Development and validation of a novel ten-gene signature predicting prognosis in Hepatocellular carcinoma

Guanbao Zhou
Ningbo First Hospital
Genjie Lu
Ningbo Medial Center Li Huili Hospital
Liang Yang
Ningbo First Hospital
Yangfang Lu (luyangfang2020@163.com)
Ningbo Medical Centre Li Huili Hospital

Research

Keywords: Hepatocellular carcinoma; ten-gene signature; molecular classification; risk score; overall survival

Posted Date: July 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-46933/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is the most common type of liver cancer with relatively poor prognosis. Thus, we aimed to identify novel molecular biomarkers to effectively predict the prognosis of HCC patients and eventually guide treatment.

Methods: Prognosis-associated genes were determined by Kaplan-Meier and multivariate Cox regression analyses using the expression and clinical data of 373 HCC patients from The Cancer Genome Atlas (TCGA) database and validated in an independent Gene Expression Omnibus (GEO) dataset. The classification of AML was performed by unsupervised hierarchical clustering of ten gene expression levels. A prognostic risk score was established based on a linear combination of ten gene expression levels using the regression coefficients derived from the multivariate Cox regression models.

Results: A total of 183 genes were significantly associated with prognosis in HCC. *SLC25A15, RAB8A, GOT2, SORBS2, IL18RAP* were top five protective genes, while *FHL3, AMD1, DCAF13, UBE2E1, PTDSS2* were top five risk genes in HCC. *SLC25A15, GOT2, IL18RAP* were significantly down-regulated and *DCAF13, PTDSS2* and *SORBS2* were significantly up-regulated in the HCC samples and these genes exhibited high accuracy in differentiating HCC tissues from normal liver tissues. Hierarchical clustering analysis of the ten genes discovered three clusters of HCC patients. HCC tumors of cluster1 and 2 were significantly associated with more favourable OS than those of cluster3, cluster2 tumors showed higher pathologic stage than cluster3 tumors. The risk score was predictive of increased mortality rate in HCC patients.

Conclusions: The ten-gene signature and the risk score may turn out to be novel molecular biomarkers and stratification of HCC patients to considerably ameliorate the prognostic prediction.

Background

Liver cancer ranks the sixth in cancer incidence rate and the fourth in the cause of cancer-related deaths in 2018 throughout the world. Global cancer statistics revealed an estimated 841,000 new liver cancer cases and 782,000 deaths caused by the disease annually [1]. Hepatocellular carcinoma (HCC) is the most prevalent subtype of liver cancer, accounting for 75-85% of liver cancer cases. Hepatitis B and C virus infection, aflatoxin and smoking exposure, excessive alcohol drinking, chemical injury with metabolic disorder are associated with the development of HCC [2–4]. Though new therapeutic methods have been developed, the clinical outcomes remain unsatisfactory for the HCC patients after surgical resection, with a 5-year overall survival (OS) rate of as low as 30% [2,5]. Thus, the determination of novel prognostic biomarkers is critical to the early diagnostic detection and improvement of prognosis in HCC patients.

In clinical settings, tumor staging systems ,such as tumor, nodes, metastasis staging (TNM) [6], and Barcelona Clinic liver cancer staging [7], are the most common method to guide prognosis assessment and treatment selection after surgery. In addition to the above staging systems, new tumor characteristics

have been introduced to improve the prediction accuracy of tumor prognosis, such as serum alpha fetoprotein (AFP), microvascular invasion and tumor differentiation [8–10]. Currently, the gene expression signatures of large cohorts of HCC tumors have been generated by high throughput sequencing and publicly available to genomics researchers, which paves the way for developing novel and robust predictive biomarkers for clinical outcomes in HCC. Recent studies have identified several gene signatures at mRNA level that had great potential in predicting HCC prognosis[11,12]. However, there is still lack of a systematic and genome-wide investigation of prognostic biomarkers at the mRNA level in HCC.

In this study, we performed Kaplan-Meier and multivariate analyses to screen for prognosis-associated genes using the expression and clinical data of 373 HCC patients from The Cancer Genome Atlas (TCGA) database [13] and validated the results in an independent Gene Expression Omnibus (GEO) dataset[14,15]. We established a prognostic risk score based on a linear combination of ten gene expression levels to effectively predict the overall survival (OS) of HCC patients. Lastly, we utilized unsupervised hierarchical clustering of ten genes and defined HCC genomic subgroups and their relevance to clinical outcomes. The completion of our study opens the avenue for developing molecular markers in prognostication and treatment decision making for HCC patients.

Methods And Materials

Data acquisition

The training and validation datasets came from two different sources. The former was obtained from the TCGA dataset. The training dataset consisted of normalized expression counts of 20532 genes and clinical characteristics information, including age, gender, tumor weight, TNM staging, residual tumor, liver transplant, overall survival status, days to the last follow-up or death of 377 HCC patients. The validation dataset was downloaded from the GEO database (GSE14520)[14,15]. The expression values of 13238 genes in the GSE14520 dataset were relative signal intensities normalized by robust multi-array average (RMA). The clinical factors analyzed in the GEO cohort comprised age, gender, serum α-fetoprotein (AFP) level, cirrhosis, main tumor size, multinodular tumors, TNM staging, days to the last follow-up or death, overall survival status.

Bioinformatics analysis of cell cycle genes

To study the biological functions and possible signaling pathways of prognosis-associated genes, the enrichment of gene ontology (GO) terms [16] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [17] pathways was analyzed by the bioinformatics online tool of gprofiler, version 6.8 [18]. The raw P values were corrected by the g:SCS algorithm which is a tailor-made algorithm for computing multiple testing correction for p-values gained from GO and pathway enrichment analysis.

Overall survival analyses

The associations of gene expression with OS were investigated in the TCGA and GEO datasets by various statistical methods. In brief, HCC patients were split into the high and low expression groups according to the median expression values or risk score. Risk score = expression of gene 1 × β1 + expression of gene 2 × β2 +⋯+ expression of gene n × βn. β values are the regression coefficients derived from the multivariate logistic regression analysis of the TCGA dataset. We used Kaplan-Meier survival analysis and log-rank methods to compare the difference of OS rates between the two groups using the survival package [19,20]. We performed multivariate Cox regression analysis to confirm whether gene expression or risk score were independent prognostic biomarkers after adjustment of the prognosis-related clinical factors using the coxph function of the survival package. P<0.05 was predefined as statistically significant. The prognosis-associated genes were then further classified as protective genes (Hazard ratio [HR] <1) and risk genes (0< HR <1).

Differential expression analyses of prognosis-associated genes

In order to characterize the diagnostic values of prognosis-associated genes, the expression values of 50 pairs of HCC tissues and normal liver tissues were obtained from the TCGA dataset and 247 HCC tissues and 241 non-cancerous tissues were accessed from the GEO dataset for validation. Differentially expressed genes (DEGs) were determined by comparing gene expression difference between HCC tissues and normal liver tissues using the limma package[21]. Receiver operating curves (ROC) were established and then area under curve (AUC) values were computed by the R package of pROC to determine the diagnostic values of the prognosis-associated genes[22].

Unsupervised hierarchical clustering analysis

HCC patient in the TCGA and GEO cohorts were classified into three distinct subgroups by unsupervised hierarchical clustering of top five protective and risk genes using the function Pheatmap of the R package of pheatmap [23]. Difference in continuous variables was compared by the analysis of variance (ANOVA) test among three subgroups of HCC patients and by student t test between subgroups of patients. Categorical variables were compared by fisher exact test among the three subgroups of HCC patients. Kaplan-Meier curves were plotted using the R package of survival[19], and survival rates were compared among the three clusters of HCC patients using the log-rank test. P value below 0.05 was predefined as statistically significant.

Results

Clinical characteristics of HCC patients

Patient data, treatment parameters and clinical characteristics of both the TCGA and GEO cohorts are summarized in Table1 and supplementary Table1. In the TCGA cohort, patient's age, pathologic stage and pathologic stage T were negatively associated with OS (P<0.05 for all cases, student t test or fisher exact test, Table1). In the GEO cohort, patient's tumor size, multinodular, cirrhosis, tumor stage and AFP level were found to be adversely correlated with OS (P<0.05 for all cases, student t test or fisher exact test,

Table1). The other characteristics did not show a significant association with OS in the TCGA and GEO datasets (P values >0.05 for all cases, student t test or fisher exact test, Table1 and supplementary Table1).

Overall survival analysis

To evaluate the predictive capability of gene expression for patients' OS, the 373 HCC patients in the TCGA dataset were divided into low and high expression groups according to median values. Kaplan-Meier survival analysis showed that high expression levels of 1253 genes and 2082 genes were associated with favourable or poor prognosis respectively, such as solute carrier family 25 member 15(*SLC25A15*), RAB8A, member RAS oncogene family (*RAB8A*), glutamic-oxaloacetic transaminase 2(*GOT2*), sorbin and SH3 domain containing 2 (*SORBS2*), four and a half LIM domains 3(*FHL3*), adenosylmethionine decarboxylase 1(*AMD1*), DDB1 and CUL4 associated factor 13(*DCAF13*) and ubiquitin conjugating enzyme E2 E1 (*UBE2E1*) (P <0.05 for all cases, log rank test, Figure1, supplementary Figure1). Then multivariate analyses were performed between patients' OS and the mortality-associated features, including patients' age, pathologic stage and pathologic stage T and 3335 gene expression levels. Multivariate survival analyses confirmed that high expression of 823 genes was significantly associated with decreased mortality, such as *SLC25A15, RAB8A, GOT2, SORBS2.* The hazard ratios of the four genes ranged from 0.44 to 0.51, with a mean of 0.48 (P<0.05 for all cases, supplementary Table2). While high expression of 1442 genes was associated with increased mortality, such as *FHL3, AMD1, DCAF13, UBE2E1.* The hazard ratio ranged from 1.99 to 2.39 for the four genes, with a mean of 2.18 (P<0.05 for all cases, supplementary Table2, supplementary Figure1).

Validation of survival analyses

In order to validate the findings above, the association between 2265 gene expression and mortality was evaluated in 242 HCC samples of the GEO dataset. Of 2265 prognosis-associated genes, Kaplan-Meier survival analysis confirmed that high expression levels of 231 genes was associated with a favourable prognosis in HCC. In contrast, high expression of 244 genes was associated with a poor prognosis (P <0.05 for all cases, log rank test, supplementary Figure1). Then multivariate analyses were performed between patients' OS and the mortality-associated features, including tumor size, multinodular, cirrhosis, tumor stage and AFP level and 475 gene expression levels. Multivariate survival analyses confirmed that high expression of 98 genes was associated with decreased mortality, while high expression of 85 genes was associated with increased mortality. The top protective genes *SLC25A15, RAB8A, GOT2, SORBS2,* interleukin 18 receptor accessory protein (*IL18RAP*) and risk genes *FHL3, AMD1, DCAF13, UBE2*E1, phosphatidylserine synthase 2(*PTDSS2*) were confirmed to be significantly associated with OS in the validation dataset (P< 0.05 for all cases, supplementary Figure1, supplementary Table3).

Bioinformatics analysis of prognosis-associated genes.

The GO function analysis indicated that the 98 risk genes were mainly enriched in 48 GO terms, such as the regulation of DNA replication, cell cycle, DNA repair, DNA recombination and cell division

(Supplementary Table4, adjusted P values <0.05 for all cases). The 85 protective genes were significantly enriched in 26 GO terms, such as organic acid metabolic process, small molecule metabolic process, steroid metabolic process, protein localization to peroxisome and aspartate family amino acid metabolic process. The KEGG pathway analysis suggested that 98 risk genes were significantly enriched in the DNA replication and cell cycle signalling pathways, the 89 protective genes were significantly enriched in cholesterol metabolism, bile secretion, PPAR signalling pathway and metabolic pathway (Supplementary Table5, adjusted P values <0.05 for all cases).

Assessment of diagnostic value

Of the 183 prognosis-associated genes, 88 up-regulated genes and 66 down-regulated genes have been found between 50 pairs of HCC tissues and normal controls, while 93 genes were down-regulated and 81 were up-regulated in 247 HCC samples as compared to 241 normal controls (adjusted P values < 0.05 for all cases, supplementary Figure2A and B). By intersecting the DEGs between the GEO and TCGA cohorts, we found 83 common down-regulated and 64 up-regulated genes in HCC (P< 0.05 for all cases, supplementary Figure2C and 2D). ROC curves were constructed to further explore the diagnostic values of the 183 genes. 24 genes in particular exhibited high accuracy in differentiating HCC tissues from bone marrow tissues, such as DNA topoisomerase II alpha (*TOP2A*), thyroid hormone receptor interactor 13(*TRIP13*), cytochrome P450 family 2 subfamily C member 8(*CYP2C8*), Rac GTPase activating protein 1(*RACGAP1*) and cyclin dependent kinase 1(*CDK1*) (P values <0.05, AUC>0.9 for all cases, Supplementary Table6). Of the top five protective genes and top five risk genes, *SLC25A15, GOT2, IL18RAP* were significantly down-regulated and *DCAF13, PTDSS2* and *SORBS2* were significantly up-regulated in the HCC samples as compared to normal liver tissues in the two cohorts (P<0.05 for all cases, AUC> 0.8 for 50% cases, Figure2).

Unsupervised hierarchical clustering analysis

We performed the classification of 373 HCC patients using the gene panel comprising top five protective genes, *SLC25A15, RAB8A, GOT2, SORBS2, IL18RAP* and top five risk genes, *FHL3, AMD1, DCAF13, UBE2E1, PTDSS2* and found three clusters of HCC patients in the TCGA dataset (supplementary Figure3). Cluster1 and 2 tumors were significantly associated with more favourable OS than cluster3 tumors, cluster2 tumors showed higher pathologic stage than cluster3 tumors (P values <0.05 for all cases, fisher exact test or log-rank test, Supplementary table7 and supplementary Figure4). Hierarchical clustering analysis of the gene panel revealed three subgroups of HCC patients in the GEO dataset (supplementary Figure5). The cluster1 HCC patients were associated with smaller tumor size, lower cancer stage, lower AFP levels and more favourable OS than cluster2 or 3 tumors (P values <0.05 for all cases, fisher exact test or log-rank test, Supplementary table8, Figure3).

Risk score is a negative predictor for overall survival in HCC

We constructed a risk score formula by linear combination of expression values of top five protective and risk genes using the coefficients of the multivariate Cox regression models from the TCGA dataset. Risk

score =0.44×expression of *SLC25A15* + 0.47×expression of *RAB8A* +0.51×expression of *GOT2* + 0.51×expression of *SORBS2* + 0.51×expression of *IL18RAP*+2.39×expression of *FHL3* + 2.23×expression of *AMD1* +2.11×expression of *DCAF13* + 1.99×expression of *UBE2E1* + 1.97×expression of *PTDSS2*. Dead HCC patients showed significantly higher risk scores than decreased ones in both cohorts (P<0.05 for all cases, student t test, Figure4A). Kaplan-Meier survival analysis suggested that HCC patients with high risk scores showed poorer survival than those with low risk scores in the TCGA dataset (P< 0.05 for all cases, Figure4B). The multivariate analysis further confirmed that risk score was a risk factor for predicting overall survival independently of prognosis-related clinical features in HCC. Using the same methods, the negative association between risk score and overall survival was also validated in the GEO dataset (P< 0.05 for all cases, Table2 and Figure4C).

Discussion

Over the past decade, despite significant advances in the treatment of HCC, HCC remains a serious threat to public health around the world. TNM staging and AFP measurement are the conventional approaches to assess HCC prognosis in clinical practice. Whereas, given HCC samples are considerably heterogeneous, it remains critical and urgent to identify new molecular biomarkers and develop prognostic prediction models with high accuracy. In recent years, gene-signatures based on aberrant mRNA have drawn much attention and displayed great potential in prognostic prediction of HCC patients [11,24,25]

In this study, we performed Kaplan-Meier and multivariate analyses using the mRNA expression data of two independent datasets and found 183 genes significantly associated with OS of HCC patients. The GO term and KEGG pathway enrichment analyses indicated that the 98 risk genes were mainly enriched in the regulation of DNA replication, cell cycle, DNA repair signalling pathways. While, the 85 protective genes were significantly enriched in the organic acid metabolic process, small molecule metabolic process, steroid metabolic process, protein localization to peroxisome and aspartate family amino acid metabolic process, PPAR signalling pathway and metabolic pathway. The difference in the GO terms and KEGG pathway of prognosis-associated genes is of great importance to the identification of prognostic biomarkers and eventually novel therapeutic targets in HCC. For instance, given the high enrichment of risk genes in the cell cycle signalling pathway, cell cycle genes may become potential candidates for developing prognostic biomarkers or druggable targets in HCC[26].

We also identified a ten-gene panel comprising *SLC25A15, RAB8A, GOT2, SORBS2, IL18RAP*, *FHL3, AMD1, DCAF13, UBE2E1, PTDSS2* expression levels that could predict the OS of HCC patients. Furthermore, we established a risk score formula using a linear combination of 10 gene expression levels and β-values of multivariate Cox regression models. The risk score was negatively correlated with OS after adjusting for known prognosticators. The ten genes play diverse roles in the tumorigeneses of cancers. For instance, the *SORBS2* gene is a component of the acto-myosin ring at the apical junctional complex in epithelial cells and may function as a tumor suppressor in cancer. *SORBS2* binds the 3' untranslated regions of two metastasis suppressors, WAP four-disulfide core domain 1 and Interleukin-17D, which inhibits the

invasiveness of ovarian cancer and impacts the polarization from monocyte to myeloid-derived suppressor cell and M2-like macrophage [27]. The *SORBS2* gene was expressed ectopically in various cervical cancer cell lines. Enhanced *SORBS2* expression led to a marked decrease in cell proliferation, colony formation and anchorage-independent growth in cervical cancer cells [28].*UBE2E1* is a member of ubiquitin-conjugating enzyme E2 class. In line with our study, *UBE2E1* expression was adversely correlated with AML survival and patients' response to induction chemotherapy [29]. lncRNA RP11-732M18.3 promoted cell proliferation and G1/S cell cycle transition of glioma cells. Mechanistically, lncRNA RP11-732M18.3 promoted the binding of 14-3-3β/α with *UBE2E1*, causing the p21 degradation[30]. These studies combined with the results in our study suggest that *UBE2E1* may have oncogenic function in cancers.

Furthermore, the ten-gene expression signature effectively stratified HCC patients into three subgroups with different survival probabilities. In addition to prognostic value, the ten genes also showed diagnostic value for HCC patients. Our study, for the first time, reported *SLC25A15, GOT2, IL18RAP* were significantly down-regulated and *DCAF13, PTDSS2* and *SORBS2* were significantly up-regulated in the HCC samples and they exhibited high accuracy in differentiating HCC tissues from normal liver tissues. Lastly, the ten genes may also pave the way for developing targeted therapies for HCC patients. For instance, over-expression of *SORBS2* expression resulted in a marked decrease in cell proliferation, colony formation and anchorage-independent growth in cervical cancer cells [28].

Conclusion

Collectively, this study firstly revealed a novel ten-gene signature which has prognostic and diagnostic values and successfully stratifies HCC patients with different prognostic probabilities. A higher risk score indicates a poorer prognosis. These findings will help researchers identify new treatments for HCC and to provide more therapeutic targets to cure HCC patients in the future.

List Of Abbreviations

Hepatocellular carcinoma: HCC

The Cancer Genome Altas: TCGA

Gene Expression Omnibus: GEO

Serum α-fetoprotein: AFP

Tumor-node-metastasis: TNM

Gene ontology: GO

Kyoto Encyclopedia of Genes and Genomes: KEGG

Differentially expressed gene: DEG

Receiver operating curves: ROC

Area under curve: AUC

Overall survival: OS

Hazard ratio: HR

Confidence interval: CI

Declarations

Acknowledgement

None

Authors' contributions

Yangfang Lu conceived the study. Liang Yang performed the survival analyses and differentially expressed gene analysis. Guanbao Zhou, Genjie Lu conducted unsupervised hierarchical clustering analysis and development of risk score. Yangfang Lu prepared the manuscript. All authors read and approved the final manuscript.

Funding

The study was financially supported by the project of Ningbo science and technology bureau (No 2019A610212).

Ethics approval and consent to participate

None

Consent for publication

None

Competing interests

The authors declare no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available upon reasonable request.

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 Cancer J Clin. 2018;68:394–424.
2. 10.1016/j.bpg.2014.08.007
   Bosetti C, Turati F, La Vecchia C. Hepatocellular carcinoma epidemiology. Best Pract Res Clin Gastroenterol [Internet]. Elsevier Ltd; 2014;28:753–70. Available from: http://dx.doi.org/10.1016/j.bpg.2014.08.007.
3. Fan JH, Wang JB, Jiang Y, Xiang W, Liang H, Wei WQ, et al. Attributable causes of liver cancer mortality and incidence in China. Asian Pacific J Cancer Prev. 2013;14:7251–6.
4. 10.1016/j.bbadis.2018.08.009
   Zhang HE, Henderson JM, Gorrell MD. Animal models for hepatocellular carcinoma. Biochim Biophys Acta - Mol Basis Dis [Internet]. Elsevier; 2019;1865:993–1002. Available from: https://doi.org/10.1016/j.bbadis.2018.08.009.
5. V GS,V, S. K. Hepatocellular carcinoma- A review. J Pharm Sci Res [Internet]. 2017;9:1276–80. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L618105150%0Ahttp://library.deakin.edu.au/resserv?sid=EMBASE&issn=09751459&id=doi:&atitle=Hepatocellular+carcinoma-+A+review&stitle=J.+Pharm.+Sci.+Res.&title=Journal+of+Pharmaceutica.
6. Subramaniam S, Kelley RK, Venook AP. A review of hepatocellular carcinoma (HCC) staging systems. Chinese Clin Oncol. 2013;2:1–12.
7. Llovet JM, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: The BCLC staging classification. Semin Liver Dis. 1999;19:329–37.
8. Marrero JA, Kudo M, Bronowicki J. The Challenge of Prognosis and Staging for Hepatocellular Carcinoma. Oncologist. 2010;15:23–33.
9. 10.1038/s41598-017-12834-1
   Bai DS, Zhang C, Chen P, Jin SJ, Jiang GQ. The prognostic correlation of AFP level at diagnosis with pathological grade, progression, and survival of patients with hepatocellular carcinoma. Sci Rep [Internet]. Springer US; 2017;7:1–9. Available from: http://dx.doi.org/10.1038/s41598-017-12834-1.
10. 10.1038/nrc3245
    Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: Impact on clinical outcome. Nat Rev Cancer [Internet]. Nature Publishing Group; 2012;12:298–306. Available from: http://dx.doi.org/10.1038/nrc3245.
11. 10.1186/s12935-019-0858-2
    Liu GM, Zeng HD, Zhang CY, Xu JW. Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. Cancer Cell Int [Internet]. BioMed Central; 2019;19:1–13. Available from: https://doi.org/10.1186/s12935-019-0858-2.

12. 10.1186/s12967-019-1946-8
Zhu GQ, Yang Y, Chen EB, Wang B, Xiao K, Shi SM, et al. Development and validation of a new tumor-based gene signature predicting prognosis of HBV/HCV-included resected hepatocellular carcinoma patients. J Transl Med [Internet]. BioMed Central; 2019;17:1–13. Available from: https://doi.org/10.1186/s12967-019-1946-8.
13. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327–41.e23.
14. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res. 2010;70:10202–12.
15. Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology. 2012;142:957–66.
16. Consortium TGO. Gene ontology: Tool for the identification of biology. Nat Genet. 2000;25:25–9.
17. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27:29–34.
18. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35:193–200.
19. Therneau T. Survival Analysis. Cran [Internet]. 2016; Available from: https://cran.r-project.org/web/packages/survival/survival.pdf.
20. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.2264&rep=rep1&type=pdf.
21. 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 Res. 2015;43:e47.
22. 10.1186/1471-2105-12-77
Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics [Internet]. BioMed Central Ltd; 2011;12:77. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-77.
23. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data. R Packag. version. 2005.
24. Long J, Zhang L, Wan X, Lin J, Bai Y, Xu W, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. J Cell Mol Med. 2018;22:5928–38.
25. Ke K, Chen G, Cai Z, Huang Y, Zhao B, Wang Y, et al. Evaluation and prediction of hepatocellular carcinoma prognosis based on molecular classification. Cancer Manag Res. 2018;10:5291–302.
26. Liping X, Jia L, Qi C, Liang Y, Dongen L, Jianshuai J. Cell Cycle Genes Are Potential Diagnostic and Prognostic Biomarkers in Hepatocellular Carcinoma. 2020;2020.

27. Zhao L, Wang W, Huang S, Yang Z, Xu L, Yang Q, et al. The RNA binding protein SORBS2 suppresses metastatic colonization of ovarian cancer by stabilizing tumor-suppressive immunomodulatory transcripts. Genome Biol Genome Biology. 2018;19:1–20.
28. Backsch C, Rudolph B, Steinbach D, Scheungraber C, Liesenfeld M, Häfner N, et al. An integrative functional genomic and gene expression approach revealed SORBS2 as a putative tumour suppressor gene involved in cervical carcinogenesis. Carcinogenesis. 2011;32:1100–6.
29. Luo H, Qin Y, Reu F, Ye S, Dai Y, Huang J, et al. Microarray-based analysis and clinical validation identify ubiquitin-conjugating enzyme E2E1 (UBE2E1) as a prognostic factor in acute myeloid leukemia. J Hematol Oncol. 2016;9:1–8.
30. 10.1016/j.ebiom.2019.06.002
Kang CM, Bai HL, Li XH, Huang RY, Zhao JJ, Dai XY, et al. The binding of lncRNA RP11-732M18.3 with 14-3-3 β/α accelerates p21 degradation and promotes glioma growth. EBioMedicine [Internet]. Elsevier B.V.; 2019;45:58–69. Available from: https://doi.org/10.1016/j.ebiom.2019.06.002.

Tables

Table I. Association between the clinicopathologic characteristics and overall survival in the TCGA dataset

Variables	Group	Alive	Dead	P value	Statistical method
Age		58.18	61.81	0.02	Student t test
Tumour weight		307.78	305.18	0.95	Student t test
Gender	Female	73	48	0.13	Fisher’s exact test
	Male	171	78		
Pathologic stage	I	129	41	<=0.001	Fisher’s exact test
	II	61	25		
	III	41	44		
	IV	2	3		
Pathologic stage T	1	135	45	<=0.001	Fisher’s exact test
	2	66	28		
	3	38	42		
	4	3	10		
Pathologic stage N	0	168	83	0.60	Fisher’s exact test
	1	2	2		
Pathologic stage M	0	182	83	0.1	Fisher’s exact test
	1	1	3		
Residual tumor	No	217	106	0.2	Fisher’s exact test
	Yes	9	9		
Liver transplant	No	80	50	0.30	Fisher’s exact test
	Yes	4	0		

Table2. Multivariate analyses between OS and the risk score in the TCGA and GEO datasets

TCGA dataset				GEO dataset			
Clinical feature	HR	95%CI	P value	Clinical feature	HR	95%CI	P value
Age	1.01	1-1.03	0.11	Tumor size	1.03	0.6-1.78	0.91
TNM stage	1.05	0.5-2.21	0.9	Multinodular	0.58	0.34-1	0.05
Pathologic t	1.62	0.8-3.29	0.18	Cirrhosis	3.42	0.83-14.03	0.09
Risk score	2.53	1.7-3.75	<0.001	TNM stage	2.46	1.67-3.63	<0.001
				AFP	1.29	0.83-2.01	0.26
				Risk score	2.03	1.3-3.18	<0.001

Notably, HR and CI refers to hazard ratio and confidence interval respectively.

Figures

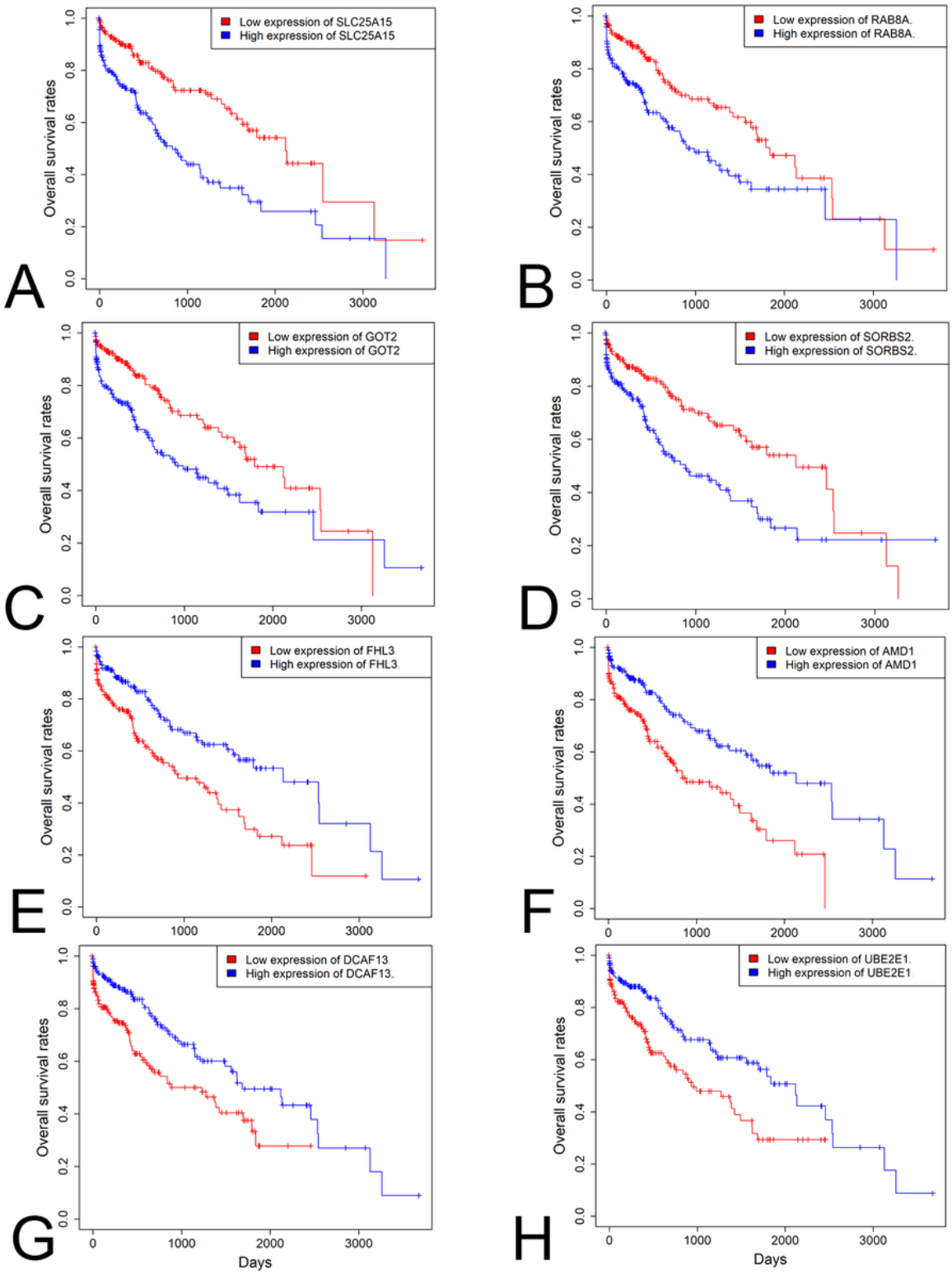

Figure 1

Kaplan-Meier survival analysis of patients' OS with SLC25A15, RAB8A, GOT2, SORBS2, FHL3, AMD1, DCAF13 and UBE2E1 (A-G) expression levels in the TCGA cohort.

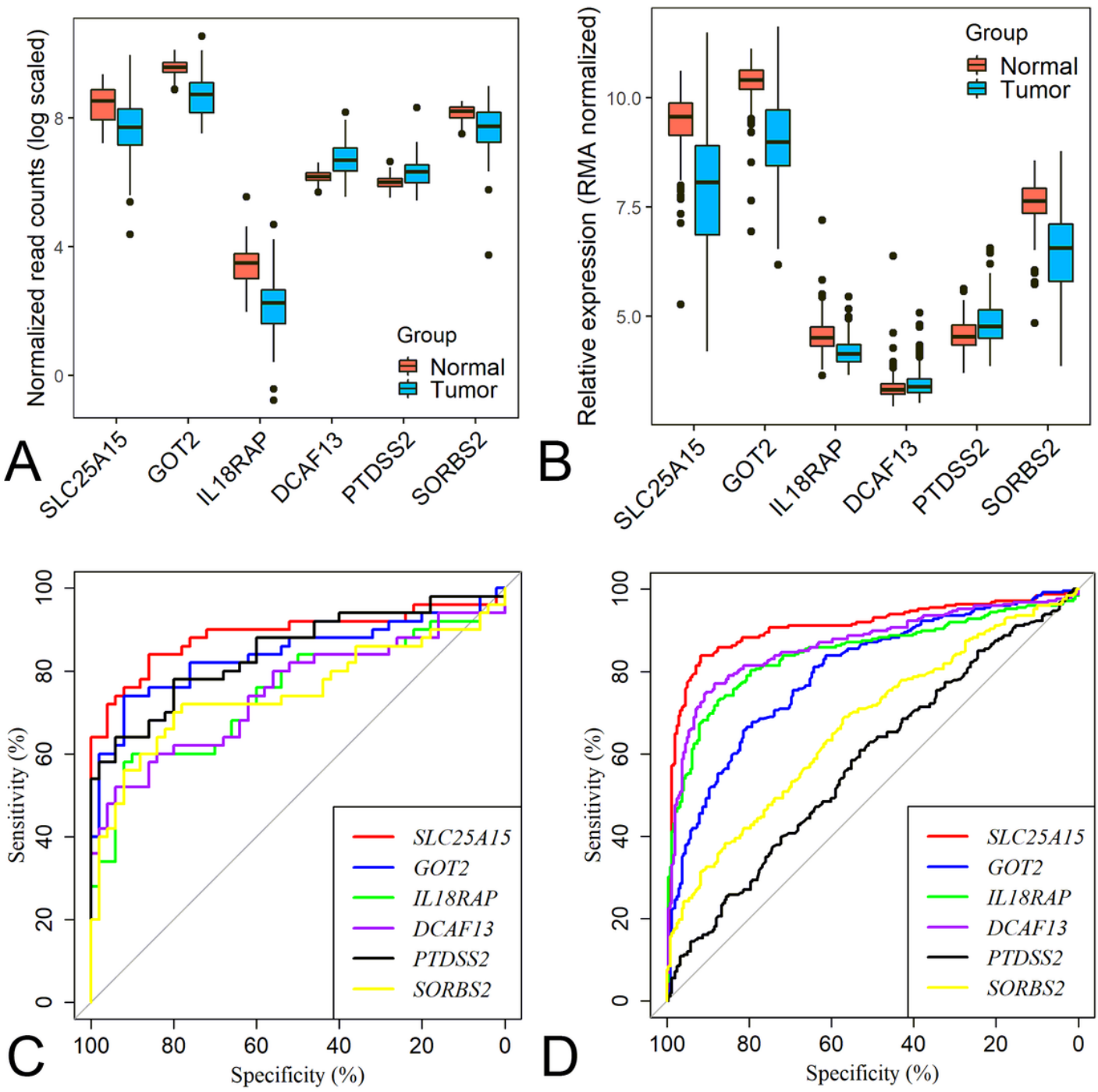

Figure 2

Differential expression gene analysis A. The expression difference of SLC25A15, GOT2, IL18RAP, DCAF13, PTDSS2 and SORBS2 in 50 pairs of HCC and normal liver samples of the TCGA dataset. B. The expression difference of SLC25A15, GOT2, IL18RAP, DCAF13, PTDSS2 and SORBS2 in 242 HCC and 247 normal liver samples of the GEO dataset. C. The ROC curves for the six genes in the TCGA cohort. D. The ROC curves for the six genes in the GEO cohort.

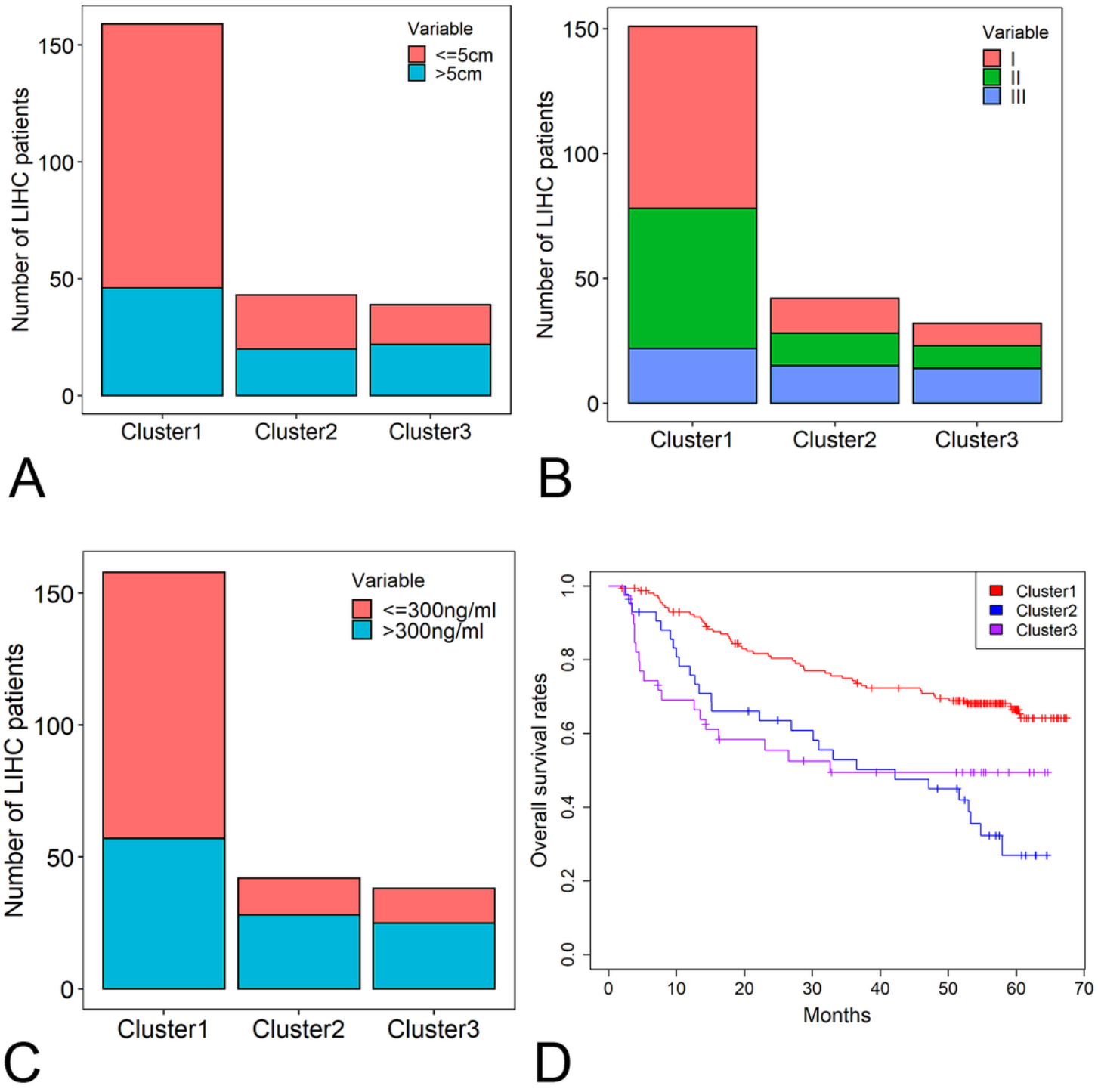

Figure 3

The three clusters (1–3) of HCC patients showed significant differences in tumor size (A), TNM stage (B), AFP level (C) and OS (D) in the GEO dataset.

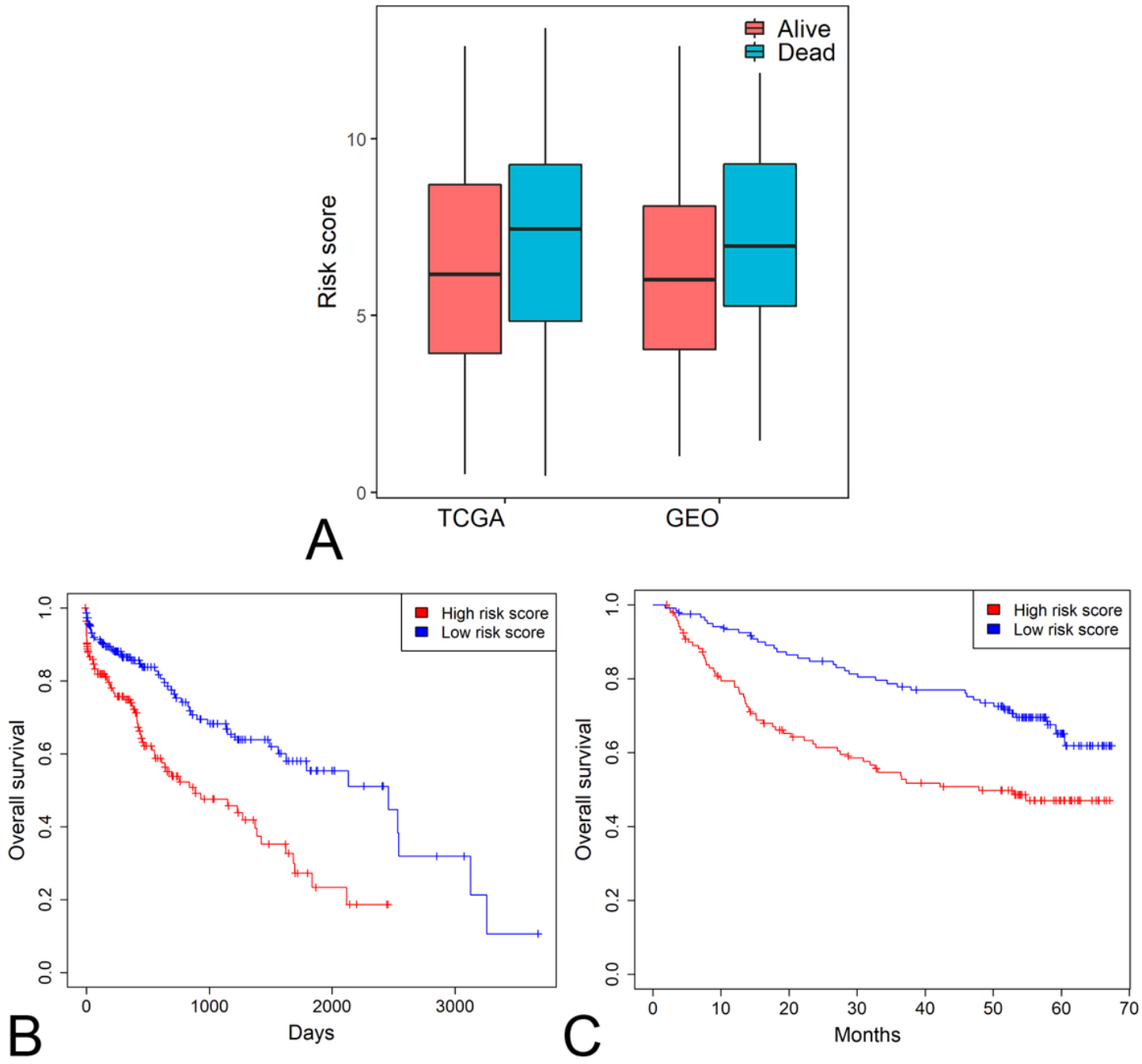

Figure 4

Risk score is a negative prognostic biomarker in HCC. A. The difference of risk scores between the alive and decreased HCC patients in the two cohorts. B. The comparison of Kaplan-Meier survival curves between high and low risk score groups in the TCGA dataset. C. The comparison of Kaplan-Meier survival curves between high and low risk score groups in the GEO dataset.

Supplementary Files

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

- Supplementarytables.docx
- Supplementarytables.docx
- SupplementaryFigure1.tif
- SupplementaryFigure1.tif
- SupplementaryFigure2.tif
- SupplementaryFigure2.tif
- SupplementaryFigure3.jpeg
- SupplementaryFigure3.jpeg
- SupplementaryFigure4.tif
- SupplementaryFigure4.tif
- SupplementaryFigure5.jpeg
- SupplementaryFigure5.jpeg