



Transcriptome Analysis Revealed a Highly Connected Gene Module Associated With Cirrhosis to Hepatocellular Carcinoma Development

Shan Shan^{1,2,3†}, Wei Chen^{4,5†} and Ji-dong Jia^{1,2,3*}

¹ Liver Research Center, Beijing Friendship Hospital, Capital Medical University, Beijing, China, ² Beijing Key Laboratory of Translational Medicine on Liver Cirrhosis, Beijing Friendship Hospital, Capital Medical University, Beijing, China, ³ National Clinical Research Center for Digestive Disease, Beijing Friendship Hospital, Capital Medical University, Beijing, China, ⁴ Experimental and Translational Research Center, Beijing Friendship Hospital, Capital Medical University, Beijing, China, ⁵ Beijing Key Laboratory of Tolerance Induction and Organ Protection in Transplantation, Beijing Friendship Hospital, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:

Alfredo Pulvirenti,
Università degli Studi di Catania, Italy

Reviewed by:

Stefano Forte,
Mediterranean Institute of Oncology
(IOM), Italy
Luciano Cascione,
Istituto Oncologico della Svizzera
Italiana, Switzerland

*Correspondence:

Ji-dong Jia
jia_jd@ccmu.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 03 December 2018

Accepted: 19 March 2019

Published: 02 April 2019

Citation:

Shan S, Chen W and Jia J-d
(2019) Transcriptome Analysis
Revealed a Highly Connected Gene
Module Associated With Cirrhosis
to Hepatocellular Carcinoma
Development. *Front. Genet.* 10:305.
doi: 10.3389/fgene.2019.00305

Introduction: Cirrhosis is one of the most important risk factors for development of hepatocellular carcinoma (HCC). Recent studies have shown that removal or well control of the underlying cause could reduce but not eliminate the risk of HCC. Therefore, it is important to elucidate the molecular mechanisms that drive the progression of cirrhosis to HCC.

Materials and Methods: Microarray datasets incorporating cirrhosis and HCC subjects were identified from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were determined by GEO2R software. Functional enrichment analysis was performed by the clusterProfiler package in R. Liver carcinogenesis-related networks and modules were established using STRING database and MCODE plug-in, respectively, which were visualized with Cytoscape software. The ability of modular gene signatures to discriminate cirrhosis from HCC was assessed by hierarchical clustering, principal component analysis (PCA), and receiver operating characteristic (ROC) curve. Association of top modular genes and HCC grades or prognosis was analyzed with the UALCAN web-tool. Protein expression and distribution of top modular genes were analyzed using the Human Protein Atlas database.

Results: Four microarray datasets were retrieved from GEO database. Compared with cirrhotic livers, 125 upregulated and 252 downregulated genes in HCC tissues were found. These DEGs constituted a liver carcinogenesis-related network with 272 nodes and 2954 edges, with 65 nodes being highly connected and formed a liver carcinogenesis-related module. The modular genes were significantly involved in several KEGG pathways, such as “cell cycle,” “DNA replication,” “p53 signaling pathway,” “mismatch repair,” “base excision repair,” etc. These identified modular gene signatures could robustly discriminate cirrhosis from HCC in the validation dataset. In contrast,

the expression pattern of the modular genes was consistent between cirrhotic and normal livers. The top modular genes TOP2A, CDC20, PRC1, CCNB2, and NUSAP1 were associated with HCC onset, progression, and prognosis, and exhibited higher expression in HCC compared with normal livers in the HPA database.

Conclusion: Our study revealed a highly connected module associated with liver carcinogenesis on a cirrhotic background, which may provide deeper understanding of the genetic alterations involved in the transition from cirrhosis to HCC, and offer valuable variables for screening and surveillance of HCC in high-risk patients with cirrhosis.

Keywords: cirrhosis, hepatocellular carcinoma, transcriptome, module, prognosis

INTRODUCTION

Hepatocellular carcinoma accounts for 90% of all primary liver malignancies (Mittal and El-Serag, 2013; Galle et al., 2018), posing a serious threat to human health and quality of life. Worldwide, most patients with HCC have underlying cirrhosis of various etiologies (Fattovich et al., 2004; Beste et al., 2015; Walker et al., 2016). Growing clinical evidence shows that removal or control of the injurious factors, such as hepatitis B or C virus, can reduce but not eliminate the risk of HCC (Casado et al., 2013; Marcellin et al., 2013; Xu et al., 2015; Sun et al., 2017). Therefore, it is important to understand the molecular mechanisms that drive the progression of cirrhosis to HCC.

Hepatocellular carcinoma occurs as a consequence of the complex interplay between multiple genetic determinants (Sanyal et al., 2010). Previous studies have found that aberrations in genetic molecules pertaining to oxidative stress, EMT, inflammatory response, cellular senescence, or telomere dysfunction may contribute to the progression of cirrhosis to HCC (Ramakrishna et al., 2013). In addition, the Wnt/ β -catenin, p53, pRb, MAPK, RAS, and JAK/STAT pathways are also reported to be canonical molecular pathways in HCC development (Aravalli et al., 2008). However, different studies often yield diverse results and the global view on the landscape of genomic changes is still not very clear.

With the aid of high-throughput detection techniques, all expressed genetic molecules in a given liver tissue sample can be simultaneously detected over a wide quantitative range (Mas et al., 2009; Villanueva et al., 2011; Yildiz et al., 2013; Wang et al., 2014; Lee, 2015; Schulze et al., 2015; Villanueva et al., 2015; Diaz et al., 2018; Shen et al., 2018). High-throughput sequencing and microarray technologies allow investigators to simultaneously measure the changes of genome-wide genes under certain biological conditions. These approaches usually generate large “interesting” gene lists. By using biological knowledge accumulated in public databases (e.g., KEGG¹), it is

Abbreviations: DEGs, differentially expressed genes; EMT, epithelial-to-mesenchymal transition; FC, fold-change; FDR, false discovery rate; HPA, Human Protein Atlas; GEO, Gene Expression Omnibus; HCC, hepatocellular carcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; MCODE, molecular complex detection; PCA, principal component analysis; RMA, robust multichip average; STRING, Search Tool for the Retrieval of Interacting Genes; TCGA, the Cancer Genome Atlas; TPM, transcripts per million.

¹<https://www.kegg.jp/>

possible to systematically dissect large gene lists in an attempt to assemble a summary of the most enriched and pertinent biology. Therefore, integrated analyses of multiple datasets generated from different studies may help us to identify reliable and reproducible genetic alterations involved in the development of HCC on a cirrhotic background.

Therefore, our present study used multiple bioinformatics tools to systematically integrate publicly available transcriptomic datasets and performed high-throughput gene expression comparisons between HCC and benign cirrhotic tissues.

MATERIALS AND METHODS

Retrieval of Microarray Datasets on Cirrhosis and HCC From Public Database

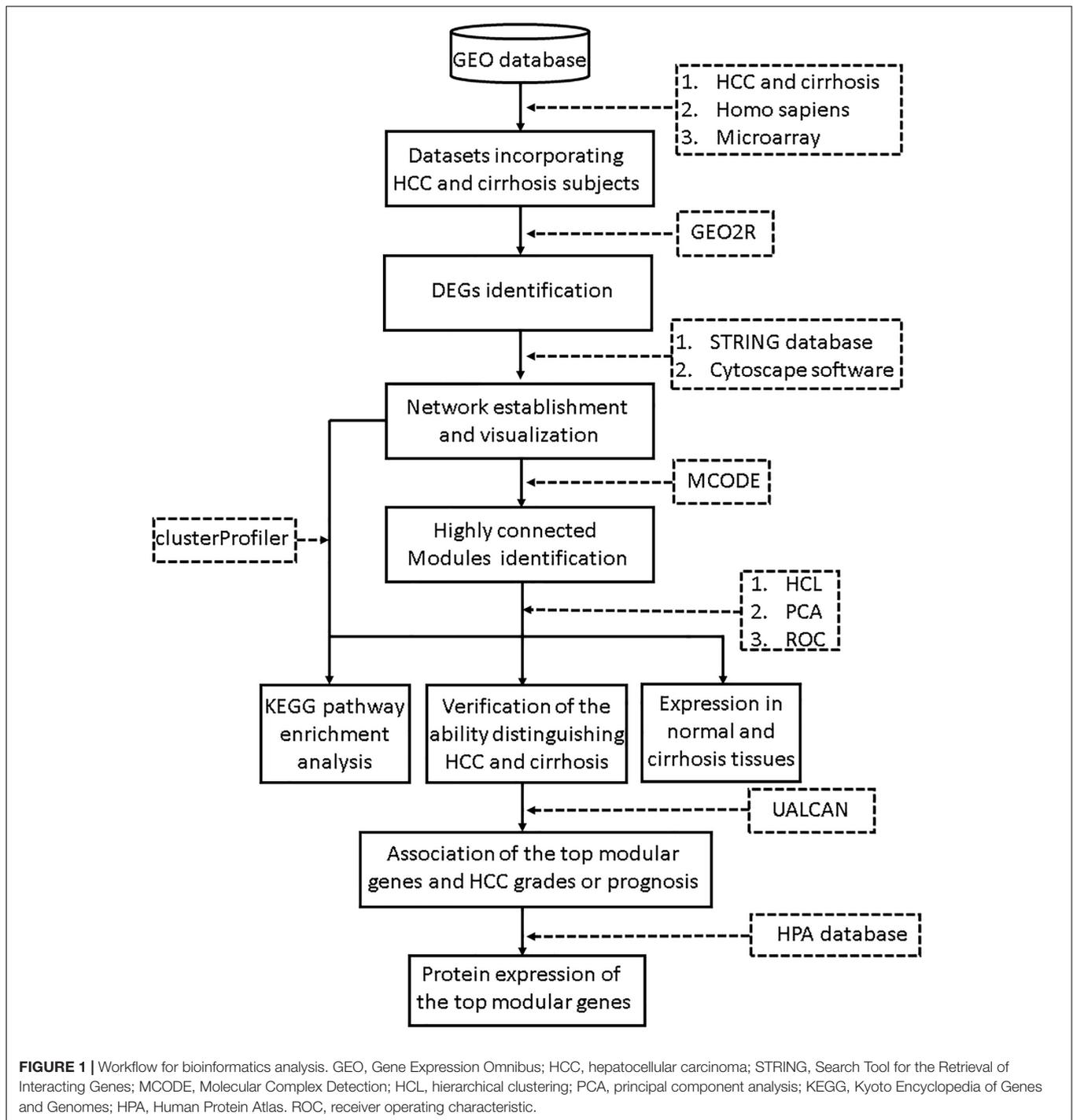
First, we searched and retrieved transcriptome profiles of cirrhotic and HCC tissues from GEO which is a public functional genomics data repository, allowing users to query, locate, review, and download studies and gene expression profiles of interest (Barrett et al., 2013).

The search terms we used included “cirrhosis” and “HCC.” Studies were considered eligible for analysis if: (1) studies contained both cirrhosis and HCC tissues; (2) species was limited to *Homo sapiens*; and (3) platform was limited to microarray. Then the retrieved datasets were further screened by manual retrieval. Our workflow for bioinformatics analysis of publicly available datasets is illustrated in **Figure 1**.

Identification of DEGs Related to Liver Carcinogenesis From the Retrieved Microarray Datasets

Differentially expressed genes between cirrhosis and HCC tissues were defined as liver carcinogenesis-related genes that may have important implications in driving cirrhosis to HCC.

Gene expression in all the datasets was normalized by the antilog-transformed RMA algorithm. GEO query and limma R packages contained in GEO2R, which allows gene expression analysis of published microarray datasets, was used to determine the DEGs between cirrhosis and HCC tissues (Davis and Meltzer, 2007). FDR < 0.05 and FC > 1.5 were considered as the cutoff values for DEG screening. The overlapping DEGs in datasets were retained for further analyses.



Functional Specification of the Identified DEGs Related to Liver Carcinogenesis

To identify and visualize enriched KEGG pathways for the candidate gene sets, clusterProfiler, which is an R package for comparing biological themes among gene clusters, was employed (Yu et al., 2012). Fisher's exact test followed by the Benjamini correction was performed and an adjusted *P*-value of <0.05 was set as the cutoff criterion.

Establishment of the Liver Carcinogenesis-Related Network and Its Modules

The internal regulatory relationships between the identified liver carcinogenesis-related genes were predicted by the STRING database (confidence score > 0.4) (Szklarczyk et al., 2017). Liver carcinogenesis-related network was established and visualized with Cytoscape software (Shannon et al., 2003).

We used the MCODE plug-in in the Cytoscape software (Bader and Hogue, 2003) to screen the modules concealed in the liver carcinogenesis-related network with the following criteria: Max. depth = 100, K-Core = 2, mode score cutoff = 0.2, and degree cutoff = 2. Likewise, the functional specification of the identified module was determined with the clusterProfiler package as mentioned above. An adjusted *P*-value of <0.05 was considered statistically significant.

Verification of the Identified Modules for Discriminating Cirrhosis From HCC

We used three of datasets (GSE89377, GSE17548, and GSE98383) to mine modules from the liver carcinogenesis-related network; and used the remaining dataset (GSE56140) to validate the findings. To verify the ability of the identified modules to discriminate cirrhosis from HCC subjects, we performed hierarchical clustering analysis by using R with the complete linkage method and the expression of modular genes as a distance metric. To verify the results of hierarchical clustering, we applied the identified modular genes that were considered as observed variables to PCA plots. PCA was conducted with the ggbiplot package in R. The first two principal components (PCs) were then subjected to binary logistic regression analysis to calculate the predicted probability which was applied to the receiver operating characteristic (ROC) curve analysis implemented by SPSS 20 (IBM, United States). Area under curve (AUC) was calculated to determine the predictive performance of the identified gene module. In order to reduce sampling bias, the modules were screened from any three out of the four datasets and repeated evaluations of their discriminant ability were performed using the remaining dataset.

Comparison of the Identified Modular Genes in Normal and Cirrhotic Samples

We used GEO2R software to determine the expression differences between any two groups in dataset. An FDR of <0.05 and an FC of > 1.5 were considered as the cutoff values for DEG screening. Modular gene expression in normal, cirrhotic, and HCC samples were visualized by using a heatmap drawn with MeV software².

Analyses of the Association Between the Top Modular Genes and HCC Histological Grade or Clinical Outcome

Modular genes with FC > 3 between cirrhosis and HCC tissues in all GEO datasets were considered as the top modular genes. Association of the top modular genes and HCC grades or prognosis was analyzed by using UALCAN (Chandrashekar et al., 2017), which is an interactive web-portal for exploring the association between tumor subgroup gene expression and survival in TCGA³. Expression differences of top modular genes between normal and different tumor grades were analyzed using the statistical method built-in the UALCAN web-software; a *P*-value of <0.05 was considered significant.

²<https://sourceforge.net/projects/mev-tm4/>

³<https://cancergenome.nih.gov/>

According to the TPM expression values, the top modular genes were divided into a high expression group (with TPM values above upper quartile) and a low expression group (with TPM values below lower quartile). With information on the association between the gene expression and survival profiles documented in TCGA, Kaplan–Meier survival analyses were performed and overall survival plots were generated. The difference between high gene expression and low gene expression was compared by log-rank test; a *P*-value of < 0.05 was considered significant.

In silico Analysis of the Top Modular Members in Normal and HCC Specimens

Protein expression and distribution of the top modular genes in human liver tissue were searched in the HPA⁴ database (Uhlen et al., 2015).

RESULTS

Retrieved Microarray Datasets Pertaining to Cirrhosis and HCC

According to the retrieval criteria, four microarray datasets (Table 1) containing a total of 95 benign (cirrhosis) and 98 malignant (HCC) subjects were found from the GEO database. GSE98983 dataset was produced by Affymetrix Human Genome U133 Plus 2.0 Array (GPL570); GSE89377 by Illumina HumanHT-12 V3.0 expression beadchip (GPL6947); GSE17548 by Affymetrix Human Genome U133 Plus 2.0 Array (GPL570); and GSE56140 by Illumina HumanHT-12 V3.0 expression beadchip.

DEGs Related to Liver Carcinogenesis

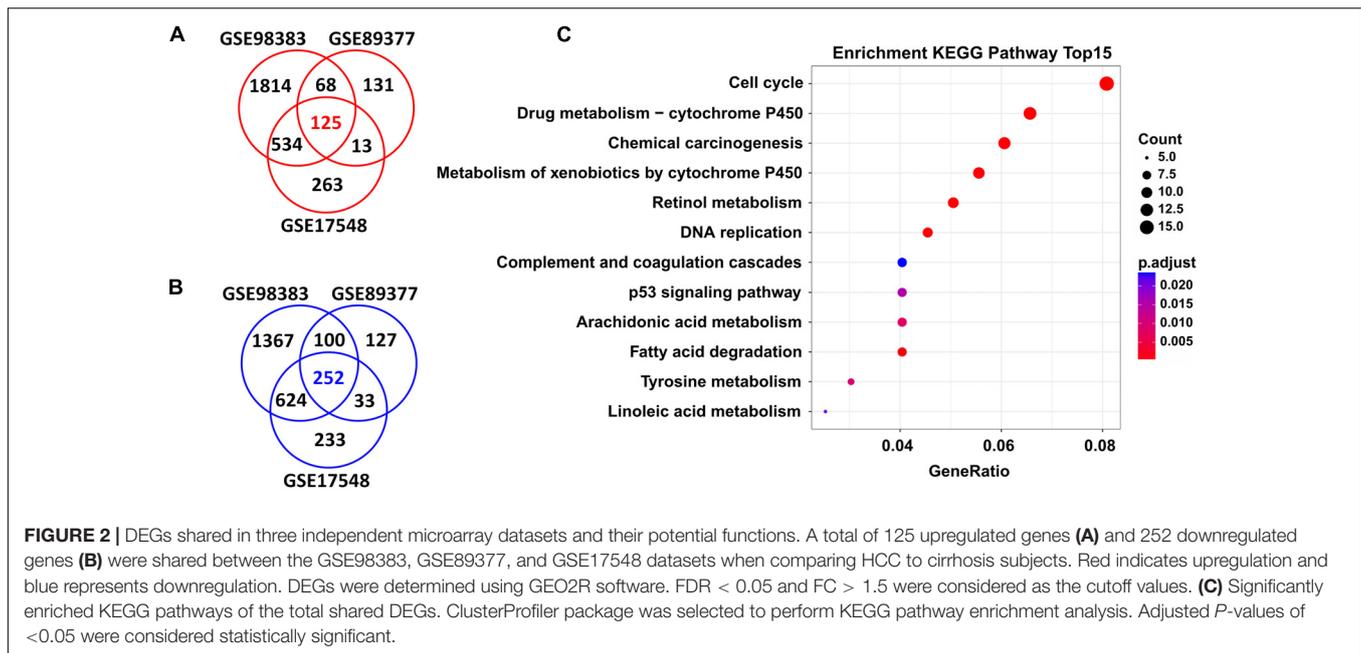
In total, we found 125 upregulated and 252 downregulated genes (adjusted *P* < 0.05 and FC > 1.5) in HCC tissues when compared with cirrhosis tissues. These DEGs were shared by the three independent datasets (GSE98383, GSE89377, and GSE17548) (Figures 2A,B and Supplementary Table S1).

⁴www.proteinatlas.org

TABLE 1 | Detailed information of the four microarray datasets used in the present study.

ID	Platform	Sample	Etiology	References
GSE98383	GPL570	LC = 29; HCC = 11	HDV	Diaz et al., 2018
GSE89377	GPL6947	LC = 12; HCC = 35	HBV, HCV, alcohol, others	Shen et al., 2018
GSE17548	GPL570	LC = 20; HCC = 17	HBV, HCV, HBV+HDV, others	Yildiz et al., 2013
GSE56140	GPL18461	LC = 34; HCC = 35	HBV, HCV	Villanueva et al., 2011

LC, liver cirrhosis; HCC, hepatocellular carcinoma; HBV, hepatic B virus; HCV, hepatic C virus; HDV, hepatic D virus.



Functional Specification of DEGs Related to Liver Carcinogenesis

Functional enrichment analysis revealed that these DEGs were significantly enriched in several KEGG pathways; as shown in **Figure 2C**, the top ones were “cell cycle,” “drug metabolism-cytochrome P450,” “chemical carcinogenesis,” “metabolism of xenobiotics by cytochrome P450,” “retinoid metabolism,” “DNA replication,” “complement and coagulation cascades,” “p53 signaling pathway,” etc. Detailed information of these pathways is listed in **Supplementary Table S2**.

Network-Based Modules Involved in Liver Carcinogenesis

By screen against the STRING database, totally 2954 interactions were found among 272 DEGs, which was visualized by using Cytoscape software. The network layout was arranged with the Allegro Spring-Electric plug-in in Cytoscape software. As shown in **Figure 3A**, the upregulated genes in the liver carcinogenesis-related network were highly connected, suggesting a core role in the whole network.

Then we used the MCODE plug-in to mine the highly clustered modules from this network. As expected, a module holding a higher connectivity (cluster score = 64.1) was identified, with 65 nodes and 1955 edges. Interestingly, all the modular genes were notably upregulated in HCC tissues compared with cirrhotic tissues, signifying their roles in liver carcinogenesis (**Figure 3B** and **Supplementary Table S3**). These modular genes were involved in several KEGG pathways including “cell cycle,” “DNA replication,” “oocyte meiosis,” “human T-cell leukemia virus 1 infection,” “progesterone-mediated oocyte maturation,” “cellular senescence,” “p53 signaling pathway,” “base excision repair,” and “mismatch repair” (**Figure 3C** and **Supplementary Table S4**).

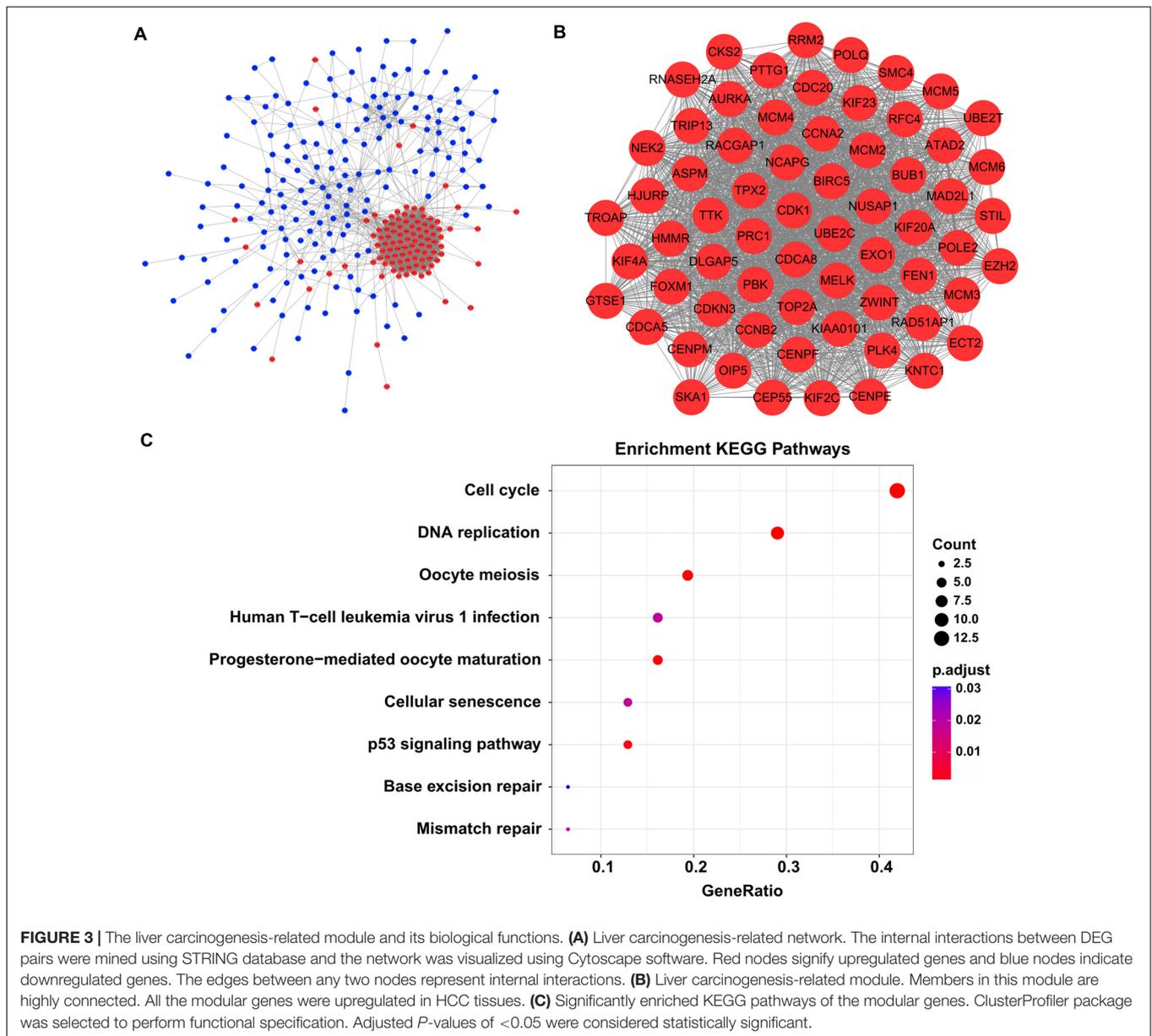
Modular Gene Signatures Discriminating Cirrhosis From HCC

Hierarchical clustering analysis of the validation dataset (GSE56140, 34 cirrhosis subjects and 35 HCC subjects) showed that subjects with the same diagnosis were inclined to be evidently more clustered (**Figure 4A**). In agreement with the result of hierarchical clustering, as shown in **Figure 4B**, the PCA plot clearly distinguished cirrhosis from HCC with a small overlap. The first two PCs that were the most informative, accounting for approximately 86.6 and 3.5% of the total observed variances, respectively. ROC analysis revealed an AUC of 0.919, indicating the identified modular genes, to some extent, could be a combined predictor of cirrhosis to HCC development (**Figure 4C**).

We next used any three out of the four datasets as training datasets to mine liver carcinogenesis-related modules. As shown in **Supplementary Figures S1–S3**, all the identified modules, as the module screened from GSE98383, GSE89377, and GSE17548, were characterized by upregulated genes and high connectivity; members in these modules identified from different training datasets were largely overlapped. In addition, hierarchical clustering, PCA and ROC analyses, to a large extent, could distinguish cirrhotic subjects from HCC subjects (**Supplementary Figures S4–S6**). These results indicated that the identified liver carcinogenesis-related module was not by chance and the ability of discriminating cirrhosis from HCC was relatively robust.

Characterization of the Modular Gene Expression Patterns in Normal, Cirrhotic, and HCC Livers

The GSE89377 dataset contained not only cirrhosis ($n = 12$) and HCC ($n = 35$) subjects but also normal ($n = 13$) subjects.



Therefore, we used this dataset to analyze the modular gene expression pattern in normal, cirrhotic, and HCC livers. As shown in **Figure 5**, while the modular genes being differentially expressed between HCC and cirrhosis/normal livers (adjusted $P < 0.05$ and $FC > 1.5$), there were no significant differences in expression between cirrhosis and normal livers (adjusted $P > 0.05$ or $FC < 1.5$).

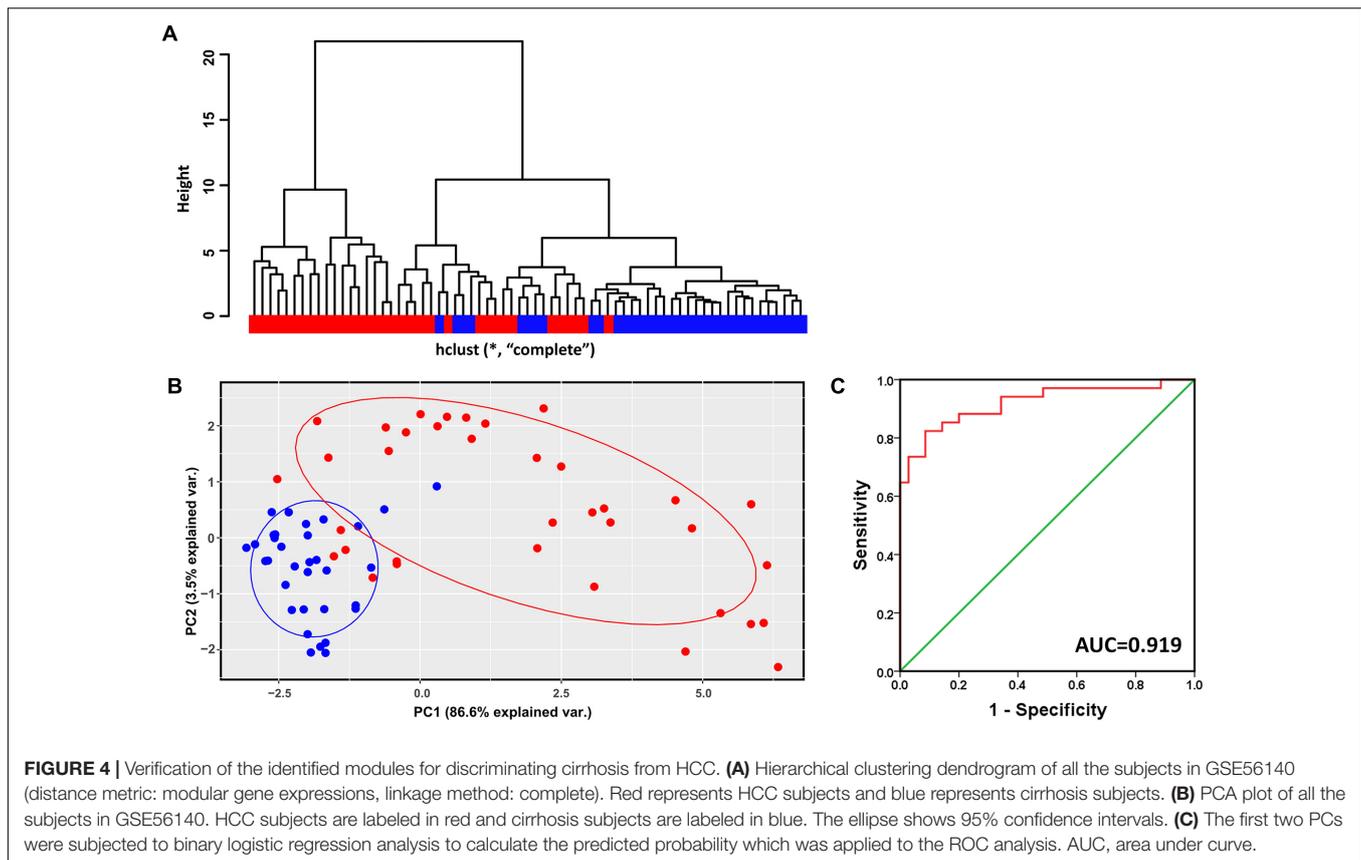
Association of Top Modular Gene Signatures With HCC Onset, Progression, and Prognosis

We next focused on the top modular gene signatures because their expression was highly dysregulated in the HCC tissues of all four datasets considered in this study. In total, five

modular genes including TOP2A, CDC20, PRC1, CCNB2, and NUSAP1 satisfied the criterion stated in Section “Materials and Methods,” therefore they were considered as the top modular gene signatures.

The liver cancer datasets in TCGA database contained 50 normal subjects, 54 HCC subjects with grade I, 173 HCC subjects with grade II, 118 HCC subjects with grade III, and 12 HCC subjects with grade IV. As shown in **Figure 6A**, all the top modular gene signatures were significantly upregulated in each HCC grade group, compared with the normal group ($P < 0.05$), and the expression of all genes increased in a stepwise manner with the HCC progress, suggesting a close relationship between the five genes and HCC onset and progression.

To investigate if the top modular gene signatures affected overall survival in patients with HCC, we performed a



survival prediction by Kaplan–Meier curve analysis embedded in UALCAN web-tool. As shown in **Figure 6B**, high expression of TOP2A, CDC20, and CCNB2 protein was significantly associated with poor survival time in HCC patients ($P < 0.01$). Although lower expression of PRC1 and NUSAP1 tended to be associated with better outcome in HCC patients, significant differences were not achieved ($P > 0.05$).

Protein Expression and Distribution of TOP2A, CDC20, PRC1, CCNB2, and NUSAP1 in HCC Livers

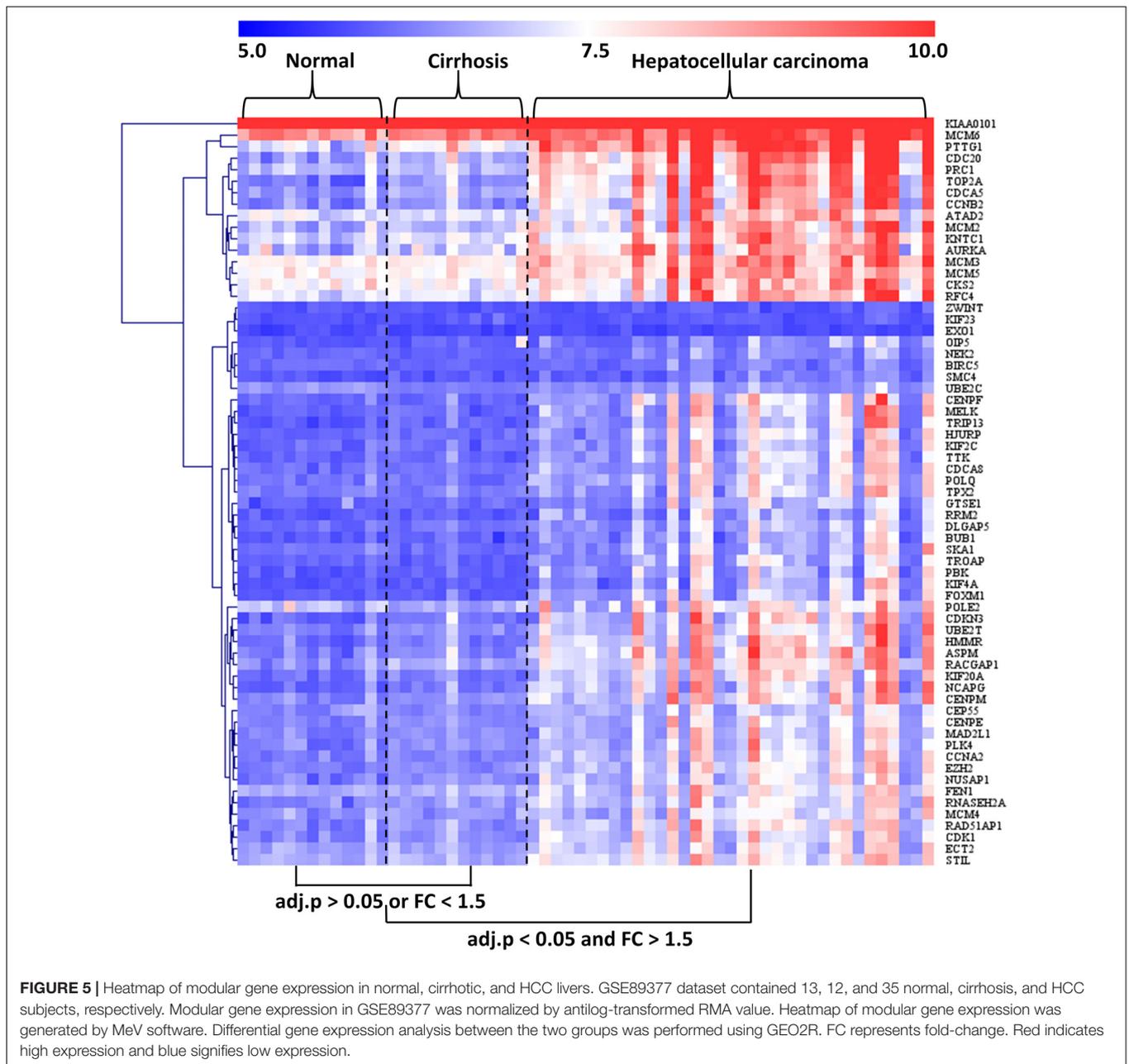
In the HPA database, we were able to find normal and HCC sections from several patients with staining for the top modular proteins. Antibodies used in the HPA database were: TOP2A (HPA006458), CDC20 (CAB004525), PRC1 (HPA034521), CCNB2 (CAB009575), and NUSAP1 (HPA043904).

Immunohistochemistry for the five members in the HPA database showed that TOP2A and NUSAP1 highly expressed in HCC cell nuclei but almost undetectable in normal tissue, whereas PRC1 highly expressed in HCC cytoplasm and plasma membrane but undetectable in normal liver tissue. Although CDC20 and CCNB2 exhibited higher rate of expression in HCC cytoplasm and plasma membrane, their abundance was very low (**Figure 7**). Hence, TOP2A, PRC1, and NUSAP1 have the potential to be liver biopsy-based markers for screening HCC high-risk patients with cirrhosis.

DISCUSSION

Our current study systematically integrated four independent microarray datasets that contains cirrhosis and HCC tissues. By performing a series of bioinformatics analyses, we found a highly connected module covering 65 HCC risky genes, which could robustly distinguish cirrhosis from HCC; the top modular genes were highly associated with HCC onset and development and prognosis.

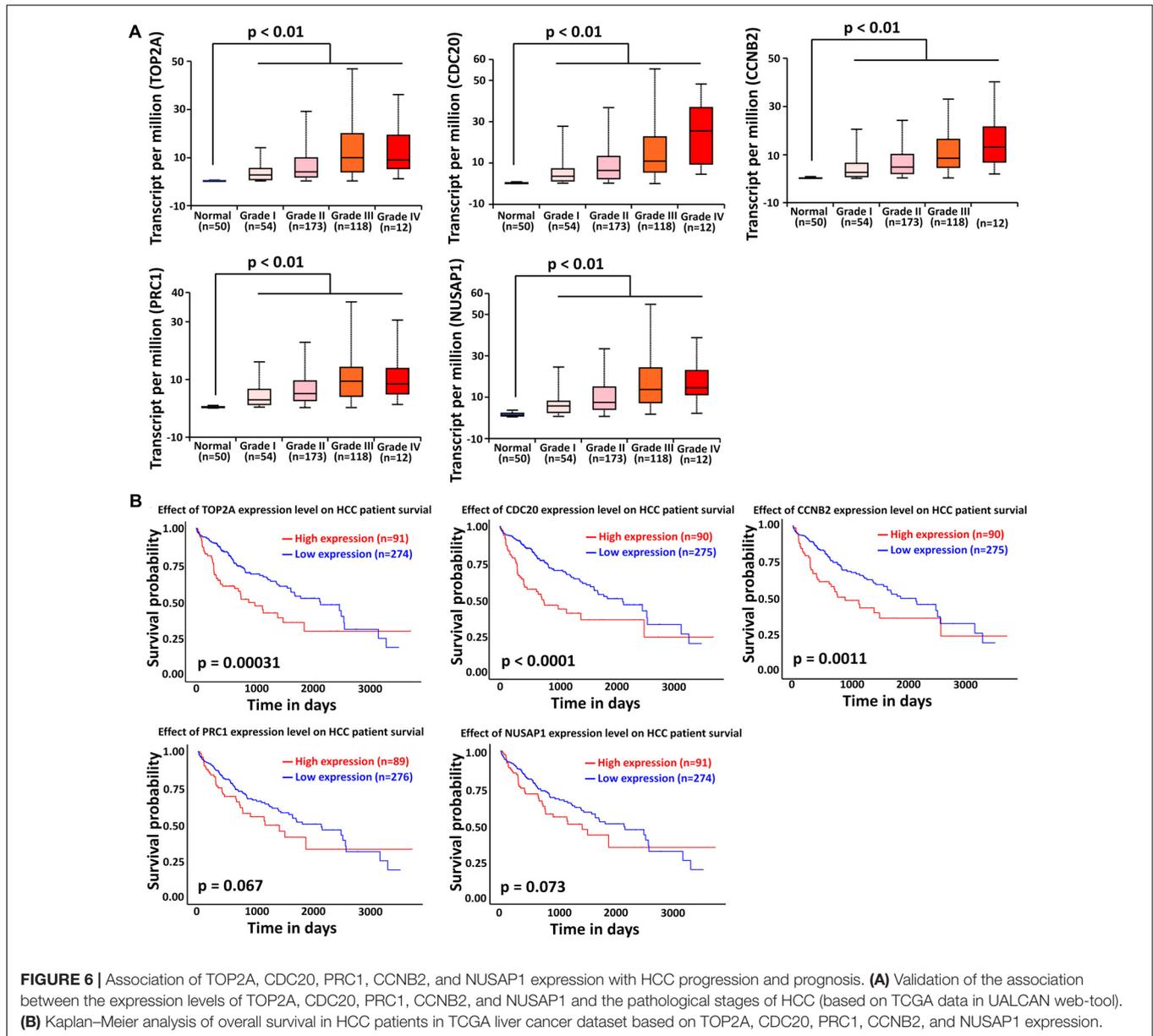
The module identified in the present study was highly connected; by functional enrichment analysis, the modular genes were found to be involved in several KEGG pathways. Mismatch repair pathway usually corrects insertion/deletion loops and base–base mismatches generated during DNA replication and recombination; base excision repair pathway is the main repair mechanism for DNA damage; the two pathways together with cell cycle and DNA replication pathways are the foundational mechanisms determining cell fate associated with carcinogenesis (Macdonald et al., 1998; Bisteau et al., 2014; Mjelle et al., 2015; Liu Z. et al., 2018). P53 signaling pathway inhibition has been widely reported to be required for liver cancer initiation (Cao et al., 2018; Dhar et al., 2018). Cellular senescence, a process of cell proliferation arrest in response to multifarious stimuli, which can modify the microenvironment of tissues, has been reported to be closely associated with cancer onset of multiple tissues including liver (Kim and Park, 2019). Since the increase of senescent cells, it is likely that the



preneoplastic setting of the cirrhotic background may provide a conducive environment for cellular transformation, which should be further investigated. Although human T-cell leukemia virus 1 infection pathway has not been reported to be linked to HCC, but the upregulation of modular genes involved in this pathway such as CDC20, MAD2L1, and PTTG1 have been confirmed to promote HCC development and progression (Cho-Rok et al., 2006; Li et al., 2014; Li Y. et al., 2017). Except the pathways mentioned above, “oocyte meiosis” and “progesterone-mediated oocyte maturation” are two main KEGG pathways which the identified modular gene were enriched in. Oocyte meiosis and progesterone-mediated oocyte maturation pathways have also been found to be associated with HCC

by integrated analysis of microarray studies (Li L. et al., 2017; Wang F. et al., 2017). However, the causal association between oocyte meiosis and HCC onset should be investigated in further study.

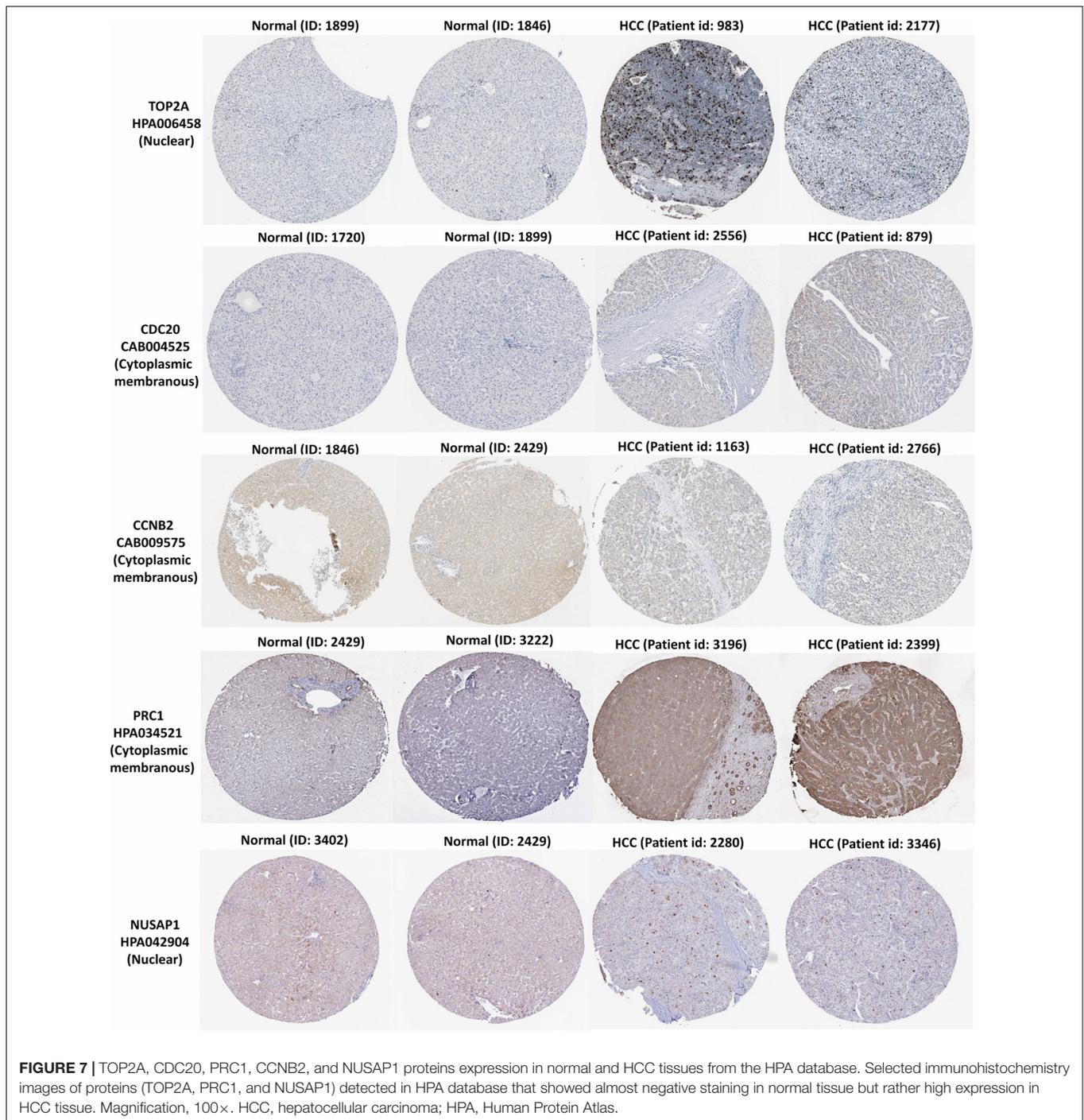
The top modular genes, TOP2A, CDC20, PRC1, CCNB2, and NUSAP1, were highly associated with HCC onset and development; high expression of TOP2A, CDC20, or CCNB2 was correlated with poor survival time in TCGA liver cancer patients, implying their potential as biopsy-based prognostic markers. Consistent with our findings in the HPA database, previous studies have also found that TOP2A (Panvichian et al., 2015; Ang et al., 2016; Xu et al., 2016; Li et al., 2018; Wang et al., 2018; Zhou et al., 2018), CCNB2 (Liu S. et al., 2018; Zhou et al., 2018),



CDC20 (Li et al., 2014; Jin et al., 2015; Li L. et al., 2017; Yan et al., 2017; Fan et al., 2018), PRC1 (Chen et al., 2016; Wang Y. et al., 2017; Liu X. et al., 2018), and NUSAP1 (Zhang et al., 2013; Roy et al., 2018; Zhou et al., 2018) are overexpressed in HCC but are almost undetectable in non-tumorous livers. TOP2A has been previously confirmed to correlate with advance histological grading, microvascular invasion, early age onset, shorter patient survival, and chemoresistance of HCC (Wong et al., 2009). High expression of CDC20 in HCC patients is associated with shorter overall survival (Fan et al., 2018); silencing of CDC20 expression significantly inhibits HCC cell proliferation and tumor growth (Li et al., 2014, Liu et al., 2015). PRC1 in HCC tissue is regulated by Wnt3a signaling, exerting an oncogenic effect by promoting cancer proliferation, stemness, metastasis, and tumorigenesis; high expression of

PRC1 was associated with early HCC recurrence and poor patient outcome (Chen et al., 2016; Wang Y. et al., 2017). Finally, NUSAP1 expression in the surgical margins of HCC is closely correlated to early postoperative recurrence and could serve as an indicator for predicting early recurrence of HCC (Zhang et al., 2013).

Despite studies devoted to decoding the process of cirrhosis development to HCC have been extensively reported, integrated studies based on multiple datasets are rare. Prior to this current work, only one study had been reported by Jiang et al. (2017) who performed a weighted gene co-expression network analysis based on five independent gene expression profiles and identified six modules contributing to HCC progression. They found hub genes in the identified modules were mainly cytokines, such as chemokine (C-C motif) ligand 22 and interleukin-19



(Jiang et al., 2017). However, our study found only one highly connected module that closely involved in the canonical carcinogenesis-associated pathways. The following reasons may at least partly explain such difference. First, cirrhotic tissues in each microarray dataset included in our study were benign tissues only and not mixture of benign and para-carcinoma tissues. Second, the microarray profiles in our study were generated by more than one platform and etiologies covered HBV, HCV, HDV, alcohol, and others. In order to obtain more reliable results,

gene expression datasets from different microarray platforms and etiologies were considered; the overlapped DEGs were retained for further analysis; datasets were separated into training and test datasets, and the ability of the identified functional module distinguishing cirrhosis from HCC was validated in the test dataset; resampling and repeated evaluations obtained robust results; moreover, the associations between the expression of the top modular genes and HCC progression and prognosis were determined in other liver cancer datasets.

CONCLUSION

Our present study systematically integrated multiple microarray gene expression profiles and found a module associated with liver carcinogenesis on a cirrhotic background that could robustly discriminate cirrhosis from HCC. The expression of top modular genes was closely associated with HCC onset, development, and prognosis. Our work may provide a deeper understanding of molecular mechanisms in HCC onset from cirrhosis and offer new insights for screening and surveillance of high-risk patients with cirrhosis during anti-viral therapy.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/>.

AUTHOR CONTRIBUTIONS

J-dJ conceived and designed the experiments. SS and WC performed data analysis and drafted the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by Key Project from Beijing Municipal Science and Technology Commission (No. D161100002716003),

REFERENCES

- Ang, C., Miura, J. T., Gamblin, T. C., He, R., Xiu, J., Millis, S. Z., et al. (2016). Comprehensive multiplatform biomarker analysis of 350 hepatocellular carcinomas identifies potential novel therapeutic options. *J. Surg. Oncol.* 113, 55–61. doi: 10.1002/jso.24086
- Aravalli, R. N., Steer, C. J., and Cressman, E. N. (2008). Molecular mechanisms of hepatocellular carcinoma. *Hepatology* 48, 2047–2063. doi: 10.1002/hep.22580
- Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2. doi: 10.1186/1471-2105-4-2
- 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, D991–D995. doi: 10.1093/nar/gks1193
- Beste, L. A., Leipertz, S. L., Green, P. K., Dominitz, J. A., Ross, D., and Ioannou, G. N. (2015). Trends in burden of cirrhosis and hepatocellular carcinoma by underlying liver disease in US veterans, 2001–2013. *Gastroenterology* 149, 1471–1482. doi: 10.1053/j.gastro.2015.07.056
- Bisteau, X., Caldez, M. J., and Kaldis, P. (2014). The complex relationship between liver cancer and the cell cycle: a story of multiple regulations. *Cancers* 6, 79–111. doi: 10.3390/cancers6010079
- Cao, P., Yang, A., Wang, R., Xia, X., Zhai, Y., Li, Y., et al. (2018). Germline duplication of SNORA18L5 increases risk for HBV-related hepatocellular carcinoma by altering localization of ribosomal proteins and decreasing levels of p53. *Gastroenterology* 155, 542–556. doi: 10.1053/j.gastro.2018.04.020
- Casado, J. L., Quereda, C., Moreno, A., Perez-Elias, M. J., Marti-Belda, P., and Moreno, S. (2013). Regression of liver fibrosis is progressive after sustained virological response to HCV therapy in patients with hepatitis C and HIV coinfection. *J. Viral. Hepat.* 20, 829–837. doi: 10.1111/jvh.12108

the National Natural Science Foundation of China (No. 81800534), and the Seed Project from Beijing Friendship Hospital (No. YYZZ2017A07).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00305/full#supplementary-material>

FIGURE S1 | Liver carcinogenesis-related module identified from GSE89377, GSE17548, and GSE56140.

FIGURE S2 | Liver carcinogenesis-related module identified from GSE98383, GSE17548, and GSE56140.

FIGURE S3 | Liver carcinogenesis-related module identified from GSE98383, GSE89377, and GSE56140.

FIGURE S4 | Verification of the identified modules for discriminating cirrhosis from HCC in GSE98383.

FIGURE S5 | Verification of the identified modules for discriminating cirrhosis from HCC in GSE89377.

FIGURE S6 | Verification of the identified modules for discriminating cirrhosis from HCC in GSE17548.

TABLE S1 | Fold changes of the differentially expressed genes shared in GSE98383, GSE89377, and GSE17548.

TABLE S2 | Significantly enriched KEGG pathways of the shared DEGs.

TABLE S3 | Significantly enriched KEGG pathways of the genes in the identified gene module.

TABLE S4 | Information of the genes in the identified module.

- Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19, 649–658. doi: 10.1016/j.neo.2017.05.002
- Chen, J., Rajasekaran, M., Xia, H., Zhang, X., Kong, S. N., Sekar, K., et al. (2016). The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/beta-catenin signalling pathway. *Gut* 65, 1522–1534. doi: 10.1136/gutjnl-2015-310625
- Cho-Rok, J., Yoo, J., Jang, Y. J., Kim, S., Chu, I. S., Yeom, Y. I., et al. (2006). Adenovirus-mediated transfer of siRNA against PTTG1 inhibits liver cancer cell growth in vitro and in vivo. *Hepatology* 43, 1042–1052. doi: 10.1002/hep.21137
- Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847. doi: 10.1093/bioinformatics/btm254
- Dhar, D., Antonucci, L., Nakagawa, H., Kim, J. Y., Glitzner, E., Caruso, S., et al. (2018). Liver cancer initiation requires p53 inhibition by CD44-enhanced growth factor signaling. *Cancer Cell* 33, 1061–1077. doi: 10.1016/j.ccell.2018.05.003
- Diaz, G., Engle, R. E., Tice, A., Melis, M., Montenegro, S., Rodriguez-Canales, J., et al. (2018). Molecular signature and mechanisms of hepatitis D virus-associated hepatocellular carcinoma. *Mol. Cancer Res.* 16, 1406–1419. doi: 10.1158/1541-7786.MCR-18-0012
- Fan, G., Tu, Y., Chen, C., Sun, H., Wan, C., and Cai, X. (2018). DNA methylation biomarkers for hepatocellular carcinoma. *Cancer Cell Int.* 18:140. doi: 10.1186/s12935-018-0629-5
- Fattovich, G., Stroffolini, T., Zagni, I., and Donato, F. (2004). Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology* 127(Suppl. 1), S35–S50. doi: 10.1053/j.gastro.2004.09.014

- Galle, P. R., Forner, A., Llovet, J. M., Mazzaferro, V., Piscaglia, F., Raoul, J. L., et al. (2018). EASL clinical practice guidelines: management of hepatocellular carcinoma. *J. Hepatol.* 69, 182–236. doi: 10.1016/j.jhep.2018.03.019
- Jiang, M., Zeng, Q., Dai, S., Liang, H., Dai, F., Xie, X., et al. (2017). Comparative analysis of hepatocellular carcinoma and cirrhosis gene expression profiles. *Mol. Med. Rep.* 15, 380–386. doi: 10.3892/mmr.2016.6021
- Jin, B., Wang, W., Du, G., Huang, G. Z., Han, L. T., Tang, Z. Y., et al. (2015). Identifying hub genes and dysregulated pathways in hepatocellular carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 19, 592–601.
- Kim, Y. H., and Park, T. J. (2019). Cellular senescence in cancer. *BMB Rep.* 52, 42–46. doi: 10.5483/BMBRep.2019.52.1.295
- Lee, J. S. (2015). The mutational landscape of hepatocellular carcinoma. *Clin. Mol. Hepatol.* 21, 220–229. doi: 10.3350/cmh.2015.21.3.220
- Li, J., Gao, J. Z., Du, J. L., Huang, Z. X., and Wei, L. X. (2014). Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. *Int. J. Oncol.* 45, 1547–1555. doi: 10.3892/ijo.2014.2559
- Li, L., Lei, Q., Zhang, S., Kong, L., and Qin, B. (2017). Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis. *Oncol. Rep.* 38, 2607–2618. doi: 10.3892/or.2017.5946
- Li, Y., Bai, W., and Zhang, J. (2017). MiR-200c-5p suppresses proliferation and metastasis of human hepatocellular carcinoma (HCC) via suppressing MAD2L1. *Biomed. Pharmacother.* 92, 1038–1044. doi: 10.1016/j.biopha.2017.05.092
- Li, N., Li, L., and Chen, Y. (2018). The identification of core gene expression signature in hepatocellular carcinoma. *Oxid. Med. Cell Longev.* 2018:3478305. doi: 10.1155/2018/3478305
- Liu, M., Zhang, Y., Liao, Y., Chen, Y., Pan, Y., Tian, H., et al. (2015). Evaluation of the antitumor efficacy of RNAi-mediated inhibition of CDC20 and heparanase in an orthotopic liver tumor model. *Cancer Biother. Radiopharm.* 30, 233–239. doi: 10.1089/cbr.2014.1799
- Liu, S., Yao, X., Zhang, D., Sheng, J., Wen, X., Wang, Q., et al. (2018). Analysis of transcription factor-related regulatory networks based on bioinformatics analysis and validation in hepatocellular carcinoma. *Biomed. Res. Int.* 2018:1431396. doi: 10.1155/2018/1431396
- Liu, X., Li, Y., Meng, L., Liu, X. Y., Peng, A., Chen, Y., et al. (2018). Reducing protein regulator of cytokinesis 1 as a prospective therapy for hepatocellular carcinoma. *Cell Death Dis.* 9:534. doi: 10.1038/s41419-018-0555-4
- Liu, Z., Li, J., Chen, J., Shan, Q., Dai, H., Xie, H., et al. (2018). MCM family in HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2 cell cycle progression. *BMC Cancer* 18:200. doi: 10.1186/s12885-018-4056-8
- Macdonald, G. A., Greenson, J. K., Saito, K., Cherian, S. P., Appelman, H. D., and Boland, C. R. (1998). Microsatellite instability and loss of heterozygosity at DNA mismatch repair gene loci occurs during hepatic carcinogenesis. *Hepatology* 28, 90–97. doi: 10.1002/hep.510280114
- Marcellin, P., Gane, E., Buti, M., Afdhal, N., Sievert, W., Jacobson, I. M., et al. (2013). Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. *Lancet* 381, 468–475. doi: 10.1016/S0140-6736(12)61425-1
- Mas, V. R., Maluf, D. G., Archer, K. J., Yanek, K., Kong, X., Kulik, L., et al. (2009). Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. *Mol. Med.* 15, 85–94. doi: 10.2119/molmed.2008.00110
- Mittal, S., and El-Serag, H. B. (2013). Epidemiology of hepatocellular carcinoma: consider the population. *J. Clin. Gastroenterol.* 47(Suppl.), S2–S6. doi: 10.1097/MCG.0b013e3182872f29
- Mjelle, R., Hegre, S. A., Aas, P. A., Slupphaug, G., Drablos, F., Saetrom, P., et al. (2015). Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair.* 30, 53–67. doi: 10.1016/j.dnarep.2015.03.007
- Panvichian, R., Tantiwetruangdet, A., Angkathunyakul, N., and Leelaudomlapi, S. (2015). TOP2A amplification and overexpression in hepatocellular carcinoma tissues. *Biomed. Res. Int.* 2015:381602. doi: 10.1155/2015/381602
- Ramakrishna, G., Rastogi, A., Trehanpati, N., Sen, B., Khosla, R., and Sarin, S. K. (2013). From cirrhosis to hepatocellular carcinoma: new molecular insights on inflammation and cellular senescence. *Liver Cancer* 2, 367–383. doi: 10.1159/000343852
- Roy, S., Hooiveld, G. J., Seehawer, M., Caruso, S., Heinzmann, F., Schneider, A. T., et al. (2018). MicroRNA 193a-5p regulates levels of nucleolar- and spindle-associated protein 1 to suppress hepatocarcinogenesis. *Gastroenterology* 155, 1951–1966.e26. doi: 10.1053/j.gastro.2018.08.032
- Sanyal, A. J., Yoon, S. K., and Lencioni, R. (2010). The etiology of hepatocellular carcinoma and consequences for treatment. *Oncologist* 15(Suppl. 4), 14–22. doi: 10.1634/theoncologist.2010-S4-14
- Schulze, K., Imbeaud, S., Letouze, E., Alexandrov, L. B., Calderaro, J., Rebouissou, S., et al. (2015). Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* 47, 505–511. doi: 10.1038/ng.3252
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shen, Q., Eun, J. W., Lee, K., Kim, H. S., Yang, H. D., Kim, S. Y., et al. (2018). Barrier to autointegration factor 1, procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3, and splicing factor 3b subunit 4 as early-stage cancer decision markers and drivers of hepatocellular carcinoma. *Hepatology* 67, 1360–1377. doi: 10.1002/hep.29606
- Sun, Y., Zhou, J., Wang, L., Wu, X., Chen, Y., Piao, H., et al. (2017). New classification of liver biopsy assessment for fibrosis in chronic hepatitis B patients before and after treatment. *Hepatology* 65, 1438–1450. doi: 10.1002/hep.29009
- 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–D368. doi: 10.1093/nar/gkw937
- Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419. doi: 10.1126/science.1260419
- Villanueva, A., Hoshida, Y., Battiston, C., Tovar, V., Sia, D., Alsinet, C., et al. (2011). Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 140, 1501–1502.e2. doi: 10.1053/j.gastro.2011.02.006
- Villanueva, A., Portela, A., Sayols, S., Battiston, C., Hoshida, Y., Mendez-Gonzalez, J., et al. (2015). DNA methylation-based prognosis and epdrivers in hepatocellular carcinoma. *Hepatology* 61, 1945–1956. doi: 10.1002/hep.27732
- Walker, M., El-Serag, H. B., Sada, Y., Mittal, S., Ying, J., Duan, Z., et al. (2016). Cirrhosis is under-recognized in patients subsequently diagnosed with hepatocellular cancer. *Aliment Pharmacol. Ther.* 43, 621–630. doi: 10.1111/apt.13505
- Wang, F., Wang, R., Li, Q., Qu, X., Hao, Y., Yang, J., et al. (2017). A transcriptome profile in hepatocellular carcinomas based on integrated analysis of microarray studies. *Diagn. Pathol.* 12:4. doi: 10.1186/s13000-016-0596-x
- Wang, Y., Shi, F., Xing, G. H., Xie, P., Zhao, N., Yin, Y. F., et al. (2017). Protein regulator of Cytokinesis PRC1 confers chemoresistance and predicts an unfavorable postoperative survival of hepatocellular carcinoma patients. *J. Cancer* 8, 801–808. doi: 10.7150/jca.17640
- Wang, J., Tian, Y., Chen, H., Li, H., and Zheng, S. (2018). Key signaling pathways, genes and transcription factors associated with hepatocellular carcinoma. *Mol. Med. Rep.* 17, 8153–8160. doi: 10.3892/mmr.2018.8871
- 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
- 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
- Xu, B., Lin, L., Xu, G., Zhuang, Y., Guo, Q., Liu, Y., et al. (2015). Long-term lamivudine treatment achieves regression of advanced liver fibrosis/cirrhosis in patients with chronic hepatitis B. *J. Gastroenterol. Hepatol.* 30, 372–378. doi: 10.1111/jgh.12718
- 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, H., Li, Z., Shen, Q., Wang, Q., Tian, J., Jiang, Q., et al. (2017). Aberrant expression of cell cycle and material metabolism related genes contributes to hepatocellular carcinoma occurrence. *Pathol. Res. Pract.* 213, 316–321. doi: 10.1016/j.prp.2017.01.019
- Yildiz, G., Arslan-Ergul, A., Bagislar, S., Konu, O., Yuzugullu, H., Gursoy-Yuzugullu, O., et al. (2013). Genome-wide transcriptional reorganization associated with senescence-to-immortality switch during human hepatocellular carcinogenesis. *PLoS One* 8:e64016. doi: 10.1371/journal.pone.0064016
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). Cluster profiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhang, M., Yang, D., Liu, X., Liu, Y., Liang, J., He, H., et al. (2013). Expression of Nusap1 in the surgical margins of hepatocellular carcinoma and its association with early recurrence. *Nan Fang Yi Ke Da Xue Xue Bao* 33, 937–938.
- Zhou, L., Du, Y., Kong, L., Zhang, X., and Chen, Q. (2018). Identification of molecular target genes and key pathways in hepatocellular carcinoma by bioinformatics analysis. *Oncol. Targets Ther.* 11, 1861–1869. doi: 10.2147/OTT.S156737

Conflict of Interest Statement: 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 © 2019 Shan, Chen and Jia. 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.