Original Article Identifying drug candidates for hepatocellular carcinoma based on differentially expressed genes

Jiyuan Xing^{1,2*}, Qingmiao Shi^{1,2*}, Junjie Zhao³, Zujiang Yu^{1,2}

¹Gene Hospital of Henan Province, Precision Medicine Center, Departments of ²Infectious Diseases, ³Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China. ^{*}Equal contributors.

Received January 29, 2020; Accepted May 23, 2020; Epub June 15, 2020; Published June 30, 2020

Abstract: The prognosis for patients with advanced hepatocellular carcinoma (HCC) is extremely poor, mainly due to rapid progression and a paucity of effective drugs. Genome-wide analysis allows for potential drugs to be explored based on differentially expressed genes (DEGs). However, drug candidates and DEGs in HCC are largely unknown. In this study, we investigated DEGs and prognostication using The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC), the Gene Expression Omnibus (GEO), and immunohistochemical staining. Protein-protein interaction networks between DEGs were also analyzed to clarify 12 hub genes and query online databases for potential HCC therapeutic drugs. We found that 885 of 3219 DEGs from a TCGA dataset were associated with prognosis. We clarified 12 hub genes that were overexpressed in tumor samples and significantly associated with poor overall survival (OS) in HCC patients. These findings were validated using GEO and ICGC cohorts. Moreover, promising drug candidates targeted against HCC were predicted using online databases. Collectively, the upregulation of 12 hub genes was associated with poor prognosis for patients with HCC, and focusing on their expression may advance efforts towards targeted HCC therapies.

Keywords: Hepatocellular carcinoma, hub genes, differential gene expression, prognosis

Introduction

Hepatocellular carcinoma (HCC) is a global lifethreatening malignancy associated with an extremely unfavorable prognosis [1]. Globally, over one million patients are diagnosed with new HCC annually. The recurrence rate of HCC after surgery is 60-80% and the five-year survival rate of liver transplants is < 80%, primarily due to various risk factors such as late diagnosis, early metastasis, and inevitable recurrence [2, 3]. Despite the development of treatment strategies for advanced HCC, its clinical prognosis remains poor and effective therapeutic drugs are limited [4-6]. Thus, an understanding of the molecular processes and mechanisms underlying HCC are critically important for the discovery of novel treatments.

Drug development, especially for cancer, relies on the identification of novel drug targets [7]. Recent innovations in biotechnology show promise for enhancing our understanding of disease biology, providing new targets, and promoting a new era of drug development. However, despite these technological advances, most attention has been directed towards generating methods to identify new cancer treatment targets [8, 9]. Although many potential targets have emerged from screening, identifying those that are actually involved in specific diseases has proven challenging.

L1000 fireworks display (L1000FWD) is an online tool enabling the interactive visualization of > 16,000 drug and gene expression signatures [8]. Potential drugs or bioactive small molecules are easily identified by searching this database for a set of differentially expressed genes (DEGs) associated with a specific disease. The Drug Gene Interactions Database (DGIdb) is a drug reuse application that contains > 40,000 genes that participate in over 15,000 drug interactions [10, 11]. These drug interactions were collected using the Drug Expert Management and Text Mining Database DrugBank, the Therapeutic Targets Database, Pharmacological Guidelines, and ClinicalTrials. gov. The Connectivity Map (CMap; http://portals.broadinstitute.org/cmap/) is another online tool for calculating drug reuse, which allows users to screen small molecules with biological activity from transcriptome expression data [12, 13]. The current version includes the expression of > 7,000 genes from five human cell lines responding to 1,309 active biomolecules at various doses. In addition, pharmacologically-related genes can be classified using bioinformatic strategies such as gene ontology and pathway analyses.

The present study analyzed DEGs from several HCC datasets and non-tumor samples including The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), and the International Cancer Genome Consortium (ICGC). Key hub genes were also selected by analyzing their protein-protein interactions (PPIs) and evaluating their expression and potential value for prognostication. We further investigated compounds using predictive informatics that may serve as potential therapeutics against HCC.

Material and methods

Datasets

To analyze mRNA expression in HCC, we downloaded the most recent expression data and clinical follow-up information about HCC samples from The Cancer Genome Atlas Project (TCGA; http://tcga-data.nci.nih.gov/) datasets, containing the expression profiles of 424 RNAseq samples, as described previously [14, 15]. Microarray data for GSE14520, GSE22058, and GSE36376 were acquired from the Gene Expression Omnibus (GEO; www.ncbi.nlm.nih. gov/geo). The GSE14520 dataset contains 445 samples, and the GSE22058 and GSE36376 datasets comprise 197 and 433 expression profiles, respectively. We also downloaded ICGC data from the HCCDB website that contains 389 expression profiles.

Functional and pathway enrichment analyses

Functional information was obtained and DEG pathway enrichment proceeded using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) [16]. Enriched pathways in TCGA normal and tumor samples were explored using Gene Set Enrichment Analysis (GSEA, version v4.0.2). Protein-protein interaction (PPI) networks of DEGs were determined using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/). The results generated by PPI analysis were filtered to select hub genes using Cytoscape (Version: 3.7.2), and the first 12 genes were selected using the degree algorithm of the plug-in cytoHubba.

Identification of candidate drugs

Potential drugs for HCC therapy were identified using L1000 Fireworks Display (L1000 FWD; http://amp.pharm.mssm.edu/L1000FWD), the Drug Gene Interaction Database (DGIdb; http:// www.dgidb.org) and the Connectivity Map (CMap; https://portals.broadinstitute.org/cmap/). A search of the NIH database (https://pubchem.ncbi.nlm.nih.gov/compound) revealed the pharmacological effects of various drugs.

Statistical analysis

Results are shown as the mean \pm standard deviation. Data were statistically analyzed using SPSS 23.0 software (SPSS, Inc., Chicago, IL, USA). Two groups of normalized data were compared using Student's t-tests. Hub gene expression and survival were compared using Kaplan-Meier curves and log-rank tests. Values with p < 0.05 (two-sided) were regarded as statistically significant.

Results

Identification and screening signature genes related to HCC

RNA-seg data was analyzed between normal liver and tumor samples derived from the TCGA database and identified DEGs based on a threshold false discovery rate (FDR) of < 1e-5 and $|\log_{2}(fold change)| > 1$ shared after filtering. We found 3229 and 926 genes that were respectively upregulated and downregulated among 3219 tumors compared with normal samples (Supplementary Table 1). A volcano map of the DEGs showed that C2 and C1 were mainly downregulated and differentially expressed (Figure 1A). We then filtered and removed samples without information about overall survival (OS) status, or with information derived from patients who did not survive or survived for < 30 days. Univariate Cox proportional hazards regression models that included filtered expression profile data for all differential gene and survival data selected 885 DEGs, including 774 and 111 that were expressed at high and low levels, respectively (Figure 1B). Moreover, 801 and 84 genes were considered as risk (hazard ratio [HR] > 1) or protective (HR < 1) factors, respectively. Collectively, our findings provide a potential strategy with which to determine the functional roles of DEGs in HCC.



Figure 1. Map of differentially expressed genes. A. Heat map shows profiles of 885 DEGs in normal and tumor samples. Red color represents high DEG expression, while blue color represents low expression. B. Volcano map of TCGA-LIHC showing DEGs. Tumor samples contained 2293 and 926 upregulated and downregulated genes, respectively, compared with normal samples. Red color indicates upregulation of genes while blue color indicates downregulation. DEG, differentially expressed gene.



Figure 2. Annotations of differentially expressed genes. A. Gene Ontology annotation diagram of DEGs. These genes are related to mitotic nuclear division, DNA replication, organelle fission, and nuclear division. B. KEGG pathway annotation diagram of DEGs. Cell cycle, DNA replication, fatty acid degradation, tyrosine metabolism, drug metabolism-cytochrome P450, and mismatch repair were significantly related to DEGs. DEGs, differentially expressed genes.

Functional analysis of differentially expressed genes

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) functional enrichment analysis of the 885 DEGs using the R package cluster profiler revealed gene enrichment for 576 GO annotations, including 328 that were significantly different. Among them, 468 GO annotations associated with biological process (BP) had 283 showing significant differences, 66 annotations associated with cellular component (CC) had 34 showing significant differences, and 42 annotations associated with molecular function (MF) had 11 showing significant differences. The GO enrichment results indicated that these genes were related to mitotic nuclear division, DNA replication, organelle fission, and nuclear division (**Figure 2A**), all which are closely associated with



Figure 3. Identification of 12 hub genes. A. Protein-protein interaction network among 12 hub genes. Darker colors indicate higher scores in degree algorithm, and thus more importance. B. Differences in the expression of 12 hub genes in normal and HCC samples from TCGA cohort. ****denotes P < 0.0001.

tumorigenesis. We found 28 enriched KEGG pathways, 15 pathways of which were related to DEGs, exhibited significant differences between normal and tumor samples, and were associated with the cell cycle, DNA replication, fatty acid degradation, tyrosine metabolism, drug metabolism-cytochrome P450, and mismatch repair (**Figure 2B**). Taken together, these data suggest that hub gene overexpression might play a pivotal role in the progression and development of human HCC.

Protein-protein interaction and hub gene analysis of differentially expressed genes

We next analyzed the protein-protein interaction (PPI) network of 885 DEGs using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING). The resulting file was screened using Cytoscape to find hub genes. We selected CDK1, CCNB1, CDC20, AURKA, AURKB, BUB1, BUB1B, PLK1, CCNB2, KIF11, TOP2A, and CDCA8 as hub genes using the degree algorithm of the cytoHubba plug-in. Figure 3A shows the relationships among these hub genes according to PPI. We then compared the expression of these 12 hub genes in TCGA normal and tumor samples and found that all of them were significantly overexpressed in tumor samples (Figure 3B). These findings indicate that the 12 hub genes play crucial roles in regulating the progression of human HCC, and that the increased expression of these hub genes are independent predictors of OS.

Immunohistochemical verification of hub gene expression

We further investigated and verified the expression of hub genes using immunohistochemical

(IHC) data from the Human Protein Atlas. IHC data in Figure 4A shows that 10 (CDK1, CCN-B1, CDC20, AURKA, AURKB, PLK1, CCNB2, KIF11, TOP2A, and CDCA8) of the 12 hub genes were expressed in normal and tumor tissues. We then verified our findings using GSE14520. GSE22058, and GSE36376 microarrays and ICGC data. All 12 hub genes were expressed at high levels, and the trend was the same in four independent datasets (Figure 4B-D). Among these hub genes, 11 were consistent across the GSE14520, GSE22058, and GSE36376 microarrays and TCGA data (without CDK1 data). The ICGC dataset, which comprises a dependent verification cohort including RNAseq data, was consistent with the TCGA dataset (Figure 4E). Collectively, the consistent expression profiles of the 12 hub genes across different datasets are suggestive of important roles in tumorigenesis.

Relationship between hub genes and overall survival

We then evaluated associations between the expression of the 12 hub genes and OS. Kaplan-Meier analysis showed that high expression of the 12 hub genes significantly associated with poor HCC patient prognosis (**Figure 5**). Univariate Cox proportional hazards regression analysis showed that the HR for all 12 hub genes was > 1 (P < 0.01), indicating that they could serve as risk factors (**Tables 1** and **2**). We investigated whether these hub genes had prognosticating value using the independent ICGC dataset. <u>Supplementary Figure 1</u> shows that the 12 hub genes clearly divided the HCC samples into high and low risk groups with an HR of > 1 by univariate Cox analysis, indicating



Figure 4. Differential expression of 12 hub genes in normal and HCC samples. (A) Protein expression of CDK1, CCNB1, CDC20, AURKA, AURKB, PLK1, CCNB2, KIF11, TOP2A, and CDCA8 genes in normal and HCC samples. Samples from the ICGC (B), GSE14520 (C), GSE22058 (D), and GSE36376 (E) cohorts. ****denotes P < 0.0001.



Figure 5. Kaplan-Meier curves of hub gene expression and overall survival. Expression levels of CDK1 (A), CCNB1 (B), CDC20 (C), AURKA (D), AUKRB (E), BUB1 (F), BUB1B (G), PLK1 (H), CCNB2 (I), KIF11 (J), TOP2A (K), and CDCA8 (L) and survival probability determined from TCGA-LIHC samples.

Gene Name	p value	HR	Lower 95% CI	Upper 95% Cl
CDK1	< 0.05	1.025204	1.015	1.035
CCNB1	< 0.05	1.017041	1.011	1.023
CDC20	< 0.05	1.009423	1.006	1.013
AURKA	< 0.05	1.008888	1.002	1.016
AURKB	< 0.05	1.014700	1.009	1.021
BUB1	< 0.05	1.084189	1.046	1.124
BUB1B	< 0.05	1.055835	1.026	1.086
PLK1	< 0.05	1.044065	1.027	1.061
CCNB2	< 0.05	1.031622	1.016	1.047
KIF11	< 0.05	1.039719	1.016	1.064
TOP2A	< 0.05	1.012318	1.005	1.019
CDCA8	< 0.05	1.042621	1.027	1.058

Table 1. Univariate Cox regression analysis of overallsurvival of 12 hub genes in TCGA cohort

Table 2. Univariate Cox regression analysis of overallsurvival of 12 hub genes in ICGC cohort

Gene Name	p value	HR	Lower 95% Cl	Upper 95% CI
CCNB1	< 0.05	2.220682	1.563	3.154
CDCA8	< 0.05	2.193199	1.561	3.080
CDK1	< 0.05	2.194094	1.499	3.212
AURKA	< 0.05	1.891931	1.411	2.536
AURKB	< 0.05	1.836473	1.373	2.457
CDC20	< 0.05	1.635094	1.275	2.097
TOP2A	< 0.05	1.677951	1.265	2.226
BUB1	< 0.05	2.056437	1.405	3.011
BUB1B	< 0.05	1.915102	1.356	2.705
CCNB2	< 0.05	1.771021	1.294	2.423
PLK1	< 0.05	1.777501	1.305	2.420
KIF11	< 0.05	1.713692	1.081	2.717



Figure 6. Annotated results of 12 hub genes in L1000FWD, DGIdb, and CMap databases.

that these genes were not only associated with tumorigenesis, but also prognosis.

Potential drug analysis of HCC-related signatures

Using the L1000FWD, a reverse screen of the 12 hub genes against the database identified 72 small molecules. We then analyzed interactions with tumorigenesis and significant relationships with prognosis among the 12 hub genes, and found 265 drugs using the Drug-Gene Interactions Database (DGIdb). Furthermore, we screened 147 drugs through the CMap database that could be associated with tumorigenesis and prognosis using the 12 hub genes. Additionally, we compared potential drug overlap among L1000FWD, DGldb, and CMap annotations. Results showed that the small molecules amsacrine, etoposide, idarubicin, teniposide, tozasertib, and vincristine overlapped between L1000 and DGldb, podophyllotoxin overlapped between L1000 and CMap, and camptothecin, doxorubicin, daunorubicin, and mitoxantrone overlapped between DGIdb and CMap (Figure 6). These findings provide a potential basis for defining therapeutic agents against HCC.

Discussion

Previous studies have suggested that DEGs contribute to the initiation and progression of cancers, including HCC. More importantly, some DEGs have been associated with patient progno-

sis. The roles of DEGs in HCC should be clarified, as they might provide a theoretical basis for the personalized treatment of HCC.

Here, univariate regression analysis models selected 3219 DEGs using the TCGA dataset and associated 885 genes with prognosis. These findings were consistent with those of Guan et al., who found that DEGs and miRNA could be promising biomarkers during tumor development and play essential roles in tumor progression for clear cell renal cell carcinoma [17]. Wu et al. developed a model based on gene-gene co-regulation and changes in gene expression levels, and identified molecular differences between smokers and non-smokers with lung cancer [18]. These studies show that DEGs, especially those associated with prognosis, could serve as novel early diagnostic biomarkers.

We identified and clarified 12 hub genes using PPI analysis, and found through the TCGA, GEO, and IHC datasets that their expression was upregulated. These hub genes also significantly associated with the prognosis of HCC patients, which we further verified using independent cohorts. Additionally, we evaluated relationships between the hub genes CDK1, CCNB1, CCNB2, CDC20, AURKA, AURKB, BUB1, BUB1B, PLK1, KIF11, TOP2A, and CDCA8, and revealed their interconnection and close association with HCC occurrence and treatment potential. All of these genes encoded proteins that were expressed at high levels in the PPI network. Wang et al. also focused on DEGs in HCC samples compared with normal controls, and found that the hub genes CDK1, CCNB1, and CCNB2 were equally expressed [19], while Chu et al. significantly related the hub genes CDK1, CCNB1, and BUB1B with poor prognosis in HCC. Our data were consistent with these findings. Chen et al. associated elevated AURKA expression in HCC samples with pathological stage and metastasis [20]. Rahul et al. found that 59 genes associated with kinases were overexpressed in HCC, and BUB1B, AURKB, and CDK1 in the present study agreed with their results [21]. Similarly, increased CDC20 [22], PLK1 [23], KIF11 [24], TOP2A [25], and CDCA8 [26] expression correlated with HCC tumor development and progression. Thus, we speculated that hub genes with high expression levels could serve as new biomarkers for prognostication, and that targeting hub genes could become a novel strategy for treating HCC.

Using the public drug databases L1000FWD, DGIdb, and CMap, which contain a substantial amount of gene expression profiles including small molecule effects on gene expression under pathological conditions, we identified a set of small molecules associated with antitumorigenesis activity that have therapeutic potential. Annotated results derived from these databases selected six, one, and four small molecules that respectively overlapped between L1000 and DGIdb, L1000 and CMap, and DGIdb and CMap, respectively.

Idarubicin is highly cytotoxic against HCC and is much more lipophilic than doxorubicin, indicat-

ing better penetration through cellular lipidic membranes [27]. Furthermore, a greater accumulation of the drug in lipiodol can overcome multidrug resistance (MDR). Tozasertib is a neuroprotective agent for injured optic nerves that can inhibit neuronal apoptosis and improve outcomes after subarachnoid hemorrhage, possibly via DLK/JIP3/MA2K7/JNK pathways [28]. Tozasertib can independently target Aurora kinases and RIPK1; thus, analogues with increased selectivity to these targets could be generated [29]. Along with the present findings, these characteristics support the notion that tozasertib could become a novel anticancer agent. Song et al. showed that vincristine sulfate (VCR) is an M phase-specific chemotherapeutic agent that has been extensively applied to treat various cancers including acute leukemia, lymphosarcoma, and Kaposi sarcoma [30]. However, the poor OS of HCC is mainly ascribed to MDR due to poor chemosensitivity to VCR. Podophyllotoxin notably exerts potent antitumor activity against HCC in vitro and in vivo, but low water solubility and toxicity limit its application [31]. Common drugs such as daunorubicin have been widely applied for HCC treatment. Camptothecin is a novel NRF2 inhibitor that could be repurposed in combination with other chemotherapeutics to enhance the treatment of cancers expressing high levels of NRF2 [32]. Camptothecin markedly suppresses NRF2 expression in various cancer cell types, including HCC. Doxorubicin could serve as a promising nanoscale drug candidate for HCC therapy by inducing HCC cell apoptosis. Mitoxantrone is an antineoplastic drug that significantly and specifically suppresses the growth and proliferation of HCC cells in vitro [33].

Teniposide (VM-26) is a broad spectrum, semisynthetic derivative of podophyllotoxin that has proven effective against several types of human cancer, particularly cerebroma [34-36]. However, its application to clinical HCC is limited by poor solubility, and the specific molecular mechanisms imparting its antitumor activity require further exploration. Amsacrine is a topoisomerase II enzyme inhibitor that is effective against acute lymphocytic leukemia [37], and its analogs are novel chemotherapeutic agents with potentially broad antitumor activities [38]. The specific antitumor molecular mechanisms of amsacrine and its analogs in HCC should be further investigated [37]. Etoposide can notably overcome the increased resistance of HCC cells expressing HBV X protein (HBx) to chemotherapy [39]. Collectively, our findings support these drugs as novel anticancer agents that may prevent HCC recurrence.

Conclusion

In conclusion, we identified 12 hub genes with upregulated expression in HCC. These genes were significantly associated with poorer OS among patients with HCC. Based on these genes, we found several candidate small molecules that may serve as promising therapeutics against HCC. We believe that our findings offer evidence towards the development of novel drugs for effective HCC treatment.

Acknowledgements

This study was supported by funds from National S&T Major Project of China (2018ZX-10725506). And Dr. Xing is supported by scholarship from the Henan Health Scholarship Council.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Zujiang Yu, Department of Infectious Diseases, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China. Tel: +86-371-66271133; E-mail: yzjzzu@126.com; Junjie Zhao, Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China. Tel: +86-371-67966942; E-mail: zhaojunjiezzu@126. com

References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.
- [2] Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A and Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol 2019; 16: 589-604.
- [3] Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019; 380: 1450-1462.
- [4] Bruix J, da Fonseca LG and Reig M. Insights into the success and failure of systemic therapy for hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol 2019; 16: 617-630.

- [5] Llovet JM, Montal R, Sia D and Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 2018; 15: 599-616.
- [6] Ma Y, Zhao C, Zhao H, Li H, Chen C, Xiang H, Zheng C, Ma C, Luo C, Qiu H, Yao Y, Hu H, Xiong B, Zhou J, Zhu H and Long Q. Comparison of treatment efficacy and safety between drugeluting bead transarterial chemoembolization with CalliSpheres microspheres and conventional transarterial chemoembolization as firstline treatment in hepatocellular carcinoma patients. Am J Transl Res 2019; 11: 7456-7470.
- [7] Cao C and Moult J. GWAS and drug targets. BMC Genomics 2014; Suppl 4: S5.
- [8] Cheng AL, Kang YK, Chen Z, Tsao CJ, Qin S, Kim JS, Luo R, Feng J, Ye S, Yang TS, Xu J, Sun Y, Liang H, Liu J, Wang J, Tak WY, Pan H, Burock K, Zou J, Voliotis D and Guan Z. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, doubleblind, placebo-controlled trial. Lancet Oncol 2009; 10: 25-34.
- [9] Zhao W, Shan B, He D, Cheng Y, Li B, Zhang C and Duan C. Recent progress in characterizing long noncoding RNAs in cancer drug resistance. J Cancer 2019; 10: 6693-6702.
- [10] Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL and Griffith M. DGldb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res 2018; 46: D1068-D1073.
- [11] Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, Krysiak K, Pan D, McMichael JF, Eldred JM, Walker JR, Wilson RK, Mardis ER, Griffith M and Griffith OL. DGldb 2.0: mining clinically relevant druggene interactions. Nucleic Acids Res 2016; 44: D1036-1044.
- [12] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES and Golub TR. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006; 313: 1929-1935.
- [13] Wang Z, Dai Z, Luo Z and Zuo C. Identification of pyrvinium, an anthelmintic drug, as a novel anti-adipogenic compound based on the gene expression microarray and connectivity map. Molecules 2019; 24: 2391.
- [14] He Y, Xue C, Yu Y, Chen J, Chen X, Ren F, Ren Z, Cui G and Sun R. CD44 is overexpressed and correlated with tumor progression in gallbladder cancer. Cancer Manag Res 2018; 10: 3857-3865.

- [15] Xue C, He Y, Zhu W, Chen X, Yu Y, Hu Q, Chen J, Liu L, Ren F, Ren Z, Cui G and Sun R. Low expression of LACTB promotes tumor progression and predicts poor prognosis in hepatocellular carcinoma. Am J Transl Res 2018; 10: 4152-4162.
- [16] He Y, Dang Q, Li J, Zhang Q, Yu X, Xue M and Guo W. Prediction of hepatocellular carcinoma prognosis based on expression of an immunerelated gene set. Aging 2020; 12: 965-977.
- [17] Guan L, Tan J, Li H and Jin X. Biomarker identification in clear cell renal cell carcinoma based on miRNA-seq and digital gene expression-seq data. Gene 2018; 647: 205-212.
- [18] Wu C, Zhu J and Zhang X. Network-based differential gene expression analysis suggests cell cycle related genes regulated by E2F1 underlie the molecular difference between smoker and non-smoker lung adenocarcinoma. BMC Bioinformatics 2013; 14: 365.
- [19] Roberts WC. Major anomalies of coronary arterial origin seen in adulthood. Am Heart J 1986; 111: 941-963.
- [20] Chen C, Song G, Xiang J, Zhang H, Zhao S and Zhan Y. AURKA promotes cancer metastasis by regulating epithelial-mesenchymal transition and cancer stem cell properties in hepatocellular carcinoma. Biochem Biophys Res Commun 2017; 486: 514-520.
- [21] Agarwal R, Narayan J, Bhattacharyya A, Saraswat M and Tomar AK. Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets. Cancer Genet 2017; 216-217: 37-51.
- [22] Li J, Gao JZ, Du JL, Huang ZX and Wei LX. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. Int J Oncol 2014; 45: 1547-1555.
- [23] He Z, Deng W, Jiang B, Liu S, Tang M, Liu Y and Zhang J. Hsa-let-7b inhibits cell proliferation by targeting PLK1 in HCC. Gene 2018; 673: 46-55.
- [24] Chen J, Rajasekaran M, Xia H, Zhang X, Kong SN, Sekar K, Seshachalam VP, Deivasigamani A, Goh BK, Ooi LL, Hong W and Hui KM. The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/β-catenin signalling pathway. Gut 2016; 65: 1522-1534.
- [25] Wong N, Yeo W, Wong WL, Wong NL, Chan KY, Mo FK, Koh J, Chan SL, Chan AT, Lai PB, Ching AK, Tong JH, Ng HK, Johnson PJ and To KF. TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance. Int J Oncol 2009; 124: 644-652.

- [26] Yan H, Li Z, Shen Q, Wang Q, Tian J, Jiang Q and Gao L. Aberrant expression of cell cycle and material metabolism related genes contributes to hepatocellular carcinoma occurrence. Pathol Res Pract 2017; 213: 316-321.
- [27] Favelier S, Boulin M, Hamza S, Cercueil JP, Cherblanc V, Lepage C, Hillon P, Chauffert B, Krausé D and Guiu B. Lipiodol trans-arterial chemoembolization of hepatocellular carcinoma with idarubicin: first experience. Cardiovasc Intervent Radiol 2013; 36: 1039-1046.
- [28] Yin C, Huang GF, Sun XC, Guo Z and Zhang JH. Tozasertib attenuates neuronal apoptosis via DLK/JIP3/MA2K7/JNK pathway in early brain injury after SAH in rats. Neuropharmacology 2016; 108: 316-323.
- [29] Martens S, Goossens V, Devisscher L, Hofmans S, Claeys P, Vuylsteke M, Takahashi N, Augustyns K and Vandenabeele P. RIPK1dependent cell death: a novel target of the Aurora kinase inhibitor Tozasertib (VX-680). Cell Death Dis 2018; 9: 211.
- [30] Song XR, Zheng Y, He G, Yang L, Luo YF, He ZY, Li SZ, Li JM, Yu S, Luo X, Hou SX and Wei YQ. Development of PLGA nanoparticles simultaneously loaded with vincristine and verapamil for treatment of hepatocellular carcinoma. J Pharm Sci 2010; 99: 4874-4879.
- [31] Wang Y, Sun H, Xiao Z, Zhang D, Bao X and Wei N. XWL-1-48 exerts antitumor activity via targeting topoisomerase II and enhancing degradation of Mdm2 in human hepatocellular carcinoma. Sci Rep 2017; 7: 9989.
- [32] Chen F, Wang H, Zhu J, Zhao R, Xue P, Zhang Q, Bud Nelson M, Qu W, Feng B and Pi J. Camptothecin suppresses NRF2-ARE activity and sensitises hepatocellular carcinoma cells to anticancer drugs. Br J Cancer 2017; 117: 1495-1506.
- [33] Xie B, He X, Guo G, Zhang X, Li J, Liu J and Lin Y. High-throughput screening identified mitoxantrone to induce death of hepatocellular carcinoma cells with autophagy involvement. Biochem Biophys Res Commun 2020; 521: 232-237.
- [34] He S, Yang H, Zhang R, Li Y and Duan L. Preparation and in vitro-in vivo evaluation of teniposide nanosuspensions. Int J Pharm 2015; 478: 131-137.
- [35] Yusenko M, Jakobs A and Klempnauer KH. A novel cell-based screening assay for smallmolecule MYB inhibitors identifies podophyllotoxins teniposide and etoposide as inhibitors of MYB activity. Sci Rep 2018; 8: 13159.
- [36] He X, Xiang N, Zhang J, Zhou J, Fu Y, Gong T and Zhang Z. Encapsulation of teniposide into albumin nanoparticles with greatly lowered toxicity and enhanced antitumor activity. Int J Pharm 2015; 487: 250-259.

- [37] Sader S and Wu C. Computational analysis of Amsacrine resistance in human topoisomerase II alpha mutants (R487K and E571K) using homology modeling, docking and all-atom molecular dynamics simulation in explicit solvent. J Mol Graph Model 2017; 72: 209-219.
- [38] Fang YP, Chuang CH, Wu PC, Huang YB, Tzeng CC, Chen YL, Liu YT, Tsai YH and Tsai MJ. Amsacrine analog-loaded solid lipid nanoparticle to resolve insolubility for injection delivery: characterization and pharmacokinetics. Drug Des Devel Ther 2016; 10: 1019-1028.
- [39] Kim HY, Jung HU, Yoo SH, Yoo KS, Cheong J, Park BS, Yun I and Yoo YH. Sorafenib overcomes the chemoresistance in HBx-expressing hepatocellular carcinoma cells through downregulation of HBx protein stability and suppresses HBV gene expression. Cancer Lett 2014; 355: 61-69.



Supplementary Figure 1. Kaplan-Meier curves of hub gene expression and survival probability. Expression levels of CDK1 (A), CCNB1 (B), CDC20 (C), AURKA (D), AUKRB (E), BUB1 (F), BUB1B (G), PLK1 (H), CCNB2 (I), KIF11 (J), TOP2A (K), and CDCA8 (L) and survival probability determined from ICGC-LIHC-JP samples.