Identification of crucial genes based on expression profiles of hepatocellular carcinomas by bioinformatics analysis

Ze-Kun Liu*, Ren-Yu Zhang*, Yu-Le Yong, Zhi-Yun Zhang, Can Li, Zhi-Nan Chen and Huijie Bian

Fourth Military Medical University, Department of Cell Biology, National Translational Science Center for Molecular Medicine, Xi'an, Shaanxi, China
* These authors contributed equally to this work.

ABSTRACT

Hepatocellular carcinoma (HCC) is one of the most heterogeneous malignant cancers with no effective targets and treatments. However, the molecular pathogenesis of HCC remains largely uncertain. The aims of our study were to find crucial genes involved in HCC through multidimensional methods and revealed potential molecular mechanisms. Here, we reported the gene expression profile GSE121248 findings from 70 HCC and 37 adjacent normal tissues, all of which had chronic hepatitis B virus (HBV) infection, we were seeking to identify the dysregulated pathways, crucial genes and therapeutic targets implicated in HBV-associated HCC. We found 164 differentially expressed genes (DEGs) (92 downregulated genes and 72 upregulated genes). Gene ontology (GO) analysis of DEGs revealed significant functional enrichment of mitotic nuclear division, cell division, and the epoxygenase P450 pathway. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed that the DEGs were mainly enriched in metabolism, cell cycle regulation and the p53 signaling pathway. The Mcode plugin was calculated to construct a module complex of DEGs, and the module was mainly enriched in cell cycle checkpoints, RHO GTPase effectors and cytochrome P450. Considering a weak contribution of each gene, gene set enrichment analysis (GSEA) was performed, revealing results consistent with those described above. Six crucial proteins were selected based on the degree of centrality, including NDC80, ESR1, ZWINT, NCAPG, ENO3 and CENPF. Real-time quantitative PCR analysis validated the six crucial genes had the same expression trend as predicted. Furthermore, the methylation data of The Cancer Genome Atlas (TCGA) with HCC showed that mRNA expression of crucial genes was negatively correlated with methylation levels of their promoter region. The overall survival reflected that high expression of NDC80, CENPF, ZWINT, and NCAPG significantly predicted poor prognosis, whereas ESR1 high expression exhibited a favorable prognosis. The identification of the crucial genes and pathways would contribute to the development of novel molecular targets and biomarker-driven treatments for HCC.

Subjects Bioinformatics, Computational Biology, Genomics
Keywords Hepatocellular carcinoma, Gene expression profile, Crucial genes, Methylation, ESR1

INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most frequent malignant cancer and the fourth-leading cause of cancer-related death worldwide (*Bray et al., 2018*). Patients with hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol consumption and aflatoxin exposure exhibit an increased risk of liver cancer (*El-Serag, 2002*). Tumorigenesis is largely ascribed to genetic changes, including structure variations, genetic copy number variants, single nucleotide variants and small insertions and deletions (*Lengauer, Kinzler & Vogelstein, 1998*). Although surgical resection, liver transplantation and sorafenib have prolonged the survival of HCC patients, high postoperative recurrence and drug resistance remain risk factors that confer a worse prognosis (*Wang et al., 2015*). Traditional treatment strategies for HCC are not optimal. Thus, an increased understanding of the molecular mechanism of HCC initiation and occurrence is of great importance and can be helpful to explore novel therapeutic targets for HCC.

Integration of molecular phenotypes at different levels of HCC subtypes can provide a genetic map for surgically resected tumors and provide information on genomic, transcriptomic,and epigenetic regulation (*Hoshida et al., 2009*). These different molecular subtypes represent diverse biologic backgrounds of HCC patients and have potential effects on drug treatment options and prognosis. The Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) public databases, with a relatively comprehensive gene expression data of HCC, provide the opportunity for the bioinformatics mining (*Jiang & Liu, 2015*). For example, a comprehensive analysis of the genomic and epigenomic landscape of HCC identified significant genes altered in tumors, such as APOB, MDM4, MET, TERT and so on (*Cancer Genome Atlas Research Network, 2017*). Lee et al. used GEO and TCGA databases to find COL1A1 was upregulated in HCC and may as a potential therapeutic target of HCC initiation and progression (*Ma et al., 2019*). CD5L, SLC22A10, UROC1, and SPP2, were identified in alcohol-related HCC datasets (*Zhang et al., 2019*). To date, numerous dysregulated and somatic mutations of genes and signaling pathways have been identified that play critical roles in HCC initiation and progression (*Kan et al., 2013*), including TP53 (*Kan et al., 2013*), UBE3C (*Jiang et al., 2014*), EGFR (*Jang et al., 2017*), SHP-1 (*Wen et al., 2018*) and JAK/STAT (*Kan et al., 2013*). The study indicated that chromosomal instability might be a driving force of TP53 inactivation and CCND1 and FGF19 amplifications (*Wang et al., 2013*). In addition, posttranscriptional modification of genes also has a cumulative effect on promoting HCC. The study found that hypomethylation of the CD147 promoter region accelerated HCC progression and was associated with worse prognosis (*Kong et al., 2011*). Unfortunately, most of the driver genes (CTNNB1, TP53, AXIN1, TERT and ARID2) previously identified do not have clinical applications (*Schulze et al., 2015*). Despite these findings, tumor heterogeneity and the small samples used in studies on HCC limit our understanding of HCC, and we are currently unable to genotype patients based on individual genome variations for drugs and early diagnosis.

In the study, we downloaded raw data for the gene expression profile GSE121248 from the GEO database, which included 70 HBV-induced HCCs and 37 adjacent normal

tissues. A Robust Multichip Averaging (RMA) normalization algorithm was used to screen differentially expressed genes (DEGs). We further utilized the centrality method of degree to identify crucial proteins and build the significant modules of molecular complex clusters in the protein-protein interaction (PPI) network. We also validated mRNA expression of six significantly crucial genes by real-time PCR, revealing results consistent with the public database analysis. Methylation of crucial genes in HCC was also assessed. Survival probability analysis of the six crucial genes was assessed using TCGA data.

MATERIALS & METHODS

Microarray data

The raw microarray data of GSE121248, including 70 HCC tissues and 37 adjacent normal tissues, were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/), which was an openly available database. Another two datasets of GSE25097 and GSE22058 were used as the testing sets to verify our results.

Data processing and identification of DEGs

The bioconductor package (version3.7, http://www.bioconductor.org/), affy package (version1.58.0, http://bioconductor.org/packages/affy/) and the RMA algorithm (*Irizarry et al., 2003*) were used for data preprocessing and normalization in R (version 3.5.1, *R Core Team, 2018*). The *t* test methods of the limma package (version 3.38.2, http://bioconductor.org/packages/limma/) (*Ritchie et al., 2015*) were used to identify DEGs in HCC tissues compared to adjacent normal tissues. Values of log Fold Change >2.0 or log Fold Change <−2.0 and *P*–value <0.05 served as the cut-off criteria.

Functional annotation of DEGs

Gene Ontology (GO) analysis was applied using the Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8, https://david.ncifcrf.gov/) to explore the biological processes, cellular components and molecular function in which the DEGs were involved (*Ashburner et al., 2000*). GeneAnswers (version 2.24.0, http://bioconductor.org/packages/GeneAnswers/) was a Kyoto Encyclopedia of Genes and Genomes (KEGG)-based R package that not only provided the annotation of biochemistry pathways associated with diseases and drugs but also facilitated visualization of the signaling pathways enriched by DEGs (*Feng et al., 2010*).

Gene set enrichment analysis of HCC

Gene Set Enrichment Analysis (GSEA) (version 3.0, the broad institute of MIT and Harvard, http://software.broadinstitute.org/gsea/downloads.jsp) was performed between HCC and adjacent normal tissues to investigate the biological characteristic of HCC (*Subramanian et al., 2005*). In detail, the 'collapse data set to gene symbols' was set to true, the 'permutation type' was set to phenotype, the 'enrichment statistic' was set to weighted, and the Signal2Noise metric was used for ranking genes. GSEA calculated a gene set Enrichment Score (ES) that analyzed genes were enriched in the biological signal conduction on the MsigDB (Molecular Signatures Database, http://software.broadinstitute.org/gsea/msigdb).

And the meandiv normalization method was used for enrichment scores in the gene sets. In this study, 1,000 gene of permutations were set to generate a null distribution for enrichment score in the hallmark gene sets and functional annotation gene sets. The gene sets database used for enrichment analysis were mainly including 'h.all.v6.2.symbols.gmt', 'c2.cp.kegg.v6.2.symbols.gmt', 'c2.cp.reactome.v6.2.symbols' and 'c5.bp.v6.2.symbols'. Nominal P–value <0.05, FDR <0.25 and gene set size >100 were defined as the cut-off criteria.

PPI network and module analysis

The PPI network was analyzed using the STRING database for *Homo sapiens* (*Szklarczyk et al., 2017*). STRING was used to calculate PPI networks of DEGs with a combined score >0.4 as the cut-off criteria. Cytoscape (version3.7.0, https://cytoscape.org/) was used to visualize the network (*Shannon et al., 2003*). Mcode (version1.5.1, Bader Lab, University of Toronto, http://www.baderlab.org/Software/MCODE/) was a Cytoscape plugin for constructing the protein modulecomplex with a degree cut-off $= 2$, node score cut-off $= 0.2$, max.depth $= 100$, and k-core $= 2$. CytoNCA (version 2.1.6, http://apps.cytoscape.org/download/stats/cytonca/) was used to analyze centrality of protein interaction networks that were representative of potentially crucial proteins in the network (*Tang et al., 2015*). The crucial proteins were identified based on four different centrality parameters (degree centrality (DC), betweenness centrality (BC), eigenvector centrality (EC) and closeness centrality (CC)). The values of centrality parameters were sorted to select the top 10 ranked proteins, and the proteins were input as the crucial candidates. ClueGO (version 2.5.3, http://apps.cytoscape.org/apps/cluego) and CluePedia (version 1.5.3, http://apps.cytoscape.org/apps/cluepedia) were employed to analyze the KEGG and Reactome pathways.

Transcriptional expression level and survival analysis of crucial genes on HCC

The mRNA expression of crucial genes of HCC and adjacent normal tissues were evaluated based on the shinyGEO database, which included two HCC datasets GSE25097 (containing of 268 HCC and 243 adjacent normal tissues) and GSE22058 (containing of 96 HCC and 96 adjacent normal tissues) (*Dumas, Gargano & Dancik, 2016*). The Human Protein Atlas (http://www.proteinatlas.org) was used for protein detection by immunohistochemistry. We performed TCGA methylation data of MEXPRESS database (https://mexpress.be/) to explain why oncogene mRNAs were highly expressed in HCC. The patients overall survival was investigated using TCGA_LIHC data and the clinicopathological features was download by TCGA database (https://www.cbioportal.org/). Statistical significance was considered at $P < 0.05$.

HCC sample collection and revalidation of mRNA expression of crucial genes

Thirty pairs of HCC and paired adjacent tissues were obtained from Xijing Hospital, Fourth Military Medical University of China. Our study was approved by the Ethical Committee and Institutional Review Board of Fourth Military Medical University (Ethical

Application Ref: KY20183288-1). The RNA was extracted using RNA extraction kit II (Omega Bio-tek, GA, USA) and cDNA was synthesized using PrimeScript^{TM} RT reagent Kit (TaKaRaBio, Otsu, Japan). The mRNA expression was assessed in 384-well plates using the CFX connect^{TM} Real-Time System (BIO-RAD, Hercules, CA, USA) with the SYBR Premix Ex Taq II (TaKaRaBio, Otsu, Japan) and primers, and qPCR data were analyzed using $^{\triangle\triangle}$Ct method with β-actin as the reference gene. Primer sequences were listed in Table S1.

Statistical analysis

Statistical analysis was performed with SPSS 19.0 software (IBM, Armonk, NY, USA). All the data were presented as mean ± SEM. Overall survival time was calculated by the Kaplan–Meier method and analyzed with the log-rank test. The univariate and multivariate analyses were calculated based on the Cox regression model. The statistical analysis was performed as appropriate by Student's t test and χ^2 test. The P-value < 0.05 were considered statistically significant.

RESULTS

Identification of DEGs

A total of 164 DEGs that met the cut-off criteria were identified, including 72 upregulated and 92 downregulated genes (Fig. 1A). The mRNA expression profiles of 30 representative DEGs in 70 HCC and 37 adjacent normal tissues were presented in Fig. 1B.

Functional annotation analysis of DEGs

GO analysis of DEGs revealed that the following GO terms were significantly enriched in condensed chromosome kinetochore ($P = 1.33E$-10), mitotic nuclear division ($P = 7.45E$-10), sister chromatid cohesion ($P = 1.97E$-09), and cell division ($P = 2.34E$-09) (Table S2). KEGG annotation analysis of DEGs in HCC revealed that the pathway mainly including retinol metabolism ($P = 4.21E$-05), drug metabolism-cytochrome P450 ($P = 5.95E$-05), cell cycle ($P = 2.44E$-04), and the p53 signaling pathway ($P = 5.59E$-04) (Fig. 2A).

Protein–protein interaction network of DEGs

The PPI network of DEGs with 113 nodes (genes) and 933 edges (interactions) was built by STRING. The CytoNCA plugin was used to analyze the centrality of nodes. The top 10 proteins, which were also defined as crucial proteins based on four different centrality parameters, were presented in Table 1, and they included TOP2A, NDC80, ESR1, ZWINT, CENPF, NCAPG, ENO3 and CCNB1. The majority of these crucial proteins participated in mitotic prometaphase ($P = 1.5E$-14) and resolution of sister chromatid cohesion ($P = 6E$-13) (Fig. 2B).

The MCODE plugin was employed to identify the modules in the PPI network. The top 2 significant modules were displayed in Fig. 2C, and the seed of two separate modules are RAD51AP1 and ESR1. KEGG and Reactome pathway analyses revealed that the two modules were mainly related to cell cycle checkpoints ($P = 7.94E$-15), RHO GTPase effectors ($P = 6.87E$-13), and Cytochrome P450 ($P = 1.37E$-12) (Table S3).
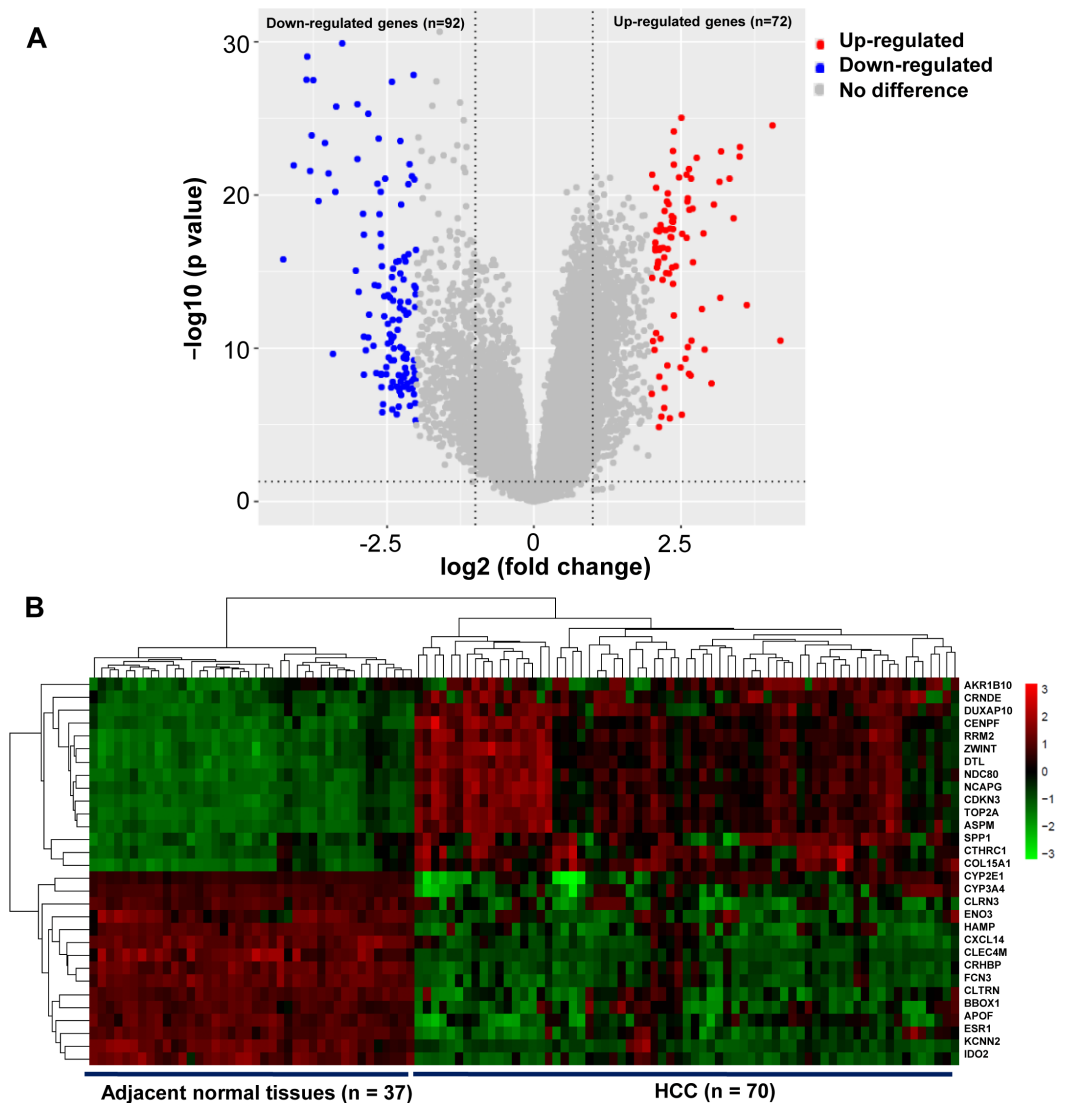
Figure 1 **Identification of DEGs in HCC.** (A) Volcano plot of gene expression profile. The upper-right red dots represented upregulated genes, and the upper-left blue dots represented downregulated genes. (B) Hierarchical clustering heatmap of the 30 representative DEGs screened on the basis of |fold change| > 2.0 and P-value < 0.05. The columns represented the samples, including 37 adjacent normal and 70 HCC tissues. The rows showed that the 30 representative DEGs, including top 15 upregulated expressed genes, and the last 15 downregulated expressed genes in HCC. Red indicated that the expression of genes was relatively upregulated, and the green indicated the expression of genes was relatively downregulated.

Full-size 🖼 DOI: 10.7717/peerj.7436/fig-1

Gene Set Enrichment Analysis of HCC

To provide further insight into the whole gene enrichment annotation of HCC and consider the potential role of undifferentiated genes or genes with smaller differences in expression in HCC and adjacent normal tissues, GSEA was performed to explore the gene expression profile based on the GO terms and pathway databases of KEGG, Reactome and Hallmark
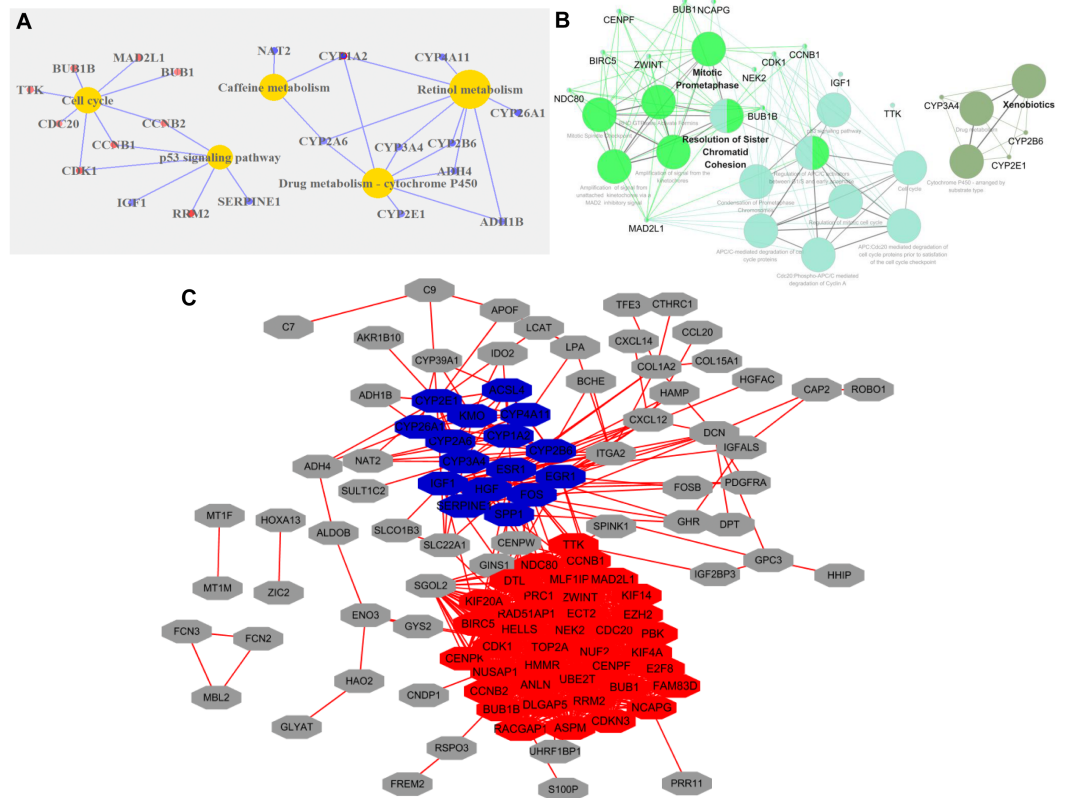
Figure 2 Functional categories of DEGs in HCC. (A) KEGG pathway analysis of DEGs. Red circles represented upregulated genes, and blue represented downregulated genes. (B) KEGG and Reactome pathways analysis of 10 top ranked crucial proteins. The crucial proteins were identified by ClueGO. The colors represented three types of enrichment. (C) The top two modules in the PPI network. The red indicated module 1, and blue indicated module 2.

Full-size ⬇ DOI: 10.7717/peerj.7436/fig-2

Table 1 The top 10 proteins ranked based on the node centrality of the PPI network.

| Rank	Degree centrality		Betweenness centrality		Closeness centrality		Eigenvector centrality	
	Gene symbol	Expression in HCC	Gene symbol	Expression in HCC	Gene symbol	Expression in HCC	Gene symbol	Expression in HCC
1	TOP2A	up-regulated	TOP2A	up-regulated	TOP2A	up-regulated	CCNB1	up-regulated
2	NDC80	up-regulated	ESR1	down-regulated	NDC80	up-regulated	CDK1	up-regulated
3	CCNB1	up-regulated	CYP3A4	down-regulated	ESR1	down-regulated	NDC80	up-regulated
4	CDK1	up-regulated	CYP2E1	down-regulated	CDKN3	up-regulated	ZWINT	up-regulated
5	CDKN3	up-regulated	NDC80	up-regulated	BIRC5	up-regulated	BUB1	up-regulated
6	ZWINT	up-regulated	IGF1	down-regulated	CENPF	up-regulated	TTK	up-regulated
7	TTK	up-regulated	CYP2B6	down-regulated	CCNB1	up-regulated	NCAPG	up-regulated
8	BUB1	up-regulated	COL1A2	up-regulated	CDK1	up-regulated	RACGAP1	up-regulated
9	CENPF	up-regulated	CENPF	up-regulated	EZH2	up-regulated	BUB1B	up-regulated
10	NCAPG	up-regulated	ENO3	down-regulated	NEK2	up-regulated	MAD2L1	up-regulated
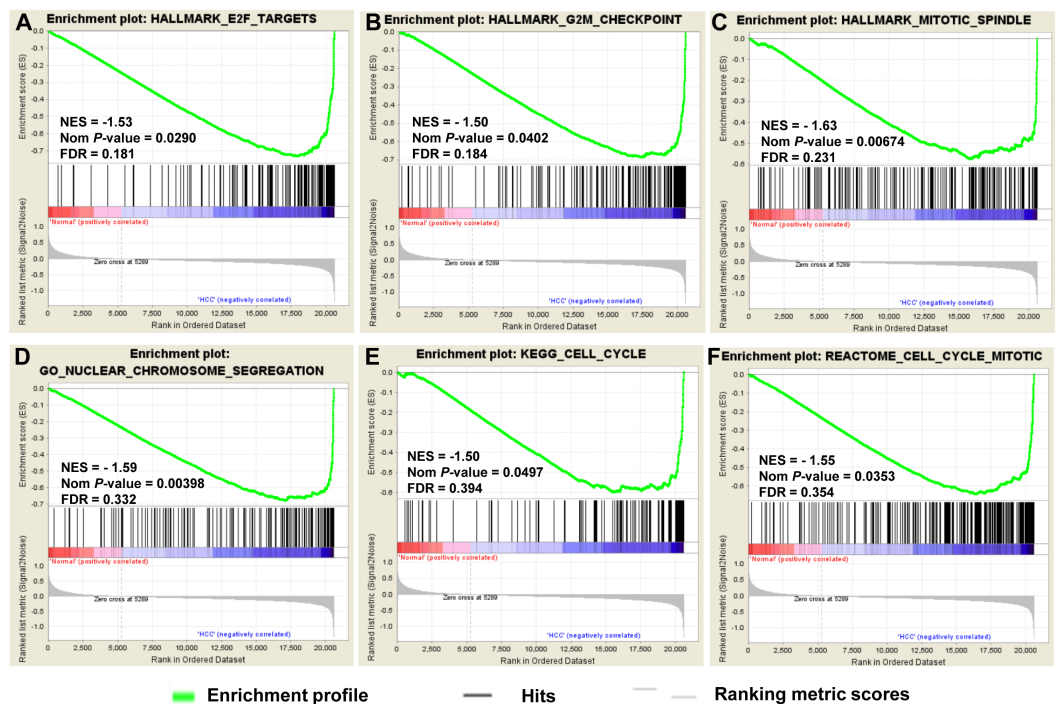
Figure 3 **GSEA of gene expression data in HCC.** The gene expression profiling identified cell cycle dys-regulation of HCC. The enrichment of whole expressed genes were analyzed using GSEA of Hallmark collection database, mainly including 'E2F_TARGETS' term (A), 'G2M_CHECKPOINT' term (B) and 'MITOTIC_SPINDLE' term (C). GO terms (D) and pathway databases of KEGG (E), Reactome (F) analysis demonstrated that the whole expressed genes were enriched in cell cycle progression-related gene sets. The false discovery rates (FDR), the nominal *P*-value and the normalized enrichment score were calculated for each gene set. Each black bar at the bottom of each panel represented a member gene of the respective gene set.

Full-size 🖼 DOI: 10.7717/peerj.7436/fig-3

gene sets. The result revealed that enrichment for three of six gene sets were significant at *FDR* <0.25, including genes involved in cell cycle and chromosome segregation (Fig. 3).

Validation of mRNA expression of crucial genes

Six significantly crucial genes, NDC80, ESR1, ZWINT, NCAPG, ENO3 and CENPF, were validated in two other HCC datasets (GSE25097 and GSE22058). The results also revealed that NDC80, ZWINT, NCAPG and CENPF were significantly upregulated in HCC, and ESR1 and ENO3 were downregulated in HCC (Fig. 4). Then, immunohistochemistry staining validated from the Human Protein Atlas database showed that ZWINT, NCAPG and CENPF protein expression were strongly upregulated in liver cancer tissues compared with normal tissues (Figs. 5A–5C). ENO3 and ESR1 protein were downregulated in liver cancer tissues compared with normal tissues (Figs. 5D & 5E). The protein expression of NDC80 was absent in the Human Protein Atlas database. Consistent with database analyses, the six crucial gene were successfully validated by qPCR in 30 paired human HCC and adjacent normal tissues, and the results revealed no difference with the analysis results of microarray profiling with the exception of ENO3 (Fig. 6).
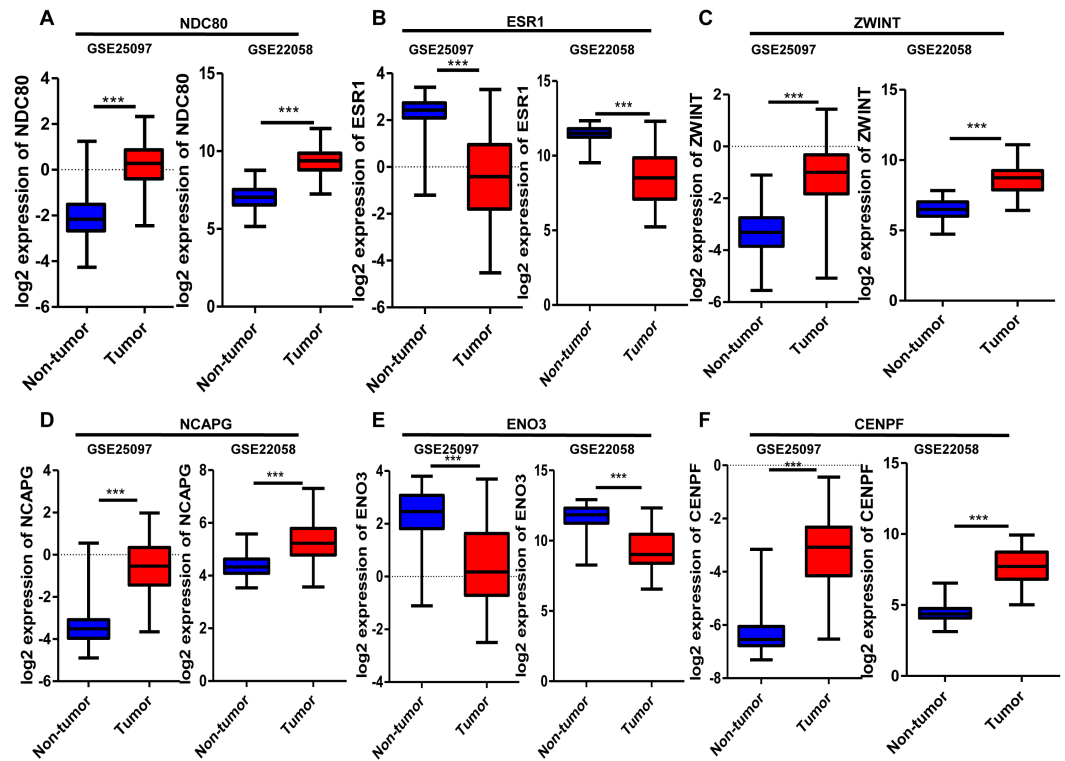
Figure 4 Validation of differential expression of six crucial genes in GEO public datasets. Increased expression of (A) NDC80, (C) ZWINT, (D) NCAPG and (F) CENPF in HCC. Reduced expression of (B) ESR1 and (E) ENO3 in HCC, *** $P < 0.001$.

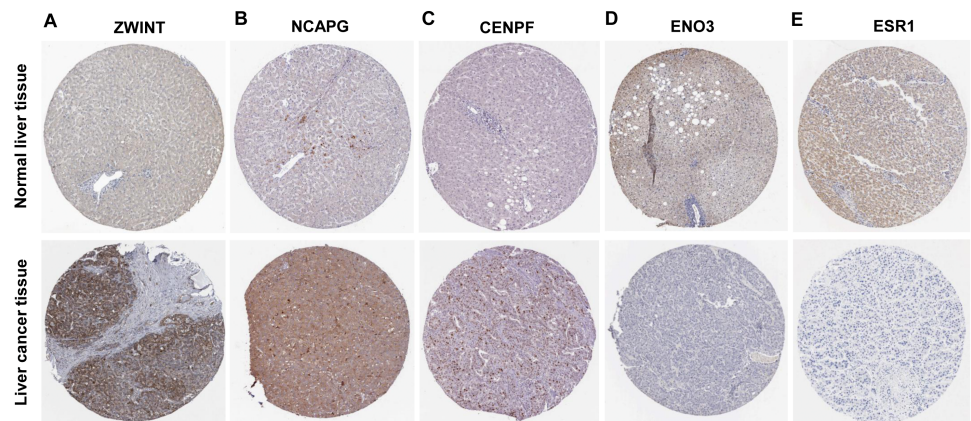Full-size ☑ DOI: 10.7717/peerj.7436/fig-4

Figure 5 Protein expression of the crucial genes in liver cancer and normal tissues based on The Human Protein Atlas database. (A) ZWINT, (B) NCAPG and (C) CENPF proteins were strongly upregulated in liver cancer tissues compared with normal tissues. (D) ENO3 and (E) ESR1 proteins were downregulated in liver cancer tissues compared with normal tissues.

Full-size ☑ DOI: 10.7717/peerj.7436/fig-5

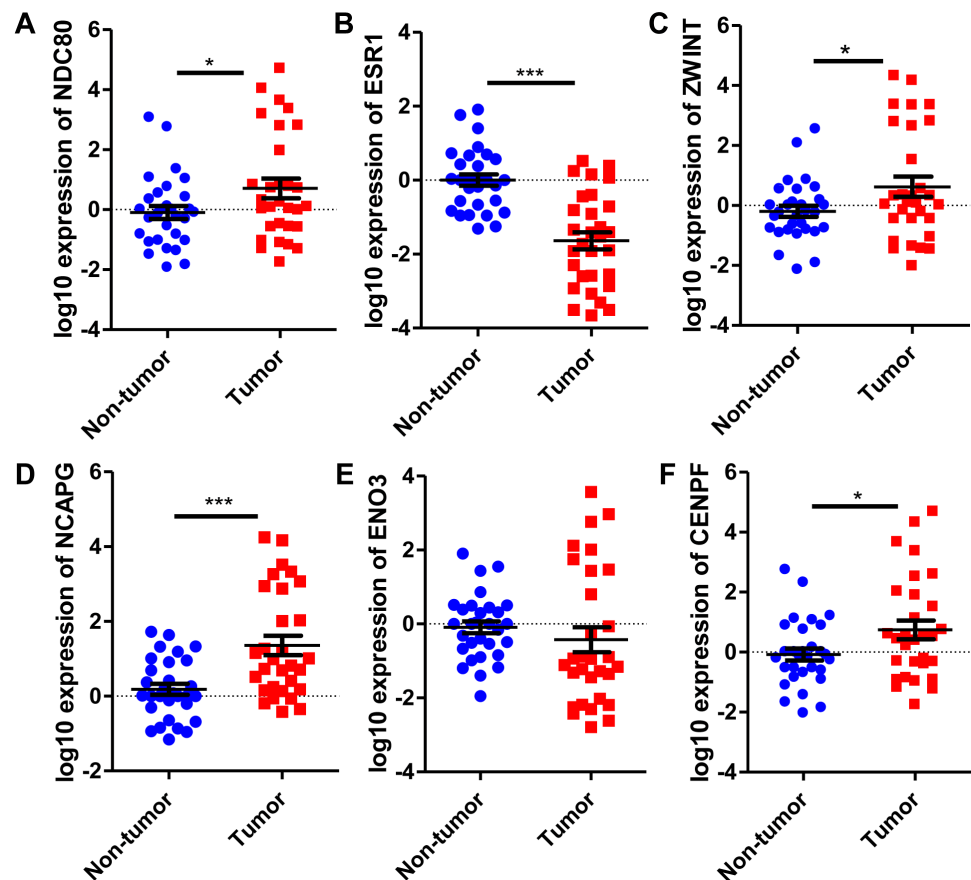Liu et al. (2019), *PeerJ*, DOI 10.7717/peerj.7436

9/20

Figure 6 Revalidation of six crucial genes in 30 pairs of HCC and paired non-tumor tissues. Quantitative real-time PCR results for the six crucial genes. (A) NDC80; (B) ESR1; (C) ZWINT; (D) NCAPG; (E) ENO3; (F) CENPF. Expression of these crucial genes was standardized against β-actin expression. The statistical analysis was calculated by the Student's t test. *$P < 0.05$, ***$P < 0.001$.

Full-size ⬛ DOI: 10.7717/peerj.7436/fig-6

The relationship between the methylation and mRNA expression of crucial genes were respectively validated by TCGA data to identify oncogenes with highly expressed mRNA in HCC. The results showed that mRNA expression of crucial genes was negatively correlated with methylation levels. The genes highly expressed or hypomethylated were mostly involved in HCC. Hypermethylated/poorly expressed crucial genes in HCC tended to exhibit high expression in adjacent normal tissues (Table 2).

Kaplan–Meier survival analysis of crucial genes in HCC patients

TCGA data were used to explore the relationship between the crucial genes and the survival of HCC patients, and high expression of NDC80 ($P = 1.10E-3$), NCAPG ($P = 2.80E-4$), CENPF ($P = 5.72E-4$) and ZWINT ($P = 2.47E-4$) significantly correlated with worse survival probability for HCC patients. ESR1 ($P = 2.02E-4$) exhibited the opposite trend, and ENO3 was not a significant marker of survival prognosis ($P = 0.186$) (Fig. 7). There we focused a crucial tumor suppressor gene, ESR1. The RNA expression data from the TCGA_LIHC cohorts revealed that increased ESR1 mRNA expression correlated

Table 2 Validation of the relationship between methylation and mRNA expression of crucial genes in TCGA.

Crucial gene	mRNA expression in HCC		Methylation status in HCC			
	mRNA expression	P-value	Methylation	Position	Pearson correlation	P-value
NDC80	High expression	<2.2E-16	Hypomethylation	chr18:2571606	−0.150	1.64E-03
				chr18:2571747	−0.282	<2.2E-16
ZWINT	High expression	<2.2E-16	Hypomethylation	chr10:58121299	−0.142	0.03
NCAPG	High expression	<2.2E-16	Hypomethylation	chr4:17812759	−0.236	2.26E-11
				chr4:17813558	−0.544	<2.2E-16
CENPF	High expression	<2.2E-16	Hypomethylation	chr1:214776904	−0.243	1.48E-12
				chr1:214777596	−0.304	1.60E-14
				chr1:214780880	−0.174	6.27E-11
				chr1:214781456	−0.154	1.38E-04
				chr1:214781663	−0.217	1.21E-05
				chr1:214788192	−0.150	7.40E-07
ESR1	Low expression	<2.2E-16	Hypermethylation	chr6:152002969	−0.139	2.58E-02
				chr6:152003038	−0.142	5.60E-05
				chr17:4850943	−0.160	1.39E-03
				chr17:4851808	−0.146	5.41E-12
ENO3	Low expression	7.46E-14	Hypermethylation	chr17:4852013	−0.139	8.45E-04
				chr17:4852349	−0.118	9.08E-12
				chr17:4854754	−0.705	8.20E-05
				chr17:4855149	−0.604	2.24E-06

significantly with gender, age, serum AFP, TNM stage, tumor recurrence and tumor differentiation (Table S4). Univariate analysis also revealed that the crucial genes ESR1 expression was significantly correlated with overall survival and recurrence (Table S5). Multivariate analysis was performed using all of the variables that were identified as significant by univariate analysis. The result showed that ESR1 mRNA expression was an independent prognostic indicator for overall survival and recurrence, and other features of patients with HCC were also analyzed (Table S5).

The mRNA expression of ER related receptor and regulatory genes in HCC

Since estrogen plays a protective role in liver cancer by binding to estrogen receptors, we also further investigated the expression of ER (estrogen receptor) regulon genes in HCC tissues compared with adjacent tissues and the correlation between ER regulon genes and ESR1 using the public GEO database (GSE22058). The mRNA expression levels of sex hormone- related genes (ESR2, ESRRB, ESRRG, PGRMC1, PGRMC2, GPER, SHBG) and ER regulatory gene, TFF1 in HCC tissues were significantly lower than those of adjacent tissues, whereas the expression of estrogen-related receptor gene, ESRRA and ER regulatory genes (MTA1 and FOXA1) were higher in tumor tissues (Fig. 8). In HCC tissues, ESR1 mRNA level was positively correlated with ESRRB ($r = 0.3441$, $P = 0.0006$), PGRMC1 ($r = 0.5124$, $P < 0.0001$) and GPER ($r = 0.2274$, $P = 0.0259$) (Fig. 9). Moreover, ESR1
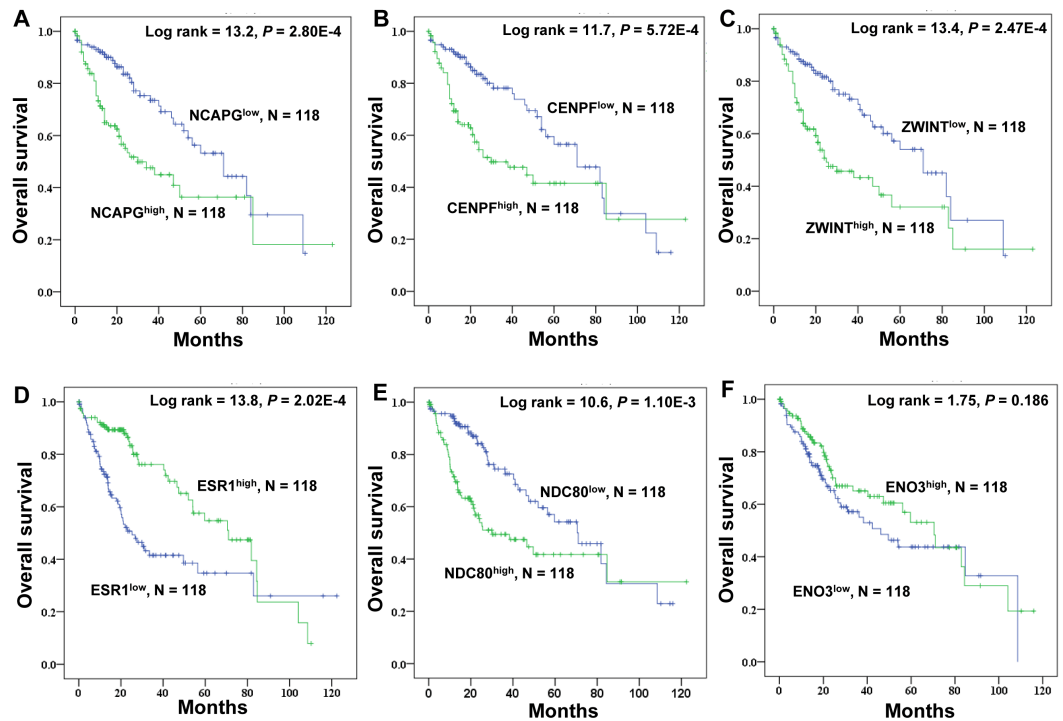
Figure 7 Kaplan–Meier survival analysis of 118 liver cancer samples using the TCGA database. The prognostic survival analysis of the six crucial genes in HCC patients. (A) NCAPG; (B) CENPF; (C) ZWINT; (D) ESR1; (E) NDC80; (F) ENO3. The green lines represent patients with high gene expression, and blue lines represent patients with low gene expression.

Full-size 🖼 DOI: 10.7717/peerj.7436/fig-7

mRNA level was negatively correlated with ER regulatory genes MTA1 ($r = -0.2714$, $P = 0.0079$) and TFF1 ($r = -0.3557$, $P = 0.0004$) (Fig. 9).

DISCUSSION

In this study, we systematically analyzed gene expression profiles of HCC, including 70 HCC and 37 adjacent normal tissues, using bioinformatics methods. A total of 167 DEGs were identified, including 72 upregulated and 92 downregulated genes. The GO terms indicated that DEGs were enriched in mitotic nuclear division, sister chromatid cohesion, cell division, oxidoreductase activity and condensed chromosome kinetochore. KEGG pathway analysis revealed that the DEGs were enriched in the cell cycle, and the p53 signaling pathway and in metabolism. To eliminate individual genetic biases, the enrichment of entire genes was performed through GSEA analysis. Whole genes were mainly enriched in cell cycle and chromosome segregation, which was consistent with the above enrichment results.

One of the distinguishing characteristics in cell division progression is the segregation of sister chromatids. Incorrect segregation of chromosomes could be caused by abnormal regulation or damage of the spindle assembly checkpoint, which subsequently increased chromosome instability and aneuploidy and accelerated tumor progress (*Carter et al.,*
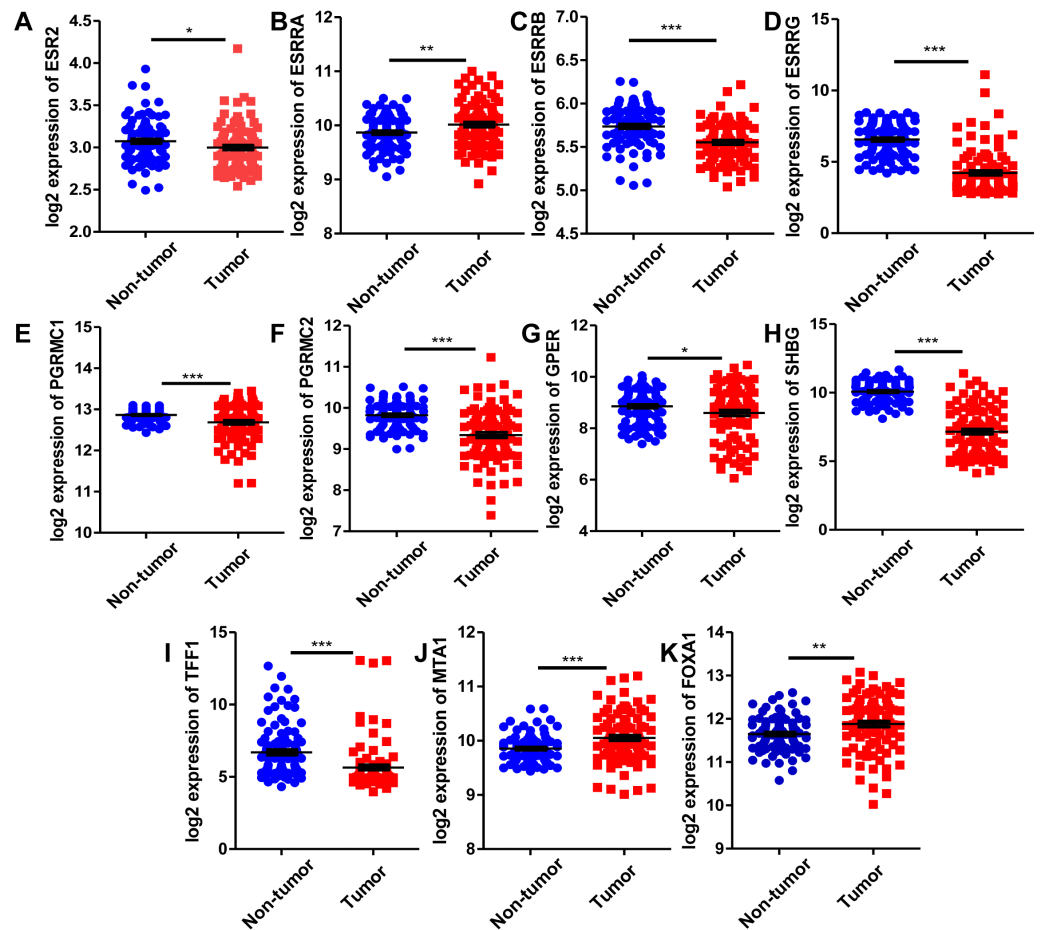
Figure 8 Comparison of mRNA expression of ER-related regulon genes in human HCC tumor tissues and adjacent tissues using GEO database GSE22058 ($n = 96$). The mRNA expression of sex hormone-related genes in HCC tissues relative to the adjacent non-tumor tissues. (A) ESR2; (B) ESRRA; (C) ESRRB; (D) ESRRG; (E) PGRMC1; (F) PGRMC2; (G) GPER; (H) SHBG. The mRNA expression of ER regulatory genes in HCC tissues relative to the adjacent non-tumor tissues. (I) TFF1; (J) MTA1; (K) FOXA1. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Full-size 🖼 DOI: 10.7717/peerj.7436/fig-8

2006). RB deficiency induced chromosomal instability to promote liver tumorigenesis (*Mayhew et al., 2007*). In addition, the cytochrome P450 family of proteins were downregulated in HCC based on microarray analysis. Jiang et al. confirmed that CYP3A5 was a suppressor by regulating mTORC2/AKT to antagonize the malignant phenotype of HCC (*Jiang et al., 2015*). RYO et al. found that CYP3A4 downregulation was related to a poor prognosis for 92 HCC patients (*Ashida et al., 2017*). Therefore, abnormal regulation of the cell cycle together with chromosome segregation or inactivation of some cytochrome P450 proteins could contribute to hepatogenesis.

In recent years, bioinformatics mining has been widely used for identifying tumor features and novel diagnosis markers (*Jiang & Liu, 2015*). For example, Agarwal et al. found that 59 kinases related genes were overexpressedin HCC, and furthermore predicted overall four distinct HCC subtypes and each subtype represented unique gene expression
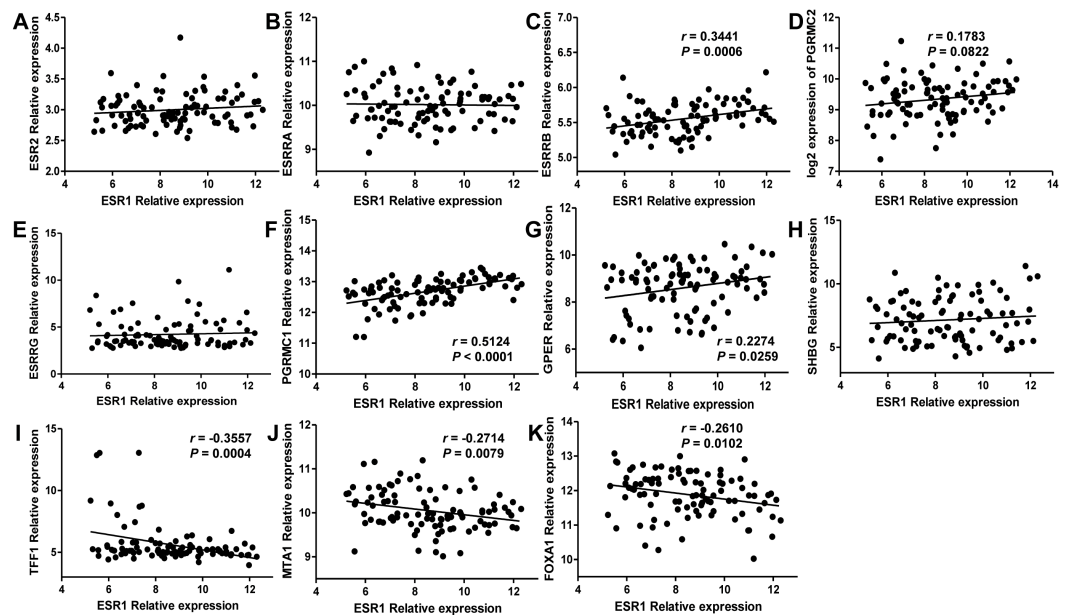
Figure 9 Correlation of ESR1 and ER regulon genes in HCC tissues based on database GSE22058 (*n = 96*). The correlations between ESR1 mRNA expression and sex hormone-related genes expression were analyzed (A–H), and the correlations between ESR1 mRNA expression and ER regulatory genes were analyzed (I–K).

Full-size ☑ DOI: 10.7717/peerj.7436/fig-9

profiling, pathway enrichment and prognosis using TCGA dataset (*Agarwal et al., 2017*). Wang et al. analyzed the potential biomarks for HCC by integrated GEO and TCGA datasets and identified AKR1B10 as a novel biomarker (*Wang et al., 2018*). Yin et al. showed that the genes involved in the G2/M checkpoint may act as biomarkers at the early HCC (*Yin, Chang & Xu, 2017*). In addition, another study found that hub genes TOP2A, PCNA and AURKA with low mutation frequency involved in HCC by bioinformatics analysis (*Xing, Yan & Zhou, 2018*). However, their study only analyzed a profile, and only identified the gene with a high node degree using the module analysis. In addition, their selected genes were validated in terms of genetic alteration method or the Kaplan–Meier plotter database. Our study combined the results of MCODE, CytoNCA and methylation modification for the identification of crucial genes. Moreover, we validated the results by the other three databases GSE25097, GSE22058, TCGA_LIHC and RT-PCR, and used clinical features and prognosis of the patients to evaluate the importance of crucial gene, thus promoting the reliability of our results.

According to degree centrality and molecular module analysis of differentially expressed proteins in the PPI network, the crucial proteins including TOP2A, NDC80, ESR1, CCNB1, ZWINT, CENPF, ENO3 and NCAPG were identified. Subsequently, the molecular module of DEGs was constructed to understand the closely related subgroups, and the seed of two separate modules was displayed as RAD51AP1 and ESR1. Furthermore, NDC80, ZWINT, NCAPG, and CENPF were upregulated in HCC as demonstrated by qPCR and associated

with poor prognosis. ESR1 was a favorable prognostic factor. ZWINT, NCAPG and CENPF were frequently identified in the related enrichment sets by GSEA analysis.

Interestingly, we found that ESR1 mRNA expression was downregulated in HCC, and ESR1 mRNA expression was negatively correlated with methylation levels of its promoter region. ESR1 was associated with a favorable prognostic factor, and may as a potential tumor suppressor in HCC. Epidemiological studies showed that the prevalence rate of HCC was lower in female than in male patients (*El-Serag & Rudolph, 2007*). Chronic hepatic disease progressed more slowly to cirrhosis and HCC in female than in male patients (*Shimizu, 2003*). Therefore, the estrogens may play a protective role in the progression of HCC. To our knowledge, ESR1 (estrogen receptor α), which encodes the estrogen receptor, has a role in hormone binding and activation of transcription. Hishida et al. previously reported that the expression level of ESR1 was downregulated in 24 (50%) patients with HCC (*Hishida et al., 2013*). Estrogen inhibited the transcription of HBV by upregulating ER α, which changed the interaction with HNF4 α to alter binding to HBV enhancer I (*Wang et al., 2012*). Studies have reported that estrogen exerted protective effects against HCC through inhibiting IL-6 production (*Naugler et al., 2007*). Moreover, we identified ESR1 mRNA level was negatively correlated with estrogen-related regulatory genes TFF1 and MTA1 using HCC dataset GSE22058.

In addition, NDC80 is a kinetochore complex componentthat can organize and stabilize microtubule-kinetochore interactions and is required for proper chromosome segregation. NDC80 was overexpressed in many solid cancers, and silencing of NDC80 inhibited cell cycle and cell proliferation (*Meng et al., 2015*). The small molecule SM15 bound to microtubules and NDC/Hec1 to control tumor growth through microtubule stabilization (*Ferrara et al., 2018*). Ying et al. confirmed that ZWINT interacted with cell cycle-related proteins to promote HCC proliferation (*Ying et al., 2018*). *Liu et al. (2018)* demonstrated that NCAPG played an oncogenic role in the development of HCC. *Shahid et al. (2018)* found that CENPF expression was increased in higher risk prostate cancer patients. ENO3 encoded beta-enolase, which was involved in glycolysis and gluconeogenesis, and little information had been reported about the function of ENO3 (*Kong et al., 2016*). Our study utilized multidimensional bioinformatics methods to seek and validate network crucial genes in association with HCC from the database of TCGA and GEO. We identified six crucial genes (NDC80, ESR1, ZWINT, NCAPG, ENO3 and CENPF) based on the degree of centrality and module analysis of DEGs network and studied the significance of the expression pattern of the crucial genes for overall survival of patients with HCC. Moreover, we used qRT-PCR to validate our bioinformatics results: NDC80, ZWINT, NCAPG and CENPF mRNA expressions were significantly upregulated in HCC, and ESR1 and ENO3 were downregulated in HCC. Furthermore, by TCGA data, mRNA expression of crucial genes were negatively correlated with methylation levels of their promoter region. Additionally, ESR1 expression correlated significantly with clinicopathological variables, indicating that it may be a crucial tumor suppressor gene of HCC. The crucial genes we

identified may be useful biomarkers for diagnosis and prognosis of HCC, even if further experimental validations are needed to confirm the findings.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare there are no competing interests.

Author Contributions
- Ze-Kun Liu performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper.
- Ren-Yu Zhang performed the experiments, analyzed the data, contributed reagents/materials/analysis tools.
- Yu-Le Yong analyzed the data.
- Zhi-Yun Zhang and Can Li prepared figures and/or tables.
- Zhi-Nan Chen and Huijie Bian conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.

Human Ethics
The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Fourth Military Medical University of China granted Ethical approval to carry out the study within its facilities (Ethical Application Ref: KY20183288-1).

Microarray Data Deposition
The following information was supplied regarding the deposition of microarray data:

Data is available at the Gene Expression Omnibus database, accession numbers: GSE25097, GSE22058 and GSE121248.

Data Availability
The following information was supplied regarding data availability:

The raw measurements are available in the Supplemental Tables.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7436#supplemental-information.

REFERENCES

Agarwal R, Narayan J, Bhattacharyya A, Saraswat M, Tomar AK. 2017. Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets. *Cancer Genetics* **216–217**:37–51 DOI 10.1016/j.cancergen.2017.06.002.

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, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nature Genetics* **25**:25–29 DOI 10.1038/75556.

Ashida R, Okamura Y, Ohshima K, Kakuda Y, Uesaka K, Sugiura T, Ito T, Yamamoto Y, Sugino T, Urakami K, Kusuhara M, Yamaguchi K. 2017. CYP3A4 gene is a novel biomarker for predicting a poor prognosis in hepatocellular carcinoma. *Cancer Genomics Proteomics* **14**:445–453 DOI 10.21873/cgp.20054.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**:394–424 DOI 10.3322/caac.21492.

Cancer Genome Atlas Research Network. 2017. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**:1327–1341 DOI 10.1016/j.cell.2017.05.046.

Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature Genetics* **38**:1043–1048 DOI 10.1038/ng1861.

Dumas J, Gargano MA, Dancik GM. 2016. shinyGEO: a web-based application for analyzing gene expression omnibus datasets. *Bioinformatics* **32**:3679–3681 DOI 10.1093/bioinformatics/btw519.

El-Serag HB. 2002. Hepatocellular carcinoma: an epidemiologic view. *Journal of Clinical Gastroenterology* **35**:S72–S78 DOI 10.1097/00004836-200211002-00002.

El-Serag HB, Rudolph KL. 2007. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* **132**:2557–2576 DOI 10.1053/j.gastro.2007.04.061.

Feng G, Du P, Krett NL, Tessel M, Rosen S, Kibbe WA, Lin SM. 2010. A collection of bioconductor methods to visualize gene-list annotations. *BMC Research Notes* **3**:10 DOI 10.1186/1756-0500-3-10.

Ferrara M, Sessa G, Fiore M, Bernard F, Asteriti IA, Cundari E, Colotti G, Ferla S, Desideri M, Buglioni S, Trisciuoglio D, Del BD, Brancale A, Degrassi F. 2018. Small molecules targeted to the microtubule-Hec1 interaction inhibit cancer cell growth through microtubule stabilization. *Oncogene* **37**:231–240 DOI 10.1038/onc.2017.320.

Hishida M, Nomoto S, Inokawa Y, Hayashi M, Kanda M, Okamura Y, Nishikawa Y, Tanaka C, Kobayashi D, Yamada S, Nakayama G, Fujii T, Sugimoto H, Koike M, Fujiwara M, Takeda S, Kodera Y. 2013. Estrogen receptor 1 gene as a tumor

suppressor gene in hepatocellular carcinoma detected by triple-combination array analysis. *International Journal of Oncology* **43**:88–94 DOI 10.3892/ijo.2013.1951.

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, Kumada H, Llovet JM, Golub TR. 2009. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Research* **69**:7385–7392 DOI 10.1158/0008-5472.CAN-09-1089.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**:249–264 DOI 10.1093/biostatistics/4.2.249.

Jang JW, Song Y, Kim SH, Kim JS, Kim KM, Choi EK, Kim J, Seo HR. 2017. CD133 confers cancer stem-like cell properties by stabilizing EGFR-AKT signaling in hepatocellular carcinoma. *Cancer Letters* **389**:1–10 DOI 10.1016/j.canlet.2016.12.023.

Jiang F, Chen L, Yang YC, Wang XM, Wang RY, Li L, Wen W, Chang YX, Chen CY, Tang J, Liu GM, Huang WT, Xu L, Wang HY. 2015. CYP3A5 functions as a tumor suppressor in hepatocellular carcinoma by regulating mTORC2/Akt signaling. *Cancer Research* **75**:1470–1481 DOI 10.1158/0008-5472.CAN-14-1589.

Jiang JH, Liu YF, Ke AW, Gu FM, Yu Y, Dai Z, Gao Q, Shi GM, Liao BY, Xie YH, Fan J, Huang XW, Zhou J. 2014. Clinical significance of the ubiquitin ligase UBE3C in hepatocellular carcinoma revealed by exome sequencing. *Hepatology* **59**:2216–2227 DOI 10.1002/hep.27012.

Jiang P, Liu XS. 2015. Big data mining yields novel insights on cancer. *Nature Genetics* **47**:103–104 DOI 10.1038/ng.3205.

Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, Hao K, Willard MD, Xu J, Hauptschein R, Rejto PA, Fernandez J, Wang G, Zhang Q, Wang B, Chen R, Wang J, Lee NP, Zhou W, Lin Z, Peng Z, Yi K, Chen S, Li L, Fan X, Yang J, Ye R, Ju J, Wang K, Estrella H, Deng S, Wei P, Qiu M, Wulur IH, Liu J, Ehsani ME, Zhang C, Loboda A, Sung WK, Aggarwal A, Poon RT, Fan ST, Wang J, Hardwick J, Reinhard C, Dai H, Li Y, Luk JM, Mao M. 2013. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Research* **23**:1422–1433 DOI 10.1101/gr.154492.113.

Kong KW, Abdul AA, Razali N, Aminuddin N, Junit SM. 2016. Antioxidant-rich leaf extract of Barringtonia racemosa significantly alters the in vitro expression of genes encoding enzymes that are involved in methylglyoxal degradation III. *PeerJ* **4**:e2379 DOI 10.7717/peerj.2379.

Kong LM, Liao CG, Chen L, Yang HS, Zhang SH, Zhang Z, Bian HJ, Xing JL, Chen ZN. 2011. Promoter hypomethylation up-regulates CD147 expression through increasing Sp1 binding and associates with poor prognosis in human hepato-cellular carcinoma. *Journal of Cellular and Molecular Medicine* **15**:1415–1428 DOI 10.1111/j.1582-4934.2010.01124.x.

Lengauer C, Kinzler KW, Vogelstein B. 1998. Genetic instabilities in human cancers. *Nature* **396**:643–649 DOI 10.1038/25292.

Liu K, Li Y, Yu B, Wang F, Mi T, Zhao Y. 2018. Silencing non-SMC chromosome-associated polypeptide G inhibits proliferation and induces apoptosis in hepatocellular carcinoma cells. *Canadian Journal of Physiology and Pharmacology* **96**:1246–1254 DOI 10.1139/cjpp-2018-0195.

Ma HP, Chang HL, Bamodu OA, Yadav VK, Huang TY, Wu A, Yeh CT, Tsai SH, Lee WH. 2019. Collagen 1A1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis. *Cancer* **11(6)**:786 DOI 10.3390/cancers11060786.

Mayhew CN, Carter SL, Fox SR, Sexton CR, Reed CA, Srinivasan SV, Liu X, Wikenheiser-Brokamp K, Boivin GP, Lee JS, Aronow BJ, Thorgeirsson SS, Knudsen ES. 2007. RB loss abrogates cell cycle control and genome integrity to promote liver tumorigenesis. *Gastroenterology* **133**:976–984 DOI 10.1053/j.gastro.2007.06.025.

Meng QC, Wang HC, Song ZL, Shan ZZ, Yuan Z, Zheng Q, Huang XY. 2015. Overexpression of NDC80 is correlated with prognosis of pancreatic cancer and regulates cell proliferation. *American Journal of Cancer Research* **5**:1730–1740.

Naugler WE, Sakurai T, Kim S, Maeda S, Kim K, Elsharkawy AM, Karin M. 2007. Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. *Science* **317**:121–124 DOI 10.1126/science.1140485.

R Core Team. 2018. R: a language and environment for statistical computing. Version 3.5.1. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/* .

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**:e47 DOI 10.1093/nar/gkv007.

Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F, Calatayud AL, Pinyol R, Pelletier L, Balabaud C, Laurent A, Blanc JF, Mazzaferro V, Calvo F, Villanueva A, Nault JC, Bioulac-Sage P, Stratton MR, Llovet JM, Zucman-Rossi J. 2015. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nature Genetics* **47**:505–511 DOI 10.1038/ng.3252.

Shahid M, Lee MY, Piplani H, Andres AM, Zhou B, Yeon A, Kim M, Kim HL, Kim J. 2018. Centromere protein F (CENPF), a microtubule binding protein, modulates cancer metabolism by regulating pyruvate kinase M2 phosphorylation signaling. *Cell Cycle* **17**:2802–2818 DOI 10.1080/15384101.2018.1557496.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504 DOI 10.1101/gr.1239303.

Shimizu I. 2003. Impact of oestrogens on the progression of liver disease. *Liver International* **23**:63–69 DOI 10.1034/j.1600-0676.2003.00811.x.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide

expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**:15545–15550 DOI 10.1073/pnas.0506580102.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, Von Mering C. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* **45**:D362–D368 DOI 10.1093/nar/gkw937.

Tang Y, Li M, Wang J, Pan Y, Wu FX. 2015. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* **127**:67–72 DOI 10.1016/j.biosystems.2014.11.005.

Wang K, Lim HY, Shi S, Lee J, Deng S, Xie T, Zhu Z, Wang Y, Pocalyko D, Yang WJ, Rejto PA, Mao M, Park CK, Xu J. 2013. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* **58**:706–717 DOI 10.1002/hep.26402.

Wang CH, Wey KC, Mo LR, Chang KK, Lin RC, Kuo JJ. 2015. Current trends and recent advances in diagnosis, therapy, and prevention of hepatocellular carcinoma. *Asian Pacific Journal of Cancer Prevention* **16**:3595–3604 DOI 10.7314/APJCP.2015.16.9.3595.

Wang SH, Yeh SH, Lin WH, Yeh KH, Yuan Q, Xia NS, Chen DS, Chen PJ. 2012. Estrogen receptor alpha represses transcription of HBV genes via interaction with hepatocyte nuclear factor 4alpha. *Gastroenterology* **142**:989–998 DOI 10.1053/j.gastro.2011.12.045.

Wang J, Zhou Y, Fei X, Chen X, Chen Y. 2018. Biostatistics mining associated method identifies AKR1B10 enhancing hepatocellular carcinoma cell growth and degenerated by miR-383-5p. *Scientific Reports* **8**:11094 DOI 10.1038/s41598-018-29271-3.

Wen LZ, Ding K, Wang ZR, Ding CH, Lei SJ, Liu JP, Yin C, Hu PF, Ding J, Chen WS, Zhang X, Xie WF. 2018. SHP-1 acts as a tumor suppressor in hepatocarcinogenesis and HCC progression. *Cancer Research* **78**:4680–4691 DOI 10.1158/0008-5472.CAN-17-3896.

Xing T, Yan T, Zhou Q. 2018. Identification of key candidate genes and pathways in hepatocellular carcinoma by integrated bioinformatical analysis. *Experimental and Therapeutic Medicine* **15**:4932–4942 DOI 10.3892/etm.2018.6075.

Yin L, Chang C, Xu C. 2017. G2/M checkpoint plays a vital role at the early stage of HCC by analysis of key pathways and genes. *Oncotarget* **8**:76305–76317 DOI 10.18632/oncotarget.19351.

Ying H, Xu Z, Chen M, Zhou S, Liang X, Cai X. 2018. Overexpression of Zwint predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins. *OncoTargets and Therapy* **11**:689–702 DOI 10.2147/OTT.S152138.

Zhang X, Kang C, Li N, Liu X, Zhang J, Gao F, Dai L. 2019. Identification of special key genes for alcohol-related hepatocellular carcinoma through bioinformatic analysis. *PeerJ* **7**:e6375 DOI 10.7717/peerj.6375.