

Original Article

Identification and validation of a ten-gene set variation score as a diagnostic and prognostic stratification tool in hepatocellular carcinoma

Chanhua Zou^{1*}, Chunling Yuan^{2*}, Jiazhou Ye³, Ziyu Liu², Xing Gao², Xuemin Piao², Rongyun Mai², Yan Lin², Donghua Zou⁴, Zhaoshan Fang⁴, Rong Liang²

Departments of ¹General Medicine, ²Medical Oncology, ³Hepatobiliary Surgery, Guangxi Medical University Cancer Hospital, Nanning 530021, Guangxi, People's Republic of China; ⁴The Fifth Affiliated Hospital of Guangxi Medical University, Nanning 530022, Guangxi, People's Republic of China. *Equal contributors.

Received April 9, 2020; Accepted August 1, 2020; Epub September 15, 2020; Published September 30, 2020

Abstract: We aimed to identify a hepatocellular carcinoma (HCC)-specific gene set during progression. Using the HCC data set from The Cancer Genome Atlas, we found that 10 genes were gradually upregulated with the progression of HCC and associated with survival, classified as HCC-unfavorable genes; 29 genes were gradually downregulated and associated with survival, classified as HCC-favorable genes. Gene set variation analysis (GSVA) was used to score individual samples against the two gene sets. Receiver operating characteristic (ROC) curve analysis showed that both the HCC-unfavorable GSVA score and HCC-favorable GSVA score were reliable biomarkers for diagnosing HCC. Moreover, tROC curve analysis and univariate/multivariate Cox proportional hazards analyses indicated that the HCC-unfavorable GSVA score was an independent prognostic biomarker. The results were validated in an external independent data set. Our results support a ten-gene set variation score as a diagnostic and predictive strategy tool in HCC.

Keywords: Hepatocellular carcinoma, prognostic stratification system, HCC, TCGA

Introduction

According to global cancer statistics from 2018, hepatocellular carcinoma (HCC) has become the sixth most common cancer and the fourth leading cause of cancer death in the world [1]. The main causes of HCC include chronic hepatitis B virus (HBV), hepatitis C virus (HCV) infection, aflatoxin contaminated food, heavy drinking, obesity, smoking, and type 2 diabetes [2, 3]. Approximately 80-90% of HCC patients have potential cirrhosis [4]. Although there are many treatments, such as hepatectomy, liver transplantation, radiofrequency ablation, embolization therapy, and molecule-targeted chemotherapy, therapeutic efficacy in advanced HCC is still limited [5]. Therefore, it is essential to explore molecular mechanisms in HCC and robust diagnostic and prognostic markers.

With the development of high-throughput sequencing technology, an increasing number

of molecular diagnostic markers have been identified for HCC. Most studies on the prognosis of HCC focused on a single or several molecules [6-9], whereas less attention has been paid to characteristic gene sets related to HCC progression. To date, there is no widely accepted molecular prognostic biomarker for HCC.

In this study, we identified two HCC progression characteristic gene sets: an HCC-unfavorable gene set and an HCC-favorable gene set. Gene set variation analysis (GSVA) was used to score individual samples against the two gene sets. Both the HCC-unfavorable GSVA score and HCC-favorable GSVA score may serve as biomarkers for HCC prognosis.

Materials and methods

Materials acquiring

In The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/>) [10], there are 171 HCC sam-

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

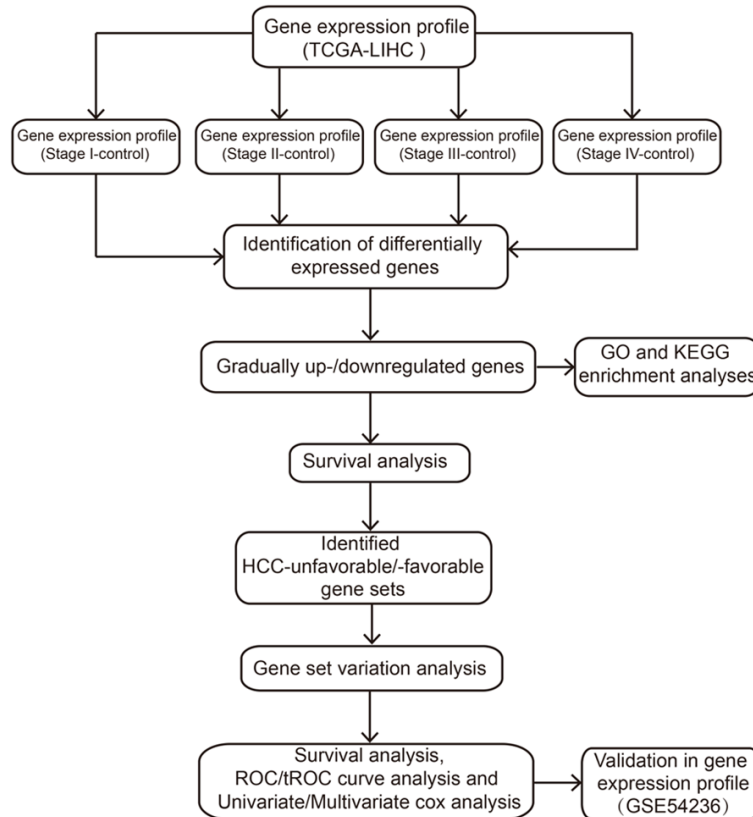


Figure 1. Flow chart of this study.

ples at stage I, 86 at stage II, 83 at III, and 5 at stage IV as well as 42 healthy liver tissue samples. In addition, GSE54236 based on the GPL6480 platform was downloaded from Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) [11], including 81 HCC samples and 80 healthy liver tissue samples, and was used to verify the prognostic value of the identified gene sets. The “normalize Between Arrays” function in the limma package [12] was applied to normalize gene expression profiles. If a gene responded to multiple probes, the average value of these probes was considered to be the expression value of the corresponding gene. The workflow of the present study is shown in **Figure 1**.

Differentially expressed gene (DEG) analysis

The RNA sequencing expression profile (displayed as read counts) of HCC in TCGA was downloaded. The voom function [13] in the limma package was employed to normalize the RNA sequencing data; the limma package [12, 14] was also used to identify DEGs in the 4

stages of HCC and healthy liver tissue samples. A DEG was considered at $P < 0.01$ after FDR correction and $|\log_{2}FC| > 1.5$ as the threshold. During HCC progression, if a DEG was gradually upregulated ($\log_{2}FC_{\text{stage I vs control}} < \log_{2}FC_{\text{stage II vs control}} < \log_{2}FC_{\text{stage III vs control}} < \log_{2}FC_{\text{stage IV vs control}}$) or gradually downregulated ($\log_{2}FC_{\text{stage I vs control}} > \log_{2}FC_{\text{stage II vs control}} > \log_{2}FC_{\text{stage III vs control}} > \log_{2}FC_{\text{stage IV vs control}}$), then it was considered to be a gene characteristic of HCC progression.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the gradually upregulated and downregulated genes were performed using the clusterProfiler package [15] in R. $P < 0.05$ was considered significant.

Survival analysis and HCC-unfavorable/favorable gene sets

The median expression value for each gradually upregulated and gradually downregulated gene was used as the cutoff to dichotomize patients into high-expression and low-expression groups. We applied Kaplan-Meier survival analysis with the log-rank method to evaluate the association of a gene with prognosis. Survival analysis was performed using the survival package ([https://CRAN.R-project.org/package = survival](https://CRAN.R-project.org/package=survival)) in R. $P < 0.01$ was considered to be significant. A gene that was gradually upregulated in HCC progression and associated with a poor prognosis was defined as an HCC-unfavorable gene; a gene that was gradually downregulated in HCC progression and associated with a good prognosis was defined as an HCC-favorable gene. Two HCC progression characteristic gene sets were established: an HCC-unfavorable gene set and an HCC-favorable gene set.

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

Calculation of HCC-unfavorable/favorable GSVA scores

The GSVA package implements a nonparametric unsupervised method, called Gene Set Variation Analysis (GSVA), for assessing gene set enrichment (GSE) in gene expression microarray data or RNA-seq data. The GSVA package [16] in R was used to calculate HCC-unfavorable GSVA and HCC-favorable GSVA scores for an individual sample.

Receiver operating characteristic (ROC) curve analysis, univariate/multivariate Cox proportional regression analysis and time-dependent ROC (tROC) curve analysis

The pROC package [17] was utilized to conduct ROC curve analysis to evaluate the ability of gene sets to diagnose HCC. Univariate/multivariate Cox proportional hazards analyses were used to compare the relative prognostic value of the two GSVA score systems with that of routine clinicopathological features. $P < 0.05$ was considered significant. tROC curve analysis was applied to evaluate the prognostic value regarding the 2-year survival rate of the independent prognostic factors.

Validation of the GSVA score system in an independent data set

The HCC-unfavorable/favorable GSVA score was calculated using the HCC cohort from TCGA, and ROC curve, tROC curve and survival analyses were performed in GSE54236.

Validation of aberrant expression of HCC-unfavorable genes at the protein level

The Human Protein Atlas (<https://v15.protein-atlas.org/>) [18] provides information on the tissue and cell distribution of all 24,000 human proteins. We scanned the Human Protein Atlas web tool to validate differential expression of HCC-unfavorable genes at the protein level.

Results

Various genes differentially expressed with HCC progression

PCA analysis of TCGA data showed that the expression patterns of global genes (**Figure 2A**) could not distinguish HCC from controls.

Compared to control samples, there were 2114 DEGs in stage I HCCs (**Figure 2B**), 2714 DEGs in stage II HCCs (**Figure 2C**), 2871 DEGs in stage III HCCs (**Figure 2D**) and 3718 DEGs in stage IV HCCs (**Figure 2E**). A total of 1273 common DEGs were identified in stage I-IV HCCs (**Figure 2F**). Among them, 82 DEGs were gradually upregulated and 176 DEGs gradually downregulated with HCC progression. PCA analysis showed that the expression patterns of these genes could distinguish HCC from controls (**Figure 2G**).

Gradually upregulated/downregulated genes involved in multiple HCC-related pathways

Functional enrichment analysis was used to explore the biological functions and related pathways of gradually the upregulated and downregulated genes. The results of GO analysis revealed that the gradually upregulated genes were significantly enriched in negative regulation of megakaryocytes, olfactory bulb interneuron differentiation, endothelial growth factor stimulus and other biological processes (**Figure 3A**). In contrast, the gradually downregulated genes are mainly involved in xenobiotic metabolic processes, responses to xenobiotic stimuli, cellular responses to xenobiotic stimuli and other biological processes (**Figure 3B**). Furthermore, the gradually upregulated genes are significantly involved in multiple pathogens of HCC-related pathways, such as viral carcinogenesis and alcoholism (**Figure 3C**) and the gradually downregulated genes in the PPAR signaling pathway, retinol metabolism, steroid hormone biosynthesis, bile secretion, and ABC transporter pathways (**Figure 3D**).

HCC-unfavorable/favorable gene set

A total of 10 gradually upregulated genes (ACP4, ATP6 VOD2, BRSK1, CHGA, CLEC2 L, CREG2, CYP19A1, PNCK, STEAP1B, TMC7) were associated with poor overall survival and classified as an HCC-unfavorable gene set (**Figure 4A**). Moreover, 4 genes (BRSK1, CLEC2 L, PNCK and TMC7) were included in The Human Protein Atlas, and all of them were highly expressed in HCC compared to normal liver (**Figure 4B**), which is consistent with our findings. Twenty-nine gradually downregulated genes were associated with good prognosis and classified as the HCC-favorable gene set (**Table 1**). CAMK4, DMGDH, IYD, CCDC42,

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

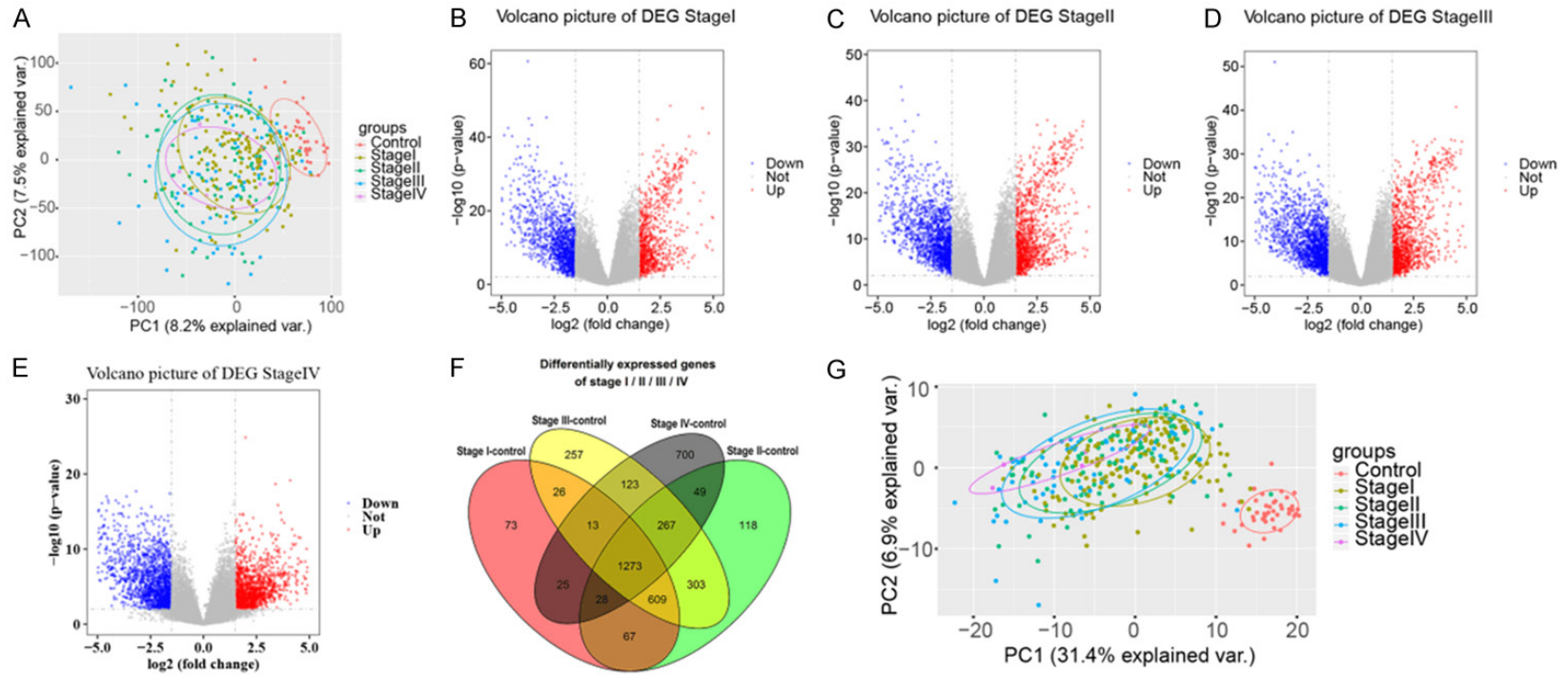
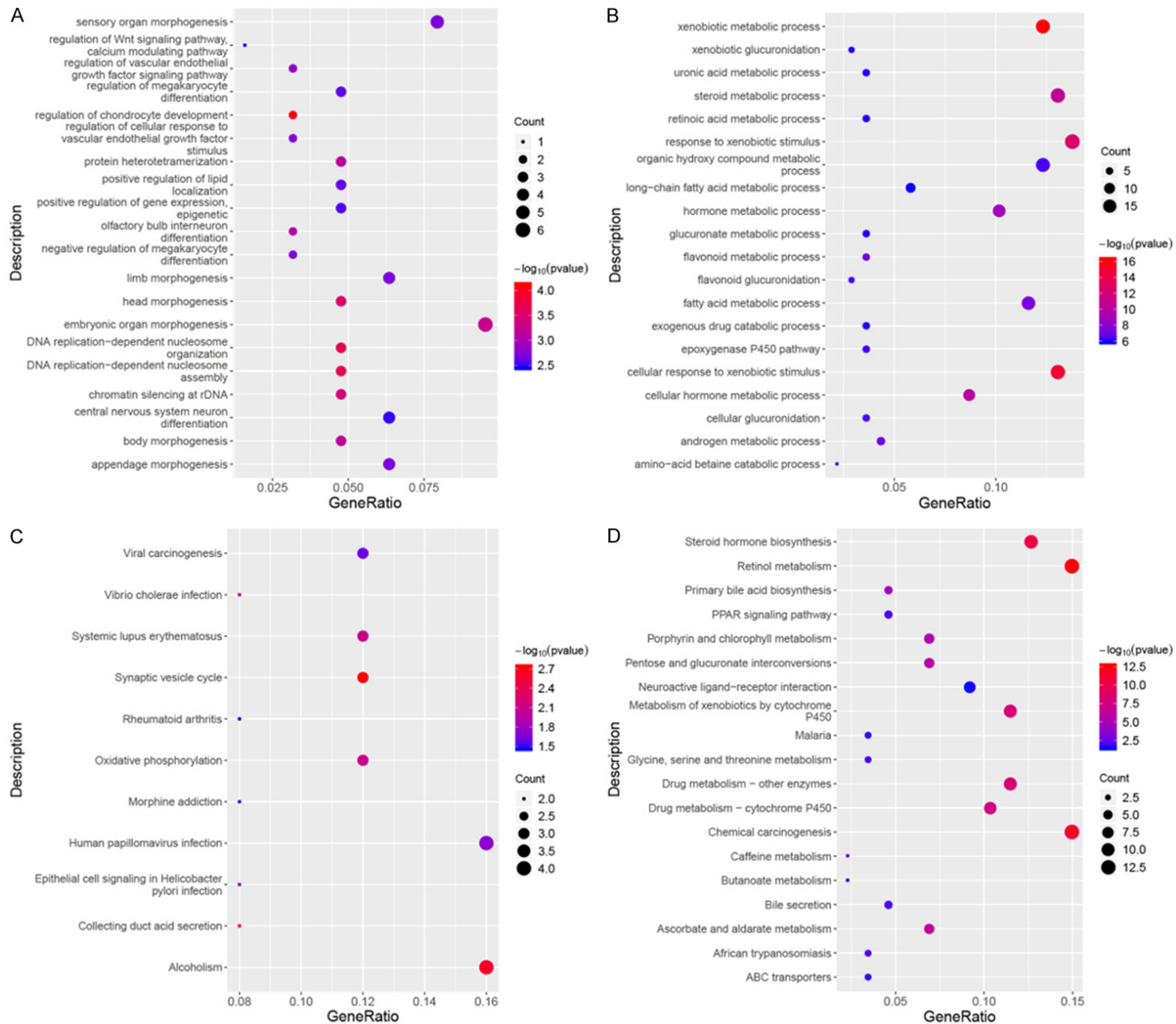


Figure 2. Differential expression gene (DEG) analysis and principal component analysis (PCA). A. PCA of TCGA HCC gene expression profiles. B. Volcano plot of differentially expressed genes between stage I HCC and normal liver tissue. C. Volcano plot of differentially expressed genes between stage II HCC and normal liver tissue. D. Volcano plot of differentially expressed genes between stage III HCC and normal liver tissue. E. Volcano plot of differentially expressed genes between stage IV HCC and normal liver tissue. Red represents upregulated genes, blue represents downregulated genes, and gray represents no significantly differentially expressed genes. F. Common differentially expressed genes in HCC stages I-IV. Intersected genes represent genes that are dysregulated in all four stages of liver cancer development. G. PCA of gradually upregulated and downregulated genes.

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma



A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

Figure 3. Biological processes and KEGG pathway enrichment analysis of gradually upregulated/downregulated genes. A. Biological process of gradually upregulated genes. B. Biological process of gradually downregulated genes. C. KEGG pathway analysis of gradually upregulated genes. D. KEGG pathway analysis of gradually downregulated genes.

ESR1, CPEB3, CYP3A43, VIPR1, AKR1D1 and ADRA1A were the ten genes with the most significant association with a good prognosis (**Figure 4C**).

HCC-unfavorable and HCC-favorable GSVAs are biomarkers of HCC, and the HCC-unfavorable GSVAs score is an independent prognostic factor

The GSVAs package was applied to calculate the HCC-unfavorable GSVAs score and HCC-favorable GSVAs score for all samples. Obviously, the HCC-favorable GSVAs score decreased but the HCC-unfavorable GSVAs score increased with HCC progression (**Figure 5A**). ROC curve analysis indicated that both HCC-unfavorable and -favorable GSVAs scores are biomarkers for HCC, with AUC = 0.962 and AUC = 0.992, respectively (**Figure 5B**), as verified in GSE54236, with AUC = 0.679 and AUC = 0.862, respectively (**Figure 5C**). All HCC samples in TCGA were separated into low- and high-score groups according to the median GSVAs score. Both GSVAs score systems were associated with prognosis in univariate Cox proportional regression analysis. In addition, multivariate Cox proportional regression analysis indicated that the HCC-unfavorable GSVAs score was an independent prognostic factor of HCC compared to clinicopathological features (**Table 2**), with AUC = 0.704 for the 2-year tROC curve (**Figure 5D**). As expected, a high HCC-unfavorable GSVAs score was associated with poorer overall survival (**Figure 5E**), which was validated in GSE54236 (**Figure 5F**).

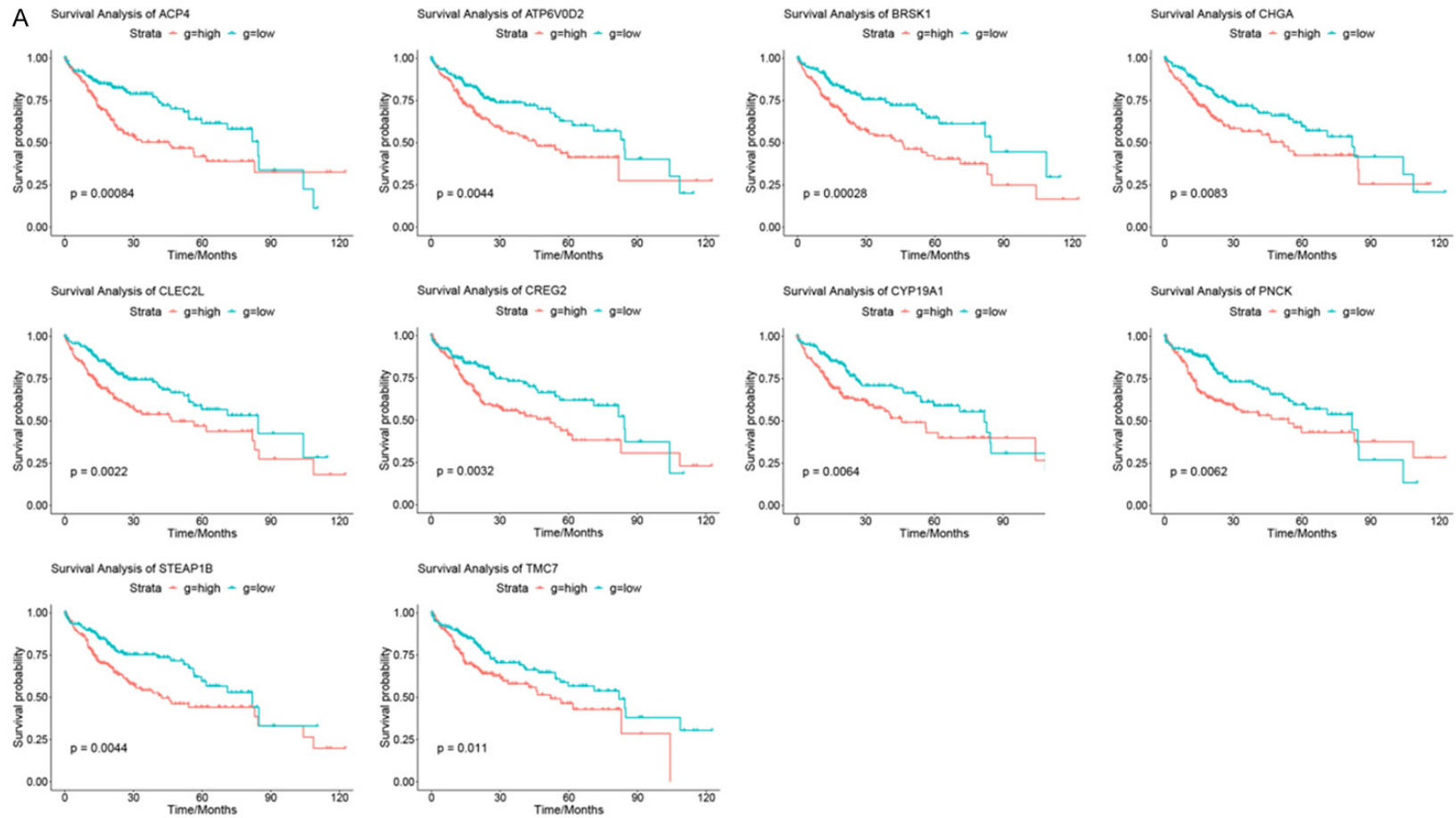
Discussion

HCC is one of the most lethal malignant tumors in the world [19]. Most HCCs are diagnosed at stages III and IV, resulting in poor prognoses. Furthermore, the pathological mechanism of HCC remains elusive, and there are no reliable biomarkers for use in the clinic to predict the survival of patients with HCC. Many previous studies have mainly focused on a single gene or molecule, without considering simultaneous changes in multiple genes [20-

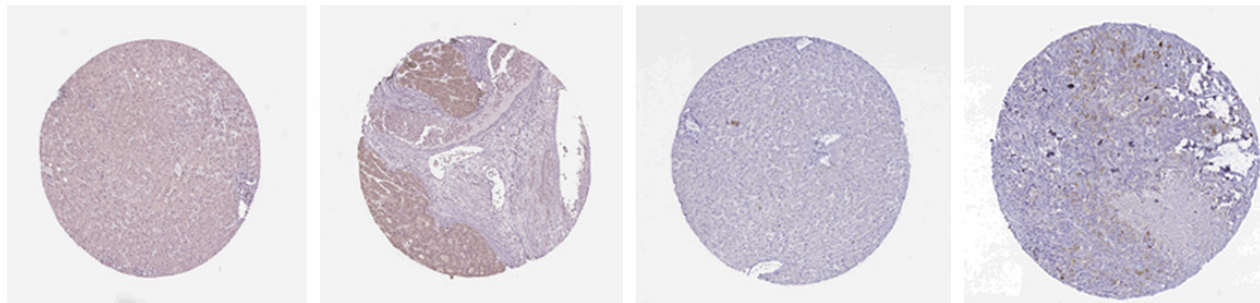
22]. In the present study, we identified 82 gradually upregulated genes and 176 gradually downregulated genes with HCC progression, which revealed that the development of HCC results from synergistic effects of multiple genes. Functional enrichment analysis indicated that the gradually upregulated genes are significantly involved in multiple pathogens of HCC-related pathways, such as viral carcinogenesis [23] and alcoholism [24]. This may indicate that expression of pathogen-related genes in HCC reflects the progression of HCC, and it thus would be crucial to eliminate pathogens in the management of HCC.

Survival analysis showed that only a few upregulated/downregulated genes are associated with prognosis. To the best of our knowledge, this is the first report of an HCC-unfavorable gene set including 10 gradually upregulated genes and an HCC-favorable gene set including 29 gradually downregulated genes. Not surprisingly, we found that some of these genes have been associated with cancer. In the HCC-unfavorable gene set, STEAP1B, TMC7, CYP19A1 and PNCK have been associated with prostate cancer, pancreatic carcinoma, and breast cancer, respectively [25-28]. In our study, we found that these genes may also be associated with HCC. ACP4, BRSK1, CHGA, ATP6 VOD2, CLEC2 L, and CREG2 may also be associated with HCC, which has been rarely reported. Many genes in the HCC-favorable gene set have been identified to be associated with HCC, such as VIPR1, CPEB3, HTR2A-AS1, ACSM3, ADRA1A, AKR1D1, BHMT, CD226, CD5 L, CYP3A4, CYP3A43, DMGDH, ESR1, GLYATL1 and RDH16. For example, low expression of VIPR1 has an adverse prognostic impact on HCC [29], loss of ACSM3 expression correlates with advanced HCC stages and poor survival [30], downregulation of BHMT in HCC is associated with poor prognosis [31], low expression of CD226 promotes proliferative, migrating, and invasive activities of HCC cells [32], and downregulation of CYP3A4 is an independent predictor of early recurrence of HCC [33]. The results of previous studies are consistent with our findings.

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma



B



BRSK1: Antibody HPA061719

CLEC2L: Antibody HPA045050

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

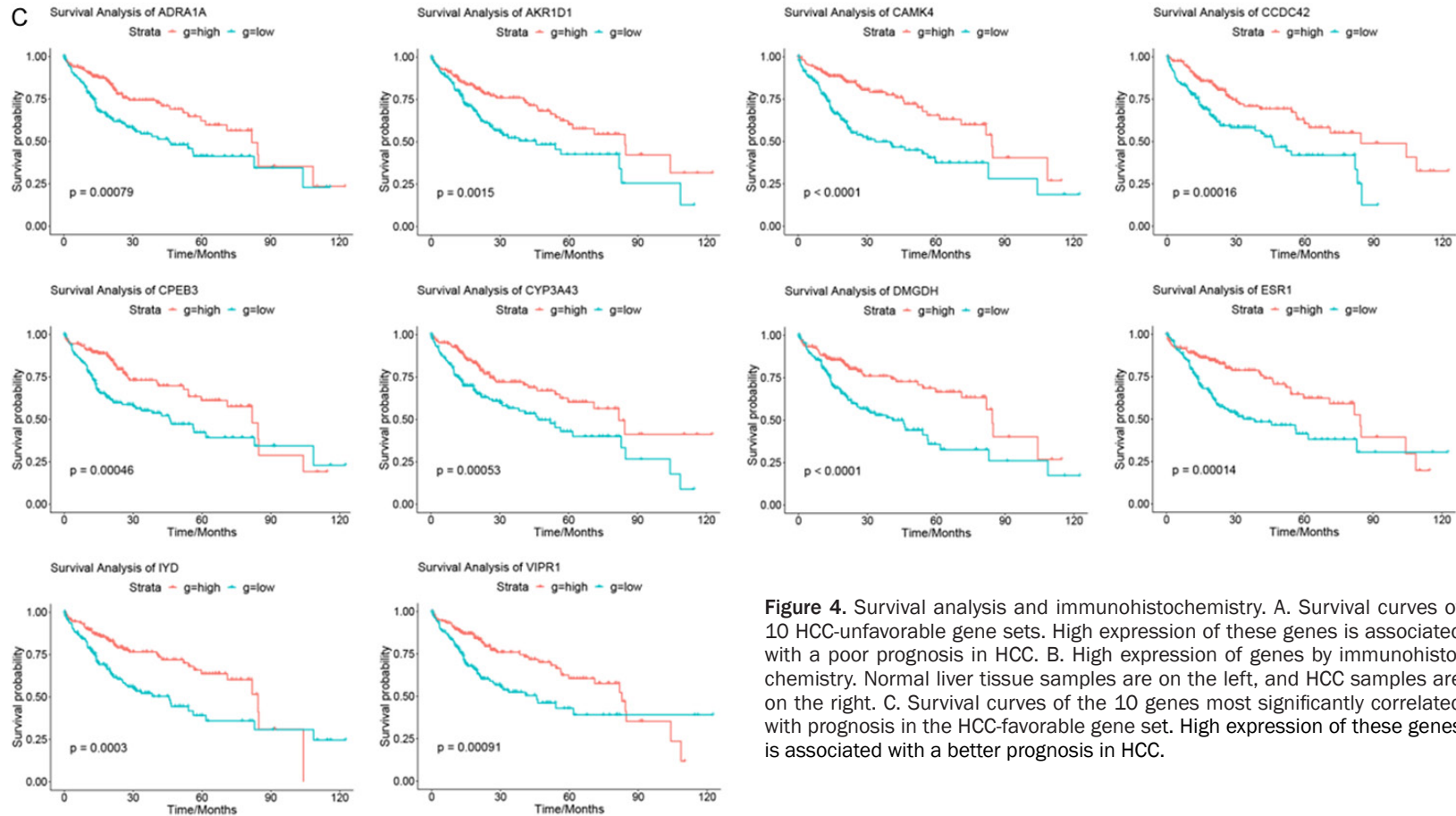
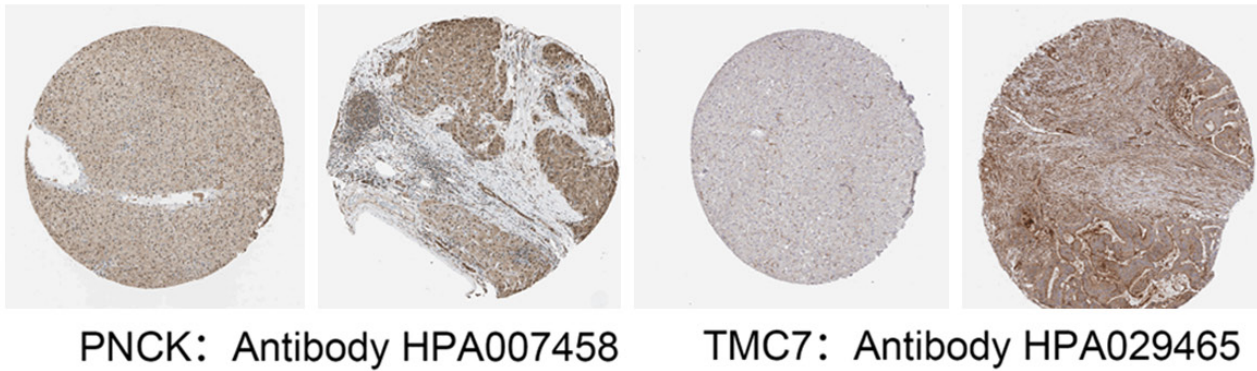


Figure 4. Survival analysis and immunohistochemistry. A. Survival curves of 10 HCC-unfavorable gene sets. High expression of these genes is associated with a poor prognosis in HCC. B. High expression of genes by immunohistochemistry. Normal liver tissue samples are on the left, and HCC samples are on the right. C. Survival curves of the 10 genes most significantly correlated with prognosis in the HCC-favorable gene set. High expression of these genes is associated with a better prognosis in HCC.

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

Table 1. The HCC-unfavorable gene set and HCC-favorable gene set

Gene set	Gene symbol
HCC-unfavorable gene set	ACP4, ATP6 V0D2, BRSK1, CHGA, CLEC2 L, CREG2, CYP19A1, PNCK, STEAP1B, TMC7
HCC-favorable gene set	ACSM3, ADRA1A, AKR1D1, BHMT, CAMK4, CCDC42, CD226, CD5 L, CLEC12A, CPEB3, CYP3A4, CYP3A43, DMGDH, ESR1, ETFDH, GHR, GLYATL1, GRAMD1C, HTR2A-AS1, IYD, LGI1, LINC00885, NDST3, NR1I2, NUGGC, RANBP3 L, RDH16, SRD5A2, VIPR1

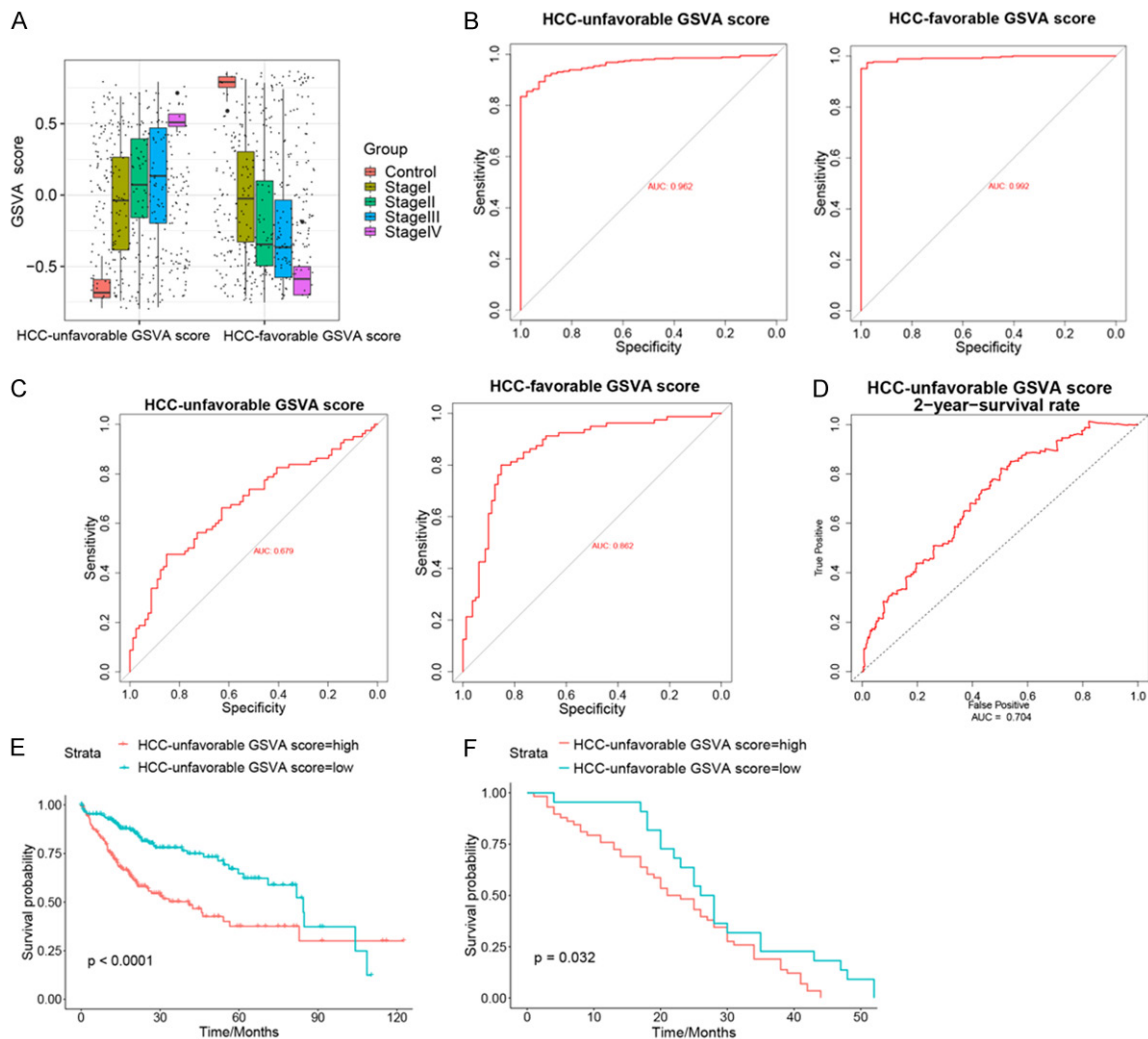


Figure 5. Evaluating the diagnostic and prognostic abilities of HCC-unfavorable and HCC-favorable GSVA scores. A. The HCC-unfavorable GSVA score gradually increased and the HCC-favorable GSVA score gradually decreased with HCC progression. B. ROC curves of HCC-unfavorable and HCC-favorable GSVA scores. C. ROC curves of HCC-unfavorable and HCC-favorable GSVA scores in GSE54236. D. tROC curves of the HCC-unfavorable GSVA score in TCGA. E. Survival analysis of the HCC-unfavorable GSVA score in the HCC cohort from TCGA. A high HCC-unfavorable GSVA score is associated with a poor prognosis in HCC. F. Survival analysis of the HCC-unfavorable GSVA score in the HCC cohort from TCGA and GSE54236. A high HCC-unfavorable GSVA score is associated with a poor prognosis in HCC.

Several gene signatures have been reported to predict prognosis in HCC [34–37]. In these studies, a coefficient for a gene in the training set

was based on Cox regression analysis or another method, and the coefficient was various. However, due to the limitations of the sample

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

Table 2. Univariate and multivariate analyses of the two GSVA scores

Factor	Univariate Cox analysis			Multivariate Cox analysis		
	β	P Value	HR (95% CI)	β	P Value	HR (95% CI)
Sex (male/female)	-0.232	0.228	0.544-1.156			
Age (> 65 years/ \leq 65 years)	0.239	0.202	0.88-1.833			
T stage (T3-4/T1-2)	0.919	0.000	1.723-3.649	0.748	0.465	0.284-15.71
Lymph node stage (N2-3/N0-1)	0.684	0.341	0.485-8.086			
Metastasis (M1/M0)	1.382	0.019	1.252-12.679	0.314	0.610	0.409-4.583
Pathological stage (III-IV/I-II)	0.911	0.000	1.711-3.613	0.418	0.683	0.204-11.281
Grade (G3-4/G1-2)	0.131	0.498	0.781-1.663			
HCC-unfavorable GSVA score (high/low)	0.820	0.000	1.553-3.32	0.582	0.036	1.039-3.086
HCC-favorable GSVA score (high/low)	-0.795	0.000	0.308-0.663	-0.508	0.072	0.346-1.047

size in previous studies as well as tumor heterogeneity, we may never obtain the real coefficient of a gene. Therefore, GSVA was used in our study to score individual samples against gene sets (HCC unfavorable and favorable). ROC curve analysis suggested that both the HCC-unfavorable GSVA score and HCC-favorable GSVA score have strong diagnostic capacity in HCC, and tROC curve analysis showed that the HCC-unfavorable GSVA score can be a prognostic biomarker. Univariate and multivariate Cox regression analyses further suggested that the HCC-unfavorable GSVA score is an independent factor for overall survival in HCC.

Tumor markers play a crucial role in the diagnosis of HCC, especially for early asymptomatic microfocal tumors. When imaging cannot be obtained, the abnormality of tumor markers plays an important role as a reference value. Various diagnostic markers, such as alpha-fetoprotein (AFP), α -L-fucosidase (AFU), desaturated- γ -carboxy-prothrombin (DCP) and phosphatidylinositol proteoglycan-3 (GPC-3), have been developed for HCC [38, 39]. AFP is currently the most widely used HCC tumor marker worldwide, but its sensitivity and specificity are not very satisfactory, especially in the early stage [40]. As AFU will increase to a certain extent in diabetes, pancreatitis, and hypothyroidism, it is not very specific for the early diagnosis of HCC, and it needs to be combined with other tumor markers to effectively detect HCC [41]. DCP is an abnormal prothrombin lacking coagulation activity detected in patients with liver cancer: it is related to the size and grade of the tumor and can be used to determine patient prognosis [42, 43].

GPC-3 is a type of heparan sulfate glycoprotein on the surface of the cell membrane that is not expressed in the normal human liver but is found at significantly higher levels in HCC patients than in those with benign liver disease [44]. Therefore, GPC-3 is helpful for the early diagnosis of HCC and differential diagnosis of benign and malignant liver tumors [45, 46]. In this study, the gene set score we developed was based on multiple genes. It can be used not only for the diagnosis of patients with early HCC but also for predicting the survival status of all HCC patients, and it is easy to obtain from peripheral blood samples.

Although we provide new insight into the HCC prognostic stratification system, several limitations were notable in the present study. First, the molecular mechanism requires experimental verification. Second, it is not clear whether the two gene sets are causal or merely markers for HCC and its prognosis. In addition, whether the gene set can distinguish other diseases, such as hepatitis or cirrhosis, remains to be further studied.

In conclusion, we identified an HCC-unfavorable gene set and an HCC-favorable gene set. The HCC-unfavorable and -favorable GSVA scores may serve as new biomarkers of HCC, and the HCC-unfavorable GSVA score is an independent biomarker for predicting prognosis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NO. 8166-0498 and 81803007), China Postdoctoral Science Foundation (NO. 2019M663412), Guangxi Key Research and Development Plan (NO.

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

GUIKEAB19245002), Guangxi Scholarship Fund of Guangxi Education Department, Guangxi Medical University Training Program for Distinguished Young Scholars, Youth Talent Fund Project of Guangxi Natural Science Foundation (NO. 2018GXNSFBA281030, 2018-GXNSFBA281091), Guangxi Medical and Health Appropriate Technology Development and Application Project (No. S2017101, S2018062) and the Project of Nanning Scientific Research and Technology Development Plan (20153012).

Disclosure of conflict of interest

None.

Address correspondence to: Zhaoshan Fang, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning 530022, Guangxi, People's Republic of China. E-mail: fangzhsh_dr@126.com; Rong Liang, Department of Medical Oncology, Guangxi Medical University Cancer Hospital, NO. 71, Hedi Road, Nanning, Guangxi, People's Republic of China. E-mail: ronglianggx@foxmail.com

References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and 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] Marengo A, Rosso C and Bugianesi E. Liver cancer: connections with obesity, fatty liver, and cirrhosis. *Annu Rev Med* 2016; 67: 103-117.
- [3] McGlynn KA, Petrick JL and London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. *Clin Liver Dis* 2015; 19: 223-238.
- [4] Davis GL, Dempster J, Meler JD, Orr DW, Walberg MW, Brown B, Berger BD, O'Connor JK and Goldstein RM. Hepatocellular carcinoma: management of an increasingly common problem. *Proc (Bayl Univ Med Cent)* 2008; 21: 266-280.
- [5] Fong ZV and Tanabe KK. The clinical management of hepatocellular carcinoma in the United States, Europe, and Asia: a comprehensive and evidence-based comparison and review. *Cancer* 2014; 120: 2824-2838.
- [6] Ma L and Deng C. Identification of a novel four-lncRNA signature as a prognostic indicator in cirrhotic hepatocellular carcinoma. *PeerJ* 2019; 7: e7413.
- [7] Zhao QJ, Zhang J, Xu L and Liu FF. Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World J Gastroenterol* 2018; 24: 3426-3439.
- [8] Sui J, Miao Y, Han J, Nan H, Shen B, Zhang X, Zhang Y, Wu Y, Wu W, Liu T, Xu S, Yang S, Yin L, Pu Y and Liang G. Systematic analyses of a novel lncRNA-associated signature as the prognostic biomarker for hepatocellular carcinoma. *Cancer Med* 2018; 7: 3240-56.
- [9] Gu JX, Zhang X, Miao RC, Xiang XH, Fu YN, Zhang JY, Liu C and Qu K. Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. *World J Gastroenterol* 2019; 25: 220-232.
- [10] Tomczak K, Czerwinska P and Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 2015; 19: A68-77.
- [11] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S and Soboleva A. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013; 41: D991-995.
- [12] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
- [13] Law CW, Chen Y, Shi W and Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15: R29.
- [14] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; 3: Article3.
- [15] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-287.
- [16] Hanzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.
- [17] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC and Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77.
- [18] Colwill K; Renewable Protein Binder Working Group, Gräslund S. A roadmap to generate renewable protein binders to the human proteome. *Nat Methods* 2011; 8: 551-558.
- [19] Bertuccio P, Turati F, Carioli G, Rodriguez T, La Vecchia C, Malvezzi M and Negri E. Global

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

- trends and predictions in hepatocellular carcinoma mortality. *J Hepatol* 2017; 67: 302-309.
- [20] Lu LL, Sun J, Lai JJ, Jiang Y, Bai LH and Zhang LD. Neuron-glia antigen 2 overexpression in hepatocellular carcinoma predicts poor prognosis. *World J Gastroenterol* 2015; 21: 6649-6659.
- [21] Wu Y, Zheng S, Yao J, Li M, Yang G, Zhang N, Zhang S and Zhong B. Decreased expression of protocadherin 20 is associated with poor prognosis in hepatocellular carcinoma. *Oncotarget* 2017; 8: 3018-3028.
- [22] Guo X, Xiong L, Zou L, Sun T, Zhang J, Li H, Peng R and Zhao J. L1 cell adhesion molecule overexpression in hepatocellular carcinoma associates with advanced tumor progression and poor patient survival. *Diagn Pathol* 2012; 7: 96.
- [23] Ringelhan M, McKeating JA and Protzer U. Viral hepatitis and liver cancer. *Philos Trans R Soc Lond B Biol Sci* 2017; 372: 20160274.
- [24] Seitz HK, Bataller R, Cortez-Pinto H, Gao B, Gual A, Lackner C, Mathurin P, Mueller S, Szabo G and Tsukamoto H. Alcoholic liver disease. *Nat Rev Dis Primers* 2018; 4: 16.
- [25] Gomes IM, Santos CR and Maia CJ. Expression of STEAP1 and STEAP1B in prostate cell lines, and the putative regulation of STEAP1 by post-transcriptional and post-translational mechanisms. *Genes Cancer* 2014; 5: 142-151.
- [26] Cheng Y, Wang K, Geng L, Sun J, Xu W, Liu D, Gong S and Zhu Y. Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma. *EBioMedicine* 2019; 40: 382-393.
- [27] Magnani L, Frige G, Gadaleta RM, Corleone G, Fabris S, Kempe MH, Verschure PJ, Barozzi I, Virchillo V, Hong SP, Perone Y, Saini M, Trumpp A, Viale G, Neri A, Ali S, Colleoni MA, Pruneri G and Minucci S. Acquired CYP19A1 amplification is an early specific mechanism of aromatase inhibitor resistance in ERalpha metastatic breast cancer. *Nat Genet* 2017; 49: 444-450.
- [28] Gardner HP, Ha SI, Reynolds C and Chodosh LA. The caM kinase, Pnck, is spatially and temporally regulated during murine mammary gland development and may identify an epithelial cell subtype involved in breast cancer. *Cancer Res* 2000; 60: 5571-5577.
- [29] Lu S, Lu H, Jin R and Mo Z. Promoter methylation and H3K27 deacetylation regulate the transcription of VIPR1 in hepatocellular carcinoma. *Biochem Biophys Res Commun* 2019; 509: 301-305.
- [30] Gopal R, Selvarasu K, Pandian PP and Ganesan K. Integrative transcriptome analysis of liver cancer profiles identifies upstream regulators and clinical significance of ACSM3 gene expression. *Cell Oncol (Dordr)* 2017; 40: 219-233.
- [31] Jin B, Gong Z, Yang N, Huang Z, Zeng S, Chen H, Hu S and Pan G. Downregulation of betaine homocysteine methyltransferase (BHMT) in hepatocellular carcinoma associates with poor prognosis. *Tumour Biol* 2016; 37: 5911-5917.
- [32] Jia B, Tan L, Jin Z, Jiao Y, Fu Y and Liu Y. MiR-892a promotes hepatocellular carcinoma cells proliferation and invasion through targeting CD226. *J Cell Biochem* 2017; 118: 1489-1496.
- [33] Ashida R, Okamura Y, Ohshima K, Kakuda Y, Uesaka K, Sugiura T, Ito T, Yamamoto Y, Sugino T, Urakami K, Kusuhara M and Yamaguchi K. CYP3A4 gene is a novel biomarker for predicting a poor prognosis in hepatocellular carcinoma. *Cancer Genomics Proteomics* 2017; 14: 445-453.
- [34] Chen P, Wang F, Feng J, Zhou R, Chang Y, Liu J and Zhao Q. Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. *Oncotarget* 2017; 8: 48948-48958.
- [35] Lu M, Kong X, Wang H, Huang G, Ye C and He Z. A novel microRNAs expression signature for hepatocellular carcinoma diagnosis and prognosis. *Oncotarget* 2017; 8: 8775-8784.
- [36] Wang Z, Wu Q, Feng S, Zhao Y and Tao C. Identification of four prognostic lncRNAs for survival prediction of patients with hepatocellular carcinoma. *PeerJ* 2017; 5: e3575.
- [37] Li B, Feng W, Luo O, Xu T, Cao Y, Wu H, Yu D and Ding Y. Development and validation of a three-gene prognostic signature for patients with hepatocellular carcinoma. *Sci Rep* 2017; 7: 5517.
- [38] Tsuchiya N, Sawada Y, Endo I, Saito K, Uemura Y and Nakatsura T. Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J Gastroenterol* 2015; 21: 10573-10583.
- [39] Wong RJ, Ahmed A and Gish RG. Elevated alpha-fetoprotein: differential diagnosis - hepatocellular carcinoma and other disorders. *Clin Liver Dis* 2015; 19: 309-323.
- [40] Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, Reddy KR, Harnois D, Llovet JM, Normolle D, Dalhgren J, Chia D, Lok AS, Wagner PD, Srivastava S and Schwartz M. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 2009; 137: 110-118.
- [41] Xing H, Qiu H, Ding X, Han J, Li Z, Wu H, Yan C, Li H, Han R, Zhang H, Li C, Wang M, Wu M, Shen F, Zheng Y and Yang T. Clinical performance of alpha-L-fucosidase for early detection of hepatocellular carcinoma. *Biomark Med* 2019; 13: 545-555.

A ten-gene set variation score as a prognostic tool in hepatocellular carcinoma

- [42] Naraki T, Kohno N, Saito H, Fujimoto Y, Ohhira M, Morita T and Kohgo Y. gamma-Carboxyglutamic acid content of hepatocellular carcinoma-associated des-gamma-carboxy prothrombin. *Biochim Biophys Acta* 2002; 1586: 287-298.
- [43] Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, Morgan TR, Kim HY, Lee WM, Bonkovsky HL and Dienstag JL; HALT-C Trial Group. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology* 2010; 138: 493-502.
- [44] Filmus J. The contribution of in vivo manipulation of gene expression to the understanding of the function of glypicans. *Glycoconj J* 2002; 19: 319-323.
- [45] Libbrecht L, Severi T, Cassiman D, Vander Borgh S, Pirenne J, Nevens F, Verslype C, van Pelt J and Roskams T. Glypican-3 expression distinguishes small hepatocellular carcinomas from cirrhosis, dysplastic nodules, and focal nodular hyperplasia-like nodules. *Am J Surg Pathol* 2006; 30: 1405-1411.
- [46] Shafizadeh N, Ferrell LD and Kakar S. Utility and limitations of glypican-3 expression for the diagnosis of hepatocellular carcinoma at both ends of the differentiation spectrum. *Mod Pathol* 2008; 21: 1011-1018.