



# Identification of Hub Genes Associated With Hepatocellular Carcinoma Using Robust Rank Aggregation Combined With Weighted Gene Co-expression Network Analysis

Hao Song<sup>1,2</sup>, Na Ding<sup>3</sup>, Shang Li<sup>3</sup>, Jianlong Liao<sup>3</sup>, Aimin Xie<sup>3</sup>, Youtao Yu<sup>2\*</sup>, Chunlong Zhang<sup>3\*</sup> and Caifang Ni<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Bailiang Li,  
Stanford University, United States

### Reviewed by:

Ankush Sharma,  
University of Oslo, Norway  
Jiri Vohradsky,  
Academy of Sciences of the Czech  
Republic, Czechia

### \*Correspondence:

Youtao Yu  
yuyoutao@126.com  
Chunlong Zhang  
zhangchunlong@hrbmu.edu.cn  
Caifang Ni  
caifangnisdfyy@163.com

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 22 April 2020

**Accepted:** 20 July 2020

**Published:** 30 September 2020

### Citation:

Song H, Ding N, Li S, Liao J,  
Xie A, Yu Y, Zhang C and Ni C (2020)  
Identification of Hub Genes  
Associated With Hepatocellular  
Carcinoma Using Robust Rank  
Aggregation Combined With  
Weighted Gene Co-expression  
Network Analysis.  
*Front. Genet.* 11:895.  
doi: 10.3389/fgene.2020.00895

<sup>1</sup> Department of Interventional Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China, <sup>2</sup> Department of Intervention Therapy, The Fourth Medical Center of PLA General Hospital, Beijing, China, <sup>3</sup> Department of Computational Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

**Background:** Bioinformatics provides a valuable tool to explore the molecular mechanisms underlying pathogenesis of hepatocellular carcinoma (HCC). To improve prognosis of patients, identification of robust biomarkers associated with the pathogenic pathways of HCC remains an urgent research priority.

**Methods:** We employed the Robust Rank Aggregation method to integrate nine qualified HCC datasets from the Gene Expression Omnibus. A robust set of differentially expressed genes (DEGs) between tumor and normal tissue samples were screened. Weighted gene co-expression network analysis was applied to cluster DEGs and the key modules related to clinical traits identified. Based on network topology analysis, novel risk genes derived from key modules were mined and biological verification performed. The potential functions of these risk genes were further explored with the aid of miRNA-mRNA regulatory networks. Finally, the prognostic ability of these genes was assessed by constructing a clinical prediction model.

**Results:** Two key modules showed significant association with clinical traits. In combination with protein-protein interaction analysis, 29 hub genes were identified. Among these genes, 19 from one module showed a pattern of upregulation in HCC and were associated with the tumor node metastasis stage, and 10 from the other module displayed the opposite trend. Survival analyses indicated that all these genes were significantly related to patient prognosis. Based on the miRNA-mRNA regulatory network, 29 genes strongly linked to tumor activity were identified. Notably, five of the novel risk genes, ABAT, DAO, PCK2, SLC27A2, and HAO1, have rarely been reported in previous studies. Gene set enrichment analysis for each gene revealed regulatory roles in proliferation and prognosis of HCC. Least absolute shrinkage and selection operator

regression analysis further validated DAO, PCK2, and HAO1 as prognostic factors in an external HCC dataset.

**Conclusion:** Analysis of multiple datasets combined with global network information presents a successful approach to uncover the complex biological mechanisms of HCC. More importantly, this novel integrated strategy facilitates identification of risk hub genes as candidate biomarkers for HCC, which could effectively guide clinical treatments.

**Keywords:** weighted gene co-expression network analysis (WGCNA), hub genes, hepatocellular carcinoma (HCC), biomarker, progression and prognosis

## INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most common malignant tumor type and the fourth leading cause of cancer-related deaths worldwide, with approximately 841,000 new cases and 782,000 deaths each year (Bray et al., 2018). Although multiple therapies have been recently developed for HCC, prognosis remains unsatisfactory due to disease progression, recurrence, and metastasis (Budhu et al., 2006). Abnormal expression of several genes is critical in tumorigenesis and development of HCC. Recent research has shown that tumor necrosis factor- $\alpha$ -induced protein 8 (TNFAIP8) increases HCC cell survival by blocking apoptosis, promoting greater resistance to the anticancer drugs sorafenib and regorafenib (Niture et al., 2020). High expression of ATP/GTP binding protein like 2 (AGBL2) is associated with significantly enhanced survival and proliferation of HCC cells *in vitro* and tumor growth *in vivo* (Wang L. L. et al., 2018). Although these single genes affect the phenotype of HCC, it is not known whether they constitute the hub genes. Integration of multiple datasets and network topology structures may therefore facilitate the identification of more robust biomarkers.

Owing to the substantial improvements in high-throughput gene microarray and next-generation sequencing technologies, bioinformatics analyses are increasingly applied to explore the biological characteristics of cancers. To avoid the potential large bias caused by analysis of a single dataset, many researchers have focused on analysis of multiple datasets for HCC. Recently, Li and colleagues examined the intersection of differentially expressed genes (DEGs) of three datasets (Li and Xu, 2020) and merged the multiple datasets for analysis (Li and Xu, 2020; Li et al., 2020). In the current study, we adopted the Robust Rank Aggregation (RRA) method for the analysis of multiple integrated datasets (Kolde et al., 2012).

We downloaded nine eligible microarray datasets from the Gene Expression Omnibus (GEO), which were subjected to meta-analysis to identify robust DEGs between HCC and matched normal tissues using the RRA method. Next, weighted gene co-expression network analysis (WGCNA) was performed with the DEGs to identify the most significant modules related to clinical traits of HCC. After screening the protein-protein interaction (PPI) network (Szkłarczyk et al., 2015), the 29 hub genes uploaded to miRNet<sup>1</sup> were

screened to construct miRNA-mRNA regulatory networks and explore their potential functions. In an external test dataset from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) collection, 28 of these hub genes were associated with the prognosis and progression of HCC. Gene set enrichment analysis (GSEA) was further performed on the independent dataset (TCGA-LIHC) to determine the potential functions of the identified hub genes. Least absolute shrinkage and selection operator (LASSO) regression was applied to construct clinical predictive models with the aim of verifying the prognostic capability of these genes in patients with HCC. In summary, integrated analysis of multiple datasets was initially conducted, followed by comprehensive screening of hub genes strongly related to HCC using a variety of efficient bioinformatics methods and verification of the results in an external dataset. Our overall findings contribute to the elucidation of the molecular mechanisms underlying pathogenesis and identification of novel prognostic biomarkers for HCC.

## MATERIALS AND METHODS

### Data Sources

We downloaded nine microarray datasets from the GEO database for RRA<sup>2</sup>. Access numbers of the included datasets are as follows: GSE36376 (Lim et al., 2013), GSE39791 (Kim et al., 2014), GSE45114 (Wei et al., 2014), GSE57957 (Mah et al., 2014), GSE60502 (Wang et al., 2014), GSE76297 (Chaisaingmongkol et al., 2017), GSE76427 (Grinchuk et al., 2018), GSE84005, and GSE14520 (Roessler et al., 2010). Datasets were collected up to February 1, 2020, and were included based on the following criteria: (1) gene expression data from HCC and adjacent normal tissue samples were evaluated; (2) at least 15 pairs of tumor and paracancerous tissue samples were assessed; and (3) the number of genes in a single dataset was >10,000. GSE14520 contained adequate clinical information and the largest HCC sample number (471 samples) for WGCNA and LASSO regression. Detailed information on these datasets is provided in **Table 1**. Additionally, the TCGA-LIHC dataset containing 374 HCC and 50 normal samples was utilized as the external validation dataset and GSEA was performed.

<sup>1</sup><https://www.mirnet.ca/>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/gds/?term=>

**TABLE 1** | Details of the eight GEO datasets about HCC.

GEO	No. of samples		Platform	References
	T	N		
GSE36376	249	193	GPL10558	Lim et al., 2013
GSE39791	72	72	GPL10558	Kim et al., 2014
GSE45114	24	25	GPL5918	Wei et al., 2014
GSE57957	39	39	GPL10558	Mah et al., 2014
GSE60502	18	18	GPL96	Wang et al., 2014
GSE76297	61	58	GPL17586	Chaisaingmongkol et al., 2017
GSE76427	115	52	GPL10558	Grinchuk et al., 2018
GSE84005	38	38	GPL5175	NA
GSE14520	471	459	GPL571&GPL3921	Roessler et al., 2010

GEO, Gene Expression Omnibus; GPL, Gene Expression Omnibus Platform; GSE, Gene Expression Omnibus Series; T, tumor samples; N, paracancerous normal samples. There is no reference information in GSE84005. NA, not available.

## Identification of Robust DEGs

The input data of WGCNA is usually less than 5000 genes. Therefore, preliminary screening of genes is required. In addition, DEGs (tumor vs normal tissue) can better reflect the differences in biological characteristics between tumors and normal liver tissues (Sarathi and Palaniappan, 2019). We employed “limma” (R package) to normalize and analyze the differences of each dataset downloaded from the GEO (HCC and normal samples) under a false discovery rate threshold (FDR) < 0.05 (Ritchie et al., 2015). Results from each dataset were ranked according to the fold change value of each gene. Next, “RobustRankAggreg” (R package) was implemented to analyze the results of the nine datasets for the identification of robust DEGs with adjusted *P*-values < 0.05 (Kolde et al., 2012).

## Construction of the WGCNA Network and Enrichment Analysis of Key Modules

Weighted gene co-expression network analysis was used to identify modules highly correlated with clinical traits. We applied “WGCNA” (R package) to cluster all the robust DEGs identified from the GSE14520 HCC dataset with the largest sample size (471 HCC samples) and sufficient clinical information (Langfelder and Horvath, 2008). The resulting adjacency matrix was transformed into a topological overlap matrix (TOM). Differentially expressed genes were subsequently grouped into different modules based on the TOM-based dissimilarity measure. A soft-thresholding power of 7 (scale-free *R* = 0.90) and minimal module size of 30 were applied. The cut height was set as 0.4 to merge similar modules.

After clustering the genetic modules, key modules associated with clinicopathological variables were determined using Pearson’s correlation coefficient, including age, hepatitis B virus (HBV) activity, alanine aminotransferase (ALT) level ( $\leq$  and  $>50$  U/L), primary tumor size ( $\leq$  and  $>5$  cm), multinodular characteristics, cirrhosis, tumor node metastasis (TNM) stage, Barcelona Clinic Liver Cancer (BCLC) stage, Cancer of the Liver Italian Program (CLIP) stage, AFP level ( $\leq$  and  $>300$  ng/mL), survival status, survival time (months), recurrence status, and recurrence time (months). We selected the modules that were

highly correlated with clinical traits. To establish the biological functions of the key modules, R package “clusterprofiler” was applied to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses on individual genes. *P*-values < 0.05 were indicative of significant enrichment.

## Identification of Hub Genes Based on WGCNA Combined With PPI and Construction of miRNA–mRNA Regulatory Networks

After the identification of the key modules, genes with gene significance (GS) > 0.3 and module membership (MM) > 0.8 were taken as core genes in WGCNA. Initially, the top 100 genes with high connectivity from each module were screened, of which the top 30 were marked as “hub genes in WGCNA”. Next, we uploaded the top 100 connectivity genes to the STRING<sup>3</sup> database for PPI network analysis (Szklarczyk et al., 2017). The “TSV: tab separated values” file was downloaded in the “Exports” option and imported into the Cytoscape software (version 3.7.0), whereby the top 30 genes were screened as “hub genes in PPI” by “Degree” using the “cytoHubba” (Chin et al., 2014) app. GeneMANIA is a common tool for PPI network analysis and predicting the functions of preferred genes (Warde-Farley et al., 2010). The program displays genes or gene lists using bioinformatics methods, including gene co-expression, physical interactions, gene co-location, gene enrichment analysis, and website prediction. We observed the interaction types among the hub genes and visualized the gene networks with the aid of GeneMANIA. Finally, the intersecting results of both analytical methods were used to obtain hub genes, which were uploaded to miRNet<sup>4</sup> to generate a miRNA–mRNA regulatory network for establishing their potential functions.

## Verification of Hub Genes

First, GEPIA2 was employed to visualize the differential expression of the hub genes between HCC and normal tissues (one-way ANOVA). Next, we used “ggpubr” (R package) to analyze the expression patterns at different TNM stages (Kruskal–Wallis test). Stage IV samples were excluded owing to a small size of less than five samples. In addition, we used the “survival” R package to perform Kaplan–Meier (K–M) survival or unique Cox regression analysis. Results were validated in the external verification dataset TCGA-LIHC.

## GSEA and LASSO

To further explore the potential functions of the genes rarely reported in HCC, we utilized the “clusterprofiler” R package to perform GSEA for each gene. In the TCGA-LIHC dataset (normalized with the “edgeR” package), 374 HCC samples were used as the gene expression matrix. Gene lists were generated according to the order of correlation with the expression of each hub gene. The C2 reference gene sets were downloaded

<sup>3</sup><https://string-db.org/>

<sup>4</sup><https://www.mirnet.ca/>

from the Molecular Signatures Database (MSigDB)<sup>5</sup>. We set an adjusted *P*-value < 0.05 as the cut-off criterion. LASSO regression is widely used in the construction of clinical prediction models (Tibshirani, 1997). Next, “glmnet” (R package) was applied to verify the potential of these genes as biomarkers. GSE14520 was used as the training set and TCGA-LIHC as the test set for the LASSO regression analysis. Each cohort was divided into two groups according to the best cutoff risk score. Finally, results were visualized with K-M and ROC curves.

## RESULTS

### Overall Study Design

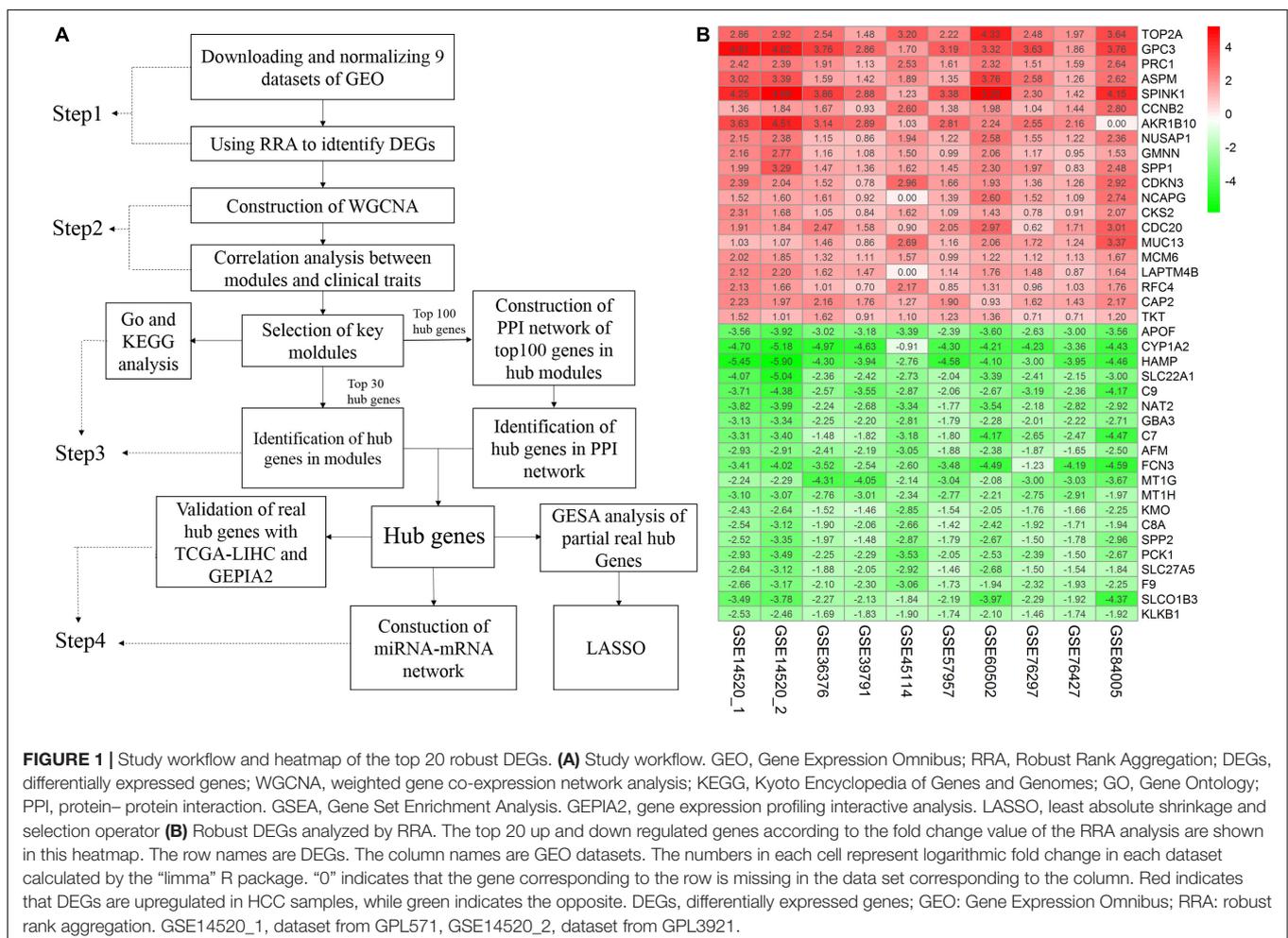
A flow chart of the study, divided into four steps, is presented in **Figure 1A**. Firstly, we used the RRA method to integrate and analyze the nine GEO datasets to obtain robust DEGs (Step 1). These DEGs were used to construct a WGCNA network using the GSE14520 dataset, and the key modules displaying a significant correlation with clinical traits were identified (Step 2). Hub genes

were screened according to the WGCNA and PPI networks (Step 3). Finally, the hub genes were validated (Step 4).

### RRA-Based Identification of Robust DEGs Between HCC and Normal Tissues

A total of 4244 robust DEGs (2674 significantly upregulated and 1570 significantly downregulated) were identified from the nine datasets integrated using RRA (adjusted *P*-value < 0.05). As shown in **Figure 1B**, the 20 most significant DEGs were consistently identified among most of the datasets evaluated, signifying the robustness of the results. The majority of these genes are associated with HCC. For example, TOP2A displaying the most significant upregulation has been identified as a biomarker for HBV-related HCC (Liao et al., 2019) and APOF with the most significant downregulation is considered a tumor suppressor in HCC (Wang Y. B. et al., 2019). Significantly, AKR1B10 was not included in the GSE84005 dataset or NCAPG and LAPTM4B in the GSE45114 dataset. However, close association of these three genes with the progression of HCC has been recently reported (DiStefano and Davis, 2019; Gong et al., 2019; Wang F. et al., 2019). The RRA method effectively maximizes the retention of hub genes.

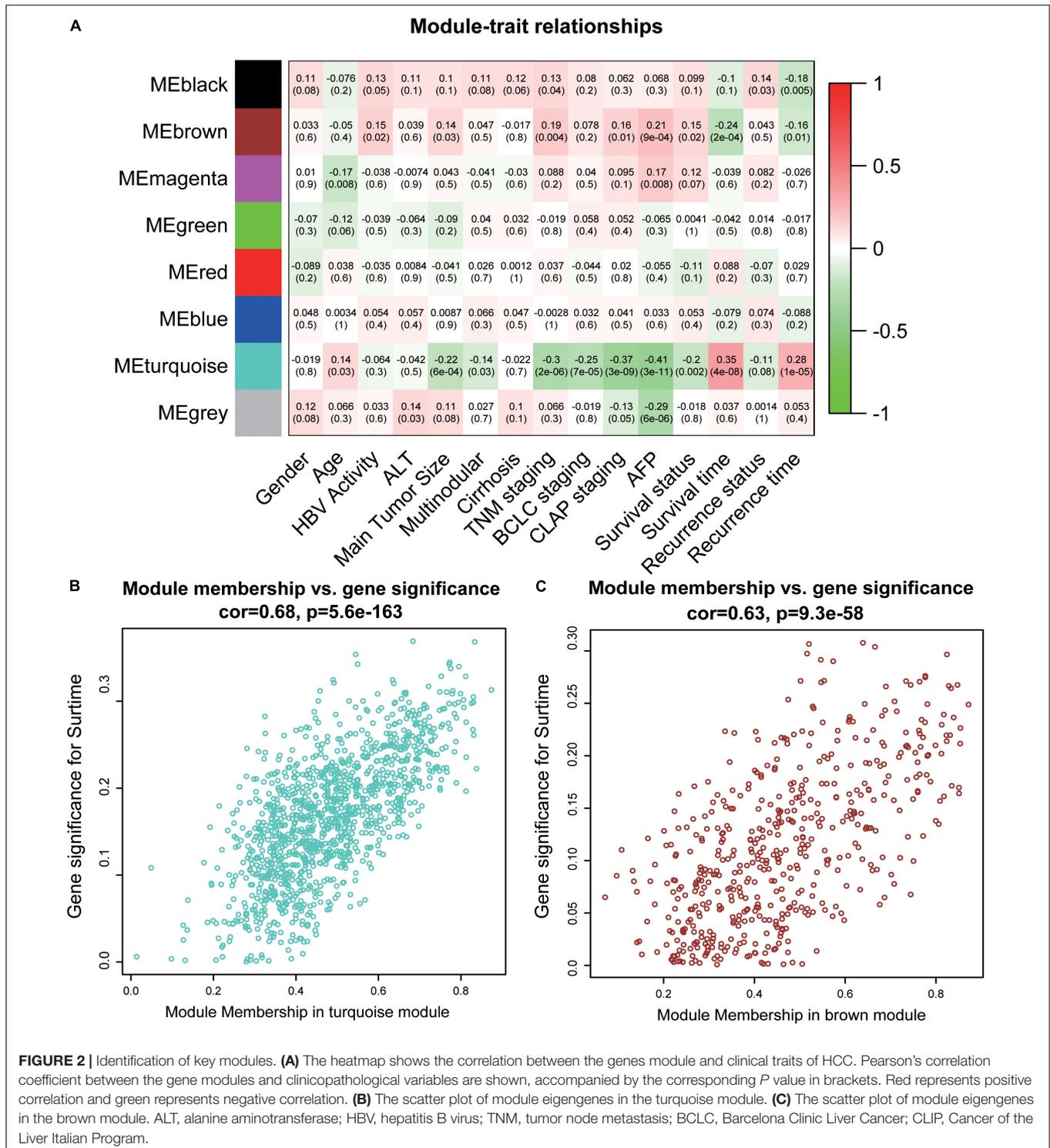
<sup>5</sup><http://software.broadinstitute.org/gsea/msigdb/index.jsp>



### Identification of Key Modules

To acquire the key modules, “WGCNA” (R package) was used to examine the co-expression network with the GSE14520 dataset. All DEGs derived from the RRA analysis were used as input. As shown in **Supplementary Figure 1A**, when the soft-thresholding power was 7 or 8,  $R^2$  was  $>0.9$  (red line). Here, a power of

$\beta = 7$  (scale-free  $R^2 = 0.9$ ) was selected as the soft-thresholding power to ensure a scale-free network. After applying threshold values, a total of eight modules were obtained for subsequent analysis (**Supplementary Figures 2C,D**). As determined from evaluation of module-trait relationships (**Figure 2A**), the brown and turquoise modules showed greater significance in relation



to clinical information, compared with the other modules, in particular, main tumor size, TNM stage, and AFP level critical for prognosis of HCC patients (Han et al., 2014; Zhang et al., 2016) (Figures 2B,C).

## Functional Enrichment Analysis of Genes Within the Key Modules and Identification of Hub Genes

To clarify the functions of genes from the two modules, we performed separate GO and KEGG analyses. In the brown module, “DNA replication,” “cell cycle,” “p53 signaling pathway,” and “cellular senescence” were enriched in the KEGG pathway analysis (Supplementary Figure 2A) while in the turquoise module, “drug metabolism – cytochrome P450” and “chemical carcinogenesis” were enriched (Supplementary Figure 2B). These findings were consistent with previous studies reporting the involvement of the above functions in tumorigenesis of HCC. For example, Xie et al. (2018) showed that DNA replication is associated with tumor cell proliferation and prognosis of patients with HCC. Moreover, genetic variations in cell cycle pathway genes affect the disease-free survival of patients with HCC (Liu et al., 2017). The TP53 mutation is considered one of the molecular mechanisms of HCC pathogenesis (Hussain et al., 2007). Abnormal cellular senescence is a characteristic phenotype of various cancers (Chen S. L. et al., 2019). Cytochrome P450 is severely damaged and dysregulated in HCC (Yan et al., 2015). The collective findings validate the functional association of the key modules in this study with HCC. The significant biological process (BP), cellular component (CC), and molecular function (MF) GO terms of the two modules are presented in Supplementary Tables 1–6.

To further screen for the most significant hub genes, we used a combination of two methods (WGCNA and PPI networks, see section “Materials and Methods”). The PPI network of the top 100 connectivity genes from the brown module is shown in Figure 3A. According to degree (high to low), the positions of genes are arranged from the inside to outside, and the top 30 considered “hub genes in PPI”. Interaction analysis of hub genes in PPI was further performed using GeneMANIA to clarify the correlations among colocalization, shared protein domains, co-expression, prediction, and pathways. As revealed by the protein–protein interaction network generated with GeneMANIA (Figure 3C), co-expression interactions accounted for the largest proportion (83.83%), consistent with the results of WGCNA. “Hub genes in WGCNA” and their correlated expression levels are shown in Figure 3B. The hub genes were obtained by selecting the intersecting results with the two methods (Figure 3D). The hub genes of the turquoise module were obtained with the same method (Supplementary Figures 3A–D). Overall, we identified a total of 29 core genes from the two key modules.

## Construction of the miRNA–mRNA Regulatory Network

Interactions between miRNA and mRNA are an increasing focus of research attention. To further explore the functions of hub

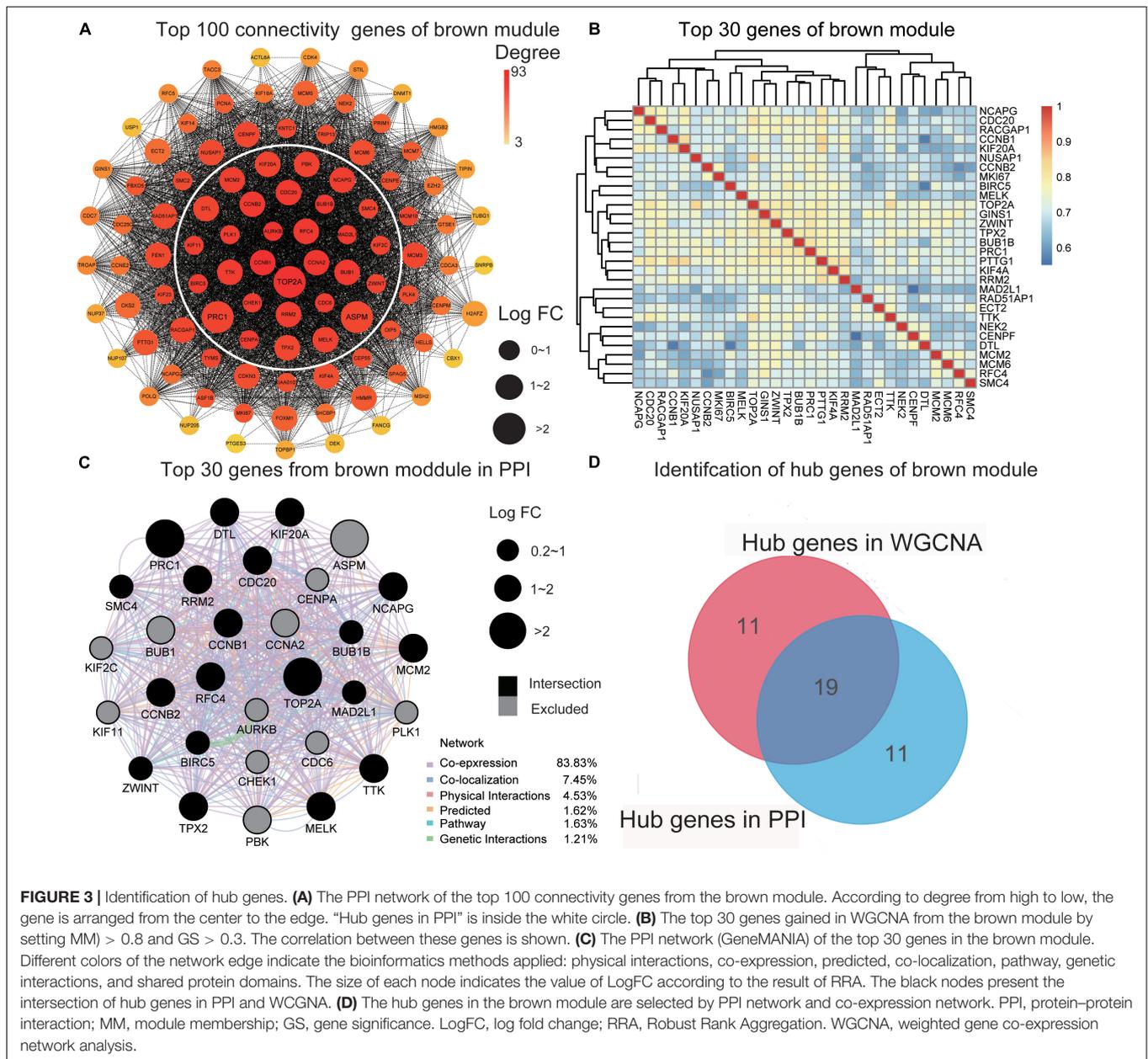
genes from a global perspective, a miRNA–mRNA regulatory network was constructed via miRNet (Figure 4). Previous studies suggest that a number of these miRNAs are related to HCC. For example, exosome hsa-miR-335 was identified as a therapeutic target for HCC (Wang F. et al., 2018). Furthermore, according to web-based KEGG analysis, this network is enriched in multiple tumor-related pathways (Supplementary Table 7), such as cell cycle and p53 signaling (Hussain et al., 2007; Sanchez-Vega et al., 2018; Ikeno et al., 2019). Thus our group of hub genes may play important roles in HCC through the miRNA–mRNA regulatory network.

## Verification of Hub Genes Based on the TCGA-LIHC Dataset

In total, 29 hub genes were obtained. Interestingly, TOP2A was consistently ranked first. Ten of the genes were filtered from the turquoise module (Figure 3C), which were further verified in TCGA-LIHC and GEPIA2 based on three parameters: (1) differential expression (HCC sample vs paracancerous sample), (2) TMN staging, and (3) survival analysis. In terms of expression, hub genes from the turquoise module were downregulated in HCC relative to normal samples. Notably, F13B was excluded due to lack of statistical significance. These results were validated using an external dataset (Figure 5A and Supplementary Figure 4A). Additionally, genes were differentially expressed in HCC samples with different TNM stages to a significant extent. A higher expression of these genes was correlated with an earlier TNM stage (Figure 5B and Supplementary Figure 4B). Survival analysis revealed an association of low expression of these genes with poor prognosis (Figure 5C and Supplementary Table 8). Using the same method, hub genes of the brown module (Supplementary Figure 3D) were validated, which showed an opposite trend to genes of the turquoise module (Supplementary Figures 5, 6 and Supplementary Table 8). Our collective data support critical roles of 28 of the 29 hub genes in HCC.

## GSEA of Tumor Suppressor Roles of Hub Genes

The majority of the hub genes for HCC have already been reported (Supplementary Table 9). However, DAO, SLC27A2, GYS2, HAO1, and PCK2 have not been previously studied in association with HCC. To analyze their potential functions in HCC, we performed GSEA on TCGA-LIHC RNA sequencing data. As shown in Supplementary Figure 7, three gene sets associated with tumors were defined. In samples showing a significant negative correlation of HAO1 and SLC27A2 expression with HCC, “epithelial-mesenchymal transition” (EMT) and “PI3K/Akt/mTOR” were enriched. “Wnt/beta-catenin signaling” and “MYC target v1” were significantly enriched in samples showing a negative correlation of PCK2 and DAO expression with HCC. The gene set “DNA repair” was enriched in samples showing negative correlation of ABAT and SLC27A2 expression with HCC. These mechanisms are typical tumor-associated pathways. For instance, EMT is reported to coordinate the occurrence of liver fibrosis, carcinogenesis, and proliferation and invasion of HCC cells



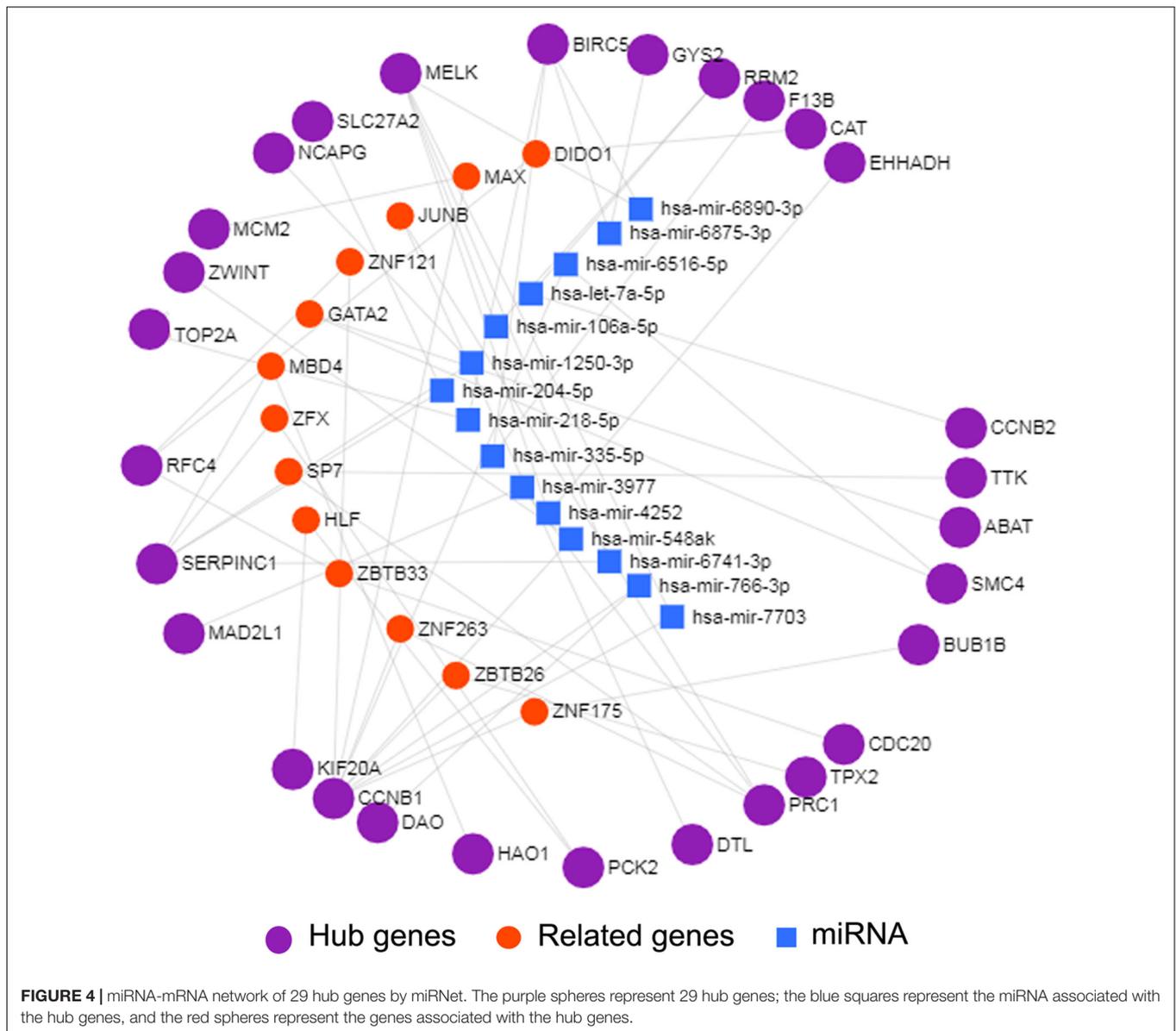
**FIGURE 3 |** Identification of hub genes. **(A)** The PPI network of the top 100 connectivity genes from the brown module. According to degree from high to low, the gene is arranged from the center to the edge. “Hub genes in PPI” is inside the white circle. **(B)** The top 30 genes gained in WGCNA from the brown module by setting  $MM > 0.8$  and  $GS > 0.3$ . The correlation between these genes is shown. **(C)** The PPI network (GeneMANIA) of the top 30 genes in the brown module. Different colors of the network edge indicate the bioinformatics methods applied: physical interactions, co-expression, predicted, co-localization, pathway, genetic interactions, and shared protein domains. The size of each node indicates the value of LogFC according to the result of RRA. The black nodes present the intersection of hub genes in PPI and WGCNA. **(D)** The hub genes in the brown module are selected by PPI network and co-expression network. PPI, protein–protein interaction; MM, module membership; GS, gene significance. LogFC, log fold change; RRA, Robust Rank Aggregation. WGCNA, weighted gene co-expression network analysis.

(Giannelli et al., 2016). The activation of PI3K/AKT signaling has been shown to promote EMT (Liu et al., 2018). “Wnt/beta-catenin signaling”, “MYC target v1”, and “DNA repair” are closely related to tumorigenesis and the development of HCC (Dolezal et al., 2017; Dimri and Satyanarayana, 2020; Pardini et al., 2020). Taken together, the findings clearly suggest that these genes are closely associated with the mechanisms underlying HCC cell proliferation.

### Construction of the Novel Hub Gene Signature for Survival Prediction

Finally, we included the above five hub genes in the LASSO regression analysis to construct a survival prediction model

for HCC patients. GSE14520 was used as the training set to generate a prediction model comprising three of the genes, specifically, *OSPCK2*, *DAO*, and *HAO1*. The formula for calculating the prognostic risk score was as follows:  $(-0.0179 \times \text{expression HAO1}) + (-0.0221 \times \text{expression PCK2}) + (-0.1209 \times \text{expression DAO})$ . The results of this scoring system were depicted using a K–M curve (Figures 6A,B). The high-risk group had shorter OS, both in the training ( $P = 0.002$ ) and test ( $P < 0.001$ ) datasets. In addition, we generated time-dependent ROC curves to evaluate the predictive effects of the three-gene signature based on the area under the curve (AUC) value. In the training cohort, one-year and three-year AUC values were 0.673 and 0.605, respectively. In the verification cohort, AUC for one year was 0.605 and that for



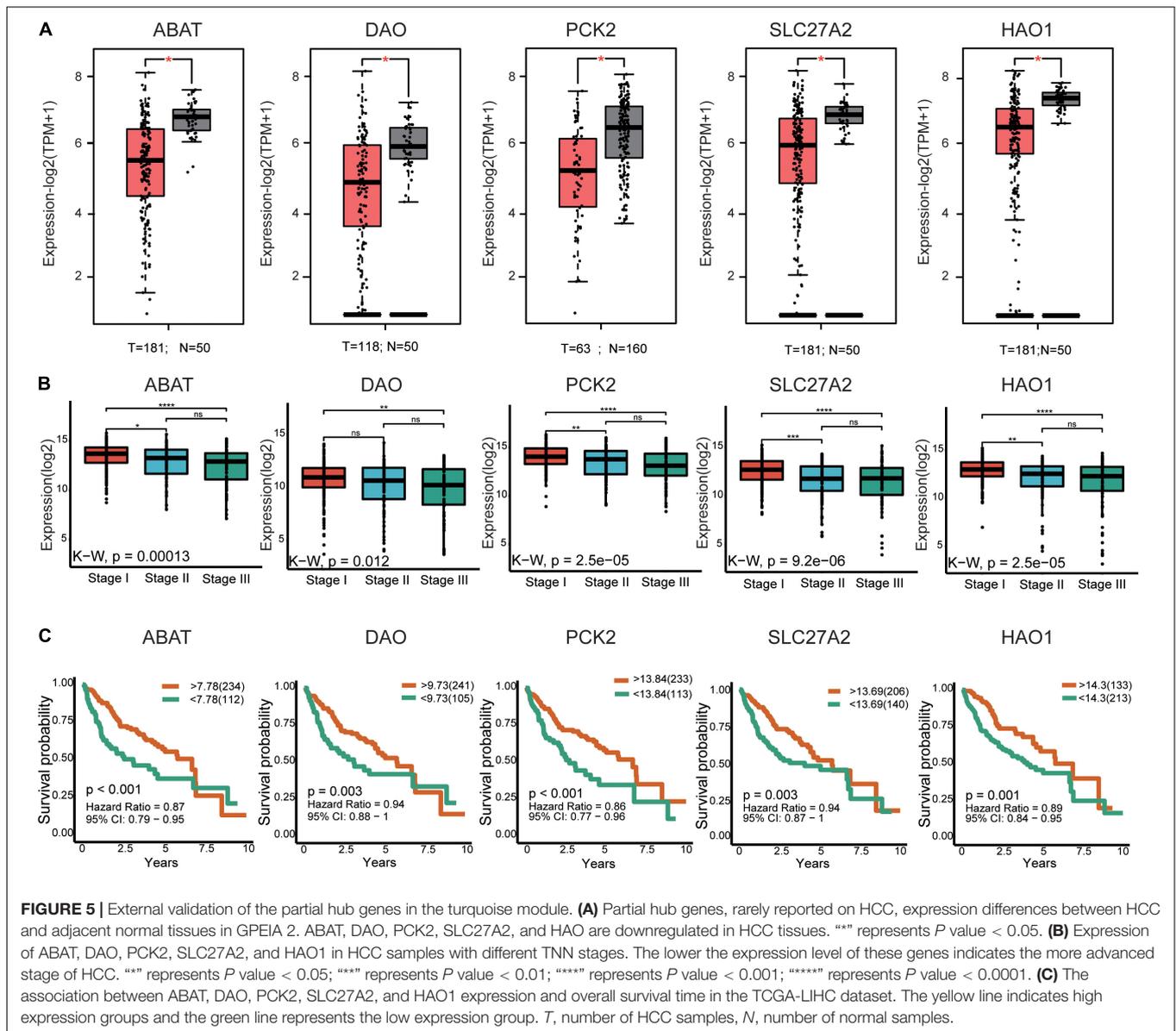
three years was 0.672 (Figures 6C,D). Based on the results, we propose that this novel three-gene signature can serve as a reliable predictor of OS in HCC patients.

## DISCUSSION

In this study, we used multiple bioinformatics methods to establish the biological mechanisms of HCC. To avoid the potential bias caused by DEGs in a single database, numerous studies have focused on evaluating multiple datasets (Xu et al., 2016). In the process of merging data, gene symbols that are not detected in only one dataset may be lost. For example, as shown in Figure 1B, AKR1B10, NCAPG, and LAPTM4B exist in multiple datasets and would therefore be lost if the datasets were simply merged. However, these genes are closely

related to the progression of HCC (DiStefano and Davis, 2019; Gong et al., 2019; Wang F. et al., 2019). Furthermore, in dataset GSE39791, logFC values of some of the top 20 DEGs were less than 1. However, in combination with other datasets, the RRA method suggests that these genes are robust DEGs. Potential bias of results due to inclusion of only one dataset should be avoided. In the current investigation, RRA was applied to analyze nine groups of datasets to minimize bias, avoid missing hub genes, and obtain the most robust DEGs.

Weighted gene co-expression network analysis is based on the correlation between modules and clinical features, and the screening results are highly reliable and biologically meaningful (Yin et al., 2018). To our knowledge, the current study is the first to combine the RRA method with WGCNA for efficient identification of hub genes associated with HCC. Among the

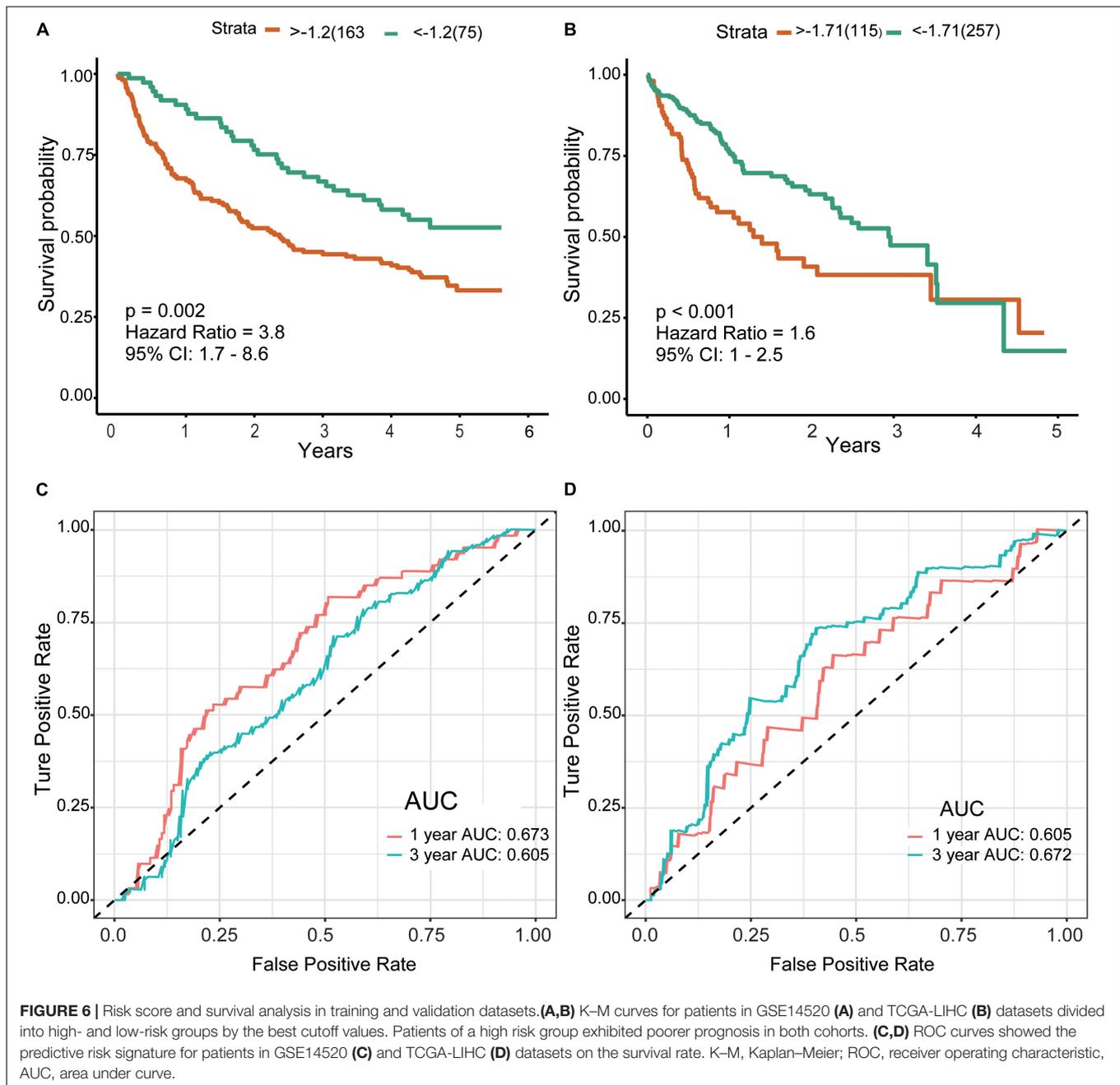


eight gene modules, brown and turquoise modules were closely related to clinical characteristics, such as primary tumor size, AFP level, TNM stage, and overall survival time. In addition, GO and KEGG analyses showed enrichment of both modules in multiple tumor-related pathways. For instance, DNA replication is associated with tumor cell proliferation and the prognosis of HCC (Xie et al., 2018), variations in cell cycle pathway genes affect disease-free survival of patients with HCC (Liu et al., 2017), TP53 mutation is considered one of the critical molecular mechanisms of HCC pathogenesis (Hussain et al., 2007), an abnormal cellular senescence phenotype is observed in various cancer types (Chen S. L. et al., 2019), and cytochrome P450 is severely damaged and dysregulated in HCC (Yan et al., 2015).

Next, we combined co-expression and PPI networks to screen for hub genes. After a series of strict screening steps, 29 hub genes (10 from the turquoise module and 19 from the brown

module) were isolated. To explore the functions of this group of genes from the global network, miRNA-mRNA regulatory networks were generated using miRNet (Figure 4). As shown in Supplementary Table 7, specific pathways, such as cell cycle, and p53 signaling, were highlighted, both of which are closely related to tumorigenesis and the development of HCC (Hussain et al., 2007; Sanchez-Vega et al., 2018; Ikeno et al., 2019). Importantly, we used TCGA-LIHC, a dataset containing 374 HCC samples, to validate the predictive power of these hub genes in the progression and prognosis of HCC. Among the genes examined, only one (F13B) failed verification.

The involvement of the majority of these genes in HCC has been confirmed in earlier experiments (Supplementary Table 9), supporting the efficacy of our screening strategy. Among the hub genes, TOP2A, RFC, and the CCMB family have received considerable research attention. DNA topoisomerase



II alpha (TOP2A) is abundantly expressed in testis, lymph node tissues, and a variety of tumor tissues, including liver cancer. Several bioinformatics analyses have validated TOP2A as a biomarker for HCC, in particular, HBV-related HCC (Liao et al., 2019). Panvichian et al. (2015) reported overexpression of TOP2A in 72.5% of tumor tissues and its significant association with the hepatitis B surface antigen (HBsAg) in serum. In addition, results of a phase III prospective randomized study showed that TOP2A is associated with the histological grade of liver cancer, microvascular invasion, early onset of malignant tumors ( $\leq 40$  years), and chemotherapy resistance (Wong et al., 2009). Replication factor C subunit 4 (RFC4) has recently been

identified as a hub gene affecting prognosis of patients with HCC (Kong et al., 2019). The knockdown of endogenous RFC4 suppresses HCC cell growth and enhances the chemosensitivity of HepG2 cells (Arai et al., 2009). Cyclin B1 (CCNB1) and TOP2A are considered key genes for early diagnosis of HCC (Wu et al., 2019).

Interestingly, DAO, ABAT, SL27AL, PCK2, and HAO1, all from the turquoise module, have not been shown to be associated with HCC to date, either *in vivo* or *in vitro*. However, several studies support inhibitory roles of these genes in other tumors. The peroxisomal enzyme D-amino acid oxidase (DAO) is highly expressed in the kidney, liver, and brain in mammals

(Fang et al., 2008) and plays a critical role in the pathophysiology of schizophrenia (Liu et al., 2016). Earlier reports suggest that DAO inhibits glioma cell growth by inhibiting angiogenesis (El Sayed et al., 2012) and inducing apoptosis (Li et al., 2008). 4-Aminobutyrate aminotransferase (ABAT) is mainly responsible for decomposing  $\gamma$ -aminobutyric acid (GABA), an inhibitory neurotransmitter, into succinic semialdehyde. In basal-like breast cancer (BLBC) cells, GABA increases the intracellular  $\text{Ca}^{2+}$  concentration and effectively activates nuclear factor 1-4 (NFAT1). Consequently, ABAT expression inhibits the tumorigenicity and metastasis of BLBC cells *in vitro* and *in vivo*. Conversely, the downregulation of ABAT promotes the progression of BLBC (Chen X. et al., 2019) and resistance to endocrine therapy of inflammatory breast cancer (Jansen et al., 2015). Moreover, ABAT has been identified as a prognostic factor for renal cell carcinoma and hepatic adenocarcinoma (Reis et al., 2015; Lu et al., 2020). Chen et al. (2017) screened six genes related to HCC metastasis and prognosis through a co-expression network analysis, which led to the identification of DAO and ABAT. However, their mechanisms of action in HCC have not been clarified. Solute carrier family 27 member 2 (SLC27A2), also designated FATP2, improves the efficiency of cancer therapy by inhibiting the activity of polymorphonuclear myeloid-derived suppressor cells (PMN-MDSCs) (Veglia et al., 2019). Phosphoenolpyruvate carboxykinase 2 (PCK2) encodes a key mitochondrial enzyme for gluconeogenesis in the liver. The overexpression of PCK2 inhibits melanoma cell growth *in vitro* and prevents tumorigenesis *in vivo* (Luo et al., 2017). More recent experiments have demonstrated an association of decreased PCK2 expression with metastasis and the recurrence of osteosarcoma (Zhang et al., 2019). Upon suppression of autophagy, levels of glucose-6-phosphatase (G6Pase) and phosphoenolpyruvate carboxykinase (PEPCK, a protein encoded by PCK2) are reduced in the human HCC cell line HepG2 (Jeon et al., 2015). Hypoxia-inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) can promote the growth of human breast tumor-repopulating cells by downregulating PCK2 (Tang et al., 2019). However, a number of studies have reported that PEPCK coordinates the regulation of central carbon metabolism to promote tumor cell growth (Montal et al., 2015). Therefore, the biological characteristics of PCK2 in HCC requires further investigation. Hydroxyacid oxidase 1 (HAO1) is expressed mainly in the liver and pancreas. An earlier genome-wide association study in Korea showed that a single nucleotide polymorphism in HAO1 is one of the risk factors for childhood acute lymphoblastic leukemia (Han et al., 2010).

In our study, GSEA consistently supported the tumor suppressor roles of these genes in multiple carcinogenic pathways in HCC datasets. Further research is warranted to establish the mechanisms of action of these genes in HCC. The collective evidence to date suggests that these genes play a suppressive roles in the biological processes of tumors. In addition, the clinical prediction model generated using a three-gene signature showed efficacy in predicting the survival of patients with HCC and the potential as a robust biomarker. Our study has some limitations, such as the fact that the nine datasets of the training set are all

microarrays and lack RNA-seq datasets. The data diversity is insufficient.

## CONCLUSION

Systematic analysis of the genes involved in pathogenesis of HCC using a novel integrated strategy led to the identification of two risk modules and several representative hub genes. Among these, HAO1, SCL27A2, DAO, ABAT, and PCK2, rarely reported in HCC to date, were validated as novel hub genes that may serve as effective clinical diagnostic and prognostic markers as well as therapeutic targets for HCC.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The datasets analyzed in the current study are available in the TCGA repository (<http://cancergenome.nih.gov/>) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

## AUTHOR CONTRIBUTIONS

YY contributed to the project design. CN and CZ revised and verified the article. ND, SL, JL, and AX provided the bioinformatics method support and reviewed the manuscript. HS analyzed the data and wrote the article. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by findings from the National Natural Scientific Foundation of China (No. 31701145).

## ACKNOWLEDGMENTS

In this study, the authors express their gratitude to the scholars and institutions that submitted the raw datasets.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00895/full#supplementary-material>

**Supplementary Figure 1** | Determination of soft-thresholding power and cut height in the WGCNA. **(A)** Analysis of scale-free index (left) and mean connectivity (right) for different soft-thresholding power ( $\beta$ ) red line indicates signed  $R^2 = 0.9$ . **(B)** When  $\beta = 7$  histogram of connectivity distribution (left) and scale-free topology  $R^2 = 0.9$  (right). **(C)** Clustering of module eigengenes. Set the cut height as 0.4 (red line) to merge similar modules. **(D)** Dendrogram of all DEGs clustered based on a dissimilarity measure (1-TOM). Each color represents a set of gene modules.

**Supplementary Figure 2** | The screening of hub genes in the turquoise module. **(A)** The PPI network of top 100 connectivity genes from the brown module. "Hub

genes in PPI" is inside the black circle. **(B)** The top 30 hub genes gained in WGCNA from the brown module by setting  $MM > 0.8$  and  $GS > 0.3$ . Correlation between these genes is shown. **(C)** PPI network (GeneMANIA) of the top 30 genes in the brown module. **(D)** Selection of hub genes that occur in both the PPI network and WGCNA. PPI, protein–protein interaction; MM, module membership; GS, gene significance. WGCNA, weighted gene co-expression network analysis.

**Supplementary Figure 3** | KEGG analysis for the key modules. **(A)** Brown module. **(B)** Turquoise module. KEGG, Kyoto Encyclopedia of Genes and Genomes.

**Supplementary Figure 4** | External validation of the rest of the hub genes in the turquoise module. **(A)** The rest of the hub genes in the turquoise module expression differences between HCC and adjacent normal tissues in GPEIA2. "\*" represents  $P$  value  $< 0.05$ . **(B)** Expression of CAT, F13B, EHHADH, GYC2, and SERPINC1 in HCC samples with different TNN stages. "\*" represents  $P$  value  $< 0.05$ ; "\*\*\*" represents  $P$  value  $< 0.01$ ; "\*\*\*\*" represents  $P$  value  $< 0.001$ .

**Supplementary Figure 5** | External validation of hub genes in the brown module. The hub genes from the brown module expression differences between HCC and adjacent normal tissues in GPEIA2. "\*" represents  $P$  value  $< 0.05$ .

**Supplementary Figure 6** | External validation of hub genes in the brown module. Expression of these genes in HCC samples with different TNN stages. "\*" represents  $P$  value  $< 0.05$ ; "\*\*\*" represents  $P$  value  $< 0.01$ ; "\*\*\*\*" represents  $P$  value  $< 0.001$ .

## REFERENCES

- Arai, M., Kondoh, N., Imazeki, N., Hada, A., Hatsuse, K., Matsubara, O., et al. (2009). The knockdown of endogenous replication factor C4 decreases the growth and enhances the chemosensitivity of hepatocellular carcinoma cells. *Liver Int.* 29, 55–62. doi: 10.1111/j.1478-3231.2008.01792.x
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492
- Budhu, A., Forgues, M., Ye, Q. H., Jia, H. L., He, P., Zanetti, K. A., et al. (2006). Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. *Cancer Cell* 10, 99–111. doi: 10.1016/j.ccr.2006.06.016
- Chaisaingmongkol, J., Budhu, A., Dang, H., Rabibhadana, S., Pupacdi, B., Kwon, S. M., et al. (2017). Common molecular subtypes among asian hepatocellular carcinoma and cholangiocarcinoma. *Cancer Cell* 32, 57.e3–70.e3. doi: 10.1016/j.ccell.2017.05.009
- Chen, P., Wang, F., Feng, J., Zhou, R., Chang, Y., Liu, J., et al. (2017). Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. *Oncotarget* 8, 48948–48958. doi: 10.18632/oncotarget.16896
- Chen, S. L., Zhang, C. Z., Liu, L. L., Lu, S. X., Pan, Y. H., Wang, C. H., et al. (2019). A GYS2/p53 negative feedback loop restricts tumor growth in HBV-related hepatocellular carcinoma. *Cancer Res.* 79, 534–545. doi: 10.1158/0008-5472.CAN-18-2357
- Chen, X., Cao, Q., Liao, R., Wu, X., Xun, S., Huang, J., et al. (2019). Loss of ABAT-mediated GABAergic system promotes basal-like breast cancer progression by activating Ca<sup>2+</sup>-NFAT1 axis. *Theranostics* 9, 34–47. doi: 10.7150/thno.29407
- Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8(Suppl. 4):S11. doi: 10.1186/1752-0509-8-S4-S11
- Dimri, M., and Satyanarayana, A. (2020). Molecular signaling pathways and therapeutic targets in hepatocellular carcinoma. *Cancers* 12:491. doi: 10.3390/cancers12020491
- DiStefano, J. K., and Davis, B. (2019). Diagnostic and prognostic potential of akr1b10 in human hepatocellular carcinoma. *Cancers* 11:486. doi: 10.3390/cancers11040486
- Dolezal, J. M., Wang, H., Kulkarni, S., Jackson, L., Lu, J., Ranganathan, S., et al. (2017). Sequential adaptive changes in a c-Myc-driven model of hepatocellular carcinoma. *J. Biol. Chem.* 292, 10068–10086. doi: 10.1074/jbc.M117.782052
- Supplementary Figure 7** | Gene sets related to cancer. Results of GSEA related to cancer in samples negatively correlated with PCK2 **(A)**, ABAT **(B)**, HAO1 **(C)**, SLC27A2 **(D)**, and DAO **(E)** expression. Highlight 3 gene sets for each gene.
- Supplementary Figure 8** | Risk score distribution, survival status, and heatmaps for patients in the GSE14520 **(A)** and TCGA-LIHC **(B)** datasets divided into high- and low-risk groups.
- Supplementary Figure 9** | Node degree distribution plot of differentially expressed genes.
- Supplementary Table 1** | BP of GO analysis for brown module.
- Supplementary Table 2** | CC of GO analysis for brown module.
- Supplementary Table 3** | MF of GO analysis for brown module.
- Supplementary Table 4** | BP of GO analysis for turquoise module.
- Supplementary Table 5** | CC of GO analysis for turquoise module.
- Supplementary Table 6** | MF of GO analysis for turquoise module.
- Supplementary Table 7** | Enriched function of the miRNA-mRNA network by miRNet.
- Supplementary Table 8** | Survival analysis for the rest of the hub genes.
- Supplementary Table 9** | Biological functions of the hub genes in HCC.
- El Sayed, S. M., El-Magd, R. M., Shishido, Y., Yorita, K., Chung, S. P., Tran, D. H., et al. (2012). D-Amino acid oxidase-induced oxidative stress, 3-bromopyruvate and citrate inhibit angiogenesis, exhibiting potent anticancer effects. *J. Bioenerg. Biomembr.* 44, 513–523. doi: 10.1007/s10863-012-9455-y
- Fang, J., Deng, D., Nakamura, H., Akuta, T., Qin, H., Iyer, A. K., et al. (2008). Oxystress inducing antitumor therapeutics via tumor-targeted delivery of PEG-conjugated D-amino acid oxidase. *Int. J. Cancer* 122, 1135–1144. doi: 10.1002/ijc.22982
- Giannelli, G., Koudelkova, P., Dituri, F., and Mikulits, W. (2016). Role of epithelial to mesenchymal transition in hepatocellular carcinoma. *J. Hepatol.* 65, 798–808. doi: 10.1016/j.jhep.2016.05.007
- Gong, C., Ai, J., Fan, Y., Gao, J., Liu, W., Feng, Q., et al. (2019). NCAPG promotes the proliferation of hepatocellular carcinoma through PI3K/AKT signaling. *Onco. Targets Ther.* 12, 8537–8552. doi: 10.2147/OTT.S217916
- Grinchuk, O. V., Yenamandra, S. P., Iyer, R., Singh, M., Lee, H. K., Lim, K. H., et al. (2018). Tumor-adjacent tissue co-expression profile analysis reveals prognostic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. *Mol. Oncol.* 12, 89–113. doi: 10.1002/1878-0261.12153
- Han, J. H., Kim, D. G., Na, G. H., Kim, E. Y., Lee, S. H., Hong, T. H., et al. (2014). Evaluation of prognostic factors on recurrence after curative resections for hepatocellular carcinoma. *World J. Gastroenterol.* 20, 17132–17140. doi: 10.3748/wjg.v20.i45.17132
- Han, S., Lee, K. M., Park, S. K., Lee, J. E., Ahn, H. S., Shin, H. Y., et al. (2010). Genome-wide association study of childhood acute lymphoblastic leukemia in Korea. *Leuk. Res.* 34, 1271–1274. doi: 10.1016/j.leukres.2010.02.001
- Hussain, S. P., Schwank, J., Staib, F., Wang, X. W., and Harris, C. C. (2007). TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene* 26, 2166–2176. doi: 10.1038/sj.onc.1210279
- Ikeno, S., Nakano, N., Sano, K., Minowa, T., Sato, W., Akatsu, R., et al. (2019). PDZK1-interacting protein 1 (PDZK1IP1) traps Smad4 protein and suppresses transforming growth factor- $\beta$  (TGF- $\beta$ ) signaling. *J. Biol. Chem.* 294, 4966–4980. doi: 10.1074/jbc.RA118.004153
- Jansen, M. P., Sas, L., Sieuwerts, A. M., Van Cauwenberghe, C., Ramirez-Ardila, D., Look, M., et al. (2015). Decreased expression of ABAT and STC2 hallmarks ER-positive inflammatory breast cancer and endocrine therapy resistance in advanced disease. *Mol. Oncol.* 9, 1218–1233. doi: 10.1016/j.molonc.2015.02.006
- Jeon, J. Y., Lee, H., Park, J., Lee, M., Park, S. W., Kim, J. S., et al. (2015). The regulation of glucose-6-phosphatase and phosphoenolpyruvate carboxykinase by autophagy in low-glycolytic hepatocellular carcinoma cells. *Biochem. Biophys. Res. Commun.* 463, 440–446. doi: 10.1016/j.bbrc.2015.05.103

- Kim, J. H., Sohn, B. H., Lee, H. S., Kim, S. B., Yoo, J. E., Park, Y. Y., et al. (2014). Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med.* 11:e1001770. doi: 10.1371/journal.pmed.1001770
- Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* 28, 573–580. doi: 10.1093/bioinformatics/btr709
- Kong, J., Wang, T., Zhang, Z., Yang, X., Shen, S., and Wang, W. (2019). Five core genes related to the progression and prognosis of hepatocellular carcinoma identified by analysis of a coexpression network. *DNA Cell Biol.* 38, 1564–1576. doi: 10.1089/dna.2019.4932
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Li, C., and Xu, J. (2020). Identification of potentially therapeutic target genes of hepatocellular carcinoma. *Int. J. Environ. Res. Public Health* 17:1053. doi: 10.3390/ijerph17031053
- Li, H., Wei, N., Ma, Y., Wang, X., Zhang, Z., Zheng, S., et al. (2020). Integrative module analysis of HCC gene expression landscapes. *Exp. Ther. Med.* 19, 1779–1788. doi: 10.3892/etm.2020.8437
- Li, J., Shen, Y., Liu, A., Wang, X., and Zhao, C. (2008). Transfection of the DAAO gene and subsequent induction of cytotoxic oxidative stress by D-alanine in 9L cells. *Oncol. Rep.* 20, 341–346.
- Liao, X., Yu, T., Yang, C., Huang, K., Wang, X., Han, C., et al. (2019). Comprehensive investigation of key biomarkers and pathways in hepatitis B virus-related hepatocellular carcinoma. *J. Cancer* 10, 5689–5704. doi: 10.7150/jca.31287
- Lim, H. Y., Sohn, I., Deng, S., Lee, J., Jung, S. H., Mao, M., et al. (2013). Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. *Ann. Surg. Oncol.* 20, 3747–3753. doi: 10.1245/s10434-013-3070-y
- Liu, S., Yang, T. B., Nan, Y. L., Li, A. H., Pan, D. X., Xu, Y., et al. (2017). Genetic variants of cell cycle pathway genes predict disease-free survival of hepatocellular carcinoma. *Cancer Med.* 6, 1512–1522. doi: 10.1002/cam4.1067
- Liu, X., Liao, W., Yuan, Q., Ou, Y., and Huang, J. (2015). TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. *Oncotarget* 6, 34309–34320. doi: 10.18632/oncotarget.5295
- Liu, Y. L., Wang, S. C., Hwu, H. G., Fann, C. S., Yang, U. C., Yang, W. C., et al. (2016). Haplotypes of the D-amino acid oxidase gene are significantly associated with schizophrenia and its neurocognitive deficits. *PLoS One* 11:e0150435. doi: 10.1371/journal.pone.0150435
- Liu, Z., Chen, M., Xie, L. K., Liu, T., Zou, Z. W., Li, Y., et al. (2018). CLCA4 inhibits cell proliferation and invasion of hepatocellular carcinoma by suppressing epithelial-mesenchymal transition via PI3K/AKT signaling. *Aging* 10, 2570–2584. doi: 10.18632/aging.101571
- Lu, J., Chen, Z., Zhao, H., Dong, H., Zhu, L., Zhang, Y., et al. (2020). ABAT and ALDH6A1, regulated by transcription factor HNF4A, suppress tumorigenic capability in clear cell renal cell carcinoma. *J. Transl. Med.* 18:101. doi: 10.1186/s12967-020-02268-1
- Luo, S., Li, Y., Ma, R., Liu, J., Xu, P., Zhang, H., et al. (2017). Downregulation of PCK2 remodels tricarboxylic acid cycle in tumor-repopulating cells of melanoma. *Oncogene* 36, 3609–3617. doi: 10.1038/ncr.2016.520
- Mah, W. C., Thurnherr, T., Chow, P. K., Chung, A. Y., Ooi, L. L., Toh, H. C., et al. (2014). Methylation profiles reveal distinct subgroup of hepatocellular carcinoma patients with poor prognosis. *PLoS One* 9:e104158. doi: 10.1371/journal.pone.0104158
- Montal, E. D., Dewi, R., Bhalla, K., Ou, L., Hwang, B. J., Ropell, A. E., et al. (2015). PEPCk coordinates the regulation of central carbon metabolism to promote cancer cell growth. *Mol. Cell* 60, 571–583. doi: 10.1016/j.molcel.2015.09.025
- Niture, S., Gyamfi, M. A., Lin, M., Chimeh, U., Dong, X., Zheng, W., et al. (2020). TNFAIP8 regulates autophagy, cell steatosis, and promotes hepatocellular carcinoma cell proliferation. *Cell Death Dis.* 11:178. doi: 10.1038/s41419-020-2369-4
- Panvichian, R., Tantiwetueangdet, A., Angkathunyakul, N., and Leelaudomlpi, S. (2015). TOP2A amplification and overexpression in hepatocellular carcinoma tissues. *Biomed. Res. Int.* 2015:381602. doi: 10.1155/2015/381602
- Pardini, B., Corrado, A., Paolicchi, E., Cugliari, G., Berndt, S. I., Bezieau, S., et al. (2020). DNA repair and cancer in colon and rectum: Novel players in genetic susceptibility. *Int. J. Cancer* 146, 363–372. doi: 10.1002/ijc.32516
- Reis, H., Padden, J., Ahrens, M., Pütter, C., Bertram, S., Pott, L. L., et al. (2015). Differential proteomic and tissue expression analyses identify valuable diagnostic biomarkers of hepatocellular differentiation and hepatoid adenocarcinomas. *Pathology* 47, 543–550. doi: 10.1097/PAT.0000000000000298
- 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:e47. doi: 10.1093/nar/gkv007
- Roessler, S., Jia, H. L., Budhu, A., Forgues, M., Ye, Q. H., Lee, J. S., et al. (2010). A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* 70, 10202–10212. doi: 10.1158/0008-5472.CAN-10-2607
- Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173, 321–337.e10. doi: 10.1016/j.cell.2018.03.035
- Sarathi, A., and Palaniappan, A. (2019). Novel significant stage-specific differentially expressed genes in hepatocellular carcinoma. *BMC Cancer* 19:663. doi: 10.1186/s12885-019-5838-3
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D362. doi: 10.1093/nar/gkw937
- Tang, K., Yu, Y., Zhu, L., Xu, P., Chen, J., Ma, J., et al. (2019). Hypoxia-reprogrammed tricarboxylic acid cycle promotes the growth of human breast tumorigenic cells. *Oncogene* 38, 6970–6984. doi: 10.1038/s41388-019-0932-1
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3
- Veglia, F., Tyurin, V. A., Blasi, M., De Leo, A., Kossenkov, A. V., Donthireddy, L., et al. (2019). Fatty acid transport protein 2 reprograms neutrophils in cancer. *Nature* 569, 73–78. doi: 10.1038/s41586-019-1118-2
- Wang, F., Li, L., Piontek, K., Sakaguchi, M., and Selaru, F. M. (2018). Exosome miR-335 as a novel therapeutic strategy in hepatocellular carcinoma. *Hepatology* 67, 940–954. doi: 10.1002/hep.29586
- Wang, F., Wu, H., Zhang, S., Lu, J., Lu, Y., Zhan, P., et al. (2019). LAPTM4B facilitates tumor growth and induces autophagy in hepatocellular carcinoma. *Cancer Manag. Res.* 11, 2485–2497. doi: 10.2147/CMAR.S201092
- Wang, L. L., Jin, X. H., Cai, M. Y., Li, H. G., Chen, J. W., Wang, F. W., et al. (2018). AGLB2 promotes cancer cell growth through IRGM-regulated autophagy and enhanced Aurora A activity in hepatocellular carcinoma. *Cancer Lett.* 414, 71–80. doi: 10.1016/j.canlet.2017.11.003
- Wang, Y. B., Zhou, B. X., Ling, Y. B., Xiong, Z. Y., Li, R. X., Zhong, Y. S., et al. (2019). Decreased expression of ApoF associates with poor prognosis in human hepatocellular carcinoma. *Gastroenterol. Rep.* 7, 354–360. doi: 10.1093/gastro/goz011
- Wang, Y. H., Cheng, T. Y., Chen, T. Y., Chang, K. M., Chuang, V. P., and Kao, K. J. (2014). Plasmalemmal Vesicle Associated Protein (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma. *BMC Cancer* 14:815. doi: 10.1186/1471-2407-14-815
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220. doi: 10.1093/nar/gkq537
- Wei, L., Lian, B., Zhang, Y., Li, W., Gu, J., He, X., et al. (2014). Application of microRNA and mRNA expression profiling on prognostic biomarker discovery for hepatocellular carcinoma. *BMC Genomics* 15(Suppl. 1):S13. doi: 10.1186/1471-2164-15-S1-S13
- Wong, N., Yeo, W., Wong, W. L., Wong, N. L., Chan, K. Y., Mo, F. K., et al. (2009). TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance. *Int. J. Cancer* 124, 644–652. doi: 10.1002/ijc.23968
- Wu, M., Liu, Z., Li, X., Zhang, A., Lin, D., and Li, N. (2019). Analysis of potential key genes in very early hepatocellular carcinoma. *World J. Surg. Oncol.* 17:77. doi: 10.1186/s12957-019-1616-6

- Xie, X. W., Wang, X. Y., Liao, W. J., Fei, R., Cong, X., Chen, Q., et al. (2018). Effect of upregulated DNA replication and sister chromatid cohesion 1 expression on proliferation and prognosis in hepatocellular carcinoma. *Chin. Med. J.* 131, 2827–2835. doi: 10.4103/0366-6999.246076
- Xu, X., Zhou, Y., Miao, R., Chen, W., Qu, K., Pang, Q., et al. (2016). Transcriptional modules related to hepatocellular carcinoma survival: coexpression network analysis. *Front. Med.* 10:183–190. doi: 10.1007/s11684-016-0440-4
- Yan, T., Lu, L., Xie, C., Chen, J., Peng, X., Zhu, L., et al. (2015). Severely Impaired and dysregulated cytochrome P450 expression and activities in hepatocellular carcinoma: implications for personalized treatment in patients. *Mol. Cancer Ther.* 14, 2874–2886. doi: 10.1158/1535-7163.MCT-15-0274
- Yang, J., Xie, Q., Zhou, H., Chang, L., Wei, W., Wang, Y., et al. (2018). Proteomic analysis and NIR-II imaging of MCM2 protein in hepatocellular carcinoma. *J. Proteome Res.* 17, 2428–2439. doi: 10.1021/acs.jproteome.8b0181
- Yin, L., Cai, Z., Zhu, B., and Xu, C. (2018). Identification of Key Pathways and Genes in the Dynamic Progression of HCC Based on WGCNA. *Genes* 9:92. doi: 10.3390/genes9020092
- Zhang, T. T., Zhao, X. Q., Liu, Z., Mao, Z. Y., and Bai, L. (2016). Factors affecting the recurrence and survival of hepatocellular carcinoma after hepatectomy: a retrospective study of 601 Chinese patients. *Clin. Transl. Oncol.* 18, 831–840. doi: 10.1007/s12094-015-1446-0
- Zhang, Y., Zhao, H., Xu, W., Jiang, D., Huang, L., and Li, L. (2019). High expression of PQBP1 and low expression of PCK2 are associated with metastasis and recurrence of osteosarcoma and unfavorable survival outcomes of the patients. *J. Cancer* 10, 2091–2101. doi: 10.7150/jca.28480

**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.

Copyright © 2020 Song, Ding, Li, Liao, Xie, Yu, Zhang and Ni. 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.