



# Construction and Validation of Two Hepatocellular Carcinoma-Progression Prognostic Scores Based on Gene Set Variation Analysis

Qifan He, Baorui Fan, Peng Du and Yonghai Jin\*

Department of Interventional Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China

**Background:** Liver hepatocellular carcinoma (LIHC) remains a global health challenge with a low early diagnosis rate and high mortality. Therefore, finding new biomarkers for diagnosis and prognosis is still one of the current research priorities.

**Methods:** Based on the variation of gene expression patterns in different stages, the LIHC-development genes (LDGs) were identified by differential expression analysis. Then, prognosis-related LDGs were screened out to construct the LIHC-unfavorable gene set (LUGs) and LIHC-favorable gene set (LFGs). Gene set variation analysis (GSVA) was conducted to build prognostic scoring models based on the LUGs and LFGs. ROC curve analysis and univariate and multivariate Cox regression analysis were carried out to verify the diagnostic and prognostic utility of the two GSVA scores in two independent datasets. Additionally, the key LCGs were identified by the intersection analysis of the PPI network and univariate Cox regression and further evaluated their performance in expression level and prognosis prediction. Single-sample GSEA (ssGSEA) was performed to understand the correlation between the two GSVA enrichment scores and immune activity.

**Result:** With the development of LIHC, 83 LDGs were gradually upregulated and 247 LDGs were gradually downregulated. Combining with LIHC survival analysis, 31 LUGs and 32 LFGs were identified and used to establish the LIHC-unfavorable GSVA score (LUG score) and LIHC-favorable GSVA score (LFG score). ROC curve analysis and univariate/multivariate Cox regression analysis suggested the LUG score and LFG score could be great indicators for the early diagnosis and prognosis prediction. Four genes (ESR1, EHHADH, CYP3A4, and ACADL) were considered as the key LCGs and closely related to good prognosis. The frequency of TP53 mutation and copy number variation (CNV) were high in some LCGs. Low-LFG score patients have active metabolic activity and a more robust immune response. The high-LFG score patients characterized immune activation with the higher infiltration abundance of type I T helper cells, DC, eosinophils, and neutrophils, while the high-LUG score patients characterized immunosuppression with the higher infiltration abundance of type II T helper cells, TRegs, and iDC. The high- and

## OPEN ACCESS

### Edited by:

Jiang Chen,  
Zhejiang University, China

### Reviewed by:

Xinwei Han,  
Zhengzhou University, China  
Yilin Pang,  
Wenzhou Medical University, China

### \*Correspondence:

Yonghai Jin  
jinyonghai\_dc@163.com

### Specialty section:

This article was submitted to  
Molecular and Cellular Oncology,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

**Received:** 01 November 2021

**Accepted:** 01 February 2022

**Published:** 09 March 2022

### Citation:

He Q, Fan B, Du P and Jin Y (2022)  
Construction and Validation of Two  
Hepatocellular Carcinoma-  
Progression Prognostic Scores Based  
on Gene Set Variation Analysis.  
*Front. Cell Dev. Biol.* 10:806989.  
doi: 10.3389/fcell.2022.806989

low-LFG score groups differed significantly in immunotherapy response scores, immune checkpoints expression, and IC50 values of common drugs.

**Conclusion:** Overall, the LIHC-progression characteristic genes can be great diagnostic and prognostic signatures and the two GSVA score systems may become promising indices for guiding the tumor treatment of LIHC patients.

**Keywords:** hepatocellular carcinoma, prognostic stratification system, gene set variation analysis, PPI, tumor infiltrating immune cell

## INTRODUCTION

Liver cancer is the sixth most common malignant tumor and the fourth most common cause of cancer-related death (Villanueva 2019). Cirrhosis, mostly as a result of hepatitis virus infection or alcohol abuse, is currently considered to be the main cause of liver cancer (Marengo et al., 2016). Liver hepatocellular carcinoma (LIHC) is the most common type of liver cancer. The main treatment strategies for LIHC are surgery, radiotherapy, chemotherapy, and palliative therapies (Llovet et al., 2015). Regrettably, these treatments are less effective in patients with advanced LIHC (Tian et al., 2018). Therefore, it is urgent to explore significant diagnosis and prognosis indicators of LIHC. The wide use of high-throughput sequencing technology in Liver cancer research has revealed many promising targets for the early diagnosis and evaluation of prognosis (Zhang et al., 2017; Calderaro et al., 2019). AFP is the most common biomarker in LIHC for early diagnosis and tumor recurrence surveillance (Pinerio et al., 2020). DKK1 has been found highly expressed in HCC tissue and proposed to be a novel HCC biomarker with a very good diagnostic performance (Shen et al., 2012). Higher expression of Glypican-3 was significantly associated with a worse prognosis in LIHC (Xiao et al., 2014). The upregulated expression of TBK1 can enhance tumor immune infiltration and predict the poor prognosis of patients with LIHC (Jiang et al., 2021). Other studies develop prognostic models based on gene sets and carry out validation analyses (Huo et al., 2020; Zhou et al., 2020). Nevertheless, most studies do not take the changes of gene expression patterns in different stages of tumors into account.

In our study, we identified LIHC-unfavorable gene set and LIHC-favorable gene set by integrating gene expression data and corresponding clinical data from TCGA. Gene set variation analysis (GSVA) was used to calculate the enrichment score of LIHC patients and construct two scoring systems. The diagnostic and prognostic capability of two scoring systems were verified in multiple datasets. By integrating the PPI network and univariate Cox regression analysis of all LCGs, ESR1, EHHADH, CYP3A4, and ACADL were determined as the key LCGs. Subsequently, we investigate the expression level and prognostic correlation of the key LDGs in different HCC datasets. Additionally, ssGSEA analysis was used to explore the correlation of the two gene sets with gene alteration and immune infiltration. These findings indicate that the two GSVA scoring systems may become reliable molecular markers and provide targets for the diagnosis and prognosis of LIHC.

## MATERIALS AND METHODS

### Data Collection

The gene expression data and corresponding clinical features of LIHC patients were downloaded from International Cancer Genome Consortium (ICGC) (Zhang et al., 2019), The Cancer Genome Atlas (TCGA) (Tomczak et al., 2015), and Gene Expression Omnibus (GEO) (Barrett et al., 2013). TCGA LIHC cohorts containing 50 control samples and 374 HCC samples (175 stage I samples, 87 stage II samples, 86 stage III samples, and 26 stage IV samples) were collected for subsequent analyses. In addition, we obtained gene expression array and prognostic information of GSE14520 cohorts (374 HCC samples and 50 control samples) and ICGC LIHC cohorts (212 HCC samples and 177 control samples) as validation sets. The genes with lower expression and samples with no prognostic information were excluded.

### Identification of LIHC-Development Genes

The “normalizeBetweenArrays” function in “limma” R package was performed to background adjustment and quantile normalization. In TCGA datasets, DEGs between normal group and I-IV HCC stage groups were respectively identified utilizing the “limma” package with a fold-change of 1.5 and an adjusted  $p$ -value of  $<0.05$  (Ritchie et al., 2015). We defined LDG as gradually upregulated DEGs (logFCstage I vs. control  $<$  logFCstage II vs. control  $<$  logFCstage III vs. control  $<$  logFCstage IV vs. control) and downregulated DEGs (logFCstage I vs. control  $>$  logFCstage II vs. control  $>$  logFCstage III vs. control  $>$  logFCstage IV vs. control). Potential functions and enriched pathways of LDGs were further explored by the “clusterProfiler” package (Yu et al., 2012), and  $p < 0.05$  was considered as significant.

### Establishment of the LIHC-Progression Gene Set Variation Analysis Score

According to the median expression level of LDGs, all samples were divided into high/low groups and subjected to Kaplan–Meier survival curves analyses, and  $p < 0.05$  was considered to be statistically significant. Those LDGs that drastically influenced survival were considered as LIHC-progression characteristic genes (LCGs) and established two prognostic gene sets, including the LIHC-unfavorable gene set (LUGs, related to poor prognosis) and the LIHC-favorable gene set (LFGs, related to good prognosis). Several external microarray

datasets (GSE10143, GSE14520, GSE22058, GSE25097, GSE36376, GSE46444, GSE54236, GSE63898, GSE64041, and GSE76427) were performed to validate the differential expression of LCGs between HCC samples and adjacent normal samples.

Gene Set Variation Analysis (GSVA) is a non-parametric, unsupervised algorithm for calculating Normalized Enrichment score (NES) of pathways and functional annotation based on gene expression array, which was extensively utilized in cancer-related studies (Liu et al., 2021a; Liu et al., 2021b; Liu et al., 2021c). Next, we further performed GSVA approach based on the two prognostic gene sets to calculate the NES of each patient as LIHC-unfavorable GSVA score (LUG score) and LIHC-favorable GSVA score (LFG score) using the “GSVA” R package (Hanzelmann et al., 2013). Receiver operating characteristic curve (ROC) analysis was employed to illustrate the diagnostic veracity of two GSVA scores in different HCC cohorts (TCGA, ICGC, and GSE14520). Patients in TCGA, ICGC, and GSE14520 cohorts were divided into high/low-risk groups according to the median scores and subsequently carried out to Kaplan–Meier survival analysis.

## Clinical Correlation Analyses of the LUG Score and LFG Score

To investigate the impact of the two GSVA scores on clinical characteristics, we further explore the relationship of the LUG score and LFG score with other clinical characteristics (age, gender, Child grade, T stage, M stage, N stage, and race). In addition, the univariate Cox regression analysis was employed to evaluate the correlation between prognosis and clinical characteristics, and multivariate Cox regression analysis was applied to analyze the independent prognostic ability of the risk factors.

## Mutation and Immunohistochemistry Analyses of LCGs

To determine the somatic mutations of HCC patients between high- and low-GSVA score groups, the mutation annotation format (MAF) from the TCGA database was generated using the “maftools” R package (Mayakonda et al., 2018). The Human Protein Atlas is a human protein online database including normal and neoplastic tissues (Uhlen et al., 2010). We utilized the Human Protein Atlas web tool to validate the abnormal expression of LCGs between HCC and liver tissues at the protein level.

## Exploration of the Molecular Mechanism

The GSVA method was used to quantify the activity of molecular pathways and find significantly correlated pathways with two GSVA scores. The differences in NES between the high- and low-GSVA scores groups were compared by independent-samples *t*-tests, and  $p < 0.05$  was regarded as statistically significant. Gene Ontology (GO) enrichment analysis was performed on the DEGs identified by the “limma” R package between the high- and low-LUG score groups. Gene set enrichment analysis (GSEA) was

applied to evaluate the immune response between the high- and low-LUG score groups, and adjusted *p*-value  $< 0.05$  was considered to be different (Subramanian et al., 2005). The gene set “c2.cp.kegg.v6.2.symbols.gmt” and “h.all.v7.2.symbols.gmt” were chosen as the reference gene set.

## Construction of PPI Network and Identification of the Hub LCGs

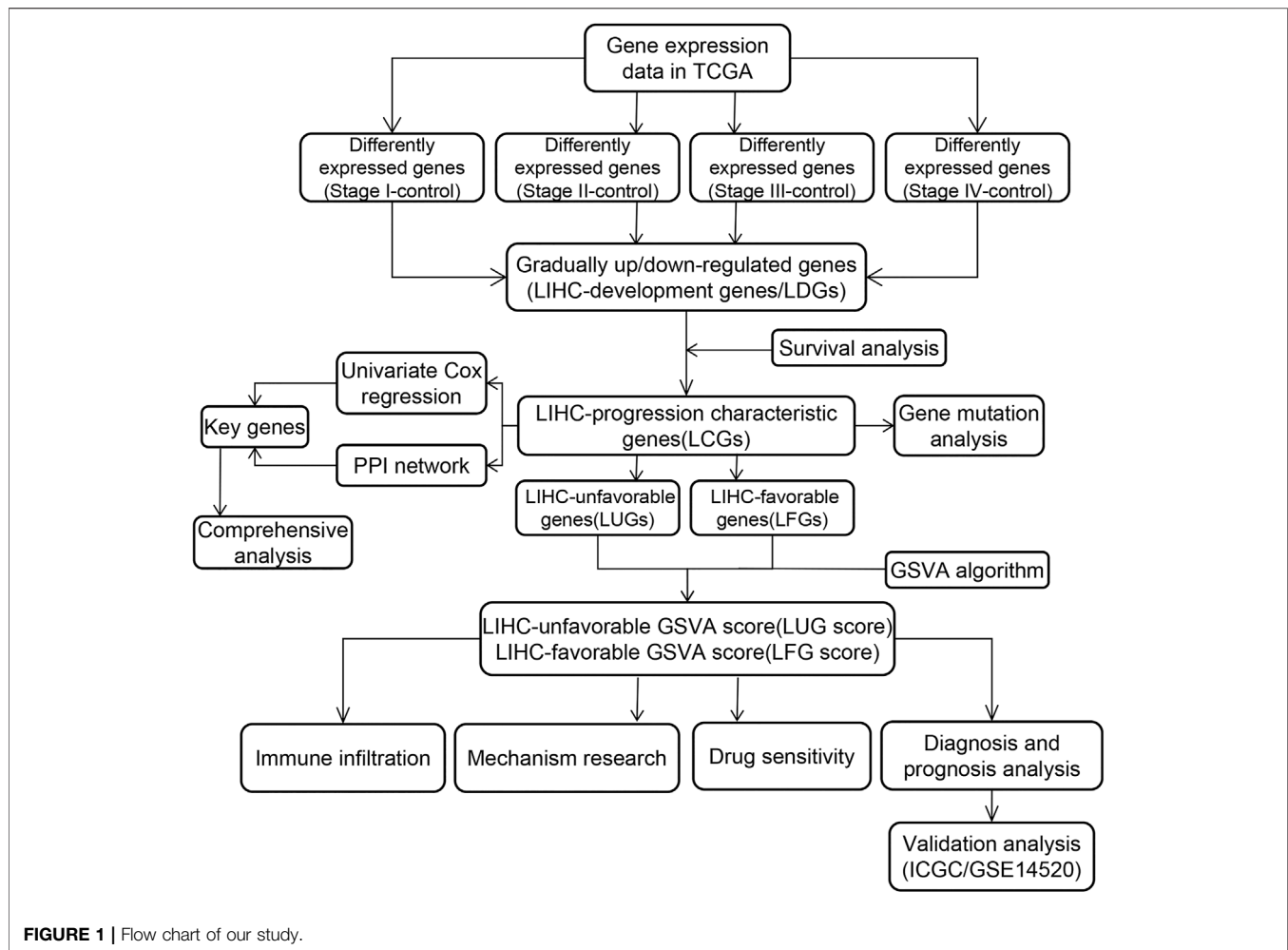
A PPI network between LDGs was constructed through the Search Tool for the Retrieval of Interacting Genes (STRING) online tool (Szklarczyk et al., 2021). Nodes with interaction scores  $> 0.9$  and containing LCGs were imported to the Cytoscape, a software for visualizing complex networks. Additionally, univariate regression analysis was utilized to evaluate the prognostic relevance of the LCGs. The key LCGs were screened out and the selection criteria was the number of adjacent nodes  $> 4$  in the network and *p*-value  $< 0.05$  in prognostic analysis. The Gene Set Cancer Analysis (GSCA) database integrates comprehensive cancer information from TCGA (Liu et al., 2018). We explored aberrant LCG expression in several types of cancer utilizing the GSCA online tool.

## Comprehensive Analysis of the Key LCGs

The difference in expression level of key LCGs between tumor and normal samples were validated in various datasets using independent-samples *t*-test procedure. And the variation of the key LCGs’ expression pattern as tumor stage increased was verified by the Gene Expression Profiling Interactive Analysis (GEPIA) database (Tang et al., 2017). Simultaneously, the external validation sets (ICGC and GSE14520) were carried out to Kaplan–Meier survival analysis between high- and low-expression groups, which were divided by the median expression value of the key LCGs. In order to further confirm the independent prognostic ability of each key LCGs, we combined the clinical features with the key LCGs to perform multivariate analyses based on TCGA and ICGC data. Furthermore, the GSCA database was employed to investigate the potential mechanism of abnormal expression of key LCGs in multiple aspects, including pathway activity and methylation. Respective co-expression networks of the key LCGs in HCC were achieved through the HCCDB online database (Lian et al., 2018), and then input into Metascape for gene annotation (Zhou et al., 2019).

## Immune Infiltration Analysis and Drug Susceptibility Analysis

Single-sample gene set enrichment analysis (ssGSEA) (Barbie et al., 2009) was conducted to quantify infiltration levels for 24 different immune cell types in TCGA HCC samples (Bindea et al., 2013). The correlation between prognostic signatures and immunocyte infiltration levels was evaluated using the “Pearson” approach. The difference in the distribution of immunocyte infiltrating levels between high- and low-GSVA groups was analyzed by Wilcoxon test. The ESTIMATE score



of each sample, comprising StromalScore and ImmuneScore, was calculated using the R package “ESTIMATE” (Yoshihara et al., 2013). The distinction in immune infiltrating level and the ESTIMATE score between high- and low-score groups were analyzed by Wilcoxon test. Immune checkpoint inhibitor (ICI) was an advanced method for activating antitumor immunity (Topalian et al., 2015). Hence, the relationship between the GSVA scores and six common inhibitory checkpoint molecules (CD274, CTLA4, HAVCR2, LAG3, PDCD1, and TIGIT) was assessed to speculate the immunotherapy response targeting ICIs. The Tumor Immune Dysfunction and Exclusion (TIDE) score and Tumor microenvironment evaluation (TME) score are two different computational models for predicting response to immune checkpoint blockade (ICB) (Jiang et al., 2018; Zeng et al., 2021). We uploaded the TCGA transcriptome profiles to the TIDE web and then obtained every patient’s TIDE score, and TME score was computed by “TMEScore” R packages. Moreover, to compare the therapeutic effects of chemotherapeutic drugs in the different score groups, we measure the semi-inhibitory concentration (IC50) values of commonly used chemotherapeutic drugs for LIHC by the “pRRophetic” package (Geeleher et al., 2014).

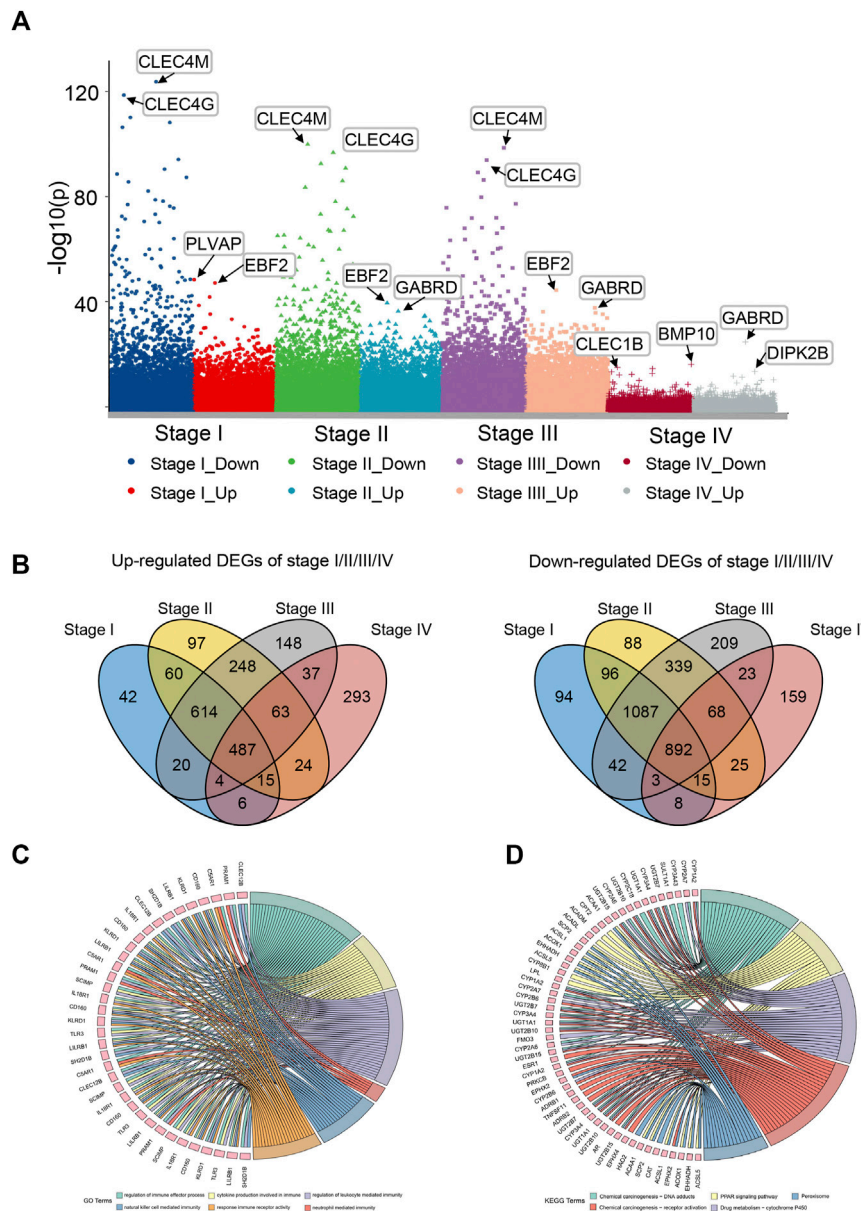
## Statistical Analyses

All statistical analyses were conducted *via* R software (Version 3.6.7). The Student’s *t*-test was used for statistical comparisons. Spearman’s correlation was applied for the analysis of the correlation. The Benjamini–Hochberg false discovery rate (FDR) method was used for *p*-value adjustment. Fisher’s test was used to identify the significant GO terms. A *p*-value <0.05 was regarded as statistically significant. The cut-off value of continuous variables, such as gene expression and immune infiltration level, was median.

## RESULTS

### Identification of the LIHC-Development Gene

The general analysis flow of our study is shown in **Figure 1**. We screened out a total of 487 common upregulated DEGs, and 892 common downregulated DEGs were identified by the intersection of DEGs between different subgroups (**Figure 2B**). Among them, 83 DEGs were gradually upregulated and 247 DEGs were gradually downregulated as the stage evolved. These DEGs



**FIGURE 2 |** Differential expression gene analysis and functional enrichment analysis. **(A)** Manhattan plot showed differentially expressed genes (DEGs) in different stages of LIHC. **(B)** Venn plot of up/downregulated common DEGs in LIHC stage I–IV. **(C)** GO enrichment analysis of LIHC-development genes. **(D)** KEGG pathway analysis of LIHC-development genes.

may have a sustained effect on HCC progression so they are considered as the LIHC-development genes (LDGs). In the result of the GO analysis, the TRGs were mainly associated with the regulation of cell cycle, chromosome segregation, mitotic nuclear division, regulation of inflammatory response and immune effector process, response to drug, and organelle fission (Figure 2F). The result of GO analysis showed that the LDGs were enriched in several immunoregulation ways, such as regulation of the immune effector process, cytokine production involved in immune response, regulation of leukocyte-mediated immunity, and neutrophil-mediated

immunity (Figure 2C). As for the KEGG pathway enrichment, the LDGs were mainly associated with the chemical carcinogenesis, PPAR signaling pathway, peroxisome, and drug metabolism of cytochrome P450 (Figure 2D).

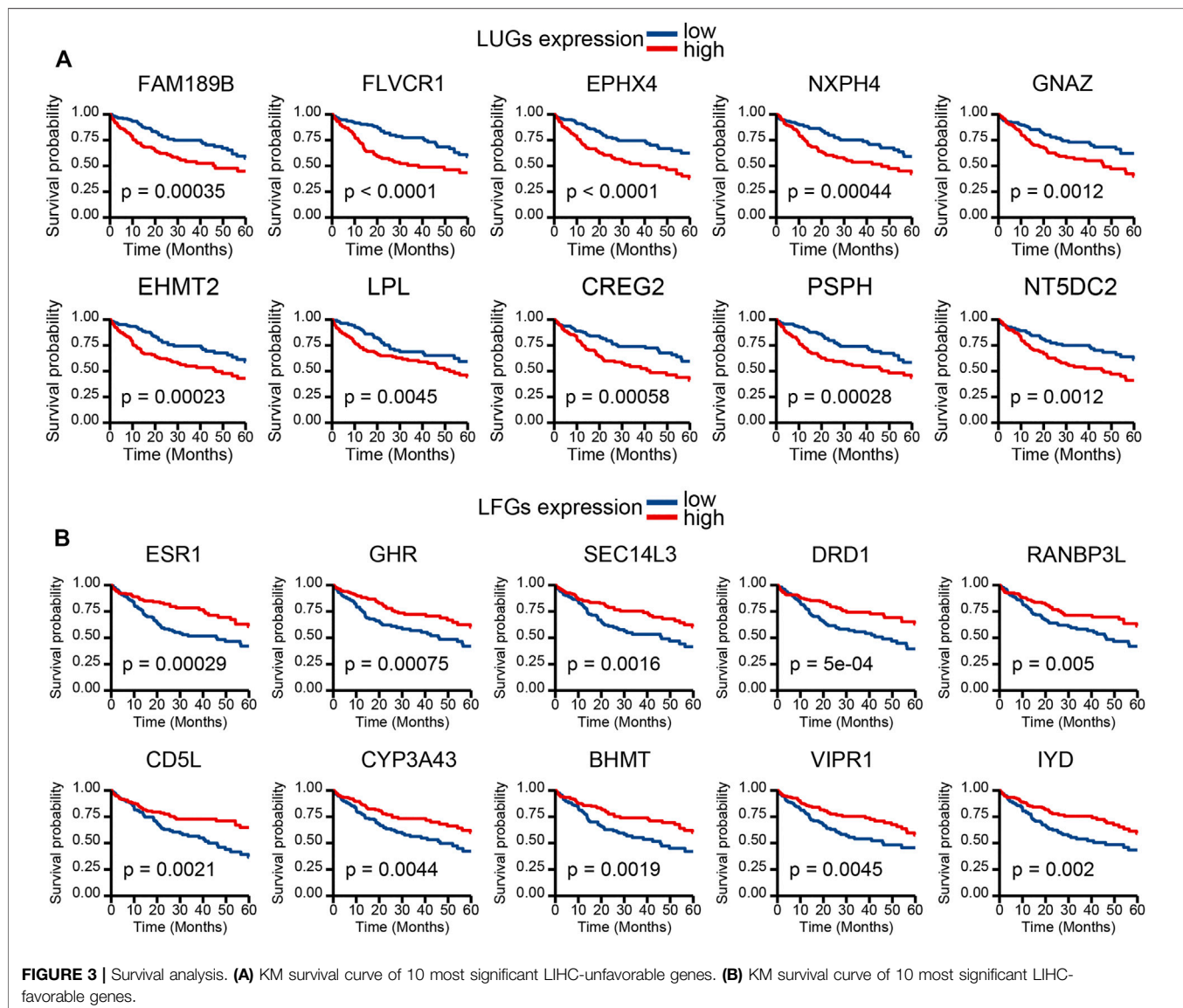
### Two Groups of LDGs With Opposite Prognostic Characteristics Were Picked out

Kaplan–Meier (KM) curve analysis discovered that 63 LDGs were prominently associated with clinical outcome and named LIHC-progression characteristic genes (LCGs). Among them, the

**TABLE 1** | LIHC-unfavorable gene set and LIHC-favorable gene set.

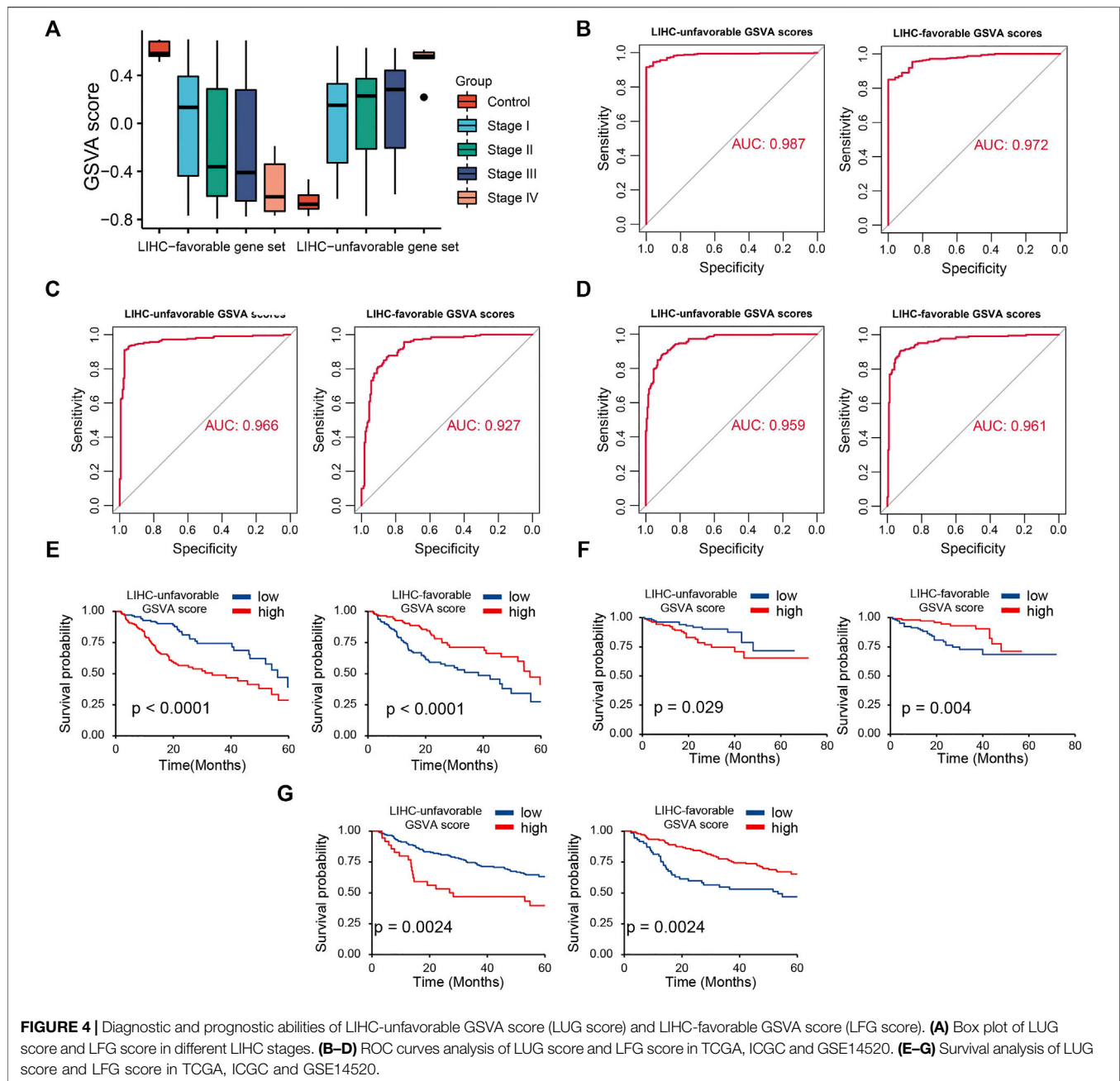
Gene set	Gene symbol
LIHC unfavorable genes	PYGO2, FAM189B, EHMT2, TARBP1, FLVCR1, ADAM15, TIGD1, LAMC1, LPL, EPHX4, EGFL6, CREG2, NXPH4, CEP72, HEY1, PSPH, H4C8, HPDL, GNAZ, NT5DC2, ATP6V0D2, NANOS1, MEX3A, HES2, CHML, GNG4, CYP19A1, ATP8A2, STK39, PNCK, ETV4
LIHC-favorable genes	VIPR1, CPEB3, ESR1, ADRA1A, CD5L, RANBP3L, GHR, HAO2, CYP3A43, ACADL, EPHX2, TERB2, IYD, CCT6B, DMGDH, GBP7, RDH16, SEC14L3, ABCA9, EHHADH, DHRS1, CYP3A4, MOGAT1, BHMT, SLC38A4, PACRG, ACOT12, TTPA, HDC, CYP8B1, HLF, DRD1

HCC, hepatocellular carcinoma; TCGA, the cancer genome atlas; ICGC, international cancer genome consortium; GEO, gene expression omnibus; LDGs, LIHC-development gene; LCGs, LIHC-progression characteristic gene; LUGs, LIHC-unfavorable gene set; LFGs, LIHC-favorable gene set; LUG score, LIHC-unfavorable GSVAs score; LFG score, LIHC-favorable GSVAs score; DEGs, differential expressed genes; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; KM, Kaplan–Meier; ROC, receiver operating characteristic; AUC, area under curve; OS, overall survival; GSVAs, gene set variation analysis; NES, normalized enrichment score; GSEA, gene set enrichment analysis; ssGSEA, single sample gene set enrichment analysis.



LIHC-unfavorable gene set (LUGs) contained 31 LCGs related to poor prognosis, while the LIHC-favorable gene set (LFGs) incorporated 32 LCGs linked to good prognosis (Table 1).

Kaplan–Meier (KM) curves based on TCGA cohorts of LUGs and LFGs are shown in Figures 3A,B. Additionally, all LCGs were differentially expressed between HCC and adjacent noncancerous

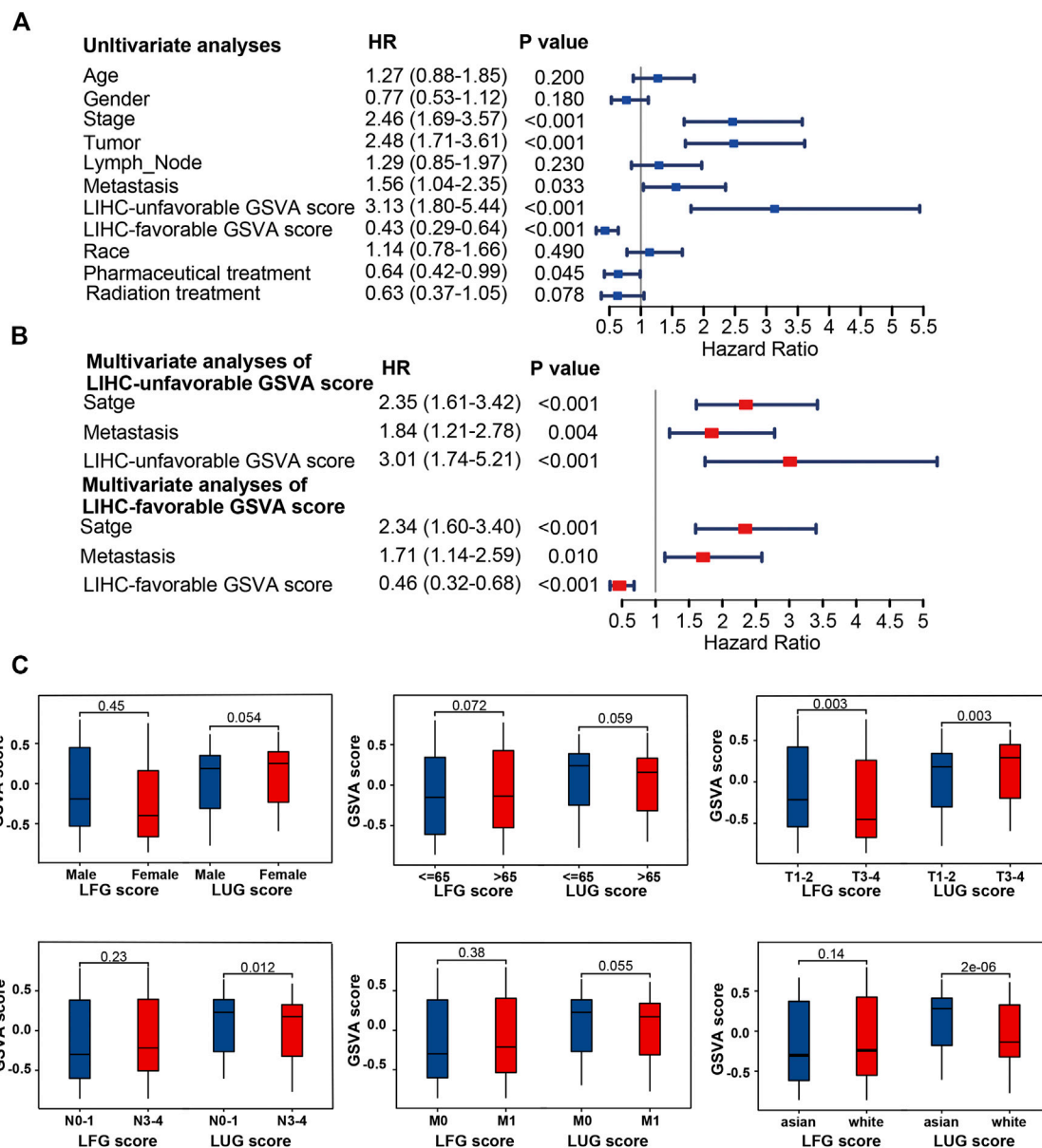


tissue in multiple validation datasets from different platforms (**Supplementary Figure S1**). IHC analyses from HPA database also confirmed aberrant expression of LCGs in tumor tissue (**Supplementary Figure S2**).

## LIHC-Progression GSVAs Score Could Effectively Predict Prognosis for LIHC Patients

Based on two prognosis-related gene sets (LUGs and LFGs), we used GSVAs algorithm to construct two LIHC-progression GSVAs scores, named LIHC-unfavorable GSVAs score (LUG score) and LIHC-

favorable GSVAs score (LFG score) respectively. Obviously, the LUG score gradually increased as the tumor progresses in HCC patients, while the LFG score was complete opposite (**Figure 4A**). ROC analysis proved that both LUGs and LFGs had great diagnostic accuracy in diverse independent verification datasets, among which  $AUC = 0.987$  and  $0.972$  in TCGA,  $AUC = 0.966$  and  $0.927$  in GSE14520, and  $AUC = 0.959$  and  $0.961$  in ICGC (**Figures 4B–D**). As shown in **Figures 4E–G**, survival analyses indicated patients from the low-LFG score group or high-LFG score group had a longer OS than those from the high-LFG score group or high-LFG score group. According to the univariate/multivariate Cox regression analysis, TNM stage, LUG score, and LFG score can serve as independent



**FIGURE 5** | Clinical correlation analyses and of LUG score and LFG score. (A,B) Univariate and multivariate Cox regression analysis. (C) Correlation of LUG score and LFG score with clinical features.

predictors to evaluate the prognosis of HCC patients (Figures 5A,B). Subsequently, we explored the relevance between the GSVA scores and other clinicopathological parameters. The result indicated the LFG score was significantly related to T stage, and the LUG score has a marked correlation with T stage, N stage, and race (Figure 5C).

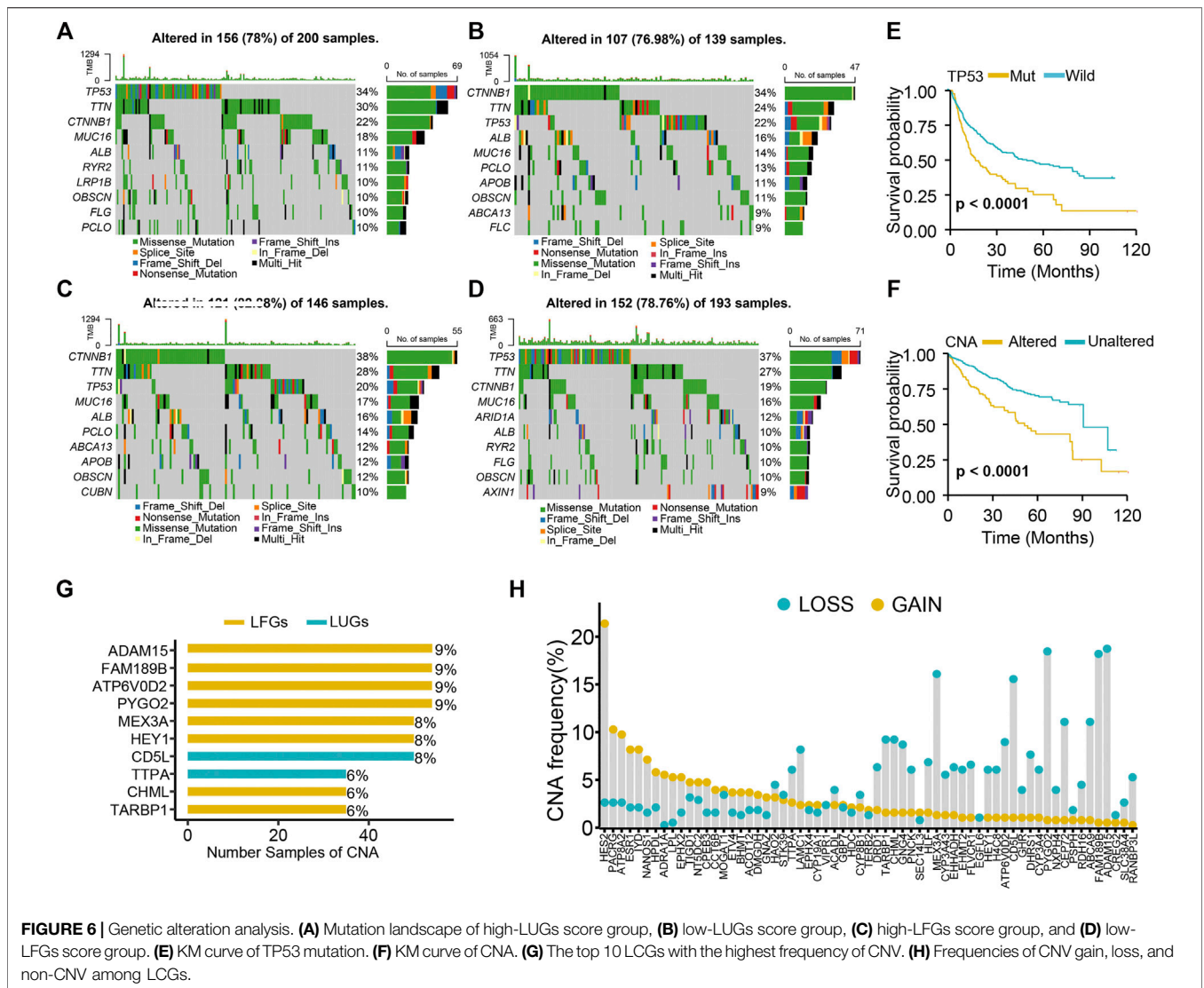
## Genetic and Transcriptional Alterations of GSVA Scores and LCGs in LIHC

Both the high-LUG score group and low-LFG score group had a higher TP53 mutation rate than the low-score groups (Figures 6A–D). The prognosis of patients with TP53 mutations was

significantly worse than those with wild TP53 (Figure 6E). Because of the high mutation frequency and poor prognostic feature of TP53, we evaluated the relationship between TP53 mutation and LCGs expression. The results showed that the expression levels of 21 of the 63 LCGs were significantly associated with TP53 mutation status (Supplementary Figure S3).

We found high CNA frequency in patients who seemed to presage poor prognosis (Figure 6F) and prevalent copy number alterations in all LCGs (Figure 6G). LCGs with CNV gain, such as NT5DC2, GNAZ, and HPDL, were significantly elevated in LIHC samples, while LCGs with CNV loss, such as CYP3A4, GHR, and





HLF, were decreased in LIHC samples, suggesting that CNV might regulate the mRNA expression of LCGs (Figure 6H). However, some LCGs with CNV loss, such as EHMT2 and HEY1, showed upregulated expression, while other LCGs with abnormal expression showed no differences of frequency between CNV gain and loss. Hence, although CNV can explain expression variation in many LCGs, CNV is not the only factor involved in the regulation of mRNA expression (Sebestyen et al., 2016).

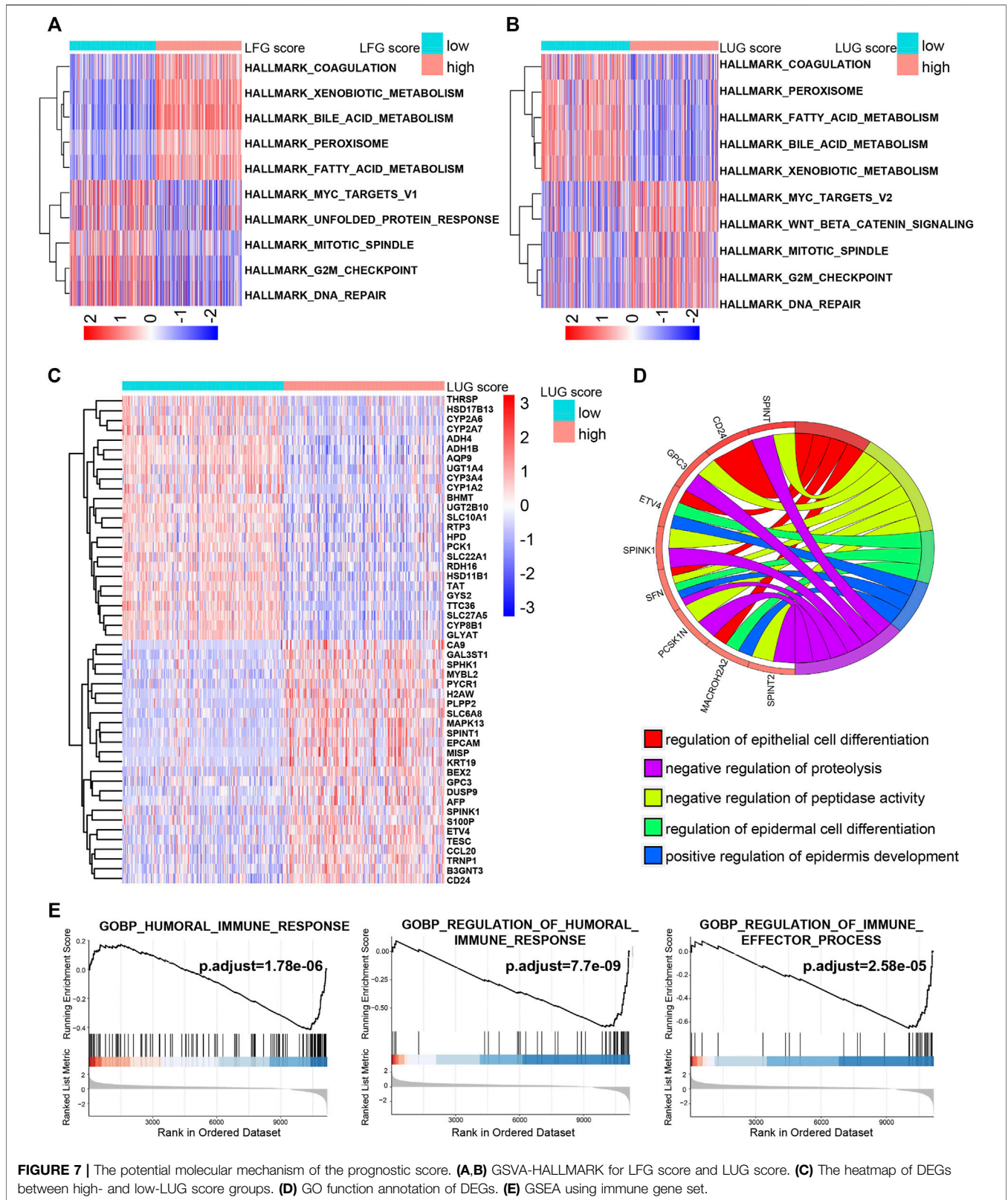
### Potential Molecular Mechanism of Two GSA Scores

Both the high-LFG score group and low-LUG score group were significantly enriched for metabolisms, such as fatty acid metabolism, bile acid metabolism, and xenobiotic metabolism, while the activity of pathways related to cell cycle, such as G2M checkpoint, mitotic spindle, and DNA repair mitotic spindle enriched significantly in the low-LFG score group and high-

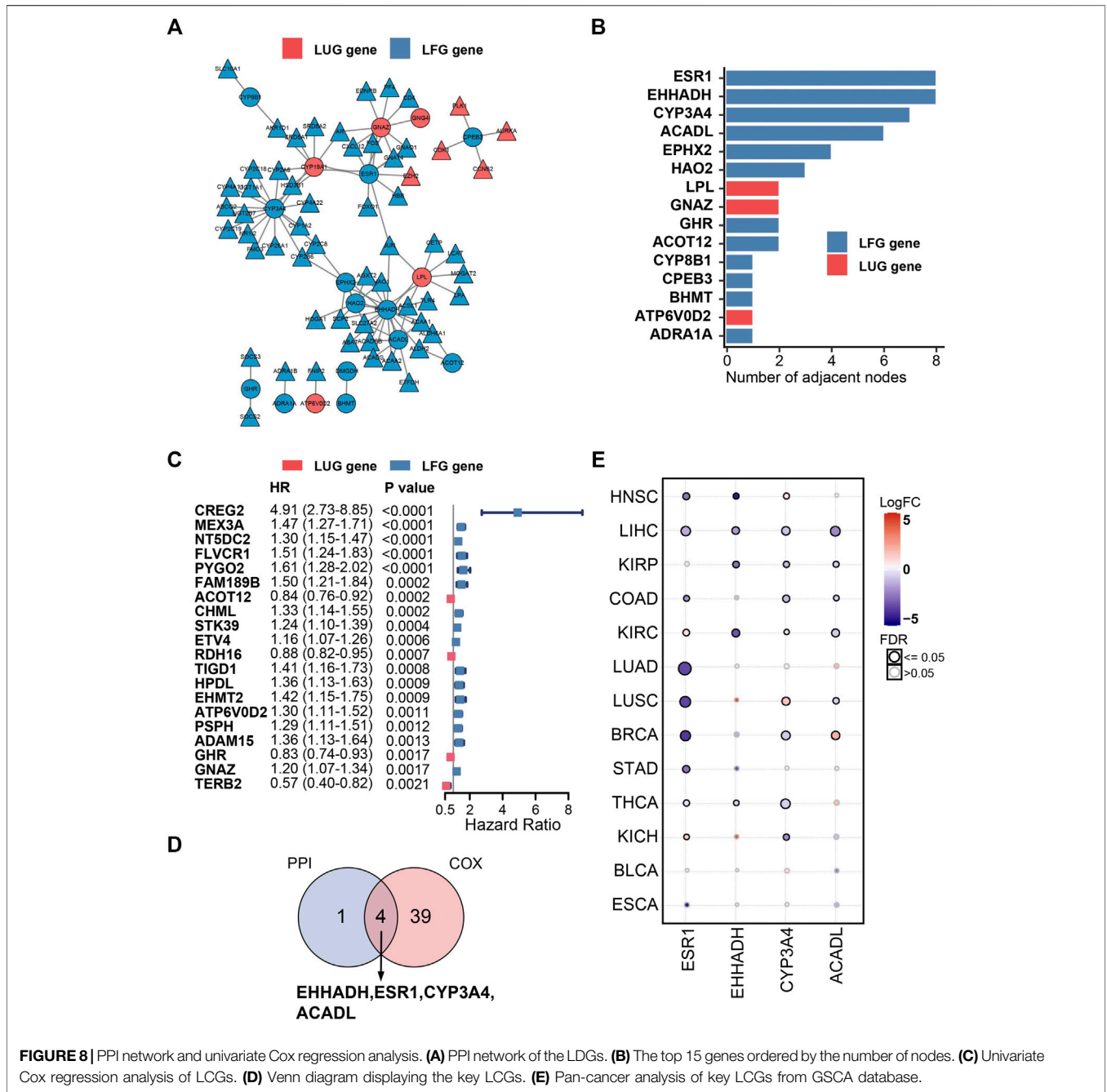
LUG score group (Figures 7A,B). GO enriched the annotation of upregulated DEGs in the high-LUG score group showed the significant activated functional pathways related to cell differentiation, including differentiation regulation of the epidermal cell and epithelial cell (Figures 7C,D). It is worth noting that the immune responses were mainly active in the high-LUG score group, as revealed by GSEA (Figure 7E).

### Four Key LCGs Were Screened out by PPI Network Analysis and Univariate Cox Regression Method

A PPI network, composed of 77 nodes and 152 edges, was built using the STRING database (Figure 8A). As shown in Figure 8B, the importance of LCGs was ordered by their number of adjacent nodes in the network. On the other hand, a total of 40 LCGs could affect the outcome of HCC patients according to univariate Cox regression analysis (Figure 8C). Eventually, four LFGs (ESR1,



**FIGURE 7 |** The potential molecular mechanism of the prognostic score. **(A,B)** GSEA-HALLMARK for LFG score and LUG score. **(C)** The heatmap of DEGs between high- and low-LUG score groups. **(D)** GO function annotation of DEGs. **(E)** GSEA using immune gene set.



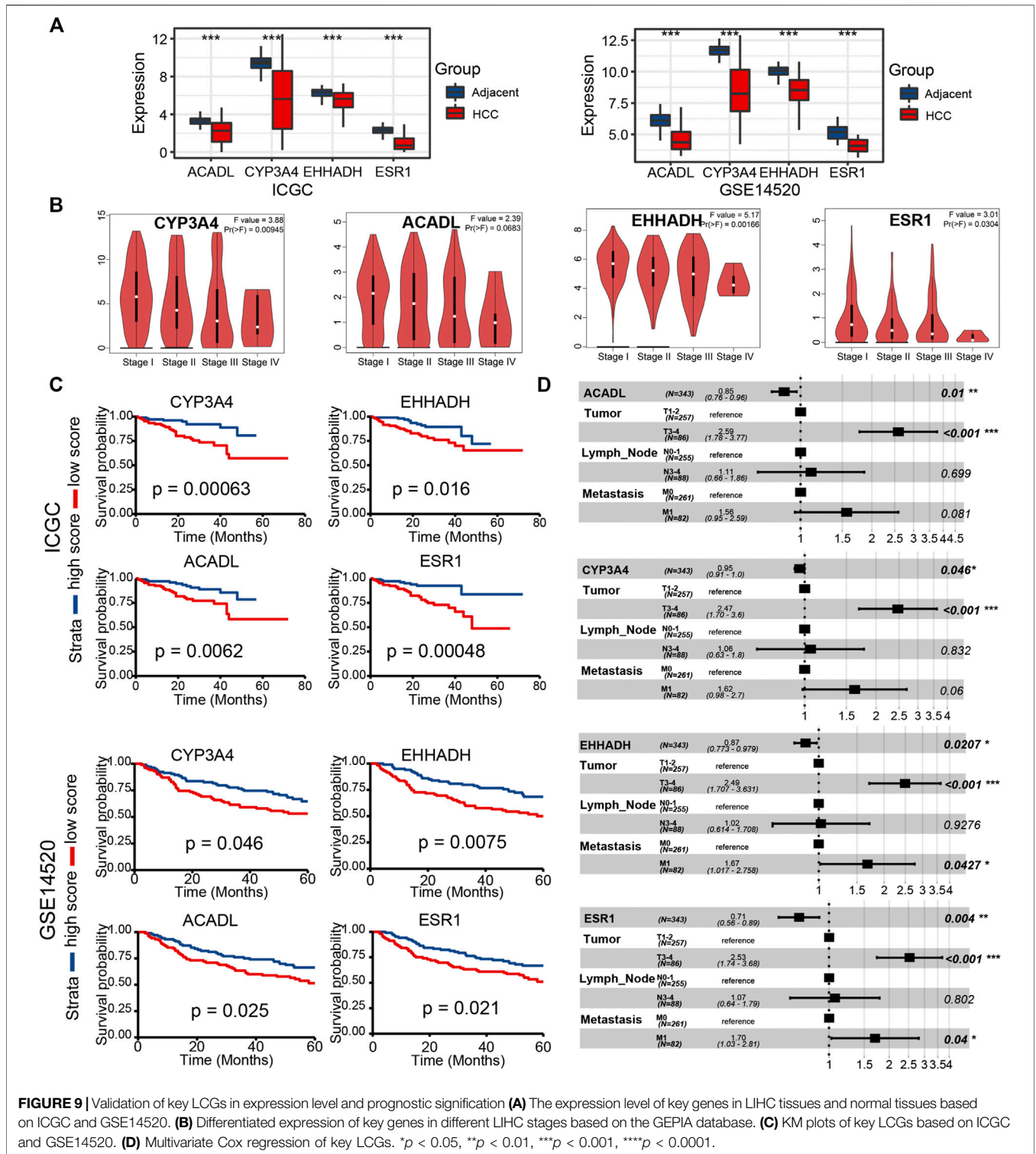
**FIGURE 8** | PPI network and univariate Cox regression analysis. **(A)** PPI network of the LDGs. **(B)** The top 15 genes ordered by the number of nodes. **(C)** Univariate Cox regression analysis of LCGs. **(D)** Venn diagram displaying the key LCGs. **(E)** Pan-cancer analysis of key LCGs from GSCA database.

EHHADH, CYP3A4, and ACADL) were selected as the key LCGs by integrated analysis of survival evaluation and PPI network (Figure 9D). Pan-cancer research indicated these four key LCGs also apparently decreased in a variety of cancers (Figure 8E).

### Validation of Key LCGs Expression and Prognosis in External Data

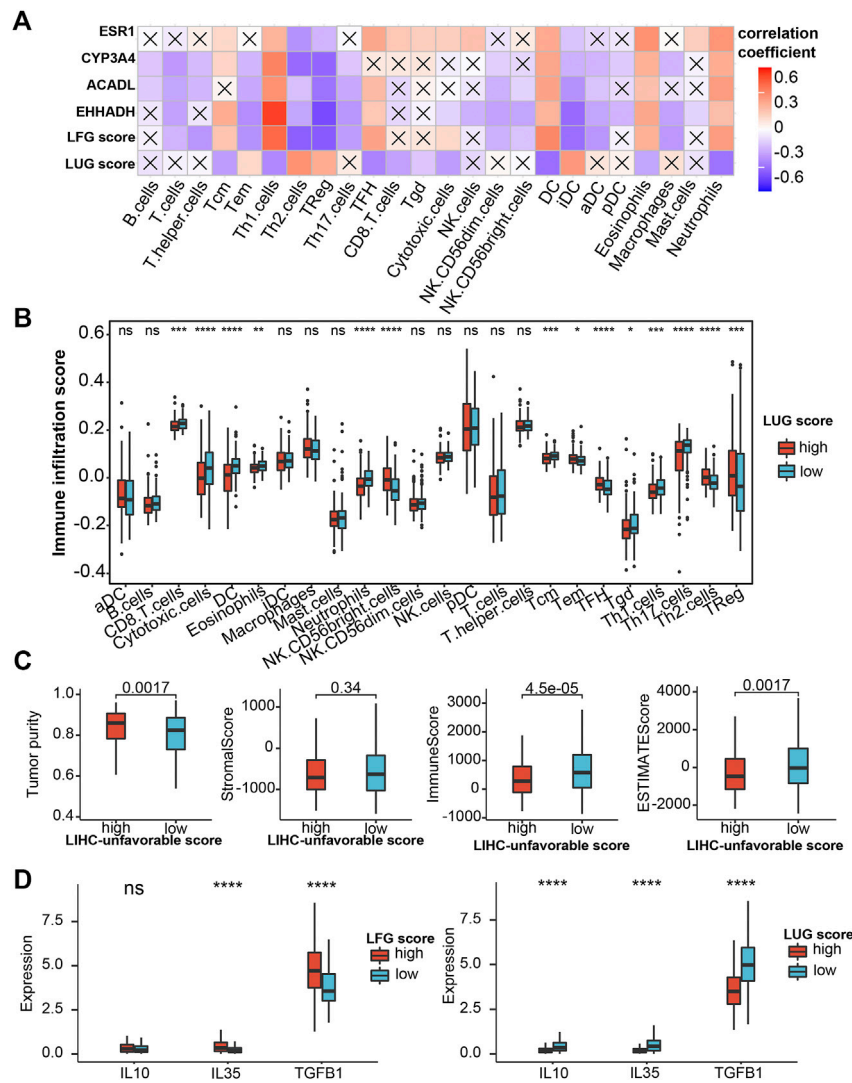
In the two external datasets (ICGC and GSE14520), gene expression levels of four key LCGs were also lower in the liver cancer tissue than adjacent tissue, which is in line with

previous researches (Figure 9A). Moreover, the box plot of gene expression at different stages obtained from GEPIA proved that four key LCGs possessed similar expression patterns in the HCC progression (Figure 9B). KM survival curves based on ICGC and GSE14520 cohorts demonstrated key LCGs performed great efficiency for distinguishing prognostic different HCC patients (Figure 9C). Combining the clinical features with the key LCGs expression, multivariate Cox regression validated that the key LCGs were independent prognostic factors and protective factors (Figure 9D; Supplementary Figure S4). Pathway analyses of



GSCA showed that all key LCGs might participate in Hormone pathways, and EHHADH is probably connected with the RTK pathway (Supplementary Figure S5A). Meanwhile, the methylation level of ESR1 and CYP3A4 in tumor samples was significantly higher than that in normal

samples, implying that methylation could be one of the factors leading to abnormal gene expression (Supplementary Figure S5B). Single gene enrichment analysis based on HCCDB and Metascape revealed that the key LCGs had remarkable correlation with metabolism pathways, which was in



**FIGURE 10 |** Evaluation of immune infiltration **(A)** Correlation heatmap of LFG score, LUG score and key LCGs with 24 immune cells. The cross indicates no significance. **(B)** Immune infiltration score in the high- and low-LUG score group. **(C)** Correlations of LUG score with immune score, stromal score, ESTIMATEScore and tumor purity. **(D)** Expression of the immune suppressive cytokines between high- and low-LUG/LFG score group. \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001, \*\*\*\**p* < 0.0001.

keeping with previous results on GSVA scores (Supplementary Figure S6).

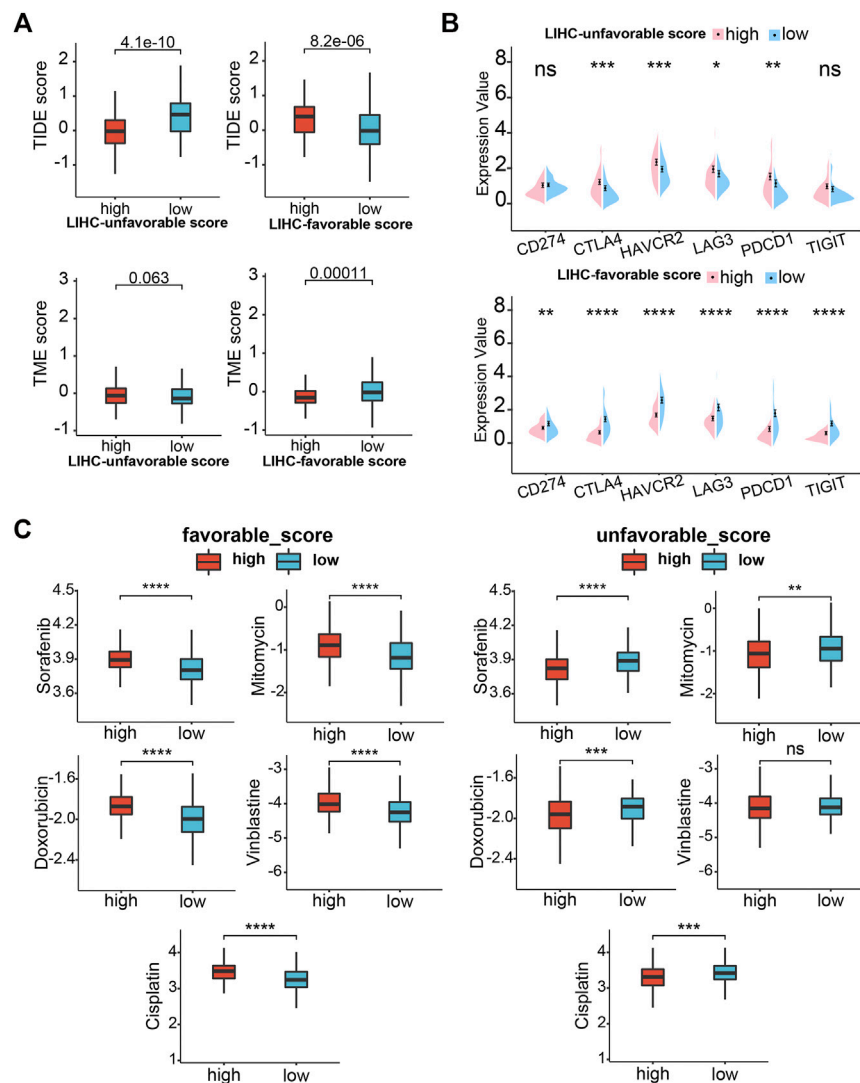
### Immune Infiltration and Drug Susceptibility Analysis

We performed the ssGSEA algorithm to assess the association of the abundance of immune cells with two GSVA scores and the key LCGs. As shown in Figures 10A,B, the LFGs score similar to the key LCGs, were positively correlated with Th1 cells, DC, Eosinophils and Neutrophils, while negatively correlated with Th2 cells, TReg and iDC. Interestingly, the LFGs score performed oppositely in these immune cells compared with the LUGs score (Figures 10A,B). Tumor purity in the high-LFG score group was significantly higher

than those in the low-LFG score group, and StromalScore, ImmuneScore, and ESTIMATEScore in the low-LFG score group were significantly higher than those in the high-LFG score group (Figure 10C). Chemokines involved in the immunosuppressive process induced by Tregs (IL-4, IL-35, and TGF-β) were also significantly upregulated in the high-LUG score group and low-LFG score group.

### Immunotherapy Response and Drug Susceptibility Analysis

Subsequently, we analyzed the correlation between GSVA scores and multiple immunotherapy response-related indices to assess their impacts on immunotherapy. Patients with low-LUG scores get a higher TIDE score and lower TME score than those with



**FIGURE 11** | Drug sensitivity analysis. **(A)** TIDE scores and TME score between high- and low-LUG/LFG score group. **(B)** Correlation analysis of two GSVA scores and six immune checkpoint genes. **(C)** Relationships between chemotherapeutic sensitivity and both LFG score and LUG score. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

high-LUG score (Figure 11A). In addition, we investigated the associations between immune checkpoints and our GSVA scores. Figure 10E shows that several immune checkpoints were differentially expressed in the two groups, including PD-1, PD-L1, and CTLA-4 (Figure 11B). These results demonstrated that patients with low LFG scores tended to have a better immunotherapy response. We next selected chemotherapy drugs recommended for liver cancer treatment by AJCC guidelines to evaluate the sensitivities of patients in the low- and high-GSVA score groups to these drugs. Interestingly, we found that the patients in the high-LUG score group or low-LFG score group had lower IC50 values for Sorafenib, Doxorubicin, Doxorubicin, and Cisplatin. Together, these results showed that LUG score and LFG score were related to drug sensitivity (Figure 11C).

## DISCUSSION

LIHC is a common digestive system tumor with high aggressiveness and poor prognosis. LIHC is insensitive to conventional radiotherapy and chemotherapies; consequently, surgery becomes the main treatment (Novikova et al., 2017). Unfortunately, only 30%–40% of LIHC patients are eligible for surgical resection, and the recurrence rate after surgery is very high (Cao et al., 2012). Therefore, it is urgently needed to explore reliable biomarkers that can be regarded as potential diagnostic and therapeutic targets.

With the rapid progress and widespread application of high-throughput sequencing technology, integrated bioinformatics analysis has emerged as a promising approach to explore various diagnostic and prognostic biomarkers for different tumors. In our research, LIHC data from TCGA were used for

bioinformatics analysis to identify genes that were differentially expressed in different stages. Interestingly, we found gene expression patterns of some DEGs incrementally or digressively changed with LIHC development. For example, a gene may be obviously differentially expressed in the advanced stage but not in the early stage. Thus, we considered these LIHC-development genes may have an impact on cell malignant transformation and tumor evolution. With tumor deterioration, there were 330 LDGs screened out, including 83 LUGs gradually upregulated and 247 LFGs gradually downregulated. Additionally, GO functional enrichment analysis indicated that LDGs were significantly involved in the regulation of immunity. Results from KEGG pathway enrichment analysis manifested LDGs were enriched for the chemical carcinogenesis and PPAR signaling pathway.

After considering the prognostic factors, the number of LDGs was further reduced to 31 LUGs and 32 LFGs. It has been reported that certain LUGs and LFGs are related to the formation and regulation of tumor progression. EMHT2 encodes a methyltransferase that is significantly associated with HCC progression and aggression (Wei et al., 2017). CHML promotes HCC metastasis and leads to poor survival, early recurrence, and more satellite nodules (Chen et al., 2019). STK39 contributes to the progression of HCC by the PLK1/ERK signaling pathway (Zhang et al., 2021). ARID2 expression significantly decreased in metastatic HCC tissues, showing a negative correlation with pathological grade and organ metastasis, and a positive association with survival of HCC patients (Jiang et al., 2020). These results confirmed the possibility that LUGs and LFGs can be used as a prognostic model for LIHC.

In the previous studies, it is the common way that a gene set is analyzed by Cox regression and every gene can get a coefficient that can construct the prognostic model. Nevertheless, because of the heterogeneity of the tumor and the limitations of the sample size, the coefficient of a gene is almost impossible to determine. Thus, we took advantage of GSVA methods to calculate individual samples' NES as prognostic features based on LUGs and LFGs. ROC curve analysis and KM analysis suggested the two GSVA scores had precise diagnosis and prognosis capacity, which were verified in the other two independent LIHC datasets. Univariate and multivariate Cox regression analysis also substantiated that LUG score and LFG score were independent prognostic factors for LIHC.

Four LCGs (ESR1, EHHADH, CYP3A4, and ACADL) were identified as key prognosis-related LCGs based on a combination of the PPI network and univariate Cox regression analysis. CYP3A4 encodes a member of the cytochrome P450 superfamily of enzymes and can influence the chemoresistance of LIHC thus leading to a poor prognosis (Ashida et al., 2017). ESR1 has been a focus in breast cancer, and its mutation is a common cause of acquired resistance (Dustin et al., 2019). ACADL restrains hepatocellular carcinoma by targeting Hippo/YAP signaling (Zhao et al., 2020). We have reason to believe the potential effects of these genes to LIHC, although exploration is still insufficient now.

Through the research on the molecular mechanism of prognostic signatures and score models, we found that the high-LFG score group with a poor prognosis was remarkably

enriched in the active metabolism, while the high-LUG score group with a poor prognosis not only exhibited low immune response and metabolic activity but also involved cell cycle regulation. The key LCGs belonged to protective factors and were involved in the metabolic process in HCC. Active metabolism was considered as one of the important signatures of a good prognosis of HCC (Yang et al., 2020; Liu et al., 2021c).

As a continuous breakthrough in the field of immunotherapy, emerging research shows that the tumor microenvironment can regulate cancer progression (Hinshaw and Shevde 2019). Increasing evidence shows that LIHC tissue is often infiltrated by many types of Immune cells (Ringelhan et al., 2018). Th1 cells participate in effective anti-cancer response but Th2 cells show a low cytolytic and antigen-presenting activity. Increase of T2 cells and decrease of T1 cells in intra-tumor are inversely associated with HCC patient survival (Foerster et al., 2018). Our research showed Th2 cells were significantly reduced in the high-LUG score group with a poor prognosis.

DCs play a key role in the initiation and regulation of the immune response. Mature DCs can guide the body to produce a specific immune response and play an anti-tumor role. On the other hand, immature DCs can lead to immune tolerance by activating the body to produce regulatory T cells, anergic T cells, or tolerant T cells (Dhodapkar et al., 2001). In this study, we found that high infiltration of immature DCs mainly happened in the high-LUG score group, while LUG score was negative with infiltration of Mature DCs.

Tregs can promote immunosuppression via secreting immune suppressive cytokines (IL-10, IL-35, TGF- $\beta$ ) or expressing co-inhibitory molecules such as CTLA-4, PD-1, LAG-3, and TIGIT (Josefowicz et al., 2012; Kumar et al., 2018). In the present study, Tregs are upregulated in the high-LUG score group and low-LFG score group. Additionally, cytokines (IL-10, IL-35, TGF- $\beta$ ) related to the immunosuppression process and co-inhibitory checkpoints (CTLA-4, PD-1, and LAG-3) were all upregulated in the high-LUGs score group, which validated that the immunosuppression induced by Tregs exists in high-LUGs score tumors.

Immune checkpoint inhibitors can block immune checkpoints on the cell membrane, which become a promising strategy in the treatment of cancer. Although a variety of immune checkpoint inhibitors has been widely applied in the front-line treatment of HCC, many advanced LIHC patients are resistant to immune checkpoint therapy (Donisi et al., 2020). Our study reveals multiple immune checkpoints (like PD-1, PD-L1, and CTLA4) expression upregulated in high-risk groups. Low TIDE score and high TME score mean a high probability of response to immune checkpoint blockade therapy. We observed that TME scores were significantly higher in high-LFG score groups than those with low-LFG score groups and TME score is completely opposite. Meanwhile, patients with low LFG scores had high expression of multiple immune checkpoints (CTLA4, CD27, HAVCR2, LAG3, PDCD1, and TIGIT). Therefore, we estimate the LFG score possibly can predict the response of immune checkpoint therapy, and combined immunotherapy may be a better choice for the treatment of LIHC.

Nonetheless, several limitations were notable in our study. First, since all data were collected retrospectively, the potential

bias of clinicopathological features is inevitable. Second, the two gene sets may be too large to economize on the sequencing costs. Finally, large-scale prospective studies and functional and mechanistic experimental studies are needed to support our findings.

## CONCLUSION

In summary, we discover two LIHC-progression characteristic gene sets and created two LIHC-progression GSVA scores with great diagnostic and prognostic values for hepatocellular carcinoma. Our findings are of great importance in developing new prognostic markers and molecular targets for LIHC.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## REFERENCES

- Ashida, R., Okamura, Y., Ohshima, K., Kakuda, Y., Uesaka, K., Sugiura, T., et al. (2017). CYP3A4 Gene Is a Novel Biomarker for Predicting a Poor Prognosis in Hepatocellular Carcinoma. *Cancer Genomics Proteomics* 14 (6), 445–453. doi:10.21873/cgp.20054
- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA Interference Reveals that Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* 462 (7269), 108–112. doi:10.1038/nature08460
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets-Update. *Nucleic Acids Res.* 41 (Database issue), D991–D995. doi:10.1093/nar/gks1193
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenaus, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003
- Calderaro, J., Ziol, M., Paradis, V., and Zucman-Rossi, J. (2019). Molecular and Histological Correlations in Liver Cancer. *J. Hepatol.* 71 (3), 616–630. doi:10.1016/j.jhep.2019.06.001
- Cao, H., Phan, H., and Yang, L. X. (2012). Improved Chemotherapy for Hepatocellular Carcinoma. *Anticancer Res.* 32 (4), 1379–1386.
- Chen, T.-W., Yin, F.-F., Yuan, Y.-M., Guan, D.-X., Zhang, E., Zhang, F.-K., et al. (2019). CHML Promotes Liver Cancer Metastasis by Facilitating Rab14 Recycle. *Nat. Commun.* 10 (1), 2510. doi:10.1038/s41467-019-10364-0
- Dhodapkar, M. V., Steinman, R. M., Krasovsky, J., Munz, C., and Bhardwaj, N. (2001). Antigen-specific Inhibition of Effector T Cell Function in Humans after Injection of Immature Dendritic Cells. *J. Exp. Med.* 193 (2), 233–238. doi:10.1084/jem.193.2.233
- Donisi, C., Puzzone, M., Ziranu, P., Lai, E., Mariani, S., Saba, G., et al. (2020). Immune Checkpoint Inhibitors in the Treatment of HCC. *Front. Oncol.* 10, 601240. doi:10.3389/fonc.2020.601240
- Dustin, D., Gu, G., and Fuqua, S. A. W. (2019). ESRI Mutations in Breast Cancer. *Cancer* 125 (21), 3714–3728. doi:10.1002/cncr.32345
- Foerster, F., Hess, M., Gerhold-Ay, A., Marquardt, J. U., Becker, D., Galle, P. R., et al. (2018). The Immune Contexture of Hepatocellular Carcinoma Predicts Clinical Outcome. *Sci. Rep.* 8 (1), 5351. doi:10.1038/s41598-018-21937-2
- Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468

## AUTHOR CONTRIBUTIONS

QH designed the experiments and wrote the paper. BF was mainly responsible for data analysis, and PD was mainly responsible for data collection. YJ was responsible for manuscript review and providing constructive comments. All authors approved the final manuscript.

## ACKNOWLEDGMENTS

We gratefully acknowledge The Cancer Genome Atlas, International Cancer Genome Consortium, Gene Expression Omnibus, cBioPortal for Cancer Genomics, GEPIA, and Gene Set Cancer Analysis databases.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.806989/full#supplementary-material>

- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *Bmc Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7
- Hinshaw, D. C., and Shevde, L. A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 79 (18), 4557–4566. doi:10.1158/0008-5472.CAN-18-3962
- Huo, J., Wu, L., and Zang, Y. (2020). A Prognostic Model of 15 Immune-Related Gene Pairs Associated with Tumor Mutation Burden for Hepatocellular Carcinoma. *Front. Mol. Biosci.* 7, 581354. doi:10.3389/fmolb.2020.581354
- Jiang, H., Cao, H.-J., Ma, N., Bao, W.-D., Wang, J.-J., Chen, T.-W., et al. (2020). Chromatin Remodeling Factor ARID2 Suppresses Hepatocellular Carcinoma Metastasis via DNMT1-Snail axis. *Proc. Natl. Acad. Sci. USA* 117 (9), 4770–4780. doi:10.1073/pnas.1914937117
- Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1
- Jiang, Y., Chen, S., Li, Q., Liang, J., Lin, W., Li, J., et al. (2021). TANK-binding Kinase 1 (TBK1) Serves as a Potential Target for Hepatocellular Carcinoma by Enhancing Tumor Immune Infiltration. *Front. Immunol.* 12, 612139. doi:10.3389/fimmu.2021.612139
- Josefowicz, S. Z., Lu, L.-F., and Rudensky, A. Y. (2012). Regulatory T Cells: Mechanisms of Differentiation and Function. *Annu. Rev. Immunol.* 30, 531–564. doi:10.1146/annurev.immunol.25.022106.141623
- Kumar, P., Bhattacharya, P., and Prabhakar, B. S. (2018). A Comprehensive Review on the Role of Co-signaling Receptors and Treg Homeostasis in Autoimmunity and Tumor Immunity. *J. Autoimmun.* 95, 77–99. doi:10.1016/j.jaut.2018.08.007
- Lian, Q., Wang, S., Zhang, G., Wang, D., Luo, G., Tang, J., et al. (2018). HCCDB: A Database of Hepatocellular Carcinoma Expression Atlas. *Genomics, Proteomics & Bioinformatics* 16 (4), 269–275. doi:10.1016/j.gpb.2018.07.003
- Liu, C.-J., Hu, F.-F., Xia, M.-X., Han, L., Zhang, Q., and Guo, A.-Y. (2018). GSCALite: a Web Server for Gene Set Cancer Analysis. *Bioinformatics* 34 (21), 3771–3772. doi:10.1093/bioinformatics/bty411
- Liu, Z., Liu, L., Jiao, D., Guo, C., Wang, L., Li, Z., et al. (2021a). Association of RYR2 Mutation with Tumor Mutation Burden, Prognosis, and Antitumor Immunity in Patients with Esophageal Adenocarcinoma. *Front. Genet.* 12, 669694. doi:10.3389/fgene.2021.669694
- Liu, Z., Lu, T., Li, J., Wang, L., Xu, K., Dang, Q., et al. (2021b). Clinical Significance and Inflammatory Landscape of a Novel Recurrence-Associated Immune Signature in Stage II/III Colorectal Cancer. *Front. Immunol.* 12, 702594. doi:10.3389/fimmu.2021.702594



- Liu, Z., Lu, T., Wang, L., Liu, L., Li, L., and Han, X. (2021c). Comprehensive Molecular Analyses of a Novel Mutational Signature Classification System with Regard to Prognosis, Genomic Alterations, and Immune Landscape in Glioma. *Front. Mol. Biosci.* 8, 682084. doi:10.3389/fmolb.2021.682084
- Llovet, J. M., Villanueva, A., Lachenmayer, A., and Finn, R. S. (2015). Advances in Targeted Therapies for Hepatocellular Carcinoma in the Genomic Era. *Nat. Rev. Clin. Oncol.* 12 (7), 408–424. doi:10.1038/nrclinonc.2015.103
- Marengo, A., Rosso, C., and Bugianesi, E. (2016). Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. *Annu. Rev. Med.* 67, 103–117. doi:10.1146/annurev-med-090514-013832
- Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118
- Novikova, M. V., Khromova, N. V., and Kopnin, P. B. (2017). Components of the Hepatocellular Carcinoma Microenvironment and Their Role in Tumor Progression. *Biochem. Mosc.* 82 (8), 861–873. doi:10.1134/S0006297917080016
- Piñero, F., Dirchwolf, M., and Pessôa, M. G. (2020). Biomarkers in Hepatocellular Carcinoma: Diagnosis, Prognosis and Treatment Response Assessment. *Cells* 9 (6), 1370. doi:10.3390/cells9061370
- Ringelhan, M., Pfister, D., O'Connor, T., Pikarsky, E., and Heikenwalder, M. (2018). The Immunology of Hepatocellular Carcinoma. *Nat. Immunol.* 19 (3), 222–232. doi:10.1038/s41590-018-0044-z
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Sebestyén, E., Singh, B., Miñana, B., Pagés, A., Mateo, F., Pujana, M. A., et al. (2016). Large-scale Analysis of Genome and Transcriptome Alterations in Multiple Tumors Unveils Novel Cancer-Relevant Splicing Networks. *Genome Res.* 26 (6), 732–744. doi:10.1101/gr.199935.115
- Shen, Q., Fan, J., Yang, X.-R., Tan, Y., Zhao, W., Xu, Y., et al. (2012). Serum DKK1 as a Protein Biomarker for the Diagnosis of Hepatocellular Carcinoma: a Large-Scale, Multicentre Study. *Lancet Oncol.* 13 (8), 817–826. doi:10.1016/S1470-2045(12)70233-4
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic Acids Res.* 49 (D1), D605–D612. doi:10.1093/nar/gkaa1074
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247
- Tian, G., Yang, S., Yuan, J., Threapleton, D., Zhao, Q., Chen, F., et al. (2018). Comparative Efficacy of Treatment Strategies for Hepatocellular Carcinoma: Systematic Review and Network Meta-Analysis. *Bmj Open* 8 (10), e021269. doi:10.1136/bmjopen-2017-021269
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review the Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *Contemp Oncol (Pozn)* 1A (1A), 68–77. doi:10.5114/wo.2014.47136
- Topalian, S. L., Drake, C. G., and Pardoll, D. M. (2015). Immune Checkpoint Blockade: A Common Denominator Approach to Cancer Therapy. *Cancer Cell* 27 (4), 450–461. doi:10.1016/j.ccell.2015.03.001
- Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., et al. (2010). Towards a Knowledge-Based Human Protein Atlas. *Nat. Biotechnol.* 28 (12), 1248–1250. doi:10.1038/nbt1210-1248
- Villanueva, A. (2019). Hepatocellular Carcinoma. *N. Engl. J. Med.* 380 (15), 1450–1462. doi:10.1056/NEJMra1713263
- Wei, L., Chiu, D. K.-C., Tsang, F. H.-C., Law, C.-T., Cheng, C. L.-H., Au, S. L.-K., et al. (2017). Histone Methyltransferase G9a Promotes Liver Cancer Development by Epigenetic Silencing of Tumor Suppressor Gene RARRES3. *J. Hepatol.* 67 (4), 758–769. doi:10.1016/j.jhep.2017.05.015
- Xiao, W.-K., Qi, C.-Y., Chen, D., Li, S.-Q., Fu, S.-J., Peng, B.-G., et al. (2014). Prognostic Significance of Glypican-3 in Hepatocellular Carcinoma: a Meta-Analysis. *Bmc Cancer* 14, 104. doi:10.1186/1471-2407-14-104
- Yang, C., Huang, X., Liu, Z., Qin, W., and Wang, C. (2020). Metabolism-associated Molecular Classification of Hepatocellular Carcinoma. *Mol. Oncol.* 14 (4), 896–913. doi:10.1002/1878-0261.12639
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118
- Zeng, D., Wu, J., Luo, H., Li, Y., Xiao, J., Peng, J., et al. (2021). Tumor Microenvironment Evaluation Promotes Precise Checkpoint Immunotherapy of Advanced Gastric Cancer. *J. Immunother. Cancer* 9 (8), e002467. doi:10.1136/jitc-2021-002467
- Zhang, C., Peng, L., Zhang, Y., Liu, Z., Li, W., Chen, S., et al. (2017). The Identification of Key Genes and Pathways in Hepatocellular Carcinoma by Bioinformatics Analysis of High-Throughput Data. *Med. Oncol.* 34 (6), 101. doi:10.1007/s12032-017-0963-9
- Zhang, C., Wang, X., Fang, D., Xu, P., Mo, X., Hu, C., et al. (2021). STK39 Is a Novel Kinase Contributing to the Progression of Hepatocellular Carcinoma by the PLK1/ERK Signaling Pathway. *Theranostics* 11 (5), 2108–2122. doi:10.7150/thno.48112
- Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., et al. (2019). The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* 37 (4), 367–369. doi:10.1038/s41587-019-0055-9
- Zhao, X., Qin, W., Jiang, Y., Yang, Z., Yuan, B., Dai, R., et al. (2020). ACADL Plays a Tumor-Suppressor Role by Targeting Hippo/YAP Signaling in Hepatocellular Carcinoma. *Npj Precis. Onc.* 4, 7. doi:10.1038/s41698-020-0111-4
- Zhou, T., Cai, Z., Ma, N., Xie, W., Gao, C., Huang, M., et al. (2020). A Novel Ten-Gene Signature Predicting Prognosis in Hepatocellular Carcinoma. *Front. Cel. Dev. Biol.* 8, 629. doi:10.3389/fcell.2020.00629
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 He, Fan, Du and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.