Identification of a thirteen‐gene signature predicting overall survival for hepatocellular carcinoma

Xiaohan Zhou¹*, Chengdong Liu²*, Hanyi Zeng², Dehua Wu¹, Li Liu²

¹Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou 510515 Guangdong Province People’s Republic of China;

²Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou 510515 Guangdong Province People’s Republic of China.

*These authors contributed equally to this work.

Corresponding Authors: Li Liu, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou 510515 Guangdong Province People’s Republic of China; e-mail: liuli@i.smu.edu.cn; and Dehua Wu, Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou 510515 Guangdong Province People’s Republic of China; e-mail: 18602062748@163.com

Abstract

Background: Hepatocellular carcinoma (HCC) is a malignant tumor of the digestive system characterized by mortality rate and poor prognosis. To indicate the prognosis of HCC patients, lots of genes have been screened as prognostic indicators. However, the predictive efficiency of single gene is not enough. Therefore, it is essential to identify a risk‐score model based on gene signature to elevate predictive efficiency.
**Methods:** lasso regression analysis followed by univariate cox regression was employed to establish a risk-score model for HCC prognosis prediction based on The Cancer Genome Atlas (TCGA) dataset and Gene Expression Omnibus (GEO) dataset GSE14520. R package “clusterProfiler” was used to conduct function and pathway enrichment analysis. The infiltration level of various immune and stromal cells in the tumor microenvironment (TME) were evaluated by ssGSEA of R package “GSVA”.

**Results:** This prognostic model is an independent prognostic factor for predicting the prognosis of HCC patients and can be more effective combining with clinical data through the construction of nomogram model. Further analysis showed patients in high-risk group possess more complex TME and immune cell composition.

**Conclusions:** Taken together, our research suggests the thirteen-gene signature to possess potential prognostic value for HCC patients and provide new information for immunological research and treatment in HCC.

**Keywords:** HCC, Risk-score model, Gene signature, Tumor microenvironment, Infiltration of immune cells.

**Introduction**

Hepatocellular carcinoma (HCC) is a common malignant tumor of human
digestive system. Its early diagnosis and treatment are becoming a worldwide problem.[1]. Therefore, seeking prognostic markers is urgent for identifying high-risk patients of HCC patients, which may be helpful to implement individualized treatment.

Increasing studies have shown that gene signatures have an advantage in predicting the overall survival (OS). Long et al. constructed a four-gene signature based on CENPA, SPP1, MAGEB6, and HOXD9 expression to predict the OS of HCC [2]. Zheng et al. also established a four-gene signature based on SPINK1, TXNRD1, LCAT, and PZP to predict the OS of HCC patients in TCGA cohort [3]. However, little research has been done to study causes of different risk between patients, such as the composition of tumor microenvironment (TME).

In our study, we constructed a thirteen-gene signature-based risk-score model that included LDHA, GPC1, GRM8, PPAT, SLC29A3, EMCN, GDI2, CBX2, LILRA2, ADAMTS5, GSR, WEE1 and SLC1A5. The risk-score model was validated in GEO database. Furthermore, we built and validated a nomogram in the TCGA cohort to predict OS of HCC patients. Functional analysis of the differentially expressed genes (DEGs) between high- and low-risk groups was conducted. The infiltration level of various immune and stromal cells in the TME was evaluated. In general, the prognostic model might help effectively predict OS of HCC patients and evaluate the infiltration level of immune and stromal cells in HCC tissue.

Materials and methods
Data preparation

TCGA and GEO database were used to establish and validate risk score model. For TCGA cohort, 371 patients possess RNA seq data, in which 365 patients with overall survival data were selected for the construction of prognostic model. The TCGA mRNA expression was generated with HTseq-FPKM. Then, 365 patients were randomly grouped into 183 cases as internal training set and 182 cases as internal testing set. For external testing set, we obtained GEO HCC dataset (access no: GSE14520, platform: GPL571) which included 247 HCC samples. In which 221 patients possess available overall survival information for validation of prognostic model. The clinical baseline data of TCGA and GSE14520 were provided in supplementary Table 1.

Construction of the risk-score prognostic model

Lasso regression analysis followed by univariate cox regression was conducted to investigate and narrow OS related genes. A total of 436 genes that meet the standard of p value <0.001 by univariate cox regression analysis were selected for lasso-penalized regression. Lasso-penalized regression was conducted by R package “glmnet” with 10-fold cross-validation. According to the expression of each gene, we screened eligible genes and calculated corresponding coefficients and risk score for each patient. A thirteen-gene signature was established based on univariate and lasso regression analyses. The prognosis index (PI) = (0.094 * LDHA expression) + (0.076 * of GPC1 expression) + (0.087 * GRM8 expression) + (0.279 * PPAT expression) +
(0.189 * SLC29A3 expression) + (-0.147 * EMCN expression) + (0.016 * GDI2 expression) + (0.187 * CBX2 expression) + (0.637 * LILRA2 expression) + (0.368 * ADAMTS5 expression) + (0.056 * GSR expression) + (0.028 * WEE1 expression) + (0.008 * SLC1A5 expression). Patients of TCGA HCC cohort were grouped into high- and low-risk groups by the optimal cut-off value. In addition, Kaplan-Meier survival and receiver operating characteristic (ROC) analysis were conducted to estimate the predictive efficiency of the risk-score model.

**Construction and validation of a predictive nomogram**

To evaluate the probability of five-year OS, we constructed a nomogram in TCGA HCC cohort[4] based on multivariable cox regression analysis of risk-score model and clinical characteristic parameters. The calibration curve was used to show the nomogram prediction probabilities against the observed rates. We used DCA plot to calculate the clinical net benefit and net reduction. The best model possesses the highest net benefit and net reduction as calculated[5].

**Differentially expressed gene screening**

Median risk score was used to group the patients into the high- and low-risk group. The DEGs were selected by the R package “Limma”. DEGs with an absolute log2FC (fold change) > 0.5 and the p < 0.05 were selected for subsequent analysis.

**Enrichment Analyses**
Gene Ontology (GO) and Encyclopedia of Genes and Genomes (KEGG) analysis were conducted using R package “clusterProfiler” [6]. Gene Set Enrichment Analyses (GSEA) was conducted using gseGO, gseKEGG and gsePathway function in R package “clusterProfiler”, with the parameters: nPerm = 1000, minGSSize = 10, maxGSize = 1000, pvalue-Cutoff=0.05.

Generation of TME gene signatures and signature score computation

We used TME-related gene sets the scoring method curated by Mariathasan et al. [7]. Briefly, a principal component analysis (PCA) was performed, and principal component 1 was considered to serve as gene signature score.

Imune cell infiltration

The ssGSEA (single sample GSEA) was used to evaluate the infiltration level of 28 immune cells. The gene sets of various immune cells were obtained from a recent publication[8]. The infiltration score of each immune cell was calculated through ssGSEA analysis of “GSVA” R package in the R software. The ssGSEA score was normalized to uniform distribution, for which the ssGSEA score is distributed between 0 to 1.

Human tissue samples

The experiments involving human samples in our study were in accordance with the principles of the Declaration of Helsinki, and approved by the Institutional
Review Board of Nanfang Hospital, Southern Medical University of Guangdong, China (NFEC-201208-K3). A total of 10 liver cancer and paired non-cancerous tissues were collected and used for qRT-PCR.

**qRT-PCR**

TRIzol® reagent (TaKaRa, Tokyo, Japan), PrimeScript™ 1st Strand cDNA Synthesis Kit (TaKaRa, Tokyo, Japan), and SYBR ® Green PCR kit (TaKaRa, Tokyo, Japan) were used to perform the extraction of total RNA, synthesization of First-strand cDNA, and Real-time PCR respectively. Primers are provided in supplementary Table 2.

**Statistical analysis**

All data are presented as the mean ± standard deviation (SD). Student’s t-test (two-tailed) was used to analyze differences between two groups using R software (version: 3.6.2). p < 0.05 was considered statistically significant: *p < 0.05; **p < 0.01; ***p < 0.001 and ****p < 0.0001.

**Results**

**Risk-score model construction**

365 HCC patients in TCGA cohort were included. These patients were randomly separated into two groups, including internal training set (n = 183) and internal testing set (n = 182). Univariate cox regression analysis identified 436 OS related genes
(P<0.001) (Supplementary Table 3). Then, lasso analysis was conducted in the internal training set to further shrink the OS related genes. Thirteen genes (including LDHA, GPC1, GRM8, PPAT, SLC29A3, EMCN, GDI2, CBX2, LILRA2, ADAMTS5, GSR, WEE1 and SLC1A5) were screened and subsequently used to construct a prognostic risk-score model (Fig 1). We further calculated the risk score for each patient and seek the best optimal cut-off using the R package “Survminer”. Kaplan-Meier curve and ROC analysis were performed to estimate the efficiency of the thirteen-gene signature in the internal training cohort (Fig. 2A), internal testing cohort (Fig. 2B), and external testing cohort (Fig. 2C). Patients showed significantly poorer OS in the high-risk group compared to be in low-risk group (p < 0.001). The area under the ROC curve (AUCs) for 1-, 3-, and 5-year respectively for internal training cohort, internal testing cohort, and external testing cohort were 0.769, 0.746, 0.763; 0.757, 0.635, 0.601; 0.637, 0.666, 0.646.

**Correlation of risk score model and clinical features**

Forest map showed the univariate cox regression analysis results of these thirteen genes (Fig 3A). The Kaplan-Meier survival analysis based on TCGA cohort revealed that EMCN was a favorable factor for prognosis (sup-Fig 1A), while the other 12 genes were risk factors for prognosis (sup-Fig 1B-M), which was consistent with the result of univariate cox regression analysis. The mRNA expression level of each mRNA except EMCN, was higher in the high-risk group (sup-Fig 2A-M). Through correlation analysis, we found that the mRNA level of these 13 genes has a strong
correlation: EMCN was negative correlated with other 12 mRNAs, while these 12 mRNAs have positive correlation with each other (Fig 3B). In addition, we found that among the 13 genes in the prognosis model, 8 genes have been studied and proved to promote HCC progress, including CBX2[9], SLC1A5[10], LDHA [11], GSR[12], GPC1[13], GDI2 [14], ADAMTS5[15], and WEE1 [16]. While the other five genes have not been experimentally studied in HCC, including EMCN, LILRA2, PPAT, SLC29A3, and GRM8. We further verified the expression of these five genes using qRT-PCR. Our results showed that SLC29A3 and PPAT were high-expressed and EMCN was low-expressed in HCC tissues compared with adjacent normal tissues (sup-Fig 3A-C). There was no difference in LILRA2 and GRM8 expression between HCC tissues and adjacent normal tissues (sup-Fig 3D-W).

Moreover, we found the risk score was higher in the high stage of HCC both in TCGA and GSE14520 cohort (Fig 3C-D). Univariate and multivariate cox regression analyses were performed to evaluate the independent predictive ability of the risk-score model in TCGA HCC cohort. Results suggested the risk-score status, person neoplasm cancer status, and tumor stage to act as independent prognostic factors (Table 1).

**Construction and validation of predictive nomogram**

We further built a nomogram to predict OS according to the results of multivariate cox regression in TCGA HCC cohort (Fig 4A). The calibration curves for the 1-, 3-, and 5-year OS showed good agreement between the prediction from the nomogram and the actual observations (Fig 4B). The AUCs for 1-, 3-, and 5-year
respectively for risk-score model and nomogram model in TCGA cohort were 0.807, 0.736, 0.709 (Fig 4C) and 0.778, 0.766, 0.763 (Fig 4D). The clinical usefulness was assessed using DCA. The results showed the nomogram model to perform better to help identify high-risk patients for intervention and low-risk patients to avoid over treatment, especially for 3- and 5-year prediction (Fig 4E-J). In addition, the C-index of our nomogram model was 0.729, which was higher than that in other five published models (Fig 4K) [17].

**Functional enrichment analysis**

To explore the differences between high- and low-risk group, we divided the samples of TCGA HCC cohort into high- and low-risk group according to the median of risk score. The volcano plot and heatmap showed the DEGs between high- and low-risk group (Fig 5A-B). GSEA was conducted to search GO (Gene Ontology) (Fig 5C) and Reactome Pathways (Fig 5D). Cell cycle related terms were significantly enriched in high-risk group, such as G1/S transition of mitotic cell cycle, cell cycle, and S phase terms. To further explore the potential functional pathways, functional pathway analysis was conducted based on the DEGs using R packages “clusterProfiler”. Results revealed that these genes were mainly associated with carboxylic acid biosynthetic process, organic acid biosynthetic process and steroid metabolic process in BP (biological process) category. For CC (cellular component) category, the DEGs mainly enriched in collagen-containing extracellular matrix and endoplasmic reticulum lumen. In MF (molecular function) category, the DEGs mainly enriched in oxidoreductase activity, acting on CH−OH group of donors (Fig 5E).
Moreover, KEGG pathway analysis suggested the DEGs to be enriched in various crosstalk in malignancy, including Metabolism of xenobiotics by cytochrome P450, Retinol metabolism, Drug metabolism–cytochrome P450, Glycolysis/Gluconeogenesis and PPAR signaling pathway (Fig 5F).

**Tumor microenvironment analysis**

TME especially immune microenvironment is a key role for the survival of tumor patients. We hypothesized that high-risk patients have more complex TME. To our surprise, our analyses revealed that high-risk group was significantly associated with all indicated TME signatures (Fig 6A-B). To evaluate the infiltration level of immune cell, we employed ssGSEA to get the relative infiltration score of 28 immune cells (Fig 7A). We observed two major features. (1) The high-risk group had more complex immune cell components than the low-risk group, including cells executing anti-tumor reactivity (activated CD4 T cells, activated dendritic cells, central memory CD4 T cells, central memory CD8 T cells, Natural killer cell and Natural killer T cell) and cells delivering pro-tumor suppression (immature dendritic cells, MDSCs, Plasmacytoid dendritic cell, regulatory T cells, and Type 2 T helper cell) (Fig 7B). (2) Pearson’s correlation analysis revealed the infiltration level of cells executing anti-tumor reactivity and delivering pro-tumor suppression was positively correlated within a local environment (Fig 7C). This observation suggests a presence of a feedback mechanism such that the recruitment or differentiation of cells specialized for immune suppression may be facilitated by anti-tumor inflammation.
Discussion

HCC is one of the most malignant tumor types worldwide with its poor prognosis and lack of effective early detection methods[8,18,19]. Therefore, it is urgent to identify new prognostic markers and establish more accurate prognostic models. Compared with a single biomarker, the prognostic gene signatures and conventional clinical parameters may have better prediction effect. In recent years, gene markers based on mRNAs have been widely used in tumor prognosis prediction[20-23]. However, prognostic gene signatures that could be used to show therapeutic effects is still very few.

In our study, we built a thirteen-gene prognostic model that included LDHA, GPC1, GRM8, PPAT, SLC29A3, EMCN, GDI2, CBX2, LILRA2, ADAMTS5, GSR, WEE1 and SLC1A5. The prediction performance of this thirteen-gene prognostic model was eligible not only in internal training set and internal testing set from TCGA HCC cohort, but also in external testing set from GSE14520 HCC cohort. The thirteen-gene risk model was also proved to be an independent prognostic factor for HCC. Patients in high-risk group has significantly worse OS than patients in low-risk group.

Nomogram is a tool to provide individual patients with the overall survival probability [24,25]. In our study, we built a nomogram model using the thirteen-gene signature combined with clinical prognostic factors. The calibration plots showed the actual survival rate to be in good agreement with the predicted survival rate,
illuminating that the prediction performance of nomogram is eligible. We further proved the nomogram to be better than a single risk score factor by DCA plot. In addition, the C-index of our nomogram model was 0.729, which was higher than that in other five published models.

Since our thirteen-gene signature could effectively predict the OS in HCC patients, we further explore the differences between high- and low-risk samples. We divided the samples of TCGA HCC cohort into high- and low-risk group according to the median of risk score. GSEA results showed that cell cycle related pathways were significantly enriched in high-risk group. Cell cycle block or misoperation play an important role in the development of HCC[26-29]. Our results suggested the poor prognosis of high-risk group to be mainly due to the high level of cell cycle block or misoperation.

TME cells are an important part of tumor tissue. Increasing evidence has demonstrated their clinicopathological significance in predicting prognosis and efficacy [30,31]. TME signature based on characteristic gene expression is a new tool for evaluating comprehensive TME and a powerful biomarker for predicting survival and guiding more effective immunotherapy strategies [32]. Our results proved that high-risk group was significantly associated with many TME signatures, especially immune relevant signatures. By investigating the regulation of immune cell infiltration level, we observed two major features. (1) The high-risk group had more complex immune cell components than the low-risk group, including cells executing anti-tumor reactivity (activated CD4 T cells, activated dendritic cells, central memory
CD4 T cells, central memory CD8 T cells, Natural killer cell and Natural killer T cell) and cells delivering pro-tumor suppression (immature dendritic cells, MDSCs, Plasmacytoid dendritic cell, regulatory T cells, and Type 2 T helper cell). (2) Pearson's correlation analysis revealed the infiltration level of these two categories of cells to be positively correlated within a local environment. Based on the results of TME analysis, we believed that it is not enough to use only one or a class of TME components to predict the prognosis, but also depends on the overall situation of TME. The more complex the TME, the higher the risk of tumor and the worse the prognosis. Our conclusion is consistent with that of the previously reported article[33].

Overall, our research indicated that the thirteen-gene signature prognostic model is a trustworthy tool for the prediction of OS in HCC. Further construction of nomogram with this thirteen-gene signature prognostic model could assist clinicians to choose personalized treatment for patients with HCC.

Acknowledgements

This work was supported by the National Nature Science Foundation of China [grant numbers 81773008, 81672756, 81872399, 81972897], the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2015), the Natural Science Foundation of Guangdong Province [grant number 2017A030311023], the Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program:2017BT01S131 and the Guangzhou Technology Project [grant number 201804010044].
Conflict of interest

The authors declare no conflicts of interest.

Author contributions

L.L., D.W., X.Z. and C.L. designed the study. X.Z. and C.L. performed data analysis.

X.Z. and H.Z. wrote the manuscript and helped with validation.

Data Availability Statement

All data, models, or code used in this study are available from the corresponding author by request.

References

1 Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin, 10.3322/caac.21492

2 Long, J., Zhang, L., Wan, X., Lin, J., Bai, Y. and Xu, W. et al. (2018) A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. J. Cell. Mol. Med., 10.1111/jcmm.13863

3 Zheng, Y., Liu, Y., Zhao, S., Zheng, Z., Shen, C. and An, L. et al. (2018) Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma. Cancer Manag Res 10, 6079-6096, 10.2147/CMAR.S181396

4 Iasonos, A., Schrag, D., Raj, G.V. and Panageas, K.S. (2008) How to build and interpret a nomogram for cancer prognosis. J. Clin. Oncol. 26, 1364-1370, 10.1200/JCO.2007.12.9791

5 Vickers, A.J. and Elkin, E.B. (2006) Decision curve analysis: a novel method for evaluating prediction models. Med. Decis. Making 26, 565-574, 10.1177/0272989X06295361

6 Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16, 284-287, 10.1089/omi.2011.0118

7 Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K. and Wang, Y. et al. (2018) TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature, 10.1038/nature25501
8 Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M. and Rieder, D. et al. (2017) Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 18, 248-262, 10.1016/j.celrep.2016.12.019
9 Mao, J., Tian, Y., Wang, C., Jiang, K., Li, R. and Yao, Y. et al. (2019) CBX2 Regulates Proliferation and Apoptosis via the Phosphorylation of YAP in Hepatocellular Carcinoma. J. Cancer 10, 2706-2719, 10.7150/jca.31845
10 Zhang, P., Wang, Q., Lin, Z., Yang, P., Dou, K. and Zhang, R. (2019) Berberine Inhibits Growth of Liver Cancer Cells by Suppressing Glutamine Uptake. Onco Targets Ther 12, 11751-11763, 10.1016/j.celrep.2016.12.019
11 Guo, Y., Li, X., Sun, X., Wang, J., Yang, X. and Zhu, X. et al. (2019) Combined Aberrant Expression of NDRG2 and LDHA Predicts Hepatocellular Carcinoma Prognosis and Mediates the Anti-tumor Effect of Gemcitabine. Int. J. Biol. Sci. 15, 1771-1786, 10.1016/j/ijsb.35094
12 McLoughlin, M.R., Orlicky, D.J., Prigge, J.R., Krishna, P., Talago, E.A. and Caviglio, I.R. et al. (2019) TrxR1, Gsr, and oxidative stress determine hepatocellular carcinoma malignancy. Proc Natl Acad Sci U S A 116, 11408-11417, 10.1073/pnas.1903244116
13 Chen, G., Wu, H., Zhang, L. and Wei, S. (2020) High glypican-1 expression is a prognostic factor for predicting a poor clinical prognosis in patients with hepatocellular carcinoma. Oncol. Lett. 20, 197, 10.3892/ol.2020.12058
14 Bayo, J., Fiore, E., Aquino, J.B., Malvicini, M., Rizzo, M. and Peixoto, E. et al. (2014) Increased migration of human mesenchymal stromal cells by autocrine motility factor (AMF) resulted in enhanced recruitment towards hepatocellular carcinoma. PLoS One 9, e95171, 10.1371/journal.pone.0095171
15 Zhu, Z., Xu, J., Wu, X., Lin, S., Li, L. and Ye, W. et al. (2021) In Silico Identification of Contradictory Role of ADAMTS5 in Hepatocellular Carcinoma. Technol Cancer Res Treat 20, 1079254474, 10.1177/1533033820986826
16 Jiang, S.P. and Li, Z.R. (2019) MiR-503-5p regulates cell epithelial-to-mesenchymal transition, metastasis and prognosis of hepatocellular carcinoma through inhibiting WEE1. Eur Rev Med Pharmacol Sci 23, 2028-2037, 10.26355/eurrev_201903_17242
17 Zhao, K., Xu, L., Li, F., Ao, J., Jiang, G. and Shi, R. et al. (2020) Identification of hepatocellular carcinoma prognostic markers based on 10-immune gene signature. Biosci Rep 40, 10.1042/BSR20200894
18 Bruix, J., Gores, G.J. and Mazzaferro, V. (2014) Hepatocellular carcinoma: clinical frontiers and perspectives. Gut 63, 844-855, 10.1136/gutjnl-2013-306627
19 Liu, C., Zhou, X., Long, Q., Zeng, H., Sun, Q. and Chen, Y. et al. (2020) Small extracellular vesicles containing miR-30a-3p attenuate the migration and invasion of hepatocellular carcinoma by targeting SNAP23 gene. Oncogene, 10.1038/s41388-020-01521-7
20 Wang, Z., Teng, D., Li, Y., Hu, Z., Liu, L. and Zheng, H. (2018) A six-gene-based prognostic signature for hepatocellular carcinoma overall survival prediction. Life Sci., 10.1016/j.lfs.2018.04.025
21 Ke, K., Chen, G., Cai, Z., Huang, Y., Zhao, B. and Wang, Y. et al. (2018) Evaluation and prediction of hepatocellular carcinoma prognosis based on molecular classification. Cancer Manag Res 10, 5291-5302, 10.2147/CMAR.S178579
22 Liu, S., Miao, C., Liu, J., Wang, C.C. and Lu, X.J. (2017) Four differentially methylated gene pairs to predict the prognosis for early stage hepatocellular carcinoma patients. J. Cell. Physiol., 10.1002/jcp.26256
23 Liu, G.M., Zeng, H.D., Zhang, C.Y. and Xu, J.W. (2019) Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. Cancer Cell Int. 19, 138, 10.1186/s12935-019-0858-2

24 Kim, S.Y., Yoon, M.J., Park, Y.I., Kim, M.J., Nam, B.H. and Park, S.R. (2017) Nomograms predicting survival of patients with unresectable or metastatic gastric cancer who receive combination cytotoxic chemotherapy as first-line treatment. Gastric Cancer, 10.1007/s10120-017-0756-z

25 Borque, Á., Rubio-Briones, J., Esteban, L.M., Sanz, G., Domínguez-Escrig, J. and Ramírez-Backhaus, M. et al. (2014) Implementing the use of nomograms by choosing threshold points in predictive models: 2012 updated Partin Tables vs a European predictive nomogram for organ-confined disease in prostate cancer. BJU Int. 113, 878-886, 10.1111/bju.12532

26 Liu, Z., Li, J., Chen, J., Shan, Q., Dai, H. and Xie, H. et al. (2018) MCM family in HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2 cell cycle progression. BMC Cancer 18, 200, 10.1186/s12885-018-4056-8

27 Quetglas, I.M., Moeini, A., Pinyol, R. and Llovet, J.M. (2014) Integration of genomic information in the clinical management of HCC. Best Pract Res Clin Gastroenterol 28, 831-842, 10.1016/j.bpg.2014.08.004

28 Hua, S., Liu, C., Liu, L. and Wu, D. (2018) miR-142-3p inhibits aerobic glycolysis and cell proliferation in hepatocellular carcinoma via targeting LDHA. Biochem. Biophys. Res. Commun., 10.1016/j.brc.2018.01.112

29 Greenbaum, L.E. (2004) Cell cycle regulation and hepatocarcinogenesis. Cancer Biol. Ther. 3, 1200-1207

30 Greten, F.R. and Grivennikov, S.I. (2019) Inflammation and Cancer: Triggers, Mechanisms, and Consequences. Immunity 51, 27-41, 10.1016/j.immuni.2019.06.025

31 Vitale, I., Manic, G., Coussens, L.M., Kroemer, G. and Galluzzi, L. (2019) Macrophages and Metabolism in the Tumor Microenvironment. Cell Metab. 30, 36-50, 10.1016/j.cmet.2019.06.001

32 Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H. and Shi, M. et al. (2019) Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. Cancer Immunol Res, 10.1158/2326-6066.CIR-18-0436

33 Jia, Q., Wu, W., Wang, Y., Alexander, P.B., Sun, C. and Gong, Z. et al. (2018) Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. Nat. Commun. 9, 5361, 10.1038/s41467-018-07767-w

Figure legends

**Figure 1:** The work flow showing the scheme of our study.

**Figure 2:** Construction of risk-score predictive model. A, Kaplan–Meier curve, heatmap of normalized mRNA expression, scatterplot of risk score and patient survival status, and ROC curve of the risk-score model in the internal training set of TCGA cohort. B, Kaplan–Meier curve,
heatmap of normalized mRNA expression, scatterplot of risk score and patient survival status, and ROC curve of the risk-score model in the internal testing set of TCGA cohort. C, Kaplan–Meier curve, heatmap of normalized mRNA expression, scatterplot of risk score and patient survival status, and ROC curve of the risk-score model in the external testing set of GSE14520 cohort.

**Figure 3:** Correlation of risk-score model and WHO stages. A, The hazard ratios (HR) with 95% confidence interval (95%CI) and p value of univariate cox regression were showed by forest plot. B, Correlation between thirteen genes, red and green curve respectively represents positive and negative correlation. C, Risk score of samples in indicated stages in the TCGA cohort. D, Risk score of samples in indicated stages in the GSE14520 cohort.

**Figure 4:** Construction of Nomogram to predict OS of HCC patients. A, Nomogram for predicting probabilities of HCC patients with 1-, 3-, and 5-year OS of HCC patients. B, Calibration plot for predicting 1-, 3-, and 5-year OS of HCC patient. Nomogram-based probability of survival is plotted on the x-axis; actual survival is plotted on the y-axis. C-D, The ROC curve of the risk-score model (C) and nomogram model (D) for 1-, 3-, and 5-year OS of HCC patient. E-J, Net benefit (E-G) and net reduction (H-J) (Y - axis) as calculated are plotted against the threshold probabilities of patients having 1-, 3-, and 5-year survival on the X-axis. The green line represents the assumption that all patients have indicated survival time. The black solid line represents the assumption that no patients have indicated survival time.

**Figure 5:** Function and pathway enrichment analysis of DEGs. A, The volcano plot of DEGs
between high- and low-risk groups. B, The heatmap showed the top 50 overexpressed genes in high-risk group and top 50 overexpressed genes in low-risk group. C-D, Significant GSEA results of DEGs, including GO (C), and Reactome Pathways (D). E, Significantly GO terms of DEGs, including molecular function (MF), cell component (CC), and biological processes (BP). F, Significantly KEGG pathways of DEGs.

**Figure 6:** TME analysis of high- and low-risk groups. A, Heatmap of normalized (Z-score) score of TME signature. B, The TME signature score of samples in high- and low-risk groups.

**Figure 7:** ssGSEA analysis of high- and low-risk groups. A, Heatmap of normalized (Z-score) score of 28 infiltrating immune cell populations. B, The signature score of 28 infiltrating immune cells in high and low risk groups. C, Correlation between infiltration of cell types executing anti-tumor immunity and cell types executing pro-tumor immune suppressive functions. R coefficient of pearson’s correlation. The shaded area represents 95% confident interval. Red points represent high-risk patients, green points represent low-risk patients.
Univariate Cox Regression

TCGA HCC patients n=371
Screening patients
365 patients with OS data
Random
Internal training set 183 patients
Internal testing set 182 patients
436 genes correlated with OS (p<0.001)
13-gene signature
Kaplan-Meier analysis
ROC analysis
Validation in Internal and External testing set
Univariate and Multivariate Cox Regression
Independence of Prognostic Model Building and Validating a predictive nomogram
High risk vs Low risk
Enrichment analysis
Tumor microenvironment analysis
Immune cell infiltration analysis

GSE14520 HCC patients n=247
Screening patients
221 patients with OS data
External testing set 221 patients
Sup-Figure 1: A-M, Kaplan-Meier analysis of OS in line with the expression of thirteen genes in TCGA cohort.
Sup-Figure 2: The expression of 13 genes in TCGA cohort.

Sup-Figure 2: A-M, The expression value of thirteen genes in high-risk, low-risk, and normal sample groups in TCGA cohort.
Sup-Figure 3: The expression of 5 genes in HCC patients.

Sup-Figure 3: A-W, The relative mRNA expression of indicated genes in HCC tumor and adjacent normal tissues.
### Supplementary Table 1: Baseline data

| Overall status | GSE14520 (221 with RNA seq) | TCGA (371 with RNA seq) |
|----------------|----------------------------|-------------------------|
| Dead           | 85                         | 130                     |
| Alive          | 136                        | 241                     |
| WHO stages     |                            |                         |
| Stage I        | 93                         | 171                     |
| Stage II       | 77                         | 86                      |
| Stage III      | 49                         | 85                      |
| Stage IV       | 0                          | 5                       |
| Unknown        | 2                          | 24                      |
| Age            |                            |                         |
| >60            | 43                         | 201                     |
| <60            | 178                        | 169                     |
| Unknown        | 0                          | 1                       |
| Gender         |                            |                         |
| Male           | 191                        | 250                     |
| Female         | 30                         | 121                     |

**Supplementary Table 1:** Baseline data of patients from GSE14520 and TCGA

### Supplementary Table 2: Primer sequences of 5 genes

| Gene names               | Primer sequences                  |
|--------------------------|-----------------------------------|
| ACTB Forward Primer      | CATGTACGTTGCTATCCAGGC             |
| ACTB Reverse Primer      | CTCCCTAATGTTCACGCACCAT            |
| EMCN Forward Primer      | AGCAAACCAGCCCGTCTTTATTC           |
| EMCN Reverse Primer      | AGCACATTCGTTACAAACCCA             |
| LILRA2 Forward Primer    | AGCCCCAGGAAGAAGACGTG              |
| LILRA2 Reverse Primer    | GCTGAGTGAGCTGTAGCATCT             |
| PPAT Forward Primer      | GATGGGAGTGCTGCTGCTG              |
| PPAT Reverse Primer      | CAACGAAGGGGCTGACAATTTC           |
| SLC29A3 Forward Primer   | TTGAGAGCTGAGCTGCGGTTG             |
| SLC29A3 Reverse Primer   | CAGTGTCAGTATCACCCATGAAA           |
| GRM8 Forward Primer      | CGAGGGAAAGCGATGCAAGCC            |
| GRM8 Reverse Primer      | CCCATCCACCGGATGGA                |

**Supplementary Table 2:** The primer sequences of indicated genes.
### Supplementary Table 3: Univariate cox results

| gene name  | Hazard_Ratio | CI95    | P_Value     |
|------------|--------------|---------|-------------|
| LDHA       | 1.91         | 1.48 - 2.47 | 0           |
| GPC1       | 1.34         | 1.14 - 1.56 | 0.00027     |
| GRM8       | 1.82         | 1.29 - 2.56 | 0.00059     |
| PPAT       | 2.13         | 1.49 - 3.06 | 0.00E-05    |
| SLC29A3    | 1.84         | 1.38 - 2.46 | 0.00E-05    |
| EMCN       | 0.58         | 0.43 - 0.77 | 0.00E-04    |
| GDI2       | 2.32         | 1.53 - 3.53 | 8.00E-05    |
| CBX2       | 1.81         | 1.45 - 2.26 | 0           |
| LILRA2     | 3.45         | 1.75 - 6.79 | 0.00034     |
| ADAMTS5    | 2.38         | 1.66 - 3.4  | 0           |
| GSR        | 1.47         | 1.2 - 1.8   | 0.00019     |
| WEE1       | 1.48         | 1.18 - 1.85 | 0.00067     |
| SLC1A5     | 1.37         | 1.22 - 1.54 | 0           |
| AAMP       | 1.91         | 1.3 - 2.8   | 0.00092     |
| ABCC5      | 1.93         | 1.35 - 2.75 | 0.00028     |
| ACYP1      | 1.87         | 1.34 - 2.59 | 0.00E-04    |
| ADM        | 1.31         | 1.13 - 1.52 | 0.00044     |
| ADSL       | 1.65         | 1.23 - 2.22 | 0.00E-04    |
| AMD1       | 2.02         | 1.51 - 2.71 | 0           |
| ANP32B     | 1.62         | 1.24 - 2.12 | 0.00042     |
| ANXA5      | 1.36         | 1.15 - 1.6  | 0.00032     |
| ARL8B      | 1.81         | 1.28 - 2.57 | 0.00085     |
| ARMC9      | 2.19         | 1.39 - 3.44 | 0.00067     |
| ASF1A      | 1.64         | 1.22 - 2.19 | 0.00095     |
| ATIC       | 1.55         | 1.2 - 2.01  | 0.00E-04    |
| ATP13A2    | 1.56         | 1.23 - 1.97 | 0.00024     |
| ATP1B3     | 1.33         | 1.14 - 1.56 | 0.00E-04    |
| ATP6V0A4   | 93.65        | 63.6 - 1379.86 | 0.00094   |
| B4GALT5    | 1.48         | 1.22 - 1.81 | 0.00E-05    |
| BARD1      | 1.89         | 1.34 - 2.66 | 0.00028     |
| BBS7       | 2.34         | 1.54 - 3.55 | 0.00E-05    |
| BCAT1      | 1.52         | 1.21 - 1.92 | 0.00038     |
| BCL10      | 1.8          | 1.31 - 2.46 | 0.00026     |
| BCORL1     | 1.62         | 1.24 - 2.13 | 0.00047     |
| BDH1       | 0.76         | 0.66 - 0.88 | 0.00014     |
| BFS1P1     | 2.93         | 1.76 - 4.86 | 0.00E-05    |
| BICD1      | 2.73         | 1.61 - 4.63 | 0.00E-04    |
| BIRC5      | 1.25         | 1.1 - 1.42  | 0.00063     |
| BRPF1      | 2.02         | 1.36 - 2.99 | 0.00045     |
| BUB1       | 1.57         | 1.26 - 1.96 | 0.00E-05    |
| Gene      | Value | 95% CI   | P-value |
|-----------|-------|----------|---------|
| BUB1B     | 1.49  | 1.22 - 1.83 | 0.00011 |
| C11orf16  | 3296.14 | 77.63 - 139961.77 | 2.00E-05 |
| C19orf26  | 12.35  | 5.5 - 27.71 | 0       |
| C2orf43   | 2.26   | 1.51 - 3.38 | 7.00E-05 |
| C2orf44   | 2.07   | 1.36 - 3.13 | 0.00065 |
| C3orf52   | 2.19   | 1.46 - 3.31 | 0.00018 |
| C5orf30   | 1.56   | 1.27 - 1.91 | 3.00E-05 |
| CACNA1B   | 4.74   | 2.09 - 10.78 | 2.00E-04 |
| CALCR     | 2.12   | 1.38 - 3.28 | 0.00068 |
| CAP1      | 1.78   | 1.31 - 2.43 | 0.00026 |
| CARD9     | 2.44   | 1.47 - 4.03 | 0.00052 |
| CASP2     | 1.64   | 1.22 - 2.19 | 0.00092 |
| CCNA2     | 1.27   | 1.11 - 1.46 | 0.00053 |
| CCNB1     | 1.29   | 1.12 - 1.49 | 0.00039 |
| CCNF      | 1.63   | 1.26 - 2.11 | 0.00022 |
| CCNJL     | 2.38   | 1.75 - 3.22 | 0       |
| CCT2      | 1.73   | 1.3 - 2.31  | 0.00016 |
| CCT4      | 1.78   | 1.31 - 2.43 | 0.00026 |
| CCT5      | 1.72   | 1.34 - 2.2  | 2.00E-05 |
| CCT7      | 1.71   | 1.24 - 2.36 | 0.00099 |
| CD3EAP    | 1.76   | 1.27 - 2.45 | 0.00077 |
| CD68      | 1.55   | 1.2 - 2     | 0.00078 |
| CD97      | 1.35   | 1.15 - 1.59 | 0.00025 |
| CDC20     | 1.27   | 1.13 - 1.43 | 5.00E-05 |
| CDC25A    | 1.58   | 1.24 - 2.01 | 0.00025 |
| CDC25B    | 1.38   | 1.17 - 1.64 | 0.00014 |
| CDC42EP2  | 1.51   | 1.2 - 1.9   | 4.00E-04 |
| CDC7      | 1.61   | 1.26 - 2.07 | 0.00016 |
| CDC8      | 1.53   | 1.23 - 1.92 | 0.00018 |
| CDCA4     | 1.55   | 1.23 - 1.96 | 0.00024 |
| CDCA8     | 1.49   | 1.26 - 1.76 | 0       |
| CDH10     | 1.69   | 1.24 - 2.3  | 8.00E-04 |
| CDK4      | 1.44   | 1.18 - 1.77 | 0.00038 |
| CDR1      | 526.59 | 16.2 - 17121.84 | 0.00042 |
| CDV3      | 1.84   | 1.32 - 2.55 | 3.00E-04 |
| CENPA     | 1.59   | 1.31 - 1.92 | 0       |
| CENPE     | 2.09   | 1.53 - 2.86 | 0       |
| CENPI     | 2.01   | 1.44 - 2.82 | 4.00E-05 |
| CENPO     | 1.78   | 1.33 - 2.37 | 9.00E-05 |
| CEP55     | 1.64   | 1.36 - 1.99 | 0       |
| CFHR4     | 0.8    | 0.72 - 0.89 | 5.00E-05 |
| CHEK1     | 1.79   | 1.35 - 2.38 | 5.00E-05 |
| CHODL     | 1.54   | 1.24 - 1.91 | 1.00E-04 |
| Gene    | Value | Lower Limit | Upper Limit | p-value |
|---------|-------|-------------|-------------|---------|
| CHORDC1| 2.26  | 1.58 - 3.24 | 1.00E-05    |
| CIT    | 2.16  | 1.51 - 3.08 | 2.00E-05    |
| CKAP5  | 1.68  | 1.27 - 2.23 | 0.00029     |
| CLCF1  | 1.41  | 1.15 - 1.71 | 0.00073     |
| CLCN2  | 1.69  | 1.27 - 2.25 | 0.00037     |
| CLDN6  | 1.71  | 1.29 - 2.28 | 0.00022     |
| CLEC3B | 0.73  | 0.62 - 0.86 | 0.0012      |
| CLSPN  | 2.17  | 1.52 - 3.09 | 2.00E-05    |
| CNOT6  | 1.8   | 1.33 - 2.44 | 0.00016     |
| CORO1C | 1.59  | 1.22 - 2.07 | 0.00054     |
| CPB2   | 0.85  | 0.77 - 0.94 | 0.00086     |
| CPSF6  | 1.75  | 1.27 - 2.4  | 0.00061     |
| CRH    | 2.06  | 1.39 - 3.06 | 3.00E-04    |
| CRISP2 | 1.99  | 1.36 - 2.92 | 0.00043     |
| CSF1   | 1.39  | 1.16 - 1.66 | 0.00027     |
| CSF3R  | 1.52  | 1.21 - 1.9  | 0.00026     |
| CSH1   | 1.09E+25 | 1207625079509.6 - 9.88344118950835e+36 | 4.00E-05 |
| CSN1S1 | 391.95 | 34.32 - 4476.56 | 0         |
| CSTF2  | 1.63  | 1.26 - 2.11 | 0.00017     |
| CYP26B1| 1.55  | 1.29 - 1.88 | 0          |
| CYP2C9 | 0.88  | 0.82 - 0.95 | 0.00051     |
| DAB2   | 1.4   | 1.19 - 1.64 | 3.00E-05    |
| DBF4   | 2.12  | 1.51 - 2.99 | 2.00E-05    |
| DDN    | 8.81  | 2.91 - 26.7 | 0.00012     |
| DFNA5  | 1.65  | 1.28 - 2.14 | 0.00014     |
| DNAJB4 | 1.49  | 1.18 - 1.89 | 0.00081     |
| DPH2   | 1.61  | 1.22 - 2.13 | 0.00069     |
| DRI    | 1.81  | 1.32 - 2.49 | 0.00024     |
| DSG2   | 1.26  | 1.1 - 1.43  | 6.00E-04    |
| DYNC1H1| 1.64  | 1.24 - 2.19 | 0.00065     |
| DYNC1LI1| 2.6  | 1.79 - 3.77 | 0          |
| E2F2   | 1.94  | 1.36 - 2.76 | 0.00024     |
| E2F5   | 1.81  | 1.34 - 2.45 | 0.00011     |
| E2F6   | 2.27  | 1.53 - 3.36 | 5.00E-05    |
| ECM2   | 0.73  | 0.61 - 0.88 | 0.00077     |
| ECT2   | 1.41  | 1.18 - 1.68 | 0.00017     |
| EGLN3  | 1.32  | 1.13 - 1.55 | 0.00036     |
| EIF1B  | 1.98  | 1.35 - 2.92 | 0.00052     |
| EIF2B5 | 2.2   | 1.45 - 3.35 | 0.00022     |
| EIF2S1 | 2.1   | 1.37 - 3.21 | 0.00062     |
| EIF5B  | 1.86  | 1.35 - 2.57 | 0.00015     |
| Gene  | Fold Change | 95% Confidence Interval | p-value |
|-------|-------------|-------------------------|---------|
| EMR3  | 3.01        | 1.67 - 5.43             | 0.00026 |
| ENO1  | 1.52        | 1.27 - 1.82             | 0       |
| EPO   | 1.29        | 1.16 - 1.44             | 1.00E-05|
| ESPL1 | 1.51        | 1.18 - 1.92             | 0.00087 |
| ETF1  | 2.12        | 1.43 - 3.14             | 0.00118 |
| ETV5  | 1.58        | 1.28 - 1.95             | 2.00E-05|
| EXO1  | 1.57        | 1.24 - 2               | 0.00021 |
| EZH2  | 1.58        | 1.27 - 1.96             | 5.00E-05|
| FABP6 | 1.46        | 1.2 - 1.78              | 0.00019 |
| FAM90A1 | 2.72      | 1.54 - 4.81             | 6.00E-04|
| FANCE | 1.6         | 1.25 - 2.06             | 0.00025 |
| FANCL | 2.02        | 1.43 - 2.85             | 7.00E-05|
| FBXO5 | 2.03        | 1.52 - 2.7              | 0       |
| FGF9  | 11.2        | 3.55 - 35.38            | 4.00E-05|
| FHOD3 | 1.67        | 1.29 - 2.15             | 8.00E-05|
| FOXI1 | 6.47        | 2.5 - 16.78             | 0.00012 |
| FOXJ3 | 1.85        | 1.32 - 2.6              | 0.00033 |
| FOXK2 | 1.81        | 1.32 - 2.49             | 0.00024 |
| FOXM1 | 1.31        | 1.12 - 1.52             | 0.00054 |
| FTCD  | 0.86        | 0.79 - 0.94             | 4.00E-04|
| FUBP1 | 1.6         | 1.23 - 2.07             | 0.00037 |
| FZD3  | 4.01        | 1.85 - 8.69             | 0.00043 |
| FZD7  | 1.5         | 1.24 - 1.82             | 4.00E-05|
| G6PD  | 1.33        | 1.19 - 1.5              | 0       |
| GAD1  | 2.23        | 1.42 - 3.51             | 0.00049 |
| GALNS | 1.67        | 1.26 - 2.21             | 0.00038 |
| GINS1 | 1.47        | 1.22 - 1.76             | 5.00E-05|
| GLMN  | 2.17        | 1.46 - 3.22             | 0.00013 |
| GLP1R | 1.7         | 1.26 - 2.29             | 0.00055 |
| GMEB1 | 2.22        | 1.4 - 3.53              | 0.00069 |
| GMPS  | 1.73        | 1.28 - 2.33             | 0.00034 |
| GNAI3 | 2.11        | 1.38 - 3.25             | 0.00064 |
| GNG5  | 1.71        | 1.28 - 2.29             | 0.00032 |
| GLN2  | 2.05        | 1.45 - 2.89             | 5.00E-05|
| GNPDA1 | 1.56       | 1.25 - 1.95             | 1.00E-04|
| GOT2  | 0.68        | 0.55 - 0.84             | 0.00049 |
| GPD2  | 1.77        | 1.32 - 2.37             | 0.00015 |
| GPRC5D | 3.54        | 2.07 - 6.08             | 0       |
| GPSM2 | 2.06        | 1.49 - 2.86             | 2.00E-05|
| GRID2 | 15047.13    | 173.48 - 1305138.65     | 2.00E-05|
| GRM4  | 18.63       | 4.23 - 82.01            | 0.00011 |
| GTDC1 | 3.33        | 1.85 - 5.99             | 6.00E-05|
| GTF2H1| 2.16        | 1.45 - 3.23             | 0.00017 |
| Gene   | Fold Change | 95% CI       | P-Value   |
|--------|-------------|--------------|-----------|
| GTF3C2 | 1.77        | 1.29 - 2.44  | 0.00044   |
| GTPBP4 | 1.89        | 1.42 - 2.51  | 1.00E-05  |
| GTSE1  | 1.6         | 1.29 - 1.97  | 1.00E-05  |
| H2AFY  | 1.7         | 1.26 - 2.3   | 6.00E-04  |
| HAND1  | 17.05       | 3.91 - 74.31 | 0.00016   |
| HAT1   | 1.82        | 1.28 - 2.6   | 0.00085   |
| HAVCR1 | 1.5         | 1.28 - 1.75  | 0         |
| HDAC1  | 1.67        | 1.27 - 2.2   | 0.00024   |
| HDAC2  | 2.1         | 1.54 - 2.87  | 0         |
| HECTD3 | 1.73        | 1.25 - 2.37  | 0.00079   |
| HMMR   | 1.47        | 1.22 - 1.77  | 5.00E-05  |
| HN1    | 1.34        | 1.14 - 1.57  | 0.00036   |
| HOXC5  | 104.46      | 11.41 - 956.53 | 4.00E-05 |
| HOXD10 | 1.95        | 1.38 - 2.76  | 0.0015    |
| HOXD13 | 2.26        | 1.5 - 3.4    | 1.00E-04  |
| HRASLS | 2.47        | 1.46 - 4.17  | 0.00077   |
| HSF2   | 1.7         | 1.26 - 2.3   | 0.00056   |
| HSPA14 | 1.88        | 1.32 - 2.69  | 0.00047   |
| HTR5A  | 32.24       | 6.31 - 164.64 | 3.00E-05 |
| HTRA2  | 2.1         | 1.4 - 3.15   | 0.00036   |
| ICMT   | 1.67        | 1.24 - 2.25  | 0.00084   |
| IGFBP3 | 1.23        | 1.09 - 1.39  | 0.00082   |
| IGSF3  | 1.35        | 1.15 - 1.57  | 0.00018   |
| IL12A  | 4.08        | 1.93 - 8.61  | 0.00023   |
| IL15RA | 1.55        | 1.28 - 1.89  | 1.00E-05  |
| IMPDH1 | 1.4         | 1.2 - 1.64   | 2.00E-05  |
| IQCC   | 1.99        | 1.36 - 2.92  | 0.00044   |
| KATNA1 | 1.91        | 1.34 - 2.74  | 0.00041   |
| KIF11  | 1.47        | 1.2 - 1.81   | 0.00022   |
| KIF14  | 1.66        | 1.23 - 2.23  | 0.00092   |
| KIF18A | 2.07        | 1.56 - 2.75  | 0         |
| KIF20A | 1.52        | 1.28 - 1.82  | 0         |
| KIF23  | 1.59        | 1.27 - 1.98  | 5.00E-05  |
| KIF2A  | 1.69        | 1.24 - 2.28  | 0.00076   |
| KIF2C  | 1.46        | 1.25 - 1.72  | 0         |
| KIF4A  | 1.36        | 1.16 - 1.61  | 0.00022   |
| KIFC1  | 1.29        | 1.12 - 1.5   | 0.00055   |
| KLHL18 | 2.7         | 1.53 - 4.77  | 0.00062   |
| KRT17  | 1.29        | 1.12 - 1.49  | 0.00032   |
| KRT32  | 14892196    | 1550.71 - 143016857397.09 | 0.00042 |
| KRT34  | 5.49        | 2.17 - 13.87 | 0.00031   |
| KRT38  | 159370.5    | 404.22 - 62834351.15 | 9.00E-05 |
| Gene      | Median Log2 Fold Change | 95% CI | p-value |
|-----------|------------------------|--------|---------|
| LAPTMB4B  | 1.25                   | 1.09 - 1.42 | 0.00096 |
| LHFPL2    | 1.53                   | 1.21 - 1.92 | 0.00031 |
| LIMS2     | 0.62                   | 0.48 - 0.79 | 0.00014 |
| LMNB1     | 1.36                   | 1.16 - 1.59 | 0.00019 |
| LMNB2     | 1.45                   | 1.19 - 1.78 | 0.00026 |
| LMO4      | 1.6                    | 1.24 - 2.07 | 0.00029 |
| LRCH3     | 2.21                   | 1.39 - 3.53 | 0.00085 |
| LRP10     | 1.5                    | 1.19 - 1.89 | 0.00052 |
| LRP12     | 1.85                   | 1.37 - 2.51 | 6.00E-05 |
| LRRC41    | 2.04                   | 1.43 - 2.92 | 9.00E-05 |
| MAD2L1    | 1.53                   | 1.21 - 1.94 | 0.00042 |
| MAPK7     | 1.87                   | 1.33 - 2.64 | 0.00036 |
| MAPKAPK5  | 2.41                   | 1.45 - 4.01 | 0.00072 |
| MARCKSL1  | 1.28                   | 1.12 - 1.47 | 0.00039 |
| MCM10     | 1.97                   | 1.5 - 2.57  | 0        |
| MCM2      | 1.31                   | 1.13 - 1.52 | 0.00046 |
| MCM6      | 1.37                   | 1.15 - 1.64 | 5.00E-04 |
| ME2       | 1.73                   | 1.26 - 2.36 | 6.00E-04 |
| MED8      | 2.22                   | 1.56 - 3.16 | 1.00E-05 |
| MELK      | 1.41                   | 1.18 - 1.69 | 2.00E-04 |
| MGMT      | 0.7                    | 0.56 - 0.86 | 0.00086 |
| MKI67     | 1.37                   | 1.15 - 1.63 | 0.00032 |
| MMP1      | 1.52                   | 1.29 - 1.79 | 0        |
| MPP2      | 3.47                   | 1.97 - 6.13 | 2.00E-05 |
| MRPL3     | 1.89                   | 1.32 - 2.7  | 0.00055 |
| MTBP      | 3.29                   | 1.76 - 6.14 | 0.00018 |
| MTCR1     | 2.04                   | 1.42 - 2.95 | 0.00014 |
| MTF2      | 1.78                   | 1.3 - 2.43  | 0.00034 |
| MTM2      | 1.86                   | 1.41 - 2.45 | 1.00E-05 |
| MUTYH     | 1.84                   | 1.34 - 2.53 | 0.00018 |
| MYBL2     | 1.22                   | 1.1 - 1.36  | 0.00021 |
| MYH6      | 10.74                  | 3.97 - 29.02 | 0        |
| MYOD1     | 32.51                  | 8.67 - 121.83 | 0        |
| NAP1L1    | 1.43                   | 1.16 - 1.77 | 0.00098 |
| NAP1L4    | 1.87                   | 1.32 - 2.66 | 0.00048 |
| NASP      | 1.51                   | 1.19 - 1.93 | 0.00079 |
| NAT10     | 1.84                   | 1.29 - 2.63 | 0.00085 |
| NCAPD2    | 1.44                   | 1.18 - 1.76 | 0.00038 |
| NCAPG     | 1.51                   | 1.24 - 1.87 | 3.00E-05 |
| NCAPG2    | 1.56                   | 1.22 - 1.99 | 0.00044 |
| NCAPH     | 1.42                   | 1.17 - 1.73 | 0.00037 |
| NCBP2     | 1.73                   | 1.27 - 2.35 | 0.00054 |
| NCF2      | 1.32                   | 1.12 - 1.54 | 0.00064 |
| Gene  | Ratio | 95% CI  | p-value |
|-------|-------|---------|---------|
| NCL   | 1.92  | 1.39 - 2.66 | 9.00E-05 |
| NDRG1 | 1.26  | 1.1 - 1.44  | 0.00075  |
| NDUFAF1 | 0.5  | 0.34 - 0.74 | 0.00054  |
| NEIL3 | 1.98  | 1.51 - 2.58 | 0  |
| NME7  | 1.94  | 1.32 - 2.84 | 0.00069  |
| NOL10 | 2.04  | 1.4 - 2.98  | 0.00022  |
| NPC1  | 1.58  | 1.21 - 2.06 | 0.00082  |
| NR0B1 | 1.46  | 1.21 - 1.76 | 8.00E-05 |
| NRAS  | 1.58  | 1.22 - 2.05 | 0.00061  |
| NUP107| 1.76  | 1.29 - 2.4  | 0.00034  |
| NUP155| 1.76  | 1.29 - 2.4  | 0.00039  |
| NUP205| 1.75  | 1.32 - 2.33 | 0.00011  |
| NUP43 | 1.78  | 1.3 - 2.43  | 0.00032  |
| NUP85 | 1.76  | 1.32 - 2.36 | 0.00014  |
| OPRK1 | 2.72  | 1.67 - 4.43 | 6.00E-05 |
| OR10J1| 1.22E+19 | 239704824.17 - 625689795763118e+29 | 0.00048  |
| OSBP2 | 1.63  | 1.22 - 2.18 | 0.00085  |
| P2RY4 | 2.28  | 1.49 - 3.47 | 0.00013  |
| P2RY6 | 1.64  | 1.28 - 2.1  | 8.00E-05 |
| PAGE1 | 1.2   | 1.1 - 1.3   | 2.00E-05 |
| PBK   | 1.37  | 1.16 - 1.62 | 0.00018  |
| PCBP2 | 2.22  | 1.54 - 3.19 | 2.00E-05 |
| PDE6A | 611.5 | 36.74 - 10178.24 | 1.00E-05 |
| PDK3  | 1.68  | 1.27 - 2.22 | 0.00025  |
| PFKFB4| 1.81  | 1.42 - 2.31 | 0  |
| PFN2  | 1.35  | 1.19 - 1.53 | 0  |
| PGLYRP4 | 13.47 | 3.6 - 50.43 | 0.00011  |
| PHC2  | 1.69  | 1.28 - 2.25 | 0.00024  |
| PHKA2 | 1.72  | 1.29 - 2.3  | 0.00023  |
| PIAS2 | 2.76  | 1.52 - 5.02 | 0.00087  |
| PLAUR | 1.38  | 1.15 - 1.66 | 0.00063  |
| PLK1  | 1.45  | 1.22 - 1.72 | 2.00E-05 |
| PLOD2 | 1.51  | 1.25 - 1.83 | 2.00E-05 |
| PNMA2 | 2.11  | 1.43 - 3.12 | 0.00017  |
| POF1B | 1.35  | 1.15 - 1.58 | 0.00018  |
| POLA1 | 1.78  | 1.29 - 2.45 | 4.00E-04 |
| POLQ  | 2.39  | 1.52 - 3.77 | 0.00017  |
| POLR3G| 2.52  | 1.57 - 4.04 | 0.00013  |
| PON1  | 0.86  | 0.8 - 0.93  | 0.00011  |
| POU3F2| 4.58  | 1.94 - 10.78| 5.00E-04 |
| PPM1G | 1.79  | 1.33 - 2.41 | 0.00013  |
| Gene    | Log2FoldChange | Lower CI | Upper CI | p-value |
|---------|----------------|----------|----------|---------|
| PPP1CB  | 1.79           | 1.32 - 2.42 | 0.00018  |
| PPT1    | 1.54           | 1.24 - 1.91  | 9.00E-05 |
| PRDM14  | 16917824       | 28399.58 - 10078062417.21 | 0 |
| PRDX1   | 1.53           | 1.2 - 1.96   | 0.00063  |
| PRKCD   | 1.42           | 1.16 - 1.75  | 0.00067  |
| PRKRA   | 1.99           | 1.37 - 2.89  | 0.00032  |
| PRPF19  | 2.12           | 1.43 - 3.14  | 0.00019  |
| PRR11   | 1.65           | 1.34 - 2.03  | 0 |
| PSMB2   | 1.76           | 1.28 - 2.42  | 0.00055  |
| PSMC3IP | 1.92           | 1.4 - 2.63   | 6.00E-05 |
| PSD1    | 2.03           | 1.37 - 2.99  | 0.00038  |
| PSMD14  | 1.76           | 1.28 - 2.44  | 6.00E-04 |
| PSRC1   | 1.72           | 1.36 - 2.16  | 0 |
| PTP4A2  | 1.92           | 1.4 - 2.64   | 5.00E-05 |
| RAB32   | 1.68           | 1.34 - 2.1   | 1.00E-05 |
| RAB33L1 | 1.42           | 1.2 - 1.68   | 6.00E-05 |
| RACGAP1 | 1.39           | 1.14 - 1.69  | 0.00089  |
| RAD1    | 1.84           | 1.3 - 2.6    | 6.00E-04 |
| RAD51AP1| 1.42           | 1.16 - 1.76  | 0.00092  |
| RAD54B  | 4.46           | 2.2 - 9.03   | 3.00E-05 |
| RAD54L  | 1.57           | 1.25 - 1.99  | 0.0013   |
| RAE1    | 1.79           | 1.28 - 2.52  | 0.00077  |
| RAMP3   | 0.72           | 0.61 - 0.85  | 9.00E-05 |
| RAN     | 1.66           | 1.29 - 2.13  | 8.00E-05 |
| RANBP1  | 1.67           | 1.29 - 2.17  | 0.00012  |
| RAX     | 291.4          | 11.5 - 7384.74 | 0.00058 |
| RBBP4   | 1.61           | 1.23 - 2.1   | 0.00055  |
| RBM14   | 2.05           | 1.36 - 3.07  | 0.00054  |
| RBM17   | 1.94           | 1.43 - 2.64  | 2.00E-05 |
| RBM28   | 2.72           | 1.61 - 4.6   | 2.00E-04 |
| REEP4   | 1.43           | 1.16 - 1.76  | 0.00076  |
| RGS17   | 3.23           | 1.62 - 6.43  | 0.00088  |
| RGS2    | 1.21           | 1.08 - 1.35  | 0.00079  |
| RIBC2   | 1.53           | 1.21 - 1.95  | 0.00052  |
| RIMS3   | 2.53           | 1.57 - 4.09  | 0.00015  |
| RIT1    | 1.66           | 1.26 - 2.19  | 3.00E-04 |
| RNF186  | 2.26           | 1.53 - 3.32  | 4.00E-05 |
| RNF2    | 1.96           | 1.44 - 2.68  | 2.00E-05 |
| RORC    | 0.79           | 0.7 - 0.9    | 0.00052  |
| RPS6KA6 | 3.11           | 1.59 - 6.09  | 0.00091  |
| RQCD1   | 1.79           | 1.31 - 2.43  | 0.00024  |
| RRAGC   | 1.81           | 1.35 - 2.43  | 6.00E-05 |
| Gene    | Ratio | 95% CI  | P-value |
|---------|-------|---------|---------|
| RTN3    | 1.88  | 1.4 - 2.51 | 3.00E-05 |
| RUVBL1  | 1.95  | 1.42 - 2.68 | 3.00E-05 |
| S100A8  | 1.24  | 1.09 - 1.4 | 0.00064 |
| S100A9  | 1.25  | 1.15 - 1.36 | 0.00049 |
| S100BP  | 1.99  | 1.33 - 2.96 | 3.00E-05 |
| SAP30   | 1.66  | 1.3 - 2.14 | 7.00E-05 |
| SCRN1   | 1.31  | 1.12 - 1.52 | 0.00099 |
| SDC3    | 1.41  | 1.15 - 1.74 | 0.00099 |
| SEMA6A  | 1.32  | 1.13 - 1.54 | 0.00045 |
| SEPHS1  | 2.26  | 1.56 - 3.27 | 2.00E-05 |
| SERBP1  | 1.98  | 1.36 - 2.89 | 4.00E-04 |
| SF3A3   | 2.05  | 1.47 - 2.86 | 2.00E-05 |
| SPQ     | 1.88  | 1.4 - 2.53 | 3.00E-05 |
| SGCB    | 1.35  | 1.13 - 1.62 | 0.00093 |
| SHCBP1  | 1.56  | 1.23 - 1.97 | 0.00024 |
| SIGLEC9 | 1.75  | 1.27 - 2.4  | 0.00062 |
| SLC11A1 | 1.53  | 1.2 - 1.96  | 0.00067 |
| SLC16A3 | 1.35  | 1.17 - 1.55 | 3.00E-05 |
| SLC1A6  | 74.66 | 6.18 - 902.28 | 0.00069 |
| SLC1A7  | 1.27  | 1.12 - 1.44 | 0.00024 |
| SLC25A22| 1.72  | 1.28 - 2.3  | 0.00031 |
| SLC25A24| 1.54  | 1.24 - 1.91 | 9.00E-05 |
| SLC2A1  | 1.49  | 1.27 - 1.74 | 0.00069 |
| SLC2A2  | 0.87  | 0.8 - 0.94  | 8.00E-04 |
| SLC36A1 | 1.86  | 1.35 - 2.55 | 0.00013 |
| SLC38A1 | 1.28  | 1.12 - 1.48 | 0.00049 |
| SLC7A8  | 1.53  | 1.25 - 1.86 | 3.00E-05 |
| SMOX    | 1.37  | 1.16 - 1.62 | 0.00017 |
| SMS     | 1.82  | 1.42 - 2.34 | 0.00069 |
| SNX5    | 1.96  | 1.37 - 2.82 | 0.00026 |
| SNX7    | 1.45  | 1.21 - 1.74 | 5.00E-05 |
| SOX11   | 7.34  | 2.4 - 22.48 | 0.00048 |
| SPA17   | 2.08  | 1.41 - 3.07 | 0.00021 |
| SPANXC  | 33.52 | 6.48 - 173.47 | 3.00E-05 |
| SPINK4  | 1.37  | 1.14 - 1.66 | 0.00093 |
| SPP1    | 1.11  | 1.05 - 1.17 | 0.00026 |
| SSRP1   | 1.82  | 1.36 - 2.43 | 5.00E-05 |
| STAM    | 1.72  | 1.25 - 2.36 | 0.00087 |
| STC2    | 1.39  | 1.18 - 1.64 | 1.00E-04 |
| STIP1   | 1.56  | 1.2 - 2.01  | 0.00074 |
| STMN1   | 1.29  | 1.11 - 1.5  | 0.00075 |
| STX3    | 1.48  | 1.21 - 1.82 | 0.00014 |
| STXB5L  | 2003.57 | 41.07 - 97732.78 | 0.00013 |
| Gene    | Fold Change | 95% CI       | p-value |
|---------|-------------|--------------|---------|
| SUV39H2 | 1.94        | 1.37 - 2.73  | 0.00017 |
| TACR3   | 214.82      | 17.16 - 2688.4 | 3.00E-05 |
| TAF1B   | 2.09        | 1.42 - 3.07  | 0.00017 |
| TARDBP  | 2.65        | 1.56 - 4.5   | 3.00E-04 |
| TARS    | 1.8         | 1.34 - 2.42  | 1.00E-04 |
| TBL1X1R1| 1.73        | 1.27 - 2.38  | 0.00062 |
| TBX19   | 2.16        | 1.39 - 3.38  | 0.00068 |
| TCOF1   | 1.73        | 1.3 - 2.29   | 0.00013 |
| TEX15   | 7.02        | 3.39 - 14.57 | 0       |
| TFDP1   | 1.45        | 1.18 - 1.78  | 0.00044 |
| TIMM23  | 2.11        | 1.41 - 3.16  | 0.00029 |
| TMEM39B | 2.17        | 1.45 - 3.24  | 0.00016 |
| TPRSS6  | 0.84        | 0.76 - 0.93  | 0.00055 |
| TNFRSF11A| 1.92       | 1.37 - 2.69  | 0.00013 |
| TNFRSF11B| 1.26      | 1.1 - 1.44   | 0.00068 |
| TOE1    | 2.15        | 1.48 - 3.13  | 6.00E-05 |
| TPX2    | 1.36        | 1.18 - 1.57  | 2.00E-05 |
| TRAPPC4 | 1.66        | 1.24 - 2.23  | 0.00059 |
| TREM1   | 1.59        | 1.27 - 1.99  | 4.00E-05 |
| TRIM36  | 3.12        | 1.8 - 5.4    | 5.00E-05 |
| TRIP13  | 1.59        | 1.32 - 1.91  | 0       |
| TROAP   | 1.33        | 1.12 - 1.57  | 0.00099 |
| TRPC4AP | 1.96        | 1.33 - 2.89  | 7.00E-04 |
| TTC26   | 3.16        | 1.99 - 5.02  | 0       |
| TTK     | 1.74        | 1.39 - 2.17  | 0       |
| TUBG1   | 1.51        | 1.21 - 1.87  | 0.00022 |
| TXLNA   | 1.79        | 1.3 - 2.47   | 0.00036 |
| UAP1L1  | 1.4         | 1.19 - 1.64  | 3.00E-05 |
| UBAP2   | 1.78        | 1.27 - 2.48  | 0.00075 |
| UBE2E1  | 1.87        | 1.38 - 2.54  | 5.00E-05 |
| UCK2    | 1.64        | 1.3 - 2.06   | 3.00E-05 |
| UPB1    | 0.84        | 0.76 - 0.93  | 6.00E-04 |
| USP29   | 5.99E+42    | 2.14458024058968e+21 | 9.00E-05 |
| UTP11L  | 2.72        | 1.87 - 3.98  | 0       |
| VRK2    | 1.98        | 1.43 - 2.74  | 4.00E-05 |
| WASF1   | 1.53        | 1.24 - 1.89  | 9.00E-05 |
| WDHD1   | 1.95        | 1.42 - 2.68  | 4.00E-05 |
| WDR12   | 1.94        | 1.32 - 2.87  | 0.00083 |
| WDR77   | 1.91        | 1.39 - 2.62  | 7.00E-05 |
| XRCC2   | 2.01        | 1.41 - 2.85  | 0.00011 |
| XRCC5   | 1.82        | 1.29 - 2.56  | 0.00061 |
| Gene   | Value | Confidence Interval | P-value |
|--------|-------|---------------------|---------|
| YARS   | 2     | 1.51 - 2.65         | 0       |
| YBX1   | 1.97  | 1.5 - 2.58          | 0       |
| YEATS2 | 1.81  | 1.31 - 2.5          | 0.00034 |
| YRDC   | 2.15  | 1.46 - 3.18         | 1.00E-04|
| YTHDF1 | 2.14  | 1.42 - 3.21         | 0.00026 |
| YTHDF2 | 2.34  | 1.51 - 3.62         | 0.00013 |
| YWHAB  | 1.91  | 1.36 - 2.67         | 0.00017 |
| YWHAQ  | 1.53  | 1.21 - 1.94         | 0.00043 |
| ZBTB40 | 1.89  | 1.31 - 2.72         | 0.00068 |
| ZDHHC18| 2     | 1.45 - 2.77         | 3.00E-05|
| ZMYM1  | 1.97  | 1.35 - 2.88         | 0.00044 |
| ZMYM4  | 1.89  | 1.39 - 2.59         | 6.00E-05|
| ZNF131 | 2.27  | 1.51 - 3.43         | 9.00E-05|
| ZNF239 | 1.83  | 1.42 - 2.36         | 0       |
| ZNF248 | 2.33  | 1.45 - 3.76         | 0.00052 |
| ZNF408 | 1.84  | 1.31 - 2.59         | 0.00044 |
| ZWINT  | 1.3   | 1.12 - 1.51         | 0.00067 |
Table 1: The results of Univariate and multivariate analyses

| Variable                                      | Univariate analysis |          |          | Multivariate analysis |          |
|-----------------------------------------------|---------------------|----------|----------|-----------------------|----------|
|                                               | HR (95% CI)         | P value  | HR (95% CI) | P value               |
| TCGA LIHC cohort (n=366)                      |                     |          |          |                       |
| Neoplasm_histologic_grade (G3-4/G1-2)         | 1.083(0.839-1.397)  | 0.542    |          |                       |
| Person_neoplasm_cancer_status (With tumor/Tumor free) | 1.877(1.42-2.48)   | <0.001   | 1.461(1.081-1.975) | 0.014    |
| Relative_family_cancer_history (Yes/No)       | 1.125(0.867-1.458)  | 0.375    |          |                       |
| Vascular_tumor_cell_type (None/ Micro or Macro) | 0.808(0.602-1.084)  | 0.155    |          |                       |
| Gender (Male / Female)                        | 0.867(0.675-1.113)  | 0.263    |          |                       |
| Age (>median/<median)                         | 1.13(0.881-1.448)   | 0.337    |          |                       |
| Tumor_stage (Stage III+IV/I+II)               | 1.883(1.448-2.448)  | <0.001   | 1.423(1.056-1.918) | 0.02     |
| Riskscore (Low / High)                        | 0.476(0.364-0.614)  | <0.001   | 0.503(0.369-0.686)  | <0.001   |
| BCLC (B+C/A)                                  | 1.403(0.851-2.311)  | 0.184    |          |                       |
