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

Jinmin Zhao (zhaojinmin2019@163.com)
Guangxi Medical University https://orcid.org/0000-0001-6478-5394
Jiazhou Ye
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Yan Lin
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Kunpeng Bu
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Rongyun Mai
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Ziyu Liu
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Xing Gao
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Xuemin Piao
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
Rong Liang
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Research article

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

Posted Date: February 4th, 2020

DOI: https://doi.org/10.21203/rs.2.22542/v1

Abstract

We aimed to identify the progress of hepatocellular carcinoma (HCC)-specific gene set. Using the HCC data set from The Cancer Genome Atlas, we found that 10 genes were gradually up-graduated with the progress of HCC and associated with survival and classed as HCC-unfavorable gene set, while 29 genes were gradually down-graduated and associated with survival and classed as HCC-favorable gene set. Gene Set Variation Analysis (GSVA) was used to score individual samples against the two gene sets. ROC curve analysis showed both of HCC-unfavorable GSVA score and HCC-favorable GSVA score were reliable biomarkers for diagnosing HCC, tROC curve analysis and univariate/multivariate Cox proportional hazards analyses indicated that HCC-unfavorable GSVA score was an independently prognostic biomarkers. Moreover, the results were validated in an external independent data set. In addition, according to mutation and methylation analysis, we proposed that the aberrant expression of HCC-unfavorable gene may be driven by hypomethylation, not mutation.

Introduction

According to global cancer statistics of 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, type 2 diabetes[2,3]. About 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, the therapeutic effect of advanced HCC is still limited[5]. Therefore, it is essential to explore its molecular mechanism and robust diagnostic and prognostic markers.

With the development of high throughput sequencing technology, more and more molecular diagnostic markers of HCC have been identified. Most of these studies on the prognosis of HCC focused on a single or several molecules[6–9], while less attention was paid to the characteristic gene set related to HCC progress. And so far, there is no widely accepted molecular prognostic biomarker for HCC.

Hereon, we identified two HCC-progression characteristic gene sets named as HCC-unfavorable gene set and HCC-favorable gene set. Gene set variation analysis (GSVA) was used to score individual samples against the two gene sets. Both of HCC-unfavorable GSVA score and HCC-favorable GSVA score may be biomarker for HCC and HCC's prognosis.

Results

Various genes were differentially expressed with the progression of HCC

PCA analysis of TCGA data showed that the expression patterns of the global genes (Figure 2A) could not distinguish HCC from control. 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 stage IV HCCs (Figure 2E). There are 1273 common DEGs in stage I-IV HCC (Figure 2F). Among of them, 82 DEGs were gradually up-regulated and 176 DEGs were gradually down-regulated with HCC progress. And PCA analysis showed that the expression patterns of these genes could distinguish HCC from control (Figure 2G).

The gradually up-regulated/down-regulated genes involved in multiple HCC-related pathways

Functional enrichment analysis was used to explore the biological functions and related pathways of gradually up-regulated and down-regulated genes. The results of GO analysis revealed that the gradually up-regulated genes were significantly involved in negative regulation of megakaryocyte, olfactory bulb interneuron differentiation, endothelial growth factor stimulus and other biological processes (Figure 3A), while that gradually down-regulated genes are mainly involved in xenobiotic metabolic process, response to xenobiotic stimulus, cellular response to xenobiotic stimulus and other biological processes (Figure 3B). The gradually up-regulated genes significantly involved in multiple pathogen of HCC-related pathways, such as viral carcinogenesis and alcoholism (Figure 3C), while the gradually down-regulated genes significantly involved in PPAR signaling pathway, Retinol metabolism, Steroid hormone biosynthesis, Bile secretion, and ABC transporters pathways (Figure 3D).

HCC-unfavorable/favorable gene set

A total of 10 gradually up-regulated genes (ACP4, ATP6V0D2, BRSK1, CHGA, CLEC2L, CREG2, CYP19A1, PNCK, STEAP1B, TMC7) were associated with poor overall survival, classed as HCC-unfavorable gene set (Figure 4A). Moreover, 4 genes (BRSK1, CLEC2L, PNCK and TMC7) were included in The Human Protein Atlas, all of them were high expressed in HCC compared to normal liver (Figure 4B) which is consistent with our findings. Twenty-nine gradually down-regulated genes were associated with good prognosis, classed as the HCC-favorable gene set (Table 1). CAMK4, DMGDH, IYD, CCDC42, ESR1, CPEB3, CYP3A43, VIPR1, AKR1D1 and ADRA1A were the ten genes with the most significant association with good prognosis (Figure 4C).

HCC-unfavorable GSVA score and HCC-favorable GSVA score are biomarkers of HCC and HCC-unfavorable GSVA score was an independent prognostic factor

The GSVA package was applied to calculate HCC-unfavorable GSVA scores and HCC-favorable GSVA score for all samples. Obviously, HCC-favorable GSVA score was decreasing, while HCC-unfavorable GSVA score was increasing with HCC progress (Figure 5A). ROC curve analysis indicated that both HCC-unfavorable GSVA score and HCC-favorable GSVA score are biomarker of HCC with AUC = 0.962 and AUC = 0.992, respectively (Figure 5B), which was verified in GSE54236 with AUC = 0.679 and AUC = 0.862 (Figure 5C). All HCCs in TCGA were separated into low- and high-score groups according to the median GSVA score. Both the GSVA score system were associated with prognosis in univariate Cox proportional regression analysis, moreover, the multivariate Cox proportional regression analysis indicated the HCC-unfavorable GSVA score was an independent prognostic factor of HCC compared to clinicopathological features (Table 2) and with AUC = 0.704 in the 2-year tROC curve (Figure 6D). As we expected, a high HCC-unfavorable GSVA score was associated with a poorer overall survival (Figure 6E), and that was validated in GSE54236. (Figure 6F)

The aberrant expression of HCC-unfavorable genes may result from hypomethylation

Only 16 (4.4%) of 364 samples had a alteration in one or several HCC-unfavorable genes and most samples did not have genetical alteration (Fig.6A). Comparison to normal liver tissue, HCC-unfavorable genes in multiple CpG islands revealed significantly (Fig.6B) low methylation level, especially ACPT, ATP6V0D2, CREG2, CYP19A1 and CLEC2L. Thus, the aberrant expression of HCC-unfavorable genes may result from hypomethylation. Among them, HTR2A-AS1 is a non-coding RNA gene, and not available in the wanderer.

Discussion

HCC is one of the most lethal malignant tumors in the world[10]. Most HCCs were diagnosed at stage III and IV, resulting in poor prognosis. The pathological mechanism of HCC is still elusive, and so far, there has been no reliable biomarker used in clinic to predict the survival of patients with HCC. Many previous studies are mainly focused on a single gene or molecule, and did not take the simultaneous changes of multiple genes into account[11–13]. In the present study, we identified 82 gradually up-regulated genes and 176 gradually down-regulated genes with HCC-progression, which revealed the development of HCC results from synergistic effects of multiple genes. Function enrichment analysis indicated that the gradually up-regulated genes significantly involved in multiple pathogen of HCC-related pathways, such as viral carcinogenesis[14] and alcoholism[15]. This may indicate that the expression of pathogen-related gene of HCC may reflect the progress of HCC and it was crucial to eliminate the pathogens in the management of HCC.

Survival analysis showed only a few up-regulated/down-regulated genes are associated with the prognosis. It may be the first time, as our best knowledge, a HCC-unfavorable gene set including 10 gradually up-regulated genes and a HCC-favorable gene set including 29 gradually down-regulated genes

was collected. Not surprisingly, we found that some of these genes were reported and associated with cancer. In HCC-unfavorable gene set, STEAP1B, TMC7, CYP19A1 and PNCK associated with prostate cancer, pancreatic carcinoma, breast cancer, respectively[16–19]. Here, we found them may also be associated with HCC. Moreover, genes ACP4, BRSK1, CHGA, ATP6V0D2, CLEC2L, and CREG2 may be associated with HCC and this was few reported so far. While in HCC-favorable gene set, many genes have been identified to be associated with HCC, such as VIPR1, CPEB3, HTR2A-AS1, ACSM3, ADRA1A, AKR1D1, BHMT, CD226, CD5L, CYP3A4, CYP3A43, DMGDH, ESR1, GLYATL1 and RDH16. Low expression of VIPR1 had an adverse prognostic impact on HCC[20], loss of ACSM3 expression was found to correlate with advanced HCC stages and a poor survival[21], down-regulation of BHMT in HCC associates with poor prognosis[22], low-expressed of CD226 could promote proliferative, migrating, and invasive activities of HCC cells[23], down-regulation of CYP3A4 gene is an independent predictor of early recurrence of HCC[24]. The results of the previous study were consistent with our findings.

Previously, several gene signatures were reported to predict prognosis in HCC[25–28]. In these studies, a gene often got a coefficient from a Cox regression analysis or other method in the training set and the coefficient was various. However, due to the limitations of the sample size in the previous studies and the tumor heterogeneity, we may never get the real coefficient of a gene. Therefore, GSVA was used to score individual samples against gene sets (HCC-unfavorable gene set and HCC-favorable gene set) in our study. ROC curve analysis suggested that both HCC-unfavorable GSVA score and HCC-favorable GSVA score exhibited strong diagnostic capacity of HCC. And tROC curve analysis showed that HCC-unfavorable GSVA score can be a prognostic biomarker. Univariate and multivariate Cox regression analysis suggested that HCC-unfavorable GSVA score was an independent factor for HCC's overall survival.

Moreover, through genetical alteration and methylation analysis, we found few samples had a alteration in one or several HCC-unfavorable genes. Comparison to normal liver tissue, HCC-unfavorable genes in multiple CpG islands revealed significantly low methylation level. Thus, the aberrant expression of HCC-unfavorable genes may result from hypomethylation, not mutation. This indicated that the pathological process of HCC may be more relevant to epigenetic than genetic mutations[29].

Although we provided new insights into the HCC prognostic stratification system, several limitations were notable in the present study. Firstly, the molecular mechanism requires experimental verification. Secondly, it is not clear whether the two gene sets are causal or merely markers for HCC and its prognosis.

Conclusion

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

Materials And Methods

Materials Acquiring

In The Cancer Genome Atlas (TCGA, https://www.cancer.gov/)[30], there are 171 HCCs with stage I, 86 HCCs stage II, 83 HCCs with stage III, 5 HCCs with stage IV and 42 healthy liver tissue samples. In addition, GSE54236 based on GPL6480 platform was download from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/)[31], included 81 HCC samples and 80 healthy liver tissue samples. GSE54236 was used to verify the prognostic value. The “normalizeBetweenArrays” function in the limma package[32] was used to normalize the gene expression profiles. If a gene responds to a multiple probes, the average value of these probes is considered to be the expression value of the corresponding gene. The workflow of the present study was shown in Figure 1.

Differentially expressed gene (DEG) analysis

The RNA sequencing expression profile (displayed as read counts) of HCC in TCGA was download. The voom function[33] in limma package was used to normalized the RNA sequencing data and limma package[32,34] was used to identify DEG of 4 stage HCC and healthy liver tissue samples respectively. DEG was set as $P < 0.01$ after FDR correction and $|logFC| > 1.5$ as the threshold. In the progress of HCC, if a DEG was gradually up-regulated (logFCstage I vs control < logFCstage II vs control < logFCstage III vs control < logFCstage IV vs control) or gradually down-regulated (logFCstage I vs control > logFCstage II vs control > logFCstage III vs control > logFCstage IV vs control), then it was considered to be HCC-progression characteristic gene.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the gradually up-regulated and gradually down-regulated genes were performed respectively using the clusterProfiler package[35] in R. $P < 0.05$ was considered significant.

Survival analysis and HCC- unfavorable/ favorable gene set

The median expression value of each gradually up-regulated and gradually down-regulated gene was used as a threshold value to divide patients into high- and low-expression groups. We applied for a Kaplan–Meier survival analysis with the log-rank method to evaluate the association with a gene and prognosis. Survival analysis was performed using survival package (https://CRAN.R-project.org/package = survival) in R. $P < 0.01$ was considered to be significant. A gene was gradually up-regulated in HCC progress and associated with poor prognosis is defined as HCC-unfavorable gene, while a gene was gradually down-regulated in HCC progress and associated with a good prognosis is defined as HCC-

favorable gene. Subsequently, two HCC-progression characteristic gene sets were collected, including HCC-unfavorable gene set and HCC-favorable gene set.

Calculation of HCC-unfavorable/favorable GSVA score

The GSVA package implements a non-parametric unsupervised method, called Gene Set Variation Analysis (GSVA), for assessing gene set enrichment (GSE) in gene expression micro array data or RNA-seq data. GSVA package[36] in R was used to calculate HCC-unfavorable GSVA score and HCC-favorable GSVA score 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[37] was used to conduct ROC curve analysis evaluate their ability 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. The tROC curve analysis was used to evaluate the prognostic value for the 2-year survival rate of the independent prognostic factors.

Validate the GSVA score system in an independent data set

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

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

The Human Protein Atlas(https://v15.proteinatlas.org/)[38] can provide information on the tissue and cell distribution of all 24,000 human proteins. We scanned The Human Protein Atlas web tool to validate the differential expression of the HCC-unfavorable genes at the protein level.

Mutation and methylation analysis

In order to explore the potential mechanism of differential expression of HCC-unfavorable genes, we scanned the mutation and methylation of these genes. The TCGAbiolinks package[39] was used to download and scan the alteration statuses of HCC-unfavorable genes. At the same time, we explored DNA methylation of these genes using Wanderer (http://maplab.imppc.org/wanderer/)[40], which is an

intuitive network tool that can be used to retrieve DNA methylation and gene expression in different tumor types in TCGA databases.

Declarations

Authors' contributions

JZ and RL designed the study. YL, JY, and KB drafted the manuscript; ZL, XG, XP and RM performed the data All authors contributed to the writing and approved the final version.

Funding

This research was funded by Regional science fund project of China natural science foundation (NO.81660498); Youth talent fund project of China natural science foundation (NO. 81803007); Youth talent fund project of GuangXi natural science foundation (NO. 2016GXNSFBA380090, NO.2018GXNSFBA281030, NO.2018GXNSFBA281091); Guangxi Medical and Health Appropriate Technology Development and Application Project (NO.S2017101, NO.S2018062); Guangxi Scholarship Fund of Guangxi Education Department; Tianqing Liver Diseases Research Fund (NO. TQGB20200192); The China Postdoctoral Science Foundation (No 2019M663412).

Availability of data and materials

We did not use new software, databases, or applications/tools in the manuscript, all results and figures have already provided in the manuscript.

Competing interests

The authors declare that they have no competing interests.

Contributor Information

Jiazhou Ye, Email: Yejiazhou2019@163.comYan Lin, Email: linyanmgx@163.comKunpeng Bu, Email: bukunpeng2008@163.comRongyun Mai, Email: rongyunmai@163.comZiyu Liu, Email: lzy80852@foxmail.comXing Gao, Email: xinggaogx@163.comXuemin Piao, Email: Piaoxuemin@outlook.comRong liang, Email: ronglianggx@126.comJinmin Zhao, Email: Zhaojinmin2019@163.com

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(6):394–424.

[2]Marengo A, Rosso C, Bugianesi E. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. Annu Rev Med. 2016;67:103–17.

[3]McGlynn KA, Petrick JL, London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. Clin Liver Dis. 2015;19(2):223–38.

[4]Davis GL, Dempster J, Meler JD, Orr DW, Walberg MW, Brown B, et al. Hepatocellular carcinoma: management of an increasingly common problem. Proc (Bayl Univ Med Cent). 2008;21(3):266–80.

[5]Fong ZV, 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(18):2824–38.

[6]Ma L, 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, 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(30):3426–39.

[8]Sui J, Miao Y, Han J, Nan H, Shen B, Zhang X, et al. Systematic analyses of a novel lncRNA-associated signature as the prognostic biomarker for Hepatocellular Carcinoma. Cancer Med. 2018.

[9]Gu JX, Zhang X, Miao RC, Xiang XH, Fu YN, Zhang JY, et al. Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. World J Gastroenterol. 2019;25(2):220–32.

[10]Bertuccio P, Turati F, Carioli G, Rodriguez T, La Vecchia C, Malvezzi M, et al. Global trends and predictions in hepatocellular carcinoma mortality. J Hepatol. 2017;67(2):302–9.

[11]Lu LL, Sun J, Lai JJ, Jiang Y, Bai LH, Zhang LD. Neuron-glial antigen 2 overexpression in hepatocellular carcinoma predicts poor prognosis. World J Gastroenterol. 2015;21(21):6649–59.

[12]Wu Y, Zheng S, Yao J, Li M, Yang G, Zhang N, et al. Decreased expression of protocadherin 20 is associated with poor prognosis in hepatocellular carcinoma. Oncotarget. 2017;8(2):3018–28.

[13]Guo X, Xiong L, Zou L, Sun T, Zhang J, Li H, et al. L1 cell adhesion molecule overexpression in hepatocellular carcinoma associates with advanced tumor progression and poor patient survival. Diagn Pathol. 2012;7:96.

[14]Ringelhan M, McKeating JA, Protzer U. Viral hepatitis and liver cancer. Philos Trans R Soc Lond B Biol Sci. 2017;372(1732).

[15]Seitz HK, Bataller R, Cortez-Pinto H, Gao B, Gual A, Lackner C, et al. Alcoholic liver disease. Nat Rev Dis Primers. 2018;4(1):16.

[16]Gomes IM, Santos CR, 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(3–4):142–51.

[17]Cheng Y, Wang K, Geng L, Sun J, Xu W, Liu D, et al. Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma. EBioMedicine. 2019;40:382–93.

[18]Magnani L, Frige G, Gadaleta RM, Corleone G, Fabris S, Kempe MH, et al. Acquired CYP19A1 amplification is an early specific mechanism of aromatase inhibitor resistance in ERα metastatic breast cancer. Nat Genet. 2017;49(3):444–50.

[19]Gardner HP, Ha SI, Reynolds C, 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(19):5571–7.

[20]Lu S, Lu H, Jin R, Mo Z. Promoter methylation and H3K27 deacetylation regulate the transcription of VIPR1 in hepatocellular carcinoma. Biochem Biophys Res Commun. 2019;509(1):301–5.

[21]Gopal R, Selvarasu K, Pandian PP, Ganesan K. Integrative transcriptome analysis of liver cancer profiles identifies upstream regulators and clinical significance of ACSM3 gene expression. Cell Oncol (Dordr). 2017;40(3):219–33.

[22]Jin B, Gong Z, Yang N, Huang Z, Zeng S, Chen H, et al. Downregulation of betaine homocysteine methyltransferase (BHMT) in hepatocellular carcinoma associates with poor prognosis. Tumour Biol. 2016;37(5):5911–7.

[23]Jia B, Tan L, Jin Z, Jiao Y, Fu Y, Liu Y. MiR–892a Promotes Hepatocellular Carcinoma Cells Proliferation and Invasion Through Targeting CD226. J Cell Biochem. 2017;118(6):1489–96.

[24]Ashida R, Okamura Y, Ohshima K, Kakuda Y, Uesaka K, Sugiura T, et al. CYP3A4 Gene Is a Novel Biomarker for Predicting a Poor Prognosis in Hepatocellular Carcinoma. Cancer Genomics Proteomics. 2017;14(6):445–53.

[25]Chen P, Wang F, Feng J, Zhou R, Chang Y, Liu J, et al. Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. Oncotarget. 2017;8(30):48948–58.

[26]Lu M, Kong X, Wang H, Huang G, Ye C, He Z. A novel microRNAs expression signature for hepatocellular carcinoma diagnosis and prognosis. Oncotarget. 2017;8(5):8775–84.

[27]Wang Z, Wu Q, Feng S, Zhao Y, Tao C. Identification of four prognostic LncRNAs for survival prediction of patients with hepatocellular carcinoma. PeerJ. 2017;5:e3575.

[28]Li B, Feng W, Luo O, Xu T, Cao Y, Wu H, et al. Development and Validation of a Three-gene Prognostic Signature for Patients with Hepatocellular Carcinoma. Sci Rep. 2017;7(1):5517.

[29]Wahid B, Ali A, Rafique S, Idrees M. New Insights into the Epigenetics of Hepatocellular Carcinoma. Biomed Res Int. 2017;2017:1609575.

[30]Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68–77.

[31]Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41(Database issue):D991–5.

[32]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(7):e47.

[33]Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.

[34]Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3.

[35]Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.

[36]Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.

[37]Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

[38]Colwill K, Renewable Protein Binder Working Group, Gräslund S. A roadmap to generate renewable protein binders to the human proteome. Nat Methods. 2011;8(7):551–8.

[39]Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS Comput Biol. 2019;15(3):e1006701.

[40]Díez-Villanueva A, Mallona I, Peinado MA. Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. Epigenetics Chromatin. 2015;8:22.

Tables

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

Gene set	Gene symbol
HCC-unfavorable gene set	ACP4, ATP6V0D2, BRSK1, CHGA, CLEC2L, CREG2, CYP19A1, PNCK, STEAP1B, TMC7
HCC-favorable gene set	ACSM3, ADRA1A, AKR1D1, BHMT, CAMK4, CCDC42, CD226, CD5L, CLEC12A, CPEB3, CYP3A4, CYP3A43, DMGDH, ESR1, ETFDH, GHR, GLYATL1, GRAMD1C, HTR2A-AS1, IYD, LGI1, LINC00885, NDST3,NR1I2, NUGGC, RANBP3L, RDH16, SRD5A2,VIPR1

Table 2 Univariate and multivariate analyses of two GSVA score

Factor	Univariate Cox analysis			Multivariate Cox analysis		
	β	P.Value	HR(95% CI)	β	P.Value	HR(95% CI)
T stage(T3-4 / T1-2)	0.919	0.000	1.723-3.649	0.748	0.465	0.284-15.71
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
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
Lymph node stage(N2-3 / N0-1)	0.684	0.341	0.485-8.086			
Grade(G3-4 / G1-2)	0.131	0.498	0.781-1.663			
Age(>65 years / <=65 years)	0.239	0.202	0.88-1.833			
Gender(male / female)	-0.232	0.228	0.544-1.156			

Figures

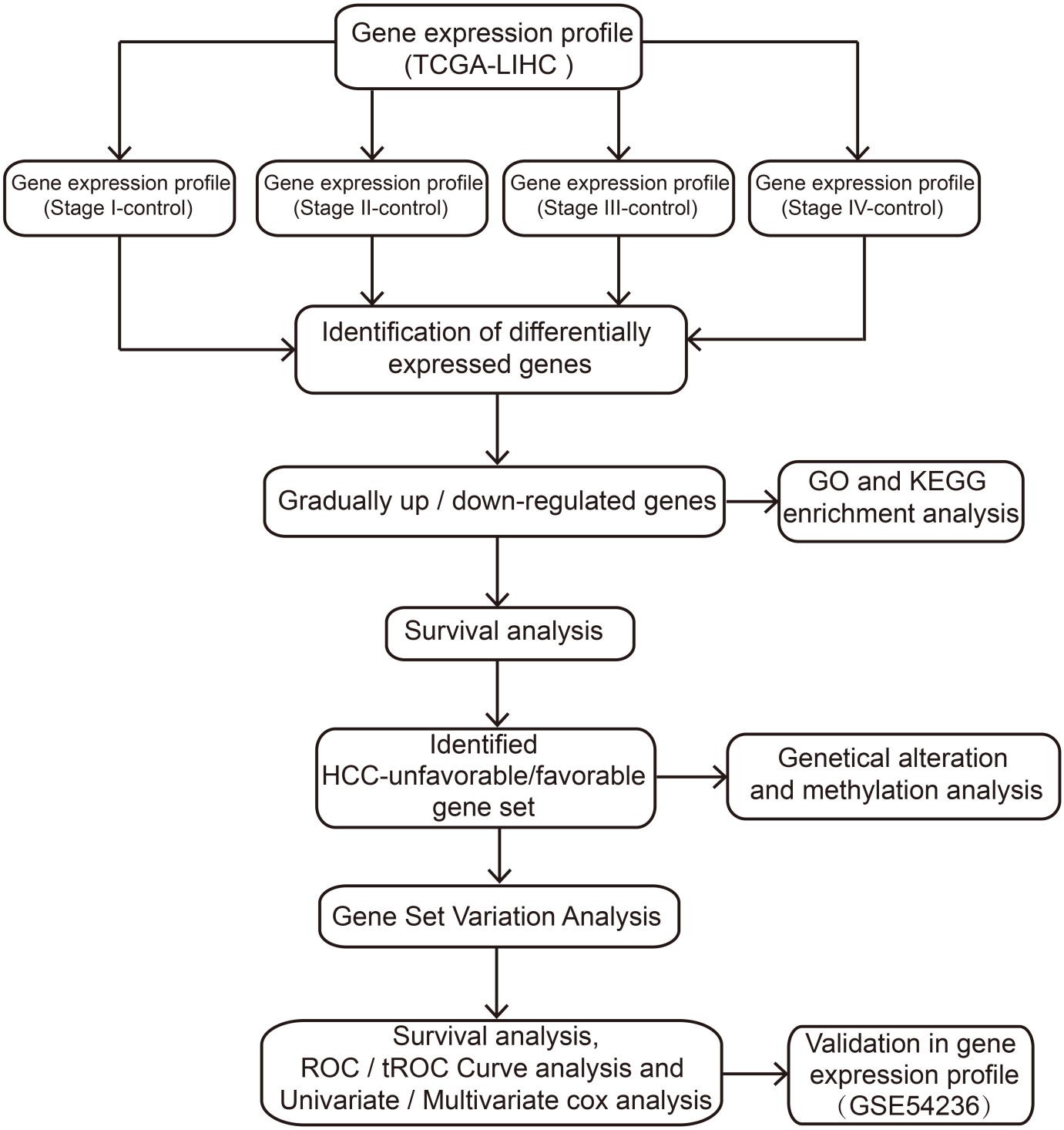

Figure 1

Flow Chart of this study.

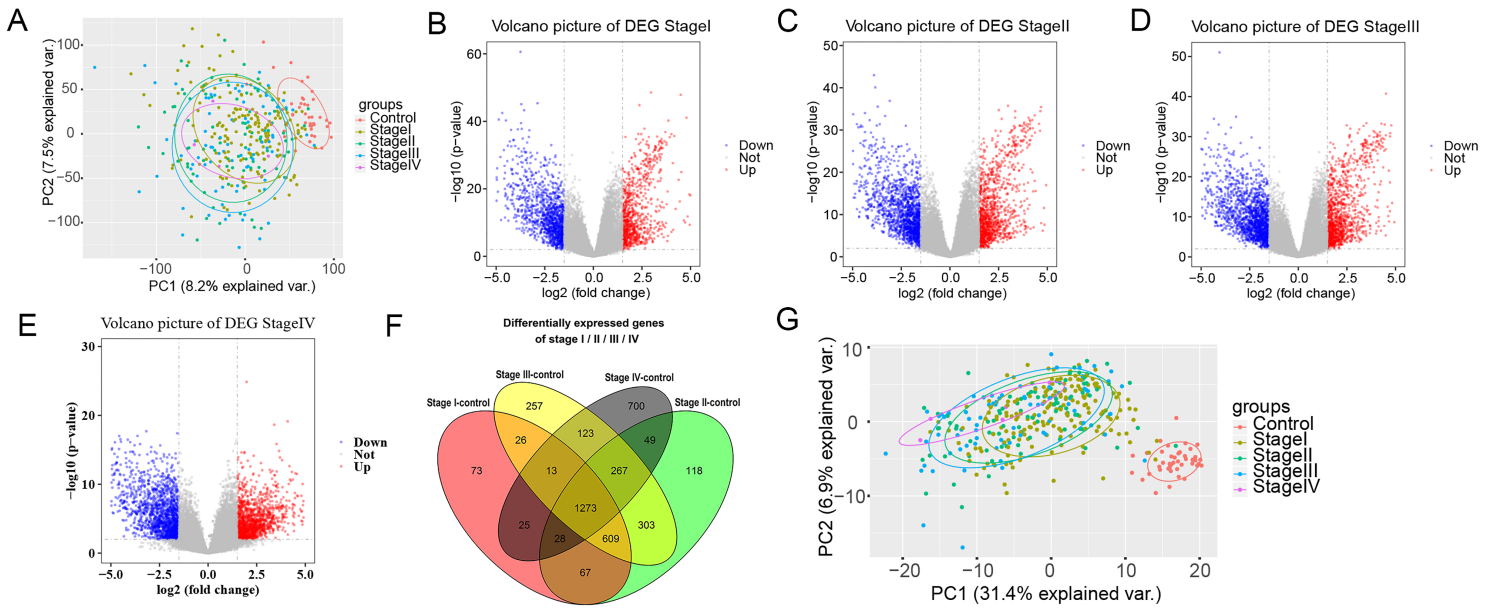

Figure 2

Differential expression gene (DEG) analysis and principal component analysis (PCA). (A) PCA of TCGA's HCC gene expression profile. (B) Volcano plot of differentially expressed gene between HCC with stage I and normal liver tissue. (C) Volcano plot of differentially expressed gene between HCC with stage II and normal liver tissue. (D) Volcano plot of differentially expressed gene between HCC with stage III and normal liver tissue. (E) Volcano plot of differentially expressed gene between HCC with stage IV and normal liver tissue. red represents up-regulated genes, blue represents down-regulated genes, and gray represents no significantly differentially expressed genes. (F) Common DEGs in HCC stage I-IV. (G) PCA of gradually up-regulated and down-regulated genes.

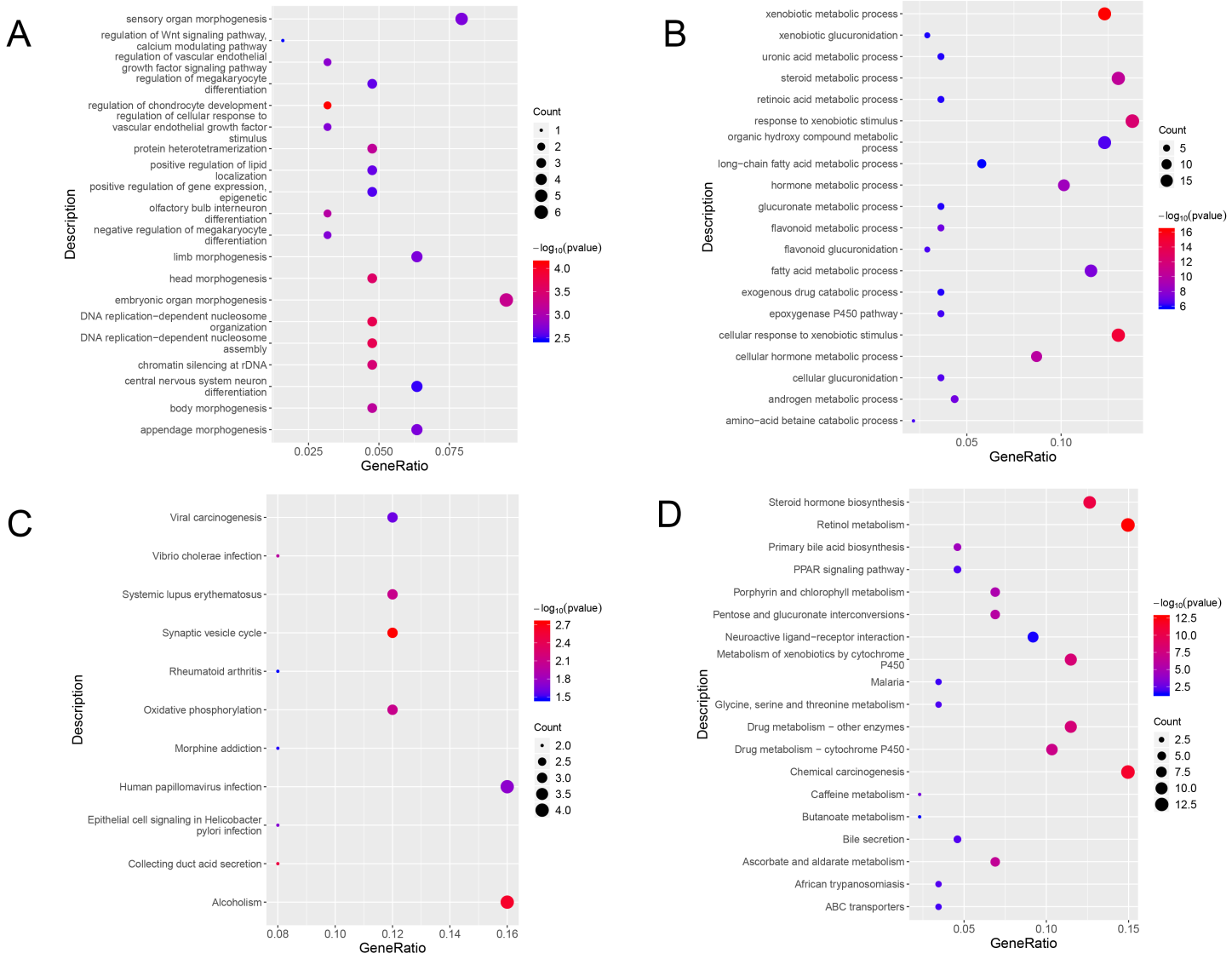

Figure 3

Biological processes and KEGG pathways enrichment analysis of gradually up-regulated/down-regulated genes. (A) Biological process of gradually up-regulated genes. (B) Biological process of gradually down-regulated genes. (C) KEGG pathway analysis of gradually up-regulated genes. (D) KEGG pathway analysis of gradually down-regulated genes.

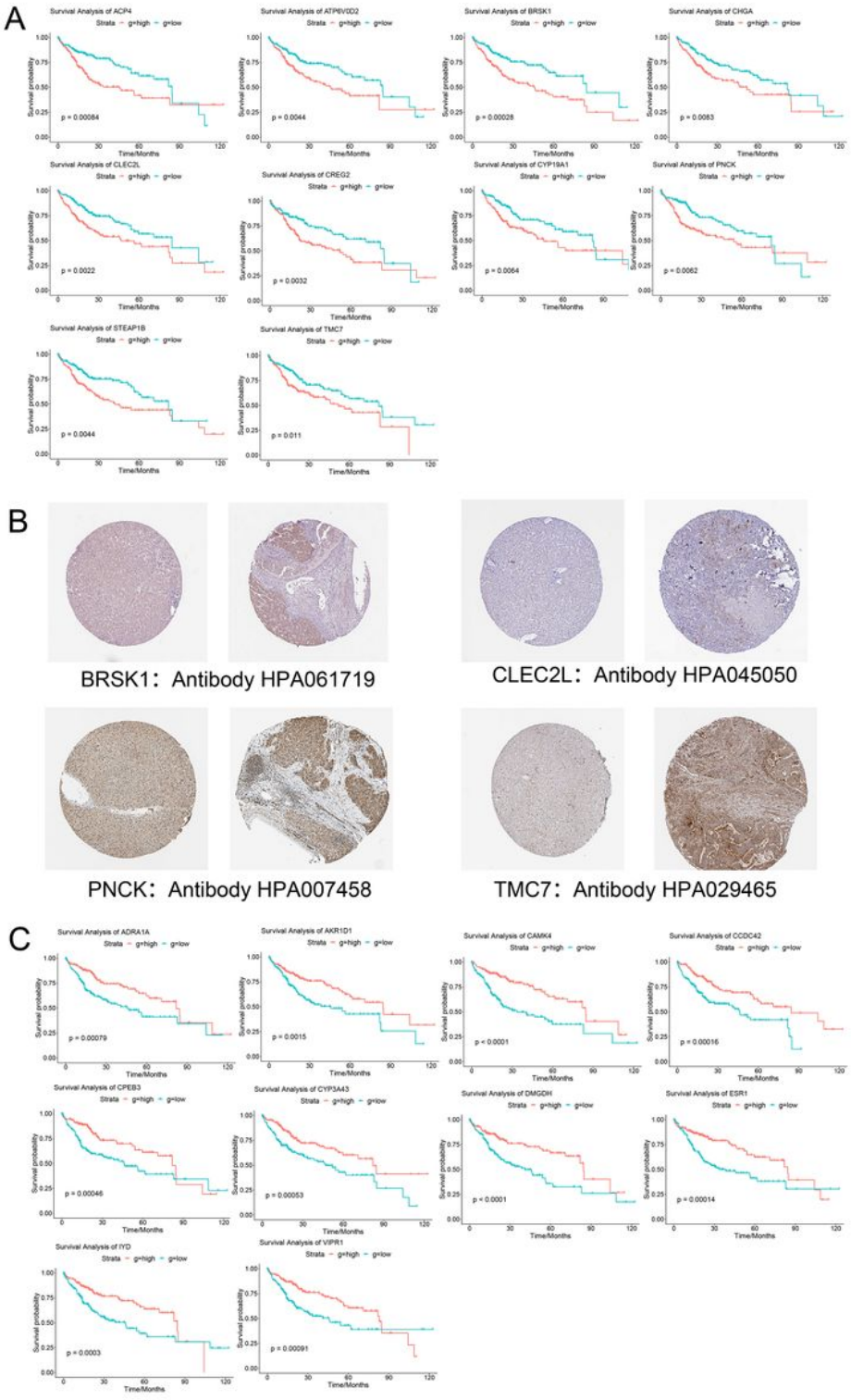

Figure 4

Survival analysis and immunohistochemistry. (A) Survival curves of 10 HCC-unfavorable gene set. (B) High expression of genes in immunohistochemistry, normal liver tissue samples are on the left and HCC samples are on the right. (C) Survival curves of 10 genes most significantly correlated with prognosis in HCC-favorable gene set.

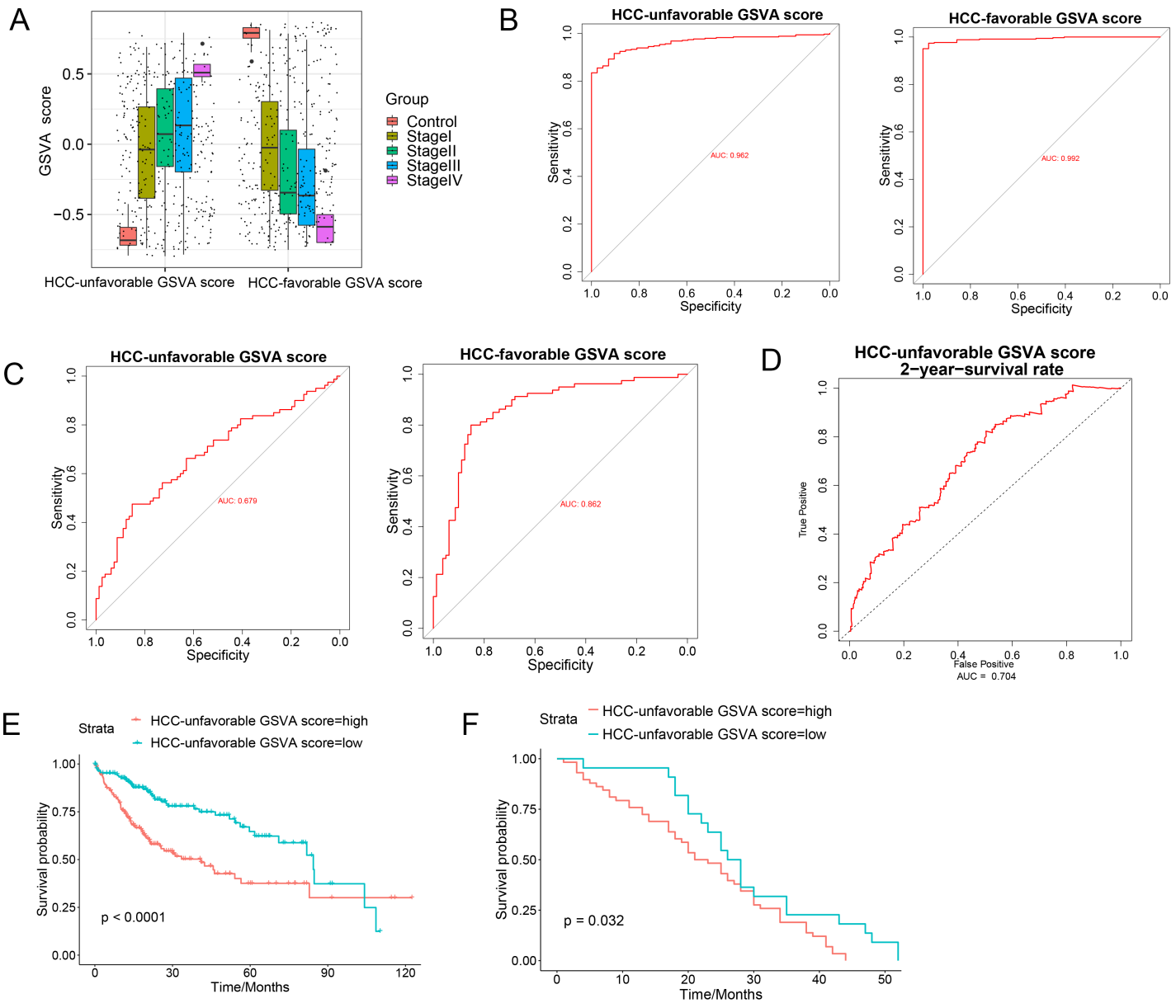

Figure 5

Evaluating the diagnostic and prognostic abilities of HCC-unfavorable GSVA score and HCC-favorable GSVA score. (A) HCC-unfavorable GSVA score was gradually increased and HCC-favorable GSVA score was gradually reduced with HCC progress. (B) ROC curves of HCC-unfavorable GSVA score and HCC-favorable GSVA score. (C) ROC curves of HCC-unfavorable GSVA score and HCC-favorable GSVA score in GSE54236. (D) tROC curves of HCC-unfavorable GSVA score in TCGA. (E) Survival analysis of HCC-unfavorable GSVA score in HCC from TCGA. (F) Survival analysis of HCC-unfavorable GSVA score in HCC from TCGA and GSE54236.

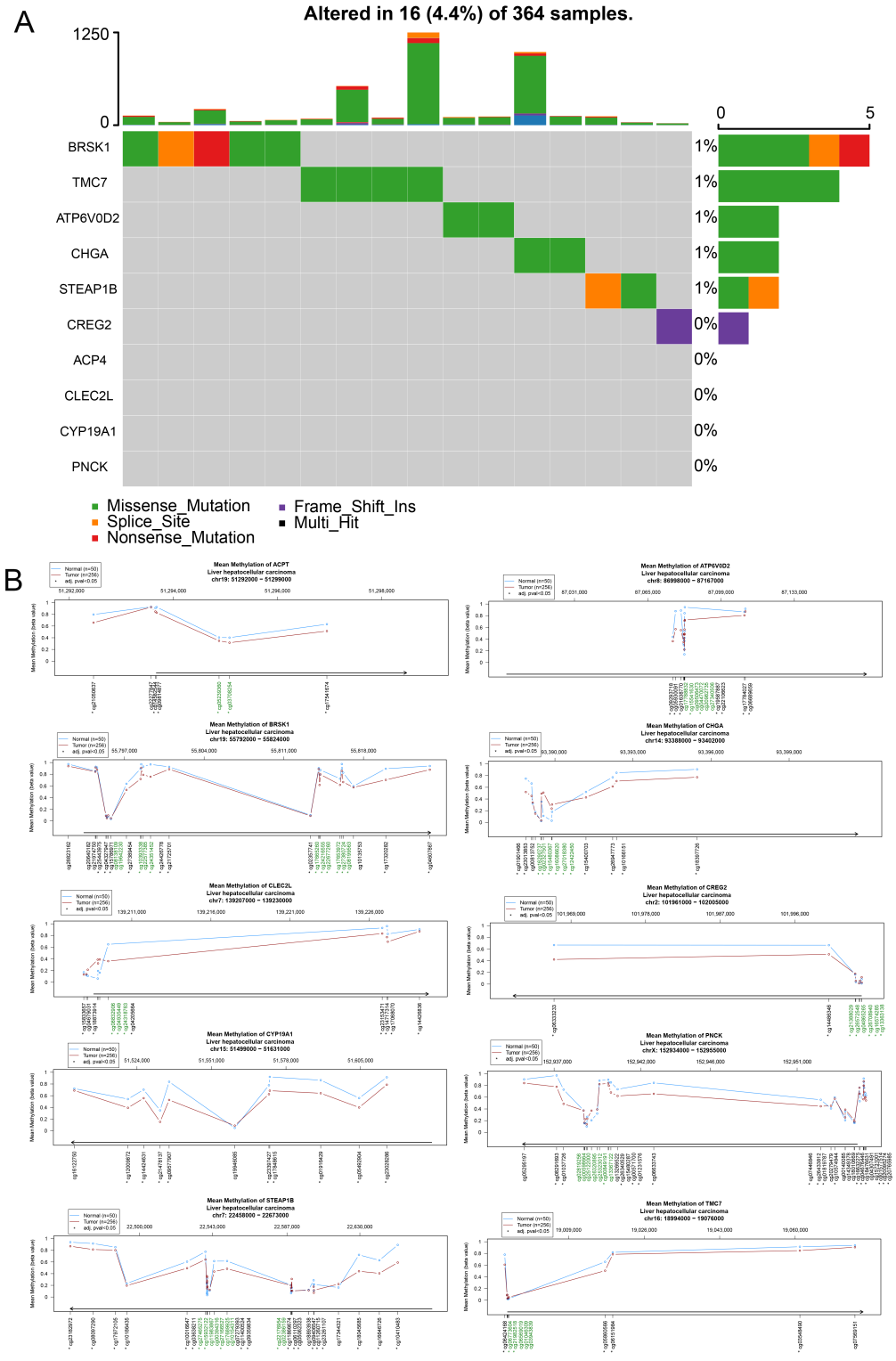

Figure 6

Genetical alteration and methylation analysis of HCC-unfavorable gene set. (A) Genetical alteration analysis of HCC-unfavorable genes. (B) Methylation analysis of HCC-unfavorable gene set.