Identification of abnormally expressed genes and their roles in invasion and migration of hepatocellular carcinoma

Jia-Ning Zhang1,*, Xia Qin2,*, Xi-Gao Liu3,*, Xiao-Nan Ma4, Zhi-Ping Huang5, Wei Zhang4, Chong-Hai Zhang6, Sheng-Xian Yuan1, Wei-Ping Zhou1

1The Third Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Shanghai 200438, China  
2The Graduate School of Second Military Medical University, Shanghai 200433, China  
3Department of Urology, Qilu Hospital of Shandong University, Jinan, Shandong 250012, P.R. China  
4The First affiliated Hospital of Anhui Medical University, Hefei 230022, China  
5Department of Hepatobiliary surgery, Southern Theater General Hospital, Guangzhou 440104, China  
6People's Hospital, Xuecheng, Zaozhuang 277100, China  
*Equal contribution

Correspondence to: Sheng-Xian Yuan, Wei-Ping Zhou; email: yuanshengx@126.com, ehphwp@126.com  
Keywords: hepatocellular carcinoma, hub genes, AURKA, CDC20, invasion and migration  
Received: October 18, 2019  
Accepted: January 2, 2020  
Published: February 2, 2020

Copyright: Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Hepatocellular carcinoma (HCC) is the third leading cause of cancer death in the world, mainly attributable to its high migration tendency. Most patients with HCC miss the benefit of radical resection or liver transplantation because of distant metastasis of the tumor. Therefore, identification of diagnostic and therapeutic targets by analyzing abnormally expressed genes associated with tumor invasion and migration in HCC could be useful. In this study, we carried out a meta-analysis of three gene expression profile data sets (GSE45050, GSE84598 and GSE89377), including 112 liver cancer and 77 normal tissue samples, to identify candidate genes and pathways. The three gene expression omnibus (GEO) data sets had 155 differentially expressed genes (DEGs) in common, including 25 up-regulated and 129 down-regulated genes. Protein-protein interaction (PPI) networks were generated in the STRING database and explored further in Cytoscape to identify network hubs. Moreover, the statistically significant (p < 0.05) functions and signaling pathways enriched by the shared DEGs were identified. The Kaplan-Meier curve was applied to analyze univariate survival outcomes of the hub genes, which suggested AURKA and CDC20 as independent prognostic factors in HCC. Additionally, we found that AURKA and CDC20 were markedly upregulated in HCC cell lines. Quantitative PCR, western blot, and Immunohistochemistry (IHC) was performed to determine mRNA and protein expression of AURKA and CDC20 in 20 normal liver tissues and 66 HCC tissues from patients with HCC who underwent complete surgical resection in stage 1 to 4. Results showed that expression of AURKA and CDC20 were significantly higher in HCC tissues than normal tissues and were increased from HCC stage 1 to 4. Importantly, through loss of function, we showed that AURKA and CDC20 both significantly promote HCC migration and invasion. In summary, we identified the key candidate DEGs and dysregulated pathways in HCC through bioinformatic analysis and experimental validation. These candidate DEGs and pathways enhance our understanding of the potential pathogenesis of HCC and may hold promise as markers for the diagnosis, treatment, and prognosis of HCC.
INTRODUCTION

Hepatocellular carcinoma (HCC), a common malignant neoplasm in the digestive system, is the third leading cause of cancer mortality worldwide [1] and ranks second in China [2]. Treatment options for HCC are limited by its complexity, heterogeneity, high recurrence rate after surgical resection, and diagnosis being primarily dependent on serological changes during disease progression [3]. Therefore, it is critical to understand the causative molecular mechanisms underlying HCC and any associations between them, to identify novel biomarkers such as target genes that can be used for early diagnosis, disease progression, and personalized therapy.

Recently, RNA sequencing and bioinformatics have been extensively utilized to identify general genomic changes during cancer development and to discover biomarkers for the diagnosis and prognosis of tumors [4]. Gene expression microarrays have been used in many studies to identify differentially expressed genes (DEGs) and pathways associated with HCC [5]. A comprehensive as well as systematic study of the interactions between DEGs and the pathways enriched by them can accurately identify the biological changes that occur during the development of HCC.

With the increasing number of datasets generated by microarray and RNA sequencing technologies, Gene expression omnibus (GEO) has become a significant resource for bioinformatics [6]. Datasets GSE45050, GSE84598 and GSE89377 from GEO were selected for this study. The DEGs shared by the three gene expression datasets were identified using the R package ‘LIMMA’ from the bioconductor project [7] and Venn diagram software. Subsequently, functional annotation of DEGs by GO categories, including biological processes (BP), cellular components (CC), molecular function (MF), as well as KEGG (Kyoto Encyclopedia of Gene and Genomes) pathway enrichment analysis was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID). Thereafter, protein–protein interaction (PPI) networks were generated in the STRING database while Cytotype MCODE (Molecular Complex Detection) was used for the identification of significant modules. Ten hub genes were selected following exploration of the network with the Cytoscape plugin, cytoHubba. The Kaplan-Meier plotter online database was performed to obtain important prognostic information associated with these hub genes (P < 0.05) was studied using the online Kaplan Meier plotter. Expression level of hub DEGs between HCC samples and paired normal samples by gene expression profiling analysis (GEPIA) (P <0.05), 2 hub genes (AURKA and CDC20) were screened out. Although it has been reported that AURKA and CDC20 were overexpressed in HCC [8, 9], which was consistent with our results of qPCR, western blot, and IHC, the study whether the upregulation of AURKA and CDC20 is related to the invasion and migration of HCC still needs further elucidation. Unexpectedly, we found that knockdown of AURKA or CDC20 both attenuated the invasion and migration of HCC cell lines. In conclusion, our study is aimed to screen and identify key candidate genes and pathways via bioinformatic analysis and experimental validation of function.

RESULTS

Identification of DEGs in HCC

A total of 112 cases of HCC and 77 cases of normal liver tissue were analyzed in this study. The online tool, GEO2R, identified 902 DEGs in GSE45050, 3499 DEGs in GSE84598 and 491 DEGs in GSE89377. DEGs in two representative samples from each of the three expression profiling datasets are shown in Figure 1A–1C). A total of 155 DEGs were common to the three datasets, including 129 downregulated (logFC < 1) and 25 upregulated (logFC > 1) genes in the HCC samples (Table 1 and Figure 1D).

Gene ontology and KEGG pathway analysis of DEGs in HCC

DAVID was used for functional annotation and pathway enrichment analysis of identified DEGs. In GO analysis, DEGs were identified to be significantly enriched in BP such as cellular response to cadmium ion, oxidation-reduction process, epoxygenase P450 pathway, negative regulation of growth, and cellular response to zinc ion (Table 2), CC such as extracellular region, organelle membrane, extracellular exosome, extracellular space, and blood microparticle (Table 2), and MF terms such as monoxygenase activity, heme binding, serine-type endopeptidase activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (Table 2). KEGG pathways analysis showed that the DEGs were primarily enriched in metabolic pathways, complement and coagulation cascades, chemical carcinogenesis, retinol metabolism, and mineral absorption (Table 2).

Key candidate DEGs and pathways identified through protein–protein interaction network (PPI) and module analysis

The PPI network of DEGs generated in the online STRING database is shown in Figure 1E and the most distinctive module identified in Cytoscape is shown in Figure 1F. Thirty genes that matched the degree criteria...
(degree >10, that is the node had more than 10 interactions) were identified as hubs and the 10 most significant genes according to the node degree were SPP2, FETUB, C8A, KLKB1, F9, MBL2, PLG, PTGS2, AURKA, CDC20 (Figure 2A). The full names, abbreviations and physical functions for these hub genes are presented in Table 3. The network of the selected hub genes and their co-expression genes generated in the cBioPortal online platform is shown in Figure 2B. Hierarchical clustering showed that the hub genes could distinctly categorize tumor tissues and normal tissues (Figure 2C).

**Hub gene screening and survival analysis of the cohorts**

A correlation analysis between the expression profiles of hub genes and outcomes in liver cancer data sets from TCGA was carried out to study the clinical relevance of the hub genes in HCC. GEPIA analysis identified that HCC patients with genomic alterations in AURKA and CDC20 had poor overall and disease-free survival (Overall survival: AURKA \( P=0.00022 \), CDC20 \( P=3.8e^{-06} \); Disease-free survival: AURKA \( P=0.0012 \), CDC20 \( P=0.0026 \)) (Figure 3). Therefore, hub genes...
**Table 1.** All 155 commonly differentially expressed genes (DEGs) were detected from three profile datasets, including 129 down-regulated genes and 25 up-regulated genes in the HCC tissues compared to normal liver tissues.

| DEGs  | Genes name                  |
|-------|-----------------------------|
| Up-regulated | SPINK1  | RACGAP1  | AURKA  | SQUE  | MELK  | CDC20  | STMN1  | UBE2T  | CD34  |
|       | ASPM    | CCNB2    | PRC1    | MDK    | TOP2A  | RAD51AP1 | CAP2    | HMMR   | KIF20A|
|       | LCN2    | AKR1C3   | KIAA0101 | CDKN3  | NCAPG  | CENPF  | NUSAP1  |        |       |
|       | BBOX1   | SXCL14   | HSD17B2 | CYP39A1 | PROZ   | C8A    | MBL2    | ACSM3   | A2M    |
| Down-regulated | GHR     | CLEC1B   | BMT    | STEAP3 | SHBG   | DNASE1L3 | BCHE   | C8orf4 |        |
|       | ID1     | CRHB1    | F9     | SLC38A4 | DBH    | MT1M   | COLEC11 | SLC39A5 | SRD5A2 |
|       | EGR1    | ECM1     | FCN2    | KLKB1  | F11    | MT1G   | SLC22A1 | IL13RA2 | NDRG2  |
|       | FETUB   | CHST4    | PTGS2   | PCK1   | ADH1C  | CYP2C9  | GCGR    | CLEC4M  | RND3   |
|       | IL1RAP  | RDIH16   | TMEM27  | AFM    | CSRNP1 | THRSP  | CYP4A11 | AGXT2   | MT1X   |
|       | C7      | AADAT    | NNMT    | ACAD8  | APOF   | HSD17B13 | DPT    | SLC25A47 | MFAP4  |
|       | SLC10A1 | CYP1A2   | MT1E    | TTC36  | LPA    | CXCL2  | TAT    | CYP4F2  | COLEC10|
|       | LYVE1   | CYP2C18  | ASPDH   | FOS    | SERPINA11 | ALDH8A1 | NAT2    | AKR1D1  | PAMR1  |
|       | CXCL12  | GNMT     | ADH4    | OIT3   | GLYAT  | CETP   | SRPX    | MAT1A   | LECT2  |
|       | PLG     | SOCS2    | CYP2C8  | SPP2   | HA02   | MT1F   | PEMT    | KCN2    | LY6E   |
|       | CPS1    | TKFC     | FCN3    | GBA3   | PDGFRA | CLEC4G  | FXYD1   | FOSB   | CLR3   |
|       | MARCO   | MT1H     | PRSS8   | TDO2   | VPR1   | IGBP3  | PLAC8   | HAMP    | DCN    |
|       | CYP8B1  | FTCD     | TIMD4   | STAB2  | HGFAC  | ADGRG7 | CRISPLD2 | SLC22A10 | PZP    |
|       | SLC27A5 | GLS2     | C9      |        |        |        |        |        |        |

*AURKA* and *CDC20* may have an important impact on the progression of hepatocellular carcinoma. Moreover, to explore prognostic value of *AURKA* and *CDC20* expression profiles in more tumor types, we analyzed the expression of *AURKA* and *CDC20* in other types of tumors other than HCC and assessed the prognostic value based on TCGA cohort by the tumor-immune system interactions (TISIDB, [http://cis.hku.hk/TISIDB/index.php](http://cis.hku.hk/TISIDB/index.php)) online database. As shown in Supplementary Figure 1, the high expression of *AURKA* is not only associated with the poor prognosis of patients with hepaticellular carcinoma, but also associated with the poor prognosis of adrenocortical carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, brain lower grade glioma, and lung adenocarcinoma, etc. In addition, the high expression of *CDC20* is related to the poor prognosis of patients with adrenocortical carcinoma, kidney renal clear cell carcinoma, brain lower grade glioma, lung adenocarcinoma, mesothelioma, and pancreatic adenocarcinoma, etc (Supplementary Figure 2).

**Hub gene analysis**

Gene expression profiles of *AURKA* and *CDC20* were compared between HCC tumor samples and adjacent normal tissues in RNA-sequencing datasets from the TCGA database. The expression of *AURKA* and *CDC20* was significantly higher in 369 HCC tissues compared with 160 normal tissues (Figure 4A and 4D). Moreover, correlation analysis revealed that *AURKA* mRNA expression was remarkably correlated with tumor stages and grades (Figure 4B and 4C). Similarly, *CDC20* mRNA expression levels were also associated with clinical HCC stages and grades (Figure 4E and 4F). Overall, elevated expression of *AURKA* and *CDC20* mRNA showed a notable association with advanced clinicopathological parameters in HCC patients.

A total of 100 candidate genes with positive or negative correlation were identified using GSEA. Importantly, GSEA with the hallmark gene sets showed that the significantly enriched pathways for *AURKA* included cell cycle, DNA replication, endocytosis, P53 signaling pathway, progesterone mediated oocyte maturation, RNA degradation, spliceosome, and ubiquitin mediated proteolysis (Figure 5A–5H). A heatmap of the transcriptional expression profiles of the 100 significant genes is shown in Figure 5I. The enriched hallmark gene sets enriched in *CDC20* included base excision repair, cell cycle DNA replication, mismatch repair,
Table 2. GO and KEGG pathway enrichment analysis of DEGs in HCC samples.

| Term                              | Description                               | Count in gene set | P-value     |
|-----------------------------------|-------------------------------------------|-------------------|-------------|
| Gene Ontology                     |                                           |                   |             |
| GO:0071276                        | Cellular response to cadmium ion          | 7                 | 4.84E-09    |
| GO:0055114                        | Oxidation-reduction process               | 23                | 1.12E-08    |
| GO:0019373                        | Epoxide P450 pathway                      | 6                 | 3.88E-07    |
| GO:0045926                        | Negative regulation of growth             | 6                 | 5.23E-07    |
| GO:0071294                        | Cellular response to zinc ion             | 6                 | 5.23E-07    |
| GO:0005576                        | Extracellular region                      | 44                | 2.33E-12    |
| GO:0031090                        | Organelle membrane                        | 12                | 1.43E-10    |
| GO:0070062                        | Extracellular exosome                     | 45                | 1.32E-05    |
| GO:0004497                        | Monooxygenase activity                    | 9                 | 3.43E-08    |
| GO:0020037                        | Heme binding                              | 11                | 3.09E-07    |
| Biological pathway                |                                           |                   |             |
| hsa01100                          | Metabolic pathways                        | 35                | 1.38E-06    |
| hsa04610                          | Complement and coagulation cascades       | 9                 | 1.93E-06    |
| hsa05204                          | Chemical carcinogenesis                   | 8                 | 5.67E-05    |
| hsa00830                          | Retinol metabolism                        | 7                 | 1.34E-04    |
| hsa04978                          | Mineral absorption                        | 6                 | 2.01E-04    |
| hsa00120                          | Primary bile acid biosynthesis            | 4                 | 0.00113     |
| hsa00140                          | Steroid hormone biosynthesis              | 5                 | 0.00578     |

Figure 2. Protein-protein interaction network and analysis of the hub genes. (A) The 10 most important hub genes identified using the Cytoscape software plugin cytoHubba. (B) Genes co-expressed with hub genes identified by cBioPortal. Nodes with bold black outline represent hub genes. Nodes with thin black outline represent the co-expressed genes. (C) Hierarchical clustering heatmap of 10 most important hub genes was constructed depend on TCGA cohort. Red indicates that the expression of genes is relatively upregulated, green indicates that the expression of genes is relatively downregulated, and black indicates no significant changes in gene expression; gray indicates that the signal strength of genes was not high enough to be detected. Abbreviation: TCGA, The Cancer Genome Atlas.
Table 3. Functional roles of 10 hub genes.

| No. | Gene symbol | Full name                          | Function                                                                                                                                                                                                                                                                                                                                 |
|-----|-------------|------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | SPP2        | Secreted phosphoprotein 2          | Pathways: Hemostasis, organism-specific biosystem. GO: endopeptidase inhibitor activity and cellular protein metabolic process                                                                                      |
| 2   | FETUB       | Fetuin B                           | Pathways: cysteine-type endopeptidase inhibitor activity. GO: metallocendopeptidase inhibitor activity and binding of sperm to zona pellucida.                                                                                                                                   |
| 3   | C8A         | Complement C8 alpha chain          | Pathways: Amoebiasis, organism-specific biosystem and complement Activation, organism-specific biosystem GO: complement binding                                                                                                                                               |
| 4   | KLKB1       | Kallikrein B1                      | Pathways: Activation of Matrix Metalloproteinases, organism-specific biosystem. GO: protein binding and serine-type endopeptidase activity.                                                                                                                                   |
| 5   | F9          | Coagulation factor IX              | Pathways: Blood Clotting Cascade, organism-specific biosystem. GO: calcium ion binding and endopeptidase activity.                                                                                                                                                           |
| 6   | MBL2        | Mannose binding lectin 2           | Pathways: Complement and coagulation cascades, organism-specific biosystem. GO: calcium ion binding and mannose binding.                                                                                                                                                    |
| 7   | PLG         | Plasminogen                        | Pathways: Angiopoietin receptor Tie2-mediated signaling, organism-specific biosystem. GO: apolipoprotein binding and serine-type endopeptidase activity.                                                                                                                  |
| 8   | PTGS2       | Prostaglandin-endoperoxide synthase 2 | Pathways: GPCR Pathway and Matrix Metalloproteinases. GO: cytokine activity and protease binding.                                                                                                                                                         |
| 9   | AURKA       | Aurora kinase A                    | Pathways: APC/C-mediated degradation of cell cycle proteins, organism-specific biosystem. GO: ATP binding and histone serine kinase activity.                                                                                                                              |
| 10  | CDC20       | Cell division cycle 20             | Pathways: APC/C complex, organism-specific biosystem. GO: anaphase-promoting complex binding and enzyme binding.                                                                                                                                                        |

Figure 3. Univariate survival analysis of the hub genes was performed using the Kaplan-Meier curve. (A and C) The expression of AURKA gene showed significantly poorer DFS and OS in HCC samples (Log rank P <0.05). (B and D) The expression of the CDC20 gene showed a significantly poorer OS and DFS in HCC samples (Log rank P <0.05). Abbreviation: DFS, disease-free survival; OS, overall survival; HCC, hepatocellular carcinoma.
oocyte meiosis, pyrimidine metabolism, RNA degradation, and spliceosome (Figure 6A–6H) and the heatmap showing transcriptional expression profiles of the 100 significant genes are shown in Figure 6I.

**Abnormal expression of AURKA and CDC20 in human HCC cells lines and clinical samples**

To validate the results from bioinformatics analysis, we determined the expression of *AURKA* and *CDC20* in two human HCC cells (Huh7 and SK-Hep-1) as well as normal human hepatic cells (QSG-7701) by qRT-PCR. As shown in Figure 7A and 7B, *AURKA* and *CDC20* mRNA expression was markedly upregulated in both the HCC cell lines. This was further confirmed by western blot which showed high expression levels of *AURKA* and *CDC20* protein in the two HCC cell lines (Figure 7C). Furthermore, we collected HCC tissues (n=66) and nontumor samples (n=20) from patients with HCC who underwent surgical resection. According to Barcelona Clinic Liver Cancer staging system, these patients have been clinically divided into stage 1 (n = 20), stage 2 (n = 20), stage 3 (n = 20), and stage 4 (n = 6). We firstly validated the mRNA and protein levels of *AURKA* and *CDC20* in each stage of patients with HCC as well as normal liver tissues by qRT-PCR and western blot. As shown in Figure 8A–8C, with the development of liver cancer, the mRNA levels of *AURKA* and *CDC20* were 5-20 folds higher than that in nontumor tissues, and the protein expression of *AURKA* and *CDC20* were also significantly upregulated in HCC tissues than normal tissues. Then, IHC was applied to analyze the proteins levels and the localization of *AURKA* and *CDC20* in 5 paraffin-embedded normal liver tissues and 17 paraffin-embedded archived HCC tissue samples, which included 5 cases of stage 1, 5

---

**Figure 4. Transcriptional expression of AURKA and CDC20 in HCC tumor tissues and adjacent normal liver tissues.** (A) Transcriptional level of *AURKA* expression was found to be significantly higher in 369 HCC tissues compared with 160 normal tissues (p<0.0001). (B) Transcriptional expression of *AURKA* was significantly correlated with AJCC stages, patients in more advanced stages expressed higher levels of *AURKA* mRNA. (C) Transcriptional expression of *AURKA* was significantly correlated with ISUP grade, patients in more advanced grade score expressed higher levels of *AURKA* mRNA. Highest *AURKA* mRNA expression was found in stage 3 or grade 4. (D) Transcriptional level of *CDC20* expression was found to be significantly higher in 369 HCC tissues compared with 160 normal tissues (p<0.0001). (E) Transcriptional expression of *CDC20* was significantly correlated with AJCC stages, patients in more advanced stages tended to express higher levels of *CDC20* mRNA. (F) Transcriptional expression of *CDC20* was significantly correlated with ISUP grade, patients in more advanced grade score expressed higher levels of *CDC20* mRNA. Highest *CDC20* mRNA expression was found in stage 3 or grade 4.
cases of stage 2, 5 cases of stage 3, 2 cases of stage 4 tumor tissues. Consistent with the results of qRT-PCR and western blot, expression of AURKA and CDC20 were remarkably increased along with the progression of HCC from stage 1 to 4. In contrast, AURKA and CDC20 were weakly detected in 5 cases of nontumor tissues (Figure 8D). Moreover, both the AURKA and CDC20 were mainly localized in cytoplasm/membrane of malignant cells. These results were further confirmed by quantitatively analysis of the H-scores of AURKA and CDC20 staining (Figure 8E). Our results are consistent with previously reported studies: Li et al reported a higher level of CDC20 mRNA in three HCC cell lines (HepG2, SMMC-7721 and Huh7) as compared with the normal cell line (LO2) and in 14 pairs of clinical HCC tissues as compared with normal adjacent tissues from the same patients [9], whereas Jeng et al reported AUKRA overexpression in 61% of HCC cases and eight liver cancer cell lines [8], both of their studies are consistent with our results of qRT-PCR, western blot, and IHC, suggesting the accuracy of our experimental results. Taking together, our results suggest that AURKA and CDC20 are upregulated in HCC, thus suggesting that they may have an important role in the progression of HCC.

Figure 5. (A–H) significantly correlated genes and hallmarks gene sets enriched in HCC DEGs obtained by GSEA. A total of 100 significant genes were obtained by GSEA with positive and negative correlation. The most significant pathways included cell cycle, DNA replication, endocytosis, P53 signaling pathway, progesterone mediated oocyte maturation, RNA degradation, spliceosome, and ubiquitin mediated proteolysis. (I) Transcriptional expression profiles of the 100 most significant genes are shown as the heat map.
In vitro effects of AURKA and CDC20 on HCC migration and invasion

Although Li et al have investigated the effect of CDC20 on cell proliferation and cell cycle of HCC cell lines [9], whether AURKA and CDC20 have a role in the invasion and migration of HCC cells is still unknown. Therefore, we transfected Huh7 and SK Hep-1 cells with siRNA specific for AURKA (siAURKA) and CDC20 (siCDC20), at 50 nM concentration, to study the functions AURKA and CDC20 in HCC. As shown in Figure 7D and 7E, expression of AURKA was significantly silenced in Huh-7 and SK-Hep-1 cells by siAURKA-1 but not siAURKA-2, the CDC20 level was also remarkably down-regulated in Huh-7 and SK-Hep-1 cells by siCDC20-1 but not siCDC20-2. First, a wound healing assay was performed to investigate the roles of AURKA and CDC20 in the migration of HCC cell lines. The results presented in Figure 9 showed that knockdown of AURKA or CDC20 could markedly impair cell migration compared with normal control cells. Next, we conducted the transwell invasion assay to evaluate the roles of AURKA and CDC20 in invasion of HCC cells. As shown in Figure 10, knockdown of AURKA or CDC20 could markedly

Figure 6. (A–H) significantly correlated genes and hallmarks pathways in HCC obtained by GSEA. A total of 100 significant genes were obtained by GSEA with positive and negative correlation. The most involved significant pathways included base excision repair, cell cycle, DNA replication, mismatch repair, oocyte meiosis, pyrimidine metabolism, RNA degradation, and spliceosome. (I) Transcriptional expression profiles of the 100 significant genes are shown in the heat map.
inhibit cell invasion. All these results suggested that AURKA and CDC20 may be key factors in the invasion and migration of HCC. Therefore, targeting AURKA or CDC20 can be a promising strategy for treating HCC patients with intrahepatic invasion or distant metastasis, thus ameliorating the prognosis of patients with HCC.

DISCUSSION

HCC is one of the leading causes of cancer-induced death worldwide. Although remarkable advances have been made in investigating the epidemiology, risk factors, and molecular mechanisms of HCC, and translated to the clinic, the prevalence and mortality rate from HCC continue to rise considerably with large numbers of HCC patients in advanced stages [10, 11]. Similar to other solid tumors, the development of HCC is also characterized by abnormal genetic and epigenetic alterations. Numerous studies spanning across exome sequencing, transcriptome analysis, and genomic characterization have demonstrated that HCC tumors are heterogeneous at multiple molecular levels, with variable phenotypes and clinical outcomes [12–14]. Thus, it is imperative to identify a series of well-selected candidate genes and pathways associated with HCC for early diagnosis, prevention, and treatment. Bioinformatic approaches have led to the identification of several biomarkers in HCC, but few studies have verified bioinformatic analysis by experimentally.

Figure 7. The enhanced expression of AURKA and CDC20 in HCC cell lines and knockdown of AURKA and CDC20 in Huh-7 and SK Hep-1 by siRNAs. (A and C) The expression of AURKA in two HCC cell lines (Huh-7 and SK Hep-1) and one liver cell line (QSG-7701) was evaluated by qRT-PCR (A) and Western blots (C); (B and C) The expression of CDC20 in Huh-7, SK Hep-1 cells and QSG-7701 cells was evaluated by qRT-PCR (B) and Western blots (C); (D and E) qRT-PCR and Western blots shows the effects of siRNAs on expression of AURKA and CDC20 in Huh-7 and SK Hep-1 cells. *p<0.05, **p<0.01, ***p<0.001.
Figure 8. The increased expression of AURKA and CDC20 along with the progression of HCC from stage 1 to 4. (A and B) The expression of AURKA and CDC20 in 20 nontumor samples and 66 HCC tissues from patients with HCC who had been clinically divided into stage 1 (n = 20), stage 2 (n = 20), stage 3 (n = 20), and stage 4 (n = 6) by Barcelona Clinic Liver Cancer staging system. (C) Analysis of AURKA and CDC20 protein level in representative 2 normal liver tissues and 12 HCC tissues from patients in HCC stage 1 to 4 by western blots. (D) Representative images (200×) from immunohistochemistry staining analysis of AURKA and CDC20 expression in normal liver tissues and different stages of HCC tissues. (E) Quantification analyses of H-score of AURKA and CDC20 staining in normal liver tissues and different stages of HCC tissues. *p<0.05, **p<0.01, ***p<0.001.
In this study, we have analyzed three mRNA microarray datasets (GSE45050, GSE84598 and GSE89377) and identified a total of 155 DEGs, including 129 down-regulated and 25 up-regulated genes, in HCC. Functional annotation using DAVID identified that the DEGs in HCC were enriched in GO terms such as cellular response to cadmium ion, oxidation-reduction process, extracellular region, extracellular exosome, and heme binding, and KEGG pathways including metabolic pathways, complement and coagulation cascades, and chemical carcinogenesis. The PPI network of DEGs created in the STRING database followed by a module analysis using the Cytoscape software identified ten hub genes, including \(\text{SPP2, FETUB, C8A, KLKB1, F9, MBL2, PLG, PTGS2, AURKA, CDC20}\). The GEPIA analysis identified that HCC patients with genomic alterations in the hub genes AURKA and CDC20 had worse outcomes of overall and disease-free survival.

\textit{AURKA} belongs to the aurora kinase family, a major regulator of mitotic progression, which is upregulated in multiple human cancers [15–17]. \textit{AURKA} is required for centrosome maturation and mitotic commitment at the late G2 phase, especially for centrosome replication, bipolar spindle formation, chromosomal segregation, and cytokinesis [18–21]. Abnormal expression of \textit{AURKA} has been shown to cause chromosome instability and contribute to tumorigenesis [22, 23]. In addition, a non-canonical function of \textit{AURKA} in DNA duplication has also been demonstrated. \textit{AURKA} promotes the development of cancers by phosphorylating Akt, mammalian target of the rapamycin (mTOR), and mitogen-activated protein kinase (MAPK), and thereby activating epithelial–mesenchymal transition (EMT) reprogramming [24]. Moreover, overexpression of \textit{AURKA} has been observed to be correlated with advanced HCC grades and stages [8]. Additionally, \textit{AURKA} has been shown to activate EMT and reprogram tumor cell stemness, and thereby contribute to cancer metastases [25, 26]. Thus, the findings from these studies are consistent with the results of our study and suggest that \textit{AURKA} may indeed be a key regulator of HCC invasion and metastasis.

\textit{CDC20} (cell division cycle 20) encodes a regulatory protein which interacts with the anaphase-promoting complex/cyclosome (APC/C) in the cell cycle and is involved in development of solid tumors [27, 28].

![Figure 9. Knockdown of AURKA and CDC20 attenuated HCC cell motility.](image-url)

(A) Huh7 cells was transfected with NC and siAURKA-1, respectively. Wound healing assay was performed after 24h recovery period. (B) SK Hep-1 cells was transfected with NC and siAURKA-1, respectively. Wound healing assay was performed after 24h recovery period. (C) Quantification of motility from (A) and (B). (D) Huh7 cells was transfected with NC and siCDC20-1, respectively. Wound healing assay was performed after 24h recovery period. (E) SK Hep-1 cells was transfected with NC and siCDC20-1, respectively. Wound healing assay was performed after 24h recovery period. (F) Quantification of motility from (D) and (E). *p<0.05, **p<0.01, ***p<0.001.
is evidence to support that $CDC20$ is involved in the progression of a variety of malignancies including pancreatic cancer, oral squamous cell carcinoma, gastric carcinoma, cervical cancer [29–32]. Moreover, $CDC20$ was shown to be associated with cell proliferation and cell cycle in liver cancer cells [9]. However, the relationship between $CDC20$ and the metastasis of HCC cells is not yet well understood. Knockdown of both, $AURKA$ and $CDC20$ by siRNA was observed to attenuate invasion and migration of HCC cell lines. Thus, $AURKA$ and $CDC20$ can be potentially used as indicators of intrahepatic invasion or distant metastasis in HCC patients.

The aim of our study was to identify the key candidate DEGs and gene regulatory networks based on DEGs in HCC. We identified the hub genes associated with poor prognosis, $AURKA$ and $CDC20$, and explored their roles in migration and invasion of HCC cell lines. However, there are a few shortcomings in our study. First, the datasets utilized in this study are from the GEO database and not directly obtained from patient samples thus limiting our control on the experimental conditions and the quality of the results. Second, the identification of abnormally expressed genes associated with tumor invasion and migration may be best identified by comparing tissues with and without invasion and metastasis. Third, we have not explored the signaling mechanisms involving $AURKA$ and $CDC20$ and additional research will be needed to further elucidate the detailed mechanisms by which these key candidate DEGs modulate HCC.

In conclusion, this study has identified key candidate DEGs and hub genes associated with HCC tumors, enhancing our understanding of the underlying molecular mechanisms, and which can be potentially used for early diagnosis, treatment and prognosis of HCC. Additional research is imperative to clearly elucidate the roles hub genes in HCC.

**Figure 10. Knockdown of $AURKA$ and $CDC20$ attenuated HCC cell invasion.** (A) Huh7 and SK Hep-1 cells transfected with siAURKA-1 were less invasive as compared to control NC cells. (B) Huh7 and SK Hep-1 cells transfected with siCDC20-1 were less invasive as compared to control NC cells. (C) Quantification of invasion from (A). (D) Quantification of invasion from (B). *p<0.05, **p<0.01, ***p<0.001.
MATERIALS AND METHODS

Gene expression datasets

The three datasets analyzed in this study, GSE45050, GSE84598, and GSE89377, were downloaded from NCBI-GEO, a free public database of transcriptional expression datasets. GSE45050 data was obtained with the GPL6244 Platform (Affymetrix Human Gene 1.0 ST Array, with transcript gene version) and sourced from six liver tumors and ten non-tumor liver tissue samples (including five liver cirrhosis, two fatty changes of liver, and three normal liver tissues), GSE84598 data was based on the GPL10558 platform (Illumina HumanHT-12 V4.0 expression beadchip) and collected from 44 hepatocellular carcinoma and 22 non-tumor tissues. GSE89377 data was obtained from the GPL6947 platform (Illumina HumanHT-12 V3.0 expression beadchip), and consisted of 13 normal liver samples, 20 chronic hepatitis, 12 cirrhosis, and 62 liver tumors tissue samples.

Data preprocessing and identification of DEGs

Raw microarray data was preprocessed to eliminate noise ensure integrity. Background correction of probe data, standardization, and summarization was performed by the robust multiarray average analysis algorithm [33] in the affy package of R [34].

DEGs between HCC and non-cancerous samples were identified using GEO2R, an interactive web tool that allows users to compare two or more datasets in a GEO series in order to identify DEGs across experimental conditions. The adjusted P-values (adj. P) and Benjamini and Hochberg false discovery rate were applied to provide a balance between statistically significant genes and false-positives. Probe sets without corresponding gene symbols or genes with more than one probe set were removed or averaged, respectively. Absolute logFC > 1 and P value < 0.05 were considered statistically significant. The raw data in .TXT format were compared using an online Venn software to identify the DEGs common to the three datasets. DEGs with log FC < 0 were considered as down-regulated genes, while the DEGs with log FC > 0 were considered as up-regulated gene.

GO and KEGG pathway enrichment analyses of DEGs

DAVID (http://david.ncifcrf.gov) (version 6.7) is an online biological information database that integrates biological data and analysis tools, and provides a comprehensive set of functional annotation information of genes and proteins for users to extract biological information [35]. KEGG is a database resource for understanding advanced functions and biological systems from large-scale molecular data generated by high-throughput experimental techniques [36]. GO is a major bioinformatics tool for annotating genes and analyzing the biological processes of these genes [37]. We used DAVID to visualize the DEGs enrichment of BP, MF, CC and pathways (P < 0.05).

PPI network and module analysis

DEG-encoded proteins and the PPI were obtained using the online database STRING (available online: http://string-db.org) [38]. An interaction with a combined score > 0.4 was considered statistically significant. Cytoscape (version 3.7.1), a public source bioinformatics software platform, was used to visualize and analyze molecular interaction networks [39]. The plug-in MCODE (version 1.5.1) of Cytoscape can be used to identify the most dense and significant module in PPI networks [40]. MCODE was used with the following criteria: degree cut-off = 2, node score cut-off = 0.2, max depth = 100, and K-score = 2, to cluster the PPI network based on the topology to find densely connected regions.

Survival analysis

The Kaplan-Meier method was used to analyze differences in survival between groups. The primary endpoint was disease-free survival (DFS), the duration from the onset of curative treatment to the date of progression or the start date of the second-line treatment or the date of death, whichever occurs first. The overall survival (OS) as a secondary endpoint was the length of time from the date of diagnosis or first treatment to the date of death or last follow-up. The Kaplan-Meier method was used to estimate and account for the duration of follow-up, the 95% confidence interval (95% CI) and the log-rank test in the separation curve. The overall score was determined as the sum of the weights of each important central gene.

Selection of hub genes and analysis

The Cytoscape plug-in cytohubba (version 0.1) was used to identify hub genes based on degree thresholds. Genes in the network with degrees ≥10 (a node had 11 or more connections/interactions) were identified as hubs. The online platform cBioPortal (http://www.cbioportal.org) was used to identify the co-expressed genes [41]. ClueGO is a plug-in of Cytoscape that can visualize the nonredundant biological terms for large clusters of genes in a functionally grouped network [42]. The biological process from GO analysis of hub genes was performed and visualized by ClueGO.
Lipofectamine 3000 (Invitrogen, Shanghai, China) was transfected into Huh7 and SK-Hep-1 cells using the small interfering RNA (siRNA) was synthesized and transfected under the same conditions. Cell transfection 7701 cells were cultured in RPMI 1640 (Gibco, Shanghai, China) and 1% penicillin/streptomycin in a humidified incubator in an atmosphere of 5% CO2 at temperature of 37°C. QSG-7701 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM; Gibco, Shanghai, China), with 10% fetal bovine serum (FBS, Gibco, Shanghai, China). Cell culture Human HCC cells (Huh7 and SK-Hep-1) and normal human hepatic cells (QSG-7701) were used in this study. Huh7 and SK-Hep-1 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM; Gibco, Shanghai, China), with 10% fetal bovine serum (FBS, Gibco, Shanghai, China) and 1% penicillin/streptomycin in a humidified incubator in an atmosphere of 5% CO2 at temperature of 37°C. QSG-7701 cells were cultured in RPMI 1640 (Gibco, Shanghai, China) under the same conditions.

Cell transfection

The small interfering RNA (siRNA) was synthesized and transfected into Huh7 and SK-Hep-1 cells using Lipofectamine 3000 (Invitrogen, Shanghai, China) according to the manufacturer’s instructions. The sequences of AURKA specific siRNA1 (siAURKA-1) are: sense 5′-GAAUCUACUUUGGAUGACATT-3′ and antisense 5′-UAGUCCAGAAGUAAGUUCCTT-3′, siRNA2 (siAURKA-2) are: sense 5′-CAGUUAGGAGCUGUGCAATT-3′ and antisense 5′-UUGACAGCUCUUAAACUGTT-3′; and the sequences of CDC20 specific siRNA1 (siCDC20-1) are: sense 5′-GGAGCUCAUCUCAAGCCAUUU-3′ and antisense 5′-AUGGCCUGAGAUGAGCUCUUU-3′, siRNA2 (siCDC20-2) are: sense 5′-GGGAAUAUAUAUCUCUGUTT-3′ and antisense 5′-ACAGAGGAUAUAUUCCTT-3′. The corresponding scramble siRNA was used as negative control (Ribobio, China).

Gene set enrichment analysis (GSEA)

The TCGA database was used with the GSEA method using the Category version 2.10.1 package. For each separate analysis, Student’s-t-test statistical score was performed in consistent pathways and the mean of the differential expression genes was calculated. A permutation test with 1000 times was used to identify the significantly changed pathways. The adjusted P value (adj. P) using Benjamini and Hochberg (BH) false discovery rate (FDR) correction was applied by default to handle the false positive results [44]. An adj. P less than 0.01 and FDR less than 0.25 was considered to be statistically significant. Statistical analysis and graphical plotting were carried out using R software (Version 3.6.1).

Patients and tissue samples

20 normal liver tissues and 66 HCC tissues were obtained from patients with HCC who underwent surgical resection at the Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (Shanghai, China). According to Barcelona Clinic Liver Cancer staging system, these patients have been clinically divided into stage 1 (n = 20), stage 2 (n = 20), stage 3 (n = 20), and stage 4 (n = 6). Tissues frozen in liquid nitrogen were used for qRT-PCR and western blot, and tissues embedded in paraffin were used for IHC. Informed consent was obtained from all patients, and the study protocol was approved by our institutional review board of the Second Military Medical University.

Cell culture

Human HCC cells (Huh7 and SK-Hep-1) and normal human hepatic cells (QSG-7701) were used in this study. Huh7 and SK-Hep-1 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM; Gibco, Shanghai, China), with 10% fetal bovine serum (FBS, Gibco, Shanghai, China) and 1% penicillin/streptomycin in a humidified incubator in an atmosphere of 5% CO2 at temperature of 37°C. QSG-7701 cells were cultured in RPMI 1640 (Gibco, Shanghai, China) under the same conditions.

Cell transfection

The small interfering RNA (siRNA) was synthesized and transfected into Huh7 and SK-Hep-1 cells using Lipofectamine 3000 (Invitrogen, Shanghai, China) according to the manufacturer’s instructions. The sequences of AURKA specific siRNA1 (siAURKA-1) are: sense 5′-GAAUCUACUUUGGAUGACATT-3′ and antisense 5′-UAGUCCAGAAGUAAGUUCCTT-3′, siRNA2 (siAURKA-2) are: sense 5′-CAGUUAGGAGCUGUGCAATT-3′ and antisense 5′-UUGACAGCUCUUAAACUGTT-3′; and the sequences of CDC20 specific siRNA1 (siCDC20-1) are: sense 5′-GGAGCUCAUCUCAAGCCAUUU-3′ and antisense 5′-AUGGCCUGAGAUGAGCUCUUU-3′, siRNA2 (siCDC20-2) are: sense 5′-GGGAAUAUAUAUCUCUGUTT-3′ and antisense 5′-ACAGAGGAUAUAUUCCTT-3′. The corresponding scramble siRNA was used as negative control (Ribobio, China).

RNA extraction and quantitative-PCR (qPCR)

Total RNA was extracted from cells and clinical samples using Trizol reagent according to the manufacturer’s instructions (Invitrogen, Shanghai, China). A total of 2µg RNA was reverse transcribed into cDNA using the PrimeScript® RT reagent Kit (ThermoFisher, China). Quantitative real-time PCR (qRT-PCR) was performed using the LightCycler®96 (Roche diagnostics, Germany) with SYBR Green (Roche diagnostics, Germany). The 18S ribosomal RNA was used as the internal control. Primers of RNAs reverse transcription and q-PCR were purchased from Ribobio Co. Ltd (Guangzhou, China). The Primer Premier 5 software was performed to design PCR primers and the sequences were AURKA forward, 5′-GAGGTCCAAAACGTGTTCTCG-3′ and reverse, 5′-ACAGGATGAGGTACACTGGTTG-3′; CDC20 forward, 5′-GCACAGTTCGCGTTCGAGA-3′ and reverse, 5′-CTGGATTTGCCAGGAGTTCGG-3′; Rn18s forward, 5′-GGAGAGGGAGCCTGAGAAA-3′; and reverse, 5′-TTACAGGGCCTCGAAAGAG-3′. The 2−ΔΔCT formula was used to calculate relative expression of the RNA level of each sample compared with the reference sample.

Western blot

Tissues were fully grinded in liquid nitrogen and then cracked with appropriate amount of RIPA buffer (50mM Tris, pH 7.4, 150mM NaCl, 1%Trition-X-100, 1% sodium deoxycholate, 0.1% SDS, 2mM sodium pyrophosphate, 25mM β-glycerophosphate, 1mM EDTA, 1mM Na3VO4, and 0.5 µg/mL leupeptin). Cells were collected and lysed using RIPA buffer. Tissue or Cell lysates were separated by SDS-PAGE, and electrophoretically transferred to NC membranes (GE Health Care Life Science). After blocking using 5% nonfat milk in PBST, the membranes were incubated at 4°C overnight with indicated antibodies. Primary antibodies were from commercial sources: anti-AURKA (Abcam, ab13824),
anti-CDC20 (Abcam, ab26483), and anti-GAPDH (EpiZyme, Cat# LF205, China). The secondary antibodies against rabbit and mouse are diluted at 1:5000, and the membranes were incubated with secondary antibodies for 1 h and visualized by chemiluminescence.

**Immunohistochemistry (IHC) and scoring**

IHC for AURKA and CDC20 expression in normal liver tissues and HCC samples from patients in stage 1-4 was performed using standard methods. Briefly, paraffin sections (4 μm) were deparaffinised and antigen retrieval by citrate buffer. After blocking using bovine serum albumin (1%; BSA) in PBST, tissue sections were incubated with rabbit anti-AURKA or anti-CDC20 diluted 1:100 overnight at 4°C. The secondary poly-horseradish peroxidase (HRP) anti-rabbit IgG antibody (Abcam, China) was incubated in room temperature for 1 hour. The immunostained sections were double-blindly reviewed and scored by two experienced pathologists. Based on the H-score method, the percentage and staining intensity of positive cells were evaluated by semi-quantitative results. Five visual fields (× 200) were observed on each section, and the total number of cells and cells stained at each intensity were counted in each field. Positive staining intensity: 0 = colorless; 1 = light yellow; 2 = medium brown; 3 = dark brown. The H-score was calculated following the formula: (% of cells stained at intensity category 1x1) + (% of cells stained at intensity category 2x2) + (% of cells stained at intensity category 3x3). H-scores varied from 0 to 300 where 300 represented 100% of cells strongly stained (+++). High CDC20 expression was defined as staining H-scores of cells ≥200.

**Wound healing assays**

Cells were seeded in six-well plates at suitable density and cultured until 100% confluence. A yellow pipette tip was used to make a straight scratch, simulating a wound. Microscopy based photographs were taken at 0h or 24 h after wounding.

**Transwell assay**

The transwell invasion assay was performed using 24-well transwell chambers (Greiner bio-one, Switzerland). First, the upper ventricular surface of the basal membrane of transwell chamber was coated with Matrigel (BD Bioscience, USA). Thereafter, 2 × 10^5 cells were suspended in 0.2 mL serum-free medium and added to the inserts. The lower compartment was supplemented with 0.5 mL medium containing 20% FBS as a chemoattractant. After incubation for 48 h at 37 °C, the cells on the upper surface of the membrane were carefully removed using a cotton swab and cells on the lower surface were continuously fixed with 100% methanol and then stained with 0.1% crystal violet. Five random visual fields of 200× magnification of each insert were selected and counted for cell numbers under a light microscope (Olympus, Japan).

**Statistical analysis**

The results are shown as mean ± SD. Differences between two groups were estimated using unpaired Student’s t-test. A two-tailed value of P < 0.05 was considered as statistically significant.

**Ethics approval**

The Ethics Committee of The Third Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital approved the study.

**Abbreviations**

HCC: Hepatocellular carcinoma; PPI: Protein-protein interactions; OS: Overall survival; DFS: Disease-free survival; qRT-PCR: Real-time quantitative reverse transcription polymerase chain reaction; DEGs: Differentially expressed genes; GEO: Gene expression omnibus; DAVID: Database for annotation, visualization, and integrated discovery; BP: Biological process; CC: Cellular component; MF: Molecular function; KEGG: Kyoto Encyclopedia of Gene and Genomes; GEPIA: Gene expression profiling interactive analysis; MCODE: Molecular complex detection; GSEA: Gene set enrichment analysis; IHC: Immunohistochemistry.

**AUTHOR CONTRIBUTIONS**

The work presented here is a collaborative effort of all the authors. ZWP and YSX defined the research theme, guided the analyses, interpretation, and presentation. ZJN, QX, and LXG carried out the cell biology and molecular biology experiments, carried out the sequence alignment, and drafted the manuscript. MXN and HZP were involved in data collection, cohort validation, and drafting of the manuscript. ZW and ZCH helped to perform the statistical analysis. All the authors have read and approved the final manuscript.

**CONFLICTS OF INTEREST**

The authors declare no Conflicts of interest.

**FUNDING**

This work is supported by Grants from the National Natural Science Foundation of China (No. 81972575)
and Shanghai Science and Technology Commission (19QA1408700).

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011; 61:69–90. 
   https://doi.org/10.3322/caac.20107 PMID:21296855

2. Chen W, Zheng R, Zuo T, Zeng H, Zhang S, He J. National cancer incidence and mortality in China, 2012. Chin J Cancer Res. 2016; 28:1–11. 
   https://doi.org/10.3978/j.issn.1000-9604.2016.02.08 PMID:27041922

3. Yang N, Ekanem NR, Sakyi CA, Ray SD. Hepatocellular carcinoma and microRNA: new perspectives on therapeutics and diagnostics. Adv Drug Deliv Rev. 2015; 81:62–74. 
   https://doi.org/10.1016/j.addr.2014.10.029 PMID:25450260

4. Iizuka N, Oka M, Yamada-Okabe H, Nishida M, Maeda Y, Mori N, Takao T, Tamesa T, Tangoku A, Tabuchi H, Hamada K, Nakayama H, Ishitsuka H, et al. Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. Lancet. 2003; 361:923–29. 
   https://doi.org/10.1016/S0140-6736(03)12775-4 PMID:12648972

5. Li L, Lei Q, Zhang S, Kong L, Qin B. Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis. Oncol Rep. 2017; 38:2607–18. 
   https://doi.org/10.3892/or.2017.5946 PMID:28901457

6. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashovsky M, Marshall KA, Philippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013; 41:D991–95. 
   https://doi.org/10.1093/nar/gks1193 PMID:23193258

7. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. 2005; 21:2067–75. 
   https://doi.org/10.1093/bioinformatics/bti270 PMID:15657102

8. Jeng YM, Peng SY, Lin CY, Hsu HC. Overexpression and amplification of Aurora-A in hepatocellular carcinoma. Clin Cancer Res. 2004; 10:2065–71. 
   https://doi.org/10.1158/1078-0432.CCR-1057-03 PMID:15041727

9. Li J, Gao JZ, Du JL, Huang ZX, Wei LX. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. Int J Otolaryngol. 2013; 2014; 45:1547–55. 
   https://doi.org/10.3892/ijo.2014.2559 PMID:25069850

10. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007; 132:2557–76. 
    https://doi.org/10.1053/j.gastro.2007.04.061 PMID:17570226

11. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, Hancock J, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 2017; 3:524–48. 
   https://doi.org/10.1001/jamaoncol.2016.5688 PMID:27918777

12. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu; Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017; 169:1327–41.e23. 
   https://doi.org/10.1016/j.cell.2017.05.046 PMID:28622513

13. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebourissou S, Couchy G, Meiller C, Shinde J, Soyosouvahn F, Calatayud AL, Pinyol R, Pelletier L, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet. 2015; 47:505–11. 
   https://doi.org/10.1038/ng.3252 PMID:25822088

14. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, Watanabe G, Gabriel S, Friedman SL, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res. 2009; 69:7385–92. 
   https://doi.org/10.1158/0008-5472.CAN-09-1089 PMID:19723656

15. Kollareddy M, Dzubak P, Zheleva D, Hajduc M. Aurora kinases: structure, functions and their association with cancer. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2008; 152:27–33. 
   https://doi.org/10.5507/bp.2008.004 PMID:18795071

16. Giet R, Prigent C. Aurora/Ipl1p-related kinases, a new oncogenic family of mitotic serine-threonine kinases.
17. Carmena M, Earnshaw WC. The cellular geography of aurora kinases. Nat Rev Mol Cell Biol. 2003; 4:842–54. https://doi.org/10.1038/nrm1245 PMID:14625535

18. Roghi C, Giet R, Uzbekov R, Morin N, Cha train I, Le Guellec R, Couturier A, Dorée M, Philippe M, Prigent C. The Xenopus protein kinase pEg2 associates with the centrosome in a cell cycle-dependent manner, binds to the spindle microtubules and is involved in bipolar mitotic spindle assembly. J Cell Sci. 1998; 111:557–72. https://doi.org/10.1006/excr.2000.4903 PMID:9454730

19. Giet R, Prigent C. The Xenopus laevis aurora/Ip11p-related kinase pEg2 participates in the stability of the bipolar mitotic spindle. Exp Cell Res. 2000; 258:145–51. https://doi.org/10.1006/excr.2000.4903 PMID:10912796

20. Shannon KB, Salmon ED. Chromosome dynamics: new light on Aurora B kinase function. Curr Biol. 2002; 12:R458–60. https://doi.org/10.1016/S0960-9822(02)00945-4 PMID:12446569

21. Castro A, Vignon S, Bernis C, Labbé JC, Prigent C, Lorca T. The D-Box-activating domain (DAD) is a new proteolysis signal that stimulates the silent D-Box sequence of Aurora-A. EMBO Rep. 2002; 3:1209–14. https://doi.org/10.1093/embo-reports/kvf241 PMID:12446569

22. Fu J, Bian M, Jiang Q, Zhang C. Roles of Aurora kinases in mitosis and tumorigenesis. Mol Cancer Res. 2007; 5:1–10. https://doi.org/10.1158/1541-7786.MCR-06-0208 PMID:17259342

23. D’Assoro AB, Haddad T, Galanis E. Aurora-A Kinase as a Promising Therapeutic Target in Cancer. Front Oncol. 2016; 5:295. https://doi.org/10.3389/fonc.2015.00295 PMID:26779440

24. Tsunematsu T, Arakaki R, Yamada A, Ishimaru N, Kudo Y. The Non-Canonical Role of Aurora-A in DNA Replication. Front Oncol. 2015; 5:187. https://doi.org/10.3389/fonc.2015.00187 PMID:26380219

25. Chou CH, Yang NK, Liu TY, Tai SK, Hsu DS, Chen YW, Chen YJ, Chang CC, Tzeng CH, Yang MH. Chromosome instability modulated by BM11-AURKA signaling drives progression in head and neck cancer. Cancer Res. 2013; 73:953–66. https://doi.org/10.1158/0008-5472.CAN-12-2397 PMID:23204235

26. D’Assoro AB, Liu T, Quatraro C, Amato A, Opyrchal M, Leontovich A, Ikeda Y, Ohmine S, Lingle W, Suman V, Ecsey J, Iankov I, Di Leonardo A, et al. The mitotic kinase Aurora—a promotes distant metastases by inducing epithelial-to-mesenchymal transition in ERα(+) breast cancer cells. Oncogene. 2014; 33:599–610. https://doi.org/10.1038/onc.2012.628 PMID:2334326

27. Fung TK, Poon RY. A roller coaster ride with the mitotic cyclins. Semin Cell Dev Biol. 2005; 16:335–42. https://doi.org/10.1016/j.semcdb.2005.02.014 PMID:15840442

28. Weinstein J, Jacobsen FW, Hsu-Chen J, Wu T, Baum LG. A novel mammalian protein, p55CDC, present in dividing cells is associated with protein kinase activity and has homology to the Saccharomyces cerevisiae cell division cycle proteins Cdc20 and Cdc4. Mol Cell Biol. 1994; 14:3350–63. https://doi.org/10.1128/MCB.14.5.3350 PMID:7513050

29. Chang DZ, Ma Y, Ji B, Liu Y, Hwu P, Abbruzzese JL, Logsdon C, Wang H. Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression. J Hematol Oncol. 2012; 5:15. https://doi.org/10.1186/1756-8722-5-15 PMID:22475664

30. Kim JM, Sohn HY, Yoon SY, Oh JH, Yang JO, Kim JH, Song KS, Rho SM, Yoo HS, Kim YS, Kim JG, Kim NS. Identification of gastric cancer-related genes using a cDNA microarray containing novel expressed sequence tags expressed in gastric cancer cells. Clin Cancer Res. 2005; 11:473–82. PMID:15701830

31. Espinosa AM, Alfaro A, Roman-Basaure E, Guardado-Estrada M, Palma Í, Serralde C, Medina I, Juárez E, Bermúdez M, Márquez E, Borges-Ibáñez M, Muñoz-Cortez S, Alcántara-Vázquez A, et al. Mitosis is a source of potential markers for screening and survival and therapeutic targets in cervical cancer. PLoS One. 2013; 8:e55975. https://doi.org/10.1371/journal.pone.0055975 PMID:23405241

32. Ouellet V, Guyot MC, Le Page C, Filali-Mouhim A, Mes-Masson AM. Tissue array analysis of expression microarray candidates identifies markers associated with tumor grade and outcome in serous epithelial ovarian cancer. Int J Cancer. 2006; 119:599–607. https://doi.org/10.1002/jjc.21902 PMID:16572426
normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. 
https://doi.org/10.1093/biostatistics/4.2.249
PMID:12925520

34. Zhong X, Li H. Score tests of genetic association in the presence of linkage based on the additive genetic gamma frailty model. Biostatistics. 2004; 5:307–27. 
https://doi.org/10.1093/biostatistics/5.2.307
PMID:15054033

35. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007; 8:R183. 
https://doi.org/10.1186/gb-2007-8-9-r183
PMID:17784955

36. Kanehisa M. The KEGG database. Novartis Found Symp. 2002; 247:91–101; discussion 101–3, 119–28, 244–52. 
PMID:12539951

37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al, and The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. Nat Genet. 2000; 25:25–29. 
https://doi.org/10.1038/75556
PMID:10802651

38. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43:D447–52. 
https://doi.org/10.1093/nar/gku1003
PMID:25352553

39. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011; 27:431–32. 
https://doi.org/10.1093/bioinformatics/btq675
PMID:21149340

40. Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson. 2012; 14:83. 
https://doi.org/10.1186/1532-429X-14-83
PMID:23199362

41. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401–04. 
https://doi.org/10.1158/2159-8290.CD-12-0095
PMID:22588877

42. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009; 25:1091–93. 
https://doi.org/10.1093/bioinformatics/btp101
PMID:19237447

43. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics. 2013; 29:661–63. 
https://doi.org/10.1093/bioinformatics/btt019
PMID:23325622

44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–50. 
https://doi.org/10.1073/pnas.0506580102
PMID:16199517
Supplementary Figure 1. (A) The expression and prognostic value of AURKA in tumors based on TCGA cohort. (B–G) High expression of AURKA is associated with the poor prognosis of patients with hepatocellular carcinoma, adrenocortical carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, brain lower grade glioma, and lung adenocarcinoma.
Supplementary Figure 2. (A) The expression and prognostic value of CDC20 in tumors based on TCGA cohort. (B–G) High expression of CDC20 is related to the poor prognosis of patients with adrenocortical carcinoma, kidney renal clear cell carcinoma, brain lower grade glioma, lung adenocarcinoma, mesothelioma, and pancreatic adenocarcinoma.