

Prediction and analysis of weighted genes in hepatocellular carcinoma using bioinformatics analysis

QIFAN ZHANG^{1*}, SHIBO SUN^{1*}, CHEN ZHU², YUJIAN ZHENG³, QING CAI³,
XIAOLU LIANG¹, HAORONG XIE¹ and JIE ZHOU¹

Departments of ¹Hepatobiliary Surgery and ²Anesthesia, Nanfang Hospital, Southern Medical University; ³Department of Hepatobiliary Surgery and Liver Transplantation Center, Guangzhou General Hospital of Guangzhou Military Area, Guangzhou, Guangdong 510515, P.R. China

Received December 28, 2017; Accepted July 5, 2018

DOI: 10.3892/mmr.2019.9929

Abstract. The aim of the present study was to identify the differentially expressed genes (DEGs) between primary tumor tissue and adjacent non-tumor tissue of hepatocellular carcinoma (HCC) samples in order to investigate the mechanisms of HCC. The microarray data of the datasets GSE76427, GSE84005 and GSE57957 were downloaded from the Gene Expression Omnibus database. DEGs were identified using the limma package in the R programming language. Following the intersection of the DEGs screened from the three datasets, 218 genes were selected for further study. A protein-protein interaction (PPI) network was constructed using the Search Tool for the Retrieval of Interacting Genes database. The construction and analysis of modules were performed using Cytoscape and the module with the highest score was selected for further analysis. Gene Ontology enrichment analysis and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis were conducted for genes involved in the PPI network and the selected subnetwork. The network of the enriched pathways and their associated genes was constructed using Cytoscape. For the genes in the global PPI network, metabolism-associated pathways were significantly enriched; whereas, for the genes in the subnetwork, 'cell cycle', 'oocyte meiosis' and 'DNA replication' pathways were significantly enriched. To demonstrate the portability and repeatability of the prognostic value of the weighted genes, a validation cohort was obtained from datasets of The Cancer Genome Atlas and Kaplan-Meier survival analysis was conducted.

Evidence is presented that the expression levels of aldehyde dehydrogenase 2 family member, cytochrome P450 family 2 subfamily C member 8, alcohol dehydrogenase 4 (class II), pi polypeptide, alcohol dehydrogenase 1B (class I), β polypeptide and cytochrome P450 family 2 subfamily C member 9 were associated with the overall survival of patients with HCC and that the expression levels of pituitary tumor-transforming 1, cell division cycle 20, DNA topoisomerase II α and cyclin B2 were negatively associated with the overall survival of patients with HCC. In conclusion, 9 weighted genes, involved in the development and progression of HCC, were identified using bioinformatics and survival analyses.

Introduction

Cancer is a principal public health problem globally and hepatocellular carcinoma (HCC) is one of the most frequently diagnosed types of cancer. HCC is the most common form of liver cancer. It is estimated that there will be 42,220 newly diagnosed cases of liver and intrahepatic bile duct cancer in the United States in 2018 (1). HCC is a complex and heterogeneous malignancy that arises in the context of progressive underlying liver dysfunction. Hepatitis B and C viruses are the primary risk factors for HCC and 80-90% of the incidence of HCC is associated with chronic viral hepatitis B or C (2,3). Recurrence is the principal cause of HCC-associated death. Five-year recurrence rates >70% have been reported despite the use of surgical or locoregional therapies in the earlier stages (4). In addition, the prognosis of patients with advanced-stage HCC is poor, with an overall survival rate <5% (5). Due to the great threat of HCC to human health, novel diagnostic and therapeutic methods are required for early cancer detection and effective treatment.

In previous years, a large number of genomic and proteomic studies have been conducted in order to examine the molecular mechanisms underlying the development and progression of HCC. The characterization of HCC has provided valuable information regarding this complex disease. Previous advances in high-throughput microarrays have received a large amount of attention and have made substantial progress in reconstructing the gene regulatory networks involved in medical biology (6). Using microarray analysis, significant differences in the levels

Correspondence to: Professor Jie Zhou, Department of Hepatobiliary Surgery, Nanfang Hospital, Southern Medical University, 1838 North of Guangzhou Avenue, Guangzhou, Guangdong 510515, P.R. China
E-mail: jacky_smu@163.com

*Contributed equally

Key words: hepatocellular carcinoma, differentially expressed genes, enrichment analysis, survival analysis, bioinformatics analysis

of gene expression in normal and diseased tissues have been observed. However, as a result of the underlying shortcomings of microarray technology, including small sample size, measurement error and information insufficiency, unveiling the disease mechanism has remained a principal challenge to HCC research (7). Therefore, Gene Ontology (GO), pathway information, network-based approaches and machine learning algorithms have been employed to identify the mechanisms underlying the development of HCC (7).

In the present study, microarray data were obtained from the Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) between primary tumor (PT) tissue and adjacent non-tumor tissue (ANTT) were identified from samples of HCC. In total, nine significant target genes for the diagnosis of HCC were identified based on GO processes, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, protein-protein interaction (PPI) networks and prognosis analysis of the clinical information from The Cancer Genome Atlas (TCGA) database.

Materials and methods

Datasets. All of the datasets included in the present study were obtained from the National Center of Biotechnology Information's GEO database (www.ncbi.nlm.nih.gov/gds/). The original mRNA expression profiles were obtained from the GSE76427 (8), GSE84005 (not published) and GSE57957 (9) datasets. In the GSE76427 dataset, there were 167 total RNA samples consisting of 115 PT tissue samples and 52 ANTT samples that were derived from 115 patients with HCC. The platform used was GPL10558 Illumina HumanHT-12 V4.0 Expression BeadChip. In the GSE84005 dataset, PT tissue samples and ANTT samples were from 38 patients with primary HCC and the platform used was GPL5175 Affymetrix Human Exon 1.0 ST Array. In the GSE57957 dataset, total RNA was obtained from 39 PT tissue samples and ANTT samples and the platform used was GPL10558 Illumina HumanHT-12 V4.0 Expression BeadChip. The clinical information of the patients was included in the three datasets (data not shown).

Identification of DEGs. Background correction and quartile data normalization of the downloaded data were performed using the robust multi-array average algorithm (10). Probes without a corresponding gene symbol were subsequently filtered and the average value of the gene symbols with multiple probes was calculated. Student's t-test and fold change (FC) filtering were conducted to screen the DEGs between two groups by using the R software (version 3.4.2; www.R-project.org/) limma (version 3.32.3) package (11,12). With a threshold of P-value <0.05 and absolute value of FC >2, volcano plot filtering was performed using the R software ggplot2 package to identify the DEGs with statistical significance between PT tissue samples and ANTT samples. The DEGs from the three datasets were intersected and the DEGs with different expression tendency were eliminated.

PPI and the module analysis. The Search Tool for the Retrieval of Interacting Genes (STRING) database (string-db.org) provides uniquely comprehensive coverage of, and ease of access to experimental and predicted interaction information.

To better understand the DEGs from an interactive perspective, a PPI network was built based on information from the STRING database. A combined score of >0.4 was set as the reliability threshold. Cytoscape is a useful tool for integrated analysis and visualization of biological networks. Cytoscape software (version 3.4) was used to visualize the PPI network (13). The network topology was analyzed using the CentiScaPe app (14) and the module analysis was conducted using the MCODE app (15).

Gene function analysis. GO enrichment analysis of the DEGs was implemented using DAVID (<http://david.abcc.ncifcrf.gov/>). GO terms ('molecular function', 'biological processes' and 'cellular components') with a P-value <0.05 were considered significantly enriched by the DEGs. KEGG is a database resource for understanding high-level functions and effects of the biological system (<http://www.genome.jp/kegg/>). DAVID was additionally used to test the statistical enrichment of genes or target genes of microRNA that were differentially expressed in KEGG pathways. The networks of the pathways (P<0.05) and pathway-associated genes were constructed by using the Cytoscape (version 3.4.0) plugin ClueGO (16) in addition to the Cluepedia (17) app. The network topology was analyzed using the CentiScaPe app (14). The genes that were associated with at least three pathways (degree ≥ 3) were defined as cross-talk genes.

TCGA datasets analysis. TCGA is a platform for researchers to download and assess free public datasets. In the present study, the prognostic value of the weighted genes was confirmed by Kaplan-Meier survival analysis based on the clinical information from TCGA datasets using OncoLnc (18). For the statistical analysis, the patients were divided into two groups based on gene expression values. Specifically, patients with expression values greater than the median value were classified into the high expression group, while the rest were classified into the low expression group.

Results

Screening of DEGs. With a threshold of P-value <0.05 and absolute value of FC >1, DEGs were identified from each of the three datasets. 494 DEGs were screened from the GSE76427 dataset, which consisted of 92 upregulated genes and 402 downregulated genes. A total of 1,005 DEGs were screened from the GSE84005 dataset, consisting of 478 upregulated genes and 527 downregulated genes. 417 DEGs were screened from the GSE57957 dataset, consisting of 109 upregulated genes and 308 downregulated genes. Volcano plots were used to visualize differential expression of genes between the tumor group and non-tumor group (Fig. 1A-C). Subsequent to the DEGs from the three datasets being intersected, 218 DEGs were selected for further analysis (Fig. 1D).

Construction of the PPI network and module analysis. Following elimination of the DEGs with different expression tendency and isolated nodes in the network, a total of 170 nodes and 666 edges were included in the PPI network (Fig. 2A). The top five DEGs with the highest degrees were DNA topoisomerase II α (TOP2A; degree=42),

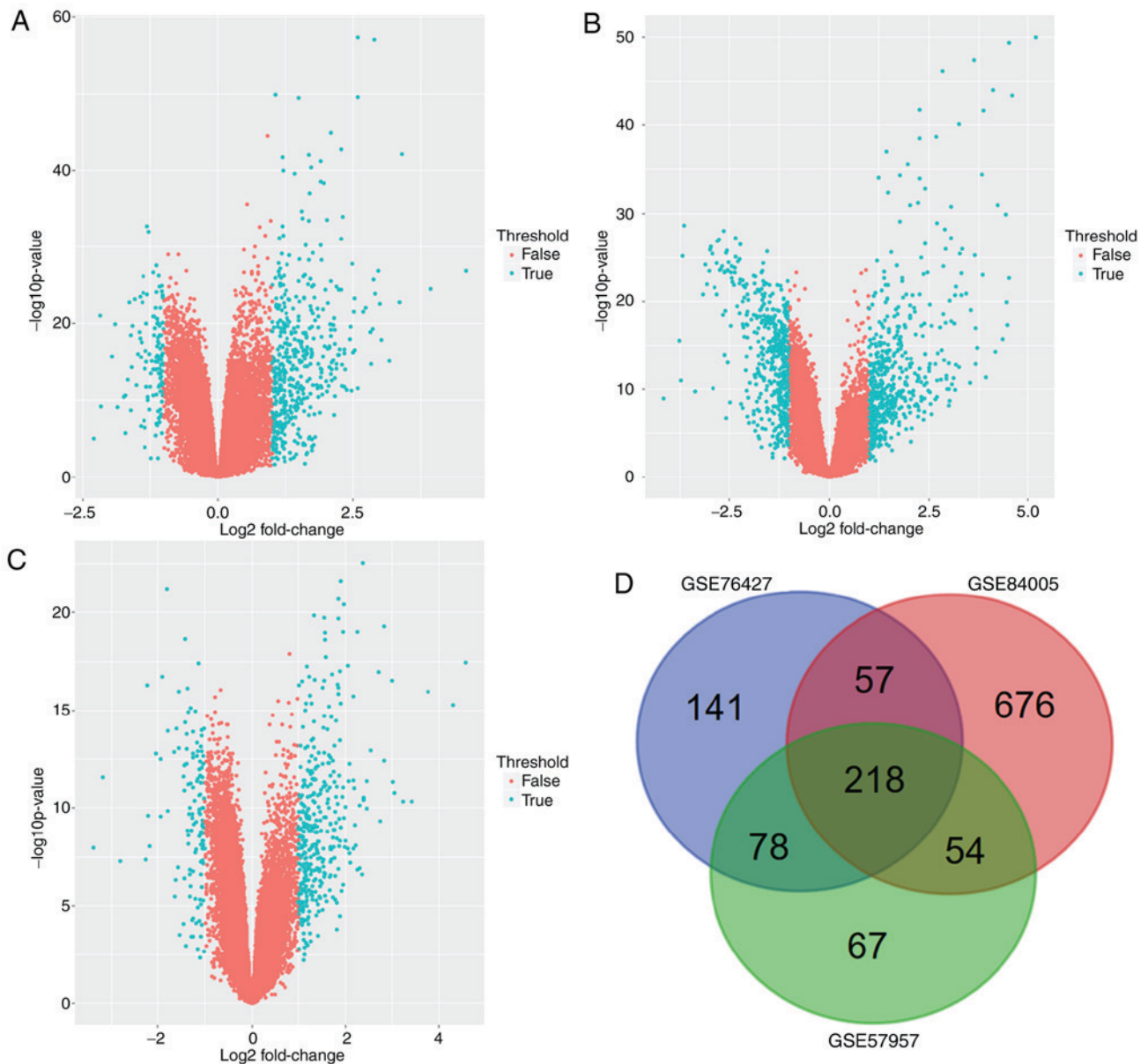


Figure 1. Identification of DEGs. Volcano plots of the gene expression data from the (A) GSE76427, (B) GSE84005 and (C) GSE57957 datasets. The horizontal axis represents the \log_2 (fold change) and the vertical axis represents the $-\log_{10}$ (P-value). The blue plots represent the selected DEGs. (D) 218 DEGs were selected for further analysis after the DEGs from three datasets were intersected. DEGs, differentially expressed genes.

CYP2E1 (cytochrome P450 family 2 subfamily E member 1; degree=27), cell division cycle 20 (CDC20; degree=22), cytochrome P450 family 2 subfamily C member 9 (CYP2C9; degree=22) and kynurenine 3-monooxygenase (degree=22). Furthermore, a total of 12 modules were separated, and the module with the highest MCODE score (of 17.765) was selected for further analysis (Fig. 2B). In this module, a total of 18 DEGs, including TOP2A, were involved and all of them were downregulated genes.

GO and KEGG enrichment analysis of the selected DEGs. GO and KEGG enrichment analysis were performed for the selected 218 mRNAs to investigate the biological functions of these genes. In the GO analysis, all of the results were ranked by enrichment score [$-\log(P\text{-value})$] and the top 10 results of each category are presented in Fig. 3. In the analysis of the

'biological processes', 'oxidation-reduction process', 'drug metabolic process' and 'epoxygenase P450 pathway' were the top three enriched terms. In the 'cellular component' analysis, 'organelle membrane', 'extracellular region' and 'extracellular space' were the top three enriched terms. In the 'molecular function' analysis, 'heme binding', 'oxidoreductase activity and acting on paired donors with incorporation or reduction of molecular oxygen' and 'monooxygenase activity' were the top three enriched terms. The results of the KEGG pathway analysis were additionally ranked by enrichment score and the pathways with $P\text{-value} < 0.05$ are demonstrated in Fig. 4A. The top three enriched pathways were 'metabolic pathways', 'retinol metabolism' and 'chemical carcinogenesis'. The KEGG pathway network composed of the significantly enriched pathways is presented in Fig. 4B and it revealed that a number of metabolism-associated

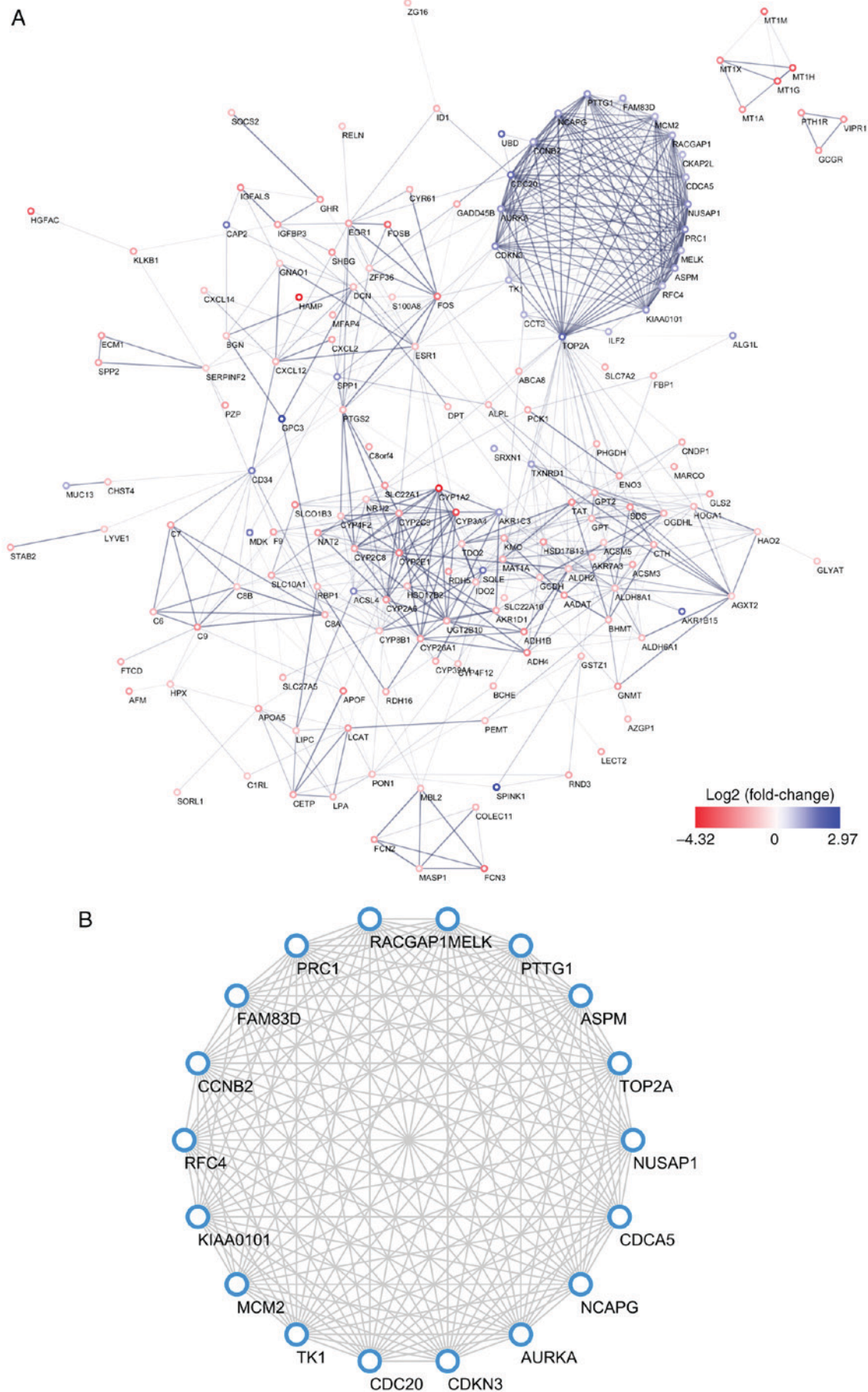


Figure 2. Protein-protein interaction network of the DEGs. (A) The global network and (B) the module obtained from the global network with the highest score. The red nodes represent the downregulated DEGs and the blue nodes indicate the upregulated DEGs. The molecular interactions between the DEGs are indicated by edges and the deeper the color of edges, the higher the combined score. DEGs, differentially expressed genes.

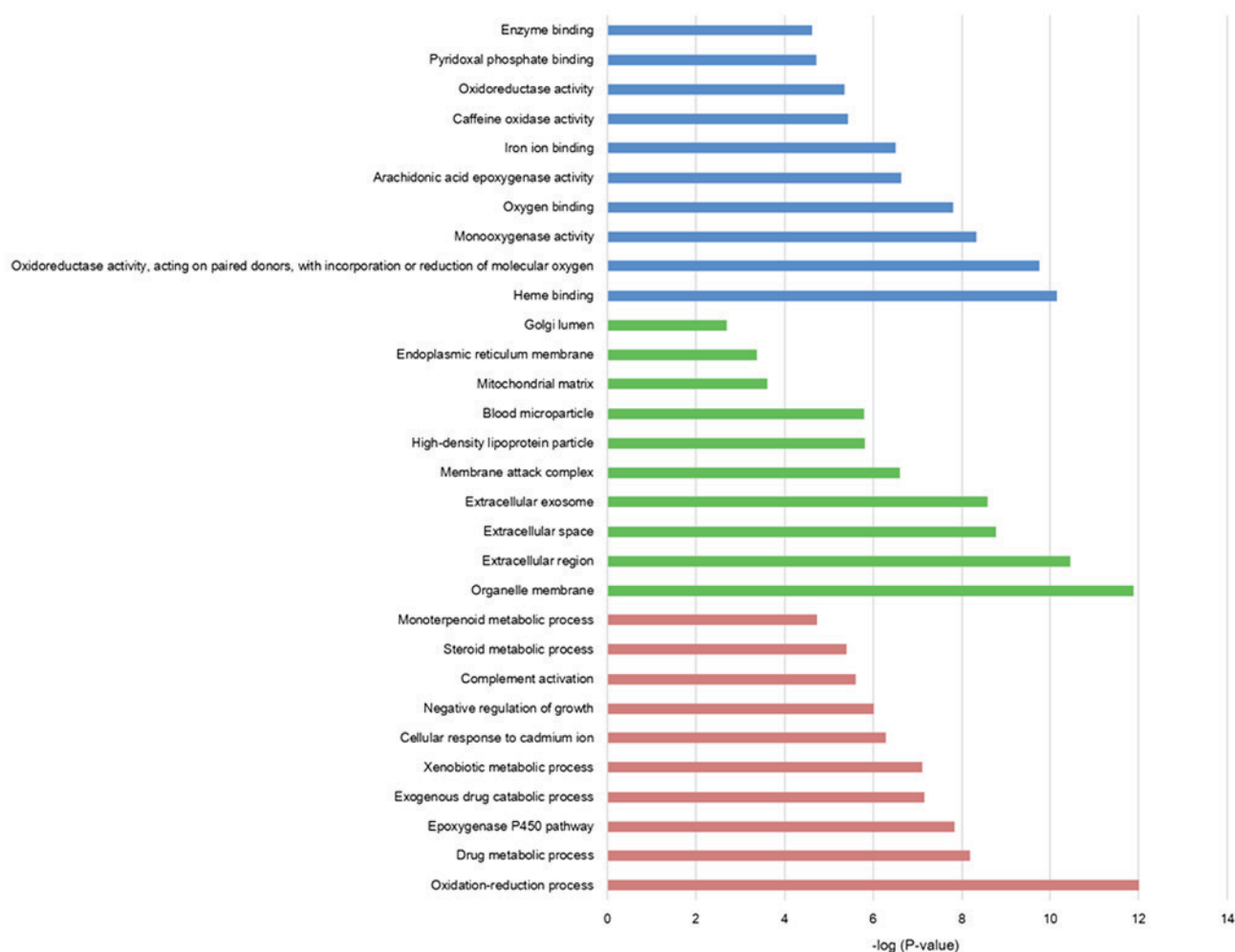


Figure 3. Top 10 enrichment scores in the Gene Ontology enrichment analysis of the selected mRNAs. Red bars represent 'biological process' terms, green bars represent 'cell component' terms and blue bars represent 'molecular function' terms.

pathways were enriched. The network constructed with the pathways and their associated genes (Fig. 4C) revealed that CYP1A2, alcohol dehydrogenase 1B (class I), β polypeptide (ADH1B), alcohol dehydrogenase 4 (class II), pi polypeptide (ADH4), cytochrome P450 family 3 subfamily A member 4, cytochrome P450 family 2 subfamily A member 6 (CYP2A6), CYP2C9, CYP2E1, cytochrome P450 family 2 subfamily C member 8 (CYP2C8), alanine-glyoxylate aminotransferase 2, aldehyde dehydrogenase 2 family member (ALDH2), N-acetyltransferase 2 and UDP glucuronosyltransferase family 2 member B10 were cross-talk genes associated with at least three pathways. KEGG pathway analysis was additionally conducted for the 18 DEGs from the subnetwork and it revealed that 'cell cycle', 'oocyte meiosis' and 'DNA replication' pathways were significantly enriched (Fig. 5). In the KEGG pathway network, CDC20, cyclin B2 (CCNB2) and pituitary tumor-transforming 1 (PTTG1) were associated with 'cell cycle' and 'oocyte meiosis', and minichromosome maintenance complex component 2 was associated with 'cell cycle' and 'DNA replication'.

TCGA datasets analysis. To demonstrate the portability and repeatability of the prognostic value of the weighted genes, a validation cohort was obtained from TCGA datasets and

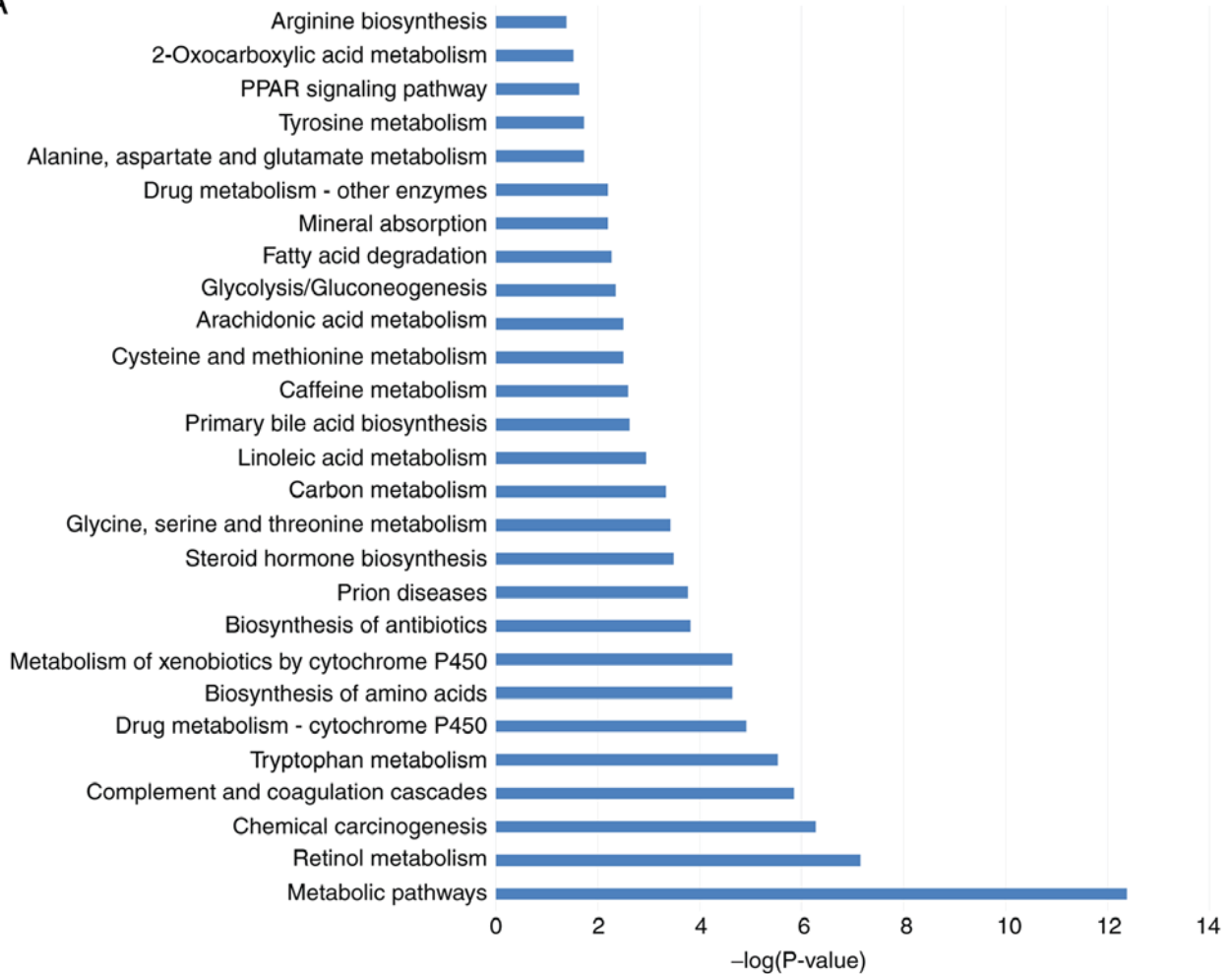
Kaplan-Meier survival analysis was performed. The log-rank test confirmed that high expression levels of PTTG1, CDC20, TOP2A and CCNB2 and low expression levels of ALDH2, CYP2C8, ADH4, ADH1B and CYP2C9 were negatively associated with the overall survival of patients with HCC (Fig. 6).

Discussion

HCC is the leading cause of cancer-associated mortality worldwide, owing to limited insights into the pathogenesis and the unsatisfactory efficacy of current therapies. Even in cases of curative surgical treatment, recurrence is common. Sorafenib and regorafenib, two oral multi-kinase inhibitors, are the only therapeutic agents that have been demonstrated to be effective in the treatment of advanced HCC (19,20), thus novel curative approaches are urgently required.

The human cytochrome P450 (CYP) enzymes are a superfamily comprised of >50 different genes categorized into 18 families, which share ~40% sequence homology (21). CYPs are primarily expressed in the liver and their primary role is the metabolism of xenobiotics in order to protect the organism from xenobiotics and environmental toxins (22). Due to the important role of CYPs in drug metabolism, the alterations

A



B

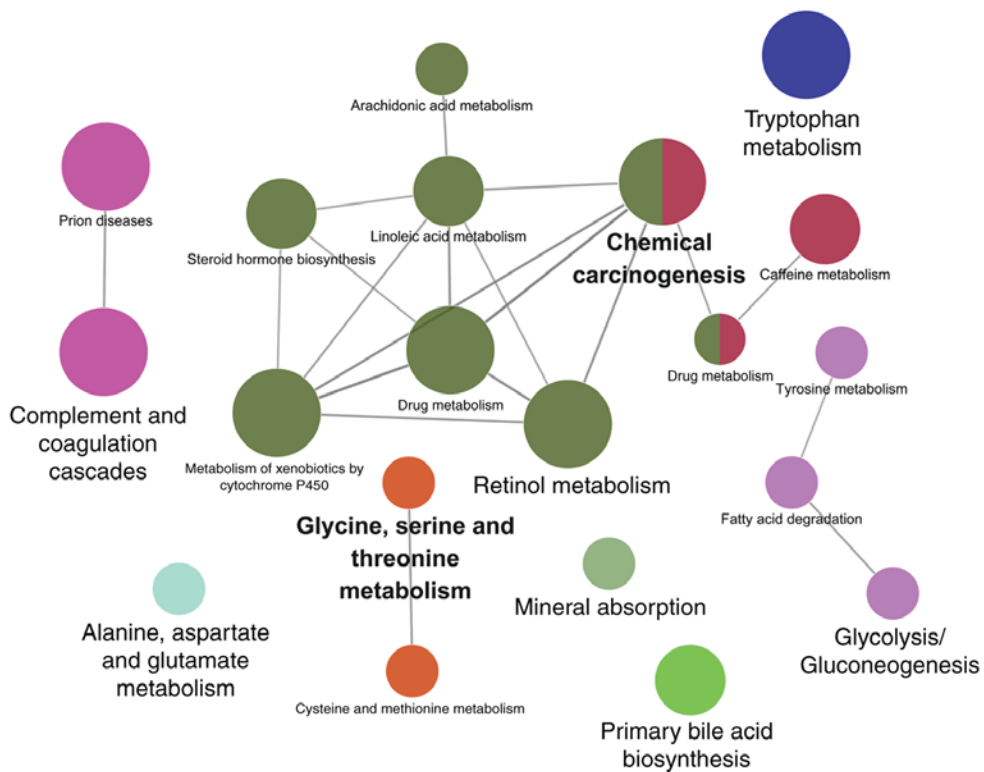


Figure 4. KEGG pathway enrichment analysis of the selected mRNAs. (A) The significantly enriched pathways ($P < 0.05$) in the KEGG pathway analysis of the selected mRNAs. The network of significantly enriched pathways ($P < 0.05$) (B) without associated genes. Larger nodes represent larger enrichment scores and the different colors represent different enrichment modules.

C

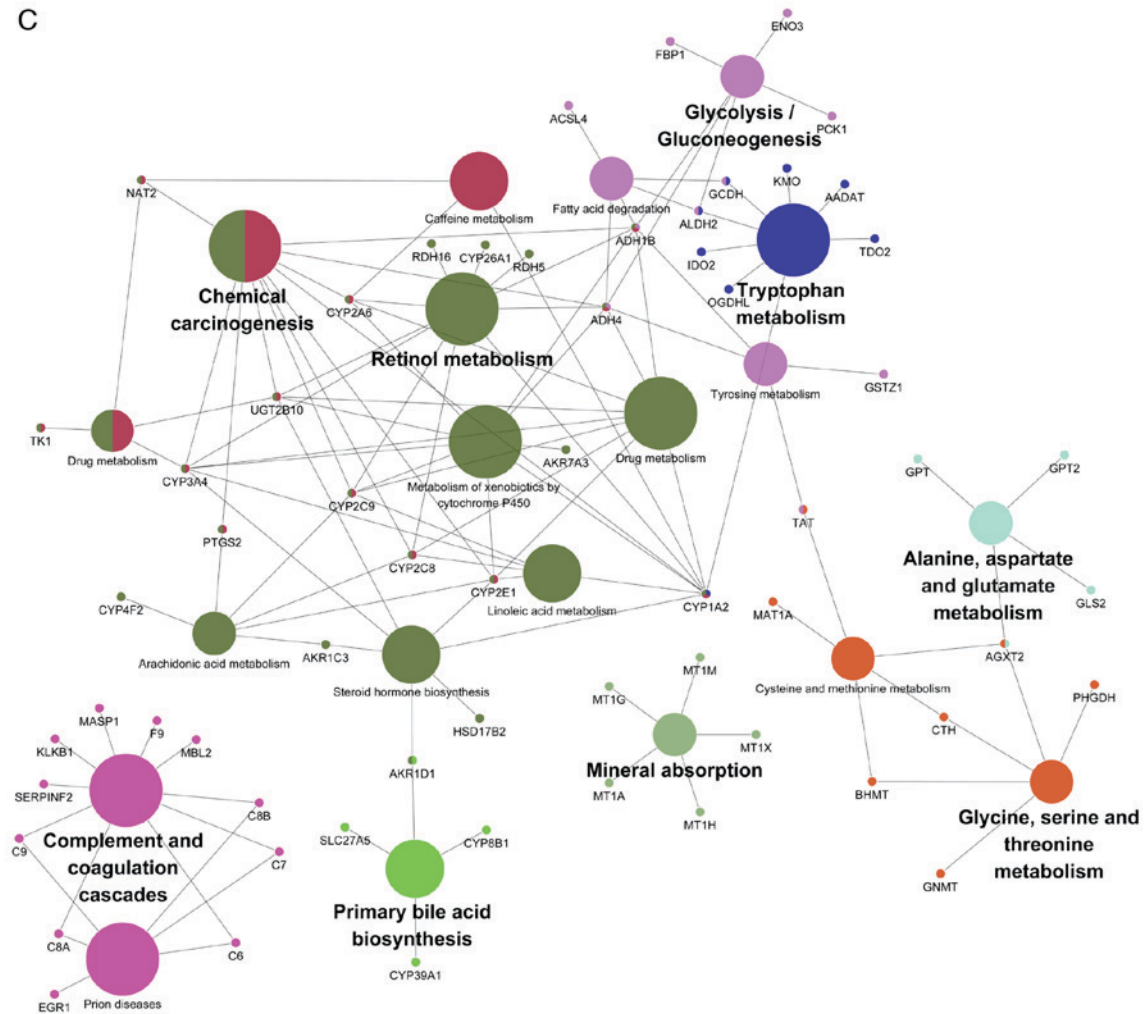


Figure 4. Continued. KEGG pathway enrichment analysis of the selected mRNAs. (C) The network of significantly enriched pathways ($P < 0.05$) with associated genes. Larger nodes represent larger enrichment scores and the different colors represent different enrichment modules. KEGG, Kyoto Encyclopedia of Genes and Genomes.

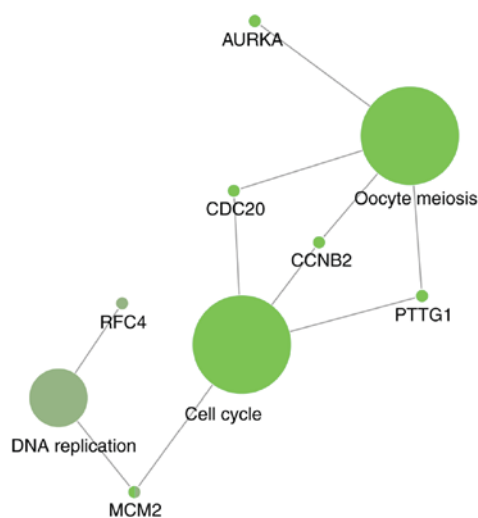


Figure 5. KEGG pathway enrichment analysis of the genes involved in the subnetwork. The network of the significantly enriched pathways ($P < 0.05$) with their associated genes. Larger nodes represent larger enrichment scores and the different colors represent different enrichment modules. KEGG, Kyoto Encyclopedia of Genes and Genomes; AURKA, aurora kinase A; CDC20, cell division cycle 20; CCNB2, cyclin B2; PTTG1, pituitary tumor-transforming 1; RFC4, replication factor C subunit 4; MCM2, minichromosome maintenance complex component 2.

in CYP activity caused by HCC may influence the pharmacokinetics of drugs used in treating HCC. A number of studies have reported that dysregulation of CYPs, including CYP2A6, CYP2C9, CYP2E1 and CYP3A5, serves an important role in the development of HCC (23-25). In the present study, evidence is presented that downregulation of CYP2C8 and CYP2C9 is associated with poor prognosis and all of the CYPs in the PPI network were downregulated.

The expression levels of four of the genes screened from the subnetwork, PTTG1, CDC20, TOP2A and CCNB2, negatively associated with the overall survival of patients with HCC. The prognostic value of TOP2A in HCC has been highlighted prior; Wong *et al* (26) identified that overexpression of TOP2A in HCC was associated with early-age onset, shorter survival and chemo-resistance. The results of the KEGG pathway analysis in the present study demonstrated that CDC20, CCNB2 and PTTG1 were associated with 'cell cycle' and 'oocyte meiosis', suggesting that overexpression of these genes may be responsible for the proliferation of the tumor cells. It has been reported in previous studies that increased expression levels of CDC20 and PTTG1 are associated with the development and progression of HCC (27,28); however, the association

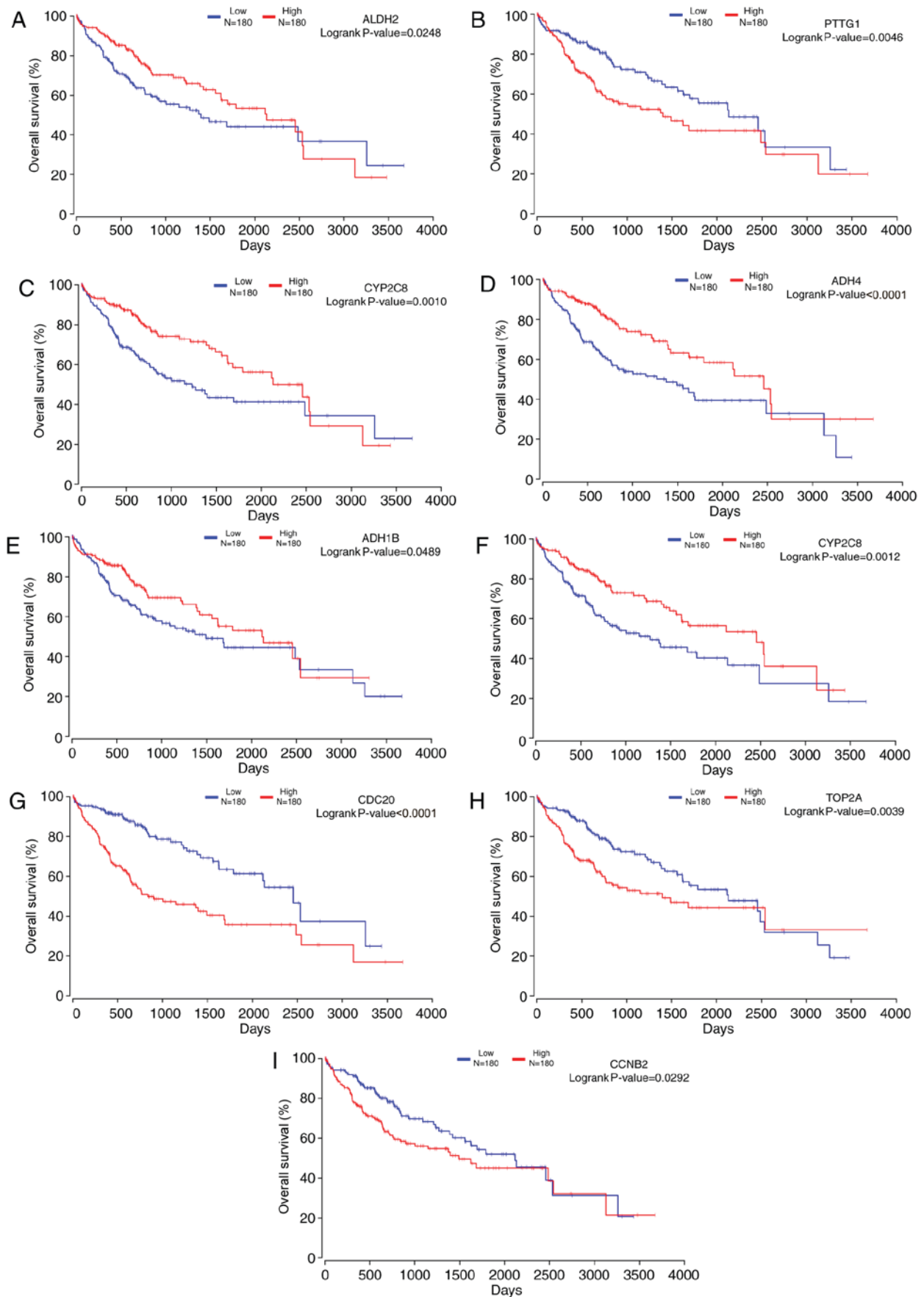


Figure 6. Kaplan-Meier survival analysis of the expression levels of the weighted genes and the overall survival of patients with HCC. Expression levels of (A) ALDH2, (B) PTTG1, (C) CYP2C8, (D) ADH4, (E) ADH1B, (F) CYP2C8, (G) CDC20, (H) TOP2A and (I) CCNB2 are presented. HCC, hepatocellular carcinoma; ALDH2, aldehyde dehydrogenase 2 family member; CYP2C8, cytochrome P450 family 2 subfamily C member 8; ADH4, alcohol dehydrogenase 4 (class II), pi polypeptide; ADH1B, alcohol dehydrogenase 1B (class I), β polypeptide; CYP2C9, cytochrome P450 family 2 subfamily C member 9; PTTG1, pituitary tumor-transforming 1; CDC20, cell division cycle 20; TOP2A, DNA topoisomerase II α ; CCNB2, cyclin B2.

between CCNB2 and HCC is still largely unknown. The prognostic value of PTTG1, CDC20 and CCNB2 requires further evaluation and the interaction between the genes in the subnetwork requires further investigation.

From the systematic bioinformatics analysis and survival analysis of the clinical information from TCGA, another three key genes involved in HCC, ALDH2, ADH4 and ADH1B, were screened. ADH4 is an important member of the ADH family that metabolizes a wide variety of substrates, including ethanol and retinol (29). Previous studies have demonstrated that ADH4 is involved in cancer, including HCC (30,31). Mitochondrial ALDH2 in the liver removes toxic aldehydes including acetaldehyde, an intermediate of ethanol metabolism, and ALDH2 mutation increases protein turnover and promotes murine HCC (32). **To the best of the authors' knowledge**, there is no study regarding the association between ADH1B and HCC.

In conclusion, nine weighted genes involved in the development and progression of HCC were identified using bioinformatics analysis and survival analysis. However, further experimental verification is required to confirm the potential effects of the weighted genes in HCC.

Acknowledgements

Not applicable.

Funding

The present study was supported by the National Key Clinical Specialty Discipline Construction Program, National Natural Science Foundation of China (grant no. 81500398); Natural Science Foundation of Guangdong Province (grant no. 2015A030310480); Guangdong Key Laboratory of Liver Disease Research (grant no. GS2017101003) and President Foundation of Nanfang Hospital, Southern Medical University (grant no. 2016C013).

Availability of data and materials

The datasets analyzed during the current study are available in National Center of Biotechnology Information's GEO database (www.ncbi.nlm.nih.gov/gds/) with the accession numbers GSE76427, GSE84005 and GSE57957.

Authors' contributions

JZ designed the study. QZ and SS performed the bioinformatics analysis. QC and XL analyzed the data. CZ and YZ collected the data for analysis and wrote the manuscript. HX participated in the design of the study and created the figures.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD and Jemal A: Cancer statistics, 2018. *CA Cancer J Clin* 68: 7-30, 2018.
2. Walter SR, Thein HH, Gidding HF, Amin J, Law MG, George J and Dore GJ: Risk factors for hepatocellular carcinoma in a cohort infected with hepatitis B or C. *J Gastroenterol Hepatol* 26: 1757-1764, 2011.
3. Yang JD and Roberts LR: Hepatocellular carcinoma: A global view. *Nat Rev Gastroenterol Hepatol* 7: 448-458, 2010.
4. Imamura H, Matsuyama Y, Tanaka E, Ohkubo T, Hasegawa K, Miyagawa S, Sugawara Y, Minagawa M, Takayama T, Kawasaki S and Makuuchi M: Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *J Hepatol* 38: 200-207, 2003.
5. Schutte K, Bornschein J and Malfertheiner P: Hepatocellular carcinoma-epidemiological trends and risk factors. *Dig Dis* 27: 80-92, 2009.
6. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G and Kallioniemi OP: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4: 844-847, 1998.
7. Wang L, Zang W, Xie D, Ji W, Pan Y, Li Z, Shen J and Shi Y: Comparison of hepatocellular carcinoma (HCC), cholangiocarcinoma (CC), and combined HCC-CC (CHC) with each other based on microarray dataset. *Tumour Biol* 34: 1679-1684, 2013.
8. Grinchuk OV, Yenamandra SP, Iyer R, Singh M, Lee HK, Lim KH, Chow PK and Kuznetsov VA: Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. *Mol Oncol* 12: 89-113, 2018.
9. Mah WC, Thurnherr T, Chow PK, Chung AY, Ooi LL, Toh HC, Teh BT, Sauntharajah Y and Lee CG: Methylation profiles reveal distinct subgroup of hepatocellular carcinoma patients with poor prognosis. *PLoS One* 9: e104158, 2014.
10. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B and Speed TP: Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31: e15, 2003.
11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK: Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47, 2015.
12. R Core Team: R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, 2013.
13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
14. Scardoni G, Pletterlini M and Laudanna C: Analyzing biological network parameters with CentiScaPe. *Bioinformatics* 25: 2857-2859, 2009.
15. Bader GD and Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2, 2003.
16. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z and Galon J: ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25: 1091-1093, 2009.
17. Bindea G, Galon J and Mlecnik B: CluePedia cytoscape plugin: Pathway insights using integrated experimental and in silico data. *Bioinformatics* 29: 661-663, 2013.
18. Anaya J: OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *Peer J Comp Sci* 2: e67, 2016.
19. Razumilava N and Gores GJ: Sorafenib for HCC: A pragmatic perspective. *Oncology* 25: 300-302, 2011.
20. Killock D: Liver cancer: Regorafenib-A new RESORCE in HCC. *Nat Rev Clin Oncol* 14: 70-71, 2017.
21. Nebert DW and Gonzalez FJ: P450 genes: Structure, evolution, and regulation. *Annu Rev Biochem* 56: 945-993, 1987.

22. Chavan H, Li F, Tessman R, Mickey K, Dorko K, Schmitt T, Kumer S, Gunewardena S, Gaikwad N and Krishnamurthy P: Functional coupling of ATP-binding cassette transporter Abcb6 to cytochrome P450 expression and activity in liver. *J Biol Chem* 290: 7871-7886, 2015.
23. Raunio H, Juvonen R, Pasanen M, Pelkonen O, Pääkkö P and Soini Y: Cytochrome P4502A6 (CYP2A6) expression in human hepatocellular carcinoma. *Hepatology* 27: 427-432, 1998.
24. Tsunedomi R, Iizuka N, Hamamoto Y, Uchimura S, Miyamoto T, Tamesa T, Okada T, Takemoto N, Takashima M, Sakamoto K, *et al*: Patterns of expression of cytochrome P450 genes in progression of hepatitis C virus-associated hepatocellular carcinoma. *Int J Oncol* 27: 661-667, 2005.
25. Xu XR, Huang J, Xu ZG, Qian BZ, Zhu ZD, Yan Q, Cai T, Zhang X, Xiao HS, Qu J, *et al*: Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc Natl Acad Sci USA* 98: 15089-15094, 2001.
26. Wong N, Yeo W, Wong WL, Wong NL, Chan KY, Mo FK, Koh J, Chan SL, Chan AT, Lai PB, *et al*: TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance. *Int J Cancer* 124: 644-652, 2009.
27. Li J, Gao JZ, Du JL, Huang ZX and Wei LX: Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. *Int J Oncol* 45: 1547-1555, 2014.
28. Fujii T, Nomoto S, Koshikawa K, Yatabe Y, Teshigawara O, Mori T, Inoue S, Takeda S and Nakao A: Overexpression of pituitary tumor transforming gene 1 in HCC is associated with angiogenesis and poor prognosis. *Hepatology* 43: 1267-1275, 2006.
29. Edman K and Maret W: Alcohol dehydrogenase genes: Restriction fragment length polymorphisms for ADH4 (pi-ADH) and ADH5 (chi-ADH) and construction of haplotypes among different ADH classes. *Hum Genet* 90: 395-401, 1992.
30. Wei RR, Zhang MY, Rao HL, Pu HY, Zhang HZ and Wang HY: Identification of ADH4 as a novel and potential prognostic marker in hepatocellular carcinoma. *Med Oncol* 29: 2737-2743, 2012.
31. Goode EL, White KL, Vierkant RA, Phelan CM, Cunningham JM, Schildkraut JM, Berchuck A, Larson MC, Fridley BL, Olson JE, *et al*: Xenobiotic-Metabolizing gene polymorphisms and ovarian cancer risk. *Mol Carcinog* 50: 397-402, 2011.
32. Jin S, Chen J, Chen L, Histen G, Lin Z, Gross S, Hixon J, Chen Y, Kung C, Chen Y, *et al*: ALDH2 (E487K) mutation increases protein turnover and promotes murine hepatocarcinogenesis. *Proc Natl Acad Sci USA* 112: 9088-9093, 2015.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.