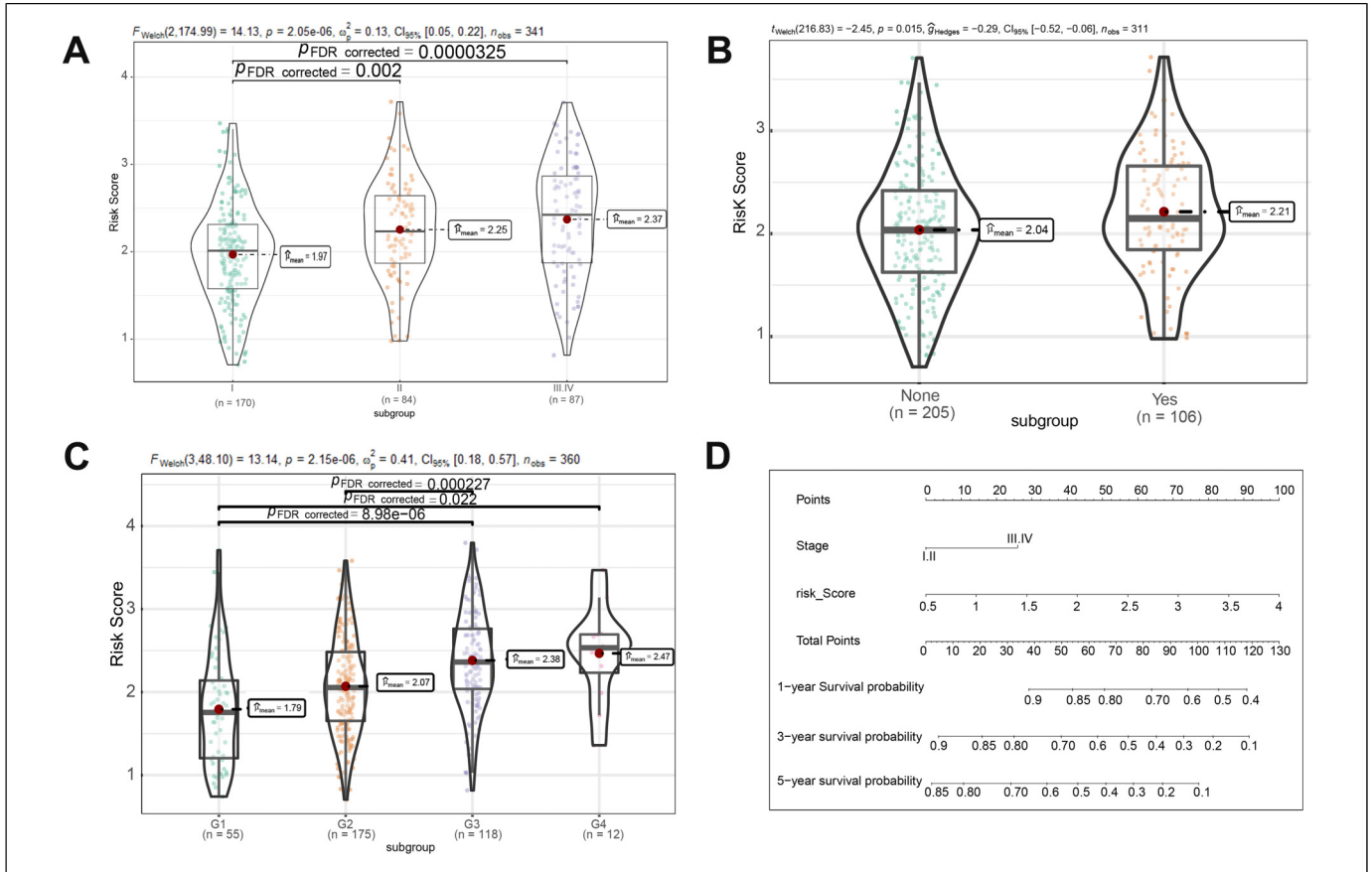


Figure 5. Clinical relevance of the LPC-related prognostic signature. (A) The distribution of the risk score in different TNM stage in TCGA. (B) The distribution of the risk score in situation of vascular invasion in TCGA. (C) The distribution of the risk score in different neoplasm histologic grade in TCGA. (D) Nomogram predicting overall survival for HCC patients from TCGA dataset. Abbreviations: LPC, liver progenitor cell; TCGA, The Cancer Genome Atlas; HCC, hepatocellular carcinoma.

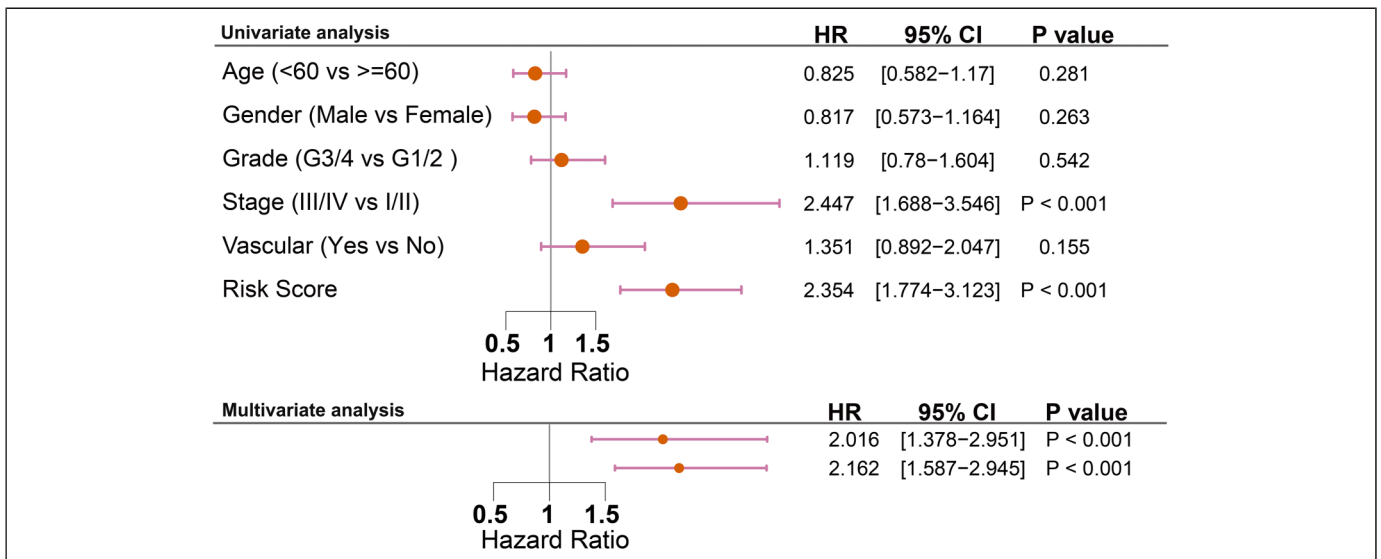


Figure 6. Results of the univariate and multivariate Cox regression analyses regarding OS in the TCGA cohort. Abbreviations: OS, overall survival; TCGA, The Cancer Genome Atlas.

observed to be significantly associated with OS. After correction for other confounding factors, the risk score and TNM stage still proved to be independent predictors of OS in the multivariate Cox regression analysis (Figure 6). Subsequently, we built a nomogram to predict 1-year, 3-year, and 5-year OS in the TCGA cohort by using TNM stage and risk score, which provided a more tailored risk prediction. The results showed that the risk score exhibited the strongest impact on prognosis among all factors (Figure 5D).

Discussion

HCC is a significant public health burden worldwide. To a certain extent, traditional clinical indicators, such as TNM stages, blood vessel invasion, and AFP value, were able to predict the prognosis of patients with HCC. However, because of heterogeneity in tumors, patients with similar clinicopathological characteristics may have different outcomes when treated with the same therapy. This presents the limitations of traditional clinical indicators in predicting the prognosis of patients with HCC. Based on gene expression profile analysis, researchers recently developed the same novel prognostic signatures focusing on certain characteristics, such as cell cycle, autophagy, and immune signature.^{3,8,18} To a certain degree, these signatures predict the mortality risk of HCC. In this study, we identified 1770 LPC-related genes based on the GSE57812 and GSE37031 datasets, and we performed GO and KEGG analyses of these genes. Analysis results showed that LPC-related genes were enriched in interferon-related pathways and the biological processes associated with antigen processing and presentation and T cell regulation, indicating a regulatory role of LPC-related genes in the tumor immune microenvironment.

To construct an LPC-related prognostic model, we first analyzed the expression levels of LPC-related genes in HCC and normal tissues from the TCGA dataset. We then constructed an LPC-related prognostic model using Cox regression and Lasso regression analyses on the TCGA dataset. The LPC-related prognostic model contained 3 genes (EZH2, PRKAA2, and SPP1), all of which negatively affected the prognosis of patients with HCC. Notably, EZH2 has been implicated in genetic, transcriptional, and posttranscriptional dysregulation in many cancer types¹⁹ and was found to promote tumorigenesis by altering the expression of numerous tumor suppressor genes. Moreover, it has been reported to participate in the self-renewal of CSCs.^{22,23} The study conducted by Zhu et al¹⁸ showed that EZH2 plays a critical role in the self-renewal maintenance of liver CSCs.²⁴ It may also participate in the self-renewal and malignant transformation of LPCs. Of note, SPP1 is a key extracellular matrix protein involved in several pathophysiological processes, including cancer progression and metastasis. Upregulation of SPP1 has been found in LPCs of several liver diseases with portal biliary proliferation.^{25,26} Moreover, Liu et al demonstrated that SPP1 promotes LPC expansion, and depletion of SPP1 significantly attenuates modified choline-deficient-induced hepatocarcinogenesis.¹⁵ As

such, these 2 genes have been shown to be closely related to tumorigenesis. However, studies on PRKAA2 in cancer are relatively rare. Thus, the role of PRKAA2 in LPCs and their relationship with HCC remain unknown. To determine the clinical prognostic risk score of each HCC patient, we calculated the risk score for each patient based on these 3 genes. The risk score could efficiently stratify the OS of HCC patients in the TCGA training dataset, TCGA testing dataset, and the ICGC dataset. The efficacy of our signature was verified in these 3 datasets, indicating a robust prognostic value of the model. Furthermore, multivariate Cox analysis showed that the risk score remained an independent prognostic factor for predicting OS. A nomogram integrated with the risk score and clinicopathological parameters was established to predict the OS of HCC patients, which helps to provide a more tailored risk prediction.

This study had several limitations. First, the LPC-related prognostic signature was constructed and validated using retrospective data from public databases. More prospective real-world data are required to verify its clinical utility. Second, we did not test the biological functions of EZH2, SPP1, and PRKAA2. The biological functions of EZH1 and SPP1 have been reported in previous studies, but studies on PRKAA2 in cancer are relatively rare. All these three genes may also play a significant role in the crosstalk between LPCs and hepatocarcinoma cells. Functional experiments are needed to clarify the underlying mechanisms of these 3 genes. In addition, because of the lack of specific biomarkers for LPCs, further research on LPC in HCC is limited. In previous studies, researchers have often used cytokeratin 19, CD133, Nestin, and CD44 as LPC biomarkers. However, these biomarkers have limitations. For example, cytokeratin 19 is also expressed in the biliary epithelial cells and CD133 has been identified as a biomarker for stem cells. As for this, the expression levels of cytokeratin 19, CD133, Nestin, and CD44 did not represent well the expression levels of LPCs. In this study, we focused on the role of LPCs in the development and progression of HCC. Therefore, we constructed an LPC-related gene signature to explore the role of LPC-related genes in predicting HCC prognosis.

In this study, we constructed and confirmed a prognostic model comprising 3 LPC-related genes. This model proved to be independently associated with OS in patients with HCC. Moreover, it could be a potential predictor for patients with HCC. The underlying mechanisms between these 3 genes and HCC remain poorly understood and warrant further investigation.

Acknowledgments

The authors would like to express their sincere thanks for sharing the data from TCGA, ICGC, and GEO database.

Author Contributions

Xiaoyong Li and Jiaqiong Lin contributed equally to this work. Xiaoyong Li contributed to conception and design. Jiaqiong Lin contributed to analysis and interpretation of data. Xiaoyong Li and Yuguo Pan contributed to writing of the manuscript. Peng Cui contributed to review of the manuscript. Jintang Xia contributed to study supervision.


Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (2019, 81902831).

ORCID iD

Xiaoyong Li  <https://orcid.org/0000-0001-7754-731X>

References

- Akinyemiju T, Abera S, et al., Global Burden of Disease Liver Cancer, C. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015. *JAMA Oncol.* 2017;3(12):1683-1691.
- Bruix J, Reig M, Sherman M. Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. *Gastroenterology.* 2016;150(4):835-853.
- Long J, Wang A, Bai Y, et al. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine.* 2019;42(6):363-374.
- Zhang FP, Huang YP, Luo WX, et al. Construction of a risk score prognosis model based on hepatocellular carcinoma microenvironment. *World J Gastroenterol.* 2020;26(2):134-153.
- Ao L, Song X, Li X, et al. An individualized prognostic signature and multiomics distinction for early stage hepatocellular carcinoma patients with surgical resection. *Oncotarget.* 2016;7(17):24097-24110.
- Miyajima A, Tanaka M, Itoh T. Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. *Cell Stem Cell.* 2014;14(5):561-574.
- Reya T, Morrison SJ, Clarke MF, Weissman IL Stem cells, cancer, and cancer stem cells. *Nature.* 2001;414(6859):105-111.
- Britton KM, Kirby JA, Lennard TW, Meeson AP Cancer stem cells and side population cells in breast cancer and metastasis. *Cancers (Basel).* 2011;3(2):2106-2130.
- Chen D, Li Z, Cheng Q, et al. Genetic alterations and expression of PTEN and its relationship with cancer stem cell markers to investigate pathogenesis and to evaluate prognosis in hepatocellular carcinoma. *J Clin Pathol.* 2019;72(9):588-596.
- Noh CK, Wang HJ, Kim CM, et al. EpCAM as a Predictive Marker of Tumor Recurrence and Survival in Patients Who Underwent Surgical Resection for Hepatocellular Carcinoma. *Anticancer Res.* 2018;38(7):4101-4109.
- Zhou L, Zhu Y. The EpCAM overexpression is associated with clinicopathological significance and prognosis in hepatocellular carcinoma patients: a systematic review and meta-analysis. *Int J Surg.* 2018;56(15):274-280.
- Tsuchiya A, Kamimura H, Tamura Y, et al. Hepatocellular carcinoma with progenitor cell features distinguishable by the hepatic stem/progenitor cell marker NCAM. *Cancer Lett.* 2011;309(1):95-103.
- Xu X, Liu RF, Zhang X, et al. DLK1 as a potential target against cancer stem/progenitor cells of hepatocellular carcinoma. *Mol Cancer Ther.* 2012;11(3):629-638.
- Li XF, Chen C, Xiang DM, et al. Chronic inflammation-elicited liver progenitor cell conversion to liver cancer stem cell with clinical significance. *Hepatology.* 2017;66(6):1934-1951.
- Liu Y, Cao L, Chen R, et al. Osteopontin Promotes Hepatic Progenitor Cell Expansion and Tumorigenicity via Activation of beta-Catenin in Mice. *Stem Cells.* 2015;33(12):3569-3580.
- Yu G, Wang LG, Han Y, et al. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-287.
- Robinson MD, McCarthy DJ, Smyth GK. Edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-140.
- Li N, Li L, Chen Y. The Identification of Core Gene Expression Signature in Hepatocellular Carcinoma. *Oxid Med Cell Longev.* 2018;2018(11):3478305.
- Li B, Chng WJ. EZH2 abnormalities in lymphoid malignancies: underlying mechanisms and therapeutic implications. *J Hematol Oncol.* 2019;12(1):118.
- Ning X, Shi Z, Liu X, et al. DNMT1 and EZH2 mediated methylation silences the microRNA-200b/a/429 gene and promotes tumor progression. *Cancer Lett.* 2015;359(2):198-205.
- Xu X, Gu J, Ding X, et al. LINC00978 Promotes the progression of hepatocellular carcinoma by regulating EZH2-mediated silencing of p21 and E-cadherin expression. *Cell Death Dis.* 2019;10(10):752.
- Fu X, Lin J, Qin F, et al. LncAPC drives Wnt/beta-catenin activation and liver TIC self-renewal through EZH2 mediated APC transcriptional inhibition. *Mol Carcinog.* 2018;57(3):408-418.
- Gorodetska I, Lukiyanchuk K, Peitzsch C, Kozeretska I, Dubrovskaya A. BRCA1 and EZH2 cooperate in regulation of prostate cancer stem cell phenotype. *Int J Cancer.* 2019;145(11):2974-2985.
- Zhu P, Wang Y, Huang G, et al. Lnc-beta-Catm elicits EZH2-dependent beta-catenin stabilization and sustains liver CSC self-renewal. *Nat Struct Mol Biol.* 2016;23(7):631-639.
- Arai M, Yokosuka O, Fukai K, et al. Gene expression profiles in liver regeneration with oval cell induction. *Biochem Biophys Res Commun.* 2004;317(2):370-376.
- Wang X, Lopategi A, Ge X, et al. Osteopontin induces ductular reaction contributing to liver fibrosis. *Gut.* 2014;63(11):1805-1818.