



A Novel Five-Gene Signature for Prognosis Prediction in Hepatocellular Carcinoma

Lisa Su^{1†}, Genhao Zhang^{2†} and Xiangdong Kong^{1*}

¹ Department of Genetic and Prenatal Diagnosis Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, ² Department of Blood Transfusion, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:

Maja Cemazar,
Institute of Oncology
Ljubljana, Slovenia

Reviewed by:

Serena Bonin,
University of Trieste, Italy
Ching-Chieh Yang,
Chi Mei Medical Center, Taiwan

*Correspondence:

Xiangdong Kong
kongxdprenatal@163.com

[†]These authors have contributed
equally to this work and
share first authorship

Specialty section:

This article was submitted to
Molecular and Cellular Oncology,
a section of the journal
Frontiers in Oncology

Received: 22 December 2020

Accepted: 06 April 2021

Published: 16 July 2021

Citation:

Su L, Zhang G and Kong X (2021)
A Novel Five-Gene Signature
for Prognosis Prediction in
Hepatocellular Carcinoma.
Front. Oncol. 11:642563.
doi: 10.3389/fonc.2021.642563

Hepatocellular carcinoma (HCC) has been a global health issue and attracted wide attention due to its high incidence and poor outcomes. In this study, our purpose was to explore an effective prognostic marker for HCC. Five cohort profile datasets from GEO (GSE25097, GSE36376, GSE62232, GSE76427 and GSE101685) were integrated with TCGA-LIHC and GTEx dataset to identify differentially expressed genes (DEGs) between normal and cancer tissues in HCC patients, then 5 upregulated differentially expressed genes and 32 downregulated DEGs were identified as common DEGs in total. Next, we systematically explored the relationship between the expression of 37 common DEGs in tumor tissues and overall survival (OS) rate of HCC patients in TCGA and constructed a novel prognostic model composed of five genes (AURKA, PZP, RACGAP1, ACOT12 and LCAT). Furthermore, the predicted performance of the five-gene signature was verified in ICGC and another independent clinical samples cohort, and the results demonstrated that the signature performed well in predicting the OS rate of patients with HCC. What is more, the signature was an independent hazard factor for HCC patients when considering other clinical factors in the three cohorts. Finally, we found the signature was significantly associated with HCC immune microenvironment. In conclusion, the prognostic five-gene signature identified in our present study could efficiently classify patients with HCC into subgroups with low and high risk of longer overall survival time and help clinicians make decisions for individualized treatment.

Keywords: HCC (hepatocellular carcinoma), prognosis, signature, risk score, overall survival

INTRODUCTION

Hepatocellular carcinoma (HCC) has been a global health issue and attracted wide attention due to its high incidence and poor outcomes (1). It is reported that HCC results in nearly 850,000 new cases and more than 600,000 death every year (2), which seriously increases the disease burden in the worldwide. In China, HCC appears to be the most common pathological type of tumors and the major cause of cancer deaths (3). Moreover, the incidence and mortality are still escalating according to the increased alcohol abuse, cirrhosis, aflatoxin exposure, diabetes, metabolic syndrome and obesity (4–9). With the rapid developments in therapeutic strategies including liver resection, liver transplantation, radiofrequency ablation, embolization therapy and immune

checkpoint inhibitors, HCC patients could access potential treatment strategies at early and intermediate stages of HCC prognosis (10). However, over half of HCC patients are in advanced stage when diagnosed, only 15% of which are suitable for curative therapies, and the five-year survival rate remains very low, no more than 20%, according to lacking of biomarkers for diagnosis at early stage and the high frequency of recurrence (11, 12). Consequently, it is necessary for us to seek novel prognostic markers for improving the poor outcomes of HCC patients.

Considering the widely exploited next-generation sequencing technology in life sciences, integrating prognosis-related gene signature shows a great advantage in the prediction of HCC prognosis by using TCGA and GEO program. In our current study, a five-mRNA signature associated with overall survival (OS) rate of HCC patients was established, and the robustness of which was externally validated by ICGC dataset and another independent cohort. We explored the association between the prognostic signature and invasion of immune cells in HCC at the same time. Initial construction of the five-gene signature for patients with HCC will help clarify the underlying mechanism of HCC, accurately predict HCC prognosis and make a meaningful contribution to therapeutic strategies.

MATERIALS AND METHODS

Public Patient Datasets and Identification of Differentially Expressed Genes (DEGs)

The mRNA expression data and clinical characteristics of patients with HCC from seven publicly available datasets including TCGA-LIHC and GTEx, GSE25097, GSE36376, GSE62232, GSE76427, GSE101685 and ICGC (LIRI-JP) were incorporated into our study. After log₂ transformed and quantile normalized, the mRNA expression data were calculated by mean expression when more than one probe were detected. The “LIMMA” package was used to screen common DEGs between normal and tumor samples in the five databases from GEO with a cut-off criterion of adjust *P* value <0.05 and |log₂FC|>1. Then “edgeR” package was used to obtain DEGs from TCGA-LIHC and GTEx cohort with the same cut-off criterion.

Signature Identification and Risk-Score Model Construction

After common DEGs were screened from GEO and TCGA-LIHC datasets, the association between common DEGs and OS rate in TCGA-LIHC cohort was calculated by a univariable Cox proportional analysis. Then LASSO-Cox regression method with the “glmnet” package was performed to determine significant prognosis-related DEGs. The stepwise Cox regression analysis was used to evaluate the above prognosis-related genes and establish a prognostic signature. Risk score was finally established based on the basis of linearly combining the formula below with the mRNA expression level multiplied the multivariate Cox regression coefficient (β) model. Risk score = ($\beta_{mRNA1} \times mRNA1$) + ($\beta_{mRNA2} \times mRNA2$) + ... + ($\beta_{mRNA_n} \times$

mRNA_n). We stratified patients in TCGA database into high-risk or low-risk score subgroups due to optimal risk score threshold determined by R package “survival” and “survminer”. The predictive power and independence of the prognostic signature in TCGA were assessed by ROC analysis, Kaplan-Meier survival analysis and Cox proportional hazards regression analysis. The genomic alterations including mutations and putative copy-number alterations of genes in the signature were analyzed on the open platform of cBioPortal (<http://www.cbioportal.org/>) (13). The mRNA and protein expression of genes in the signature were explored in Gene Expression Profiling Interactive Analysis (GEPIA) (14) (<http://gepia.cancer-pku.cn/>) and Human Protein Atlas (HPA, www.proteinatlas.org), respectively. Furthermore, the expression of five genes in HCC cell lines were explored in Cancer Cell Line Encyclopedia (CCLE) (15) (<https://portals.broadinstitute.org/ccle>).

Functional Analysis

The mRNAs positively and negatively correlated with the prognostic risk score in TCGA-LIHC dataset are calculated by Pearson correlation analysis (|*R*|>0.5 and *P*<0.001). GO and KEGG pathway analysis were used for functional annotation of the positively and negatively correlated genes by using R package “clusterProfiler”.

Clinical Specimens and Quantitative Real-Time PCR (qRT-PCR) Analysis

Previous collected 59 fresh frozen tissues from HCC patients were selected as an independent validation cohort (16). qRT-PCR was used to detect the mRNA levels of the five gene (17), after the relative mRNAs expression levels were normalized to β -ACTIN and log₂ transformed. Primer sequences are showed in **Table S1**.

External Validation of the Five-mRNA Signature

ICGC dataset was downloaded from the ICGC portal (<https://dcc.icgc.org/projects/LIRI-JP>). Risk score of patients with survival data in ICGC and our independent cohort was calculated with above formula, and we stratified patients into high-risk or low-risk score subgroups. Then predictive power and independence of the prognostic model were evaluated in the two cohorts.

Calculation of Immune Status and Immune Infiltrates Analysis

To assess the immune status of each sample, ESTIMATE algorithm was applied to estimate stromal and immune cells calculated from the TCGA cohort, measuring stromal levels (stromal scores) and degree of immunocyte infiltration (immune scores) (18). The association between prognostic risk score and immune, stromal scores were analyzed by Pearson correlation analysis. Furthermore, CIBERSORT algorithm, a kind of deconvolution algorithm based on gene expression, was applied to assess the cell composition of different tumor-

infiltrating immune cells (TIICs) and to further examine the relationship between the prognostic risk scores and the immune microenvironment in TCGA database (19) (<http://cibersort.stanford.edu/>). Finally, the predictive ability of significantly changed TIIC was assessed by Kaplan-Meier survival analysis.

Statistical Analysis

ROC curve analysis and Kaplan-Meier survival analysis were performed to assess the prediction performance of OS rate with R software (Version 4.0.3). Cox proportional model was performed to analyze relationship between prognostic signature and OS rate, together with other clinical features including age, gender, grade, TNM stage, vascular invasion, tumor status, cirrhosis, hepatitis virus infection and Child-pugh scores. Clinical characteristics of HCC patients in TCGA, ICGC and clinical validation cohorts were showed in **Table S2**. Results were considered statistically significant when P value <0.05.

RESULTS

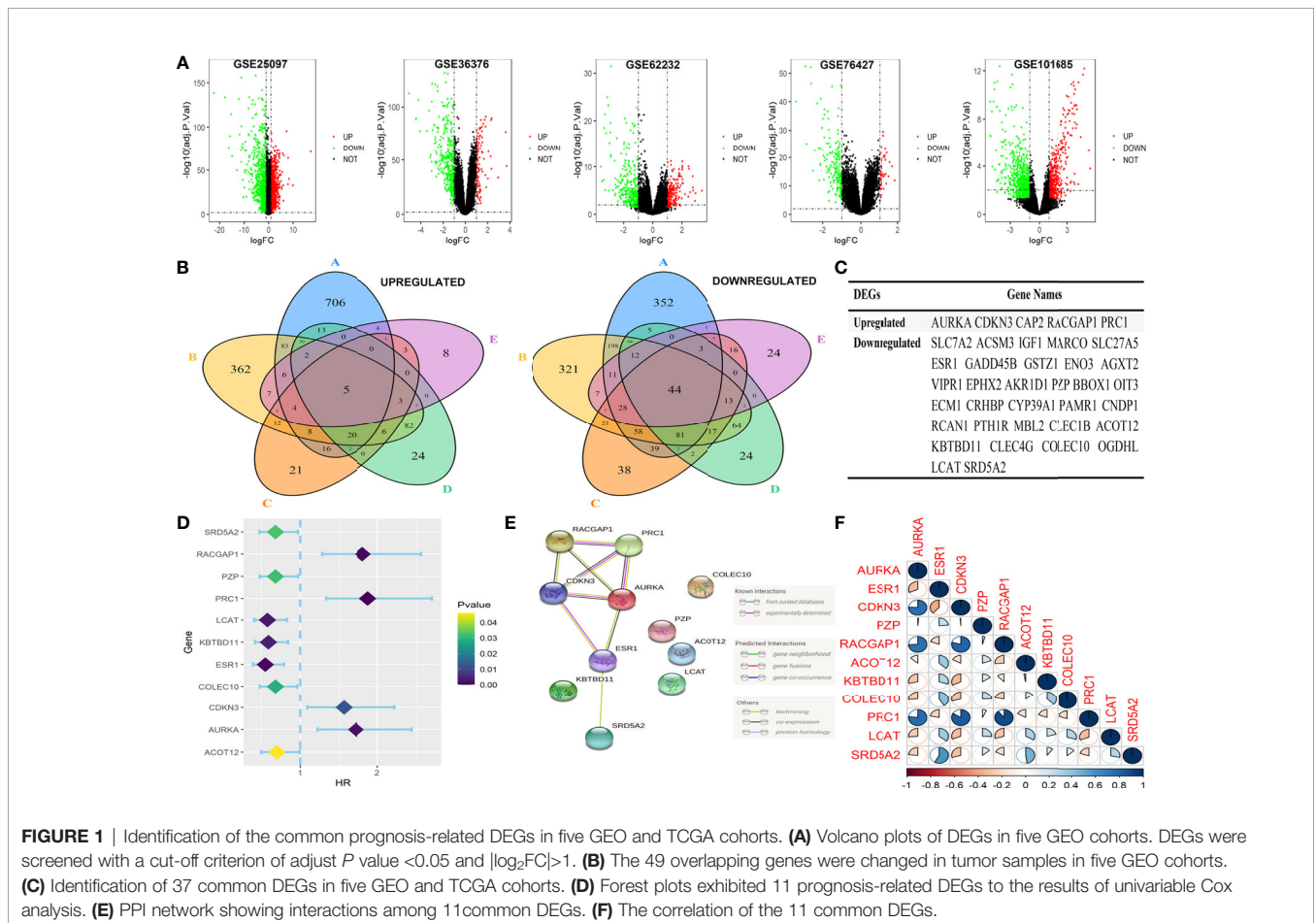
DEGs Identification in Patients With HCC

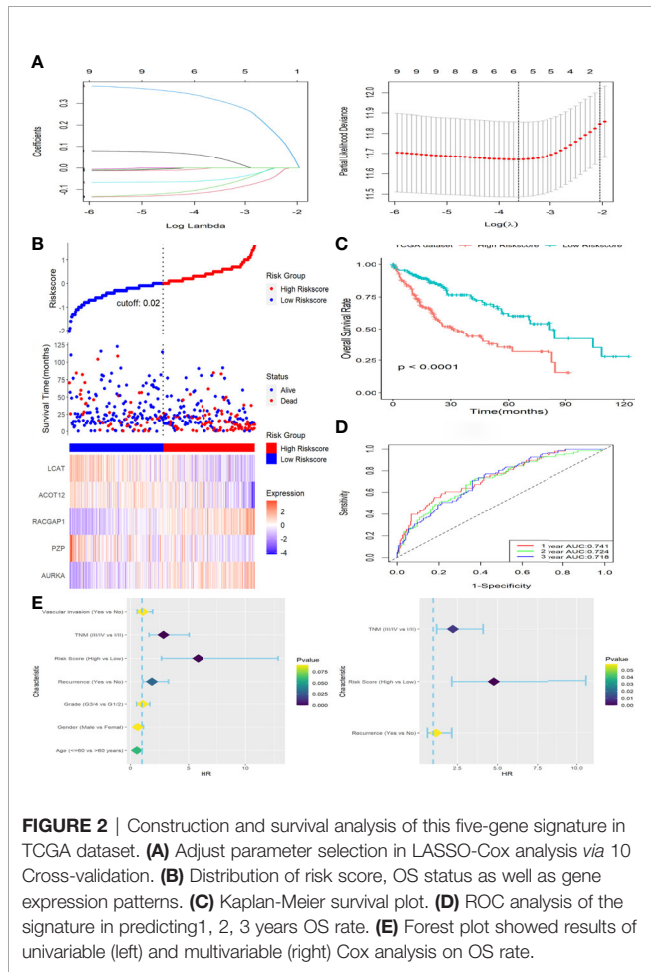
DEGs in the five GEO datasets was screened with a cut-off standard of P value <0.05 and $|\log_2FC| > 1$ first as shown in **Figure 1A**.

Five upregulated and 44 downregulated genes were identified as common DEGs in the five GEO datasets (**Figure 1B**). When considering DEGs from TCGA-LIHC and GTEx cohort with the same cut-off criterion, 37 genes (5 upregulated and 32 downregulated, **Figure 1C**) were identified as common DEGs in both GEO, TCGA-LIHC and GTEx databases and were further analyzed.

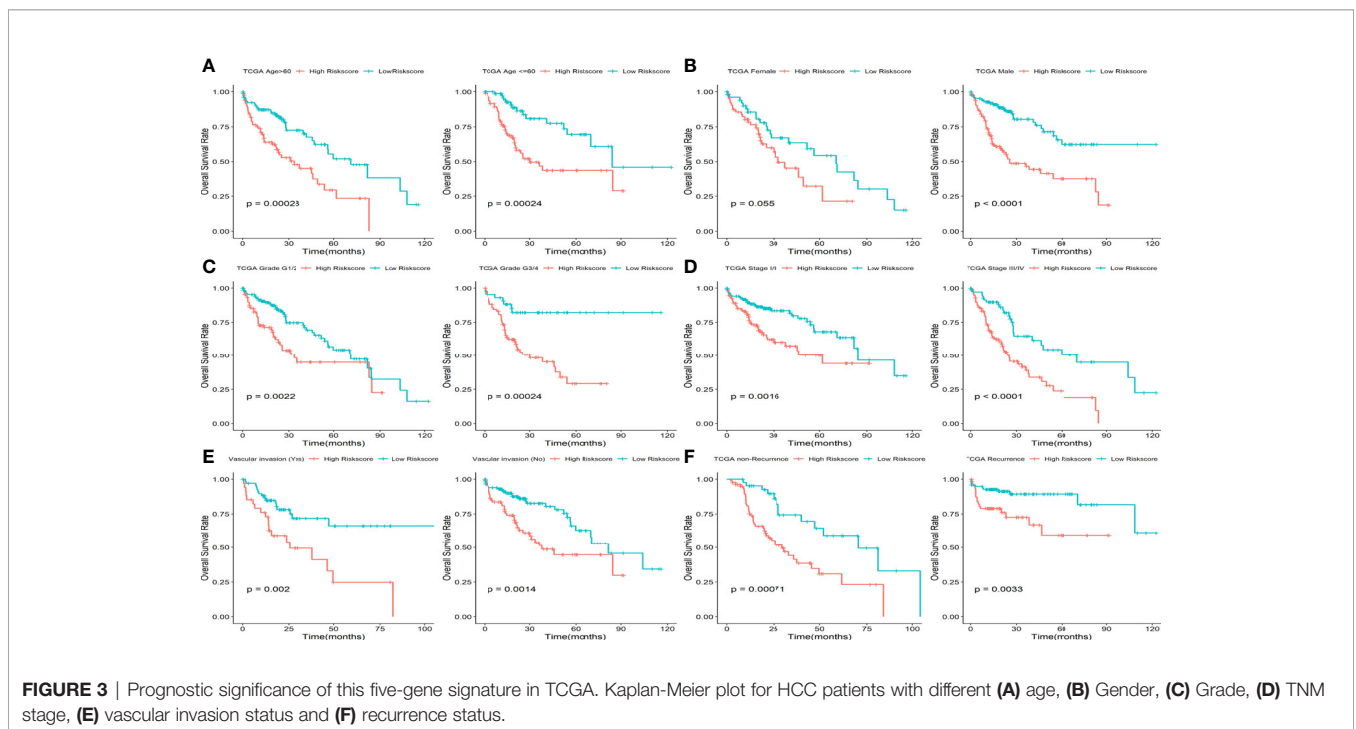
Establishment of a Prognosis-Related Signature in TCGA

Eleven of the thirty-seven common DEGs were significant related to OS rate of patients with HCC in TCGA as shown in **Figure 1D**. The interworking network showed that ESR1, AURKA and CDKN3 were the central genes (**Figure 1E**) and the relevance of the eleven genes is shown in **Figure 1F**. Then significant prognostic genes among the 11 genes above were selected by performing the LASSO-Cox regression model based on the minimum value of λ and a prognostic five-gene signature was finally identified *via* a stepwise Cox proportional model (**Figure 2A**). Risk score = $(0.0996 \times AURKA) - (0.1421 \times PZP) + (0.3809 \times RACGAP1) - (0.0742 \times ACOT12) - (0.1438 \times LCAT)$. To determine whether there is a multicollinearity problem among five genes (AURKA, PZP, RACGAP1, ACOT12 and LCAT) expression and clinicopathological features, collinearity





statistics was performed with a cut-off standard of tolerance >0.1 and variance inflation factor (VIF) <10. In fact, no multicollinearity problem exists as shown in **Table S3**. Risk score of patients in TCGA was calculated with the above formula, and patients were stratified into high-risk or low-risk subgroups with an optimal risk score threshold (**Figure 2B**). Result of Kaplan-Meier survival analysis revealed patients with higher risk score were significantly relevant to poor OS rate (**Figure 2C**). Result of ROC analysis revealed that this signature had a good prognostic performance, and the AUCs were 0.741, 0.724, 0.718 at 1-, 2-, 3-year, respectively (**Figure 2D**). Furthermore, we explored prognostic performance of the five-gene signature in patients with different clinical features including age, vascular invasion, grade, recurrence, TNM stage and gender, the results of which revealed that higher risk scores had a statistically relationship with shorter OS time in different clinical subgroups (**Figure 3**). Finally, statistically significant variables obtained from univariable Cox regression analysis were input into multivariate Cox regression analysis, and the results revealed that TNM stage, recurrence and risk score were statistically associated with OS, while TNM stage (HR = 2.228, 95%CI 1.208-4.108, $P = 0.010$) and risk score (HR = 4.773, 95%CI 2.157-10.561, $P = 0.000$) were independent prognostic factors in TCGA (**Figure 2E**). In order to discover the coefficient prediction efficiency of the prognosis-related signature, a nomogram model was established, the result of which showed that the nomogram with a C-index of 0.749 could help us provide a quantitative method for predicting the 1-, 2-, 3-year survival rate accurately (**Figure 4A**). The overlap of the forecasted probability and the actual probability of 1-, 2-, 3-year survival in calibration curve indicated a well agreement (**Figures 4B-D**).



Verification of the Signature in ICGC Cohort

To verify capability of the five-gene signature, ICGC dataset was downloaded as a validation cohort. The risk score of patients was calculated with the same formula, and we stratified patients into low-risk or high-risk subgroups (Figure 5A). Result of Kaplan-Meier survival analysis revealed that patients with higher risk score were prominently relevant to poor OS rate (Figure 5B). Result of ROC analysis revealed that the five-gene signature had a good prognostic performance, and the AUCs were 0.727, 0.720, 0.725 at 1-, 2-, 3-year, respectively (Figure 5C). Furthermore, the predictive performance of the five-gene signature was explored in patients with different clinical features such as age, gender and TNM stage, the results of which revealed that higher risk scores had a statistically relationship with shorter OS time in different clinical subgroups (Figure 6). Finally, statistically significant variables obtained from univariable Cox regression analysis were input into multivariate Cox regression analysis, and the results revealed that TNM stage, gender and risk score were statistically relevant to OS of HCC patients, furthermore, TNM stage (HR=2.246, 95%CI 1.162-4.339, $P = 0.016$), gender (HR=0.361, 95%CI 0.185-0.704, $P = 0.002$) and risk score (HR=4.662, 95%CI 2.044-10.632, $P = 0.000$) were independent prognostic factors for OS of patients with HCC in ICGC (Figure 5D).

Verification of the Five-Gene Signature in a Clinical Cohort

To verify capability of the 5-gene signature in the prediction of OS rate in actual clinical practice, qRT-PCR analysis was performed in a

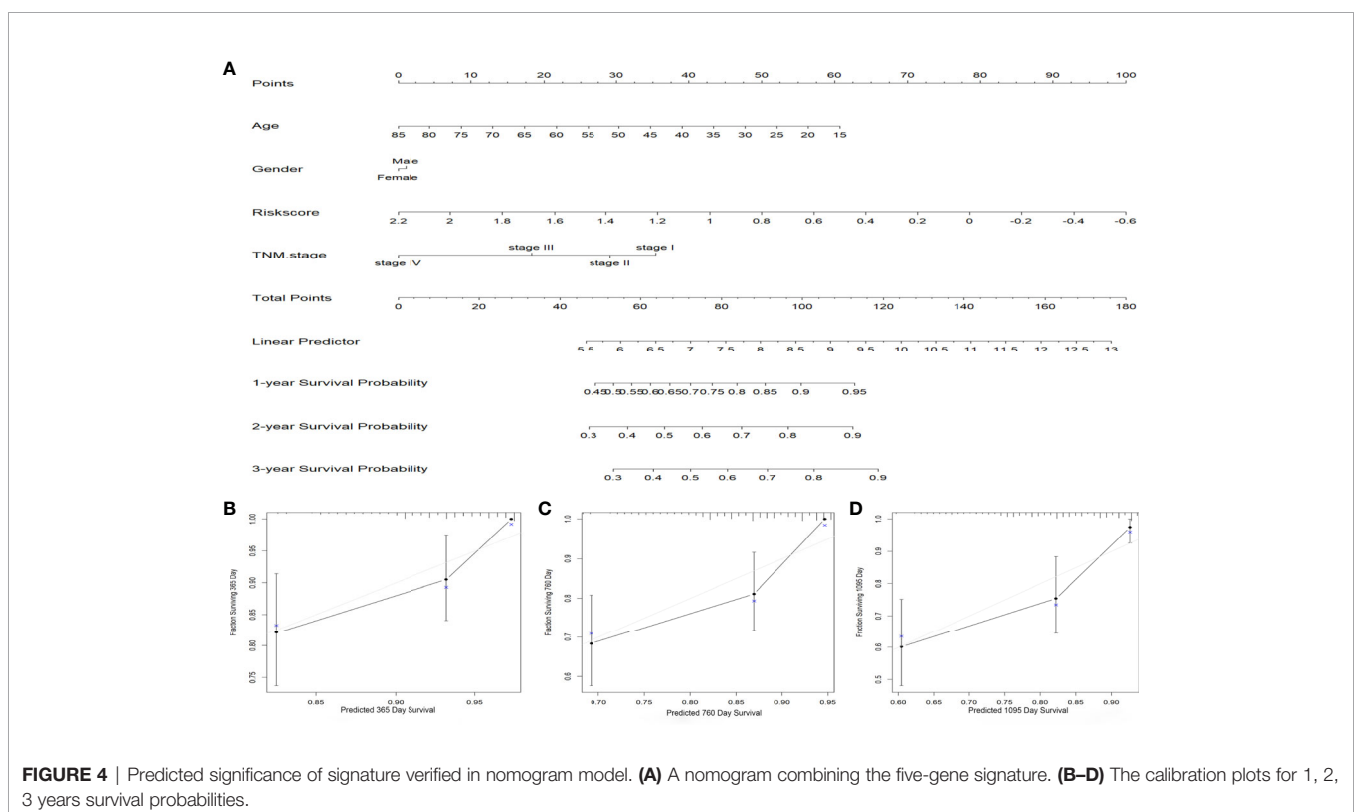
clinical cohort. Risk score of patients was calculated with the same formula, and we stratified patients into high-risk or low-risk groups (Figure 7A). Result of Kaplan-Meier survival analysis revealed that patients with higher risk score were prominently relevant to poor OS rate (Figure 7B). Result of ROC analysis revealed that this signature had a good prognostic performance, and the AUCs were 0.803, 0.707, 0.701 at 1-, 2-, 3-year, respectively (Figure 7C). Furthermore, statistically significant variables obtained from univariable Cox regression analysis were input into multivariate Cox regression analysis, and the results revealed that HBV infection, risk score, NASH and recurrence were statistically relevant to OS of HCC patients, while the risk score (HR = 4.663, 95%CI 1.716-21.387, $P = 0.047$) was only independent prognostic factors (Figure 7D).

Functional Analysis

According to the results of Pearson correlation analysis, screened 989 positively and 32 negatively correlated genes enriched GO terms could be classified into three functional clusters such as cell cycle and proliferation, lipid transport and localization, ATP metabolic process, while the mainly enriched pathway were cell cycle, spliceosome, complement and coagulation cascades (Figure 8).

Genetic Alterations and Expression Level of the Genes in the Five-Gene Signature

366 samples with mutation and CNA data are available in cBioPort platform. AURKA, PZP, RACGAP1, ACOT12 and LCAT had missense mutation in 0.3%, 2.4%, 1.3%, 1.3%, 0.8% of HCC samples, respectively. Besides, AURKA, PZP, ACOT12 and LCAT had amplification in 1.6%, 0.3%, 0.3%, 0.3% of HCC



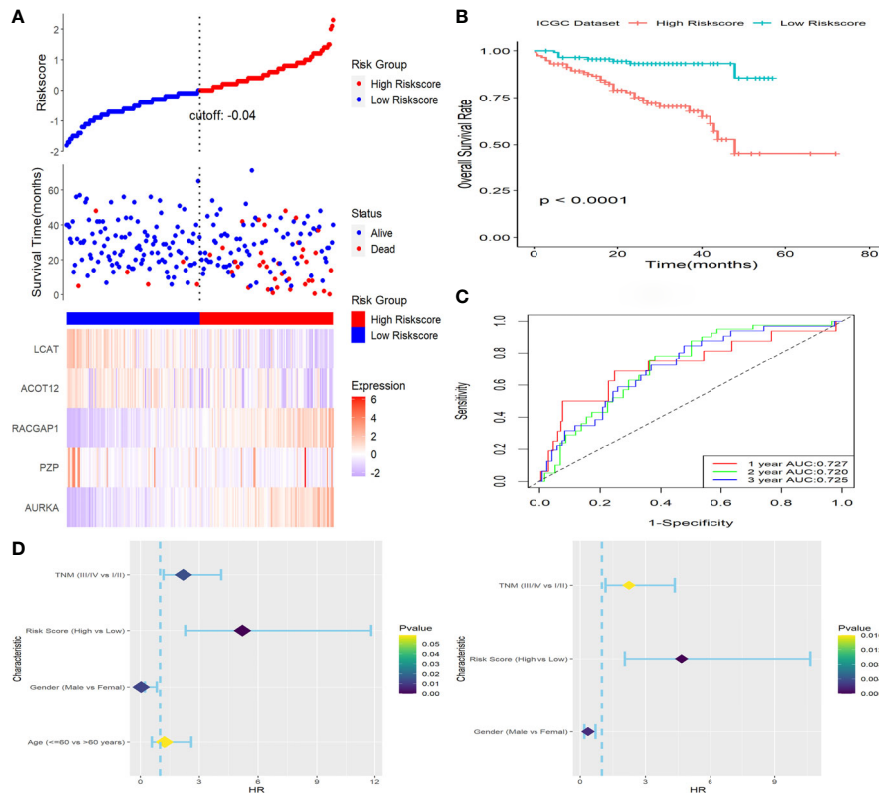


FIGURE 5 | Verification of this signature in ICGC. **(A)** Distribution of risk score, OS status as well as gene expression patterns. **(B)** Kaplan-Meier survival plot. **(C)** ROC analysis of the signature in predicting 1, 2, 3 years OS rate. **(D)** Forest plot showed results of univariable (left) and multivariable (right) Cox analysis on OS rate.

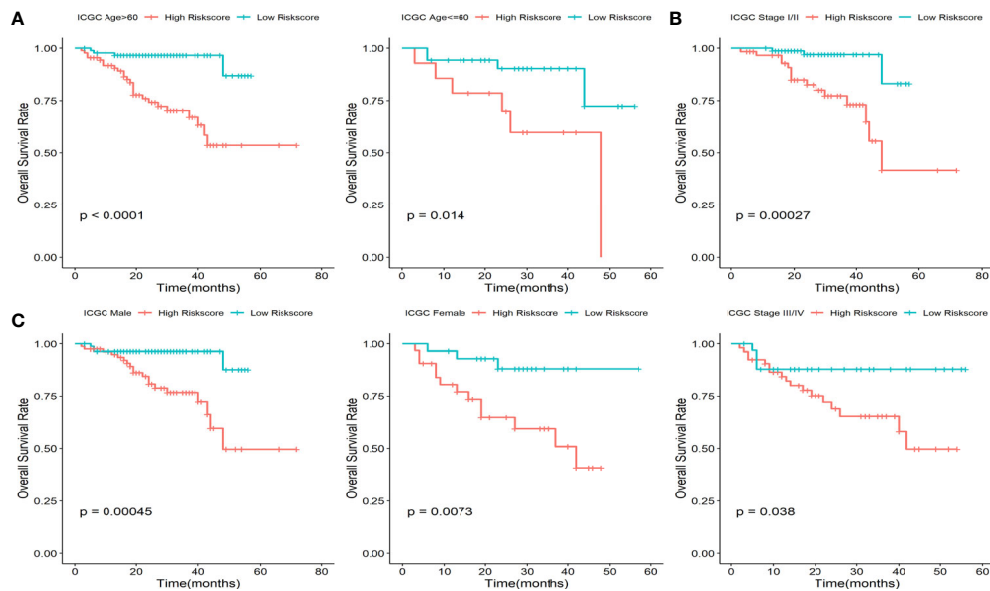


FIGURE 6 | Prognostic significance of this five-gene signature in ICGC cohort. Kaplan-Meier plot for HCC patients with different **(A)** age, **(B)** TNM stage and **(C)** Gender.

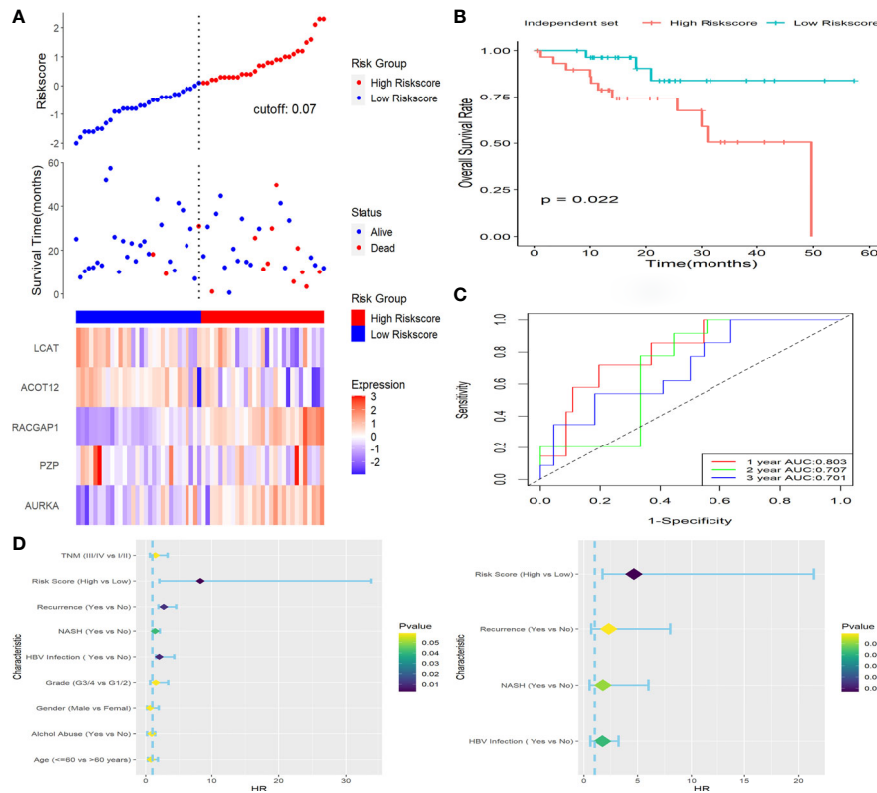


FIGURE 7 | Verification of this signature in a clinical dataset. **(A)** Distribution of risk score, OS status as well as gene expression patterns. **(B)** Kaplan-Meier survival plot. **(C)** ROC analysis of the signature in predicting 1, 2, 3 years OS rate. **(D)** Forest plot showed results of univariable (left) and multivariable (right) Cox analysis on OS rate.

samples, respectively. No putative copy-number alterations were found for RACGAP1. Amplification is the most common form of genetic alteration (**Figure S1A**). Additionally, the mRNA expression levels of the five genes differed significantly between tumor and normal specimens in GEPIA (**Figure S1B**), so as the protein expression levels of AURK, PZP and RACGAP1 in HPA (**Figure S1C**). Unfortunately, ACOT12 and LCAT were not found in HPA. Furthermore, the expression levels of these five genes were significantly different between HCC cell lines in CCLL (**Figure S1D**).

Association Between Risk Score and Infiltration of Immune Cells in TCGA

On the strength of ESTIMATE algorithm, the prognostic risk scores have a significant association with stromal scores, but not immune scores, and samples in high-risk group had lower stromal scores when compared with samples in low-risk group (**Figure S2**), indicating that the prognostic risk scores had a close connection with tumor immune status. In the following, to understand how the prognostic risk scores reflected immune status of HCC, distinctions and correlations of 22 types of TIICs in two subgroups in TCGA were assessed *via* CIBERSORT (**Figures 9A, B**). According to the difference analysis, the proportion of macrophages.M0 was down-regulated in low-risk

score group, while the proportion of resting memory CD4 T cells was up-regulated (**Figure 9C**). Furthermore, patients with decreased resting memory CD4 T cells were prominently relevant to poor prognosis (**Figure 9D**).

DISCUSSION

Although the TNM staging system and the prognostic scoring systems of American Joint Committee on Cancer are implemented to assess the prognosis of HCC patients, in current predictive methods each system cannot always be effective in predicting the prognosis according to the losing sight of different genetic and epigenetic backdrops of tumors (20). Herein, it becomes more important than ever before to explore effective biomarkers for improving the prognosis. In previous time, although many studies tried to clarify molecular mechanisms of HCC development, the achievement is still unsatisfied, because the molecular pathogenesis is extremely complex and heterogeneous (21). In this study, bioinformatics methods and expression profiling techniques have been performed to identify DEGs in patients with HCC. Five cohort profile datasets from GEO (GSE25097, GSE36376, GSE62232, GSE76427 and GSE101685) were integrated with TCGA-LIHC and GTEx dataset, then 5 upregulated DEGs and 32 downregulated DEGs

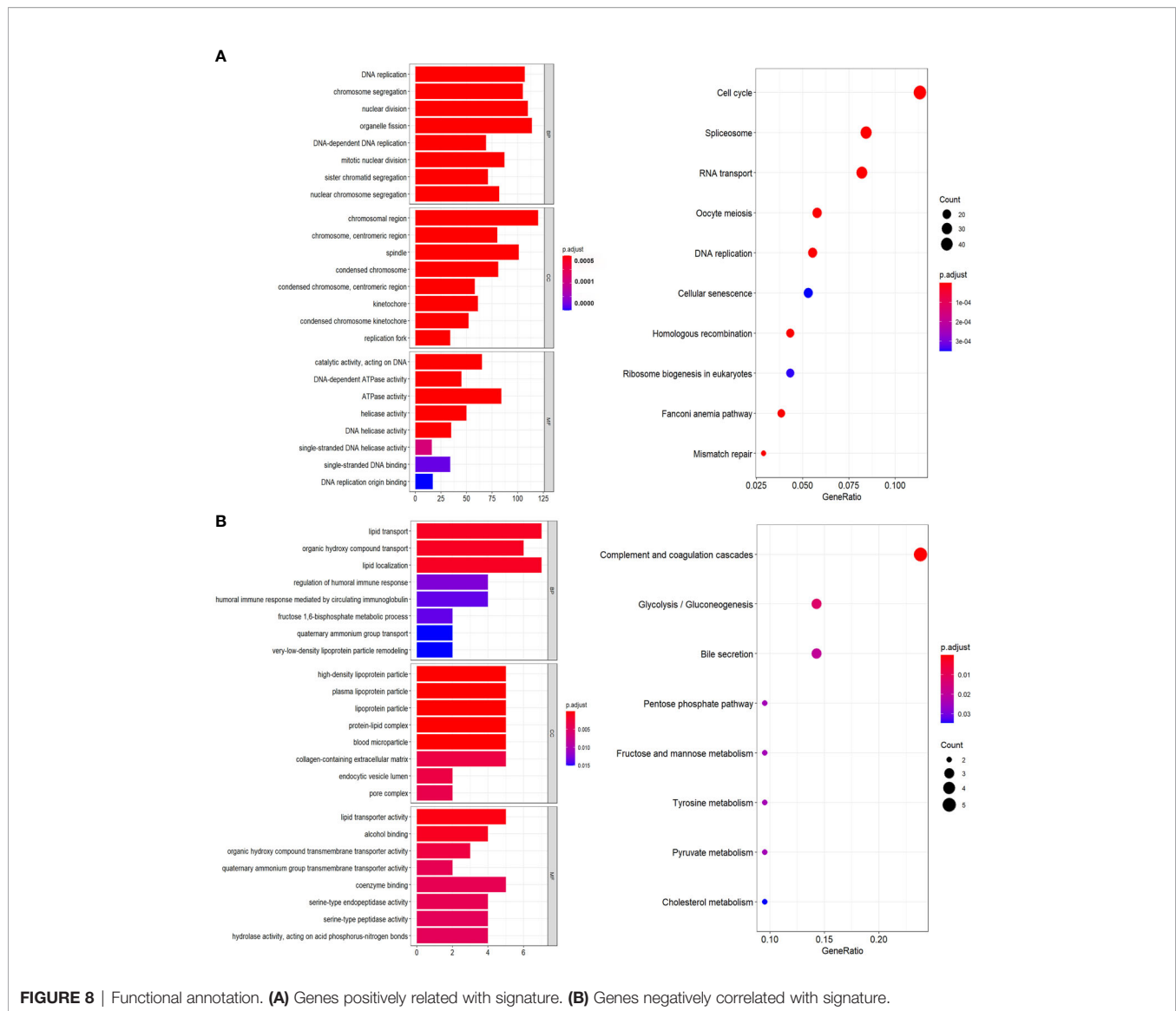
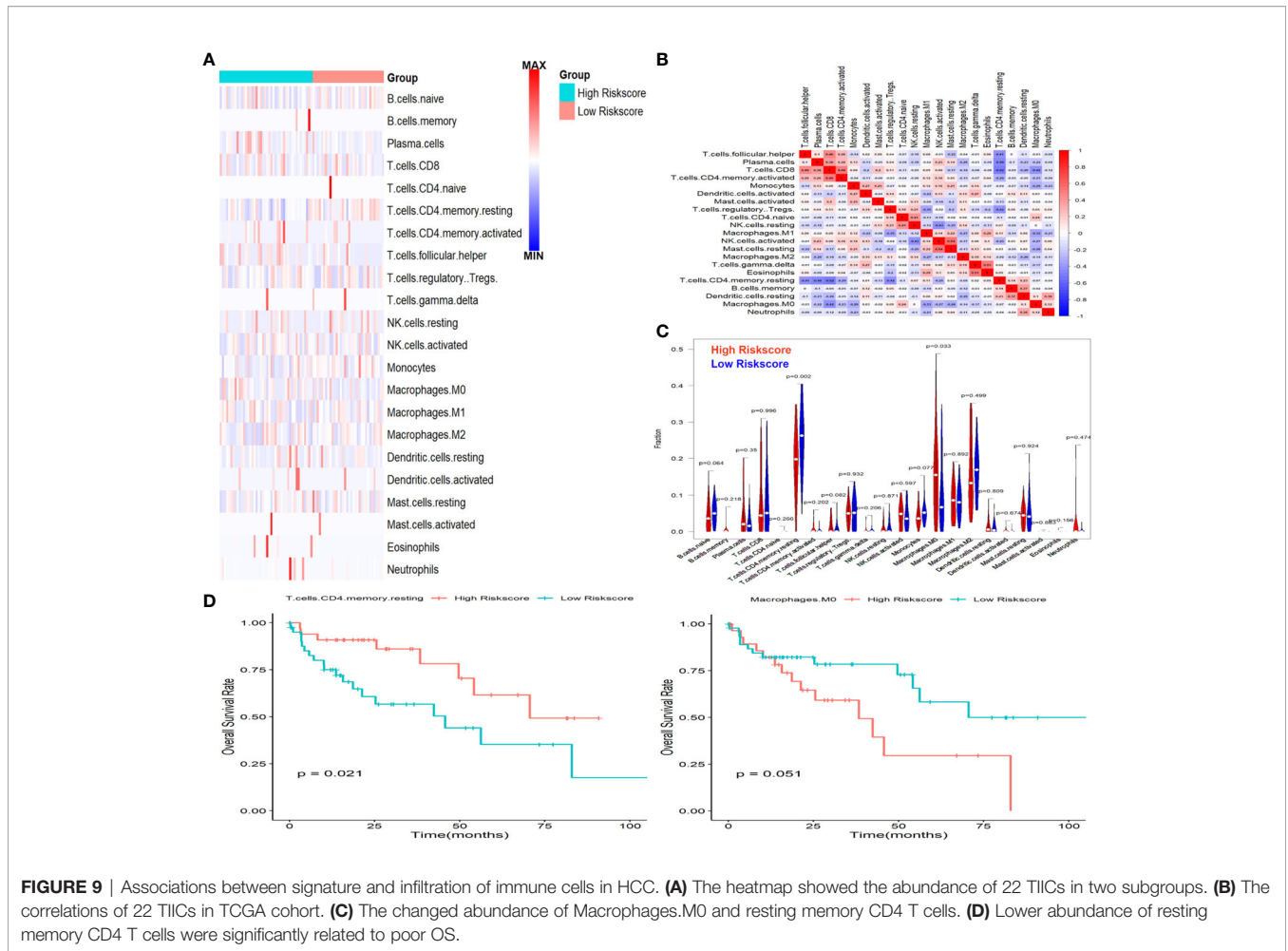


FIGURE 8 | Functional annotation. **(A)** Genes positively related with signature. **(B)** Genes negatively correlated with signature.

were identified as common DEGs in total and were further analyzed. Next, we systematically explored the relationship between the expression of 37 common DEGs in tumor tissues and OS rate in TCGA and constructed a novel prognosis-related model composed of five genes (AURKA, PZP, RACGAP1, ACOT12 and LCAT). Furthermore, the predicted value of the signature was validated in ICGC and another independent clinical samples cohort, which demonstrated that the signature had an excellent performance in the prediction of OS rate. What is more, it was an independent risk factor for patients with HCC when considering other clinical factors in the three cohorts. Finally, when we tried to explore the potential mechanisms, we found that the prognostic risk score had a close connection with stromal score, and patients in high-risk score group exhibited an increased abundance of Macrophages.M0 and decreased abundance of resting memory CD4 T cells in TCGA-LIHC dataset, which suggested that the signature was significantly associated with HCC immune microenvironment.

Most of the genes in our five-gene signature had been reported to participate in cancer development. AURKA was reported to act a pivotal part in abrogating G2 checkpoint induced by DNA damage in urothelial cell carcinoma (22, 23). Inhibition of AURKA could reduce the activity of poly (ADP-ribose) 1 level and promote non-homologous end joining repair (NHEJ) mechanisms in ovarian carcinoma cells (24). Blocking the β -catenin pathway and inhibiting AURKA activity at the same time may enhance antitumor response in adrenocortical cancer (25). AURKA was essential for mediating TGF- β induced plasticity and chemoresistance in triple-negative breast cancer (26). Liu (27) found that knockdown of AURKA could lead to increased radiotherapy efficacy in human colorectal cancer. Moreover, overexpressed AURKA might promote hepatocellular carcinoma cell growth, adhesion and migration *in vitro* (28). RACGAP1 has been identified as a hub gene and significantly associated with overall survival in patients with



various cancer types, such as pancreatic ductal adenocarcinoma (29), gastric cancer (30), cervical cancer (31) and lung adenocarcinoma (32). Furthermore, Zhao has found that depletion of RACGAP1 could lead to mitotic catastrophe and massive cell death in esophageal squamous cell carcinoma (33). Kehan confirmed that RACGAP1 played an essential role in breast cancer metastasis by modulating ECT2-dependent mitochondrial quality control (34). Yong found that RACGAP1 overexpression was significantly related to poor prognosis in HCC patients, and could promote proliferation ability of HCC cells *via* inhibiting activation of the Hippo and YAP pathways (35). Chen found that ten hub genes associated with immune infiltration, including RACGAP1 and AURKA, could predict survival outcome in HCC *via* bioinformatics analysis (36). The down-regulated ACOT12 was significantly associated with poor diagnosis in metastatic HCC patients (37). ACOT12 could facilitate metastasis through epigenetic induction of epithelial-mesenchymal transition in HCC (38). Previous study found that LCAT was overexpressed in the sera of high-grade and lymph-node-positive breast cancer and could be a common plasma protein marker in aggressive breast cancer (39). Besides, LCAT was hypermethylated and decreased in HCC

tissues (40), could act as a good biomarker at predicting HCC diagnosis, prognosis and recurrence (41), and it was confirmed in the later study by Long (42). As for PZP, in previous study we have confirmed that its depletion was significantly related to poor survival outcomes in HCC (16). All these indicated that AURKA, PZP, RACGAP1, ACOT12 and LCAT played important roles in the development of HCC and might be targets for immunotherapeutic intervention strategies in further.

In the present study, we identified a prognosis-related biomarker to stratify HCC patients and forecast the prognosis. When compared with previous signatures (43–48), some genes in the signature, such AURKA, LCAT and ACOT12, were also identified as hub genes in predicting survival outcomes of HCC patients, in agreement with our study. Additionally, the five-gene signature had some novelties although all of them could effectively predict the prognosis of HCC patients. Firstly, five GEO datasets, TCGA-LIHC and GTEx were incorporated into our study to screen common DEGs between normal and tumor samples, making the DEGs more reliable than those in previous research. Secondly, in order to ensure the clinical relevance, qRT-PCR analysis was performed to validate the signature in a clinical cohort after the signature was constructed in TCGA cohort. Thirdly, strong prognostic performance among

different clinical characters made the five-gene signature more attractive for clinical implementation. Finally, the five-gene signature contains fewer genes to make it more easily to implement in comparison with previous signature. There is no denying that our present study existed certain limitations. The great heterogeneity of HCC and the mechanisms of post-curative recurrence might decrease the performance of the prognostic model. Moreover, the association between prognostic risk score and infiltration of immune cells maybe inaccurate when considering it is based on estimated tumor characteristics. In addition, multicenter randomized controlled studies combining mRNA, single-nucleotide polymorphism, CpG and long non-coding RNA are needed to investigate the five-gene signature in the future.

In conclusion, a prognostic five-gene signature was identified in our present study, which could efficiently classify HCC patients and help clinicians make decisions for individualized treatment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committees of Zhengzhou University. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GZ designed the study, downloaded and analyzed the data. LS wrote the manuscript. XK critically reviewed the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.642563/full#supplementary-material>

Supplementary Figure 1 | Genetic alterations and expression level of genes in the prognostic significance. **(A)** The genetic alteration of the five genes in TCGA-LIHC. **(B)** The mRNA expression levels of the five genes in GEPIA. * $p < 0.05$. **(C)** The protein expression levels of the five genes in HPA. **(D)** The expression levels of the five genes in CCLE.

Supplementary Figure 2 | Relationship between prognostic significance and immune status in TCGA. Correlation **(A)** and difference **(B)** analysis of prognostic risk scores with immune status.

Supplementary Table 1 | The sequences of the qRT-PCR primers.

Supplementary Table 2 | Clinical characteristics of HCC patients involved in TCGA, ICGC and the clinical cohort.

REFERENCES

- Rao CV, Asch AS, Yamada HY. Frequently Mutated Genes/Pathways and Genomic Instability as Prevention Targets in Liver Cancer. *Carcinogenesis* (2017) 38(1):2–11. doi: 10.1093/carcin/bgw118
- Satriano L, Lewinska M, Rodrigues PM, Banales JM, Andersen JB. Metabolic Rearrangements in Primary Liver Cancers: Cause and Consequences. *Nat Rev Gastroenterol Hepatol* (2019) 16(12):748–66. doi: 10.1038/s41575-019-0217-8
- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA Cancer J Clin* (2016) 66(2):115–32. doi: 10.3322/caac.21338
- Bravi F, Tavani A, Bosetti C, Boffetta P, La Vecchia C. Coffee and the Risk of Hepatocellular Carcinoma and Chronic Liver Disease: A Systematic Review and Meta-Analysis of Prospective Studies. *Eur J Cancer Prev* (2017) 26(5):368–77. doi: 10.1097/CEJ.0000000000000252
- Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, Obesity, and Mortality From Cancer in a Prospectively Studied Cohort of U.S. Adults. *N Engl J Med* (2003) 348(17):1625–38. doi: 10.1056/NEJMoa021423
- Ioannou GN, Bryson CL, Weiss NS, Miller R, Scott JD, Boyko EJ. The Prevalence of Cirrhosis and Hepatocellular Carcinoma in Patients With Human Immunodeficiency Virus Infection. *Hepatology* (2013) 57(1):249–57. doi: 10.1002/hep.25800
- Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JP. Type 2 Diabetes and Cancer: Umbrella Review of Meta-Analyses of Observational Studies. *Bmj* (2015) 350:g7607. doi: 10.1136/bmj.g7607
- Schlesinger S, Aleksandrova K, Pischon T, Jenab M, Fedirko V, Trepo E, et al. Diabetes Mellitus, Insulin Treatment, Diabetes Duration, and Risk of Biliary Tract Cancer and Hepatocellular Carcinoma in a European Cohort. *Ann Oncol* (2013) 24(9):2449–55. doi: 10.1093/annonc/mdt204
- Yang C, Huang X, Liu Z, Qin W, Wang C. Metabolism-Associated Molecular Classification of Hepatocellular Carcinoma. *Mol Oncol* (2020) 14(4):896–913. doi: 10.1002/1878-0261.12639
- Forner A, Reig M, Bruix J. Hepatocellular Carcinoma. *Lancet* (2018) 391(10127):1301–14. doi: 10.1016/S0140-6736(18)30010-2
- Hollebecque A, Malka D, Ferté C, Ducreux M, Boige V. Systemic Treatment of Advanced Hepatocellular Carcinoma: From Disillusions to New Horizons. *Eur J Cancer* (2015) 51(3):327–39. doi: 10.1016/j.ejca.2014.12.005
- Fujiwara N, Friedman SL, Goossens N, Hoshida Y. Risk Factors and Prevention of Hepatocellular Carcinoma in the Era of Precision Medicine. *J Hepatol* (2018) 68(3):526–49. doi: 10.1016/j.jhep.2017.09.016
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. the Cbio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095
- Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A Web Server for Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res* (2017) 45(W1):W98–w102. doi: 10.1093/nar/gkx247
- Nusinow DP, Szpyt J, Ghandi M, Rose CM, McDonald ER, Kalocsay M, et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* (2020) 180(2):387–402.e316. doi: 10.1016/j.cell.2019.12.023
- Su L, Zhang G, Kong X. Prognostic Significance of Pregnancy Zone Protein and Its Correlation With Immune Infiltrates in Hepatocellular Carcinoma. *Cancer Manag Res* (2020) 12:9883–91. doi: 10.2147/CMAR.S269215
- Zhang G. Expression and Prognostic Significance of BANF1 in Triple-Negative Breast Cancer. *Cancer Manag Res* (2020) 12:145–50. doi: 10.2147/CMAR.S229022
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring Tumour Purity and Stromal and Immune Cell Admixture From Expression Data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
- Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. the Prognostic Landscape of Genes and Infiltrating Immune Cells Across Human Cancers. *Nat Med* (2015) 21(8):938–45. doi: 10.1038/nm.3909

20. Cai J, Tong Y, Huang L, Xia L, Guo H, Wu H, et al. Identification and Validation of a Potent Multi-mRNA Signature for the Prediction of Early Relapse in Hepatocellular Carcinoma. *Carcinogenesis* (2019) 40(7):840–52. doi: 10.1093/carcin/bgz018
21. Bruix J, Gores GJ, Mazzaferro V. Hepatocellular Carcinoma: Clinical Frontiers and Perspectives. *Gut* (2014) 63(5):844–55. doi: 10.1136/gutjnl-2013-306627
22. Veerakumarasivam A, Goldstein LD, Saeb-Parsy K, Scott HE, Warren A, Thorne NP, et al. AURKA Overexpression Accompanies Dysregulation of DNA-Damage Response Genes in Invasive Urothelial Cell Carcinoma. *Cell Cycle* (2008) 7(22):3525–33. doi: 10.4161/cc.7.22.7042
23. Bertolin G, Tramier M. Insights Into the Non-Mitotic Functions of Aurora Kinase a: More Than Just Cell Division. *Cell Mol Life Sci* (2020) 77(6):1031–47. doi: 10.1007/s00018-019-03310-2
24. Do TV, Hirst J, Hyter S, Roby KF, Godwin AK. Aurora a Kinase Regulates Non-Homologous End-Joining and Poly(ADP-Ribose) Polymerase Function in Ovarian Carcinoma Cells. *Oncotarget* (2017) 8(31):50376–92. doi: 10.18632/oncotarget.18970
25. Maria AG, Borges KS, Peixoto Lira RC, Thomé CH, Berthon A, Drougat L, et al. Inhibition of Aurora Kinase a Activity Enhances the Antitumor Response of Beta-Catenin Blockade in Human Adrenocortical Cancer Cells. *Mol Cell Endocrinol* (2021) 528:111243. doi: 10.1016/j.mce.2021.111243
26. Jalalirad M, Haddad TC, Salisbury JL, Radisky D, Zhang M, Schroeder M, et al. Aurora-a Kinase Oncogenic Signaling Mediates TGF- β -Induced Triple-Negative Breast Cancer Plasticity and Chemoresistance. *Oncogene* (2021) 40(14):2509–23. doi: 10.1038/s41388-021-01711-x
27. Liu F, Zhang Y, Dong Y, Ning P, Zhang Y, Sun H, et al. Knockdown of AURKA Sensitizes the Efficacy of Radiation in Human Colorectal Cancer. *Life Sci* (2021) 271:119148. doi: 10.1016/j.lfs.2021.119148
28. Zhang H, Bao J, Zhao S, Huo Z, Li B. MicroRNA-490-3p Suppresses Hepatocellular Carcinoma Cell Proliferation and Migration by Targeting the Aurora Kinase a Gene (AURKA). *Arch Med Sci* (2020) 16(2):395–406. doi: 10.5114/aoms.2019.91351
29. Wu M, Li X, Liu R, Yuan H, Liu W, Liu Z. Development and Validation of a Metastasis-Related Gene Signature for Predicting the Overall Survival in Patients With Pancreatic Ductal Adenocarcinoma. *J Cancer* (2020) 11(21):6299–318. doi: 10.7150/jca.47629
30. Chen X, Zhang D, Jiang F, Shen Y, Li X, Hu X, et al. Prognostic Prediction Using a Stemness Index-Related Signature in a Cohort of Gastric Cancer. *Front Mol Biosci* (2020) 7:570702. doi: 10.3389/fmolb.2020.570702
31. Chen T, Yang S, Xu J, Lu W, Xie X. Transcriptome Sequencing Profiles of Cervical Cancer Tissues and SiHa Cells. *Funct Integr Genomics* (2020) 20(2):211–21. doi: 10.1007/s10142-019-00706-y
32. Jin D, Song Y, Chen Y, Zhang P. Identification of Three lncRNAs as Potential Predictive Biomarkers of Lung Adenocarcinoma. *BioMed Res Int* (2020) 2020:7573689. doi: 10.1155/2020/7573689
33. Zhao W, Wang M, Wang C, Liu Y, Liu H, Luo S. RACGAP1 is Transcriptionally Regulated by E2F3, and Its Depletion Leads to Mitotic Catastrophe in Esophageal Squamous Cell Carcinoma. *Ann Transl Med* (2020) 8(15):950. doi: 10.21037/atm-20-2901
34. Ren K, Zhou D, Wang M, Li E, Hou C, Su Y, et al. RACGAP1 Modulates ECT2-Dependent Mitochondrial Quality Control to Drive Breast Cancer Metastasis. *Exp Cell Res* (2021) 400(1):112493. doi: 10.1016/j.yexcr.2021.112493
35. Yang XM, Cao XY, He P, Li J, Feng MX, Zhang YL, et al. Overexpression of Rac Gtpase Activating Protein 1 Contributes to Proliferation of Cancer Cells by Reducing Hippo Signaling to Promote Cytokinesis. *Gastroenterology* (2018) 155(4):1233–49.e1222. doi: 10.1053/j.gastro.2018.07.010
36. Chen H, Wu J, Lu L, Hu Z, Li X, Huang L, et al. Identification of Hub Genes Associated With Immune Infiltration and Predict Prognosis in Hepatocellular Carcinoma Via Bioinformatics Approaches. *Front Genet* (2020) 11:575762. doi: 10.3389/fgene.2020.575762
37. Roessler S, Lin G, Forgues M, Budhu A, Hoover S, Simpson RM, et al. Integrative Genomic and Transcriptomic Characterization of Matched Primary and Metastatic Liver and Colorectal Carcinoma. *Int J Biol Sci* (2015) 11(1):88–98. doi: 10.7150/ijbs.10583
38. Lu M, Zhu WW, Wang X, Tang JJ, Zhang KL, Yu GY, et al. ACOT12-Dependent Alteration of Acetyl-Coa Drives Hepatocellular Carcinoma Metastasis by Epigenetic Induction of Epithelial-Mesenchymal Transition. *Cell Metab* (2019) 29(4):886–900.e885. doi: 10.1016/j.cmet.2018.12.019
39. Park HM, Kim H, Kim DW, Yoon JH, Kim BG, Cho JY. Common Plasma Protein Marker LCAT in Aggressive Human Breast Cancer and Canine Mammary Tumor. *BMB Rep* (2020) 53(12):664–9. doi: 10.5483/BMBRep.2020.53.12.238
40. Cooke AL, Morris J, Melchior JT, Street SE, Jerome WG, Huang R, et al. A Thumbwheel Mechanism for APOA1 Activation of LCAT Activity in HDL. *J Lipid Res* (2018) 59(7):1244–55. doi: 10.1194/jlr.M085332
41. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, et al. DNA Methylation-Driven Genes for Constructing Diagnostic, Prognostic, and Recurrence Models for Hepatocellular Carcinoma. *Theranostics* (2019) 9(24):7251–67. doi: 10.7150/thno.31155
42. Ouyang G, Yi B, Pan G, Chen X. A Robust Twelve-Gene Signature for Prognosis Prediction of Hepatocellular Carcinoma. *Cancer Cell Int* (2020) 20:207. doi: 10.1186/s12935-020-01294-9
43. Liu GM, Zeng HD, Zhang CY, Xu JW. Identification of a Six-Gene Signature Predicting Overall Survival for Hepatocellular Carcinoma. *Cancer Cell Int* (2019) 19:138. doi: 10.1186/s12935-019-0858-2
44. Zhang BH, Yang J, Jiang L, Lyu T, Kong LX, Tan YF, et al. Development and Validation of a 14-Gene Signature for Prognosis Prediction in Hepatocellular Carcinoma. *Genomics* (2020) 112(4):2763–71. doi: 10.1016/j.ygeno.2020.03.013
45. Li N, Zhao L, Guo C, Liu C, Liu Y. Identification of a Novel DNA Repair-Related Prognostic Signature Predicting Survival of Patients With Hepatocellular Carcinoma. *Cancer Manag Res* (2019) 11:7473–84. doi: 10.2147/CMAR.S204864
46. Zhu G, Xia H, Tang Q, Bi F. An Epithelial-Mesenchymal Transition-Related 5-Gene Signature Predicting the Prognosis of Hepatocellular Carcinoma Patients. *Cancer Cell Int* (2021) 21(1):166. doi: 10.1186/s12935-021-01864-5
47. Cai JL, Zhu GQ, Du JX, Wang B, Wan JL, Xiao K, et al. Identification and Validation of a New Gene Signature Predicting Prognosis of Hepatocellular Carcinoma Patients by Network Analysis of Stemness Indices. *Expert Rev Gastroenterol Hepatol* (2021) 1–11. doi: 10.1080/17474124.2021.1845142
48. Liu GM, Xie WX, Zhang CY, Xu JW. Identification of a Four-Gene Metabolic Signature Predicting Overall Survival for Hepatocellular Carcinoma. *J Cell Physiol* (2020) 235(2):1624–36. doi: 10.1002/jcp.29081

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Su, Zhang and Kong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.