Hub gene identification and prognostic model establishment for patients with HBV-related hepatocellular carcinoma

Lianmei Wang

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing

Jing Meng

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing

Shasha Qin

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing

Aihua Liang (✉ ahliang@icmm.ac.cn)

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing

Research Article

Keywords: HBV-related hepatocellular carcinoma (HBV-HCC), Prognosis, Protein–protein interaction (PPI), Hub genes, Machine learning

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-269454/v2>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Hub gene identification and prognostic model establishment for patients with HBV-related hepatocellular carcinoma

1 Lianmei Wang¹, Jing Meng¹, Shasha Qin¹, Aihua Liang^{1*}

2 ¹Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute
3 of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

4 **Correspondence to:* Professor Aihua Liang. 16. Dongzhimen Nanxioejie, Dongcheng District,
5 Beijing, China. Email: ahliang@icmm.ac.cn.

6

7

8

9

10

11

12

13

14

15

Abstract

Hepatocellular carcinoma (HCC) is associated with poor 5-year survival. Chronic infection with hepatitis B virus (HBV) contributes to ~50% of HCC cases. Identification of biomarkers is pivotal for the therapy of HBV-related HCC (HBV-HCC). We downloaded gene-expression profiles from Gene expression omnibus (GEO) datasets with HBV-HCC patients and the corresponding controls. Integration of these differentially expressed genes (DEGs) was achieved with the RobuStrankaggreg (RRA) method. DEGs functional analyses and pathway analyses was performed using the Gene ontology (GO) database, and the Kyoto encyclopedia of genes and genomes (KEGG) database respectively. Cyclin-dependent kinase 1 (CDK1), Cyclin B1 (CCNB1), Forkhead box M1 (FOXM1), Aurora kinase A (AURKA), Cyclin B2 (CCNB2), Enhancer of zeste homolog 2 (EZH2), Cell division cycle 20 (CDC20), DNA topoisomerase II alpha (TOP2A), BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B), and ZW10 interactor (ZWINT), were identified as the top-ten hub genes. The expression of hub-genes was verified in the liver cancer-riken, JP project from international cancer genome consortium (ICGC-LIRI-JP), the cancer genome atlas (TCGA) HCC cohort, and Human protein profiles dataset. A four-gene prognostic-related model based on the expression of ZWINT, EZH2, FOXM1 and CDK1 were established through Cox regression analysis in ICGC-LIRI-JP project, and verified in TCGA-HCC cohort. Furthermore, a nomogram model based on pathology stage, gender and four-genes prognostic model was built to predict the prognosis for HBV-HCC patients. In conclusion, ZWINT, EZH2, FOXM1 and CDK1 play a pivotal role in HBV-HCC, and are potential therapeutic targets of HBV-HCC.

Keywords: HBV-related hepatocellular carcinoma (HBV-HCC), Prognosis, Protein-protein interaction (PPI), Hub genes, Machine learning

1. Introduction

HCC is the second most lethal cancer worldwide, and carries 5-year survival of 18% [1]. About 3.5% of the world population suffers chronic infection with HBV [2], and ~50% of such cases lead to HCC [3]. In addition, male sex, cirrhosis, diabetes mellitus, aflatoxin exposure, alcohol intake, and tobacco consumption are risk factors for HCC [3-5].

Sodium taurocholate co-transporting polypeptide is the receptor for HBVs entering hepatocytes [6]. Long-term HBV infection, augmentation of HBV replication, genome integration of HBV, HBV mutants, and HBV-encoded oncoproteins work in HBV-HCC [3]. Universal vaccination of newborns against the HBV has reduced the prevalence of HCC dramatically [7]. Antiviral therapy with nucleoside analogs or nucleotide analogs suppresses HBV replication significantly in patients with chronic HBV infection [2]. Sorafenib, lenvatinib, regorafenib, and HBV-related T-cell immunotherapies are potential therapeutic methods for HBV-HCC [2, 8-10]. Early detection is important for cancer treatment, so biomarker detection is pivotal for HBV-HCC prevention.

Bioinformatics analysis integrated with data on RNA expression is performed to predict the prognosis of diseases. Data on RNA expression can be obtained with next-generation sequencing and microarrays [11]. GEO is a repository of data on high-throughput gene expression and hybridization arrays and chips. To avoid inconsistent results due to different experiments, GEO datasets can be incorporated with the RRA method for comparison of sequenced genes [12].

In the present study, we downloaded datasets from GEO datasets. Microarray data were analyzed with R (<https://www.r-project.org/>). DEGs from these datasets were integrated using the RRA method (https://cran.rstudio.com/bin/windows/contrib/3.5/RobustRankAggreg_1.1.zip). DEGs were evaluated by the database for annotation, visualization and integrated discovery (DAVID; <https://david.ncifcrf.gov/>). The GO database (www.geneontology.org/) focuses on the function of the genes and gene products, and the KEGG database (www.genome.jp/kegg/pathway.html) focuses on enrichment of pathways: both were employed for DEGs. We pinpointed hub genes with Cytoscape (<https://cytoscape.org/>) from DEGs analyzed by the search tool for the retrieval of interacting genes/proteins (STRING) database (<https://string-db.org/>). Then hub genes were verified with ICGC (<https://dcc.icgc.org/>)-LIRI-JP, TCGA (<https://portal.gdc.cancer.gov/>) HCC project, and Human protein profiles dataset (www.proteinatlas.org). Furthermore, a four gene prognosis model was constructed based on the

expression of ZWINT, EZH2, FOXM1 and CDK1. This study provides us potential molecular targets for HBV-HCC treatment.

2. Materials and methods

2.1. Data source. The criteria of GEO datasets selection in this study were as follows: (1) paired tissue samples including cancer and non-cancerous from the same patient; (2) HBV positive (3) Data was achieved with the same type of arrays; and (4) patient size more than 6. The exclusion criteria were as follows: (1) dataset biased, and (2) no paired tissues from the same patients. We searched in the GEO database, and only two results (GSE121248 and GSE55092) met our criteria for HBV-associated HCC. The information of these two datasets were listed in table1. The RNA-sequencing data and clinical information of ICGC-LIRI-JP project were downloaded from ICGC (<https://icgc.org/>), 201 HBV infected patients with paired tissue samples were achieved in this cohort. TCGA HCC project were downloaded from TCGA (<http://cancergenome.nih.gov/>).

Table 1 Details of the GEO HBV-related HCC data.

GEO	Sample	Platform	NO. of Patients	reference
GSE121248	HBV-HCC	GPL570	37	PMID:17975138 [13]
GSE55092	HBV-HCC	GPL570	10	PMID: 5141867 [14]

2.2. Screening of DEGs. We analyzed the data using R 4.0.2. Normalization and DEGs analyses were done with ‘limma R’ or ‘edgeR’ within Bioconductor (www.bioconductor.org/). All gene-expression data were subjected to log₂ transformation. DEGs were screened with corrected P < 0.05 and |log fold change (FC)| > 2. RRA within R was used to integrate the DEGs in the microarray data.

2.3. Pathway analyses of DEGs using GO and KEGG databases. Gene annotation (using the GO database) and pathway enrichment (using the KEGG database) of DEGs were analyzed with DAVID 6.8 and P < 0.05 was considered significant. Visualization was undertaken with ‘ggplot2’ within R.

2.4. Construction of protein-protein interaction (PPI) networks and module analyses. PPI networks were created using the STRING database. Cytoscape 3.7.1 was applied to screen for hub genes according to degrees. Hub genes were selected with degree ≥ 40 . We analyzed the modules in PPI networks. The MCODE default parameters were ‘Degree cutoff = 2’, ‘Node score cutoff = 0.2’, ‘K-core = 2’, and ‘Max. depth = 100’ [15].

2.5. *Prognostic analyses of hub genes.* Survival analyses of hub genes was done through R based on the ICGC-LIRI-JP or TCGA-HCC dataset. Univariate Cox regression analysis and multivariate Cox regression analysis and were undertaken through genes selected from survival data with $P < 0.05$. Correlation of the prognosis with candidate genes was analyzed and visualized by ‘survival’, ‘survminer’ and ‘ggplot2’ packages within R. Time-dependent analyses of Receiver operating characteristic (ROC) curves by survival were done using the ‘ROC’ package within R.

2.8. *Nomogram.* Multivariate Cox regression analyses was undertaken to select significant factors using a four-gene prognostic model and clinical characteristics for the prognosis. The result of multivariate cox regression was visualized with forest plots. The nomogram, calculation plot, ROC plot and Decision curve analysis (DCA) were calculated and visualized with ‘rms’, ‘foreign’, ‘survival’, ‘regplot’, ‘mstate’, ‘survivalROC’, ‘survcomp’, ‘Hmisc’, ‘grid’, ‘lattice’, ‘Formula’, ‘ggplot2’ and ‘rmda’ packages within R.

2.9. *Identification of the potential drugs* The Drug Gene Interaction Database (DGIdb) (<https://www.dgiddb.org>) provides the information of drug-gene interaction. ZWINT, EZH2, FOXM1 and CDK1 targeted drugs were identified in DGIdb and Pubmed (<https://www.ncbi.nlm.nih.gov/>), and visualized with the Cytoscape software.

2.9. *Statistical analysis.* Data are shown as the mean $\pm$ SEM. Statistical analyses were performed using Prism 8.0 software (Graph-pad software inc.) or R, and P values less than 0.05 were considered significant.

3. Results

3.1. *Screening of DEGs in HBV–HCC patients.*

To clearly describe our study, the analysis procedure was shown in Figure 1. The HBV–HCC datasets GSE121248 and GSE55092 were normalized, and the DEGs were selected by corrected $P < 0.05$ and $|\log FC| > 2$ with ‘limma’ (Fig 2a, b). DEGs were integrated through ‘RRA’ according to corrected $P < 0.05$ and $|\log FC| > 1$. A total of 290 DEGs were identified, 101 of which showed upregulated expression, and 190 of which exhibited downregulated expression (Supplementary Table S1). We summarized the top-20 genes with upregulated and downregulated expression in the heatmap (Fig 2c.)

Figure 1. Overall workflow of this study.

Figure 2. Screening and integration of DEGs. DEGs of GSE121248 (a) and GSE55092 (b) were integrated with RRA method, and the top 20 upregulated and downregulated DEGs were listed in c. Red dots and blue dots in the a and b are genes with upregulated expression and downregulated expression, respectively, with $P < 0.05$, and $|\log FC| > 2$.

3.2. Enrichment analyses using GO and KEGG databases. Functional annotation of integrated DEGs was analyzed using DAVID. Functional annotation was based on three categories: biological process (BP), molecular function (MF), and cell component (CC). The top-five BP were “Oxidation-reduction process”, “Positive regulation of transcription from RNA polymerase II promoter”, “Mitotic nuclear division”, “Cell division”, and “Positive regulation of cell proliferation”. The top-five CC were “Cytoplasm”, “Extracellular space”, “Cytosol”, “Extracellular region”, and “Extracellular space”. The top-five MF were “Protein binding”, “Identical protein binding”, “Protein homodimerization activity”, “Calcium ion binding”, and “Heme binding” (Fig. 3a). All GO annotation of genes was listed in Supplementary Table S2.

Pathway analyses using the KEGG database were undertaken using DAVID and visualized with ‘ggplot2’. In these DEGs, the pathways that were augmented were “Metabolic pathways”, “Cell cycle”, “Chemical carcinogenesis”, “Oocyte meiosis”, “Biosynthesis of antibiotics”, “p53 signaling pathway”, “Retinol metabolism”, “Drug metabolism-cytochrome P450”, “Mineral absorption”, “Bile secretion”, “Complement and coagulation cascades”, “Carbon metabolism”, “Steroid hormone biosynthesis”, “Arachidonic acid metabolism”, “Glycolysis / Gluconeogenesis”, “Biosynthesis of amino acids”, “Metabolism of xenobiotics by cytochrome P450”, “Progesterone-mediated oocyte maturation”, “Tryptophan metabolism”, “PPAR signaling pathway”, “Linoleic acid metabolism”, “Fructose and mannose metabolism”, “Prion diseases”, “Glycine, serine and threonine metabolism”, “Ovarian steroidogenesis”, “Caffeine metabolism” and “Histidine metabolism” (Fig. 3b). The genes involved in these pathways were listed in Supplementary Table S3.

Figure 3. Functional analyses (using the GO database) and pathway analyses (using the KEGG database) of DEGs. (a) The top-five categories for BP, MF and CC, (b) Augmentation of pathways using the KEGG database.

3.2. Investigation of DEGs using PPI networks.

We identified 290 DEGs in the GSE121248 and GSE55092 datasets, and constructed PPI networks through the STRING database. We screened eleven functional modules from the PPI network through MCODE within Cytoscape. The main function module are shown in Fig. 4a. The genes in this module that were enriched according to the KEGG database were “Cell cycle”, “Oocyte meiosis”, “p53 signaling pathway”, “Progesterone-mediated oocyte maturation” (Fig. 4b). CDK1, CCNB1, FOXM1, AURKA, CCNB2, EZH2, CDC20, TOP2A, BUB1B and ZWINT were screened as top hub genes through Cytoscape according to the value of degree ≥ 40 . The PPI network of hub genes was shown in Fig. 4c, and the expression heatmap in GSE121248 and GSE55092 was shown in Fig. 4d.

168

169 Figure 4. PPI network of significant module and hub genes.

170 3.3. Verification of hub genes.

171 We verified these hub genes in ICGC-LIRI-JP dataset, TCGA HCC cohort, and Human
 172 Protein Profiles. We identified DEGs through ‘edgeR’ in the ICGC-LIRI-JP and TCGA cohort.
 173 Expression of these hub genes was increased dramatically in ICGC-LIRI-JP (Fig. 5a) and TCGA
 174 (Fig. 5b) project. We also confirmed these ten hub genes protein expression patterns in the Human
 175 Protein Profiles, FOXM1 and AURKA were robustly positive in HCC samples compared with
 176 normal tissues, TOP2A, CCNB1, CCNB2, CDK1, CDC20, and EZH2 were moderately increased
 177 in HCC cohort in contrast to normal patients (Fig. 5c). However, BUB1B were not found in the
 178 website.

Figure 5. mRNA and protein expression patterns of the indicated genes. The mRNA levels in ICGC-LIRI-JP cohort (a), and TCGA HCC dataset (b), and protein levels in Human Protein Profiles dataset (c). *** $P \leq 0.001$ vs. normal patients.

3.4. Construction and verification of a four-gene prognostic model.

The Kaplan–Meier curve for Overall survival (OS) of each hub gene was performed with ICGC-LIRI-JP cohort, and all of these hub genes were significantly correlated with patient survival (Fig. 6). Expression of these hub genes in ICGC-LIRI-JP project was analyzed with univariate and multivariate cox regression analyses. These patients were separated into low or high-risk groups according to the median risk score as the cut-off value. All hub genes were identified as independent prognostic factors related to patient survival with $P < 0.05$ after univariate cox regression analyses. ZWINT, EZH2, FOXM1 and CDK1 were identified after multivariate cox

regression analyses, their coefficients were -0.1396, 0.1638, 0.1217 and 0.1614, and hazard ratios were 0.8696, 1.1779, 1.1294 and 1.1752 respectively. The Kaplan–Meier curve for OS in the high-risk and low-risk groups based on ZWINT, EZH2, FOXM1 and CDK1 was dramatically different ($P = 4.956 \times 10^{-7}$) (Fig. 7a). The prognostic capacity was analyzed by the area under the curve (AUC) of ROC (Fig. 7a). The AUC for survival at 1, 2 and 3 years was 0.841, 0.774 and 0.813, respectively. A higher AUC indicates a better forecasting model than a lower AUC, so our four-gene model showed high sensitivity and specificity to assess the prognosis of HBV-HCC patients.

Furthermore, we verified the four-gene prognostic model by TCGA HCC cohort, The Kaplan–Meier curve for OS was significant different ($P = 2.718 \times 10^{-3}$) between the high- and low-risk groups (Fig. 7b). The AUC for survival at 1, 2 and 3 years was 0.636, 0.636 and 0.644, respectively (Fig. 7b). Therefore, we can conclude that ZWINT, EZH2, FOXM1 and CDK1 are prognosis molecules for ICGC-LIRI-JP and TCGA HCC project, and perform better forecasting ability in ICGC-LIRI-JP project.

Figure 6. Kaplan-meier curve of the indicated genes.

Figure 7. Survival analysis based on the gene expression of ZWINT, EZH2, FOXM1 and CDK1. Kaplan–Meier survival curves and time-dependent ROC curves of the prognostic model in the ICGC-LIRI-JP (a) and TCGA (b) dataset.

3.5. Establishment of a prognostic model using a nomogram.

Multivariate cox regression was undertaken to evaluate the four-gene prognostic model and clinical factors (age, gender, histology grade, pathology stage, cancer history and malignancy) for the prognosis of HBV-HCC patients. The pathology stage, gender and four-gene prognostic model were independent prognostic factors with $P < 0.05$ (Fig. 8).

Figure 8. Multivariate Cox regression analysis of clinical factors and four-gene prognostic model with OS. *P < 0.05, **P < 0.01, and ***P < 0.001.

We constructed a nomogram with the three independent prognostic factors selected from multivariate cox regression analyses to predict OS at 1, 2 and 3 years in the ICGC-LIRI-JP cohort (Fig. 9a). The gray line in the calculation curve indicated the best prediction, and the nomogram model (described with a red line) matched well with the gray line (Fig. 9b). Therefore, we concluded that the nomogram model performed well. The AUC of the nomogram mode for OS at 1, 2 and 3 years was 0.774, 0.841 and 0.854, respectively (Fig. 9c). We evaluated the nomogram model through DCA. The nomogram model based on a combination of pathology stage, gender and four-gene prognostic model performed better to predict OS than that using a single factor (Fig. 9d).

Figure 9. Analyses of the ICGC-LIRI-JP cohort using a nomogram. (a) Nomogram according to pathology stage, gender, prognostic model, and histology grade. (b) Calibration curves for predicting OS at 1, 2 and 3 years using the nomogram model. (c) Time-dependent ROC curves using the nomogram model. (d) DCA curves using the nomogram model.

3.6. Identification of potential inhibitors for HBV-HCC related biomarkers.

DGIdb was applied to determine the potential drug targeted. Drugs interacted with ZWINT, EZH2, FOXM1 or CDK1 were identified in DGIdb and Pubmed. Drugs targeted with ZWINT or FOXM1 were not available in DGIdb. Both EZH2 and CDK1 were upregulated in HBV-HCC tissues, and their expression patterns are negative synergy with prognostics. Hence, inhibitors for EZH2 or CDK1 are potential drugs to treat HBV-HCC. CHEMBL1236539, AG-24322, BMS-387032, RONICICLIB, AT-7519, DINACICLIB, JNJ-7706621, RIVICICLIB, RGB-286638, VORUCICLIB, CT-7001, CHIR-99021, ALVOCIDIB, ZOTIRACICLIB, BOHEMINE, SELICICLIB, ALSTERPAULLONE, PHA-793887, MILCICLIB, AZD-5438, RG-547 and RO3306 are inhibitors for CDK1 (Fig. 10). Astemizole, Tanshinone I, GSK126, Benzimidazole derivative, DZNep, TAZEMETOSTAT, CPI-1205, PF-06821497, EPZ005687, EI1, GSK343, GSK926, EPZ011989, CPI-169, ZLD1039, UNC1999, OR-S1, DS-3201b, SAH-EZH2, Wedelolactone, apomorphine, hydrochloride, oxphenbutazone, nifedipine, ergonovine, maleate, AZD9291, MAK683, A769661, GNA022, ANCR and FBW7 are inhibitors for EZH2 (Fig. 10).

Figure 10. Inhibitors of EZH2 or CDK1.

4. Discussion

We integrated DEGs obtained from HBV-HCC-related GEO datasets with the RRA method, conducted functional analyses using the GO database, pathway analyses using the KEGG database, and identified hub genes in these DEGs with PPI network and Cytoscape software. CDK1, CCNB1, FOXM1, AURKA, CCNB2, EZH2, CDC20, TOP2A, BUB1B and ZWINT were detected as the top-ten hub genes. ZWINT, EZH2, FOXM1 and CDK1 were pinpointed as prognosis-related

molecules for HBV–HCC. In addition, a nomogram model was built to predict the prognosis for HBV–HCC patients.

Integrated DEGs were clustered according to pathway analyses, and alterations were detected in “glycolysis”, “p53 signaling pathway”, “cell cycle”, and “bile secretion”. Greater consumption of glucose in tumors than that in normal tissues was discovered first by the German physiologist Otto Warburg [16]. The universal metabolic feature of cancer cells is preferential conversion of glucose to lactate instead of entry into the tricarboxylic-acid cycle [17]. This action provides carbon sources for rapid growth of tumor cells and an acidic micro-environment that aids the invasion and metastasis of cancer cells [18]. P53 is one of the most frequent mutation sites in HCC [1]. P53, as a tumor suppressor, regulates cell proliferation or activates senescence and apoptosis [19]. P53 null mice develop leukemias and sarcomas [20]. P53 contributes to the cancer hallmark of evading growth suppressors [19]. P53 deregulates aerobic glycolysis by targeting the key enzymes of glucose metabolism, including glucose transporter 3, hexokinase 2, phosphoglycerate mutase, and pyruvate kinase isoform M2 [21]. Alterations of the synthesis, transport, and metabolism of bile acids are universal in liver inflammation and HCC [22]. Secretion of bile acids from the liver into the bile duct is mediated by bile salt export pump (BSEP), expression of which is regulated by the farnesoid X receptor (FXR) [23]. Deficiency of BSEP or the FXR induces tumor genesis in humans and mice [24, 25]. We isolated hub genes through analyses of PPI networks, and CDK1, CCNB1, FOXM1, AURKA, CCNB2, EZH2, CDC20, TOP2A, BUB1B and ZWINT were pinpointed as the top ten hub genes. CDK1, CCNB1, CCNB2, CDC20 and BUB1B are involved in the cell cycle, which is usually aberrant in tumor cells.

We identified ZWINT, EZH2, FOXM1 and CDK1 as pivotal biomarkers for the prognosis of HBV–HCC patients. As a regulator of centromere function and cell growth, Zwint is associated with chromosome-instability signatures and poor clinical prognosis in tumor patients [26]. EZH2 as an enzymatic catalytic subunit of polycomb repressive complex 2 (PRC2) regulates cell cycle progression, autophagy, and apoptosis. Ectopic expression of EZH2 was observed in prostate cancer, esophageal cancer, breast cancer, gastric cancer, anaplastic thyroid carcinoma, nasopharyngeal carcinoma and endometrial carcinoma [27]. FOXM1 has multiple roles in the growth, angiogenesis, metabolism, migration, DNA-damage response, development and progression of tumor cells. FOXM1 expression is augmented in cancer of the liver, lung, breast, stomach, brain, colon, pancreas, prostate gland, and blood [28-30]. FOXM1 depletion in mice can

lead to embryonic lethality, and FOXM1 is necessary for mitosis of hepatoblast-like precursor cells and liver regeneration [31, 32]. CDK1 belongs to Ser/Thr kinase family, and is indispensable for the centrosome cycle and onset of mitosis [33]]. Deletion of CDK1 causes lethality in mouse embryos in morula and blastocyst stages [34]. CDK1 activated by cyclin A and cyclin B drives the phosphorylation of thousands of proteins to mitosis [35]. Aberrant regulation of CDK1 expression triggers genomic instability and chromosomal instability, which are signature features of chromosomally unstable tumors [36]. Due to the oncogenic role of EZH2 and CDK1, inhibitors were used in clinical trial or preclinical study. EZH2 inhibitors including TAZEMETOSTAT, CPI1205, GSK126, PF-06821497, SHR2554, DS-3201b, and MAK683 were in the process of clinical test [27]. CDK1 inhibitor RO3306 suppressed tumor growth in a preclinical model of HCC [37]. However, ZWINT and FOXM1 inhibitors were barely reported.

5. Conclusion

In the present study, we identified ZWINT, EZH2, FOXM1 and CDK1 as potential molecular biomarkers of HBV-HCC. This study provides us a proof-of-concept that targeting ZWINT, EZH2, FOXM1 or CDK1 may treat HBV-HCC.

Data availability

Publicly datasets were analyzed in the present study. The data can be found at <https://dcc.icgc.org/>; <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga> and www.ncbi.nlm.nih.gov/geo/.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgements

We thank the Research square (<https://www.researchsquare.com/>) for publishing our manuscript as preprint.

Funding

Funding information is not applicable

References

1. A. J. Craig, J. von Felden, T. Garcia-Lezana, S. Sarcognato and A. Villanueva, "Tumour

313 evolution in hepatocellular carcinoma," *Nat Rev Gastroenterol Hepatol*, vol. 17, no. 3, pp. 139-
314 152, 2020.

315 2. A. T. Tan and S. Schreiber, "Adoptive T-cell therapy for HBV-associated HCC and HBV
316 infection," *Antiviral Res*, vol. 176, pp. 104748, 2020.

317 3. Y. Xie, "Hepatitis B Virus-Associated Hepatocellular Carcinoma," *Adv Exp Med Biol*, vol. 1018,
318 pp. 11-21, 2017.

319 4. C. Global Burden of Disease Cancer, "Global, Regional, and National Cancer Incidence,
320 Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for
321 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study,"
322 *JAMA Oncol*, vol. 5, no. 12, pp. 1749-1768, 2019.

323 5. J. A. Marrero, R. J. Fontana, S. Fu, H. S. Conjeevaram, G. L. Su and A. S. Lok, "Alcohol,
324 tobacco and obesity are synergistic risk factors for hepatocellular carcinoma," *J Hepatol*, vol. 42,
325 no. 2, pp. 218-224, 2005.

326 6. H. Yan, G. Zhong, G. Xu, W. He, Z. Jing, Z. Gao, Y. Huang, Y. Qi, B. Peng, H. Wang, L. Fu, M.
327 Song, P. Chen, W. Gao, B. Ren, Y. Sun, T. Cai, X. Feng, J. Sui and W. Li, "Sodium taurocholate
328 cotransporting polypeptide is a functional receptor for human hepatitis B and D virus," *Elife*, vol.
329 3, 2012.

330 7. M. H. Chang, C. J. Chen, M. S. Lai, H. M. Hsu, T. C. Wu, M. S. Kong, D. C. Liang, W. Y. Shau
331 and D. S. Chen, "Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular
332 carcinoma in children. Taiwan Childhood Hepatoma Study Group," *N Engl J Med*, vol. 336, no.
333 26, pp. 1855-1859, 1997.

334 8. C. H. Hsu, Y. C. Shen, Y. Y. Shao, C. Hsu and A. L. Cheng, "Sorafenib in advanced
335 hepatocellular carcinoma: current status and future perspectives," *J Hepatocell Carcinoma*, vol. 1,
336 pp. 85-99, 2014.

337 9. M. Kudo, R. S. Finn, S. Qin, K. H. Han, K. Ikeda, F. Piscaglia, A. Baron, J. W. Park, G. Han, J.
338 Jassem, J. F. Blanc, A. Vogel, D. Komov, T. R. J. Evans, C. Lopez, C. Dutcus, M. Guo, K. Saito,
339 S. Kraljevic, T. Tamai, M. Ren and A. L. Cheng, "Lenvatinib versus sorafenib in first-line treatment
340 of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial,"
341 *Lancet*, vol. 391, no. 10126, pp. 1163-1173, 2018.

342 10. J. Bruix, S. Qin, P. Merle, A. Granito, Y. H. Huang, G. Bodoky, M. Pracht, O. Yokosuka, O.
343 Rosmorduc, V. Breder, R. Gerolami, G. Masi, P. J. Ross, T. Song, J. P. Bronowicki, I. Ollivier-

Hourmand, M. Kudo, A. L. Cheng, J. M. Llovet, R. S. Finn, M. A. LeBerre, A. Baumhauer, G. Meinhardt, G. Han and R. Investigators, "Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial," *Lancet*, vol. 389, no. 10064, pp. 56-66, 2017.

11. X. Gao, Y. Chen, M. Chen, S. Wang, X. Wen and S. Zhang, "Identification of key candidate genes and biological pathways in bladder cancer," *PeerJ*, vol. 6, pp. e6036, 2018.

12. R. Kolde, S. Laur, P. Adler and J. Vilo, "Robust rank aggregation for gene list integration and meta-analysis," *Bioinformatics*, vol. 28, no. 4, pp. 573-580, 2012.

13. S. M. Wang, L. L. Ooi and K. M. Hui, "Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma," *Clin Cancer Res*, vol. 13, no. 21, pp. 6275-6283, 2007.

14. M. Melis, G. Diaz, D. E. Kleiner, F. Zamboni, J. Kabat, J. Lai, G. Mogavero, A. Tice, R. E. Engle, S. Becker, C. R. Brown, J. C. Hanson, J. Rodriguez-Canales, M. Emmert-Buck, S. Govindarajan, M. Kew and P. Farci, "Viral expression and molecular profiling in liver tissue versus microdissected hepatocytes in hepatitis B virus-associated hepatocellular carcinoma," *J Transl Med*, vol. 12, pp. 230, 2014.

15. G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinformatics*, vol. 4, pp. 2, 2003.

16. O. Warburg, F. Wind and E. Negelein, "The Metabolism of Tumors in the Body," *J Gen Physiol*, vol. 8, no. 6, pp. 519-530, 1927.

17. N. N. Pavlova and C. B. Thompson, "The Emerging Hallmarks of Cancer Metabolism," *Cell Metab*, vol. 23, no. 1, pp. 27-47, 2016.

18. R. A. Gatenby and E. T. Gawlinski, "The glycolytic phenotype in carcinogenesis and tumor invasion: insights through mathematical models," *Cancer Res*, vol. 63, no. 14, pp. 3847-3854, 2003.

19. D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, no. 5, pp. 646-674, 2011.

20. N. Ghebranious and L. A. Donehower, "Mouse models in tumor suppression," *Oncogene*, vol. 17, no. 25, pp. 3385-3400, 1998.

21. A. S. Gomes, H. Ramos, J. Soares and L. Saraiva, "p53 and glucose metabolism: an orchestra to be directed in cancer therapy," *Pharmacol Res*, vol. 131, pp. 75-86, 2018.

375 22. W. Jia, G. Xie and W. Jia, "Bile acid-microbiota crosstalk in gastrointestinal inflammation and
376 carcinogenesis," *Nat Rev Gastroenterol Hepatol*, vol. 15, no. 2, pp. 111-128, 2018.

377 23. P. A. Dawson, T. Lan and A. Rao, "Bile acid transporters," *J Lipid Res*, vol. 50, no. 12, pp.
378 2340-2357, 2009.

379 24. F. Iannelli, A. Collino, S. Sinha, E. Radaelli, P. Nicoli, L. D'Antiga, A. Sonzogni, J. Faivre, M.
380 A. Buendia, E. Sturm, R. J. Thompson, A. S. Knisely, G. Natoli, S. Ghisletti and F. D. Ciccarelli,
381 "Massive gene amplification drives paediatric hepatocellular carcinoma caused by bile salt export
382 pump deficiency," *Nat Commun*, vol. 5, pp. 3850, 2014.

383 25. F. Yang, X. Huang, T. Yi, Y. Yen, D. D. Moore and W. Huang, "Spontaneous development of
384 liver tumors in the absence of the bile acid receptor farnesoid X receptor," *Cancer Res*, vol. 67, no.
385 3, pp. 863-867, 2007.

386 26. H. Ying, Z. Xu, M. Chen, S. Zhou, X. Liang and X. Cai, "Overexpression of Zwint predicts
387 poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-
388 cycle-related proteins," *Onco Targets Ther*, vol. 11, pp. 689-702, 2018.

389 27. R. Duan, W. Du and W. Guo, "EZH2: a novel target for cancer treatment," *J Hematol Oncol*,
390 vol. 13, no. 1, pp. 104, 2020.

391 28. I. M. Kim, T. Ackerson, S. Ramakrishna, M. Tretiakova, I. C. Wang, T. V. Kalin, M. L. Major,
392 G. A. Gusarova, H. M. Yoder, R. H. Costa and V. V. Kalinichenko, "The Forkhead Box m1
393 transcription factor stimulates the proliferation of tumor cells during development of lung cancer,"
394 *Cancer Res*, vol. 66, no. 4, pp. 2153-2161, 2006.

395 29. O. A. Kalinina, S. A. Kalinin, E. W. Polack, I. Mikaelian, S. Panda, R. H. Costa and G. R.
396 Adami, "Sustained hepatic expression of FoxM1B in transgenic mice has minimal effects on
397 hepatocellular carcinoma development but increases cell proliferation rates in preneoplastic and
398 early neoplastic lesions," *Oncogene*, vol. 22, no. 40, pp. 6266-6276, 2003.

399 30. X. Song, S. S. Fiati Kenston, J. Zhao, D. Yang and Y. Gu, "Roles of FoxM1 in cell regulation
400 and breast cancer targeting therapy," *Med Oncol*, vol. 34, no. 3, pp. 41, 2017.

401 31. K. Krupczak-Hollis, X. Wang, V. V. Kalinichenko, G. A. Gusarova, I. C. Wang, M. B.
402 Dennewitz, H. M. Yoder, H. Kiyokawa, K. H. Kaestner and R. H. Costa, "The mouse Forkhead
403 Box m1 transcription factor is essential for hepatoblast mitosis and development of intrahepatic
404 bile ducts and vessels during liver morphogenesis," *Dev Biol*, vol. 276, no. 1, pp. 74-88, 2004.

405 32. S. Y. Tang, Y. Jiao and L. Q. Li, "[Significance of Forkhead Box m1b (Foxm1b) gene in cell

proliferation and carcinogenesis], " *Ai Zheng*, vol. 27, no. 8, pp. 894-896, 2008.

33. B. Xie, S. Wang, N. Jiang and J. J. Li, "Cyclin B1/CDK1-regulated mitochondrial bioenergetics in cell cycle progression and tumor resistance," *Cancer Lett*, vol. 443, pp. 56-66, 2019.

34. D. Santamaria, C. Barriere, A. Cerqueira, S. Hunt, C. Tardy, K. Newton, J. F. Caceres, P. Dubus, M. Malumbres and M. Barbacid, "Cdk1 is sufficient to drive the mammalian cell cycle," *Nature*, vol. 448, no. 7155, pp. 811-815, 2007.

35. A. Crncec and H. Hochegger, "Triggering mitosis," *FEBS Lett*, vol. 593, no. 20, pp. 2868-2888, 2019.

36. M. Malumbres and M. Barbacid, "Cell cycle, CDKs and cancer: a changing paradigm," *Nat Rev Cancer*, vol. 9, no. 3, pp. 153-166, 2009.

37. C. X. Wu, X. Q. Wang, S. H. Chok, K. Man, S. H. Y. Tsang, A. C. Y. Chan, K. W. Ma, W. Xia and T. T. Cheung, "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*, vol. 8, no. 14, pp. 3737-3750, 2018.

- 422 **Description of the supplementary tables**
- 423 Table S1. Integration of DEGs
- 424 Table S2: GO enrichment terms of integrated DEGs.
- 425 Table S3. KEGG pathway analysis of integrated DEGs.