


Bioinformatic identification of differentially expressed genes associated with hepatocellular carcinoma prognosis

Xu Huang, MD^a, Xu Wang, MD^b, Ge Huang, MS^c, Ruotao Li, MD^d, Xingkai Liu, MD^a, Lidong Cao, MD^e, Junfeng Ye, MD^a, Ping Zhang, MD^{a,*} 

Abstract

Hepatocellular carcinoma (HCC) is still a significant global health problem. The development of bioinformatics may provide the opportunities to identify novel therapeutic targets. This study bioinformatically identified the differentially expressed genes (DEGs) in HCC and associated them with HCC prognosis using data from published databases. The DEGs downloaded from the Gene Expression Omnibus (GEO) website were visualized using the Venn diagram software, and then subjected to the GO and KEGG analyses, while the protein–protein interaction network was analyzed using Cytoscape software with the Search Tool for the search tool for the retrieval of interacting genes and the molecular complex detection plug-in. Kaplan–Meier curves and the log rank test were used to associate the core PPI network genes with the prognosis. There were 57 upregulated and 143 downregulated genes in HCC samples. The GO and pathway analyses revealed that these DEGs are involved in the biological processes (BPs), molecular functions (MFs), and cell components (CCs). The PPI network covered 50 upregulated and 108 downregulated genes, and the core modules of this PPI network contained 34 upregulated genes. A total of 28 of these upregulated genes were associated with a poor HCC prognosis, 27 of which were highly expressed in HCC tissues. This study identified 28 DEGs to be associated with a poor HCC prognosis. Future studies will investigate their possible applications as prognostic biomarkers and potential therapeutic targets for HCC.

Abbreviations: BPs = biological processes, BUB1B = mitotic checkpoint serine/threonine kinase B, CC = cell components, CCNB1 = cyclin-B1, CDK1 = cyclin-dependent kinase 1, DEGs = the differentially expressed genes, FC = fold change, GEO = the Gene Expression Omnibus, GEPIA = Gene Expression Profiling Interactive Analysis, GO = Gene Ontology, HCC = hepatocellular carcinoma, KEGG = Kyoto Encyclopedia of Genes and Genomes, MAD2L1 = Mitotic arrest deficient 2 like 1, miRNAs = microRNA, PPI = protein–protein interaction.

Keywords: bioinformatics analysis, differentially expressed genes, gene expression omnibus, hepatocellular carcinoma, prognosis

1. Introduction

Hepatocellular carcinoma (HCC) is the fifth most commonly diagnosed human cancer in the world,^[1] accounting for approximately 841,080 new annual cases and approximately 781,631 cancer-related deaths globally.^[2] In China, there were 461,600 new HCC cases and 422,100 HCC-related deaths in 2015.^[3] Clinically, HCC patients are usually diagnosed at an intermediate to an advanced stage of disease, thus making total surgical tumor resection or percutaneous ablation unlikely; thus, liver transplantation followed by radiotherapy and transcatheter arterial chemoembolization has become the best treatment of choice.^[4] Unfortunately, HCC has a high

recurrence rate even after a successful surgical resection of tumor lesion(s), which is likely due to the insensitivity of HCC to chemotherapy and radiotherapy after surgery.^[5] Therefore, further studies of the underlying mechanisms of HCC development could help us to identify and evaluate the novel HCC biomarkers for prediction of treatment outcomes and prognosis of HCC patients.

In recent years, the high-throughput sequencing and bioinformatics research have sped up the analytical process and understanding of gene alterations in human diseases, including cancer.^[6] Using microarray technology, researchers have revealed many differentially expressed genes (DEGs) involved

This study was supported in part by a grant from the Department of Science and Technology of Jilin Province, China (#20190304049YY). However, this funding agency has no role in our study design, data collection, analysis, and interpretation, and preparation of this manuscript.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

^a Department of Hepatobiliary and Pancreatic Surgery, The First Bethune Hospital of Jilin University, Changchun, China, ^b Department of Neurology, The First Bethune Hospital of Jilin University, Changchun, China, ^c Department of Radiology, The Second Bethune Hospital of Jilin University, Changchun, China, ^d Department of Hand and Foot Surgery, The First Bethune Hospital of Jilin University, Changchun, China, ^e Department of Hepatobiliary and Pancreatic Surgery, The Second Bethune Hospital of Jilin University, Changchun, China.

**Correspondence: Ping Zhang, Department of Hepatobiliary and Pancreatic Surgery, The First Bethune Hospital of Jilin University, 71 Xinmin Street, Changchun 130021, China (e-mail: azhangpinga@126.com).*

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Huang X, Wang X, Huang G, Li R, Liu X, Cao L, Ye J, Zhang P. Bioinformatic identification of differentially expressed genes associated with hepatocellular carcinoma prognosis. Medicine 2022;101:38(e30678).

Received: 14 April 2022 / Received in final form: 21 August 2022 / Accepted: 22 August 2022

<http://dx.doi.org/10.1097/MD.00000000000030678>

in HCC initiation, development, progression, and metastasis. For example, Ally *et al* have shown that the wingless and Int-1 signaling pathway as well as mouse double minute 4 homolog, MET, vascular endothelial growth factor, myeloid cell leukemia protein 1, isocitrate dehydrogenase 1, and telomerase reverse transcriptase are aberrantly expressed or activated, suggesting that these genes could be potential therapeutic targets for HCC; whereas aberrant expression of cytotoxic T-lymphocyte-associated protein 4, programmed cell death protein 1 and programmed death-ligand 1, which are all important immune checkpoint proteins, was not observed in the control tissues.^[7] In addition, Mou *et al* have identified differentially expressed miRNAs, like hsa-mir-221, and its interactions with estrogen receptor alpha and C-X-C motif chemokine 12 as potential targets for early HCC detection and treatment.^[8] Moreover, many other bioinformatics studies have been conducted recently, further improving the bioinformatics methods and providing a better understanding of the underlying HCC mechanisms.^[9]

Therefore, in this study, we searched and downloaded the GSE101685, GSE62232, and GSE112790 datasets from the NCBI Gene Expression Omnibus (GEO) website and identified DEGs and then performed GO and KEGG pathway enrichment analyses. Thereafter, we also established the protein–protein interaction (PPI) network using molecular complex detection for these DEGs and then associated the core DEGs with the HCC prognosis using the Kaplan–Meier plotter and re-analyzed these core DEGs. This study could provide a novel insight into HCC for the future development of novel biomarkers and therapeutic targets for HCC.

2. Methods

2.1. Collection of public gene expression data

This study searched and downloaded 3 gene expression profiles from HCC and the corresponding normal specimens (GSE101685, GSE62232, and GSE112790) from the GEO database (<https://www.ncbi.nlm.nih.gov/gds>), data of which obtained from analysis of the GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array; Affymetrix, Santa Clara, CA, USA) in 24 cases of HCCs versus 8 normal liver, 81 HCCs versus 10 normal liver, and 183 HCC versus 15 normal liver specimens, respectively, resulting in a total of 321 samples (288 HCC and 32 nontumor liver tissue samples).

2.2. Identification of DEGs

DEGs in HCC versus normal liver tissue specimens were identified by using the GEO2R online tools.^[10] Specifically, the raw data were first analyzed using the online Venn diagram software (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to identify DEGs from these 3 datasets. If the DEGs met the criteria of $|\log_2 \text{fold change (FC)}| \geq 2.0$ and adjusted P -value $< .05$, these genes were considered as upregulated genes, whereas the DEGs with a log FC value < 0 were considered as a downregulated ones.

2.3. The GO and KEGG analyses

The GO term analysis was applied to define genes and their products, mRNA, or proteins in order to evaluate the unique biological signaling of the high-throughput transcriptome or genomic data.^[11] The KEGG database was used to integrate the currently known PPI network information for the metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development^[12] using the database for annotation, visualization and integrated discovery 6.8 (<https://david.ncifcrf.gov>), an online bioinformatics tool designed to interpret

biological function annotations of genes or proteins.^[13] In our study, we utilized the database for annotation, visualization and integrated discovery to identify the enriched biological processes (BPs), molecular functions (MFs), and cellular components (CCs) as well as pathways with $P < .05$ as the cutoff criterion for these DEGs in HCC.

2.4. The PPI network construction and analysis

We used an online Search Tool of the search tool for the retrieval of interacting genes (the Retrieval of Interacting Genes) to assess the potential interactions of these DEG-related proteins.^[14] The cutoff criteria were as follows: required confidence level ≥ 0.4 , while the maximum interactor number = 0.^[15] We then used Cytoscape software to visualize the PPI network using the molecular complex detection, version 1.31 functions (the degree cutoff value = 2; the maximum depth = 100; k-core = 2; and node score cutoff = 0.2 as significance).

2.5. Analysis of HCC prognosis and GEPIA confirmation of DEGs

We conducted HCC survival analysis using the European Genome-phenome Archive, The Cancer Genome Atlas, and GEO databases^[16] using the Kaplan–Meier plotter (www.kmplot.com), which is capable of assessing the effects of 54,000 genes for association with the prognosis of 21 cancer types in the HGU133 Plus 2.0 array, with mean follow-up time of up to 69 months.^[17] After that, we performed GEPIA via an interactive website (<http://gepia.cancer-pku.cn>)^[18] and further assessed these DEGs values in HCC.

2.6. Statistical analysis

The data were downloaded and analyzed using the R statistical software (version 4.0.2^[19]). All other analyses were incorporated directly in the above subtitles. Overall, a P value $< .05$ was considered statistically significant.

3. Results

3.1. Identification of HCC-related DEGs

In this study, we searched and downloaded 3 gene expression profiles (GSE101685, GSE62232, and GSE112790 containing 288 HCC and 32 normal liver samples) from the GEO database and obtained 723, 385, and 444 DEGs, respectively. We then used the cutoff value of $\log_2 \text{FC} \geq 2.0$ or < 0 and $P < .05$ to identify a total of 200 DEGs (57 upregulated and 143 downregulated genes; Fig. 1 and Table 1).

3.2. DEGs GO terms and KEGG pathways

These 200 DEGs were analyzed for the GO terms, that is, the upregulated DEGs were enriched in the BPs terms for different cell processes, like cell division or mitotic cytokinesis. In contrast, the downregulated DEGs were enriched in other cell pathways, like the epoxygenase P450 pathway, the oxidation-reduction process. In the MFs, the upregulated DEGs were enriched in protein kinase binding or protein kinase activity, whereas the downregulated DEGs were enriched in oxygen binding, heme binding, and steroid hydroxylase activity. For the CC GO term, the upregulated DEGs were enriched in the mid-body, cytoplasm, spindle microtubule, and centrosome; whereas the downregulated DEGs were enriched in the organelle membrane, extracellular region, endoplasmic reticulum membrane, and extracellular exosome (Table 2).

Furthermore, the KEGG analysis revealed that the upregulated DEGs were enriched in chemical carcinogenesis, retinol

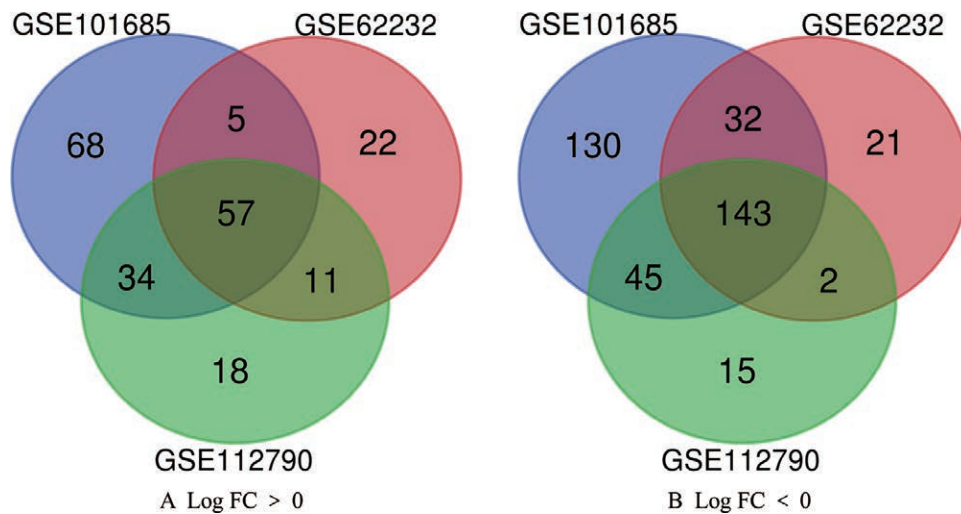


Figure 1. The Venn diagram of 200 DEGs. Illustration of the 200 DEGs from the 3 datasets (GSE101685, GSE62232, and GSE112790), which downloaded from PubMed database analyzed by using Venn diagram software (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). The various colors represent different datasets. (A) The 57 upregulated DEGs (log FC > 0). (B) The 143 downregulated DEGs (log FC < 0). DEGs = the differentially expressed genes, FC = fold change.

Table 1
Identification of DEGs in HCC tissue samples versus normal livers.

Expression	Genes
Upregulated DEGs (n = 57)	CDK1, TYMS, FAM72A//FAM72D//FAM72B//FAM72C, SPINK1, UBD//GABBR1, RCAA2, CAP2, DTL, RACGAP1, FAM83D, CTHRC1, UHRF1, RRM2, ZWINT, CCNB1, NDC80, TOP2A, KIAA0101, ASPM, HELLS, FLVCR1, HMMR, CCNA2, CD24, TTK, CDKN3, AKR1B10, PBK, NCAPG, GINS1, GPC3, CDKN2C, SULT1C2, CCL20, ROBO1, SPP1, CENPU, PRR11, LOC101930489//MIR4435-2HG//LINC00152, NEK2, ANLN, ACSL4, APOBEC3B, BIRC5, KIF20A, AURKA, UBE2T, DUXAP10, CRNDE, NUSAP1, NQO1, BUB1B, MAD2L1, COL15A1, DLGAP5, ECT2, PRC1
Downregulated DEGs (n = 143)	HBA2//HBA1, MT1G, CYP4A22//CYP4A11, LECT2, TUBE1, CYP26A1, BBOX1, PLG, CYP2A6, SOCS2, LINC01093, CYP2C8, CXCL14, SLC22A1, IGF1, SULT1E1, CYP39A1, SPP2, HAO2, LINC01554, FAM134B, MT1F, SLC25A47, MFSD2A, FLJ22763, HHIP, APOA5, ADH1B, KCND3, KCNN2, SLC01B3, SLC10A1, SLC1A2, GSTZ1, PRG4, LY6E, ASPA, CYP1A2, INS-IGF2//IGF2, MT1E, CNDP1, MAN1C1, BC02, FOLH1B//FOLH1, FCN3, ACSM3, GBA3, CYP2C19, PDGFRA, ANXA10, TTC36, LOC100287413//GLYATL1, CLEC4G, CYP2B6, GYS2, KBTBD11, FOLH1B, KMO, LPA, GHR, CLEC1B, MIR675//H19, CXCL2, LIFR, CLRN3, CYP2C9, CFHR3, MARCO, BHMT, CYP2A7, CYP2E1, EXPH5, MT1H, LCAT, CTH, CLEC4M, VNN1, LYVE1, ESR1, HSD11B1, RSP03, IGFBP3, FOS, LOC101928916//NNMT, PLAC8, ALDOB, HAMP, DNASE1L3, DCN, NAT2, BCHE, CPEB3, IL1RAP, RDH16, AKR1D1, CYP8B1, CXCL12, GNMT, TMEM27, HPGD, CRHBP, DNAJC12, MFAP3L, MME, AVPR1A, WDR72, THRSP, CYP4A11, IDO2, HGFAC, IGFALS, MT1X, MT2A, ADGRG7, S100A8, C7, CYP3A43, PZP, FBP1, AADAT, ADH4, GPM6A, OIT3, HGF, MOGAT2, MT1M, MA-GI2-AS3, CYP3A4, GLYAT, CYP2B7P//CYP2B6, CETP, GLS2, SRD5A2, ADRA1A, ECM1, APOF, HBB, MT1HL1, C9, SRPX, FCN2, OAT, LINC00844

The upregulated DEGs were identified by the criteria of log₂ fold change (FC) ≥ 2.0 and adjusted *P* value < .05, and the number of the DEGs was 57. The downregulated DEGs were identified by the criteria of log₂ fold change (FC) ≤ -2.0 and adjusted *P* value < .05, and the number of the DEGs was 143. DEGs = the differentially expressed genes

metabolism, drug metabolism - cytochrome P450, metabolism of xenobiotics by cytochrome P450, and the cell cycle; however, the downregulated DEGs did not significantly associated with (formed) any signaling (*P* < .05; Table 3).

3.3. Identification of DEGs-related PPI network

Since PPIs are key for gene-gene interactions in cells, we obtained the PPI network of these DEGs using the online search tool for the retrieval of interacting genes and Cytoscape software. This PPI network contained a total of 158 nodes and 885 edges, covering 50 upregulated and 108 downregulated genes (Fig. 2A); however, we could not relate 42 of these 200 DEGs into this PPI network complex because they did not associate with this overall PPI network. Next, we analyzed the core modules of this PPI network using the MOCDE app and found them to include 34 nodes and 537 edges (Fig. 2B), showing that all 34 core nodes were from upregulated genes of the 158 core DEGs, but nor from the downregulated DEGs, with unknown reason.

3.4. Association of these core DEGs with survival of HCC patients

We then performed Kaplan–Meier analysis (<http://kmplot.com/analysis/>) to associate these 34 main DEGs with HCC prognosis. The database is capable to assess the effect of 54k genes (i.e., mRNA, miRNA, or protein) on survival in 21 cancer types, including breast (n = 7830), ovarian (n = 2190), lung (n = 3452), and gastric (n = 1440) cancers. As shown in Figure 3 and Table 4, we found that 28 DEGs were significantly associated with a poor HCC prognosis, whereas the remaining 6 DEGs were not associated with the HCC prognosis (*P* > .05).

Next, we performed GEPIA of these 28 genes between cancer patients and healthy individuals (<http://gepia.cancer-pku.cn/>). We found that 27 of these 28 genes were highly expressed in HCC tissue samples versus normal livers (*P* < .05; Fig. 4 and Table 5). These 27 genes, particularly Cyclin B1, CDK1, MAD2L1, or BUB1B, were enriched in the cell cycle signaling (*P* = 3.09E-4; Fig. 5 and Table 6).

Table 2

GO terms of the DEGs in HCC tissue samples.

Expression	Category	Term	Count	%	P	FDR
Upregulated DEGs	GOTERM_BP_DIRECT	GO:0051301–cell division	8	9.33	6.73E-07	9.15E-04
	GOTERM_BP_DIRECT	GO:0000281–mitotic cytokinesis	4	4.66	7.09E-05	0.096
	GOTERM_BP_DIRECT	GO:0000910–cytokinesis	4	4.66	1.40E-04	0.190
	GOTERM_BP_DIRECT	GO:0051988–regulation of attachment of spindle microtubules to kinetochore	3	3.50	1.68E-04	0.229
	GOTERM_BP_DIRECT	GO:0007067–mitotic nuclear division	5	5.83	4.36E-04	0.591
	GOTERM_BP_DIRECT	GO:0007052–mitotic spindle organization	3	3.50	.003	3.656
	GOTERM_MF_DIRECT	GO:0019901–protein kinase binding	5	5.83	9.99E-05	0.105
	GOTERM_MF_DIRECT	GO:0004672–protein kinase activity	5	5.83	8.04E-04	0.843
	GOTERM_MF_DIRECT	GO:0004672–protein kinase activity	4	4.66	.001	1.367
	GOTERM_MF_DIRECT	GO:0008017–microtubule binding	4	4.66	.002	1.367
	GOTERM_MF_DIRECT	GO:0003682–chromatin binding	6	6.99	.002	2.430
	GOTERM_CC_DIRECT	GO:0003682–chromatin binding	8	9.33	1.79E-08	1.95E-05
	GOTERM_CC_DIRECT	GO:0030496–midbody	25	29.1	1.32E-06	0.001
	GOTERM_CC_DIRECT	GO:0005737–cytoplasm	4	4.67	1.51E-04	0.164
	GOTERM_CC_DIRECT	GO:0005876–spindle microtubule	19	22.1	9.04E-04	0.981
GOTERM_CC_DIRECT	GO:0005634–nucleus	6	6.99	.003	2.747	
GOTERM_CC_DIRECT	GO:0005813–centrosome	6	6.99	.003	2.747	
Downregulated DEGs	GOTERM_BP_DIRECT	GO:0019373–epoxygenase P450 pathway	9	5.72	2.53E-13	3.93E-10
	GOTERM_BP_DIRECT	GO:0055114–oxidation-reduction process	24	15.3	3.94E-11	6.11E-08
	GOTERM_BP_DIRECT	GO:0042738–exogenous drug catabolic process	7	4.45	1.16E-10	1.80E-07
	GOTERM_BP_DIRECT	GO:0017144–drug metabolic process	8	5.08	7.07E-10	1.10E-06
	GOTERM_BP_DIRECT	GO:0045926–negative regulation of growth	7	4.45	3.26E-09	5.06E-06
	GOTERM_BP_DIRECT	GO:0071294–cellular response to zinc ion	7	4.45	3.26E-09	5.06E-06
	GOTERM_BP_DIRECT	GO:0006805–xenobiotic metabolic process	10	6.35	5.01E-09	7.77E-06
	GOTERM_BP_DIRECT	GO:0008202–steroid metabolic process	8	5.08	2.33E-08	3.61E-05
	GOTERM_MF_DIRECT	GO:0019825–oxygen binding	12	7.63	2.06E-14	2.85E-11
	GOTERM_MF_DIRECT	GO:0020037–heme binding	16	10.2	4.59E-14	6.33E-11
	GOTERM_MF_DIRECT	GO:0005506–iron ion binding	16	10.2	2.39E-13	3.30E-10
	GOTERM_MF_DIRECT	GO:0004497–monooxygenase activity	12	7.63	2.52E-13	3.49E-10
	GOTERM_MF_DIRECT	GO:0008392–arachidonic acid epoxygenase activity	8	5.08	4.72E-12	6.52E-09
	GOTERM_MF_DIRECT	GO:0008392–arachidonic acid epoxygenase activity	8	5.08	4.52E-10	6.25E-07
	GOTERM_CC_DIRECT	GO:0008395–steroid hydroxylase activity	15	9.53	8.33E-16	1.02E-12
	GOTERM_CC_DIRECT	GO:0031090–organelle membrane	34	21.6	6.75E-09	7.78E-06
	GOTERM_CC_DIRECT	GO:0005576–extracellular region	9	5.72	1.00E-05	0.012
	GOTERM_CC_DIRECT	GO:0072562–blood microparticle	9	5.72	1.00E-05	0.012
	GOTERM_CC_DIRECT	GO:0005789–endoplasmic reticulum membrane	19	12.1	2.38E-05	0.027
	GOTERM_CC_DIRECT	GO:0070062–extracellular exosome	36	22.9	2.29E-04	0.263

The GO term analysis was applied to define genes and their products, mRNA, or proteins in order to evaluate the unique biological signaling of the high-throughput transcriptome or genomic data. The table was utilized DAVID to identify the enriched biological processes (BPs), molecular functions (MFs), and cellular components (CCs) as well as pathways with $P < .05$ as the cutoff criterion for these DEGs in HCC. DAVID = the database for annotation, visualization and integrated discovery, FDR = false discovery rate, GO = gene ontology.

Table 3

KEGG pathways of the DEGs in HCC tissue samples.

Pathway ID	Name	Count	%	P	Genes
hsa05204	Chemical carcinogenesis	12	4.93	2.43E-08	CYP3A4, CYP3A43, CYP2C19, CYP2C9, CYP2C8, ADH4, NAT2, HSD11B1, ADH1B, CYP2A6, CYP2E1, CYP1A2
hsa00830	Retinol metabolism	10	4.11	4.02E-07	CYP3A4, CYP2B6, CYP2C9, CYP2C8, ADH4, ADH1B, CYP26A1, CYP2A6, CYP1A2, RDH16
hsa00982	Drug metabolism - cytochrome P450	10	4.11	6.84E-07	CYP3A4, CYP2C19, CYP2B6, CYP2C9, CYP2C8, ADH4, ADH1B, CYP2A6, CYP2E1, CYP1A2
hsa00980	Metabolism of xenobiotics by cytochrome P450	9	3.37	1.38E-05	CYP3A4, CYP2B6, CYP2C9, ADH4, HSD11B1, ADH1B, CYP2A6, CYP2E1, CYP1A2
hsa04110	Cell cycle	7	2.88	0.01	CCNB1, CDK1, MAD2L1, CDKN2C, TTK, BUB1B, CCNA2

The KEGG database was used to integrate the currently known PPI network information for the metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development using DAVID 6.8 (<https://david.ncicrf.gov>), an online bioinformatics tool designed to interpret biological function annotations of genes or proteins.

DAVID = the database for annotation, visualization and integrated discovery, HCC = hepatocellular carcinoma, KEGG = Kyoto Encyclopedia of Genes and Genomes.

4. Discussion

HCC development and progression, like most other malignancies, involve the aberrant activation of oncogenes and the loss

of tumor suppressor genes, resulting in abnormal BPs, MFs, and CCs as well as alterations of cell growth and apoptosis.^[20] In the current study, altered CCNB1, CDK1, MAD2L1, and BUB1B expression was shown in HCC versus normal liver tissues, and

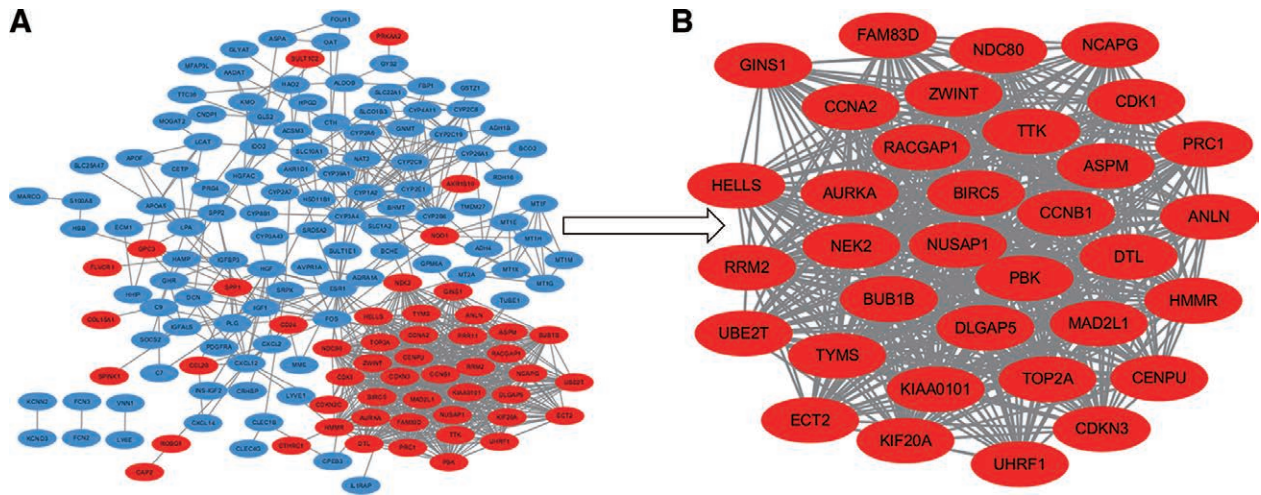


Figure 2. The PPI network. (A) The PPI network covers a total of 158 DEGs. Nodes indicate proteins; edges indicate the interaction of proteins. The blue nodes refer to downregulated DEGs; the red nodes refer to upregulated DEGs. (B) Module analysis of the DEGs using the Cytoscape software with a cutoff value of 2, the node score cutoff value = 0.2 (k -core = 2), and maximum depth = 100. DEGs = the differentially expressed genes, PPI = protein–protein interaction.

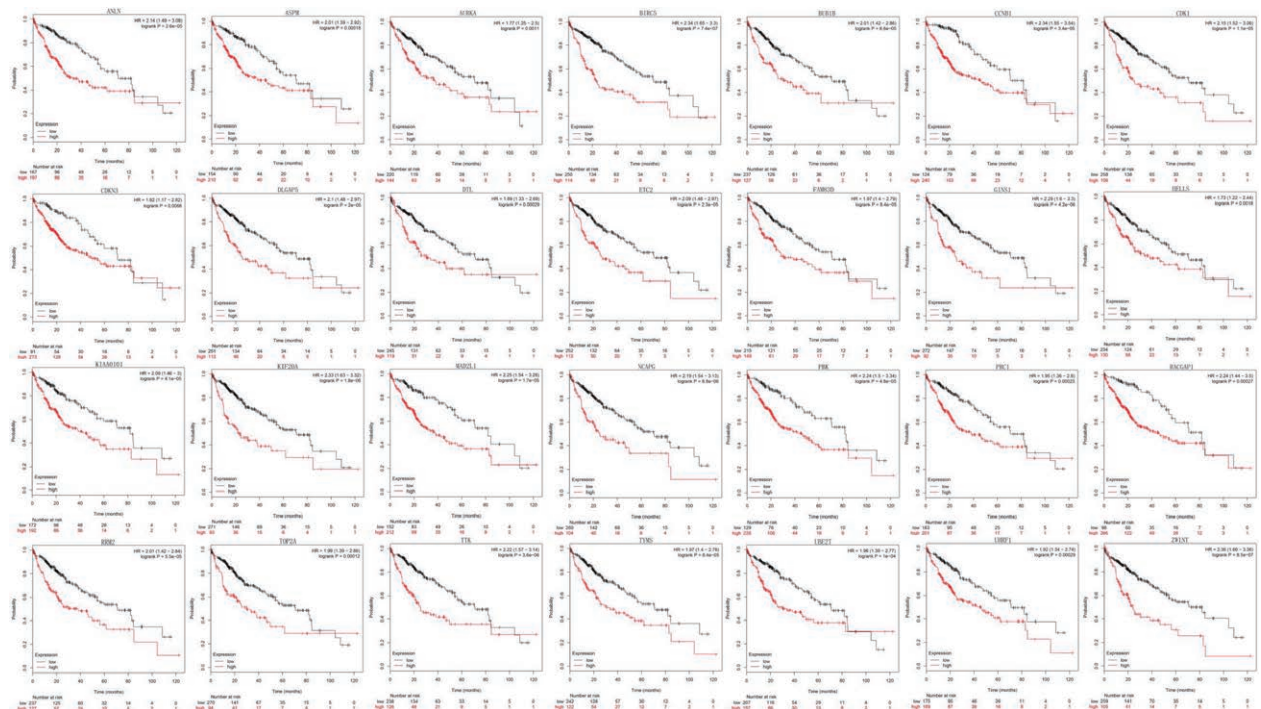


Figure 3. The Kaplan–Meier survival curves of the 28 DEGs. Association of the 34 core genes with HCC prognosis as analyzed by using the online tool Kaplan–Meier plotter. The data demonstrated that 28 of 34 genes were associated with significantly poor HCC prognosis ($P < .05$). DEGs = the differentially expressed genes, HCC = hepatocellular carcinoma.

Table 4
Survival significance of the 34 key candidate genes in HCC.

Category	Genes
Genes with a significantly worse survival ($P < .05$)	ANLN, ASPM, AURKA, BIRC5, BUB1B, CCNB1, CDK1, CDKN3, DLGAP5, DTL, ECT2, FAM83D, GINS1, HELLIS, KIAA0101, KIF20A, MAD2L1, NCAPG, PBK, PRC1, RACGAP1, RRM2, TOP2A, TTK, TYMS, UBE2T, UHRF1, ZWINT
Genes without a significantly worse survival ($P > .05$)	CCNA2, NDC80, NEK2, NUSAP1, HMMR, CENPU

HCC = hepatocellular carcinoma.

changes in expression of these genes resulted in tumor cell cycle progression and cell growth. Therefore, these genes were associated with a poor overall survival of HCC patients. A previous

study conducted by Sarathi et al reported alteration of several genes in HCC, which associated with the stage of HCC, such as NDUFA4L2, CRHBP, and PIGU. However, their data were not

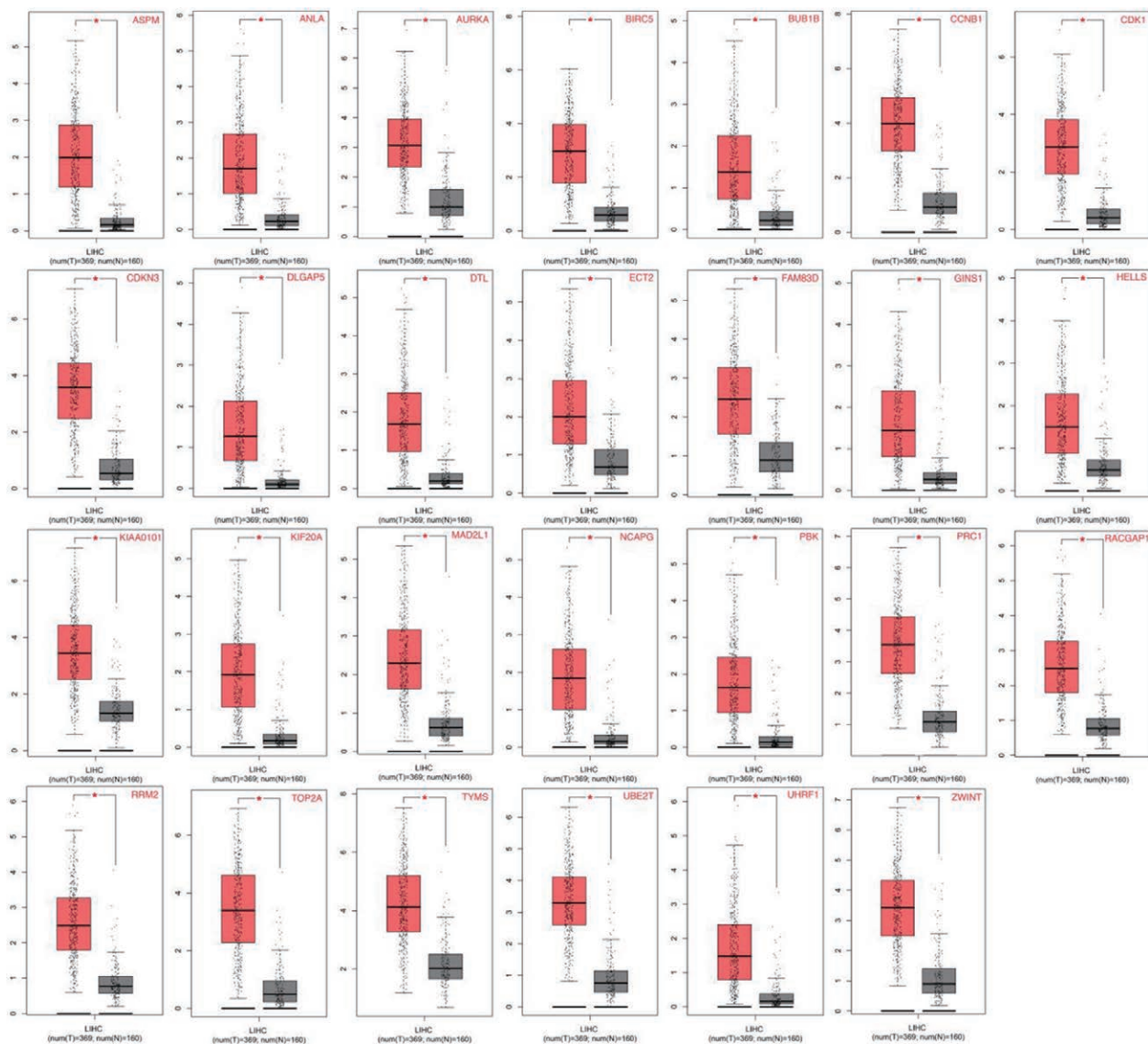


Figure 4. The expression level of the 27 DEGs between the liver tissue and the normal tissue. Identification of 27 genes that are significantly highly expressed in HCC tissues. The 28 genes that are significantly associated with a poor survival of HCC patients were analyzed by using GEPIA of HCC tissues versus normal livers. The data showed that 27 of these 28 genes were highly expressed in HCC versus normal liver samples ($P < .05$). The red color refers to HCC; the gray color refers to normal liver samples. DEGs = the differentially expressed genes, GEPIA = gene expression profiling interactive analysis, HCC = hepatocellular carcinoma.

Table 5

Validation of the 28 genes from the GEPIA data.

Category	Genes
Highly expressed genes ($P < .05$)	ANLN, ASPM, AURKA, BIRC5, BUB1B, CCNB1, CDK1, CDKN3, DLGAP5, DTL, ECT2, FAM83D, GINS1, HELLS, KIAA0101, KIF20A, MAD2L1, NCAPG, PBK, PRC1, RACGAP1, RRM2, TOP2A, TYMS, UBE2T, UHRF1, ZWINT
Nonhighly expressed genes ($P > .05$)	TTK

GEPIA = gene expression profiling interactive analysis.

shown in our current study, which might be due to the different RNA-Seq data. The heterogeneity of HCC might be the another reason and blocked the treatment in clinical.^[21]

The cell cycle is a cell process that is strictly controlled by different cyclins and cyclin-dependent kinases (CDKs).^[22] Cyclin B1 (CCNB1) is a protein associated with cell mitosis and primarily expressed in the G2/M phase of the cell cycle.

Zheng *et al* have shown that miRNA-200c expression was able to enhance the radiosensitivity of esophageal cancer by arrest of cells at cell cycle G2/M via downregulation of CCNB1 expression.^[23] Similarly, Pan *et al* have reported that eukaryotic translation initiation factor 3 subunit D silencing was able to suppress renal cancer cell tumorigenesis by induction of G2/M-phase arrest and downregulation of the CCNB1/CDK1

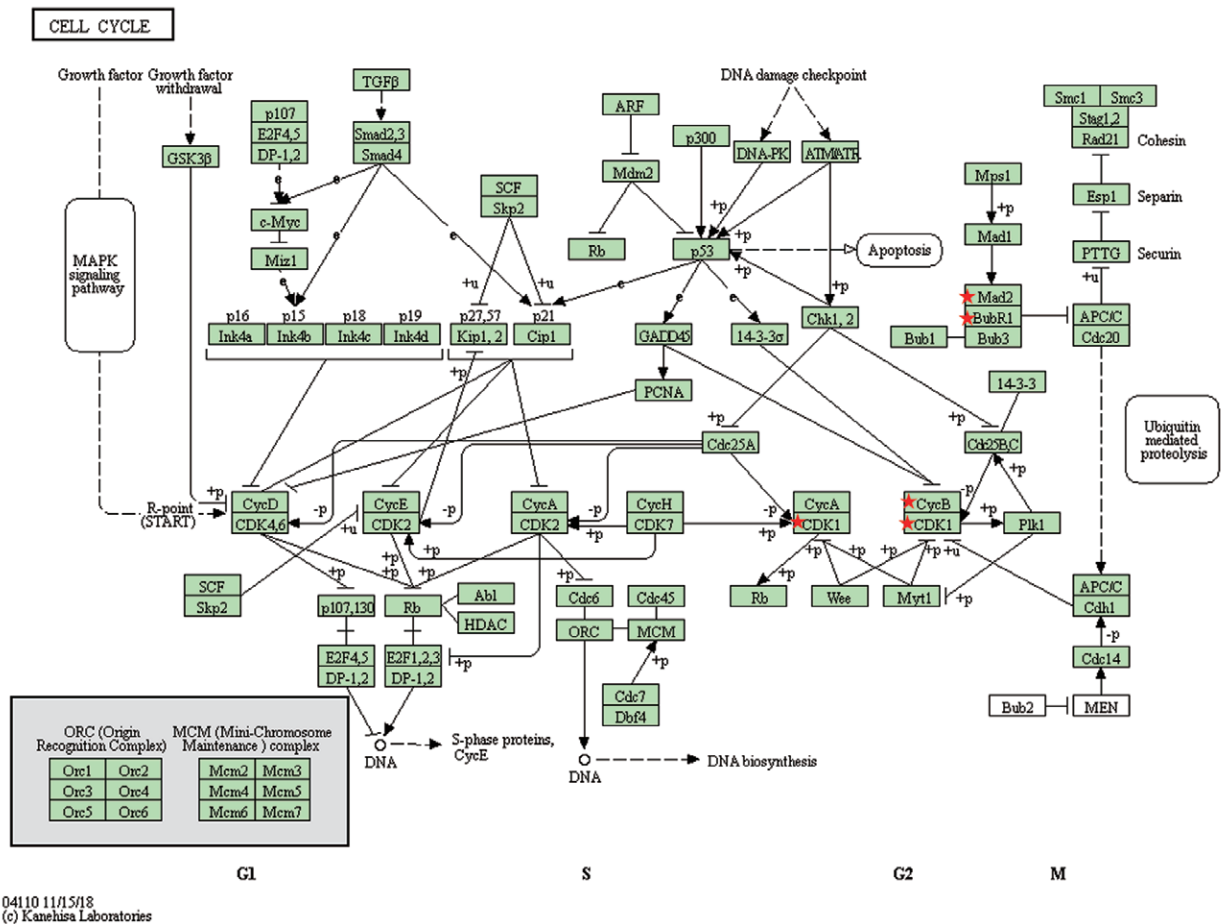


Figure 5. The KEGG pathway enrichment analysis of the 27 selected genes. The data showed that level of Cyclin B1, CDK1, MAD2L1, and BUB1B was markedly enhanced in the cell cycle pathway. The red arrows point out the position of these genes. CycB, Cyclin B1; Mad2, MAD2L1; BubR1, BUB1B. BUB1B = mitotic checkpoint serine/threonine kinase B, CDK1 = Cyclin-dependent kinase 1, KEGG = Kyoto Encyclopedia of Genes and Genomes, MAD2L1 = Mitotic arrest deficient 2 like 1.

Table 6
Re-analysis of 27 selected genes via KEGG pathway enrichment.

Pathway ID	Name	Count	%	P	Genes
cfa04110	Cell cycle	4	8.36	3.09E-04	CCNB1, CDK1, MAD2L1, BUB1B
cfa04115	p53 signaling pathway	3	6.27	.002	CCNB1, CDK1, RRM2

KEGG = Kyoto Encyclopedia of Genes and Genomes.

signaling pathway activity.^[24] Other studies have demonstrated that CCNB1 overexpression is associated with a poor prognosis of human epidermal growth factor receptor 2-positive breast cancer,^[25] colorectal cancer,^[26] nonmuscle-invasive bladder cancer,^[27] pancreatic cancer,^[28] and non-small cell lung cancer.^[29] Our current data further support these findings in other cancers. Indeed, Zhuang *et al* also have revealed that the upregulated CCNB1 expression in HCC tissues associated with a worse HCC prognosis.^[30] Furthermore, Gavet *et al* have previously demonstrated that the cyclinB1-Cdk1 complex, which functions as the master mitotic factor, was significantly activated in the late G2 phase to trigger the prophase of the cell cycle.^[31] CDK1, also known as cell division control protein 2 or CDC2, is one of the CDK family members that is frequently mutated in various tumor tissues; this mutation results in uncontrolled cell proliferation or unscheduled reentry into the

cell cycle.^[32] Zhong *et al* have revealed that the BTB (BR-C, ttk and bab) or POZ (Pox virus and Zinc finger) domain-containing protein potassium channel tetramerisation domain containing 12 was able to activate CDK1 and Aurora A to induce the G2/M transition for cell growth and tumorigenesis.^[33] In addition, a possible reason for everolimus resistance in prostate cancer cells might be due to upregulation of the CDK1-cyclin B complex, which leads to tumor cell progression towards cell cycle G2/M phase.^[34] Meanwhile, knockdown of eukaryotic translation initiation factor 3 subunit D expression notably suppressed renal cell carcinoma cell tumorigenesis by inhibition of the cyclin B1/CDK1 signaling.^[24] Thus, CDK1 not only participates in cell-cycle progression but also regulates cell apoptosis by phosphorylation of caspases,^[35] indicating that CDK1 is a promoter of tumorigenesis in different tissues and cells and could be a target for future cancer therapy.

Mitotic arrest deficient 2 like 1 (MAD2L1) is a component of the mitotic spindle assembly checkpoint that participates in cell mitosis.^[36] Foijer *et al* have demonstrated that deletion of MADL1 in mouse hepatocytes results in HCC development and progression because of chromosome instability,^[37] although another study conducted by Scintu *et al* has shown that the MAD2L1 level is not associated with intra-chromosomal instability in human ductal breast carcinoma.^[38] This discrepancy might be due to the different types of species and neoplasms. However, several studies have shown an association of high MAD1L2 expression with cancer risk and a poor prognosis.^[39–42] Additionally, Li *et al* have reported that miR-200c-5p expression leads to the reduction of HCC cell proliferation and metastasis by suppression of MAD1L2 expression.^[43] Our current data further support the oncogenic role of MAD1L2 in HCC. Furthermore, the mitotic checkpoint serine/threonine kinase B (BUB1B) is a mitotic checkpoint serine/threonine protein kinase that binds to centromeres during cell mitosis.^[44] BUB1B overexpression has been associated with the development of several human cancers^[45–48] and with a poor survival of patients. Fu *et al* also have reported that a high BUB1B expression enhanced the development of prostate cancer and was associated with a poor prognosis of patients.^[46] Moreover, mutations of the BUB1B gene are associated with mosaic variegated aneuploidy syndrome and contribute to a high risk of developing childhood cancers.^[49,50]

However, the current study is just a proof-of-principle study and lacks experimental verification, although previous studies have been conducted to associate some of these molecules with HCC tumorigenesis. Our finding of the association of 4 genes (CCNB1, CDK1, MAD2L1, and BUB1B) with a poor HCC prognosis will need to be confirmed in the future. In conclusion, the current study identified a number of DEGs (especially Cyclin B1, CDK1, MAD2L1, or BUB1B) to be associated with a poor HCC prognosis. Future studies will investigate their functions in HCC and apply them as prognostic biomarkers and potential therapeutic targets for HCC.

5. Conclusions

We assessed DEGs in 288 HCC versus 32 normal liver tissues from 3 online datasets (GSE101685, GSE62232, and GSE112790) and revealed a total of 200 DEGs (57 upregulated and 143 downregulated genes). Our GO term and pathway analysis identified that these DEGs were involved in BPs, MFs, and CCs. The PPI network analysis covered 50 upregulated and 108 downregulated genes. The core modules of this PPI network contained 34 upregulated genes, 28 of which were associated with a poor HCC prognosis, and 27 of these 28 genes were highly expressed in HCC tissues. In conclusion, our current study revealed 200 DEGs, 28 of which were associated with a poor HCC prognosis. However, further studies are needed to confirm their involvement in HCC development and to verify them as prognostic markers and potential treatment targets for HCC.

Acknowledgments

We would like to thank the individuals who helped us with this work, especially Ms. Luodan Wang in inspiring the first author for his research over the past decade.

Author contributions

Conceptualization: Xu Huang, Ping Zhang.

Data curation: Xu Huang, Xu Wang, Ge Huang.

Formal analysis: Xu Huang, Xu Wang, Ruotao Li.

Investigation: Xu Huang, Xingkai Liu; Lidong Cao.

Methodology: Xu Huang, Xu Wang, Ge Huang, Ruotao Li, Ping Zhang.

Supervision: Junfeng Ye; Ping Zhang; Validation: Junfeng Ye; Ping Zhang.

Visualization: Xu Huang.

Writing – original draft: Xu Huang, Ruotao Li, Lidong Cao.

Writing – review & editing: Xu Wang, Ge Huang, Ping Zhang.

References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69:7–34.
- [2] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
- [3] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66:115–32.
- [4] Huo YR, Eslick GD. Transcatheter arterial chemoembolization plus radiotherapy compared with chemoembolization alone for hepatocellular carcinoma: a systematic review and meta-analysis. *JAMA Oncol.* 2015;1:756–65.
- [5] Valle JW, Dangoor A, Beech J, et al. Treatment of inoperable hepatocellular carcinoma with pegylated liposomal doxorubicin (PLD): results of a phase II study. *Br J Cancer.* 2005;92:628–30.
- [6] Nair M, Sandhu SS, Sharma AK. Prognostic and predictive biomarkers in cancer. *Curr Cancer Drug Targets.* 2014;14:477–504.
- [7] Ally A, Balasundaram M, Carlsen R, et al. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell.* 2017;169:1327–1341.e1323.
- [8] Mou T, Zhu D, Wei X, et al. Identification and interaction analysis of key genes and microRNAs in hepatocellular carcinoma by bioinformatics analysis. *World J Surg Oncol.* 2017;15:63.
- [9] Li L, Lei Q, Zhang S, et al. Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis. *Oncol Rep.* 2017;38:2607–18.
- [10] Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23:1846–7.
- [11] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25:25–9.
- [12] Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45:D353–61.
- [13] Dennis G, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003;4:P3.
- [14] Yang J, Hou Z, Wang C, et al. Gene expression profiles reveal key genes for early diagnosis and treatment of adamantinomatous craniopharyngioma. *Cancer Gene Ther.* 2018;25:227–39.
- [15] Feng H, Gu ZY, Li Q, et al. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *J Ovarian Res.* 2019;12:35.
- [16] Xu Z, Zhou Y, Cao Y, et al. Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis. *Med Oncol.* 2016;33:130.
- [17] Pan JH, Zhou H, Cooper L, et al. LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers. *Front Immunol.* 2019;10:6.
- [18] Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017;45:W98–W102.
- [19] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
- [20] Shibata T, Aburatani H. Exploration of liver cancer genomes. *Nat Rev Gastroenterol Hepatol.* 2014;11:340–9.
- [21] Sarathi A, Palaniappan A. Novel significant stage-specific differentially expressed genes in hepatocellular carcinoma. *BMC Cancer.* 2019;19:663.
- [22] Hartwell LH, Weinert TA. Checkpoints: controls that ensure the order of cell cycle events. *Science.* 1989;246:629–34.
- [23] Zheng R, Liu Y, Zhang X, et al. miRNA-200c enhances radiosensitivity of esophageal cancer by cell cycle arrest and targeting P21. *Biomed Pharmacother.* 2017;90:517–23.

- [24] Pan XW, Chen L, Hong Y, et al. EIF3D silencing suppresses renal cell carcinoma tumorigenesis via inducing G2/M arrest through downregulation of Cyclin B1/CDK1 signaling. *Int J Oncol.* 2016;48:2580–90.
- [25] Sabbaghi M, Gil-Gómez G, Guardia C, et al. Defective Cyclin B1 induction in Trastuzumab-emptansine (T-DM1) acquired resistance in HER2-positive Breast cancer. *Clin Cancer Res.* 2017;23:7006–19.
- [26] Fang Y, Yu H, Liang X, et al. Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. *Cancer Biol Ther.* 2014;15:1268–79.
- [27] Kim SK, Roh YG, Park K, et al. Expression signature defined by FOXM1-CCNB1 activation predicts disease recurrence in non-muscle-invasive bladder cancer. *Clin Cancer Res.* 2014;20:3233–43.
- [28] Zhou L, Li J, Zhao YP, et al. The prognostic value of Cyclin B1 in pancreatic cancer. *Med Oncol.* 2014;31:107.
- [29] Zhang LL, Feng ZL, Su MX, et al. Downregulation of Cyclin B1 mediates nagilactone E-induced G2 phase cell cycle arrest in non-small cell lung cancer cells. *Eur J Pharmacol.* 2018;830:17–25.
- [30] Zhuang L, Yang Z, Meng Z. Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients. *Biomed Res Int.* 2018;2018:7897346.
- [31] Gheghiani L, Loew D, Lombard B, et al. PLK1 Activation in late G2 sets up commitment to mitosis. *Cell Rep.* 2017;19:2060–73.
- [32] Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer.* 2009;9:153–66.
- [33] Zhong Y, Yang J, Xu WW, et al. KCTD12 promotes tumorigenesis by facilitating CDC25B/CDK1/Aurora A-dependent G2/M transition. *Oncogene.* 2017;36:6177–89.
- [34] Tsaor I, Makarević J, Hudak L, et al. The cdk1-cyclin B complex is involved in everolimus triggered resistance in the PC3 prostate cancer cell line. *Cancer Lett.* 2011;313:84–90.
- [35] Zhang P, Kawakami H, Liu W, et al. Targeting CDK1 and MEK/ERK overcomes apoptotic resistance in BRAF-mutant human colorectal cancer. *Mol Cancer Res.* 2018;16:378–89.
- [36] Li Y, Benezra R. Identification of a human mitotic checkpoint gene: hsMAD2. *Science.* 1996;274:246–8.
- [37] Fojier F, Albacker LA, Bakker B, et al. Deletion of the spindle assembly checkpoint gene is tolerated in mouse models of acute T-cell lymphoma and hepatocellular carcinoma. *Elife.* 2017;6:e20873.
- [38] Scintu M, Vitale R, Prencipe M, et al. Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma. *Cancer Lett.* 2007;254:298–307.
- [39] Guo Y, Zhang X, Yang M, et al. Functional evaluation of missense variations in the human MAD1L1 and MAD2L1 genes and their impact on susceptibility to lung cancer. *J Med Genet.* 2010;47:616–22.
- [40] Ko YH, Roh JH, Son YI, et al. Expression of mitotic checkpoint proteins BUB1B and MAD2L1 in salivary duct carcinomas. *J Oral Pathol Med.* 2010;39:349–55.
- [41] Wang Z, Katsaros D, Shen Y, et al. Biological and clinical significance of MAD2L1 and BUB1, genes frequently appearing in expression signatures for breast cancer prognosis. *PLoS One.* 2015;10:e0136246.
- [42] Zhong R, Chen X, Chen X, et al. MAD1L1 Arg558His and MAD2L1 Leu84Met interaction with smoking increase the risk of colorectal cancer. *Sci Rep.* 2015;5:12202.
- [43] Li Y, Bai W, Zhang J. MiR-200c-5p suppresses proliferation and metastasis of human hepatocellular carcinoma (HCC) via suppressing MAD2L1. *Biomed Pharmacother.* 2017;92:1038–44.
- [44] Davenport JW, Fernandes ER, Harris LD, et al. The mouse mitotic checkpoint gene *bub1b*, a novel *bub1* family member, is expressed in a cell cycle-dependent manner. *Genomics.* 1999;55:113–7.
- [45] Ding Y, Hubert CG, Herman J, et al. Cancer-specific requirement for BUB1B/BUBR1 in human brain tumor isolates and genetically transformed cells. *Cancer Discov.* 2013;3:198–211.
- [46] Fu X, Chen G, Cai ZD, et al. Overexpression of BUB1B contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer. *Onco Targets Ther.* 2016;9:2211–20.
- [47] Hanks S, Coleman K, Reid S, et al. Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. *Nat Genet.* 2004;36:1159–61.
- [48] Yang Y, Gu C, Luo C, et al. BUB1B promotes multiple myeloma cell proliferation through CDC20/CCNB axis. *Med Oncol.* 2015;32:81.
- [49] Ochiai H, Miyamoto T, Kanai A, et al. TALEN-mediated single-base-pair editing identification of an intergenic mutation upstream of BUB1B as causative of PCS (MVA) syndrome. *Proc Natl Acad Sci USA.* 2014;111:1461–6.
- [50] Rio Frio T, Lavoie J, Hamel N, et al. Homozygous BUB1B mutation and susceptibility to gastrointestinal neoplasia. *N Engl J Med.* 2010;363:2628–37.