An apoptosis-related gene signature for the prognosis of hepatocellular carcinoma

CURRENT STATUS: UNDER REVIEW

Pengfei Zhu
Department of surgery

kissssu@126.com Corresponding Author
ORCID: https://orcid.org/0000-0003-2141-8636

Zhang Lei
Huazhong University of Science and Technology Tongji Medical College

Du Zhicheng
Zhengzhou University First Affiliated Hospital

Liao Yuan
Zhengzhou University First Affiliated Hospital

Yan Lei
Henan oilfield general hospital

Zhai Wenlong
Zhengzhou University First Affiliated Hospital

Chen Kunlun
Zhengzhou University First Affiliated Hospital

DOI: 10.21203/rs.3.rs-19086/v1

SUBJECT AREAS
Cancer Biology Oncology

KEYWORDS
hepatocellular carcinoma, apoptosis, signature, prognosis, nomogram
Abstract
Hepatocellular carcinoma (HCC) is a major public health burden worldwide owning to high incidence and poor prognosis. Although a mushrooming number of apoptosis-related genes had been disclosed in HCC, the prognostic value and clinical utility of them remain to be illustrated. Here, we defined the data from Gene Expression Omnibus (GEO) as a training cohort and data from The Cancer Genome Atlas-Liver Hepatocellular Carcinoma data set (TCGA-LIHC) as a validation cohort. The apoptosis-related differentially expressed genes (AR-DEGs) were identified with the two cohorts and the Gene Set Enrichment Analysis. Then, we constructed a Lasso-penalized Cox regression model using AR-DEGs and conducted a signature including 14 apoptotic genes to calculate the risk score. Patients with a high risk score indicated worse overall survival than those with low risk. Besides, the 3-year and 5-year area under curve (AUC) values of the signature were above 0.7 in both training and validation cohorts (0.762, 0.818, 0.717, 0.745, respectively). Moreover, a nomogram containing the signature and clinical characteristics presented reliable net benefits for the survival prediction. And the nomogram was tested by probability calibration curves and Decision Curve Analysis (DCA). Furthermore, protein-protein interaction (PPI) and Gene Ontology (GO) enrichment analysis disclosed several noticeable pathways that might clarify the hidden mechanism. Collectively, the present study formed a novel signature based on the 14 apoptotic genes and this possibly predicted prognosis and strengthened the communication with HCC patients about the likely treatment.

Background
Liver cancer is the sixth most frequently diagnosed malignant tumor and the fourth leading cause of cancer-related deaths worldwide, with about 841,000 new cases and 782,000 deaths per annum (1). Hepatocellular carcinoma (HCC) accounts for about 90% of cases of primary liver cancer; the remaining pathologic types are cholangiocarcinoma and mixed hepatocellular-cholangiocarcinoma (2). Surgical resection of the tumor completely and safely is the most effective treatment for HCC, however, most patients are diagnosed at an advanced stage and become unrespectable (3). And thus, the 5-year survival rate for HCC patients was found to be less than 12% (4). Although existing drugs such as sorafenib and immune checkpoint inhibitors could help improve patients’ prognoses,
the effect is still unsatisfactory (5). Therefore, more sensitive biomarkers are earnestly required for the early detection of HCC.

Apoptosis, a programmed cell death, is precisely regulated by numerous checkpoints to maintain equilibrium (6). Suppression of apoptotic cell death was reported to cause cancers, autoimmune diseases, propagation of intracellular pathogens, and neurodegenerative diseases (7). Remarkably, an elevated level of apoptosis could be observed in acute liver injury, but apoptosis generally occurs to a lesser extent in non-alcoholic fatty liver disease and liver cancer (8–10).

In our present research, the apoptosis-related differentially expressed genes (AR-DEGs) were defined as the intersection of the high-throughput data from TCGA, GEO, and GSEA. We then selected a 14-AR-DEGs signature to calculate the risk score with the Lasso-COX regression model. Patients were divided into the high-risk group and low-risk group based on the median risk score. Moreover, univariate and multivariate COX regression models were conducted using risk score and clinical characteristics in both TCGA and GEO data. ROC curve, Calibration curve, and DCA certified the prognostic significance of models. Finally, protein-protein interaction (PPI) and GO enrichment analysis were executed to uncover the underlying mechanism. Briefly, we identified a novel apoptotic gene signature and it might serve as a tool for survival prediction, treatment adjustment, and even clinical classification.

Materials And Methods

Data collection and process

The GSE14520 dataset was obtained from the Gene Expression Omnibus (GEO) database and taken as a training cohort, which consisted of 218 HCC and 221 control samples. All the microarray raw data in the training cohort were transformed with log2(x+1) and normalized to ensure harmonized criteria. Then, the messenger RNA (mRNA) expression profiles of 364 patients with HCC and 50 normal samples were accessed from the UCSC Xena (https://tcga.xenahubs.net) and selected as a validation cohort. Meanwhile, a list of 1982 apoptosis-related genes, named “GO_APOPTOTIC_PROCESS”, was downloaded from GSEA.

Identification of AR-DEGs
Based on the Limma R package, differentially expressed genes (DEGs) were obtained from the training cohort and validation cohort (11). And the cutoff values were set as follows: the adjusted P-value < 0.05 and |log2 fold change (FC)| > 1.5. Then, genes were defined as apoptosis-related differentially expressed genes (AR-DEGs) and enrolled in the subsequent analysis, if they were the intersection of DEGs and apoptosis-related genes.

**Construction of the apoptotic gene signature and risk score calculation**

A Lasso-penalized Cox regression model was applied in Rstudio software with AR-DEGs to shrink and choose variables further (12). The “glmnet” R package (13) helped determine the optimal penalty parameter \( \lambda \) and partial likelihood deviance. And finally, a fourteen-apoptotic-gene signature was conducted and the risk score could be calculated. The calculation formula was as follows:

\[
\sum_{i=1}^{10} (\text{Expression}_i \times \text{Coef}_i)
\]

The \( \text{Expression}_i \) was the mRNA expression of each gene and \( \text{Coef}_i \) represented the corresponding coefficient.

**Evaluation of the prognostic value of the apoptotic gene signature**

Initially, the training cohort was equally divided into the high-risk group (n=109) and low-risk group (n=109) based on the median risk score. And the Kaplan-Meier method was applied to estimate the survival differences between the high and low-risk groups. Besides, we plotted the receiver operating characteristic (ROC) curve to obtain the area under the curve (AUC) for 3-year and 5-year overall survival. Also, the validation cohort was split into the high-risk group (n=182) and low-risk group (n=182), and then survival and ROC analyses were repeated for accuracy.

**Assessment of the apoptotic gene signature in subgroup analysis**

To identify the prognostic value of the apoptotic gene signature in various conditions, we conducted the subgroup analysis in the training cohort. The clinical characteristics were stratified into the following groups: male versus female, age younger than 50 years versus 50 years and older, AFP less than 300 ng/ml (low) versus higher than 300 ng/ml (high), ALT less than 50U/L (low) versus higher than 50U/L (high), early BCLC_stage (0+A) versus advanced BCLC_stage (B+C), main tumor size less than 5 cm versus main tumor size bigger than 5 cm, multinodular negative versus multinodular
positive, cirrhosis negative versus cirrhosis positive, and early TNM_stage (I+II) versus advanced
TNM_stage (III+IV). Then, survival curves were drawn to compare the median survival time between
the high-risk and low-risk subgroups.

**Building and validating the nomogram**

To determine appropriate variables for nomogram building, univariate and multivariate Cox
regression analyses were conducted in the training and validation cohorts separately. Therefore, the
independent prognostic variables were included in the nomogram by the “rms” and “ggplot2” R
packages (14). Then, calibration of the nomogram was assessed visually using 3-year and 5-year
calibration plots. As an innovative but robust assessment method, the 3-year and 5-year decision
curve analyses (DCA) were plotted to identify the net benefits of risk score, whether alone or
cooperating with other clinical characteristics (15). And the x-axis was threshold probability, while the
y-axis was net benefits. The black and green lines represented the two extreme cases in DCA.

**Protein-protein interaction, GO and KEGG analysis**

To uncover underlying mechanisms in the apoptotic gene signature, we implemented protein-protein
interaction (PPI), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)
enrichment analysis with STRING database and “clusterprofiler” R package (16, 17). In the PPI
analysis, the potential interactions with 14 apoptotic genes were predicted under the conditions as
follows: high confidence (interaction score>0.400) and no more than 20 interactors per shell.
Furthermore, the results of PPI analysis were applied in GO and KEGG enrichment analyses and the
outcome was visualized by Cytoscape (Version 3.72) and Rstudio (Version 1.2.5033).

**Extra validation of the apoptotic gene signature with online tools**

To include more normal samples in the validation cohort, we associated TCGA_LIHC with The
Genotype-Tissue Expression (GTex) in the GEPIA database. The expression of the 14 genes in the
apoptotic gene signature was further validated at the mRNA level using GEPIA database, and the
protein level with The Human Protein Atlas (18-20). Also, the genetic alterations of the 14 apoptotic
genes were investigated in The cBioPortal for Cancer Genomics and Human Cancer Metastasis
Database (HCMDB) to explore the relationship with HCC metastasis (21, 22).
Statistical analysis

All of the statistical analyses were processed with R software or SPSS (IBM, SPSS, version 22), and P-value < 0.05 was considered statistically significant unless otherwise specified. One-way ANOVA was called in the “limma” R package for differential expression analysis in cohorts. Then, the Kaplan-Meier survival curve analysis and the log-rank test were applied to analyze overall survival. The AUC value of greater than 0.70 was considered acceptable. Univariate and multivariate Cox proportional hazard models were conducted to obtain the hazard ratios and to select independent prognostic variables.

Results

Establishment of the apoptotic gene signature and risk score calculation

Initially, there were 415 DEGs, including up and down-regulated, in the training cohort (Figure 1a, Additional file 1a), 666 DEGs in the validation cohort (Figure 1b, Additional file 1b), and 1982 apoptosis-related genes from GSEA. Then, 36 AR-DEGs were defined as the intersection of the three sets (Figure 1c) and presented in Additional file 2. Furthermore, AR-DEGs, survival status, and survival time were put into the Lasso-penalized Cox regression model and fourteen apoptotic genes were chosen to conduct the signature when the partial likelihood deviance was minimized (Figure 2a, b). Based on the apoptotic gene signature, the risk score was calculated as follows: risk score=0.0364×expression of SOX4+(-0.2997×expression of PDE1B)+ 0.2101×expression of IVNS1ABP+0.4377×expression of TNFRSF25+(-0.2879×expression of GPLD1)+0.1611×expression of SERPINE1+1.3316×expression of FGF8+0.0890×expression of NRP1+0.1165×expression of UNG+(-0.1386×expression of IL2RB)+(-0.0364×expression of CTLA4)+(-0.3726×expression of CD3E)+(-0.2001×expression of FASLG)+(-0.0560×expression of ATOX1). Moreover, the detailed description and coefficient could be found in Additional file 3.

Identification of the prognostic value of the apoptotic gene signature

Patients were subdivided into the high-risk group and low-risk group with the median risk score. In the training cohort, overall survival was significantly better in the low-risk group than the high-risk group and patients with low-risk scores tended to survive (P-value<0.0001, Figure 2c, Additional file 4). The 3-year AUC and 5-year AUC values were 0.762 and 0.818 respectively, both of which were significant.
(Figure 2d). As for the validation cohort, the high-risk group also showed greater median survival time than the low-risk group (P-value=0.0028, Figure 2e). And the time-dependent ROC revealed the prognostic accuracy of the risk score was 0.717 at 3 years, and 0.745 at 5 years (Figure 2f). Taken together, these results indeed demonstrated the prognostic value of the apoptotic gene signature in training and validation cohorts.

**Subgroup analyses in the clinical characteristics using the signature**

The purpose of subgroup analysis was to detect whether the apoptotic gene signature could benefit specific populations. Based on the risk score, patients in the training cohort were stratified according to gender, age, tumor size, cirrhosis, multinodular, AFP, ALT, TNM_stage, and BCLC_stage. The prognostic value of the signature was certified in all of the subgroup analyses (all P-value<0.01, Figure 3a, b, c, d, Additional file 5), except for female groups (Additional file 5d).

**Construction and validation of the nomogram**

In the training cohort, univariate Cox regression analysis revealed that tumor size, cirrhosis, pathological TNM_stage, BCLC_stage, AFP, and risk score of the apoptotic gene signature were the prognostic factor for overall survival (Figure 4a, Table 1). Then, the significant variables were included in multivariate Cox regression analysis, and only BCLC_stage, cirrhosis, and risk score were still independent prognostic factors (Figure 4b). The forest map and nomogram were plotted built on the three variables. When patients met specific conditions, they received corresponding points and the higher total points got, the worse outcome was (Figure 4c). Moreover, three internal validation groups (70 subjects per group) demonstrated that the nomogram approached an ideal model (Figure 4f, g) and had better predictive performance than cirrhosis or BCLC_stage at 3 years and 5 years with the decision curve analysis (Figure 4d, e).

As for the validation cohort, univariate and multivariate Cox regression analyses indicated that pathological N_stage, M_stage, and risk score were selected to build nomogram (Figure 5a, b, Table 2). Similarly, the nomogram and risk score proved the accuracy and clinical utility for 3-year and 5-year survival prediction (Figure 5c, d, e). And we also generated three internal validation groups (120 subjects per group) randomly to testify the deviations from the ideal model (Figure 5f, g).
In short, the risk score based on the apoptotic signature brought obvious net benefit for prognosis, alone or in combination with classical clinical features.

**Potential mechanisms of the apoptotic gene signature**

Subsequent exploration of the underlying mechanisms was conducted using PPI, GO, and KEGG analyses. A total of 69 genes encoding proteins were enriched and visualized in Additional file 6. Several genes were previously reported to participate in apoptosis and cancer tumorigenesis, including caspase family, FAS, VEGFA and so on. The cut-off of the combined score was set at 0.4 (high) and the detailed information was shown in Additional file 7. Then, the 69 genes were analyzed from the biological process (BP), cellular component (CC), molecular function (MF), and KEGG pathway perspectives. And genes of the signature mainly mapped to the chromosome and took part in growth factor, death receptor, and tumor necrosis factor receptor binding (Figure 6a). Additionally, some of the signaling pathways were related to multiparous tumors and enriched in analysis, including the PI3K-Akt signaling pathway, p53 signaling pathway, MAPK signaling pathway, and TNF signaling pathway. Interestingly, the Hepatitis B pathway was also associated with the genes of the apoptotic signature (Figure 6b).

**Extra validation of the 14 genes of the apoptotic signature**

In the GEPIA2 database, the mRNA expression of the 14 apoptotic genes was concordant with our previous study. The overexpressed genes included SOX4, IVNS1ABP, TNFRSF25, CTLA4, CD3E, FASLG, FGF8, ATOX1, NRP1, and UNG; while the down-regulated ones were PDE1B, GPLD1, SERPINE1, and IL2RB (Additional file 8). Moreover, the protein level expression of IVNS1ABP, NRP1, SOX4, and UNG was significantly increased in immunohistochemistry (Figure 7a). Unfortunately, the remaining 10 genes were not recorded in the protein atlas. In the cBioportal database, FASLG possessed the most frequent genetic alterations (8%), followed by IVNS1ABP (7%), GPLD1 (3%), SOX4 (2.5%) and TNFRSF25 (2.5%). The most frequent alteration type was amplification and deep deletion (Figure 7b). Notably, FALSG would be up-regulated in the primary tumor with metastasis than those without metastasis in HCC patients. And this suggested that FALSG might play a vital role in the neoplasm metastasis according to the HCMDB database (Additional file 8o).
Discussion

According to the annual projections of the World Health Organization, over 1 million people will die from liver cancer in 2030 (23). With a 5-year survival lower than 12%, HCC is the second most lethal tumor after pancreatic cancer (24). Although the molecular mechanisms of HCC remain unclear, the apoptosis played a key role in tumorigenesis and cellular immortality (8, 25). With the development of high-throughput sequencing, more and more single molecular biomarkers were reported (26), but only a few studies highlighted the predictive value of multiple gene signature, especially the apoptotic gene signature in HCC.

In our research, AR-DEGs were singled out and put into a Lasso-penalized Cox regression model, based on TCGA_LIHC, GSE14520, and GSEA data set. The Lasso-COX model avoided the possible collinearity impact and selected fourteen apoptotic genes to identify the signature and calculate the risk score. Whether in training or validation cohorts, the risk score showed a robust prognostic value. As were shown in Figure 2c and Figure 2e, the low-risk patients had longer median survival than those at high risk. Also, the AUC values of 3-year and 5-year were over 0.7 (0.762, 0.818, 0.717, 0.745, respectively), which meant the signature shown high prognostic accuracy. Moreover, the signature was equally effective in subgroup analyses of age, TNM_stage, BCLC_stage, and so on. It was noteworthy that the signature didn’t perform well in the female population. And one possible reason might be the deficiency of samples to detect the statistical difference. Furthermore, the multivariate COX regression analysis and nomogram were conducted in the training cohort and the stability of the nomogram was testified by the calibration curve internally. Particularly, the net benefit obtained by the nomogram far exceeded that in a single variable in 3-year and 5-year DCA than the traditional BCLC stage. Similarly, the nomogram consisted of the apoptotic gene signature, pathological N_stage, and M_stage in the validation cohort. Also, the nomogram exhibited higher prognostic value than that in N_stage or M_stage alone. Unlike existing prognostic models simply relied on gene expression, the individualized prediction could be realized by score calculation. Besides, in Figure 4d and Figure 5d, the nomogram got more decision profit at specific threshold probability when the risk score combined with the traditional indicators. Effective collaboration between the signature and current clinical
stages might provide new ideas for surgeons. With a nomogram, the outcome and treatment for each patient would be tailored referring to a personal score. If possible, patients with high points in the nomogram could receive a more active proposal, including surgical resection and immune checkpoint inhibitors drugs. Therefore, the risk score based on the apoptotic gene signature might have clinical application in the prognosis evaluation and individualized treatment.

In the subsequent mechanism research, the fourteen apoptotic genes were enriched in growth factor binding, death receptor binding, tumor necrosis factor receptor binding, PI3K-Akt pathway, p53 signaling pathway, and Hepatitis B pathway. It is no doubt that the above molecular functions and signaling pathways were involved in liver cancer (27-30). Indeed, most of the fourteen genes had been reported to participate in HBV infection and cancer. Over-expression of SOX4 might promote HCC intrahepatic metastasis by targeting semaphorin 3C (31). The previous study demonstrated that FALSG, also known as FasL, led to the escape from immune surveillance in HCC (32). In our study, the FASLG had the highest mutation frequency and higher mRNA expression in tumors with metastasis, which were shown in Figure 7b and Additional file 8o. Besides, TNFRSF25 was up-regulated in HCC and the silencing gene significantly inhibited liver cancer cell proliferation (33). FGF8, a member of the fibroblast growth factor family, enhanced the invasion of HCC by increasing the expression of YAP1 and EGFR (34). The upregulation of Neuropilin-1 (NRP-1) was related to the boost of arterial blood supply in primary carcinoma and converted the hepatic vasculature to the arterial phenotype (35). In breast cancer, abnormal expression of ATOX1 promoted migration and metastasis through raising the copper (Cu) transportation and Memo protein expression (36, 37). Moreover, the overexpression of UNG was proved to facilitate the HBV mutations and increase HCC risk (38), which was identified with immunohistochemistry staining in Figure 7a. Thus, the high risk score of the signature implied poor prognosis and this might be combined effects of the multiple genes.

However, there are still some weaknesses that deserve comment. First, only differentially expressed genes were defined as candidates and applied for model building. Given the complicated processes of HCC, some crucial apoptotic genes might be omitted and thus reduces the reliability of the signature. Then, more independent validation cohorts with data integrity were in urgent need to pinpoint the
utilization of the signature and address the lack of samples in the subgroup (female, non-cirrhosis). Moreover, the basic experimental studies were warranted to investigate the elusive mechanisms of HCC.

Conclusions
In brief, a novel signature was composed of fourteen apoptotic genes in our study and identified for prediction of the outcome of HCC patients. This gene signature might reflect the deregulation of apoptosis and provided potential prognostic biomarkers and personalized precision treatment in liver cancer. Further basic research and clinical trials would be conducted to confirm our findings.

List Of Abbreviations
HCC: hepatocellular carcinoma; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; DEGs: differentially expressed genes; KM: Kaplan–Meier; LASSO: Least Absolute Shrinkage and Selection Operator; OS: overall survival; FC: fold change; HR: hazard ratio; CI: confidence interval; ROC: receiver operating characteristic; GSEA: gene set enrichment analysis; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; DCA: Decision Curve Analysis.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
The datasets generated and analyzed during the current study are available from the corresponding author on a reasonable request.

Competing interests
The authors declare that they have no competing interests.

Funding
The project was supported by the National Natural Science Foundation of China (No. 81702863).

Authors' contributions
Pengfei Zhu analyzed the data and Lei Zhang wrote the manuscript; Liao Yuan and Zhicheng Du
revised the manuscript; Kunlun Chen and Wenlong Zhai designed the study and revised the manuscript.

Acknowledgment

The authors are grateful for the selfless support from other members of the Department of Hepatobiliary and Pancreatic Surgery.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68(6):394-424.

2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA: A Cancer Journal for Clinicians. 2016;66(2):115-32.

3. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol. 2019.

4. Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019;380(15):1450-62.

5. Cheng AL, Hsu C, Chan SL, Choo SP, Kudo M. Challenges of combination therapy with immune checkpoint inhibitors for hepatocellular carcinoma. J Hepatol. 2020;72(2):307-19.

6. Danial NN. BCL-2 family proteins: critical checkpoints of apoptotic cell death. Clin Cancer Res. 2007;13(24):7254-63.

7. Singh R, Letai A, Sarosiek K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. Nat Rev Mol Cell Biol. 2019;20(3):175-93.

8. Schwabe RF, Luedde T. Apoptosis and necroptosis in the liver: a matter of life and death. Nat Rev Gastroenterol Hepatol. 2018;15(12):738-52.

9. Luedde T, Kaplowitz N, Schwabe RF. Cell death and cell death responses in liver disease: mechanisms and clinical relevance. Gastroenterology. 2014;147(4):765-83
10. Guicciardi ME, Malhi H, Mott JL, Gores GJ. Apoptosis and necrosis in the liver. Compr Physiol. 2013;3(2):977-1010.

11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43(7):e47-e.

12. Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in medicine. 1997;16(4):385-95.

13. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software. 2010;33(1):1.

14. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. Journal of clinical oncology. 2008;26(8):1364-70.

15. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74.

16. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research. 2016:gkw937.

17. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284-7.

18. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic acids research. 2019;47(W1):W556-W60.

19. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. Nature genetics. 2013;45(6):580.
20. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human protein atlas. Nature biotechnology. 2010;28(12):1248-50.
21. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1-pl.
22. Zheng G, Ma Y, Zou Y, Yin A, Li W, Dong D. HCMDB: the human cancer metastasis database. Nucleic acids research. 2018;46(D1):D950-D5.
23. Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019;380(15):1450-62.
24. Craig AJ, von Felden J, Garcia-Lezana T, Sarcognato S, Villanueva A. Tumour evolution in hepatocellular carcinoma. Nature Reviews Gastroenterology & Hepatology. 2019.
25. Marquardt JU, Edlich F. Predisposition to Apoptosis in Hepatocellular Carcinoma: From Mechanistic Insights to Therapeutic Strategies. Front Oncol. 2019;9:1421.
26. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nature Genetics. 2015;47(5):505-11.
27. Jing Y, Sun K, Liu W, Sheng D, Zhao S, Gao L, et al. Tumor necrosis factor-α promotes hepatocellular carcinogenesis through the activation of hepatic progenitor cells. Cancer letters. 2018;434:22-32.
28. Min X, Wen J, Zhao L, Wang K, Li Q, Huang G, et al. Role of hepatoma-derived growth factor in promoting de novo lipogenesis and tumorigenesis in hepatocellular carcinoma. Molecular oncology. 2018;12(9):1480-97.
29. Wei J-c, Qu K, Wang Z-x, Wu Q-f, Zhang L-q, Pang Q, et al. Sorafenib inhibits proliferation and invasion of human hepatocellular carcinoma cells via up-regulation of p53 and suppressing FoxM1. Acta Pharmacologica Sinica. 2015;36(2):241-51.
30. Lim CJ, Lee YH, Pan L, Lai L, Chua C, Wasser M, et al. Multidimensional analyses
reveal distinct immune microenvironment in hepatitis B virus-related hepatocellular carcinoma. Gut. 2019;68(5):916-27.

31. Liao YL, Sun YM, Chau GY, Chau YP, Lai TC, Wang JL, et al. Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma. Oncogene. 2008;27(42):5578-89.

32. Shin EC, Shin JS, Park JH, Kim H, Kim SJ. Expression of fas ligand in human hepatoma cell lines: role of hepatitis-B virus X (HBX) in induction of Fas ligand. International journal of cancer. 1999;82(4):587-91.

33. Zhang YC, Guo LQ, Chen X, Wang GN, Ni R, Wang MC, et al. The role of death receptor 3 in the biological behavior of hepatocellular carcinoma cells. Molecular medicine reports. 2015;11(2):797-804.

34. Pei Y, Sun X, Guo X, Yin H, Wang L, Tian F, et al. FGF8 promotes cell proliferation and resistance to EGFR inhibitors via upregulation of EGFR in human hepatocellular carcinoma cells. Oncology Reports. 2017;38(4):2205-10.

35. Zhang Y, Liu P, Jiang Y, Dou X, Yan J, Ma C, et al. High Expression of Neuropilin-1 Associates with Unfavorable Clinicopathological Features in Hepatocellular Carcinoma. Pathol Oncol Res. 2016;22(2):367-75.

36. Blockhuys S, Wittung-Stafshede P. Copper chaperone Atox1 plays role in breast cancer cell migration. Biochemical and Biophysical Research Communications. 2017;483(1):301-4.

37. MacDonald G, Nalvarte I, Smirnova T, Vecchi M, Aceto N, Doelemeyer A, et al. Memo is a copper-dependent redox protein with an essential role in migration and metastasis. Sci Signal. 2014;7(329):ra56-ra.

38. Liu W, Wu J, Yang F, Ma L, Ni C, Hou X, et al. Genetic Polymorphisms Predisposing the
Interleukin 6-Induced APOBEC3B-UNG Imbalance Increase HCC Risk via Promoting the Generation of APOBEC-Signature HBV Mutations. Clin Cancer Res. 2019;25(18):5525-36.

Supplementary Information

Additional file 1
Volcano plot of Differentially expressed genes (DEGs). (a) DEGs in GSE14520. (b) DEGs in TCGA_LIHC.

Additional file 2
The apoptosis-related differentially expressed genes.

Additional file 3
Detailed information of the 14 genes in the apoptotic signature.

Additional file 4
The curve of risk score and heatmap of the fourteen apoptotic genes in high- and low-risk group.

Additional file 5
Kaplan–Meier survival analysis in subgroup. (a) Age>50. (b) Age<50. (c) Male. (d) Female. (e) Cirrhosis. (f) Non-cirrhosis. (g) Multinodular. (h) Non-multinodular. (i) Tumor size>5cm. (j) Tumor size<5cm. (k) Early TNM stage. (l) Advanced TNM stage. (m) Early BCLC stage. (n) Advanced BCLC stage.

Additional file 6
Protein-protein interaction of the fourteen apoptotic genes.

Additional file 7
The combined score of protein-protein interaction from the STRING database.

Additional file 8
The mRNA level expression of the fourteen apoptotic genes in association with TCGA_LIHC, The Genotype-Tissue Expression (GTEx) database, and the Human Cancer Metastasis Database (HCMDB). (a) The mRNA level expression of ATOX1. (b) The mRNA level expression of CD3E. (c) The mRNA level expression of CTLA4. (d) The mRNA level expression of FGF8. (e) The mRNA level expression of UNG. (f) The mRNA level expression of FASLG. (g) The mRNA level expression of IVNS1ABP. (h) The mRNA
level expression of NRP1. (i) The mRNA level expression of SOX4. (j) The mRNA level expression of TNFRSF25. (k) The mRNA level expression of GPLD1. (l) The mRNA level expression of IL2RB. (m) The mRNA level expression of PDE1B. (n) The mRNA level expression SERPINE1. (o) The mRNA level expression of FALSG.

FALSG would be up-regulated in the primary tumor with metastasis than those without metastasis in HCC patients.

Tables

Table 1. Univariate and multivariate cox regression analyses of OS in GSE14520

| Variables    | Univariate HR(95%CI) | p-value | Multivariate HR(95%CI) | p-value |
|--------------|----------------------|---------|------------------------|---------|
| Gender       | 1.708(0.824-3.540)   | 0.150   | 1.901(0.511-1.587)     | 0.717   |
| Age          | 0.962(0.794-1.166)   | 0.694   | 0.901(0.511-1.587)     | 0.717   |
| ALT          | 1.060(0.688-1.634)   | 0.791   | 0.901(0.511-1.587)     | 0.717   |
| Tumor size   | 2.001(1.299-3.081)   | 0.002   | 0.901(0.511-1.587)     | 0.717   |
| Multinodular | 1.591(0.984-2.572)   | 0.058   | 0.901(0.511-1.587)     | 0.717   |
| Cirrhosis    | 4.650(1.143-8.912)   | 0.032   | 2.990(1.451-6.160)     | 0.003   |
| TNM stage    | 3.487(2.233-5.470)   | <0.001  | 1.443(0.692-3.006)     | 0.328   |
| BCLC stage   | 3.520(2.253-5.489)   | <0.001  | 4.566(1.978-7.543)     | <0.001  |
| AFP          | 1.622(1.056-2.491)   | 0.027   | 1.011(0.702-1.729)     | 0.674   |
| Riskscore(high/low) | 4.857(2.933-8.042) | <0.001  | 4.117(2.426-6.985)     | <0.001  |

ALT: alanine aminotransferase; AFP: alpha-fetoprotein; HR: hazard ratio; CI: confidence intervals.

Table 2. Univariate and multivariate cox regression analyses of OS in TCGA
| Variables                | Univariate HR(95%CI) | p-value | Multivariate HR(95%CI) | p-value |
|-------------------------|----------------------|---------|------------------------|---------|
| Gender                  | 1.072(0.891-1.286)   | 0.488   |                        |         |
| Age                     | 1.233(0.876-1.755)   | 0.248   |                        |         |
| Race                    | 1.423(0.982-2.051)   | 0.127   |                        |         |
| T_stage                 | 1.381(1.180-1.419)   | 0.005   | 0.992(0.776-1.266)     | 0.094   |
| N_stage                 | 1.604(1.342-1.918)   | <0.001  | 1.595(1.334-1.907)     | <0.001  |
| M_stage                 | 1.331(1.107-1.601)   | 0.002   | 1.404(1.196-1.799)     | 0.007   |
| Riskscore(high/low)     | 1.712(1.198-2.455)   | <0.001  | 1.568(1.193-2.249)     | <0.001  |

HR: hazard ratio; CI: confidence intervals.

Figures
Figure 1

Differentially expressed genes (DEGs) in GSE14520 and TCGA_LIHC. (a) Heatmap of DEGs in GSE14520. (b) Chromosome Distribution Chart of DEGs in TCGA_LIHC. (c)Venn diagram of the apoptosis-related differentially expressed genes (AR-DEGs).
Lasso-penalized Cox regression model, Kaplan–Meier analysis and Time-dependent ROC analysis. (a, b) Lasso-penalized Cox regression model to shrink and choose variables. (c) Kaplan–Meier analysis and Time-dependent ROC analysis (d) in the GSE14520. (e) Kaplan–Meier analysis and Time-dependent ROC analysis (f) in the TCGA_LIHC.
Figure 3

KM survival subgroup analysis in the GSE14520. (a) AFP < 300 ng/ml. (b) AFP > 300 ng/ml.
(c) ALT < 50U/L, (d) ALT > 50U/L.

### Figure 4

Univariate and multivariate Cox regression analyses, calibration curve, and decision curve analyses (DCA) in the training cohort. (a, b) Univariate and multivariate Cox regression analyses.
models with a risk score and clinical characteristics. (c) Nomogram to predict overall survival at 3 years and 5 years. (d, e) Net benefit in the nomogram and single variables using the DCA. (f, g) Calibration curve to verify the model accuracy.

Figure 5
Construction and verification of the nomogram and in the validation cohort. (a, b) Univariate and multivariate Cox regression analyses. (c) Nomogram using risk score, N_stage, and M_stage. (d, e) The 3-year and 5-year DCA for the nomogram. (f, g) Calibration curve to validate the nomogram at 3 years and 5 years.
Figure 6

GO and KEGG analysis. (a) Molecular Function, Biological Process, Cellular Component
(b) KEGG enrichment analysis.
Extra validation with online tools. (a) Immunohistochemistry results of IVNS1ABP, SOX4, NRP1, and UNG. (b) Genetic alterations of the 14 apoptotic genes.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Supplemental Files.zip