Revealing Survival-Associated Genes in HCC Through Comprehensive In-Silico Analyses

Kiruthika Kasthuri[1], Vijayalakshmi Kumaravel[1]*, Karthikaiselvi Lakshmanan[1]

[1]*Department of Biochemistry, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203*

***Corresponding Author: Vijayalakshmi Kumaravel**

Department of Biochemistry

Faculty of Science and Humanities

SRM Institute of Science and Technology

Kattankulathur - 603203

Tamil Nadu, India

ABSTRACT

Hepatocellular carcinoma (HCC) is the most common primary malignancy of the liver and is typically associated with chronic liver disease, including hepatitis infections and cirrhosis. Despite advances in treatment, HCC remains a major global health burden due to its late diagnosis, high recurrence rate, and poor overall prognosis. Targeted gene therapy in HCC aims to modulate specific oncogenic or tumor-suppressor pathways to halt tumor progression and enhance therapeutic efficacy. By selectively correcting dysregulated genes, this approach holds promise for improving treatment precision and patient survival outcomes. The study employed GEO2R for differential expression analysis, followed by hub gene selection and construction of a protein-protein interaction (PPI) network to identify key regulatory genes. Gene ontology and KEGG enrichment analyses, along with Kaplan-Meier Plotter and GEPIA validation, were used to characterize functional pathways and assess the prognostic relevance of the hub genes. Survival analysis demonstrated that GTSE1 is associated with unfavourable prognosis, whereas RRM2, CCNB1, and CCNB2 emerge as promising prognostic biomarkers. Therefore, the in-silico findings provide a basis for advancing targeted gene therapy approaches aimed at improving survival in patients with HCC.

Keywords: Hepatocellular carcinoma, GEO2R, Gene Ontology, Hub genes, Kaplan-Meier plot, GEPIA.

1. INTRODUCTION

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy and ranks among the leading causes of cancer-related deaths worldwide [1]. Its global incidence continues to rise, largely driven by the increasing prevalence of chronic liver diseases, including viral hepatitis, alcoholic liver disease, and metabolic disorders such as non-alcoholic fatty liver disease (NAFLD) [2,3]. Despite advances in diagnostic imaging, surgical techniques, and systemic therapies, the prognosis of HCC remains poor due to late diagnosis, high recurrence rates, and limited effective treatment options [4,6]. As a result, the development of preventive strategies and early detection biomarkers remains a major clinical priority. Recent studies emphasize the need for targeted and immune-based therapeutic approaches to improve survival and treatment outcomes in patients with advanced HCC [5]. Overall, the growing burden of HCC underscores the importance of integrated approaches that combine prevention, early detection, and novel therapeutic strategies to address this global health challenge.

Hepatocellular carcinoma (HCC) is a highly heterogeneous malignancy at the molecular, genetic, and immunologic levels, and recent discoveries of molecular biomarkers have greatly advanced its classification and management. Several gene expression based prognostic models have been developed to predict survival outcomes following liver resection, highlighting the importance of transcriptomic profiling in patient stratification [7]. Molecular subclassification of HCC into proliferative and non-proliferative groups based on signaling pathway activation has improved therapeutic precision, particularly in predicting response to targeted agents [8]. Key genomic alterations, including mutations in *TP53*, *CTNNB1,* and *TERT*, as well as dysregulation of the Wnt/β-catenin, PI3K/AKT/mTOR, and MAPK pathways, have been identified as potential biomarkers associated with tumor progression and treatment response [9]. Transcriptomic and genomic studies have also revealed distinct molecular subclasses of HCC that correlate with tumor differentiation, recurrence risk, and prognosis [10,11]. Furthermore, recent research has focused on immune-related biomarkers such as PD-L1 expression, tumor mutational burden, and immune cell infiltration, which are now recognized as predictors of response to immune checkpoint inhibitors [12,13]. Collectively, the integration of genomic, transcriptomic, and immunologic biomarkers has refined the molecular understanding of HCC and offers promising avenues for precision diagnosis and personalized therapy in liver cancer management.

In this study, HCC datasets obtained from the Gene Expression Omnibus (GEO) were analyzed to identify differentially expressed genes (DEGs) by comparing human HCC tissue with the corresponding noncancerous liver tissue. Subsequent bioinformatics analyses were performed to identify hub genes and conduct a series of functional evaluations.

2. MATERIALS AND METHODS

Retrieval of hepatocellular carcinoma datasets from the GEO

The GEO (https://www.ncbi.nlm.nih.gov/geo/) is used to download two HCC datasets with accession numbers GSE121248 and GSE84402.[14] The GSE121248 expression profile contained 70 HCC samples and 37 non-cancerous samples. The GSE84402 expression profile contained 14 HCC samples and 14 non-cancerous samples. The two datasets were compared and analyzed to retrieve DEGs from these groups [15-16].

GEO2R Analysis

The GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) analysis of datasets GSE121248 and GSE84402 was performed to obtain DEGs between noncancerous liver tissue and HCC. The following inclusion criteria for the DEGs contain upregulated genes that must have a log2 fold change (logFC)≥ 1 and an adjusted p-value < 0.01. The downregulated genes must have a log2 fold change (logFC) ≤ -1 and an adjusted p-value < 0.01. [17]

Hub Gene Selection

The overlapping genes of two different datasets were identified by constructing a Venn diagram with the online tool, Venny 2.1(https://bioinfogp.cnb.csic.es/tools/venny/). The common genes of both upregulated and downregulated DEGs are identified for two different data sets.

Protein-Protein Interaction network of the DEGs

The search tool (https://string-db.org/) is used to construct a PPI network of the DEGs for the retrieval of Interacting Genes [18]. Using Cytoscape, the network was visualized and using the Molecular Complex Detection (MCODE) app to screen the nodes of hub genes and to identify clusters with higher biological relevance. MCODE was executed with the parameters set to a degree cutoff of 2, haircut cluster finding, node score cutoff of 0.2, K-core of 2, and maximum depth of 100 [19].

Gene Ontology and KEGG Pathway Enrichment

The gene ontology and KEGG pathway enrichment analysis were performed using ShinyGO tool (https://bioinformatics.sdstate.edu/go/). Through GO enrichment analysis, we can roughly compare and classify DEGs to better understand their biological characteristics [20]. The KEGG provides functional interpretation of genes and genomes as whole network.The false discovery rate cutoff was set < 0.05 for demonstrating significance. The results were visualized as lollipop charts [21].

Kaplan-Meier Plotter and GEPIA of the Hub genes

The validation of survival biomarkers and analysis is done using Kaplan-Meier plotter https://kmplot.com/analysis/) with the high and low expression of genes. The hub genes are detected. The Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku. cn/), an open-source cancer big data analysis website, was used to analyze the differential expression of HCC and normal tissues from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) portal. All hub genes were analysed individually, and |log2FC| ¼1 and p-value ¼ 0.01 were used as parameters [22].

3. RESULTS

Identification of top DEGS

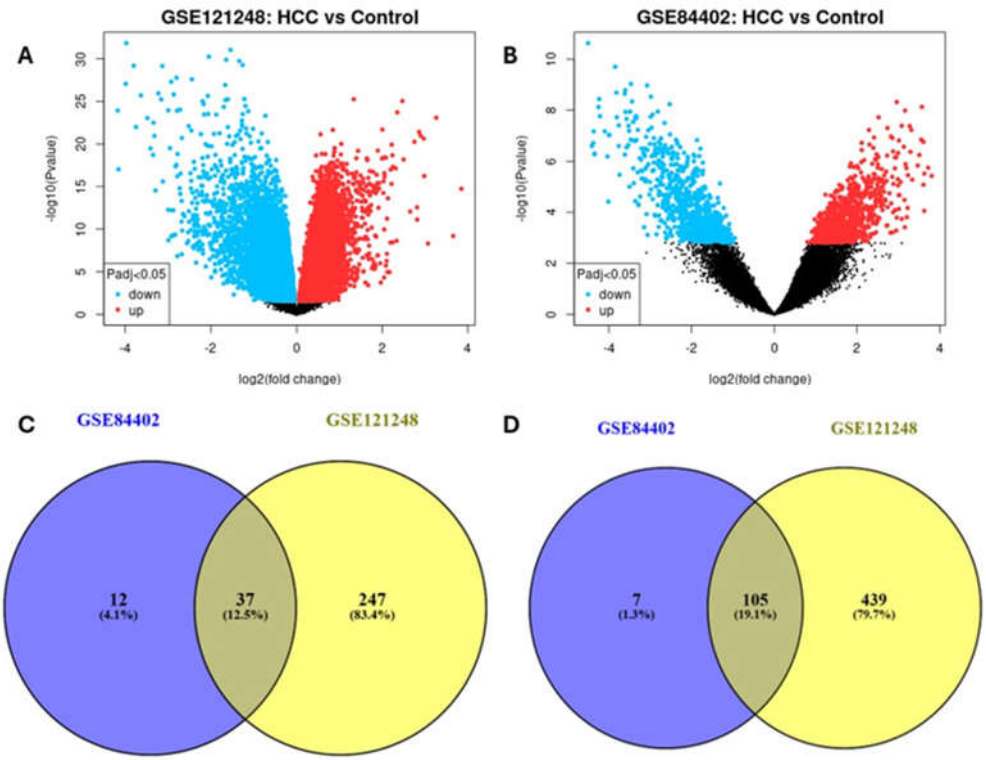

Figure 1: A and **B.** Volcano plot from GEO2R analysis for datasets GSE121248 and GSE84402. **C** and **D.** Venn diagram constructed for up regulated and down regulated genes for both datasets

The DEGs were identified by performing GEO2R analysis in two GEO datasets GSE121248 and GSE84402. The volcano plots obtained for both the datasets through GEO2R analysis are presented in Figure 1A and 1B. Dataset GSE121248 consists of 37 normal and 70 HCC samples while GSE84402 had 14 normal and 14 HCC samples. GSE121248 had 284 upregulated and 544 downregulated genes, and GSE84402 had 49 upregulated and 112 downregulated genes. A total of 37 upregulated genes and 105 downregulated genes overlapped in the Venn diagram from both datasets, as shown in Figures 1C and 1D.

PPI and Target Selection

Figure 2A consists of a PPI network image of the DEGs obtained from STRING. The MCODE plugin in Cytoscape was utilized to reveal the gene clusters that exhibit strong interconnectivity and biological relevance. The gene cluster that is formed with a high score of 32.485 with 536 edges and 34 genes- MKI67, CCNB1, TOP2A, TTK, NCAPG, DEPDC1, UBE2T, GMNN, CENPF, CDKN3, OIP5, PRC1, CCNB2, AURKA, ANLN, MELK, FAM83D, CDK1, PTTG1, PBK, GTSE1, NEK2, CENPW, RRM2, FANCI, KIF4A, CENPE, RACGAP1, BIRC5, GINS1, ASPM, UBE2C, KIF23 and NUF2 is seen in Figure 2B.

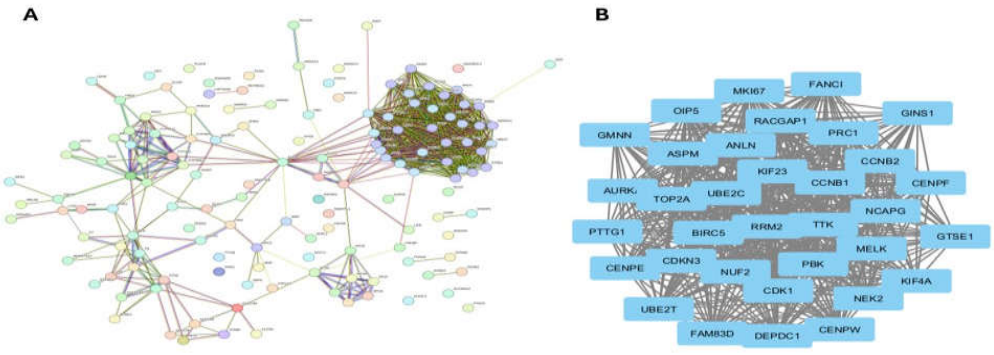

Figure 2: A. STRING PPI **B.** MCODE cluster with highest score

Gene Ontology and Pathway Enrichment

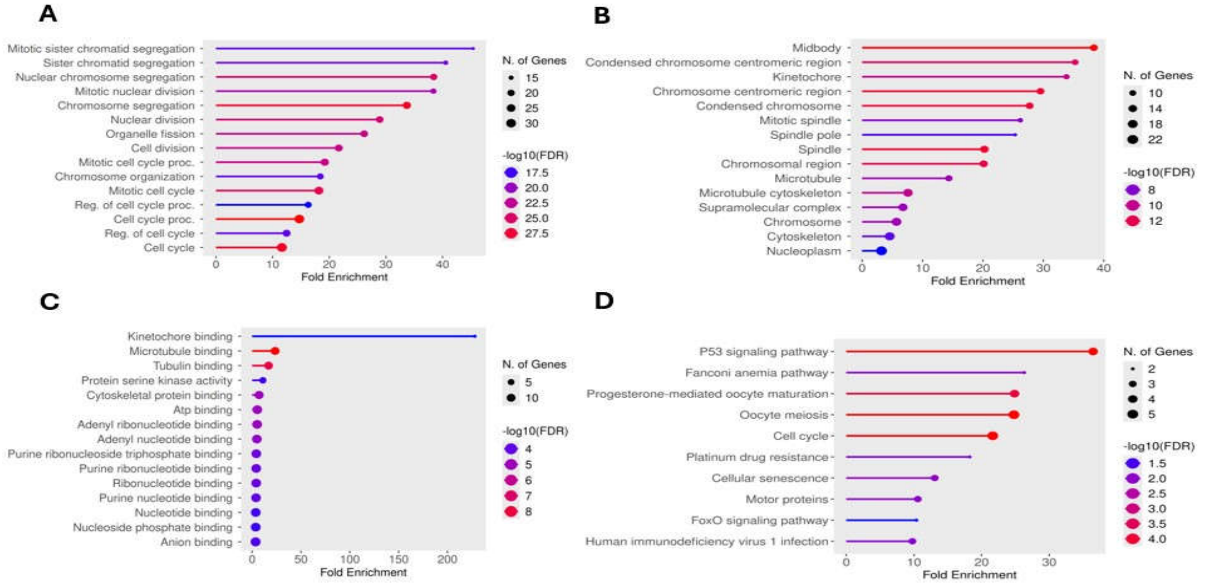

Figure 3: Lollipop charts representing Gene Ontology and KEGG Pathway Enrichment **A.** Biological processes **B.** Cellular components **C.** Molecular functions **D.** KEGG pathway enrichment

Gene ontology and KEGG pathway enrichment were conducted to understand the biological processes, cellular components, molecular functions, and signalling pathways that the target genes are associated which were identified through the MCODE cluster. Figure 3 consists of lollipop charts generated by ShinyGO for each category. GTSE1 (G2 and S-phase expressed 1), RRM2 (Ribonucleotide reductase regulatory subunit M2), CCNB1 (Cyclin B1) and CCNB2 (Cyclin B2) are associated with p53 signalling pathway and CCNB1 and CCNB2 are associated with foxO signalling pathway ($p > 0.05$). Studies have demonstrated that p53 and foxO signalling pathways contribute significantly to HCC progression.

Survival Analysis

The survival plot was constructed to determine the prognostic values of the identified targets for HCC using GEPIA and is shown in figure 4. GTSE1 has the worst overall survival with the hazard ratio (HR) 1.5 during high expression in patients and is significantly associated with poor prognosis ($p = 0.028$). Similarly, RRM2 (HR = 1.9, $p = 0.00058$) and CCNB1 (HR = 2, $p = 0.00015$) high expressions correlate with significantly lower survival and high mortality risk

since HR is closer to value 2. High CCNB2 (HR = 1.4) expression in patients is associated with poor survival but was not statistically significant (p = 0.052).

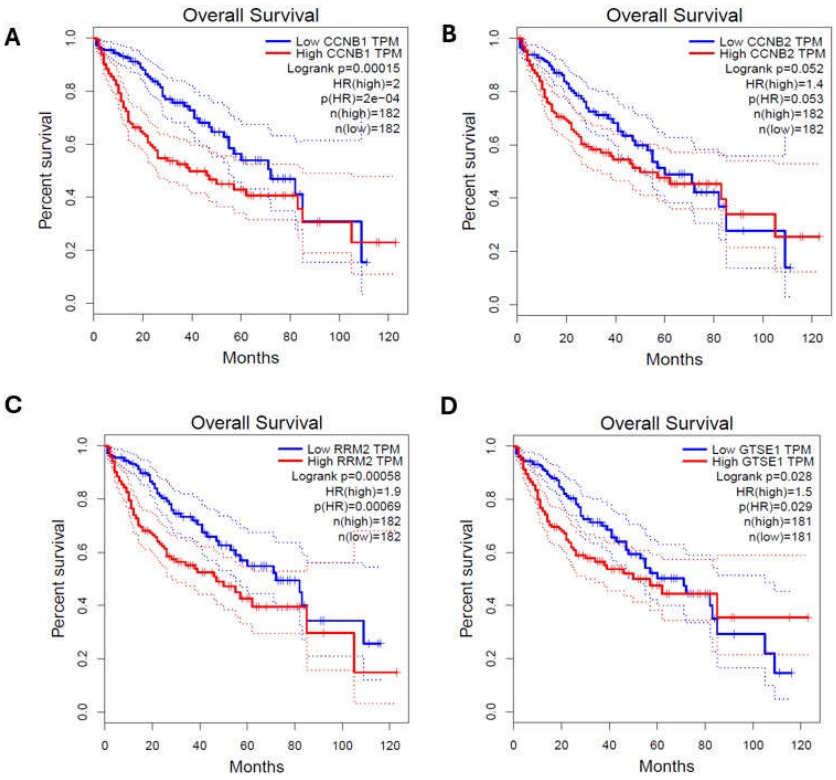

Figure 4: Survival plot for target genes plotted using GEPIA **A.** CCNB1 **B.** CCNB2 **C.** RRM2 **D.** GTSE1

4. DISCUSSION

In this study two microarray datasets from Asian countries GSE121248 and GSE84402 were used for gene screening. The DEGs were screened by GEO2R and the hub genes were found by constructing Venn diagram for the DEGs of two different HCC microarray datasets. The upregulated genes of both datasets and the downregulated genes of the datasets is taken for Venn diagram. A total of 37 genes is upregulated and 105 genes are downregulated genes is overlapped. The PPI network revealed extensive interconnectivity among the DEGs, and MCODE revealed the cluster with high biological significance. This high level of interconnectivity indicates strong functional associations among the genes within the module. The identified hub genes are involved in cell cycle regulation, mitotic division, and DNA replication, processes that are often dysregulated in cancer progression [23].

Genes such as MKI67, CCNB1, AURKA, CDK1 and TOP2A have been widely reported as proliferation markers and oncogenic drivers in HCC and other malignancies [24]. Their

clustering within the same module reinforces the hypothesis that dysregulated cell cycle associated pathways contribute to uncontrolled cell proliferation and tumor progression in HCC. The identification of these hub genes provides valuable insights into the molecular mechanisms underlying HCC progression and offers potential biomarkers or therapeutic targets for future investigation.

GO enrichment analysis enabled the classification of DEGs by biological process, cellular components and molecular functions, thereby providing insights into their potential biological roles. KEGG pathway analysis offered a systems-level understanding of the molecular interactions and functional pathways associated with the identified gene cluster. A false discovery rate (FDR) threshold of < 0.05 was applied to ensure the reliability of the enrichment results.

The enriched GO terms were primarily associated with the cell cycle regulation, mitotic spindle organization and chromosome segregation, indicating that the identified genes play a critical role in cell division and proliferation. In KEGG pathway, GO enriched cell cycle regulation, mitotic spindle organization and chromosome segregation, indicating that the identified genes play a critical role in cell division and proliferation. KEGG pathway enrichment revealed that several key genes such as GTSE1, RRM2, CCNB1 and CCNB2 were involved in the p53 signalling pathway, while CCNB1 and CCNB2 were also linked to the FoxO signalling pathway. These pathways are well known to regulate cell cycle arrest, apoptosis and DNA repair mechanisms, all of which are frequently disrupted during hepatocellular carcinoma (HCC) progression. Previous studies have demonstrated that dysregulation of the p53 pathway leads to impaired cell cycle control and evasion of apoptosis, contributing to HCC initiation and metastasis [25].

The involvement of the hub genes identified in this study within these critical oncogenic pathways underscores their potential as diagnostic biomarkers and therapeutic targets for HCC. The p53 signalling pathway plays a pivotal role in maintaining genomic stability by regulating cell cycle arrest, DNA repair, senescence and apoptosis in response to cellular stress and DNA damage. As a tumor suppressor, p53 prevents the propagation of the damaged cells. The inactivation or mutation in the p53 leads to uncontrolled proliferation and resistance to apoptosis [26,27].

The Forkhead box O (FoxO) signaling pathway has been implicated in promoting cell proliferation, survival and resistance to oxidative stress [28]. The FoxO signaling pathway

controls diverse biological processes, including cell cycle regulation, oxidative stress response, DNA repair, and apoptosis. In cancer, dysregulation of the FoxO gene through AKT-mediated phosphorylation and inactivation can promote tumorigenesis and survival of malignant cells [29].

The results are consistent with previous studies reporting that CCNB1 and CCNB2, key regulators of the G2/M transition, promote uncontrolled cell division and tumor proliferation in HCC [30]. Likewise, RRM2, an enzyme essential for DNA synthesis and repair, has been implicated in enhanced tumor cell proliferation and chemoresistance [31]. GTSE1 has also been regulated by p53 which modulates cell cycle progression and suppresses apoptosis.

Collectively, these findings highlight GTSE1, RRM2 and CCNB1 as potential prognostic biomarkers and therapeutic targets in hepatocellular carcinoma, emphasizing their involvement in pathways associated with the cell cycle regulation and tumor progression.

5. CONCLUSION

In this study, the results reveal about the hub genes related to HCC prognosis. The DEGs obtained from two different datasets show expression of genes. Key genes involved in the p53 and FoxO pathways can be identified and subsequently validated through *in-vitro* and *in-vivo* experimental approaches. GO and KEGG analyses also allow the identification of additional gene characteristics, including their associated biological process, cellular components and molecular functions. Survival analysis indicated that GTSE1 is associated with unfavourable prognosis whereas RRM2, CCNB1 and CCNB2 serves as a potential prognostic biomarker. Thus *in-silico* approach provides a foundation for targeted gene therapy strategies aimed at improving the survival of HCC patients.

REFERENCE

1. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, McCullough ML, Patel AV, Ma J, Soerjomataram I, Flanders WD, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin*. 2018;68(1):31–54.

2. Degasperi E, Colombo M. Distinctive features of hepatocellular carcinoma in non-alcoholic fatty liver disease. *Lancet Gastroenterol Hepatol*. 2016;1(2):156–64.

3. Estes C, Anstee QM, Arias-Loste MT, Bantel H, Bellentani S, Caballeria J, Colombo M, Craxi A, Crespo J, Day CP, Eguchi Y, et al. Modeling NAFLD disease burden in multiple countries during 2016–2030. *J Hepatol*. 2018;69(4):896–904.

4. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet*. 2012;379(9822):1245–55.

5. Chakraborty E, Sarkar D. Emerging therapies for hepatocellular carcinoma (HCC). *Cancers (Basel)*. 2022;14(11):2798.

6. Bruix J, Boix L, Sala M, Llovet JM. Focus on hepatocellular carcinoma. *Cancer Cell*. 2004;5(3):215–19.

7. Nault JC, De Reyniès A, Villanueva A, Calderaro J, Rebouissou S, Couchy G, Decaens T, Franco D, Imbeaud S, Rousseau F, Azoulay D, et al. A hepatocellular carcinoma 5-gene score associated with survival after liver resection. *Gastroenterology*. 2013;145(1):176–87.

8. Gores GJ. Decade in review—hepatocellular carcinoma: HCC subtypes, stratification and sorafenib. *Nat Rev Gastroenterol Hepatol*. 2014;11(11):645–47.

9. Khemlina G, Ikeda S, Kurzrock R. Biology of hepatocellular carcinoma: implications for genomic and immune therapies. *Mol Cancer*. 2017;16(1):149.

10. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, Watanabe G, et al. Integrative transcriptome analysis reveals molecular subclasses of human hepatocellular carcinoma. *Cancer Res*. 2009;69(18):7385–92.

11. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*. 2015;47(5):505–11.

12. Ding X, He M, Chan AW, Liu X, Gao Y, Deng Y, Coarfa C, Weisenberger DJ, Chen W, Yao Y, et al. Genomic and epigenomic features of primary and recurrent hepatocellular carcinomas. *Gastroenterology*. 2018;155(3):807–21.

13. Llovet JM, Castet F, Heikenwalder M, Maini MK, Mazzaferro V, Pinato DJ, Pikarsky E, Zhu AX, Finn RS. Immunotherapies for hepatocellular carcinoma. *Nat Rev Clin Oncol*. 2021;19(3):151–72.

14. Wang SM, Ooi LL, Hui KM. Identification and validation of a novel gene signature associated with recurrence of human hepatocellular carcinoma. *Clin Cancer Res*. 2007;13(21):6275–83.

15. Chen CL, Tsai YS, Huang YH, Liang YJ, Sun YY, Su CW, Chau GY, Yeh YC, Chang YS, Hu JT, Wu JC. Lymphoid enhancer factor 1 drives HCC progression through EMT regulator control. *Hepatol Commun*. 2018;2(11):1392–407.

16. Wang H, Huo X, Yang XR, He J, Cheng L, Wang N, Deng X, Jin H, Wang N, Wang C, Zhao F, et al. STAT3-mediated upregulation of lncRNA HOXD-AS1 facilitates liver cancer metastasis by regulating SOX4. *Mol Cancer*. 2017;16(1):136.

17. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, et al. NCBI GEO: archive for functional genomics datasets. *Nucleic Acids Res*. 2013;41(Database issue):D991–95.

18. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein–protein interaction networks. *Nucleic Acids Res*. 2013;41(Database issue):D808–15.

19. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein networks. *BMC Bioinformatics*. 2003;4:2.

20. Mazandu GK, Chimusa ER, Mulder NJ. Gene ontology semantic similarity tools: features and challenges. *Brief Bioinform*. 2017;18(5):886–901.

21. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes and pathways. *Nucleic Acids Res*. 2017;45(D1):D353–61.

22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for gene expression analysis. *Nucleic Acids Res*. 2017;45(W1):W98–102.

23. Guo L, Chen C, Shi M, Wang F, Chen X, Chen L. Identification of hub genes and pathways in hepatocellular carcinoma through integrated bioinformatics analysis. *Front Oncol*. 2021;11:687200. doi:10.3389/fonc.2021.687200.

24. Zhang H, Liu T, Liu S, Zhang Z, Li F, Li X, Wang Y. Integrated bioinformatics analysis identifies hub genes associated with hepatocellular carcinoma. *Front Genet*. 2020;11:571210. doi:10.3389/fgene.2020.571210.

25. Zhang X, Xu H, Yin H, Zhang J, Su J, Zhao L. Dysfunction of the p53 pathway contributes to hepatocellular carcinoma progression and chemoresistance. *Front Oncol*. 2019;9:627. doi:10.3389/fonc.2019.00627.

26. Vousden KH, Prives C. The growing complexity of p53. *Cell*. 2009;137(3):413–31. doi:10.1016/j.cell.2009.04.037.

27. Kim Y, Jang S. Role of the p53 signaling pathway in hepatocellular carcinoma. *Cancers (Basel)*. 2019;11(12):1981. doi:10.3390/cancers11121981.

28. Sun W, Zhang T, Huang L, Wang C. FoxO signaling pathway modulation in hepatocellular carcinoma. *Cancer Cell Int*. 2020;20:1–12. doi:10.1186/s12935-020-01203-1.

29. Farhan M, Wang H, Gaur U, Little PJ, Xu J, Zheng W. FOXO signaling pathways as therapeutic targets in cancer. *Int J Biol Sci*. 2017;13(7):815–27. doi:10.7150/ijbs.20052.

30. Liu C, Zhang Y, Chen H, Jiang L, Xiao D, Wang Y. CCNB1 and CCNB2 are downregulated by miR-335-5p and associated with poor prognosis in HCC. *BMC Cancer*. 2020;20:1020. doi:10.1186/s12885-020-07549-9.

31. Zhang Y, Wang S, Li X, Fan H, Zhang Y. RRM2 facilitates tumor progression through DNA synthesis and cell cycle regulation in HCC. *Front Oncol*. 2021;11:667040. doi:10.3389/fonc.2021.667040.