Identifying Potential Prognostic Biomarkers Associated With Clinicopathologic Characteristics of Hepatocellular Carcinoma by Bioinformatics Analysis

Shaojie Huang
First Affiliated Hospital of Sun Yat-sen University
Jia Yao
Third Affiliated Hospital of Sun Yat-sen University
Xiaofan Lai
Massachusetts General Hospital
Lu Yang
First Affiliated Hospital of Sun Yat-sen University
Fang Ye
First Affiliated Hospital of Sun Yat-sen University
Wenqi Huang
First Affiliated Hospital of Sun Yat-sen University
Yubin Xie
First Affiliated Hospital of Sun Yat-sen University
Zhongxing Wang (wzhxing@mail.sysu.edu.cn)
First Affiliated Hospital of Sun Yat-sen University

Research Article

Keywords: hepatocellular carcinoma, bioinformatics analysis, differentially expressed genes, biomarkers, protein□protein interaction network

Posted Date: January 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-146400/v1

Identifying potential prognostic biomarkers associated with clinicopathologic characteristics of hepatocellular carcinoma by bioinformatics analysis

Shaojie Huang[1†], Jia Yao[2†], Xiaofan Lai[3,4†], Lu Yang[3], Fang Ye[3], Wenqi Huang[3], Yubin Xie[5*] and Zhongxing Wang[3*]

[1]Department of Cardiac Surgery, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China.

[2]Department of Hepatic Surgery and Liver Transplantation Center, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[3]Department of Anesthesiology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.

[4]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA.

[5]Precision Medicine Institute, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.

Correspondence:

wzhxing@mail.sysu.edu.cn; xieyb6@mail.sysu.edu.cn

[†]Shaojie Huang, Jia Yao and Xiaofan Lai contributed equally.

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide. However, the molecular mechanisms of HCC remain largely unknown so far.

Methods: To unravel the underlying carcinogenic mechanisms, we utilized Robust Rank Aggregation analysis (RRA) to identify a set of overlapping differentially expressed genes (DEGs) from 5 microarray datasets on Gene Expression Omnibus (GEO) database. Enriched Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of DEGs were conducted. The protein-protein interaction (PPI) network was constructed and Cytoscape V3.8.0 was used for selecting hub genes. The expression of hub genes was validated in TCGA datasets and HCC samples in our center by qPCR and immunohistochemistry analysis.

Results: Totally 126 DEGs were identified. GO and KEGG pathways of DEGs mostly associated with “organelle fission”, “nuclear division” and “caffeine metabolism. Ten hub genes (BUB1B, CDKN3, CCNB1, CCNB2, CDK1, TOP2A, CDC20, MELK, NUSAP1, AURKA) were selected. Overall survival (OS) and progression-free survival (PFS) analysis suggested the good value of these genes for HCC diagnosis and prognosis. These genes were upregulated in HCC samples from TCGA, which were associated with higher tumor grades and possibly resulted from hypomethylation. Moreover, these hub genes were markedly dysregulated in HCC samples in our center and significantly associated with clinicopathologic characteristics of HCC patients.

Conclusions: In conclusion, our study identified several hub genes as novel candidate biomarkers for diagnosis and prognosis of HCC, which may provide new insight into HCC pathogenesis in order to search for better treatments.

Keywords: hepatocellular carcinoma, bioinformatics analysis, differentially expressed genes, biomarkers, protein-protein interaction network

Background

Hepatocellular carcinoma (HCC) is one of the most prevalent cancers in the clinical setting, which represents the second highest cause of cancer-related deaths worldwide [1,2]. Nowadays HCC was still associated with enormous mortality and morbidity that cause large burden to the society [3,4]. It is mainly due to the ineffective diagnostic methods at the early stage of HCC and the unclear pathologically molecular mechanisms. Therefore, it is of great importance to understand the precise molecular mechanisms underlying the carcinogenesis of HCC and develop novel potential diagnostic methods and therapies.

During the last decades, with the rapid development of second-generation sequencing technology, microarray and bioinformatic analysis have been widely applied to screen genetic dysregulations and explore the precise pathogenesis of various diseases, which have helped us identify the differentially expressed genes (DEGs) and functional pathways involved in the initiation and development of HCC [5,6]. Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) is a useful website that can be used to obtain a great number of gene profiles related with different diseases and screen out DEGs or other small molecules by integrated bioinformatics analysis [7].

In the present study, raw data from five microarray datasets (mRNA expression profile datasets) conducted on HCC and normal samples were downloaded from the GEO database. Robust differentially expressed genes (DEGs) were screened by bioinformatics analysis using R language, a set of overlapping DEGs were identified

via RRA analysis, and functional enrichment analysis were performed. The protein-protein interaction (PPI) network was constructed with the Search Tool for the Retrieval of Interacting Genes (STRING), and top 10 hub genes were identified and presented by Cytoscape V3.8.0. Overall survival (OS) and progression-free survival (PFS) analysis of the hub genes were performed by Kaplan-Meier Plotter. Besides, the gene expression differences between HCC and normal tissues, the gene expression differences in HCC samples with different tumor stages and promoter methylation levels analysis were conducted by UALCAN. Moreover, the correlation between the expression of identified 10 hub genes and tumor infiltrating immune cells as well as gene alteration were analyzed. The expression of hub genes and their association with clinicopathologic characteristics of HCC patients were investigated. In conclusion, a total of 126 DEGs and 10 hub genes were identified, which may be novel candidate biomarkers for HCC.

Materials and methods

Data collection

Gene Expression Omnibus (GEO) [8] is a widely used gene expression database for retrieving gene expression profiles from any species submitted by research institutes worldwide. Five microarray datasets GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248 were all downloaded from the GEO database. GSE41804 consisted of 20 HCC samples and 20 noncancerous samples. GSE45267 contained 46 HCC tissues and 41 normal liver tissues. GSE84402 included 14 HCC

samples and 14 noncancerous samples. GSE107170 consisted of 39 HCC samples and 80 noncancerous samples. GSE121248 consisted of 70 HCC samples and 37 adjacent normal samples.

Tissue samples

Clinical HCC specimens were histopathologically and clinically diagnosed at the third affiliated hospital, Sun Yat-sen University (Guangzhou, Guangdong, China). Each HCC tissue has paired nontumorous tissue. All patients signed informed consent for their samples to be used for research, and approval was obtained from the Committees for Ethical Review of Research.

Real-time quantitative PCR (qPCR)

Total RNA was extracted from tissue specimens using the TRIzol reagent (Molecular Research Center, Inc.) according to the manufacturer's protocol. Quantification of RNA was conducted with a NanoDrop 8000 spectrophotometer and 1 μg RNA was subjected to reverse transcription by RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, K1622). The obtained cDNAs were used to be the template for qPCR reactions with the FastStart Essential DNA Green Master Mix (Roche, 06924204001). All samples were run in triplicate and GAPDH was used as the internal control. All primers used for qPCR are described in (**Additional file 9: Table S4**).

Immunohistochemistry (IHC) assay

Paraffin-embedded human HCC tissue sections were subjected to immunostaining using an UltraSensitiveTM SP (Mouse/Rabbit) IHC Kit (MXB, KIT-9710). After deparaffinization and antigen retrieval, the tissue sections were incubated overnight with corresponding primary antibodies listed in (**Additional file 10: Table S4**). Signal amplification and detection were conducted using DAB system following the manufacturer's instructions (MXB, MAX-001).

Identification of robust DEGs

In our study, the data from five microarray datasets (GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248) were normalized and the differentially expressed genes (DEGs) between HCC and corresponding normal samples were screened using "limma" and "GEOquery" R packages provided by bioconductor. The Benjamini-Hochberg method was used to adjust original p-value of each gene and an adjusted p-value <0.05 was considered as the cut-off criterion. Moreover, Robust Rank Aggregation (RRA) analysis was adopted to identify the most significant DEGs by integrating the results of all 5 microarray datasets [9]. Genes with adjusted p-value < 0.05 were defined as significant DEGs in the RRA analysis.

Functional enrichment analysis

Gene Ontology (GO) [10] is an open database constructed by the Gene Ontology Consortium, which mainly consists of three parts: biological process (BP), cellular

component (CC) and molecular function (MF). KEGG [11] is an online database for functions and signaling pathways analysis of different gene sets based on genome sequencing datasets. In this study, the clusterProfiler v3.14.3 [12] in R language was used for function enrichment analysis of DEGs. GO terms and KEGG terms of the identified DEGs with adjusted p-value < 0.05 were considered to be significantly enriched.

PPI network construction and hub genes selection

The Search Tool for the Retrieval of Interacting Genes (STRING) [13] (http://string-db.org) was used to predict the protein-protein interaction (PPI) network of the DEGs, which was visualized by Cytoscape software V3.8.0. The CytoHubba [14], a plug-in of Cytoscape, was adopted to select top 10 hub genes. GeneCards (https://www.genecards.org/) [15] was used to search for full names and biological functions of 10 hub genes.

Correlation analysis

Pearson's correlation test was applied to do correlation analysis of the expression of 10 hub genes in five datasets (GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248) respectively. The correlation among the hub genes were visualized by heatmaps using R language.

Gene set enrichment analysis (GSEA)

We utilized Java GSEA v4.1.0 desktop program (http://software.broadinstitute.org/gsea/datasets.jsp) to conduct GSEA analysis of ten hub genes. Based on the median expression of each hub gene in GSE121248 dataset, a total of 70 HCC samples were divided into two groups (high expression vs low expression). The reference gene set "c2.all.v7.2.symbols.gmt" was downloaded from the Molecular Signature Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/index.jsp). Enriched pathways with a cutoff threshold of p-value < 0.05 and FDR < 0.25 were considered statistically significant.

Validation and survival analysis of hub genes

Overall survival (OS) and progression-free survival (PFS) analysis of the hub genes were performed by the online tool Kaplan-Meier Plotter (http://kmplot.com/analysis/). Besides, the gene expression and promoter methylation levels analysis between HCC and adjacent normal samples as well as their correlations with clinical features in liver tissues were conducted by UALCAN (UALCAN, http://ualcan.path.uab.edu/index.html) [16] based on the Cancer Genome Atlas (TCGA) datasets.

Correlation analysis of gene expression and tumor-infiltrating immune cells

The correlation between the expression of identified 10 hub genes and tumor infiltrating immune cells (B cells, CD4+ T cells, CD8+ T cells, macrophages,

neutrophils and dendritic cells) as well as tumor purity was analyzed by the online tool TIMER (https://cistrome.shinyapps.io/timer/) [17,18] based on the TCGA database.

Genetic alterations of hub genes

CBioPortal (https://www.cbioportal.org) [19] provides a publicly web resource for exploring and analyzing various cancer genomics data from cancer tissues and cell lines. In our study, the genomic profiles of 10 hub genes including mutations, putative copy-number alterations from GISTIC and mRNA Expression with a z-score threshold ± 2.0 were all analyzed via cBioPortal.

Statistical analysis

Comparisons in transcriptional expression were conducted using Student's t-test. Statistical analysis was performed using GraphPad Prism software (version 8.0.1). P-value < 0.05 was considered statistically significant, with significance defined as $P < 0.05$ (*), $P < 0.01$ (**) and $P < 0.001$ (***).

Results

Identification of robust DEGs by the RRA method

The workflow for identification, validation, and functional analysis of DEGs in our study is shown in **Fig. 1**. Main characteristics of five HCC datasets, including dataset ID, study country, sample number in each group, platform ID, and number of genes

in each platform are all presented in (**Additional file 6: Table S1**). After standardization of the microarray results, we obtained totally 1,017 DEGs (619 downregulated and 398 upregulated mRNAs), 1,088 DEGs (701 downregulated and 387 upregulated mRNAs), 1,435 DEGs (912 downregulated and 523 upregulated mRNAs), 1,196 DEGs (744 downregulated and 452 upregulated mRNAs), 666 DEGs (465 downregulated and 201 upregulated mRNAs) from GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248 respectively. By bioinformatics analysis using R language, the DEGs between HCC and normal samples in GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248 were presented in the volcano plots (**Fig. 2A-E**). Based on the results of RRA analysis, a total of 126 significant DEGs (48 upregulated and 78 downregulated DEGs) were identified (**Additional file 7: Table S2**). The top 20 upregulated and downregulated DEGs according to adjusted p-values were presented in heatmaps (**Fig. 3**). In addition, the expression of these top 20 upregulated and 20 downregulated DEGs identified from each dataset (GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248) were displayed by heatmap visualization (**Additional file 1: Fig. S1A-E**).

GO enrichment and KEGG pathway analysis

Next, we applied the most significant DEGs to do GO enrichment and KEGG pathway analysis by clusterProfiler v3.14.3 in R language. The GO enrichment analysis indicated that DEGs were significantly enriched in biological processes of "organelle fission", "nuclear division", "chromosome segregation" and "mitotic

nuclear division" (**Fig. 4A**). In terms of cellular component, "spindle", "microtubule", "chromosomal region" and "midbody" were the most significantly enriched GO terms (**Fig. 4B**). Variations in DEGs related with MF were markedly enriched in "tetrapyrrole binding", "oxidoreductase activity", "iron ion binding", "heme binding" and "monooxygenase activity" (**Fig. 4C**). Furthermore, KEGG analysis results suggested that the top canonical pathways mostly associated with DEGs included "caffeine metabolism", "TGF-beta DNA replication" and "tryptophan metabolism" (**Fig. 4D**).

Construction of PPI network and hub genes analysis

Then we constructed the PPI network of significant DEGs, which consists of 661 edges and 120 Nodes. After using cytoHubba of Cytoscape 3.8.0, 10 hub genes (BUB1B, CDKN3, CCNB1, CCNB2, CDK1, TOP2A, CDC20, MELK, NUSAP1 and AURKA) from the PPI network with relative higher degrees were selected (**Fig. 5A**). Summaries for gene symbols, full names and biological functions of 10 hub genes were presented in (**Additional file 9: Table S3**). Furthermore, through analyzing the expression data of 10 hub genes using R language, the strong correlations between the hub genes were also confirmed in GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248 respectively (**Fig. 5B-F**). To further investigate the molecular functions of these hub genes in HCC, we performed GSEA analysis of ten hub genes based on GSE121248 dataset. The results showed these genes are significantly enriched in eight signaling pathways: PID ATM pathway, regulation of

Tp53 activity, Kauffmann DNA repair genes, transcription coupled nucleotide excision repair, sumoylation of sumoylation proteins, nucleotide excision repair, regulation of Tp53 activity through phosphorylation and angiogenesis **(Fig. 6A–H)**.

Expression validation and survival analysis of hub genes

In order to further verify the reliability of the 10 hub genes, we analyzed the data on HCC and adjacent normal tissue samples as well as their correlations with clinical features by the online tool UALCAN based on TCGA datasets. We found that all of them were markedly upregulated in HCC tissue samples compared with adjacent normal samples (**Fig. 7A-D and Additional file 2: Fig. S2A-F**). Besides, the hub genes were significantly differentially expressed in HCC samples with different tumor grades, which indicated that higher expression levels were associated with higher tumor grades (**Fig. 7E-H and Additional file 2: Fig. S2G-L**).

To make sure if the methylation status of hub genes contributes to HCC pathogenesis, we explored the association between expression levels of these hub genes and their promoter methylation levels in HCC samples. As a result, we detected that the promoter methylation levels of CCNB2, CDK1, TOP2A and CCNB1 were significantly decreased in HCC samples compared with paracancerous normal samples (**Fig. 7I-L**). The above figures presented Beta values of CpG probes located up to 1500bp upstream of genes' start site (TSS200, TSS1500), which was the ratio of the methylated probe intensity and the sum of methylated and unmethylated probe intensity. In addition, we also did overall survival (OS) and progression-free survival

(PFS) analysis of the hub genes by the online tool Kaplan-Meier Plotter. Kaplan-Meier curves showed that higher expression of these genes was significantly associated with shorter OS and PFS of patients with HCC (**Fig. 8A-L and Additional file 3: Fig. S3A-H**), which suggested that their high diagnostic and prognostic values as biomarker candidates.

Association of hub genes' expression with tumor purity and immune infiltration

It was widely accepted that infiltrating immune cells, an important component of the tumor microenvironment, played a crucial role in HCC pathogenesis [20]. To better investigate immunotherapy targets, we continued to explore potential associations between HCC hub genes' expression levels and tumor purity as well as infiltration of immune cells by the online tool TIMER. The tumor purity was estimated through DNA SNP array data using previously developed methods [21, 22]. Interestingly, the analysis showed that all hub genes were positively associated with tumor purity and infiltrating immune cells including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells (**Fig. 9A-F and Additional file 4: Fig. S4A-D**).

Genetic alterations of hub genes and their associations with overall survival of HCC patients

To better evaluate the functions of selected hub genes in HCC, we continued to analyze genetic alterations in these hub genes and their associations with OS of 348 HCC patients in cBioPortal (https://www.cbioportal.org/) based on TCGA datasets.

As a result, we found that multiple hub genes had relatively high genetic alteration rates in HCC patients. Among them, CCNB2, AURKA, CCNB1 and NUSAP1 were the top four genes with high genetic alterations, and their alteration rates were 9%, 9%, 8% and 8% respectively (**Additional file 5: Fig. S5A**). Next, we selected hub genes with alteration rates over 8% and analyzed their associations with overall survival of 348 HCC patients. Interestingly, patients with CCNB2, NUSAP1, AURKA and CCNB1 alterations presented with relatively shorter OS, while the difference of NUSAP1 and AURKA was not statistically significant (**Additional file 5: Fig. S5B-E**). These results indicate that genetic alterations of hub genes in HCC may have a strong effect on HCC patients' prognosis.

The expression and clinical relevance of hub genes in HCC tissue samples

To further validate the expression of hub genes in HCC tissue samples, we performed qPCR experiment to detect the mRNA levels of ten hub genes in 25 pairs of HCC tumor tissues and adjacent non-tumor samples. The results showed that all the mRNA levels of these hub genes were significantly upregulated in HCC tissues relative to control (**Fig. 10A-J)**, which were consistent with the previous results by bioinformatics analysis. Next, we analyzed the correlation between the expression of hub genes and the clinicopathologic characteristics of these HCC patients. We divided 25 HCC patients into low and high groups using median as the cut off from the above qPCR results and performed Pearson's Chi-square test. The results indicated that the expression levels of CCNB1 (**Table 1**), CCNB2 (**Additional file 11: Table S6**) and

TOP2A (**Table 4**) were significantly correlated with vascular invasion and AFP levels of HCC patients. The expression of CDC20 (**Table 2**), BUB1B (**Table 3**), CDK1 (**Additional file 12: Table S7**) and CDKN3 (**Additional file 13: Table S8**) were significantly correlated with AFP levels of HCC patients. Besides, NUSAP1 (**Additional file 14: Table S9**) and AURKA (**Additional file 15: Table S10**) expression were significantly associated with tumor number of HCC patients. And MELK (**Additional file 16: Table S11**) expression was strongly associated with vascular invasion of HCC patients. In conclusion, these results prove that selected hub genes were indeed upregulated in HCC and significantly correlated with clinicopathologic characteristics of HCC patients, which can be used as useful biomarkers in HCC diagnosis and treatment. Moreover, immunohistochemistry staining indicated that the protein expression levels of these hub genes were much higher in tumor samples than in non-tumor samples (**Fig. 11**).

Discussion

Hepatocellular carcinoma (HCC) is the fifth most common cancers worldwide with high morbidity and mortality [23], but its molecular mechanisms still remain unclear now. Thus, further exploring the molecular mechanisms of HCC for finding better therapeutic strategies is of great significance. Though microarray technology and bioinformatic analysis [24] have provided important new insights into discovering novel biomarkers and biological pathways involved in various diseases, inconsistencies were observed between different studies [25]. In our study, we have

identified 126 significant DEGs (78 downregulated and 48 upregulated genes) after integrating microarray data from five datasets (GSE41804, GSE45267, GSE84402, GSE107170 and GSE121248) using the RRA method [26].

The GO enrichment analysis indicated that the robust DEGs identified were significantly enriched in biological processes of “organelle fission”, “nuclear division”, “chromosome segregation” and “mitotic nuclear division”. Variations in DEGs related with MF were markedly enriched in “tetrapyrrole binding”, “oxidoreductase activity”, “iron ion binding”, “heme binding” and “monooxygenase activity”. Furthermore, KEGG analysis results suggested that the top significant pathways mostly included “caffeine metabolism”, “TGF-beta DNA replication” and “tryptophan metabolism”. Li-Ting Wang et al found that intestine-specific homeobox gene ISX could regulate immune suppression in HCC by modulating tryptophan catabolism and IL-6 signaling [27]. Delia Hoffmann group have identified the expression of Tryptophan 2,3-Dioxygenase expression in human hepatocellular carcinoma cells, which indicated that TDO might be an immunotherapeutic target in hepatocellular carcinoma [28]. Besides, caffeine has shown tumouricidal activity in HCC cells and was associated with decreased risk of HCC recurrence [29]. Quinone oxidoreductase1 (NQO1) transcription might be inhibited by promoter hypermethylation in HCC, which might be involved in HCC pathogenesis [30]. Quinone oxidoreductase-1 in HCC cells protects against ferroptosis in HCC by modulating iron metabolism and lipid peroxidation [31]. In addition, Xanthine Oxidoreductase was identified as a tumor suppressor that facilitated the development

of HCC [32]. Consistent with these published data, the enriched biological functions and pathways provide basis for further research into the molecular mechanisms of HCC.

We also obtained 10 hub genes from the constructed PPI network including MELK, NUSAP1, BUB1B, CDKN3, AURKA, CCNB1, CCNB2, CDK1, TOP2A and CDC20 with relatively higher degrees. MELK, a cell-cycle-dependent protein kinase, is involved in the pathogenesis and recurrence of HCC [33]. The overall survival (OS) and recurrence-free survival (RFS) of the high MELK mRNA expression group was significantly shorter than that of the lower MELK expression group [34]. As a microtubule-associated protein with the capacity to bundle and stabilize microtubules, NUSAP1 plays essential part in diverse biological processes both physiologically and pathologically [35]. In vitro studies indicated that HepG2 and Huh-7 cell proliferation and invasion was inhibited significantly following NUSAP1 knockdown [36]. Besides, NUSAP1 was found to be a target of miR193A-5p and its increased levels correlated with HCC patients' shorter survival times [37]. The multidomain protein kinases BUBR1 encoded by BUB1B is a central component of the mitotic checkpoint for spindle assembly ensuring chromosomes segregation during mitosis [38]. BUB1B promoted mTORC1 signaling pathway in HCC, which could be blocked by rapamycin [39]. Cyclin-dependent kinase inhibitor 3 (CDKN3) is a dual-specificity phosphatase that associated with Cdk2 and CDKN3 overexpression could delay the G1-S phase transition [40]. CDKN3 expression was negatively associated with the pathological stage of liver cancer and CDKN3 inhibition could facilitate the

clonogenic capacity and chemotherapeutic tolerance in HCC tissues [41]. AURKA has been defined as a crucial regulator of mitotic chromosome segregation through its catalytic activity [42, 43]. The oncogenic phosphatase PRL-3 is upregulated in metastatic colorectal cancer and it interacted with AURKA and FZR1, thus influencing the development of colorectal cancer [44]. It was reported that thiostrepton could reduce FOXM1 target genes expression including AURKA and CCNB1, which contributed to the progression of G2/M cell cycle [45]. The coordinated actions of the cyclin B-dependent kinase (CDK1-CCNB) play central roles in promoting the error-free chromosome segregation [46, 47]. YAP promotes liver cancer cells proliferation through the induction of several cell cycle regulators including CCNB1 and CCNB2 [48]. In addition, CDK1 is a common serine/threonine kinase that has been proven to phosphorylate a number of different substrates during the process of mitosis and have a large effect on cell morphology [49, 50]. Tim Hon Man Chan lab has revealed a new pathway CHD1L/TCTP/Cdc25C/CDK1 in HCC progression [51]. Anti-CDK1 treatment can increase the efficacy of sorafenib in PDX hepatocellular carcinoma models [52]. TOP2A is a type II topoisomerase that functions during mitosis and meiosis for proper segregation of daughter chromosomes [53]. TOP2A overexpression was associated with shorter HCC patients' survival, which highlighted the potential prognostic value of TOP2A in HCC [54]. Moreover, TRRAP induced HCC cell proliferation by upregulating mitotic gene TOP2A, which provided a potential therapeutic strategy for HCC [55]. Ayuko Saeki et al have detected some polymorphic base changes of CDC20 in HCC cell lines or

specimens [56]. SIRT2 regulates the anaphase-promoting complex/cyclosome activity via deacetylation of CDC20 and thus inhibits tumorigenesis [57]. In our study, we confirmed the expression levels of 10 hub genes were indeed significantly upregulated in HCC tissues by qPCR analysis and immunohistochemistry staining. In addition, selected hub genes were significantly correlated with clinicopathologic characteristics of HCC patients, which can be used as useful biomarkers in HCC diagnosis and treatment.

Furthermore, we determined that a majority of hub genes were not only significantly upregulated in HCC samples, but also correlated with higher tumor stage, suggesting their vital roles in HCC pathogenesis. Kaplan-Meier curves showed that higher expression of all these hub genes was significantly associated with shorter OS and PFS of patients with HCC, which also confirmed their great diagnostic and prognostic values as therapeutic targets and prognosis predictors. Abundant evidence has demonstrated that DNA methylation status of genes played a key role in HCC, which help develop molecular diagnoses and individualized treatments [58, 59]. To make sure if the methylation status of hub genes contributes to HCC pathogenesis, we referred to UALCAN to explore the association between expression levels of these hub genes and their promoter methylation levels in HCC samples. As a result, we detected that CCNB2, CDK1, TOP2A and CCNB1 were hypomethylated in HCC samples compared with adjacent normal samples, which is consistent with their upregulation levels in HCC. Genetic abnormal alteration including gene amplification, deletion and mutation are reported to be closely correlated with the initiation and

progression of cancers [60]. Therefore, we analyzed genetic alterations in selected hub genes and their associations with OS of 348 HCC patients and found that multiple hub genes had relatively high genetic alteration rates in HCC patients. CCNB2, AURKA, CCNB1 and NUSAP1 were the top four genes with high genetic alterations. Interestingly, patients with CCNB2, NUSAP1, AURKA and CCNB1 alterations presented with relatively shorter OS, while the difference of NUSAP1 and AURKA was not statistically significant.

In summary, we have identified a set of robust DEGs from five microarray datasets related with HCC development by RRA analysis. Ten hub genes (BUB1B, CDKN3, CCNB1, CCNB2, CDK1, TOP2A, CDC20, MELK, NUSAP1 and AURKA) were strongly upregulated in HCC samples and their higher expression was significantly associated with shorter OS and PFS of HCC patients. Though more work needs to be done to validate their contribution to the pathogenesis of HCC by in vivo and in vitro experiments, we believe these data would help enhance the current understanding of the HCC progression to some extent.

Abbreviations

HCC, Hepatocellular Carcinoma; GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein network; GO, Gene Ontology; BP, biologic process. CC, cellular component; MF, molecular function; GSEA, Gene Set Enrichment Analysis; TCGA, The Cancer Genome Atlas.

Authors' contributions

SH and WH: research design, data collection and bioinformatics analysis. JY and FY: qPCR experiments, immunohistochemistry staining and statistical analysis. XL and LY: wrote the manuscript. ZW and YX: guided research ideas, design, research methods and manuscript revision.

Funding

This work was funded by grants from the National Natural Science Foundation of China (Grant no. 81971877) to Zhongxing Wang.

Availability of data and materials

The data (GSE41804, GSE45267, GSE84402, GSE107170, GSE121248) included in the current study are available in the GEO database (https://www.ncbi.nlm.nih.gov/geo).

Ethics approval and consent to participate

Our research was approved by the Ethics Committee of the Third Affiliated Hospital of Sun Yat-sen University ([2018]02-043-01). HCC patients (or their parents or guardians) in our research have all signed the written informed consent form. All methods were performed in accordance with the relevant guidelines and regulations.

Patient consent for publication

NA.

Competing interests

The authors declare that there are no conflict of interests.

References

1. Feng GS, Hanley KL, Liang Y, Lin X. Improving the Efficacy of Liver Cancer Immunotherapy: the Power of Combined Preclinical and Clinical Studies. Hepatology. 2020 Jul 26. doi: 10.1002/hep.31479.
2. Petrowsky H, Fritsch R, Guckenberger M, De Oliveira ML, Dutkowski P, Clavien PA. Modern therapeutic approaches for the treatment of malignant liver tumours. Nat Rev Gastroenterol Hepatol. 2020 Jul 17. doi: 10.1038/s41575-020-0314-8.
3. Finn RS, Zhu AX. Evolution of Systemic Therapy for Hepatocellular Carcinoma. Hepatology. 2020 May 7. doi: 10.1002/hep.31306.
4. Dal Bo M, De Mattia E, Baboci L, Mezzalira S, Cecchin E, Assaraf YG, et al. New insights into the pharmacological, immunological, and CAR-T-cell approaches in the treatment of hepatocellular carcinoma. Drug Resist Updat. 2020 Jul;51:100702. doi: 10.1016/j.drup.2020.100702.
5. Qi L, Chan TH, Tenen DG, Chen L. RNA editome imbalance in hepatocellular carcinoma. Cancer Res. 2014 Mar 1;74(5):1301-6. doi: 10.1158/0008-5472.CAN-13-3485.

6. Zender L, Villanueva A, Tovar V, Sia D, Chiang DY, Llovet JM. Cancer gene discovery in hepatocellular carcinoma. J Hepatol. 2010 Jun;52(6):921-9. doi: 10.1016/j.jhep.2009.12.034.
7. Tenenbaum JD, Bhuvaneshwar K, Gagliardi JP, Fultz Hollis K, Jia P, Ma L, et al. Translational bioinformatics in mental health: open access data sources and computational biomarker discovery. Brief Bioinform. 2019 May 21;20(3):842-856. doi: 10.1093/bib/bbx157.
8. Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. Nat Commun. 2016 Sep 26;7:12846. doi: 10.1038/ncomms12846.
9. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012 Feb 15;28(4):573-80. doi: 10.1093/bioinformatics/btr709.
10. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9. doi: 10.1038/75556.
11. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017 Jan 4;45(D1):D353-D361. doi: 10.1093/nar/gkw1092.

12. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284-7. doi: 10.1089/omi.2011.0118.
13. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan;43(Database issue):D447-52. doi: 10.1093/nar/gku1003.
14. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11.
15. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics. 1998;14(8):656-64. doi: 10.1093/bioinformatics/14.8.656.
16. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analysis. Neoplasia. 2017 Aug;19(8):649-658. doi: 10.1016/j.neo.2017.05.002.
17. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analysis of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016 Aug 22;17(1):174. doi: 10.1186/s13059-016-1028-7.
18. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for

549 Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 2017
550 Nov 1;77(21):e108-e110. doi: 10.1158/0008-5472.CAN-17-0307.
551 19. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio
552 cancer genomics portal: an open platform for exploring multidimensional cancer
553 genomics data. Cancer Discov. 2012 May;2(5):401-4.
554 20. Rizvi S, Wang J, El-Khoueiry AB. Liver Cancer Immunity. Hepatology. 2020 Jun
555 9. doi: 10.1002/hep.31416.
556 21. Li B, Li JZ. A general framework for analyzing tumor subclonality using SNP
557 array and DNA sequencing data. Genome Biol. 2014 Sep 25;15(9):473. doi:
558 10.1186/s13059-014-0473-4.
559 22. Li B, Senbabaoglu Y, Peng W, Yang ML, Xu J, Li JZ. Genomic estimates of
560 aneuploid content in glioblastoma multiforme and improved classification. Clin
561 Cancer Res. 2012 Oct 15;18(20):5595-605. doi: 10.1158/1078-0432.CCR-12-
562 1427. Epub 2012 Aug 21. Erratum in: Clin Cancer Res. 2014 May 1;20(9):2500.
563 23. Qiu Z, Li H, Zhang Z, Zhu Z, He S, Wang X, et al. A Pharmacogenomic
564 Landscape in Human Liver Cancers. Cancer Cell. 2019 Aug 12;36(2):179-
565 193.e11. doi: 10.1016/j.ccell.2019.07.001.
566 24. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based
567 artificial intelligence in tumor pathology. Cancer Commun (Lond). 2020
568 Apr;40(4):154-166. doi: 10.1002/cac2.12012.
569 25. Barbieri CE, Chinnaiyan AM, Lerner SP, Swanton C, Rubin MA. The Emergence
570 of Precision Urologic Oncology: A Collaborative Review on Biomarker-driven

Therapeutics. Eur Urol. 2017 Feb;71(2):237-246. doi: 10.1016/j.eururo.2016.08.024.

26. Hou Q, Bing ZT, Hu C, Li MY, Yang KH, Mo Z, et al. RankProd Combined with Genetic Algorithm Optimized Artificial Neural Network Establishes a Diagnostic and Prognostic Prediction Model that Revealed C1QTNF3 as a Biomarker for Prostate Cancer. EBioMedicine. 2018 Jun;32:234-244. doi: 10.1016/j.ebiom.2018.05.010.

27. Wang LT, Chiou SS, Chai CY, Hsi E, Yokoyama KK, Wang SN, et al. Intestine-Specific Homeobox Gene ISX Integrates IL6 Signaling, Tryptophan Catabolism, and Immune Suppression. Cancer Res. 2017 Aug 1;77(15):4065-4077. doi: 10.1158/0008-5472.CAN-17-0090.

28. Hoffmann D, Dvorakova T, Stroobant V, Bouzin C, Daumerie A, Solvay M, et al. Tryptophan 2,3-Dioxygenase Expression Identified in Human Hepatocellular Carcinoma Cells and in Intratumoral Pericytes of Most Cancers. Cancer Immunol Res. 2020 Jan;8(1):19-31. doi: 10.1158/2326-6066.CIR-19-0040.

29. Wiltberger G, Wu Y, Lange U, Hau HM, Tapper E, Krenzien F, et al. Protective effects of coffee consumption following liver transplantation for hepatocellular carcinoma in cirrhosis. Aliment Pharmacol Ther. 2019 Mar;49(6):779-788. doi: 10.1111/apt.15089.

30. Tada M, Yokosuka O, Fukai K, Chiba T, Imazeki F, Tokuhisa T, et al. Hypermethylation of NAD(P)H: quinone oxidoreductase 1 (NQO1) gene in human hepatocellular carcinoma. J Hepatol. 2005 Apr;42(4):511-9. doi:

10.1016/j.jhep.2004.11.024.

31. Sun X, Ou Z, Chen R, Niu X, Chen D, Kang R, et al. Activation of the p62-Keap1-NRF2 pathway protects against ferroptosis in hepatocellular carcinoma cells. Hepatology. 2016 Jan;63(1):173-84. doi: 10.1002/hep.28251.

32. Sun Q, Zhang Z, Lu Y, Liu Q, Xu X, Xu J, et al. Loss of Xanthine Oxidoreductase Potentiates Propagation of Hepatocellular Carcinoma Stem Cells. Hepatology. 2020 Jun;71(6):2033-2049. doi: 10.1002/hep.30978.

33. Xia H, Kong SN, Chen J, Shi M, Sekar K, Seshachalam VP, et al. MELK is an oncogenic kinase essential for early hepatocellular carcinoma recurrence. Cancer Lett. 2016 Dec 1;383(1):85-93. doi: 10.1016/j.canlet.2016.09.017.

34. Hiwatashi K, Ueno S, Sakoda M, Iino S, Minami K, Yonemori K, et al. Expression of Maternal Embryonic Leucine Zipper Kinase (MELK) Correlates to Malignant Potentials in Hepatocellular Carcinoma. Anticancer Res. 2016 Oct;36(10):5183-5188. doi: 10.21873/anticanres.11088.

35. Zhao Y, He J, Li Y, Lv S, Cui H. NUSAP1 potentiates chemoresistance in glioblastoma through its SAP domain to stabilize ATR. Signal Transduct Target Ther. 2020 Apr 22;5(1):44. doi: 10.1038/s41392-020-0137-7.

36. Wang Y, Ju L, Xiao F, Liu H, Luo X, Chen L, et al. Downregulation of nucleolar and spindle-associated protein 1 expression suppresses liver cancer cell function. Exp Ther Med. 2019 Apr;17(4):2969-2978. doi: 10.3892/etm.2019.7314.

37. Roy S, Hooiveld GJ, Seehawer M, Caruso S, Heinzmann F, Schneider AT, et al. microRNA 193a-5p Regulates Levels of Nucleolar- and Spindle-Associated

615 Protein 1 to Suppress Hepatocarcinogenesis. Gastroenterology. 2018
616 Dec;155(6):1951-1966.e26. doi: 10.1053/j.gastro.2018.08.032.
617 38. Bolanos-Garcia VM, Blundell TL. BUB1 and BUBR1: multifaceted kinases of
618 the cell cycle. Trends Biochem Sci. 2011 Mar;36(3):141-50. doi:
619 10.1016/j.tibs.2010.08.004.
620 39. Qiu J, Zhang S, Wang P, Wang H, Sha B, Peng H, et al. BUB1B promotes
621 hepatocellular carcinoma progression via activation of the mTORC1 signaling
622 pathway. Cancer Med. 2020 Sep 25. doi: 10.1002/cam4.3411.
623 40. Gyuris J, Golemis E, Chertkov H, Brent R. Cdi1, a human G1 and S phase protein
624 phosphatase that associates with Cdk2. Cell. 1993 Nov 19;75(4):791-803. doi:
625 10.1016/0092-8674(93)90498-f.
626 41. Dai W, Miao H, Fang S, Fang T, Chen N, Li M. CDKN3 expression is negatively
627 associated with pathological tumor stage and CDKN3 inhibition promotes cell
628 survival in hepatocellular carcinoma. Mol Med Rep. 2016 Aug;14(2):1509-14.
629 doi: 10.3892/mmr.2016.5410.
630 42. Bischoff JR, Plowman GD. The Aurora/Ipl1p kinase family: regulators of
631 chromosome segregation and cytokinesis. Trends Cell Biol. 1999 Nov;9(11):454-
632 9. doi: 10.1016/s0962-8924(99)01658-x.
633 43. Nigg EA. Mitotic kinases as regulators of cell division and its checkpoints. Nat
634 Rev Mol Cell Biol. 2001 Jan;2(1):21-32. doi: 10.1038/35048096.
635 44. Zhang C, Qu L, Lian S, Meng L, Min L, Liu J, et al. PRL-3 Promotes
636 Ubiquitination and Degradation of AURKA and Colorectal Cancer Progression

via Dephosphorylation of FZR1. Cancer Res. 2019 Mar 1;79(5):928-940. doi: 10.1158/0008-5472.CAN-18-0520.

45. Willems E, Dedobbeleer M, Digregorio M, Lombard A, Lumapat PN, Rogister B. The functional diversity of Aurora kinases: a comprehensive review. Cell Div. 2018 Sep 19;13:7. doi: 10.1186/s13008-018-0040-6.

46. Nurse P. Universal control mechanism regulating onset of M-phase. Nature. 1990 Apr 5;344(6266):503-8. doi: 10.1038/344503a0.

47. Sivakumar S, Gorbsky GJ. Spatiotemporal regulation of the anaphase-promoting complex in mitosis. Nat Rev Mol Cell Biol. 2015 Feb;16(2):82-94. doi: 10.1038/nrm3934.

48. Wei T, Weiler SME, Tóth M, Sticht C, Lutz T, Thomann S, et al. YAP-dependent induction of UHMK1 supports nuclear enrichment of the oncogene MYBL2 and proliferation in liver cancer cells. Oncogene. 2019 Jul;38(27):5541-5550. doi: 10.1038/s41388-019-0801-y.

49. Malumbres M. Cyclin-dependent kinases. Genome Biol. 2014;15(6):122. doi: 10.1186/gb4184.

50. Blethrow JD, Glavy JS, Morgan DO, Shokat KM. Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. Proc Natl Acad Sci U S A. 2008 Feb 5;105(5):1442-7. doi: 10.1073/pnas.0708966105.

51. Chan TH, Chen L, Liu M, Hu L, Zheng BJ, Poon VK, et al. Translationally controlled tumor protein induces mitotic defects and chromosome missegregation in hepatocellular carcinoma development. Hepatology. 2012 Feb;55(2):491-505.

659 doi: 10.1002/hep.24709.
660 52. Wu CX, Wang XQ, Chok SH, Man K, Tsang SHY, Chan ACY, et al. Blocking
661 CDK1/PDK1/β-Catenin signaling by CDK1 inhibitor RO3306 increased the
662 efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical
663 model of hepatocellular carcinoma. Theranostics. 2018 Jun 13;8(14):3737-3750.
664 doi: 10.7150/thno.25487.
665 53. Nielsen CF, Zhang T, Barisic M, Kalitsis P, Hudson DF. Topoisomerase IIα is
666 essential for maintenance of mitotic chromosome structure. Proc Natl Acad Sci U
667 S A. 2020 Jun 2;117(22):12131-12142. doi: 10.1073/pnas.2001760117.
668 54. Wong N, Yeo W, Wong WL, Wong NL, Chan KY, Mo FK, et al. TOP2A
669 overexpression in hepatocellular carcinoma correlates with early age onset,
670 shorter patients survival and chemoresistance. Int J Cancer. 2009 Feb
671 1;124(3):644-52. doi: 10.1002/ijc.23968.
672 55. Kwan SY, Sheel A, Song CQ, Zhang XO, Jiang T, Dang H, et al. Depletion of
673 TRRAP Induces p53-Independent Senescence in Liver Cancer by Down-
674 Regulating Mitotic Genes. Hepatology. 2020 Jan;71(1):275-290. doi:
675 10.1002/hep.30807.
676 56. Saeki A, Tamura S, Ito N, Kiso S, Matsuda Y, Yabuuchi I, et al. Frequent
677 impairment of the spindle assembly checkpoint in hepatocellular carcinoma.
678 Cancer. 2002 Apr 1;94(7):2047-54. doi: 10.1002/cncr.10448.
679 57. Kim HS, Vassilopoulos A, Wang RH, Lahusen T, Xiao Z, Xu X, et al. SIRT2
680 maintains genome integrity and suppresses tumorigenesis through regulating

APC/C activity. Cancer Cell. 2011 Oct 18;20(4):487-99. doi: 10.1016/j.ccr.2011.09.004.

58. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, et al. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. Theranostics. 2019 Sep 25;9(24):7251-7267. doi: 10.7150/thno.31155.

59. Midorikawa Y, Yamamoto S, Tatsuno K, Renard-Guillet C, Tsuji S, Hayashi A, et al. Accumulation of Molecular Aberrations Distinctive to Hepatocellular Carcinoma Progression. Cancer Res. 2020 Sep 15;80(18):3810-3819. doi: 10.1158/0008-5472.CAN-20-0225.

60. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu; Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017 Jun 15;169(7):1327-1341.e23. doi: 10.1016/j.cell.2017.05.046.

706 **Table 1.** Correlation between CCNB1 expression level and the clinicopathologic
707 characteristics of 25 cases of HCC patients.

Characteristics		CCNB1 expression level		Pearson's chi-square test (p-value∗)
		Low	High	
Gender	Male	10	2	0.225
	Female	8	5	
Age	<60	6	9	0.327
	⩾60	6	4	
Differentiation	High	7	8	0.174
	Moderate/low	2	8	
Tumor size	<5cm	3	4	0.856
	⩾5cm	7	11	
Tumor number	Single	1	3	0.743
	Multiple	7	14	
TNM stage	I+II	2	4	0.702
	III+IV	8	11	
Vascular	No	1	5	0.047
invasion	Yes	12	7	
AFP	<20	1	7	0.007
	⩾20	12	5	

708 ∗ Results were considered statistically significant at $p < 0.05$.
709
710
711
712
713
714
715

716 **Table 2.** Correlation between CDC20 expression level and the clinicopathologic
717 characteristics of 25 cases of HCC patients.

Characteristics		CDC20 expression level		Pearson's chi-square test (p-value∗)
		Low	High	
Gender	Male	10	2	0.114
	Female	7	6	
Age	<60	5	9	0.346
	⩾60	6	5	
Differentiation	High	7	8	0.054
	Moderate/low	1	9	
Tumor size	<5cm	4	4	0.678
	⩾5cm	7	10	
Tumor number	Single	1	4	0.408
	Multiple	8	12	
TNM stage	I+II	2	3	0.840
	III+IV	9	11	
Vascular invasion	No	1	5	0.078
	Yes	11	8	
AFP	<20	1	7	0.007
	⩾20	12	5	

718 ∗ Results were considered statistically significant at $p < 0.05$.
719
720
721
722
723
724
725

Table 3. Correlation between BUB1B expression level and the clinicopathologic characteristics of 25 cases of HCC patients.

Characteristics		BUB1B expression level		Pearson's chi-square test (p-value*)
		Low	High	
Gender	Male	9	2	0.189
	Female	8	6	
Age	<60	3	9	0.096
	⩾60	9	4	
Differentiation	High	7	10	0.432
	Moderate/low	2	6	
Tumor size	<5cm	3	4	0.856
	⩾5cm	7	11	
Tumor number	Single	1	3	0.617
	Multiple	8	13	
TNM stage	I+II	2	4	0.702
	III+IV	8	11	
Vascular invasion	No	3	5	0.471
	Yes	9	8	
AFP	<20	1	5	0.047
	⩾20	12	7	

* Results were considered statistically significant at $p < 0.05$.

Table 4. Correlation between TOP2A expression level and the clinicopathologic characteristics of 25 cases of HCC patients.

Characteristics		TOP2A expression level		Pearson's chi-square test (p-value∗)
		Low	High	
Gender	Male	10	2	0.225
	Female	8	5	
Age	<60	6	9	0.327
	⩾60	6	4	
Differentiation	High	7	8	0.174
	Moderate/low	2	8	
Tumor size	<5cm	3	4	0.856
	⩾5cm	7	11	
Tumor number	Single	1	3	0.617
	Multiple	8	13	
TNM stage	I+II	2	4	0.702
	III+IV	8	11	
Vascular	No	1	5	0.047
invasion	Yes	12	7	
AFP	<20	1	7	0.007
	⩾20	12	5	

∗ Results were considered statistically significant at $p < 0.05$.

Figure legends

Fig. 1 Study workflow.

Fig. 2 Identification of DEGs in GEO. DEGs in GSE41804 (A), GSE45267 (B), GSE84402 (C), GSE107170 (D), GSE121248 (E) dataset were presented.

Fig. 3 Identification of DEGs by RRA analysis. Top 20 upregulated and downregulated genes were presented by heatmap. The numbers represents logarithmic fold change in each dataset.

Fig. 4 GO and KEGG analysis of DEGs. Top 20 enriched GO terms related with biologic process (A), cellular component (B), molecular function (C) for robust DEGs. (D) Top 20 enriched pathways for DEGs.

Fig. 5 Hub genes network and correlation analysis. (A) Ten hub genes network by CytoHubba. Correlation analysis among hub genes in GSE41804 (B), GSE45267 (C), GSE84402 (D), GSE107170 (E), GSE121248 (F) was displayed.

Fig. 6 GSEA analysis based on GSE121248 data. Some representative enriched pathways of (A) NUSAP1, (B) CDK1, (C) MELK, (D) AURKA, (E) BUB1B, (F) CCNB2, (G) CCNB1, (H) CDKN3.

Fig. 7 Expression validation and methylation analysis of hub genes. (A-D) Expression levels of hub genes in HCC. (E-H) Expression analysis of hub genes with tumor grades. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The first layer asterisk above the error bar represents comparisons to normal samples. (I-L) Promoter methylation levels of hub genes.

Fig. 8 Survival analysis. (A-F) Overall survival analysis of hub genes. (G-L) Progression-free survival analysis of hub genes.

Fig. 9 Association of hub genes' expressions with immune infiltration. The correlation between expression of (A) BUB1B, (B) CDKN3, (C) CCNB1, (D) CDK1, (E) MELK, (F) TOP2A and immune infiltration.

Fig. 10 The mRNA expression of hub genes in 25 cases. QPCR analysis of mRNA levels of (A) CCNB1, (B) CCNB2, (C) NUSAP1, (D) BUB1B, (E) CDK1, (F) AURKA, (G) TOP2A, (H) CDC20, (I) CDKN3, (J) MELK.

Fig. 11 Immunohistochemistry of HCC tissues. Representative IHC images of hub genes in HCC samples. Scale bar = 50 μm.

789 Additional file 1. Differential heatmap of robust DEGs in each dataset. (A) The top
790 20 upregulated and 20 downregulated genes according to adjusted p-value identified
791 from GSE84402 (A), GSE41804 (B), GSE45267 (C), GSE121248 (D) and
792 GSE107170 (E) were displayed in the heatmap plot.
793 Additional file 2. Expression validation analysis of hub genes. (A-F) Expression
794 analysis of hub genes in HCC samples. (G-L) Expression analysis of hub genes with
795 different tumor grades performed.
796 Additional file 3. Survival analysis of hub genes. (A-D) Overall survival analysis of
797 hub genes. (E-H) Progression-free survival analysis of hub genes.
798 Additional file 4. Association of hub genes' expressions with tumor purity and
799 immune infiltration. The correlation between the expression of (A) AURKA, (B)
800 CCNB2, (C) CDC20, (D) NUSAP1 and tumor purity, immune infiltration analysis.
801 Additional file 5. Genetic Alterations of Hub Genes and Their Associations with
802 Overall Survival of HCC patients. (A) CCNB2, AURKA, CCNB1 and NUSAP1 were
803 the top four genes with high genetic alterations. Patients with (B) AURKA, (C)
804 CCNB1, (D) CCNB2, (E) NUSAP1 alterations presented with shorter OS.
805 Additional file 6. Detailed information about GEO data were showed in this study.
806 Additional file 7. Screening overlapping DEGs by integrated GEO data.
807 Additional file 8. Summaries for the function of 10 hub genes.
808 Additional file 9. Primers used for qPCR analysis.
809 Additional file 10. Primary antibodies for IHC staining.
810 Additional file 11. Correlation between CCNB2 expression level and the
811 clinicopathologic characteristics of 25 HCC patients.
812 Additional file 12. Correlation between CDK1 expression level and the

clinicopathologic characteristics of 25 HCC patients.

Additional file 13. Correlation between CDKN3 expression level and the clinicopathologic characteristics of 25 HCC patients.

Additional file 14. Correlation between NUSAP1 expression level and the clinicopathologic characteristics of 25 HCC patients.

Additional file 15. Correlation between AURKA expression level and the clinicopathologic characteristics of 25 HCC patients.

Additional file 16. Correlation between MELK expression level and the clinicopathologic characteristics of 25 HCC patients.

Figures

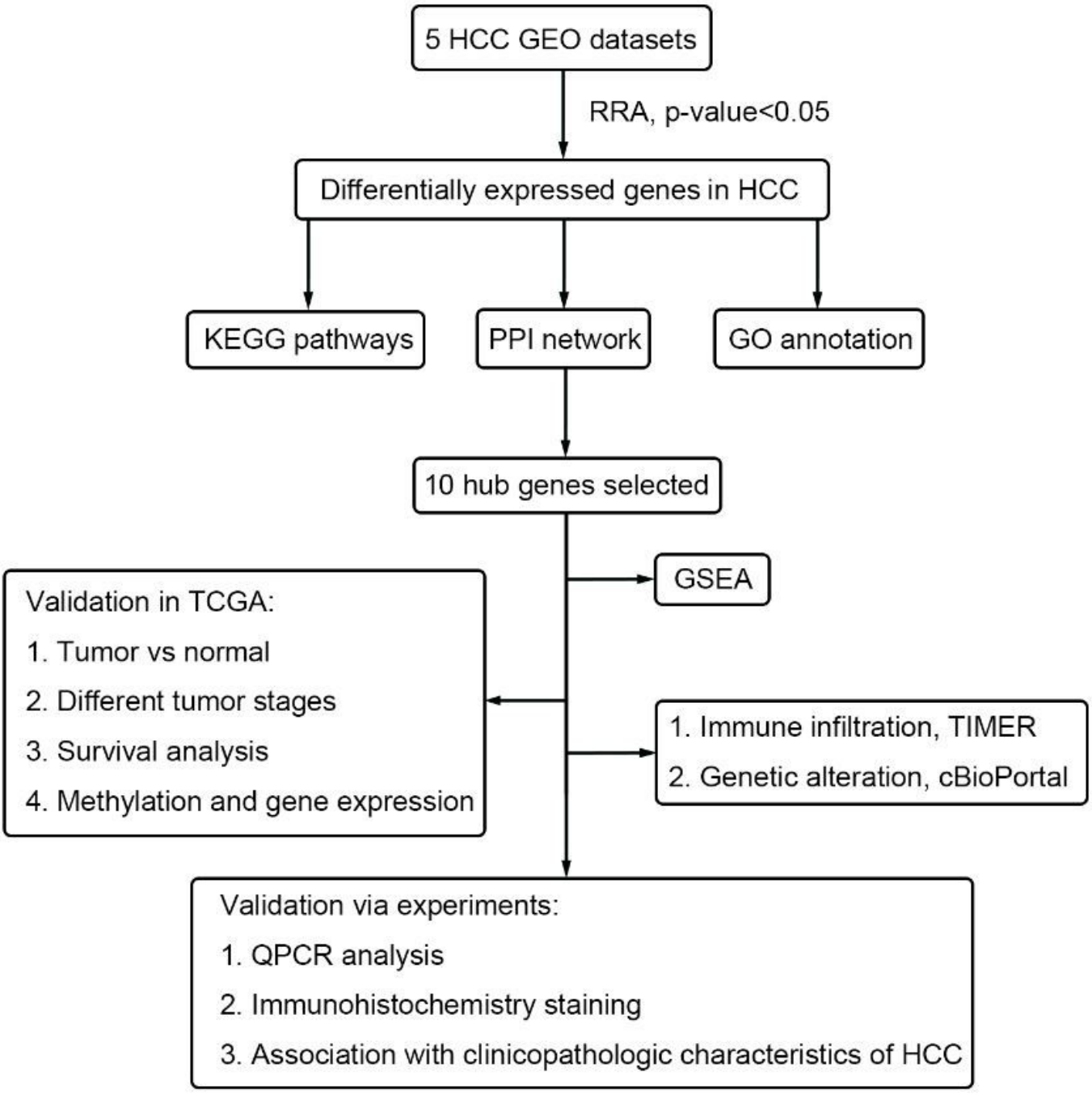

Figure 1

Study workflow.

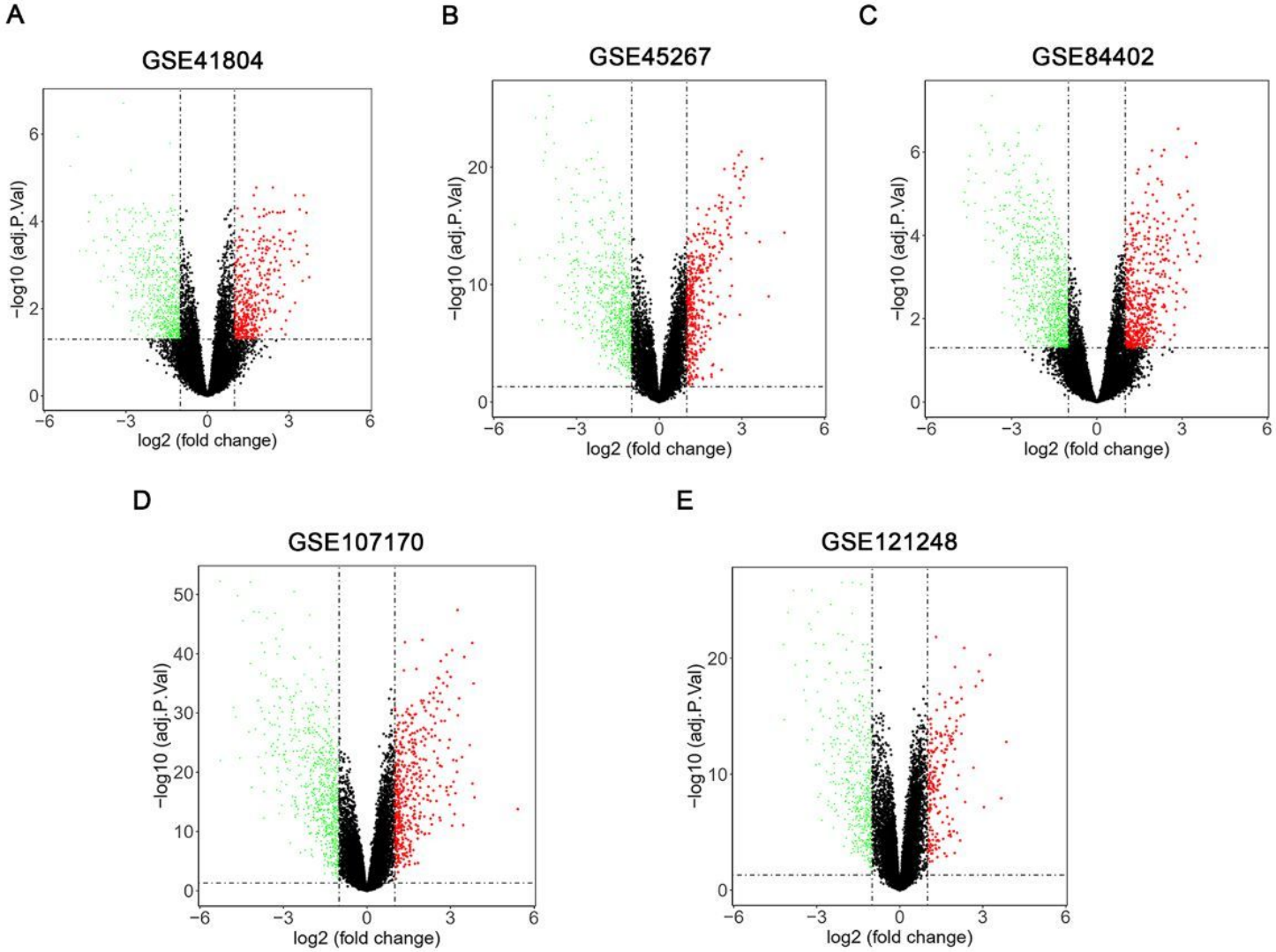

Figure 2

Identification of DEGs in GEO. DEGs in GSE41804 (A), GSE45267 (B), GSE84402 (C), GSE107170 (D), GSE121248 (E) dataset were presented.

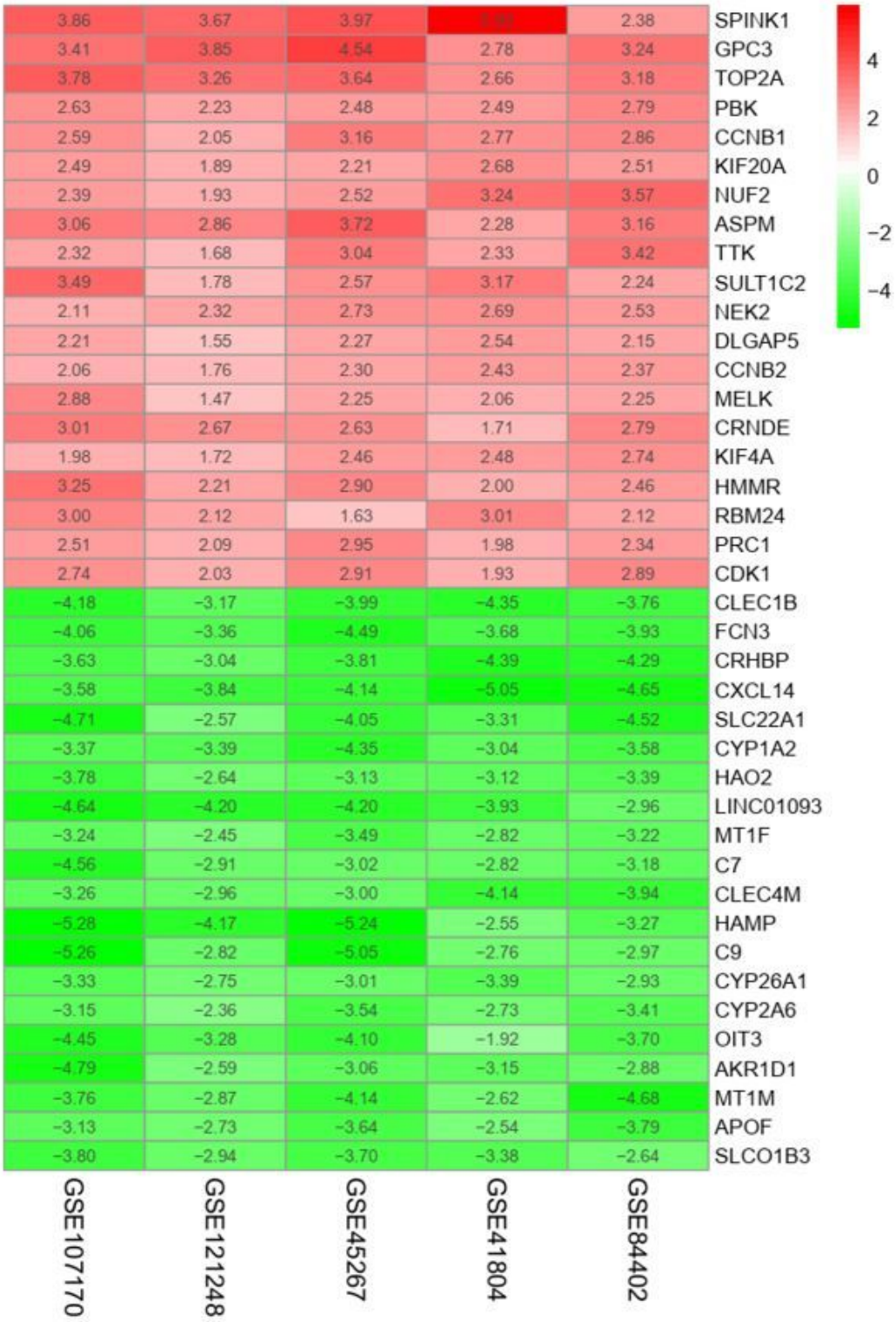

Figure 3

Identification of DEGs by RRA analysis. Top 20 upregulated and downregulated genes were presented by heatmap. The numbers represents logarithmic fold change in each dataset.

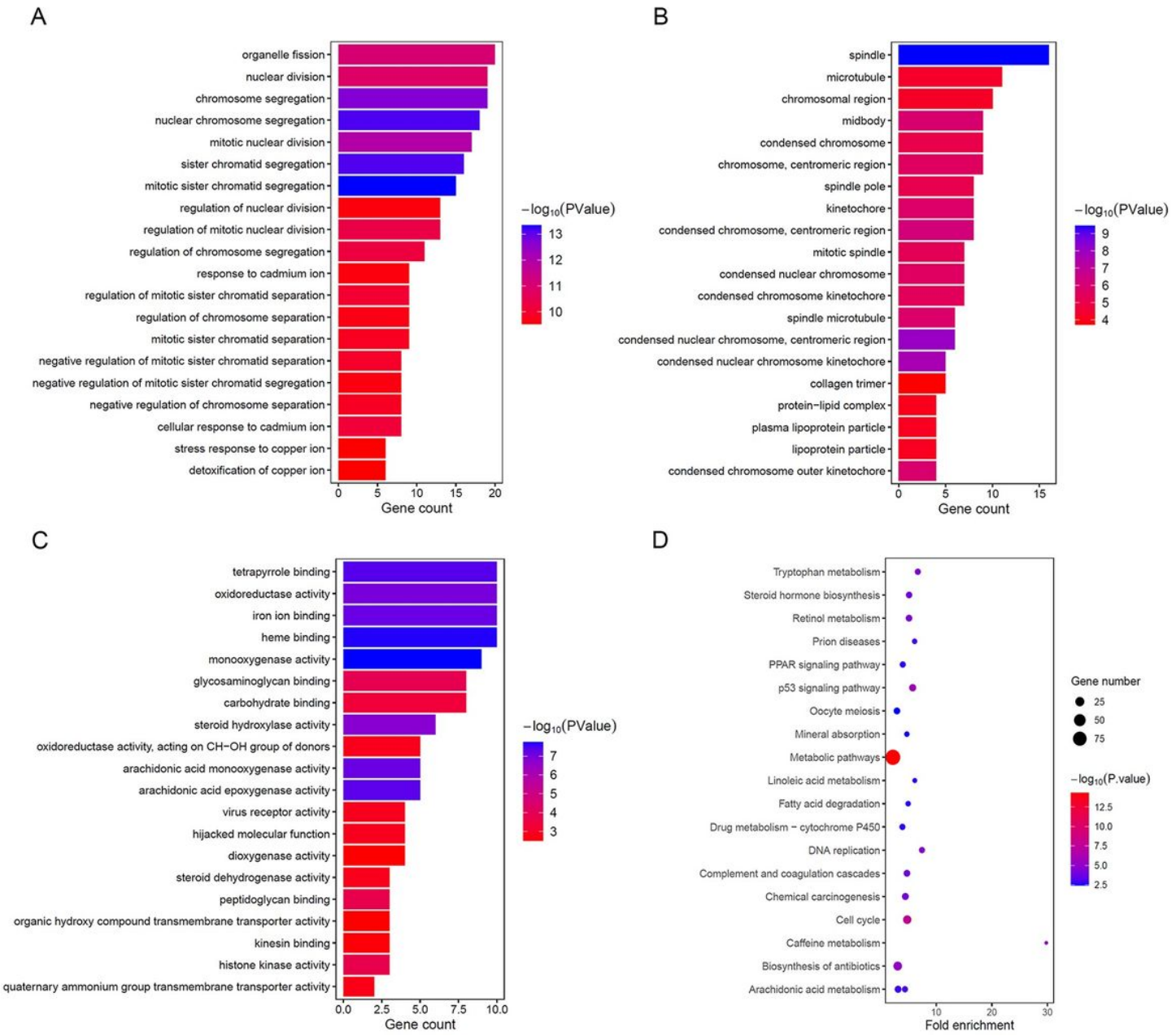

Figure 4

GO and KEGG analysis of DEGs. Top 20 enriched GO terms related with biologic process (A), cellular component (B), molecular function (C) for robust DEGs. (D) Top 20 enriched pathways for DEGs.

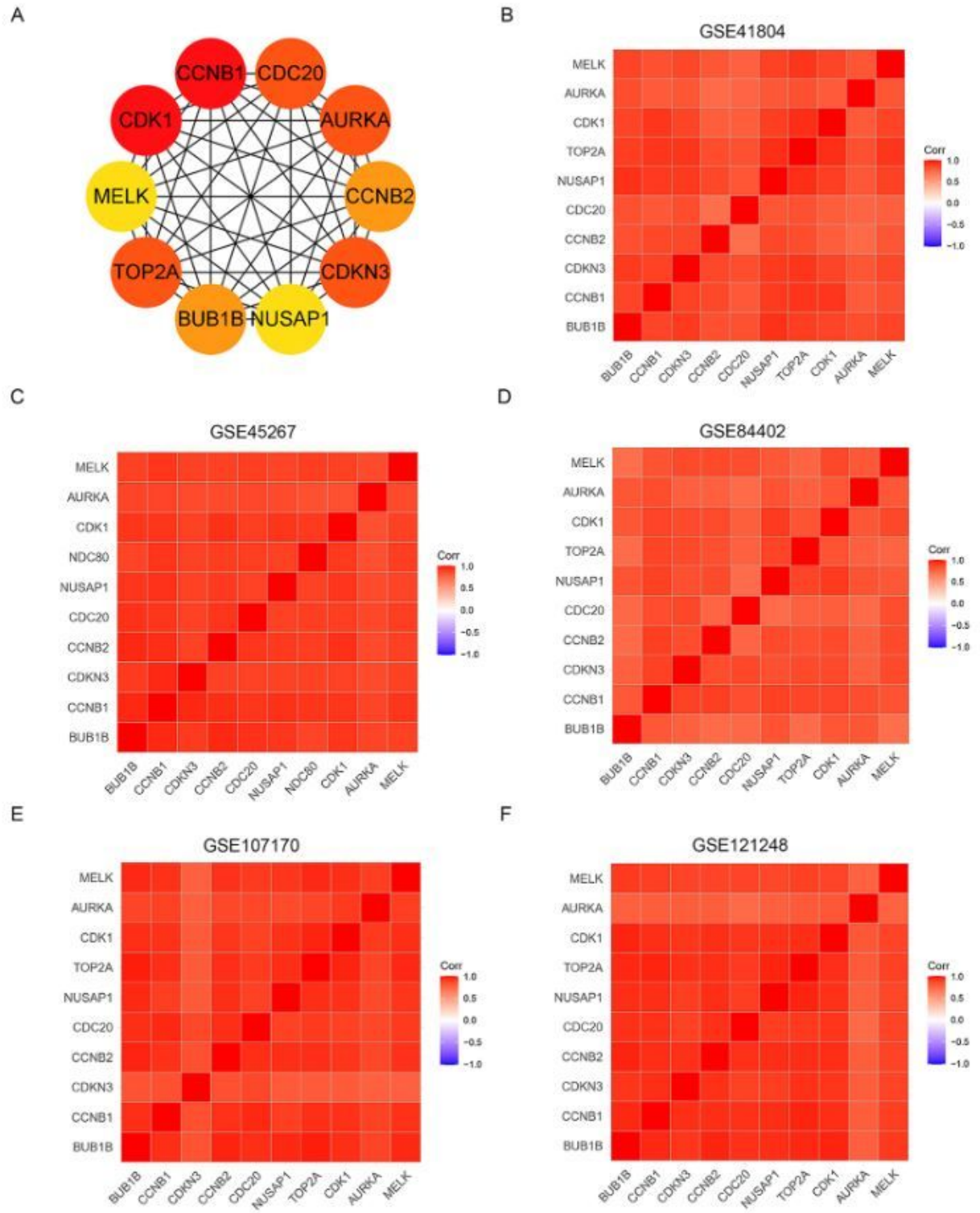

Figure 5

Hub genes network and correlation analysis. (A) Ten hub genes network by CytoHubba. Correlation analysis among hub genes in GSE41804 (B), GSE45267 (C), GSE84402 (D), GSE107170 (E), GSE121248 (F) was displayed.

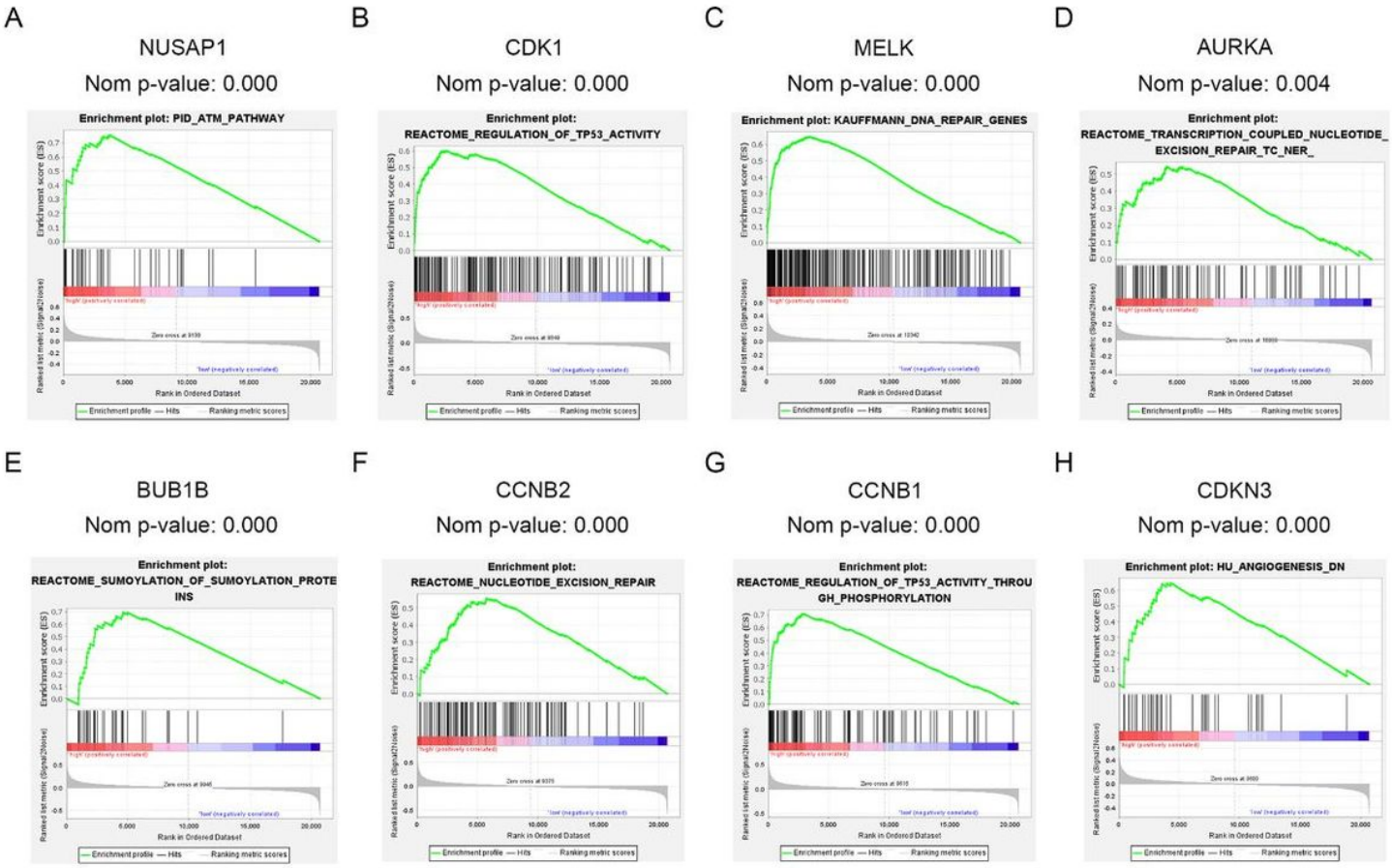

Figure 6

GSEA analysis based on GSE121248 data. Some representative enriched pathways of (A) NUSAP1, (B) CDK1, (C) MELK, (D) AURKA, (E) BUB1B, (F) CCNB2, (G) CCNB1, (H) CDKN3.

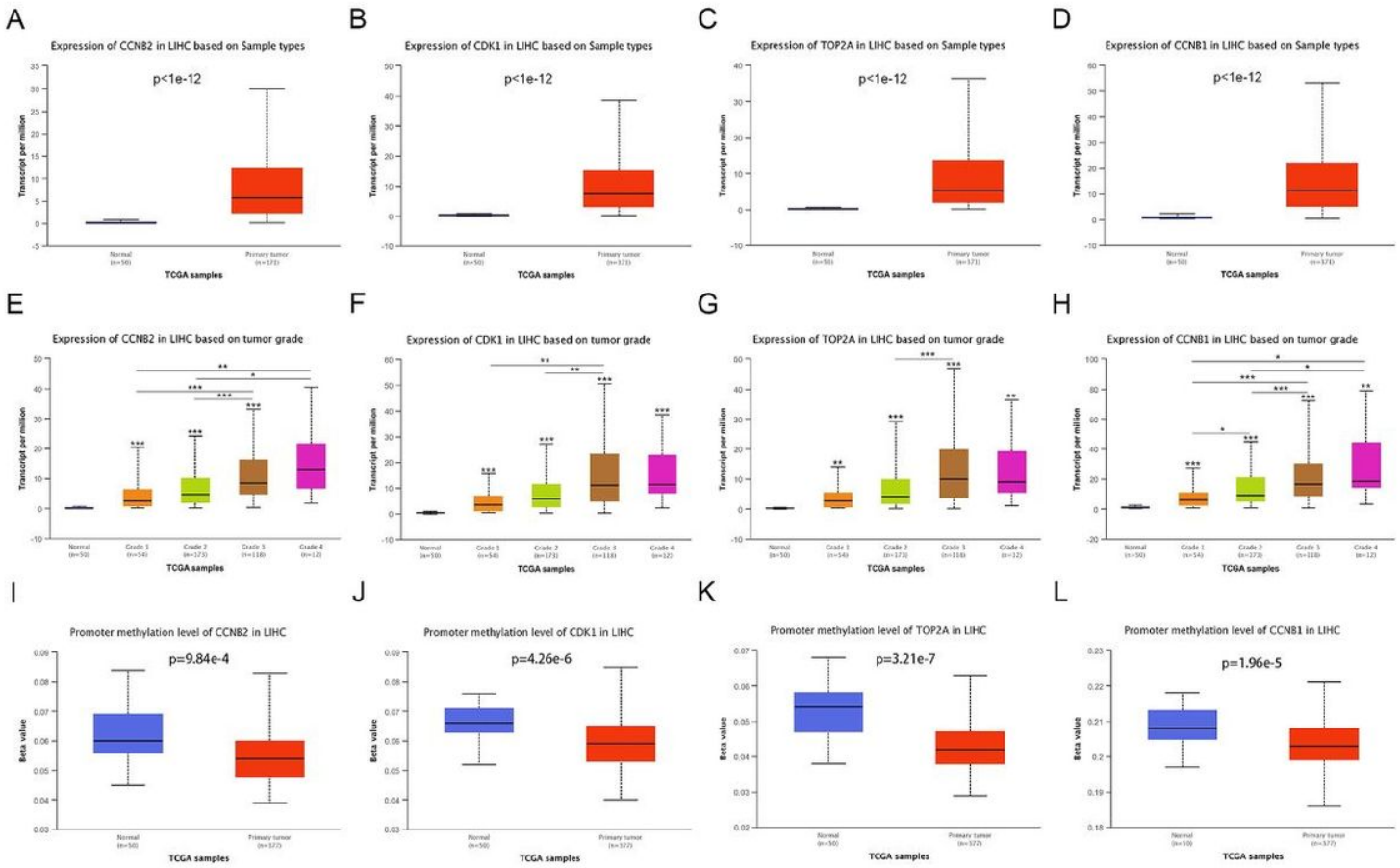

Figure 7

Expression validation and methylation analysis of hub genes. (A-D) Expression levels of hub genes in HCC. (E-H) Expression analysis of hub genes with tumor grades. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The first layer asterisk above the error bar represents comparisons to normal samples. (I-L) Promoter methylation levels of hub genes.

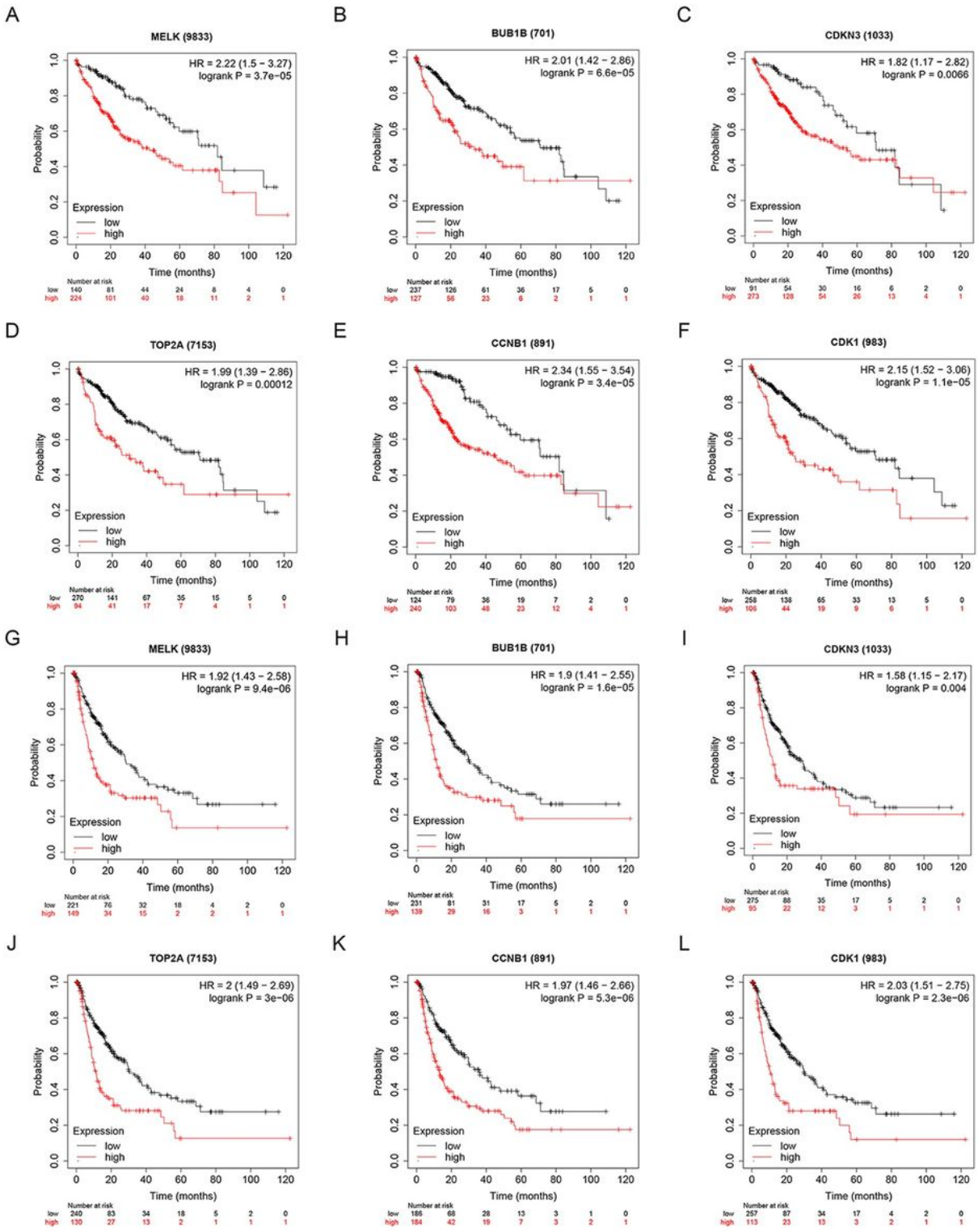

Figure 8

Survival analysis. (A-F) Overall survival analysis of hub genes. (G-L) Progression-free survival analysis of hub genes.

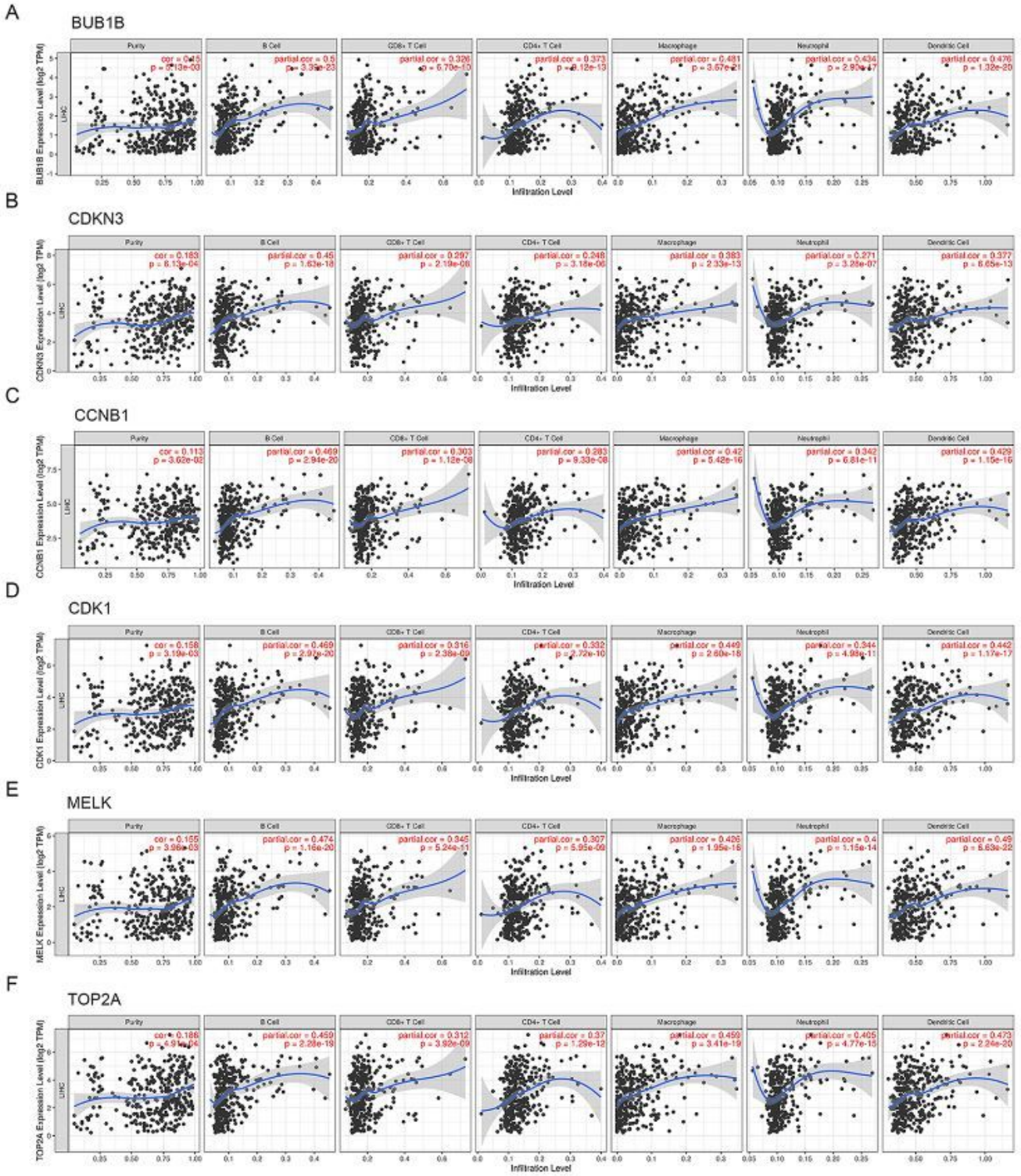

Figure 9

Association of hub genes' expressions with immune infiltration. The correlation between expression of (A) BUB1B, (B) CDKN3, (C) CCNB1, (D) CDK1, (E) MELK, (F) TOP2A and immune infiltration.

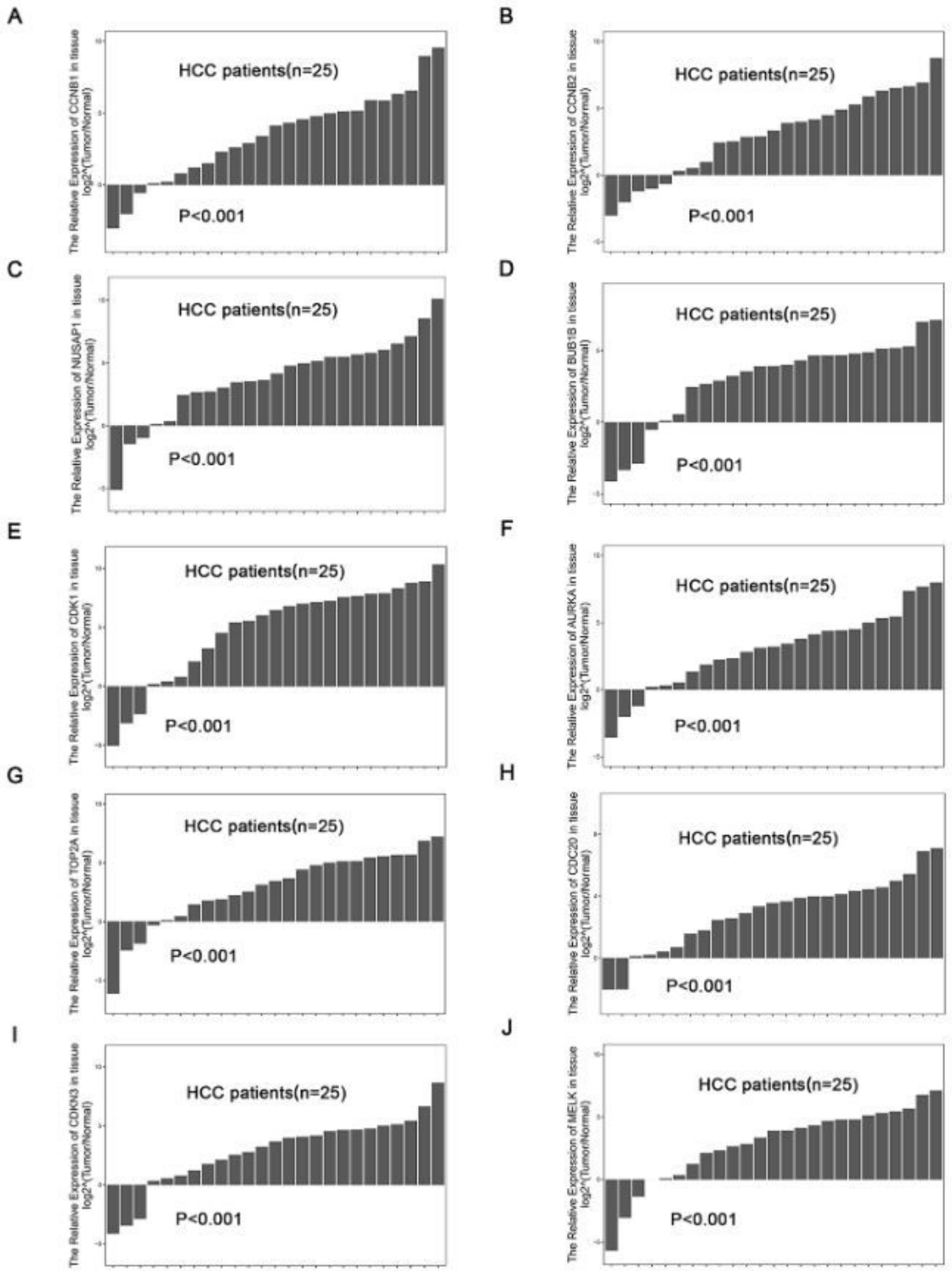

Figure 10

The mRNA expression of hub genes in 25 cases. QPCR analysis of mRNA levels of (A) CCNB1, (B) CCNB2, (C) NUSAP1, (D) BUB1B, (E) CDK1, (F) AURKA, (G) TOP2A, (H) CDC20, (I) CDKN3, (J) MELK.

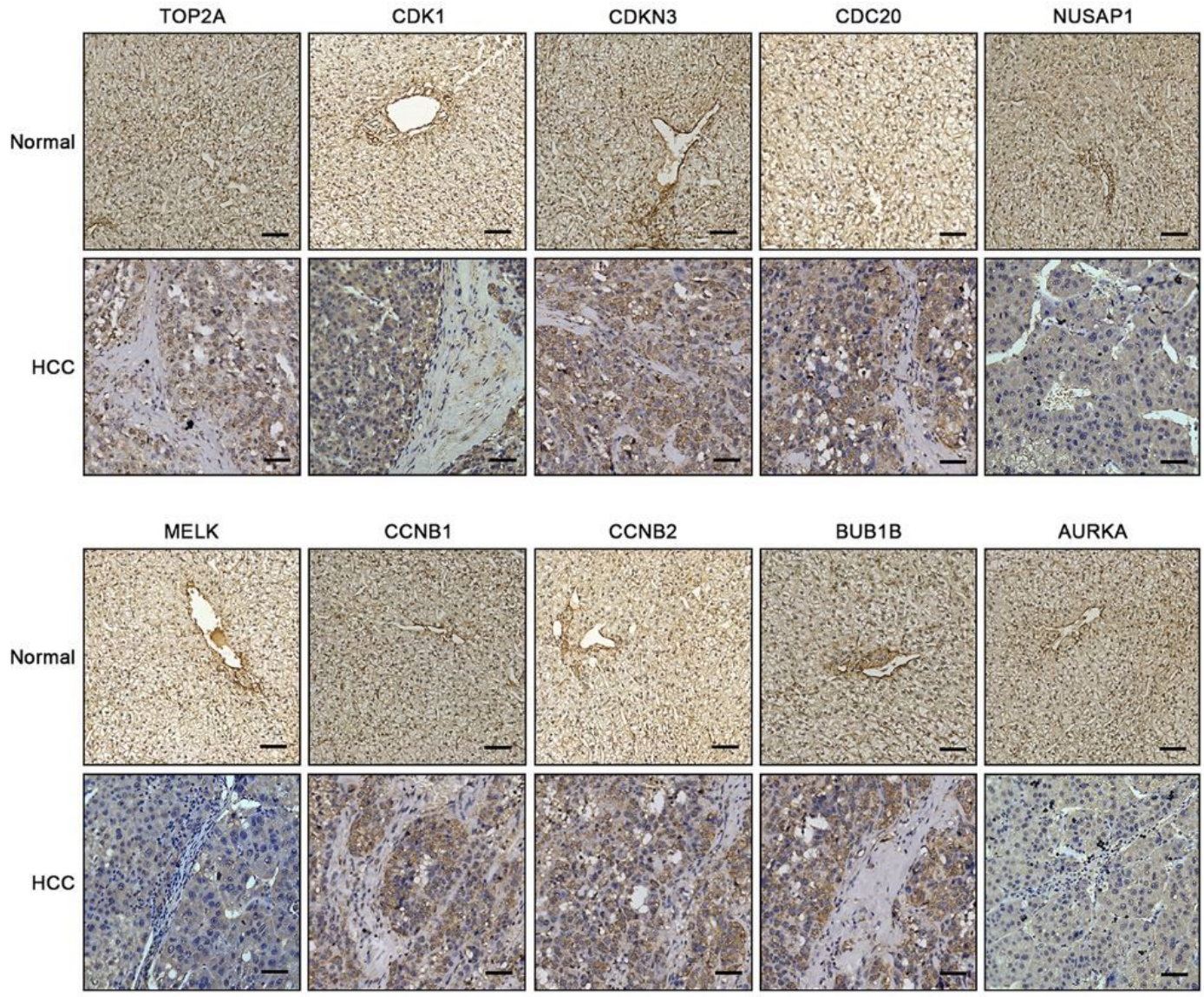

Figure 11

Immunohistochemistry of HCC tissues. Representative IHC images of hub genes in HCC samples. Scale bar = 50 μm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.tif
- Additionalfile2.tif
- Additionalfile3.tif
- Additionalfile4.tif
- Additionalfile5.tif

- Additionalfile6.docx
- Additionalfile7.docx
- Additionalfile8.docx
- Additionalfile9.docx
- Additionalfile10.docx
- Additionalfile11.docx
- Additionalfile12.docx
- Additionalfile13.docx
- Additionalfile14.docx
- Additionalfile15.docx
- Additionalfile16.docx