

Screening and Identification of Significant Genes in Hepatocellular Carcinoma via Bioinformatical Analysis and Validation in Clinical Tissue

jun guo zhou

North Sichuan Medical College [Search North Sichuan Medical College]: North Sichuan Medical University

Dongsheng Wang

Laboratory department of Sichuan Tumor Hospital

Xiaorong Zhong

North Sichuan Medical College [Search North Sichuan Medical College]: North Sichuan Medical University

Wei Ying

north sichuan medical college

Yanchao Feng

North Sichuan Medical College [Search North Sichuan Medical College]: North Sichuan Medical University

Wenbo Li

North Sichuan Medical College [Search North Sichuan Medical College]: North Sichuan Medical University

Lin Shi

north sichuan medical college

Lizheng Huang

north sichuan medical college

Jianshui Li

Affiliated Hospital of North Sichuan Medical college

Lingmi Hou

Affiliated Hospital of North Sichuan Medical College

wei Zheng Leng (✉ lengzhengwei@163.com)

Department of Hepatobiliary Surgery ☒, Affiliated Hospital of North Sichuan Medical College, Nanchong, China.

Research

Keywords: Carcinoma, Hepatocellular, Bioinformatics, CDK1, RRM2

Posted Date: November 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-99437/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: To screen out significant genes associated with occurrence and development of hepatocellular carcinoma (HCC) via bioinformatical analysis and validation using clinical specimens.

Methods: Gene expression chips were obtained from GEO database, differentially expressed genes (DEGs) between HCC and para-cancerous tissues were identified by GEO2R and Venn diagrams. What's more, Gene Ontology (GO) function analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs were carried out by DAVID. The protein-protein interaction (PPI) network and module analysis of DEGs were performed by STRING and Cytoscape to get hub genes. Subsequently, the influence of hub genes on overall survival and the expression levels were determined with Ualcan and GEPIA, and found the pathway via re-analysis of DAVID. Besides, immunohistochemistry staining were performed to verify the key genes, and follow-up results including prognoses and the clinicopathological features were statistically analyzed.

Results: 49 up-regulated DEGs and 122 down-regulated DEGs were selected. There were to tally 33 Hub genes were screened out, while 28 of them were related to prognosis and high expression in HCC. Furthermore, CDK1 and RRM2 were significantly enriched in p53 signaling pathway. Meanwhile, CDK1, RRM2 were highly expressed in HCC tissues by immunohistochemistry staining. Additionally, CDK1 and RRM2 were negatively correlated with overall survival. Tumor size and AFP were significant prognostic factors, while CDK1 and RRM2 were independent prognostic factors.

Conclusion: This study confirmed CDK1 and RRM2 as the key genes in HCC which could be potential biological target for diagnosis and treatment.

Background

HCC (hepatocellular carcinoma) is the most common malignant tumor of digestive system featured with high mortality and morbidity(1). At present, a lot progress has been achieved with the usual treatment strategies, such as surgical resection, transplantation, chemotherapy, radio frequency ablation and arterial embolization. Unfortunately, these strategies would often bring unfavorable side effects resulting poor prognosis of HCC. Targeted medicine therapy has an essential effect on prognosis that it provides a supplement and innovative option to traditional treatments(2). However, resistance of targeted drugs has been gradually increasing with emergence of target mutations and escapes. In addition, the mechanism of HCC occurrence and development has not been fully elucidated based on current researches, thus discovery of new targets would be extremely difficult and important(3) as it plays an important role in early diagnosis and personalized precision treatment of HCC(4).

Gene profile, as an efficient and reliable technology of gene detection, has been widely used that large amount of gene detection data has been accumulated, stored and shared in many public databases for further analyses and mining leading to the emergence and development of bioinformatics(5). Especially in recent years, a wealth of data have been yielded from accumulating researches on the genes for liver cancer. However, due to heterogeneity of samples, the existing data is usually biased and mixed and couldn't comprehensively and objectively reflect the actual gene expression during independent gene sequencing research. Besides, evidences for direct and reliable gene verification are insufficient in most of current studies. Therefore, joint analysis of multiple samples and verification using clinical samples is considered as an effective way to find new target genes(6,7).

In this study, we analyzed four original microarray datasets and included 336 HCC cases and 72 para-cancerous cases data. Key genes were selected and verified in clinical samples by immunohistochemistry. The obtained key genes enriched in occurrence and development of HCC potentially provide a theoretical basis for researches on occurrence and development mechanism of HCC, and identifying of new HCC targets.

Materials And Methods

2.1 Materials and patients:

CDK1 and RRM2 Antibodies purchased from SAB company, China; Immunohistochemical kit and DBA color developing solution purchased from BOSTER Biological Technology co.ltd, China. 71 cases of primary HCC and matched para-cancer tissues were collected between 2013 and 2015 from the Affiliated Hospital of North Sichuan Medical College. The study was approved by the Ethics Committee of Affiliated Hospital of North Sichuan Medical College.

2.2 Microarray data information:

We selected gene expression profiles GSE101685, GSE112791, GSE62232, GSE45267 from GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) from NCBI-GEO, is a free public gene database(<http://www.ncbi.nlm.nih.gov/geo>); Organism: Homo sapiens; Experiment type: Expression profiling by array. The microarray contains 336 HCC cases and 72 cases of para-cancerous tissue, and was published in 2019 or 2014, including Asian and European-American (Table1).

2.3 Data processing and screening of DEGs GEO2R(<http://www.ncbi.nlm.nih.gov/geo/geo2r>) is an important online tool in NCBI-GEO for comparing two or more datasets to identify DEGs. We compared cancer tissue and para-cancerous tissue DEGs in gene profile datasets GSE101685, GSE112791, GSE62232 and GSE45267 respectively by GEO2R. Defined those with $\log_{2}FC$ (fold change) >2 and adj. P-value < 0.05 as Up-regulated DEGs and those with $\log_{2}FC$ (fold change) <-2 and adj. P-value < 0.05 as Down-regulated DEGs. Then Venn online software was used to screen the co-expressed DEGs in the four profiles.

2.4 Gene ontology and KEGG pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID,<http://david.ncifcrf.gov>) (version 6.8) (8) is an integrated biological knowledgebase and analytic tool aimed at systematically extracting biological meaning from large gene/protein lists. Gene ontology(9) is a comprehensive resource of computable knowledge regarding the functions of genes and gene products. KEGG(10) is a reference knowledge base for biological interpretation of genome sequences and other high-throughput data. In this study, we selected and analyzed the biological processes (BP), molecular function (MF), cell component (CC) and KEGG pathway enrichment of DEGs in David website. P-Value <0.05 was considered statistically significant.

2.5 PPI network and module analysis

STRING(<http://string-db.org>) (version 11.0)(11) online database aims to collect, score and integrate all publicly available sources of protein-protein interaction information, and to complement these with computational predictions. Cytoscape(12) is network biology analysis and visualization tools. MCODE is an APP of Cytoscape for

finding densely connected regions from PPI networks. In this study, PPI network of DEGs was drawn by STRING and Cytoscape, the most significant hub module in the PPI networks was identified by MCODE.

2.6 Survival analysis and RNA sequencing expression of Hub genes:

UALCAN(13) is an interactive web-portal and user-friendly tool to perform in-depth analyses of TCGA gene expression data. GEPIA (Gene Expression Profiling Interactive Analysis)(<http://gepia.cancer-pku.cn>)(14) web server is a valuable resource for gene expression analysis based on tumor and normal samples from the TCGA and the GTEx databases. We used UALCAN to obtain the survival analysis results of hub genes. And the RNA sequencing expressions of Hub genes in cancer tissues and paracarcinoma tissue were analyzed with GEPIA. P-Value<0.05 was considered statistically significant.

2.7 Re-analysis core genes via KEGG pathway enrichment:

Hub genes associated with poor prognosis and high expressions in cancer tissues were selected and named as core genes, which were identified by KEGG pathway enrichment analysis with DAVID.

2.8 Immunohistochemical analysis

Immunohistochemical(IHC) staining was used to detect the expression and localization of CDK1 and RRM2 in 71 HCC tissues and matched para-cancer tissues. Then, to follow up the 5-year overall survival rate, and analysis of the correlation between age, gender, tumor size, AFP, AST, ALT, etc. and key genes.

Paraffin section was deparaffined in xylene for 30 min three times at 37°C, hydrated in a series of 100, 95, 90, 85 and 80% ethanol solutions, oxidize in 3% hydrogen peroxide solution for 30 min and washed in phosphate-buffered saline (PBS) for 2 min six times. The antigen was recovered in boiling sodium citrate buffer (10mmol/L, pH=6.0) for 5 min. Then the sections were cooled down to room temperature, BSA blocks unwanted antigenic sites for 30 min at 37°C, incubated with anti-CDK1 and anti-RRM2 monoclonal antibody (dilution 1:100) overnight at 4°C. After washing with PBS, the slides were incubated with SABC for 30 min at 37°C, DAB staining solution was used to dye the slides for 5 min and hematoxylin for counterstaining the nuclei for 1 min. The sections were then dehydrated in ethanol, coverslips were placed on the slides. Evaluation of immunohistochemical staining: staining included the intensity of staining (scored as: 0, no staining; 1, weak staining; 2, moderate staining; and 3, strong staining) and the percentage of positive tumor cells (scored as: 0, <5%; 1, 5-25%; 2, 26-50%; 3, 51-75%; and 4, 76-100%). Staining intensity and frequency were transformed into a Composite Expression Score (CES) (The formula: CES = Intensity×Frequency). The range of CES was from 0 to 12, according to the score definition that: 0-6, low expression; 7-12 high expression. These scores were independently determined by three senior pathologists. Images of sections were captured using an Olympus (100X-200X) microscope (Olympus, Tokyo, Japan).

2.9 Statistical analysis:

Data and figures were processed via GraphPad Prism7.0 software and IBM SPSS25. Comparisons of CDK1, RRM2 gene expression were performed with independent samples t-test, or using one-way of ANOVA and Bonferroni's multiple comparison tests. Correlation between the CDK1, RRM2 expression and clinicopathological features was detected with chi-square test. Kaplan-Meier and Cox regression model analyze are used to analysis the associated overall survival and clinicopathological factors. p-value< 0.05 was considered statistically significant.

Results

3.1 Identification of DEGs in HCC:

Total of 171 DEGs were screened out from GSE101685, GSE112791, GSE62232 and GSE45267 gene chips by online GEO2R tool and Venn software, which including 49 up-regulated genes and 122 down-regulated (Figure 1A).

3.2 DEGs gene ontology and KEGG pathway enrichment

We list the top five GO enrichment in P-Value order for the 171 DEGs identified by DAVID software. Table 2 demonstrates the results of GO enrichment analysis (P-Value<0.05, FDR<0.05, Table 2). The results of GO enrichment analysis includes biological processes(BP), molecular function (MF), cell component (CC) .The biological processes(BP) of up-regulated DEGs were enriched in mitotic nuclear division, cell division, mitotic sister chromatid segregation, sister chromatid cohesion, G2/M transition of mitotic cell cycle and down-regulated DEGs in epoxygenase P450 pathway, oxidation-reduction process, exogenous drug catabolic process, drug metabolic process, xenobiotic metabolic process; The molecular function (MF) of up-regulated DEGs were enriched in protein kinase binding, protein binding and down-regulated DEGs in oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, oxygen binding, heme binding, iron ion binding, monooxygenase activity; The cell component(CC) of up-regulated DEGs were enriched in midbody, condensed chromosome kinetochores, nucleus, kinetochore, cytoplasm and down-regulated DEGs in organelle membrane, extracellular region, endoplasmic reticulum membrane, blood microparticle.

KEGG pathway were enriched in Chemical carcinogenesis, Retinol metabolism, Drug metabolism - cytochrome P450, Linoleic acid metabolism (Including up-regulated genes and down-regulated genes).

3.3 PPI network and modular analysis:

We performed PPI diagram of 171 DEGs by online STRING database and Cytoscape. 132 out of 171 DEGs were identified from PPI, including 44 up-regulated genes and 88 down-regulated genes. And there were 132 nodes and 775 edges(Figure 1B). Then we use MCODE to screen out the hub modular from PPI, there were 33 nodes and 505 edges(MCODE scores=31.562, Degree Cutoff=2, Node Score Cutoff=0.2, k-Core=2 and Max. Depth=100.), and there are 33 hub genes(Figure 1B).

3.4 Survival analysis and RNA sequencing expression of hub genes:

29 genes were found to be able to cause poor overall survival(OS) in HCC patients(P<0.05) among the 33 hub genes uploaded to UALCAN website. Subsequently, 28 out of 29 genes were proved by GEPIA to have significant higher expression levels in HCC tissues than that in adjacent tissues (P<0.05)(|Log2FC| Cutoff=1; p-value Cutoff=0.01; Jitter Size=0.4)(Table 3).

3.5 Re-analysis of 28 core genes via KEGG pathway enrichment:

KEGG pathway enrichment analysis performed on the 28 key genes that related to overall survival and high expressions in HCC tissues. Fortunately, four enriched signal pathways and seven key genes were identified. Five genes (CCNB1, CDK1, AD2L1, BUB1B, CCNA2) markedly enriched in the cell cycle pathway; Four genes (CCNB1, CDK1, AD2L1, CCNA2) markedly enriched in the Progesterone-mediated oocyte maturation pathway; Four genes (CCNB1, CDK1, AD2L1, AURKA) markedly enriched in the Oocyte meiosis pathway; Three genes (CCNB1, CDK1, RRM2) markedly enriched in the p53 signaling pathway.

3.6 Immunohistochemical(IHC) staining and Prognostic analysis:

We found that CDK1 and RRM2 were expressed in the nucleus and cytoplasm, and much more highly expressed in HCC tissues than in para-cancer tissues($p < 0.05$)(Figure 2A). Expectedly, patients have poor overall survival (OS) with high expression of CDK1 or RRM2(Figure 2B)($p < 0.05$).

3.8 Clinicopathological features of HCC patients:

Our analysis showed that high expression of CDK1 was strongly correlated with tumor size and AFP ($p = 0.025, 0.020$), while RRM2 with age, tumor size and AFP ($p = 0.008, 0.005, 0.002$). But either of which were not correlated with gender, HbsAg, cirrhosis, drinking, AST or ALT (Table 4). Kaplan-Meier and Cox regression model analyze showed that CDK1 expression, RRM2 expression, tumor size and AFP were significant prognostic factors($P < 0.05$) (Table 5). Furthermore, we found that CDK1 expression and RRM2 expression were independent prognostic factors in patients with HCC($P < 0.05$)(Table 5).

Discussion

Numerous studies have been conducted to reveal the causes and underlying mechanisms of HCC formation and progression. Unfortunately, the prognosis still remains poor due to strong invasion and metastasis. And the commonly adapted treatments could not bring satisfactory overall efficacy. Early diagnosis and molecular targeted therapy based on mechanism researches are expected to improve the therapeutic effect of HCC treatment. Gene chip analysis and bioinformatics are the important and widely used ways to screen molecular target on a comprehensive scale and in an efficient and accurate manner. . This study integrated four gene expression profiles from GEO, identified 171 common DEGs. CCNB1, CDK1, AD2L1, BUB1B, CCNA2, AURKA, and RRM2 key genes were identified with a series of bioinformatics analyses to be associated with cell cycle, progesterone-mediated oocyte maturation, Oocyte meiosis and p53 signaling pathway. Moreover, we found higher expressions of CDK1 and RRM2 in HCC tissues by examining 71 paired clinical tissue samples using immunohistochemistry assay. Subsequently, CDK1 expression and RRM2 expression were identified as independent prognostic factors in HCC by Cox regression model analyze($P < 0.05$).

Interestingly, all the seven key genes screened out were constantly up-regulated in HCC and para-cancer tissues, which indicates that partial of up-regulated genes in HCC cells are responsible for occurrence and development of HCC. Accumulated researches have recognized the main factors for HCC including gene mutation, environment changes, abnormal regulation, and etc. And a lot of research indicated that HCC is caused by overexpression of proto-oncogenes(15,16). Thus it is hypothesized that HCC may promote abnormal expressions of certain oncogenes to cause production of cancerous cells. So, the proto-oncogenes should be the focus for studying on cancer prevention and treatment, especially the newly revealed 7 key genes.

Cyclin-dependent kinase 1(CDK1) was enriched in the above four pathways and governed the phosphorylation of factor 4E-binding protein 1 (4E-BP1), which is important in tumorigenesis(17). In addition, studies have shown that CDK1/EMT process promotes tumor progression via CDK1/PDK1 / β -Catenin pathways(18) and regulates tumorigenesis in tumor stem cells of liver cancer. CDK1/CCNB1 also regulating the apoptosis, invasion and cell cycle of HCC cells by blocking the p53 signaling pathway via Modulating CDK1/CCNB1 Axis(19). Therefore, CDK1 is closely associated with occurrence and development of HCC. In our study, the results of immunohistochemical staining showed that CDK1 expression in HCC was significantly increased. Moreover, results of both Kaplan-Meier

and Cox regression model analyze demonstrated that CDK1 was an independent unfavorable predictor of OS in HCC patients. These results made us convinced of the reliability of CDK1 as a biological target of HCC, and proved that the bioinformatics analysis method in this study could accurately screen out cancer-related key genes.

Ribonucleotide reductase M2 (RRM2), a critical rate-limiting enzyme for DNA synthesis and repair, plays an important role in cell division, proliferation and differentiation by promoting cells proliferation and inhibiting cellular apoptosis. Host RRM2 and NS5B can inhibit RNA replication of hepatitis B virus, thereby inhibiting the conversion of hepatitis B to liver cancer(20). RRM2 may be a key gene for liver cancer transformation from cirrhosis via P53 pathway(21), Gao(22) suggested in his study that RRM2 was a candidate target for HCC therapy as suppression of RRM2 markedly suppressed proliferation of HCC cells and RRM2 expression was higher in HCC than in non-HCC tissue. Wang Y(23) reported that SCYL1- BP1 could affect the cell cycle through increasing steady state levels of RRM2 proteins. And in our study, we confirmed the RRM2 was highly expressed in HCC tissues through immunohistochemical staining and could be an independent prognostic factors in HCC patients as CDK1.

CDK1 and RRM2 were identified to be core genes in colorectal cancer by Ding X(24) in their recent study using integrated bioinformatics analysis, which is constant with the results obtained by Ding YG(25) in anaplastic thyroid cancer, and by Zhou Y(26) in Glioblastoma. In addition, CDK1 and RRM2 were proved by Zhou Z(27) and Wu M(28) as key genes in HCC and to be potential new biological targets by bioinformatics in HCC. However, laboratory verification and further mechanism research of CDK1 and RRM2 were not performed in all the mentioned studies based on bioinformatics,. What's more, both CDK1 and RRM2 were enriched in the P53 signaling pathway that significantly contributing to reproduction, suppression of tumor expression and cell cycle regulation(29,30). The research on targeted therapy of P53 signaling pathway has a considerable future(31,32). Jin J(33) proved that LINC00346 could regulate apoptosis, invasion and cell cycle of HCC cells by promoting CDK1/CCNB1 expressions via indirectly blocking p53 signaling pathway. So far, the effect of gene therapy targeting RRM2 has been proven in many different types of tumors(26,34). However, there are only few extensive researches on its effect in HCC and RRM2 as a feasible therapeutic option has not been fully explored and recognized.

Notably, both CDK1 and RRM2 enriched in P53 signaling pathway. In our study, CDK1 and RRM2 were highly expressed in HCC as expected, and they were also identified as independent prognostic factors. Therefore, HCC diagnosis and treatment could possibly be improved by targeting both CDK1 and RRM2 although the underlying molecular mechanism linking CDK1 and RRM2 is still unclear and needs further research.

Our results are in line with previous findings and suggest that CDK1 and RRM2 can be key genes associated with HCC development. CDK1 and RRM2 combination is expected to become a reliable new biological target for diagnosis and gene therapy. More importantly, bioinformatics is further proved to be an advanced and useful method for screening key genes in cancers.

There were few limitations in the present study that should be noted. First, all the candidate genes could not be verified. Second, sample size was relatively small. Finally, molecular mechanisms should have been explored with cell experiments. So that further researches will be necessary to clarify the above issues.

Conclusions

In summary, our present study confirmed that CDK1 and RRM2 as the key genes in the development and progression of HCC. And, CDK1 and RRM2 might be utilized as prognostic indicators of HCC, providing potential new biological target for diagnosis and treatment in HCC.

Abbreviations

BP: Biological process; MF: Molecular function; CC: Cellular component; DAVID: Database for Annotation, Visualization and Integrated Discovery; DEGs: Differentially expressed genes; GEO: Gene Expression Omnibus; GEPIA: Gene Expression Profiling Interactive Analysis; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Gene and Genome; MCODE: Molecular Complex Detection; PPI: protein-protein interaction; STRING: Search Tool for the Retrieval of Interacting Genes; HCC: Hepatocellular carcinoma; IHC: immunohistochemical; CCNB1: cyclin B1; CDK1: Cyclin Dependent Kinase 1; MAD2L1: Mitotic Arrest Deficient 2 Like 1; BUB1B: BUB1 Mitotic Checkpoint Serine/Threonine Kinase B; CCNA2: Cyclin A2; AURKA: Aurora Kinase A; RRM2: Ribonucleotide Reductase Regulatory Subunit M2.

Declarations

Ethics approval and consent to participate:

The study was approved by the Ethics Committee of Affiliated Hospital of North Sichuan Medical College (2018EA030).

Consent for publication:

Not applicable.

Availability of data and materials:

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest:

The authors declare that they have no competing interests.

Funding:

This study was supported by the Foundation of the National Natural Science Foundation of China (No. 81402444), Health and Planning Committee of Sichuan province (No.18PJ005), the Foundation of the Department of Science and Technology of Sichuan province (Nos. 19YYJC0384, 2017JY0170).

Authors' contributions:

Guojun Zhou conducted the experiments and writing the manuscript. Dongsheng Wang, Xiaorong Zhong and Wei Ying collected and analyzed the data. Yanchao Feng, Wenbo Li, Lin Shi, Lizheng Huang, Jianshui Li significantly contributed to clinical data collection. Lingmi Hou and Zhengwei Leng supervised the project and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

We would like to acknowledge the contribution of Yi Qian for the language polishing of this article.

Corresponding Author's information:

Full name: Zhengwei Leng. Address: Fujiang Road 234#, Nanchong City, Sichuan Province, China. Email address: lengzhengwei@163.com. Telephone: 13980315571. & Full name: Lingmi Hou. Address: Wenhua Raod 63#, Nanchong City, Sichuan Province, China. Email address: houlingmi@163.com. Telephone: 13890706013.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: a cancer journal for clinicians*. 2020; 70(1):7-30.
2. Couri T, Pillai A. Goals and targets for personalized therapy for HCC. *Hepatol Int*. 2019; 13(2):125-137.
3. Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nature reviews. Clinical oncology*. 2018; 15(10):599-616.
4. Mu W, Jiang D, Mu S, Liang S, Liu Y, Zhang N. Promoting Early Diagnosis and Precise Therapy of Hepatocellular Carcinoma by Glypican-3-Targeted Synergistic Chemo-Photothermal Theranostics. *Acs Appl Mater Inter*. 2019; 11(26):23591-23604.
5. Tang B, Pan Z, Yin K, Khateeb A. Recent Advances of Deep Learning in Bioinformatics and Computational Biology. *Front Genet*. 2019; 10:214.
6. Parris TZ. Pan-cancer analyses of human nuclear receptors reveal transcriptome diversity and prognostic value across cancer types. *Sci Rep-Uk*. 2020; 10(1):1873.
7. Dhiman VK, Bolt MJ, White KP. Nuclear receptors in cancer - uncovering new and evolving roles through genomic analysis. *Nature reviews. Genetics*. 2018; 19(3):160-174.
8. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44-57.
9. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019; 47(D1):D330-D338.
10. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019; 47(D1):D590-D595.
11. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019; 47(D1):D607-D613.
12. Otasek D, Morris JH, Boucas J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol*. 2019; 20(1):185.
13. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia (New York, N.Y.)*. 2017; 19(8):649-658.
14. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*. 2019; 47(W1):W556-W560.
15. Muller M, Bird TG, Nault J. The landscape of gene mutations in cirrhosis and hepatocellular carcinoma. *J Hepatol*. 2020; 72(5):990-1002.

16. Cervello M, Emma MR, Augello G, Cusimano A, Giannitrapani L, Soresi M et al. New landscapes and horizons in hepatocellular carcinoma therapy. *Aging*. 2020; 12(3):3053-3094.
17. Velasquez C, Cheng E, Shuda M, Lee-Oesterreich PJ, Pogge von Strandmann L, Gritsenko MA, Jacobs JM, Moore PS, Chang Y. Mitotic protein kinase CDK1 phosphorylation of mRNA translation regulator 4E-BP1 Ser83 may contribute to cell transformation. *P Natl Acad Sci Usa*. 2016; 113(30):8466-71.
18. Wu CX, Wang XQ, Chok SH, Man K, Tsang SHY, Chan ACY, Ma KW, Xia W, Cheung TT. Blocking CDK1/PDK1/beta-Catenin signaling by CDK1 inhibitor RO3306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma. *Theranostics*. 2018; 8(14):3737-3750.
19. Jin J, Xu H, Li W, Xu X, Liu H, Wei F. LINC00346 Acts as a Competing Endogenous RNA Regulating Development of Hepatocellular Carcinoma via Modulating CDK1/CCNB1 Axis. *Frontiers in bioengineering and biotechnology*. 2020; 8:54.
20. Kitab B, Satoh M, Ohmori Y, Munakata T, Sudoh M, Kohara M, Tsukiyama-Kohara K. Ribonucleotide reductase M2 promotes RNA replication of hepatitis C virus by protecting NS5B protein from hPLIC1-dependent proteasomal degradation. *The Journal of biological chemistry*. 2019; 294(15):5759-5773.
21. Wu M, Liu Z, Li X, Zhang A, Lin D, Li N. Analysis of potential key genes in very early hepatocellular carcinoma. *World J Surg Oncol*. 2019; 17(1):77.
22. Gao J, Chen H, Yu Y, Song J, Song H, Su X et al. Inhibition of hepatocellular carcinoma growth using immunoliposomes forco-delivery of adriamycin and ribonucleotide reductase M2 siRNA. *Biomaterials*. 2013; 34(38):10084-98.
23. Wang Y, Zhi Q, Ye Q, Zhou C, Zhang L, Yan W et al. SCYL1-BP1 affects cell cycle arrest in human hepatocellular carcinoma cells via Cyclin F and RRM2. *Anti-Cancer Agent Me*. 2016; 16(4):440-6.
24. Ding X, Duan H, Luo H. Identification of Core Gene Expression Signature and Key Pathways in Colorectal Cancer. *Front Genet*. 2020; 11:45.
25. Ding Y, Ren Y, Xu Y, Wei C, Zhang Y, Zhang S, Guo C. Identification of key candidate genes and pathways in anaplastic thyroid cancer by bioinformatics analysis. *Am J Otolaryng*. 2020; 41(3):102434.
26. Zhou Y, Yang L, Zhang X, Chen R, Chen X, Tang W, Zhang M. Identification of Potential Biomarkers in Glioblastoma through Bioinformatic Analysis and Evaluating Their Prognostic Value. *Biomed Res Int*. 2019; 2019:6581576.
27. Zhou Z, Li Y, Hao H, Wang Y, Zhou Z, Wang Z, Chu X. Screening Hub Genes as Prognostic Biomarkers of Hepatocellular Carcinoma by Bioinformatics Analysis. *Cell Transplant*. 2019; 28(1):76S-86S.
28. Wu M, Liu Z, Zhang A, Li N. Identification of key genes and pathways in hepatocellular carcinoma: A preliminary bioinformatics analysis. *Medicine*. 2019; 98(5):e14287.
29. Khan H, Reale M, Ullah H, Sureda A, Tejada S, Wang Y, Zhang Z, Xiao J. Anti-cancer effects of polyphenols via targeting p53 signaling pathway: updates and future directions. *Biotechnol Adv*. 2020; 38:107385.
30. Raj N, Bam R. Reciprocal Crosstalk Between YAP1/Hippo Pathway and the p53 Family Proteins: Mechanisms and Outcomes in Cancer. *Frontiers in cell and developmental biology*. 2019; 7:159.
31. Zhu L, Dai L, Yang N, Liu M, Ma S, Li C et al. Transcription factor IRX5 promotes hepatocellular carcinoma proliferation and inhibits apoptosis by regulating the p53 signalling pathway. *Cell Biochem Funct*. 2020; 38(5):621-629.

32. Wu X, Li J, Gassa A, Buchner D, Alakus H, Dong Q et al. Circulating tumor DNA as an emerging liquid biopsy biomarker for early diagnosis and therapeutic monitoring in hepatocellular carcinoma. *Int J Biol Sci.* 2020; 16(9):1551-1562.
33. Jin J, Xu H, Li W, Xu X, Liu H, Wei F. LINC00346 Acts as a Competing Endogenous RNA Regulating Development of Hepatocellular Carcinoma via Modulating CDK1/CCNB1 Axis. *Frontiers in bioengineering and biotechnology.* 2020; 8:54.
34. Yang P, Lin L, Liu T. Sorafenib Inhibits Ribonucleotide Reductase Regulatory Subunit M2 (RRM2) in Hepatocellular Carcinoma Cells. *Biomolecules.* 2020; 10(1).

Tables

Table 1: Details of gene expression profile dataset GSE101685, GSE112791, GSE62232 and GSE45267.

DataSets	Platforms	Tumour	Normal	Country	Public time
GSE101685	GPL570	24	8	Taiwan	Jul 31, 2019
GSE112791	GPL570	183	15	Japan	Jan 03, 2019
GSE62232	GPL570	81	10	France	Dec 31, 2014
GSE45267	GPL570	48	39	Taiwan	Jul 31,2014

Table 2: Gene ontology analysis for DEGs

Expression	Category	Term	Count	%	P-Value	FDR
up-regulated	GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	12	25.00	3.81E-11	5.36E-08
	GOTERM_BP_DIRECT	GO:0051301~cell division	13	27.08	8.64E-11	1.22E-07
	GOTERM_BP_DIRECT	GO:0000070~mitotic sister chromatid segregation	5	10.42	5.46E-07	7.69E-04
	GOTERM_BP_DIRECT	GO:0007062~sister chromatid cohesion	6	12.50	7.92E-06	0.01
	GOTERM_BP_DIRECT	GO:0000086~G2/M transition of mitotic cell cycle	6	12.50	3.16E-05	0.04
	GOTERM_MF_DIRECT	GO:0019901~protein kinase binding	9	18.75	4.94E-06	0.006
	GOTERM_MF_DIRECT	GO:0005515~protein binding	38	79.17	8.65E-06	0.0097
	GOTERM_CC_DIRECT	GO:0030496~midbody	9	18.75	8.76E-10	9.90E-07
	GOTERM_CC_DIRECT	GO:0000777~condensed chromosome kinetochores	7	14.583	6.97E-08	7.88E-05
	GOTERM_CC_DIRECT	GO:0005634~nucleu	30	62.5	1.48E-06	0.0017
	GOTERM_CC_DIRECT	GO:0000776~kinetochore	6	12.5	1.63E-06	0.0018
	GOTERM_CC_DIRECT	GO:0005737~cytoplasm	29	60.417	2.83E-06	0.0032
	down-regulated	GOTERM_BP_DIRECT	GO:0019373~epoxygenase P450 pathway	9	7.7586	6.84E-14
GOTERM_BP_DIRECT		GO:0055114~oxidation-reduction process	23	19.828	9.59E-12	1.46E-08
GOTERM_BP_DIRECT		GO:0042738~exogenous drug catabolic process	7	6.035	4.37E-11	6.68E-08
GOTERM_BP_DIRECT		GO:0017144~drug metabolic process	8	6.897	2.29E-10	3.49E-07
GOTERM_BP_DIRECT		GO:0006805~xenobiotic metabolic process	10	8.621	1.21E-09	1.84E-06
GOTERM_MF_DIRECT		GO:0016705~oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	13	11.207	7.61E-16	1.05E-12
GOTERM_MF_DIRECT		GO:0019825~oxygen binding	12	10.345	3.71E-15	4.98E-12
GOTERM_MF_DIRECT		GO:0020037~heme binding	16	13.793	4.49E-	6.04E-

					15	12
GOTERM_MF_DIRECT	GO:0005506~iron ion binding	16	13.793	2.35E-14	3.20E-11	
GOTERM_MF_DIRECT	GO:0004497~monooxygenase activity	12	10.345	4.53E-14	6.17E-11	
GOTERM_CC_DIRECT	GO:0031090~organelle membrane	15	12.931	8.89E-17	1.22E-13	
GOTERM_CC_DIRECT	GO:0005576~extracellular region	30	25.862	2.90E-08	3.29E-05	
GOTERM_CC_DIRECT	GO:0005789~endoplasmic reticulum membrane	18	15.517	1.07E-05	0.012	
GOTERM_CC_DIRECT	GO:0072562~blood microparticle	8	6.897	3.18E-05	0.036	

Note: BP: biological processes, MF: molecular function, CC: cell component, FDR: False Discovery Rate

Table 3: Overall survival analysis and expression of hub genes in HCC tissues

Category	Genes
Gene with significantly Overall survival (P<0.05)	ANLN ASPM AURKA BIRC5 BUB1B CCNA2 CCNB1 CDK1 CDKN3 DLGAP5 DTL ECT2 FAM83D GINS1 HELLS HMMR KIF20A MAD2L1 NCAPG NDC80 NEK2 PBK RACGAP1 RRM2 TOP2A TTK UBE2T UHRF1 ZWINT
Gene with high expressed in HCC tissues (P < 0.05)	ANLN ASPM AURKA BIRC5 BUB1B CCNA2 CCNB1 CDK1 CDKN3 DLGAP5 DTL ECT2 FAM83D GINS1 HELLS HMMR KIF20A MAD2L1 NCAPG NDC80 NEK2 PBK RACGAP1 RRM2 TOP2A UBE2T UHRF1 ZWINT

Table 4: Relationships among CDK1, RRM2 and clinicopathological parameters in patients with HCC

Variables	CDK1 expression level				RRM2 expression level			
	High(n=60)	Low(n=11)	χ^2	P	High(n=49)	Low(n=22)	χ^2	P
Age			0	1			7.142	0.008*
<50	19	4			11	12		
≥50	41	7			38	10		
Gender			0	1			0.009	0.926
male	47	9			38	18		
female	13	2			11	4		
Tumor Size			5.022	0.025*			7.816	0.005*
<6cm	19	8			14	14		
≥6cm	41	3			35	8		
HbsAg			0.108	0.742			0.647	0.421
positive	43	9			34	18		
negative	17	2			15	4		
Cirrhosis			0.002	0.965			0.005	0.945
yes	40	8			33	15		
no	20	3			16	7		
Drinking			0.742	0.389			4.808	0.028
yes	28	3			4	7		
no	32	8			45	15		
AFP			5.443	0.020*			9.541	0.002*
<20	14	7			9	12		
≥20	46	4			40	10		
AST			2.091	0.148				
<40	22	1			9	6	0.723	0.395
≥40	38	10			40	16		
ALT			0.091	0.763			0.313	0.576
<50	28	4			21	11		
≥50	32	7			28	11		

Table 5: Kaplan-Meier and Cox regression model analyze the parameters associated to overall survival(OS) of HCC patents

Variables	OS (Kaplan-Meier)		OS (Cox regression model)		
	Log-rank χ^2	P	HR	95%CI	P
CDK1(low vs. high)	7.632	0.006*	3.128	1.058-9.250	0.039*
RRM2(low vs. high)	9.659	0.002*	2.148	1.010-4.725	0.047*
Age(<50 vs. ≥ 50)	2.785	0.095			
Gender(male vs. female)	0.001	0.970			
Tumor Size(≤ 6 cm vs. negative)	8.384	0.004*	0.709	0.364-1.382	0.313
HbsAg(positive vs. negative)	0.010	0.921			
Cirrhosis(yes vs. no)	0.907	0.635			
Drinking(yes vs. no)	0.155	0.925			
AFP(<20 vs. ≥ 20)	6.171	0.013*	1.192	0.589-2.411	0.625
AST(<40 vs. ≥ 40)	0.360	0.549			
ALT(<50 vs. ≥ 50)	0.295	0.587			

Note: * means $p < 0.05$

Figures

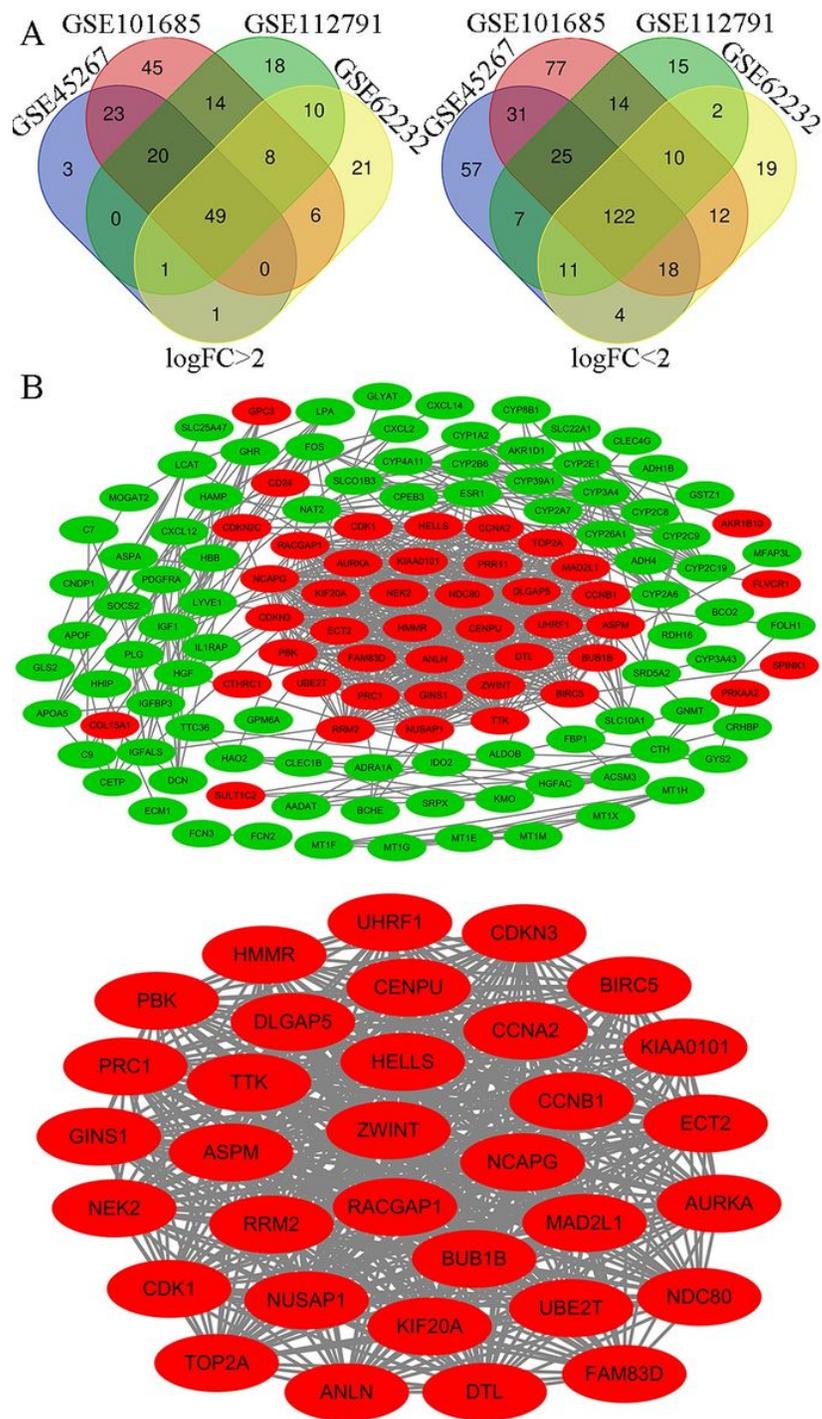


Figure 1

The key genes in hepatocellular carcinoma via bioinformatical analysis. (A): Commonly DEGs. $\log_{2}FC > 2$ as up-regulated DEGs, $\log_{2}FC < -2$ as down-regulated DEGs. (B): Protein-protein interaction network of commonly DEGs analysis by STRING and Cytoscape. The PPI network drawn with STRING and Cytoscape. 33 hub genes analyzed via MCODE. Note: the nodes meant proteins, the edges meant the interaction of proteins, green circles meant down-regulated DEGs and red circles meant up-regulated DEGs.

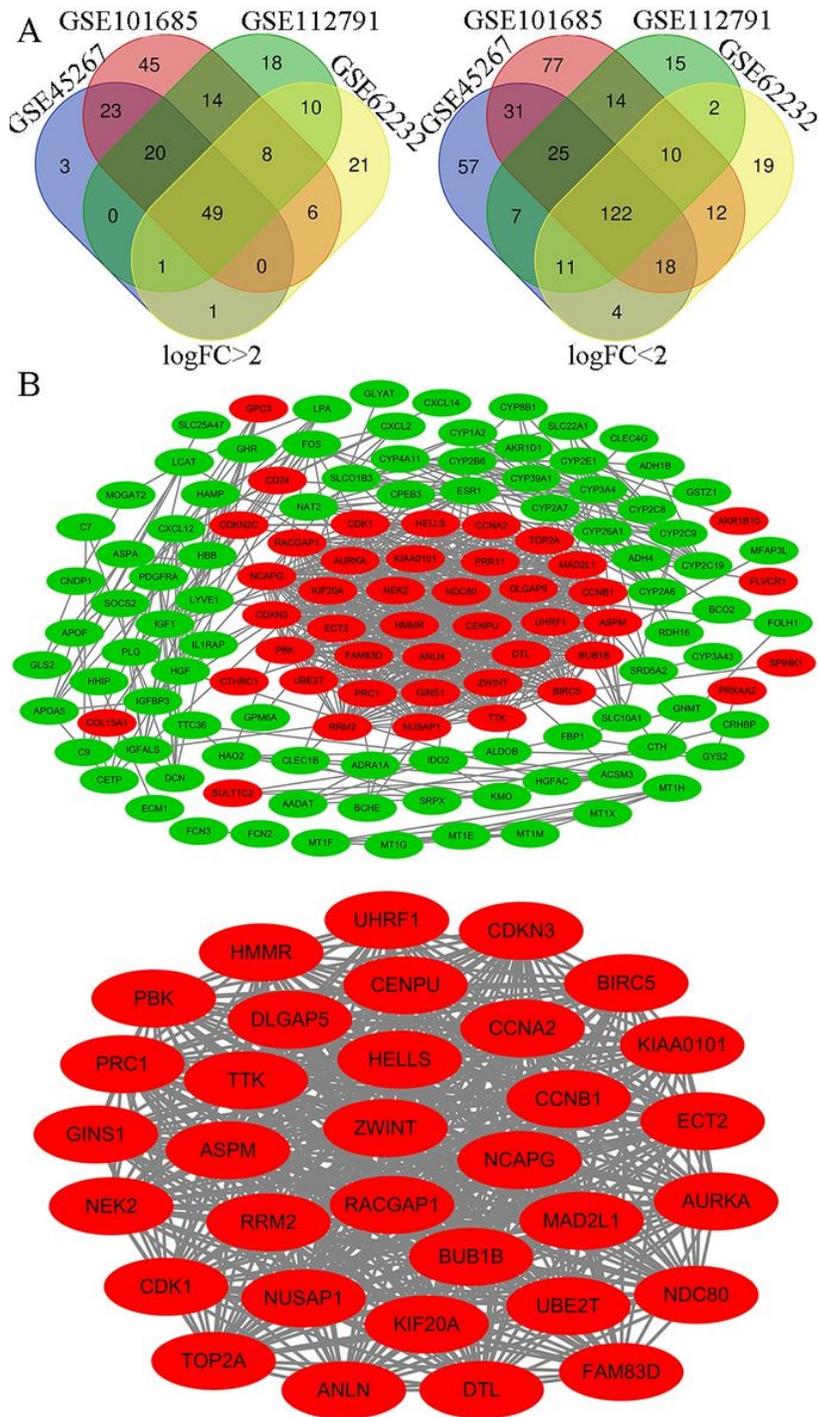
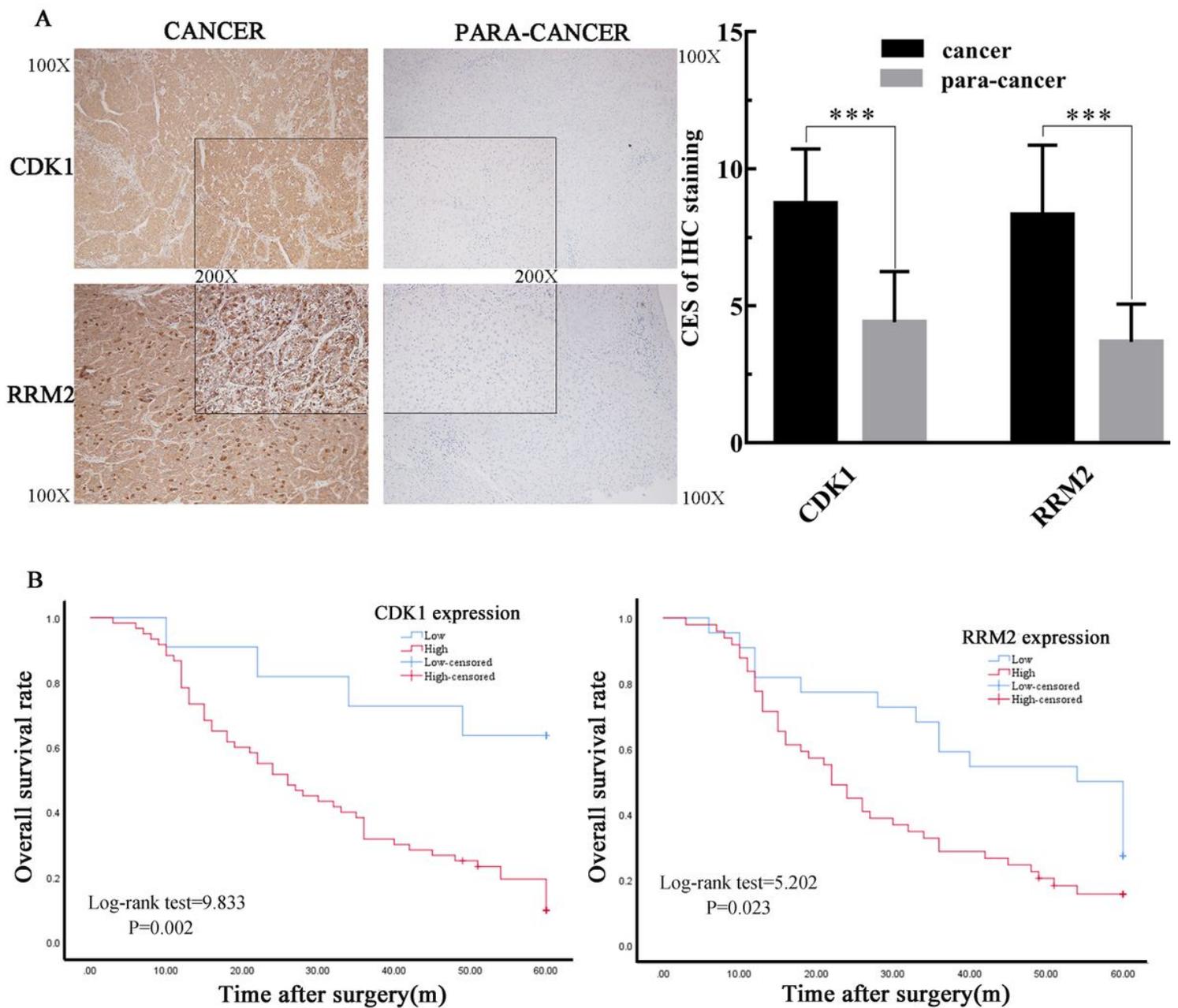


Figure 1

The key genes in hepatocellular carcinoma via bioinformatical analysis. (A): Commonly DEGs. $\log_{2}FC > 2$ as up-regulated DEGs, $\log_{2}FC < -2$ as down-regulated DEGs. (B): Protein-protein interaction network of commonly DEGs analysis by STRING and Cytoscape. The PPI network drawn with STRING and Cytoscape. 33 hub genes analyzed via MCODE. Note: the nodes meant proteins, the edges meant the interaction of proteins, green circles meant down-regulated DEGs and red circles meant up-regulated DEGs.



CDK1 expression

Log-rank test=9.833
P=0.002

RRM2 expression

Log-rank test=5.202
P=0.023

Figure 2

The expression and prognostic of CDK1 and RRM2 in hepatocellular carcinoma. (A): CDK1 and RRM2 are highly expressed in HCC tissues. (magnification, x100,x200). (B): Kaplan-Meier analysis of overall survival (OS) curves of HCC patients.

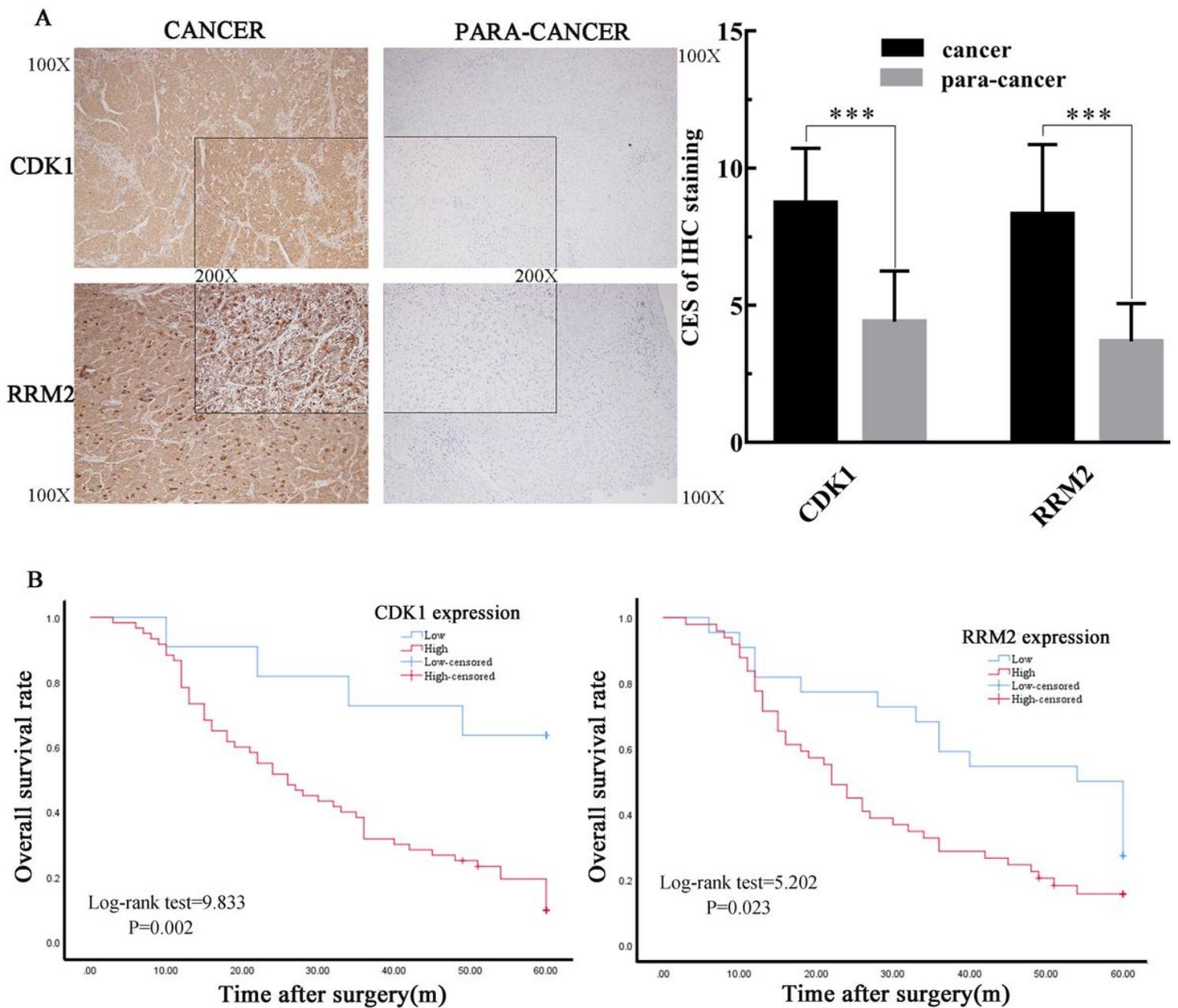


Figure 2

The expression and prognostic of CDK1 and RRM2 in hepatocellular carcinoma. (A): CDK1 and RRM2 are highly expressed in HCC tissues. (magnification, x100,x200). (B): Kaplan-Meier analysis of overall survival (OS) curves of HCC patients.