Centromere protein N may be a novel malignant prognostic biomarker for hepatocellular carcinoma

Qingqing Wang [1, 2], **Xiaoyan Yu** [1, 2], **Zhewen Zheng** [1, 2], **Fengxia Chen** [1, 2], **Ningning Yang** [1, 2], **Yunfeng Zhou** [Corresp. 1, 2]

[1] Hubei Cancer Clinical Study Center, Hubei Key Laboratory of Tumor Biological Behaviors, Zhongnan Hospital, Wuhan University, Wuhan, China

[2] Department of Radiation Oncology and Medical Oncology, Zhongnan Hospital, Wuhan University, Wuhan, China

Corresponding Author: Yunfeng Zhou
Email address: yfzhouwhu@163.com

Background. Hepatocellular carcinoma (HCC) is one of the deadliest tumors. The majority of HCC is detected in the late stage, and the clinical results for HCC patients are poor. There is an urgent need to discover early diagnostic biomarkers and potential therapeutic targets for HCC. **Methods.** The GSE87630 and GSE112790 datasets from the Gene Expression Omnibus (GEO) database were downloaded to analyze the differentially expressed genes (DEGs) between HCC and normal tissues. R packages were used for Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses of the DEGs. A Search Tool for Retrieval of Interacting Genes (STRING) database was used to develop a protein-protein interaction (PPI) network, and also cytoHubba, Molecular Complex Detection (MCODE), EMBL-EBI, CCLE, Gene Expression Profiling Interactive Analysis (GEPIA), and Oncomine analyses were performed to identify hub genes. Gene expression was verified with a third GEO dataset, GSE25097. The Cancer Genome Atlas (TCGA) database was used to explore the correlations between the hub genes and clinical indexes of HCC patients. The functions of the hub genes were enriched by gene set enrichment analysis (GSEA), and the biological significance of the hub genes was explored by real-time polymerase chain reaction (qRT-PCR), western blot, immunofluorescence, CCK-8, colony formation, Transwell and flow cytometry assays with loss-of-function experiments *in vitro*. **Results.** Centromere protein N (CENPN) was screened as a hub gene affecting HCC tumorigenesis. Evaluation by Cox regression showed that a high level of CENPN expression was an independent danger variable for poor prognosis of HCC. GSEA showed that high CENPN expression was linked to the following pathways: liver cancer subclass proliferation, cell cycle, p53 signaling pathway, Rb1 pathway, positive regulation of cell cycle G1/S phase transition, and DNA damage response signal transduction by p53 class moderators. Further cell experiments showed that knocking down CENPN expression decreased the proliferation and colony-forming abilities of HepG2 and Huh7 cells as well as Ki67 expression in these cell lines. The cell

cycle was arrested in G1 phase, which is consistent with previous experiments on CENPN downregulation., but neither migration nor invasion were significantly affected. Western blot results revealed that the expression of p53, p27, p21, CDK4, cyclin D1, CDK2, cyclin E, pRb, E2F1 and c-myc decreased after CENPN knockdown, but there was no significant change in total Rb levels. In addition, CENPN-knockdown cells subjected to irradiation showed significantly enhanced of γ-H2AX expression and reduced colony formation.
Conclusion. CENPN functions as an oncogene in HCC and may be a therapeutic target and promising prognostic marker for HCC.

1   **Centromere protein N may be a novel malignant prognostic biomarker for hepatocellular**

2   **carcinoma**

4   Qingqing Wang[1,2], Xiaoyan Yu[1,2], Zhewen Zheng[1,2], Fengxia Chen[1,2], Ningning Yang[1,2],

5   Yunfeng Zhou[1,2,*]

6   [1]Hubei Cancer Clinical Study Center, Hubei Key Laboratory of Tumor Biological Behaviors,

7   Zhongnan Hospital, Wuhan University, Wuhan, Hubei, 430071, China

8   [2]Department of Radiation Oncology and Medical Oncology, Zhongnan Hospital, Wuhan

9   University, Wuhan, Hubei, 430071, China.

10   Corresponding Author:

11   Yunfeng Zhou

12   169 Donghu Road, Wuhan, Hubei, 430071, China.

13   E-mail address: yfzhouwhu@163.com.

14   **Abstract**

15   **Background.** Hepatocellular carcinoma (HCC) is one of the deadliest tumors. The majority of

16   HCC is detected in the late stage, and the clinical results for HCC patients are poor. There is an

17   urgent need to discover early diagnostic biomarkers and potential therapeutic targets for HCC.

18   **Methods.** The GSE87630 and GSE112790 datasets from the Gene Expression Omnibus (GEO)

19   database were downloaded to analyze the differentially expressed genes (DEGs) between HCC

20   and normal tissues. R packages were used for Kyoto Encyclopedia of Genes and Genomes

21   (KEGG) and Gene Ontology (GO) enrichment analyses of the DEGs. A Search Tool for

22   Retrieval of Interacting Genes (STRING) database was used to develop a protein-protein

23   interaction (PPI) network, and also cytoHubba, Molecular Complex Detection (MCODE),

24   EMBL-EBI, CCLE, Gene Expression Profiling Interactive Analysis (GEPIA), and Oncomine

25   analyses were performed to identify hub genes. Gene expression was verified with a third GEO

26   dataset, GSE25097. The Cancer Genome Atlas (TCGA) database was used to explore the

27   correlations between the hub genes and clinical indexes of HCC patients. The functions of the

28   hub genes were enriched by gene set enrichment analysis (GSEA), and the biological

29   significance of the hub genes was explored by real-time polymerase chain reaction (qRT-PCR),

30   western blot, immunofluorescence, CCK-8, colony formation, Transwell and flow cytometry

31   assays with loss-of-function experiments *in vitro*.

32   **Results.** Centromere protein N (CENPN) was screened as a hub gene affecting HCC

33   tumorigenesis. Evaluation by Cox regression showed that a high level of CENPN expression was

34   an independent danger variable for poor prognosis of HCC. GSEA showed that high CENPN

35   expression was linked to the following pathways: liver cancer subclass proliferation, cell cycle,

36   p53 signaling pathway, Rb1 pathway, positive regulation of cell cycle G1/S phase transition, and

37   DNA damage response signal transduction by p53 class moderators. Further cell experiments

38   showed that knocking down CENPN expression decreased the proliferation and colony-forming

39   abilities of HepG2 and Huh7 cells as well as Ki67 expression in these cell lines. The cell cycle

40   was arrested in G1 phase, which is consistent with previous experiments on CENPN

41   downregulation., but neither migration nor invasion were significantly affected. Western blot

42   results revealed that the expression of p53, p27, p21, CDK4, cyclin D1, CDK2, cyclin E, pRb,

43   E2F1 and c-myc decreased after CENPN knockdown, but there was no significant change in

44   total Rb levels. In addition, CENPN-knockdown cells subjected to irradiation showed

45   significantly enhanced of γ-H2AX expression and reduced colony formation.

46   **Conclusion.** CENPN functions as an oncogene in HCC and may be a therapeutic target and

47   promising prognostic marker for HCC.

48   **Introduction**

49   Liver cancer is among the most common malignant tumors. Despite a declining mortality rate,

50   liver cancer remains to be one of the leading 10 causes of cancer-related fatalities in many

51   countries (Siegel et al. 2020). Hepatocellular carcinoma (HCC) is the most common type of liver

52   cancer, as well as its occurrence and development are closely related to genetic changes, genetic

53   susceptibility and alterations in key signaling pathways (Marquardt & Thorgeirsson 2014). HCC

54   is a highly heterogeneous disease (Calderaro et al. 2019); however, due to the limited availability

55 of HCC markers for early diagnosis, most HCC is detected at an advanced disease stage, which

56 limits the effectiveness of common treatment methods such as surgical resection and

57 chemotherapy (Forner et al. 2018). Therefore, it is vital to acquire a better understanding of HCC

58 and also to identify new early diagnostic as well as therapeutic targets.

59 Increasing knowledge of the diversity and heterogeneity of tumors and the completion of the

60 Human Genome Project have actually promoted the wider use of modern high-throughput

61 sequencing technology. In addition, researchers have begun to use bioinformatics analysis

62 techniques to determine differentially expressed genes (DEGs) and functional paths associated

63 with tumorigenesis (Can 2014).

64 In this study, by analyzing the GSE87630 and GSE112790 datasets, we found that the DEG

65 CENPN was related to the occurrence of HCC and confirmed this finding in the GSE25097 and

66 also TCGA-LIHC. We observed that upregulated CENPN expression was related to low survival

67 and the advanced T classification of HCC patients. Functional studies showed that knocking

68 down the CENPN gene could inhibit cell proliferation and increase the cytotoxic effect of X-rays

69 on cells *in vitro*. Further mechanistic studies showed that CENPN partly regulated the cell cycle

70 through the p27/p21-Rb/E2F1 axis.

71 **Materials & Methods**

72 **Microarray data**

73 The gene expression profiles of the GSE87630, GSE112790 and GSE25097 datasets were

74 downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression

75   Omnibus (GEO) database (Edgar et al. 2002). GSE87630 contains 64 HCC samples and 30

76   normal samples and is based on the GLP6947 platform (Illumina HumanHT-12 V3.0 Expression

77   BeadChip) (Woo et al. 2017); GSE112790 contains 183 HCC samples and 15 normal samples

78   and is based upon the GLP570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133

79   Plus 2.0 Array) (Shimada et al. 2019); and GSE25097 contains 268 HCC samples, 243 adjacent

80   nontumor samples, and 40 cirrhotic liver samples, and also 6 healthy liver samples and is based

81   upon the GPL10687 platform (Rosetta/Merck Human RSTA Affymetrix 1.0 microarray, Custom

82   CDF) (Tung et al. 2011).

83   **Identification of DEGs**

84   DEGs between HCC samples and normal samples were identified utilizing GEO2R (Barrett et al.

85   2013), a tool for the online analysis of differential genes between datasets in a GEO series, with

86   the following thresholds: $|\,logFC\,| > 1$ and a false discovery rate (FDR) < 0.05. Volcano plots

87   were constructed with GraphPad Prism 8 (GraphPad Software Inc.). Heatmaps were drawn with

88   the heatmap R package (Galili et al. 2018). Overlapping DEGs in the two datasets were detected

89   with the VennDiagram R package (Chen & Boutros 2011).

90   **Functional enrichment analysis**

91   Gene Ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway

92   enrichment analyses were mainly utilized to explore the functions of these identified DEGs. GO

93   consists three components: biological process (BP), cellular component (CC) as well as molecular

94   function (MF) (Ashburner et al. 2000). KEGG mainly evaluates the pathways in which DEGs may

95   be involved (Kanehisa & Goto 2000). The results were visualized using clusterProfiler and ggplot2

96   R package (threshold: $P < 0.05$) (Ito & Murphy 2013; Yu et al. 2012).

97   **Protein-protein interaction (PPI) network construction**

98   The PPI network comprising all the DEGs was predicted and analyzed by utilizing the Search

99   Tool for Retrieval of Interacting Genes (STRING) database (http://string-db.org) (Szklarczyk et

100   al. 2015) S. A combined score > 0.7 was taken into consideration significant. The PPI network

101   was used to clarify mechanisms relevant to the event and growth of HCC.

102   **Hub gene selection**

103   The public bioinformatics software Cytoscape is a system for visualizing complicated networks

104   and incorporating related information (Smoot et al. 2011). The cytoHubba plugin application of

105   Cytoscape was utilized to screen hub genes in a network through 11 topological analysis

106   approaches (Chin et al. 2014). In this study, the intersecting sets of the top 30 genes revealed

107   with the maximal clique centrality (MCC) and density of maximum neighborhood component

108   (DMNC) methods were visualized as a Venn diagram

109   (http://bioinformatics.psb.ugent.be/webtools/Venn/). Molecular Complex Detection (MCODE)

110   was utilized to screen within the PPI network with the following parameters: degree cutoff > 2,

111   node score cutoff > 0.2, K-core > 2 and max dep > 100 (Gary D Bader 2003). Gene Expression

112   Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index.html), an online tool

113   that consists of information from The Cancer Genome Atlas (TCGA) and Genotype-Tissue

114   Expression (GTEx) databases, was utilized to analyze the expression of the genes as well as their

115   association with patient survival and clinical stage (Tang et al. 2017). Then, after combining the

116   statistical results of gene expression from the Oncomine (http://www.oncomine.com) (Rhodes et

117   al. 2004), EMBL-EBI (https://www.ebi.ac.uk) and CCLE (https://www.broadinstitute.org/ccle)

118   databases, the statistically significant hub genes were screened. EMBL-EBI, a web-based tool,

119   provides a series of bioinformatics applications for sequence analysis (Li et al. 2015). CCLE

120   contains genomic data, analytical data, and visualization data for approximately 1000 cell lines

121   (Barretina et al. 2012). The levels of expression of the hub genes in clinical samples were

122   compared with those in the GSE87630, GSE112790 and GSE25097 datasets using GraphPad

123   Prism 8 (GraphPad Software Inc.). The area under the curve (AUC) assessed by the pROC R

124   package was used to validate the prognostic efficacy of the hub genes in liver cancer (Robin et

125   al. 2011).

126   **Database analysis of the differential expression and prognosis of the CENPN gene**

127   CENPN gene expression in the TCGA liver cancer database (TCGA-LIHC) was extracted. The

128   limma R package was utilized to analyze the different scatter plots, and the survival R package

129   was used to draw survival curves, and also the limma and ggpubr R packages were utilized to

130   analyze CENPN expressing in the different clinical stages (Holleczek & Brenner 2013; Ritchie et

131   al. 2015).

132   **Gene set enrichment analysis (GSEA)**

133   The TCGA-LIHC gene expression dataset was downloaded and installed from the Xena web

134   browser (https://xenabrowser.net/datapages). Two groups from the TCGA-LIHC dataset were

135   classified based on the median CENPN expression level (high and low CENPN). The GSEA

136   4.1.0 software program, which was downloaded from http://www.broad.mit.edu/gsea/, was

137   utilized to carry out GSEA with the following predefined gene sets (Subramanian et al. 2005):

138   hallmark gene sets, curated gene sets and gene ontology. The permutation number was

139   established as 1000, and an FDR <0.25 was considered significant.

140   **Cell culture and cell transfection**

141   The human HCC cell lines HepG2 and Huh7 were purchased from Procell Life Science &

142   Technology Company (Wuhan, China) and confirmed to contain no mycoplasma contamination

143   via STR analysis. Cells were cultured in DMEM supplemented with 10% fetal bovine serum and

144   1% penicillin/streptomycin at 37°C and 5% CO_2. A small interfering RNA (siRNA) targeting

145   CENPN and a negative control (NC) were purchased from GenePharma (Suzhou, China). We

146   used Lipofectamine 3000 (L3000015, Invitrogen, Waltham, USA) as the transfection reagent

147   according to the manufacturer's instructions. Forty-eight hours following transfection, the cells

148   were harvested for further experiments, including RNA or protein extraction.

149   **RNA extraction, reverse transcription as well as qRT-PCR**

150   Total RNA was extracted utilizing the traditional assay, that is to say using TRIzol reagent

151   (Invitrogen, USA). The cDNA synthesis with total RNA (1 μg) was carried out with the

152   Primescript™ RT reagent kit (Vazyme, Nanjing, China), and quantitative real-time PCR (qRT-

153   PCR) was conducted utilizing 2x SYBR® Green Supermix (Vazyme, Nanjing, China). Relative

154   gene expression was quantified via the 2-ΔΔCt approach and normalized to GAPDH expression.

155   The primer sequences used for amplification were as follows: CENPN forward, 5′-

156   CTGTGTGAGGAAAAGCGTGC-3′; CENPN reverse, 5′-TCACCTGGTCCTTTACTCATCTG-

157   3′; GAPDH forward, 5′-GGAGCGAGATCCCTCCAAAAT-3′, and GAPDH reverse, 5′-

158   GGCTGTTGTCATACTTCTCATGG-3′.

159   **X-irradiation**

160   Cells were exposed to X-rays at doses of 0, 2 and 10 Gy according to the experimental design.

161   The equipment used was a Siemens Primus Accelerator (6 Mv; Siemens AG, Munich,

162   Germany)t at Zhongnan Hospital of Wuhan University.

163   **CCK-8 and colony formation assays**

164   For the CCK-8 assay, cells were seeded into plates with 96 wells at 3000 cells per well. Ten

165   microliters of CCK-8 reagent (Dojindo Laboratories, Kumamoto, Japan) were added to each

166   well, and the plates were incubated at 37°C for 1-4 hours. The absorbance of the wells at 450 nm

167   was determined in a microplate reader (Molecular Devices, USA).

168   For the colony formation assay, 1000 HepG2 cells per 1 ml in an overall volume of 2 ml was

169   seeded into a plate with six wells. For Huh7 cells, 1500 cells per 1 ml in a total volume of 2 ml

170   was seeded per well. After 24 hours, the cells were irradiated with X-rays at doses of 0 Gy and 2

171   Gy. After 14 days, the cells were treated with 4% paraformaldehyde for 15 minutes as well as

172   stained with 0.1% crystal violet for 1 hour. Finally, the number of colonies was calculated after

173   washing the wells with water three times and drying.

174   **Flow cytometry analysis**

175   HCC cells were harvested 48 hours following transfection and washed in PBS. The cells were

176   stained with 1 ml of DNA staining solution and 10 μl of propidium iodide (PI) at room

177   temperature for 30 minutes protected from light and then were analyzed by flow cytometry (Cat.

178 #FC500, Beckman, USA).

179 **Transwell assay**

180 Cells transfected with siRNA after 24 hours were seeded in a Transwell chamber system

181 (Corning, USA). Approximately 2 x 10^5 HepG2 cells or 8 x 10^4 Huh7 cells in 200 μl of serum-

182 free medium were seeded in the upper chamber, while 600 μl of medium containing 20% serum

183 was added to the lower chamber. After a 48-hour incubation, cells in the upper chamber were

184 wiped off with cotton swabs, and those in the lower chamber were fixed and stained as described

185 in the colony formation experiment before they were photographed under an optical microscope.

186 **Western blotting**

187 Protein lysates were obtained from HepG2 and Huh7 cells with RIPA lysis buffer (#P0013B,

188 Beyotime Biotechnology, China) containing 1% cocktail (#HY-K0021, MCE, USA). The protein

189 concentration was detected with a BCA assay kit (Beyotime Biotechnology, Shanghai, China),

190 and 30 μg of protein sample was divided by SDS-PAGE through a 10% gel and then transferred

191 to PVDF membranes, which were blocked in 5% skim milk in TBST before they were blotted

192 with the appropriate primary antibody followed by the corresponding secondary antibody.

193 Detailed information on the antibodies used is listed in *Table S1*. An enhanced

194 chemiluminescence kit was utilized to develop the bands.

195 **Immunofluorescence staining**

196 The immunofluorescence method was based on a previously described method (Wang et al.

197   2017). Briefly, cells were seeded into a 6-well plate or on cell slides and cultured overnight.

198   After fixation for 15 minutes, the cells were incubated with 0.2% Triton X-100 for 15 minutes at

199   RT and blocked with 5% BSA for 30 minutes. Next, the cells were incubated successively with

200   primary antibody targeting CENPN, E-cadherin, N-cadherin, vimentin or γ-H2AX and

201   appropriate secondary antibody before they were stained with DAPI at RT for 5 minutes. Finally,

202   the cells were photographed under a fluorescence or confocal microscope.

203   **Statistical analysis**

204   All prognostic information from the TCGA-LIHC dataset was collected, and samples without

205   results were excluded. On the basis of optimal sample separation, Kaplan-Meier (K-M) survival

206   curves were generated to calculate survival, and the difference in survival between groups was

207   determined by the log-rank test. The relationships between variables and patient survival were

208   analyzed with Cox models, and the results were evaluated with SPSS 22.0 software (SPSS,

209   Chicago, IL, United States). All cell experimental results were individually replicated 3 times.

210   The values are reported as the means ± standard deviations and were statistically analyzed with

211   GraphPad Prism 8.0 software (GraphPad Software, La Jolla, CA, United States). Student's t-test

212   was utilized for comparisons between groups. $P < 0.05$ was thought about statistically

213   significant.

214   **Results**

215   **Preliminary screening and enrichment analysis of DEGs in HCC**

216   *Figure 1* illustrates the study design. The DEGs in HCC vs. normal tissues in the GSE87630 and

217   GSE112790 datasets were analyzed by processing and standardizing the gene expression profile

218   data through GEO2R (*Fig. 2A, B, Fig. S1 and S2*). GSE87630 contains 1162 DEGs, of which

219   394 are upregulated and 768 are downregulated, whereas GSE112790 contains 1713 DEGs

220   comprising 963 upregulated DEGs and 750 downregulated DEGs. Then, R was used to intersect

221   the two datasets, and a total of 532 overlapping DEGs were acquired: 171 upregulated genes and

222   361 downregulated genes (*Fig. 2C, Table S2*).

223   To clarify the functions of these 532 DEGs, GO as well as KEGG enrichment were carried out,

224   and the ggplot2 package of R was used to visualize the top 10 per classification of GO and the

225   top 35 of KEGG. In the GO BP category, the identified DEGs were related to the organic acid

226   catabolic process and epoxygenase P450 pathway (*Fig. 2D, Table S3*). In the GO CC category,

227   the DEGs were enriched in MCM complex and blood microparticle extracellular matrix (*Fig.*

228   *2D, Table S3*). Finally, in the GO MF classification, the DEGs were enriched in monooxygenase

229   activity and iron ion binding (*Fig. 2D, Table S3*). KEGG pathway enrichment result showed that

230   these DEGs play important roles in multiple key signaling pathways, including DNA replication,

231   the cell cycle as well as the p53 signaling pathway (*Fig. 2E, Table S4*).

232   **Construction of the PPI network as well as screening of hub genes**

233   The STRING database was utilized to analyze the potential interacting proteins among 532

234   DEGs with a combined score > 0.7. For the PPI network, which contained 347 nodes and 1560

235   edges, was built with Cytoscape software (*Fig. 3A*). Then, the hub genes were analyzed with the

236   cytoHubba plugin. Two topological analysis approaches, MCC and DMNC, were used to rank

237   the top 30 nodes in the created PPI network (*Fig. 3C and D*, *Table 1*). Then, the intersection was

238   calculated online using a Venn diagram (*Fig. 3E, Table S5*). Module analysis was performed

239   with MCODE, and the top 5 modules were listed (*Table S6*). CENPN was screened as the seed

240   gene in module 1 (the most significant module), which consisted of 32 nodes as well as 479

241   edges (*Fig. 3B, Table S7*). Thus, CENPN was considered the hub gene.

242   **Study of CENPN expression in HCC cells and tissues via online databases**

243   CCLE and EMBL-EBI were used to explore the RNA expression level of CENPN in HCC cell

244   lines. The results revealed that CENPN was highly expressed in liver cancer cells (*Fig. 4A and*

245   *B*). With respect to tissue expression levels, Oncomine evaluation revealed that CENPN

246   expression was significantly higher in tumor tissues than in normal tissues (*Fig. 4F*), and

247   CENPN expression in the GSE87630 and GSE112790 datasets was consistent with the above

248   results (*Fig. 4C and D*); furthermore, this result was verified in the GSE25097 dataset (*Fig. 4E*).

249   GEPIA revealed that the overall survival (OS) and disease-free survival (DFS) of liver cancer

250   patients with high CENPN expression were shorter than those of patients with low CENPN

251   expression (*Fig. 4G and H*). To better understand the accuracy of CENPN in HCC

252   tumorigenesis, receiver operating characteristic (ROC) curves were drawn, and the AUC values

253   of GSE87630, GSE112790 and GSE25097 were 0.913, 0.904 and 0.787, respectively (*Fig. 4 I-*

254   *K*). Our results revealed that CENPN was highly expressed in liver cancer as well as affected the

255   prognosis of patients.

256   **Verification of the correlation between CENPN and clinicopathological features of HCC**

257   **patients through the TCGA database**

258   The optimal cutoff value was calculated by the survminer package of R, and the patients were

259   separated into 2 groups: high CENPN expression and low CENPN expression. Compared with

260   the low expression group, the high expression group had dramatically shorter survival (*Fig. 5C*).

261   When the clinical features of CENPN were combined in the analysis, the Wilcoxon rank-sum

262   test revealed that CENPN expression in tumor tissues was dramatically higher than that in

263   normal(*Fig. 5A*). Similar results were obtained in the paired analysis of normal and tumor tissues

264   from the same patient (*Fig. 5B*). These results suggest that CENPN expression is negatively

265   associated with HCC prognosis. In particular, with the progression of grade, stage and T stage,

266   CENPN expression showed an upward trend (*Fig. 5D-F*). Furthermore, univariate Cox analysis

267   revealed that CENPN, stage, and T and M stages affected HCC prognosis (*Fig. 6A-C*, *Table 2*).

268   Multivariate Cox analysis showed that high CENPN expression was an independent predictor of

269   poor prognosis in HCC patients (*Table 2*). In summary, CENPN acts as an oncogene in liver

270   cancer.

271   **GSEA of the CENPN gene**

272   GSEA showed that high CENPN expression was mainly associated with liver cancer subclass

273   proliferation, the cell cycle, the p53 signaling pathway, the Rb1 pathway, cell cycle checkpoints,

274   positive regulation of the G1/S phase transition, reactome SCF SKP2-mediated degradation of

275   p27 p21, E2F targets and more correlated gene sets (*P*<0.05; *Fig. 7A-H*). High CENPN

276   expression was also associated with DNA damage, DNA damage detection, and DNA damage

277   response signal transduction by p53 class mediators and to other DNA damage and repair-related

278   gene sets (*P*<0.05; *Fig. 8A*).

279   **CENPN downregulation arrests the cell cycle in G1 phase and inhibits HCC cell**

280   proliferation but has no effect on migration and invasion

281   The above results suggest that CENPN functions as a biomarker for the diagnosis of HCC. To

282   further explore the mechanism by which CENPN impacts the biological behavior of HCC, we

283   established a cell model of CENPN deficiency (HepG2 and Huh7) by transfecting siRNA

284   targeting CENPN and NC siRNA. qRT-PCR, western blot and immunofluorescence analyses

285   were performed 48 hours after transfection to confirm knockdown (*Fig. 7I, J, M and O*). CCK-8

286   assays revealed that CENPN deficiency considerably inhibited the proliferation and viability of

287   HCC cells (*Fig. 7K, L*), and immunofluorescence analysis showed that when CENPN was

288   knocked down, Ki67 expression decreased significantly (*Fig. 7N, P*). Cell cycle analysis showed

289   that CENPN deficiency significantly inhibited the transition from G1 to S phase in HCC cells

290   (*Fig. 7Q-X*). However, Transwell assays revealed that there was no difference in the numbers of

291   cells that migrated or invaded after CENPN knockdown compared with those subjected to NC

292   siRNA (*Fig. S3A, B*). Immunofluorescence results indicated that there was no significant

293   difference in E-cadherin, N-cadherin and vimentin expression between the groups (*Fig. S3C-G*),

294   and the western blot results were consistent with the above data (*Fig. S3H, I*). In summary, these

295   findings indicate that CENPN knockdown inhibits HCC cell proliferation but does not affect

296   invasion or migration.

297   CENPN affects the p21-CDK2/cyclin E, p27-CDK4/cyclin D and Rb/E2F1 signaling

298   pathways in HCC

299   To elucidate the mechanism by which CENPN promotes HCC, we used western blotting to

300   analyze the expression levels of related markers. The western blot results revealed that compared

301   with the NC, the CENPN-knockdown group showed enhanced expression of P53, P27 and P21

302   and also lowered expression of CDK4, cyclin D1, CDK2, cyclin E as well as pRb, but there was

303   no significant change in total Rb levels. In addition, the expression of E2F1 and c-Myc decreased

304   (*Fig. 8F*). Taken together, these results indicate that CENPN downregulation prevents cell

305   proliferation via regulating p21 and p27.

306   **Decreased CENPN expression promotes radiation damage in HCC cells**

307   GSEA indicated that CENPN is closely related to DNA damage and repair functions. Cells

308   transfected with siRNA-CENPN and NC for 48 hours were treated with X-rays at doses of 0 Gy

309   and 10 Gy, and γ-H2AX foci were detected by immunofluorescence 24 hours later. The number

310   of γ-H2AX foci increased after CENPN knockdown and was further pronounced upon 10 Gy X-

311   ray exposure (*Fig. 8B, C*). Moreover, colony formation assays suggested that the number of HCC

312   cell colonies formed after CENPN knockdown combined with X-ray treatment was significantly

313   lower than that after siRNA-CENPN treatment alone (*Fig. 8D, E*). These results suggest that

314   interfering with CENPN expression enhances X-ray-induced radiation damage in HCC cells.

315   **Discussion**

316   The prognosis of HCC patients is poor, and a major of patients can only receive palliative

317   treatment. The effect of surgical treatment is satisfactory, but few patients benefit from it. Only

318   1/3 of patients are eligible for radical treatments such as percutaneous ablation, surgical resection

319   or liver transplantation (Forner et al. 2012). In addition, patients often miss the optimal

320   timeframe for surgery when they are diagnosed, so there is an urgent need to identify new

321   biomarkers and develop new treatment strategies.

322   In this study, through a comprehensive bioinformatics analysis, 532 DEGs were identified in the

323   GSE87630 and GSE112790 datasets. GO and KEGG analyses showed that the DEGs were closely

324   related to organic acid catabolic process, epoxygenase P450 pathway, MCM complex, blood

325   microparticle extracellular matrix, DNA replication, the cell cycle, the p53 signaling pathway and

326   additional items. PPI network and Cytoscape analyses revealed CENPN as a new DEG in HCC.

327   CENPN is located on chromosome 16q23.2 and encodes nucleosome-related complexes, which

328   are critical for dynamic assembly. CENPN can recognize the histone H3 variant CENPA (a

329   centromere-specific nucleosome that plays a fundamental role in centromere assembly) in the

330   centromeric nucleosome, and it is correctly controlled by the formation of the CENPA

331   nucleosome-associated complex (NAC) to regulate cell mitosis (Chittori et al. 2018; Foltz et al.

332   2006). This correct assembly ensures accurate chromosome separation and avoids diseases

333   caused by aneuploidy due to chromosome misseparation (Kops et al. 2005). The consumption of

334   CENPN can lead to the downregulation of several CENPs, and this downregulation is considered

335   a necessary condition for the manufacture of new centromeres (Tian et al. 2018). CENPN is

336   closely related to the occurrence and development of many kinds of cancer. Sarah An *et al*

337   showed by multivariate analysis that an increase in CENPN expression could significantly

338   increase the recurrence and mortality rates of breast cancer patients with a smoking history

339   (Andres et al. 2015). Moreover, Daniel J. Wright *et al*. identified CENPN as a gene associated

340   with aneuploidy, genomic instability and cancer susceptibility based on intensity data and

341   sequences of genotypic arrays (Wright et al. 2017). CENPN regulates cell proliferation and cell

342   cycle progression by affecting the mitotic process (Wright et al. 2017). Md. Rezanur Rahman *et*

343   *al*. identified CENPN as a poor prognostic marker for colorectal cancer by bioinformatics

344   analysis (Rahman et al. 2019). Recently, it was reported that CENPN can promote the

345   proliferation of oral cancer cells in the entrance cavity by regulating the cell cycle (Oka et al.

346   2019). However, the expression of CENPN in HCC and its role and mechanism in malignant

347   development are unclear.

348   As verified by a third GEO dataset and TCGA data, CENPN is highly expressed in HCC, and the

349   expression level of CENPN was positively associated with grade, stage and T classification and

350   negatively correlated with patient prognosis. Further Cox regression analysis showed that

351   CENPN is a potential independent prognostic variable for HCC. In summary, CENPN may be an

352   oncogene in HCC.

353   To better understand the function of CENPN in HCC, we used GSEA and found that in terms of

354   the hallmarks of HCC, CENPN was related to the cell cycle, DNA damage and repair, and also

355   other functional items. Moreover, the findings revealed that abnormal expression of the CENPN

356   gene is related to the incident and also progression of HCC. This research is the first to report the

357   promoting effect of CENPN on the growth of human HCC cells. CENPN downregulation can

358   hinder the proliferation of HCC cells and cause G0/G1 phase arrest. The results of our cell

359   experiments suggest that CENPN plays a role as an oncogene. In addition, we examined the

360   expression of epithelial-mesenchymal transition (EMT)-related markers (E-cadherin, N-cadherin

361   as well as vimentin) to understand the oncogenic function of CENPN as fully as possible (Hugo

362   et al. 2007); however, there were no difference in expression.

363   To much better understand the potential molecular mechanism of CENPN in HepG2 and Huh7

364   cells, we focused on three signaling pathways according to the GSEA results—P27, P21 and

365   Rb/E2F1, which play a vital role in HCC and various cancers (Gartel 2009; Razavipour et al.

366   2020; Rubin et al. 2020). Our results show that interfering with CENPN expression in HCC can

367   inhibit c-myc and cyclin E expression by activating the p27-CDK4/cyclin D1 and p21-

368   CDK2/cyclin E axes, thus reducing the level of phosphorylated Rb and inhibiting E2F1

369   transcription. P21 (Eldeiry et al. 1993; Harper et al. 1993) and p27 (Polyak et al. 1994;

370   Toyoshima & Hunter 1994) are well-known cyclin-dependent kinase inhibitors that mediate cell

371   cycle arrest (Nan et al. 2004). When p21 and p27 interact with their respective cyclin binding

372   companions, they block the kinase activity of CDKs (Abukhdeir & Park 2008). P21 prevents the

373   binding of cyclin E and CDK2 (Gu et al. 1993) and is directly regulated by p53 (Abukhdeir &

374   Park 2008) while p27 prevents the binding of cyclin D1 and CDK4 (Hunter & Pines 1994);

375   therefore, p21 and p27 inhibit the G1/S phase transition (Alt et al. 2002; Cheng et al. 1999). The

376   inactivation of these complexes results in other proteins not being phosphorylated, including pRb

377   (RB1), so it is unable to release E2F transcription factors to activate the expression of genes that

378   regulate S phase (Matsuoka et al. 1994). Previous research has revealed that the negative

379   regulation of cyclin D1 and CDK4 is consistent with the downregulation of CENPN (Molenaar

380   et al. 2008), and recent reports in oral squamous cell carcinoma (OSCC) suggest that

381   downregulation of CENPN can reduce the expression of cyclin D1 and CDK4 with

382   corresponding increases in the expression of p21 and p27 (Oka et al. 2019). This suggests that

383   interfering with CENPN expression in OSCC can upregulate p27 and p21 expression to inhibit

384   CDK4 and cyclin D1 expression and arrest the cell cycle in G1 phase. In addition, some

385   researches have revealed that decreased p27 and p21 expression is closely correlated with the

386   histological grade, metastasis and prognosis of HCC (Fiorentino et al. 2000; Qin & Ng 2001). In

387   papillary carcinoma (PTC), CITED1-meditaed interference of p21 and p27 expression can

388   increase the level of phosphorylated Rb and the transcriptional activity of E2F1, which leads to

389   PTC cell proliferation (Li et al. 2018). Xing et al reported that in HCC cells, haprolid upregulates

390   the expression of p21 and p27 and inhibits the G1/S phase transition of cells, which may be

391   related to the downregulation of Rb/E2F expression (Xing et al. 2020). Therefore, it can be

392   speculated that CENPN reduces the level of phosphorylated Rb and inhibits E2F1 transcription

393   by downregulating the activities of the p27-CDK4/cyclin D1 and p21-CDK2/cyclin E axes, thus

394   inhibiting c-myc as well as cyclin E, inhibiting cell proliferation, and ultimately affecting the

395   prognosis of liver cancer (*Fig. 8E*). However, the potential mechanism by which CENPN

396   inactivates p27 and p21 and whether it regulates p21 by inhibiting p53 require further study.

397   More importantly, we found that knocking down CENPN expression increased X-ray-induced γ-

398   H2AX expression. γ-H2AX is an important indicator of radiation-induced DNA double-strand

399   breaks (DSBs) (Hohmann et al. 2018). Disruptions in the p53 protein are associated with greater

400   than 50% of human cancers and, p53 is a crucial molecule involved in the cellular response to

401   radiotherapy (Cuddihy & Bristow 2004; Levine & Oren 2009). GSEA suggested that CENPN

402   could be enriched in the p53 pathway, and western blot analysis showed that p53 expression

403   increased with CENPN knockdown, so we inferred that CENPN interferes with DNA damage by

404   inhibiting p53 in HCC (*Fig. 8G*).

405   **Conclusion**

406   In conclusion, our work showed for the first time that CENPN plays a carcinogenic role in HCC

407   tumorigenesis. The upregulation of CENPN expression was significantly associated with poor

408   patient survival, and knocking down CENPN expression inhibits cell proliferation, at least in part

409   through the p27-CDK4/cyclin D1, p21-CDK2/cyclin E and Rb/E2F1 pathways. In addition,

410   CENPN knockout increases radiotherapy-induced DNA damage. Therefore, CENPN can be used

411   as not only a useful biomarker for diagnosis and prognosis but also a potential therapeutic target

412   for HCC. In future work, we will extend to the regulatory network of CENPN, such as miRNA

413   and transcription factors involved in regulation. Moreover, we will study the carcinogenic effects

414   of CENPN by overexpressing it in normal liver cells, exploring its function *in vivo*, and

415   ultimately develop a more effective strategy for the treatment of HCC patients.

416   **Acknowledgments**

417   Not applicable.

418   **Data sharing statement**

419   The datasets analyzed in this research can be downloaded from the GEO website

420   (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87630,

421   https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112790, and

422   https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE25097) and also the TCGA website

423   (https://portal.gdc.cancer.gov/cart).

424

References

Abukhdeir AM, and Park BH. 2008. P21 and p27: roles in carcinogenesis and drug resistance. *Expert Rev Mol Med* 10:e19. 10.1017/S1462399408000744

Alt JR, Gladden AB, and Diehl JA. 2002. p21(Cip1) Promotes cyclin D1 nuclear accumulation via direct inhibition of nuclear export. *J Biol Chem* 277:8517-8523. 10.1074/jbc.M108867200

Andres SA, Bickett KE, Alatoum MA, Kalbfleisch TS, Brock GN, and Wittliff JL. 2015. Interaction between smoking history and gene expression levels impacts survival of breast cancer patients. *Breast Cancer Res Treat* 152:545-556. 10.1007/s10549-015-3507-z

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, and Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-29. 10.1038/75556

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jane-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P, Jr., de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, and Garraway LA. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483:603-607. 10.1038/nature11003

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, and Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41:D991-995. 10.1093/nar/gks1193

Calderaro J, Ziol M, Paradis V, and Zucman-Rossi J. 2019. Molecular and histological correlations in liver cancer. *J Hepatol* 71:616-630. 10.1016/j.jhep.2019.06.001

Can T. 2014. Introduction to bioinformatics. *Methods Mol Biol* 1107:51-71. 10.1007/978-1-62703-748-8_4

Chen H, and Boutros PC. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. 10.1186/1471-2105-12-35

Cheng MG, Olivier P, Diehl JA, Fero M, Roussel MF, Roberts JM, and Sherr CJ. 1999. The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts. *Embo Journal* 18:1571-1583. DOI 10.1093/emboj/18.6.1571

Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, and Lin CY. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8 Suppl 4:S11. 10.1186/1752-0509-8-S4-S11

Chittori S, Hong J, Saunders H, Feng H, Ghirlando R, Kelly AE, Bai Y, and Subramaniam S. 2018. Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. *Science* 359:339-343. 10.1126/science.aar2781

462   Cuddihy AR, and Bristow RG. 2004. The p53 protein family and radiation sensitivity: Yes or no? *Cancer Metastasis*
463   *Rev* 23:237-257. 10.1023/B:CANC.0000031764.81141.e4

464   Edgar R, Domrachev M, and Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization
465   array data repository. *Nucleic Acids Res* 30:207-210. 10.1093/nar/30.1.207

466   Eldeiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, and
467   Vogelstein B. 1993. Waf1, a Potential Mediator of P53 Tumor Suppression. *Cell* 75:817-825. Doi
468   10.1016/0092-8674(93)90500-P

469   Fiorentino M, Altimari A, D'Errico A, Cukor B, Barozzi C, Loda M, and Grigioni WF. 2000. Acquired expression of
470   p27 is a favorable prognostic indicator in patients with hepatocellular carcinoma. *Clinical Cancer Research*
471   6:3966-3972.

472   Foltz DR, Jansen LE, Black BE, Bailey AO, Yates JR, 3rd, and Cleveland DW. 2006. The human CENP-A
473   centromeric nucleosome-associated complex. *Nat Cell Biol* 8:458-469. 10.1038/ncb1397

474   Forner A, Llovet JM, and Bruix J. 2012. Hepatocellular carcinoma. *Lancet* 379:1245-1255. 10.1016/S0140-
475   6736(11)61347-0

476   Forner A, Reig M, and Bruix J. 2018. Hepatocellular carcinoma. *Lancet* 391:1301-1314. 10.1016/S0140-
477   6736(18)30010-2

478   Galili T, O'Callaghan A, Sidi J, and Sievert C. 2018. heatmaply: an R package for creating interactive cluster heatmaps
479   for online publishing. *Bioinformatics* 34:1600-1602. 10.1093/bioinformatics/btx657

480   Gartel AL. 2009. p21(WAF1/CIP1) and cancer: a shifting paradigm? *Biofactors* 35:161-164. 10.1002/biof.26

481   Gary D Bader CWH. 2003. An automated method for finding molecular complexes in large protein interaction
482   networks. *BMC Bioinformatics* 4:27.

483   Gu Y, Turck CW, and Morgan DO. 1993. Inhibition of CDK2 activity in vivo by an associated 20K regulatory subunit.
484   *Nature* 366:707-710. 10.1038/366707a0

485   Harper JW, Adami GR, Wei N, Keyomarsi K, and Elledge SJ. 1993. The P21 Cdk-Interacting Protein Cip1 Is a Potent
486   Inhibitor of G1 Cyclin-Dependent Kinases. *Cell* 75:805-816.

487   Hohmann T, Kessler J, Grabiec U, Bache M, Vordermark D, and Dehghani F. 2018. Automatic detection of DNA
488   double strand breaks after irradiation using an gamma H2AX assay. *Histol Histopathol* 33:475-485.
489   10.14670/Hh-11-945

490   Holleczek B, and Brenner H. 2013. Model based period analysis of absolute and relative survival with R: data
491   preparation, model fitting and derivation of survival estimates. *Comput Methods Programs Biomed* 110:192-
492   202. 10.1016/j.cmpb.2012.10.004

493   Hugo H, Ackland ML, Blick T, Lawrence MG, Clements JA, Williams ED, and Thompson EW. 2007. Epithelial--
494   mesenchymal and mesenchymal--epithelial transitions in carcinoma progression. *J Cell Physiol* 213:374-
495   383. 10.1002/jcp.21223

496   Hunter T, and Pines J. 1994. Cyclins and cancer. II: Cyclin D and CDK inhibitors come of age. *Cell* 79:573-582.
497   10.1016/0092-8674(94)90543-6

498   Ito K, and Murphy D. 2013. Application of ggplot2 to Pharmacometric Graphics. *CPT Pharmacometrics Syst*
499   *Pharmacol* 2:e79. 10.1038/psp.2013.56

500   Kanehisa M, and Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28:4.

501   Kops GJ, Weaver BA, and Cleveland DW. 2005. On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat*
502   *Rev Cancer* 5:773-785. 10.1038/nrc1714

503 Levine AJ, and Oren M. 2009. The first 30 years of p53: growing ever more complex. *Nat Rev Cancer* 9:749-758.
504   10.1038/nrc2723
505 Li H, Guan H, Guo Y, Liang W, Liu L, He X, Ke W, Cao X, Xiao H, and Li Y. 2018. CITED1 promotes proliferation
506   of papillary thyroid cancer cells via the regulation of p21 and p27. *Cell Biosci* 8:57. 10.1186/s13578-018-
507   0256-9
508 Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, and Lopez R. 2015. The EMBL-
509   EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* 43:W580-584.
510   10.1093/nar/gkv279
511 Marquardt JU, and Thorgeirsson SS. 2014. SnapShot: Hepatocellular carcinoma. *Cancer Cell* 25:550 e551.
512   10.1016/j.ccr.2014.04.002
513 Matsuoka M, Kato JY, Fisher RP, Morgan DO, and Sherr CJ. 1994. Activation of cyclin-dependent kinase 4 (cdk4)
514   by mouse MO15-associated kinase. *Mol Cell Biol* 14:7265-7275. 10.1128/mcb.14.11.7265
515 Molenaar JJ, Ebus ME, Koster J, van Sluis P, van Noesel CJ, Versteeg R, and Caron HN. 2008. Cyclin D1 and CDK4
516   activity contribute to the undifferentiated phenotype in neuroblastoma. *Cancer Res* 68:2599-2609.
517   10.1158/0008-5472.CAN-07-5032
518 Nan KJ, Jing Z, and Gong L. 2004. Expression and altered subcellular localization of the cyclin-dependent kinase
519   inhibitor p27Kip1 in hepatocellular carcinoma. *World J Gastroenterol* 10:1425-1430.
520   10.3748/wjg.v10.i10.1425
521 Oka N, Kasamatsu A, Endo-Sakamoto Y, Eizuka K, Wagai S, Koide-Ishida N, Miyamoto I, Iyoda M, Tanzawa H,
522   and Uzawa K. 2019. Centromere Protein N Participates in Cellular Proliferation of Human Oral Cancer by
523   Cell-Cycle Enhancement. *J Cancer* 10:3728-3734. 10.7150/jca.32281
524 Polyak K, Lee MH, Erdjument-Bromage H, Koff A, Roberts JM, Tempst P, and Massague J. 1994. Cloning of
525   p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals.
526   *Cell* 78:59-66. 10.1016/0092-8674(94)90572-x
527 Qin LF, and Ng IO. 2001. Expression of p27(KIP1) and p21(WAF1/CIP1) in primary hepatocellular carcinoma:
528   clinicopathologic correlation and survival analysis. *Hum Pathol* 32:778-784. 10.1053/hupa.2001.27105
529 Rahman MR, Islam T, Gov E, Turanli B, Gulfidan G, Shahjaman M, Banu NA, Mollah MNH, Arga KY, and Moni
530   MA. 2019. Identification of Prognostic Biomarker Signatures and Candidate Drugs in Colorectal Cancer:
531   Insights from Systems Biology Analysis. *Medicina (Kaunas)* 55. 10.3390/medicina55010020
532 Razavipour SF, Harikumar KB, and Slingerland JM. 2020. p27 as a Transcriptional Regulator: New Roles in
533   Development and Cancer. *Cancer Res* 80:3451-3458. 10.1158/0008-5472.CAN-19-3663
534 Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, and Chinnaiyan AM.
535   2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6:1-6.
536   10.1016/s1476-5586(04)80047-2
537 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression
538   analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. 10.1093/nar/gkv007
539 Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, and Muller M. 2011. pROC: an open-source package
540   for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. 10.1186/1471-2105-12-77
541 Rubin SM, Sage J, and Skotheim JM. 2020. Integrating Old and New Paradigms of G1/S Control. *Mol Cell* 80:183-
542   192. 10.1016/j.molcel.2020.08.020
543 Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, Ogawa K, Ono H, Mitsunori Y, Ban D,

544   Kudo A, Arii S, Tanabe M, Wands JR, and Tanaka S. 2019. Comprehensive molecular and immunological
545   characterization of hepatocellular carcinoma. *EBioMedicine* 40:457-470. 10.1016/j.ebiom.2018.12.058
546 Siegel RL, Miller KD, and Jemal A. 2020. Cancer statistics, 2020. *CA Cancer J Clin* 70:7-30. 10.3322/caac.21590
547 Smoot ME, Ono K, Ruscheinski J, Wang PL, and Ideker T. 2011. Cytoscape 2.8: new features for data integration and
548   network visualization. *Bioinformatics* 27:431-432. 10.1093/bioinformatics/btq675
549 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Eberta BL, Gillettea MA, Paulovich A, Pomeroy SL, Golub
550   TR, Lander ES, and Mesirov JP. 2005. Gene set enrichment analysis: a knowledgebased approach for
551   interpreting genome-wide expression profiles. *PNAS* 102:6. 10.1073/pnas.0506580102
552 Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A,
553   Tsafou KP, Kuhn M, Bork P, Jensen LJ, and von Mering C. 2015. STRING v10: protein-protein interaction
554   networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447-452. 10.1093/nar/gku1003
555 Tang Z, Li C, Kang B, Gao G, Li C, and Zhang Z. 2017. GEPIA: a web server for cancer and normal gene expression
556   profiling and interactive analyses. *Nucleic Acids Res* 45:W98-W102. 10.1093/nar/gkx247
557 Tian T, Li X, Liu Y, Wang C, Liu X, Bi G, Zhang X, Yao X, Zhou ZH, and Zang J. 2018. Molecular basis for CENP-N
558   recognition of CENP-A nucleosome on the human kinetochore. *Cell Res* 28:374-378. 10.1038/cr.2018.13
559 Toyoshima H, and Hunter T. 1994. p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21.
560   *Cell* 78:67-74. 10.1016/0092-8674(94)90573-8
561 Tung EK, Mak CK, Fatima S, Lo RC, Zhao H, Zhang C, Dai H, Poon RT, Yuen MF, Lai CL, Li JJ, Luk JM, and Ng
562   IO. 2011. Clinicopathological and prognostic significance of serum and tissue Dickkopf-1 levels in human
563   hepatocellular carcinoma. *Liver Int* 31:1494-1504. 10.1111/j.1478-3231.2011.02597.x
564 Wang Q, Li A, Jin J, and Huang G. 2017. Targeted interfering DEP domain containing 1 protein induces apoptosis in
565   A549 lung adenocarcinoma cells through the NF-kappaB signaling pathway. *Onco Targets Ther* 10:4443-
566   4454. 10.2147/OTT.S142244
567 Woo HG, Choi JH, Yoon S, Jee BA, Cho EJ, Lee JH, Yu SJ, Yoon JH, Yi NJ, Lee KW, Suh KS, and Kim YJ. 2017.
568   Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. *Nat Commun*
569   8:839. 10.1038/s41467-017-00991-w
570 Wright DJ, Day FR, Kerrison ND, Zink F, Cardona A, Sulem P, Thompson DJ, Sigurjonsdottir S, Gudbjartsson DF,
571   Helgason A, Chapman JR, Jackson SP, Langenberg C, Wareham NJ, Scott RA, Thorsteindottir U, Ong KK,
572   Stefansson K, and Perry JRB. 2017. Genetic variants associated with mosaic Y chromosome loss highlight
573   cell cycle genes and overlap with cancer susceptibility. *Nat Genet* 49:674-679. 10.1038/ng.3821
574 Xing J, Bhuria V, Bui KC, Nguyen MLT, Hu Z, Hsieh CJ, Wittstein K, Stadler M, Wilkens L, Li J, Kalesse M, Bozko
575   P, and Plentz RR. 2020. Haprolid Inhibits Tumor Growth of Hepatocellular Carcinoma through Rb/E2F and
576   Akt/mTOR Inhibition. *Cancers (Basel)* 12. 10.3390/cancers12030615
577 Yu G, Wang LG, Han Y, and He QY. 2012. clusterProfiler: an R package for comparing biological themes among
578   gene clusters. *OMICS* 16:284-287. 10.1089/omi.2011.0118
579

Figure 1

Flowchart of the integrated analysis.

GO / KEGG analysis

Protein-protein interaction (PPI) network of DEGs based STRING

Module analysis by MCODE in cytoscape

hub genes selected by cytoscape

CENPN

CCLE
EMBL-EBI
Oncomine
GEPIA
ROC
GSEA

qRT-PCR
CCK-8
Cell cycle
Immunofluorescence
Western blotting

TCGA database
Bioinformatics analysis
COX regression analysis

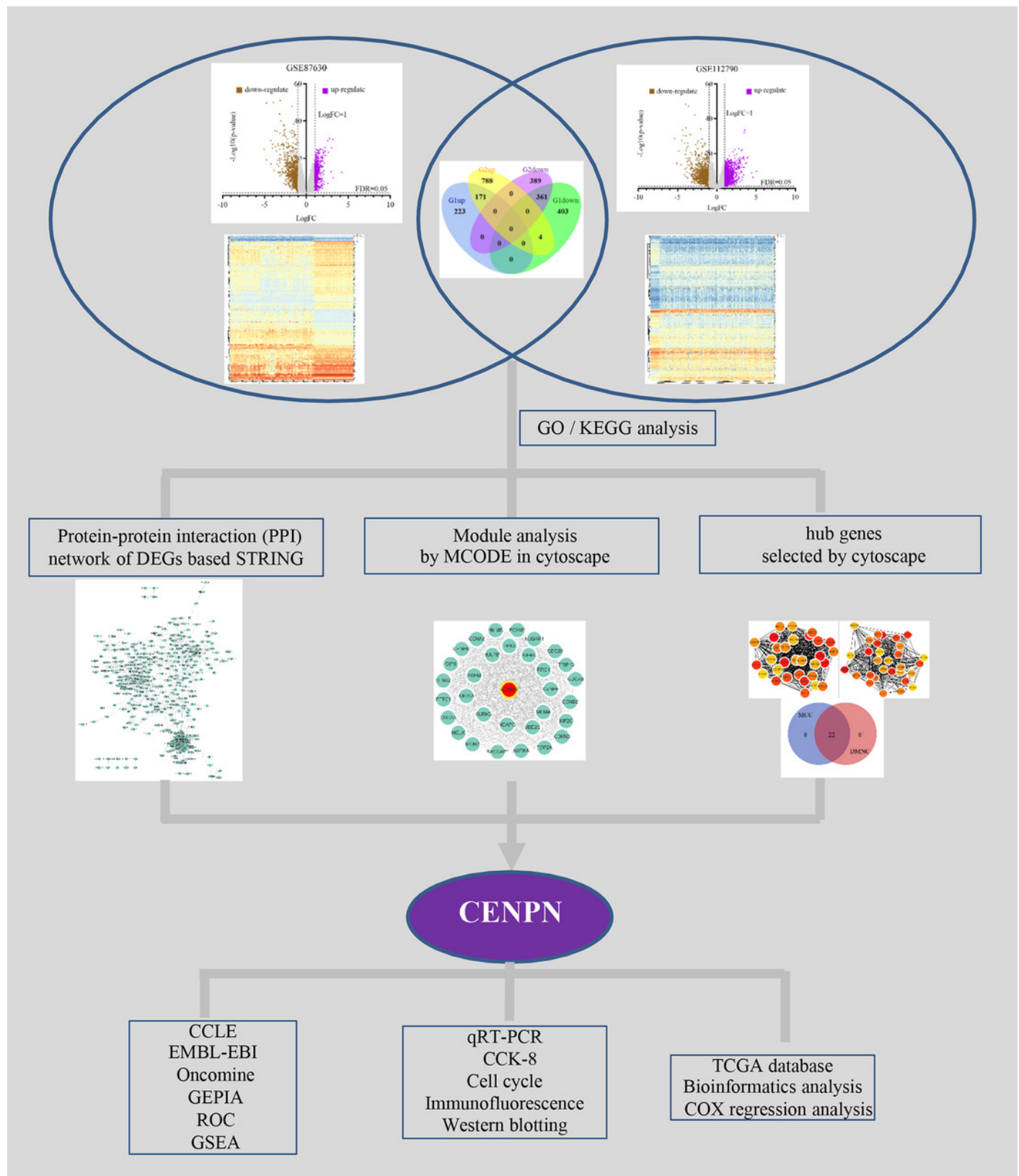



Figure 2

Differentially expressed genes (DEGs) were evaluated by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses.

(A, B) Volcano plot of the DEGs between HCC and normal liver tissues in each dataset. Purple dots: significantly upregulated genes in HCC; brown dots: significantly downregulated genes in HCC; gray dots: non-DEGs. Genes with an adjusted P value (FDR)< 0.05 and | logFC | > 1 were considered to have statistically significant differences in expression. (C) Venn diagram of 532 overlapping DEGs from G1 (GSE87630) and G2 (GSE112790): 171 upregulated genes and 361 downregulated genes. (D) GO analysis. (G) KEGG pathway analysis.

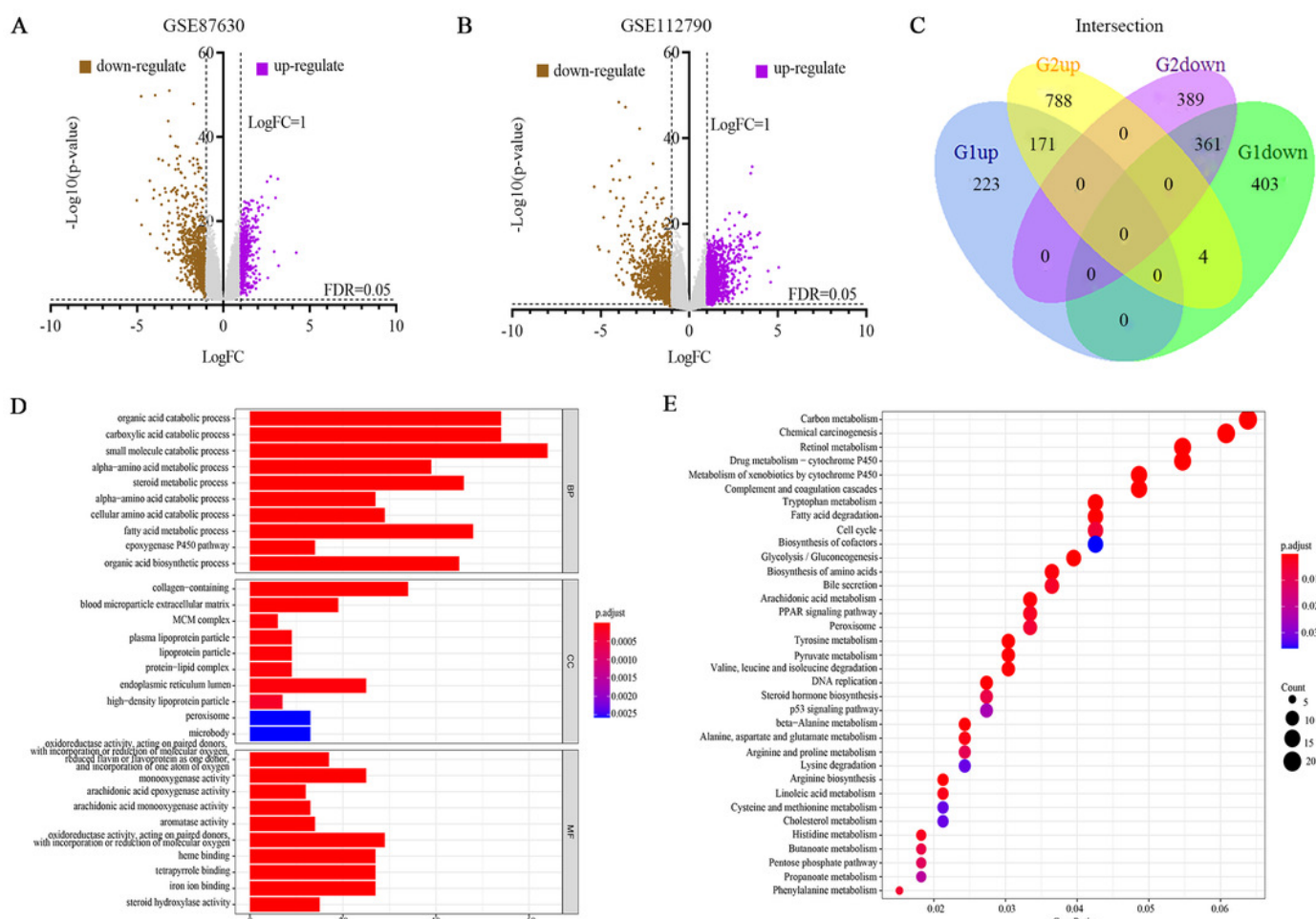



Figure 3

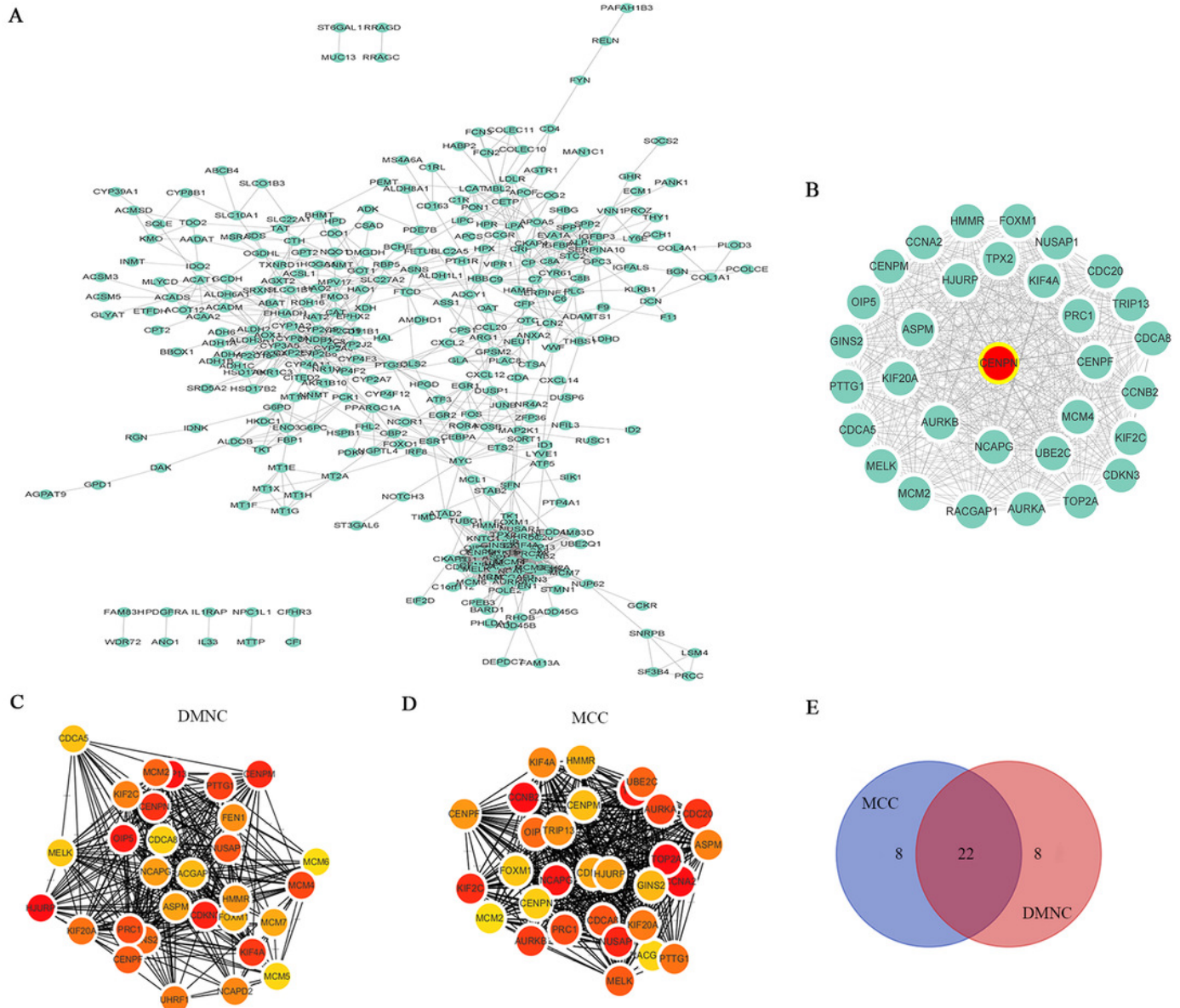
Screening the hub gene through STRING and Cytoscape.

(A) A total of 347 DEGs were visualized in a PPI network. Nodes represent proteins, and edges represent interactions among proteins. There were 347 nodes and 1560 edges in the network. (B) The most significant module in MCODE analysis. (C, D) The top 30 hub genes were searched through two ranking methods in cytoHubba. MCC, maximal clique centrality; DMNC, density of maximum neighborhood component. (E) Venn diagram of 22 overlapping hub genes from the MCC and DMNC analyses.

A

B

C
DMNC

D
MCC

E
Figure 4

CENPN expression in liver hepatocellular carcinoma and its prognostic information.

(A) CCLE revealed that CENPN was highly expressed in many cancer cells, including liver cancer cells. (B) The EMBL-EBI results, illustrated with GraphPad Prism 8, showed that CENPN was highly expressed in many HCC cell lines. (C, D and E) CENPN expression in HCC and liver tissues in the GSE87630, GSE112790 and GSE25097 datasets. (F) Analysis of CENPN in cancer vs. normal tissue from Oncomine. Heat maps of hub gene expression in clinical HCC tissues vs. normal liver tissues are shown. (G, H) The GEPIA database was used to examine the relationship between CENPN expression and overall survival and disease-free survival in LIHC patients. LIHC, liver hepatocellular carcinoma. HCC, hepatocellular carcinoma. (I, J and K) The ROC curves of GSE87630, GSE112790 and GSE25097. The AUCs were 0.913, 0.904 and 0.787, respectively. ROC, receiver operating characteristic. AUC, area under the curve.

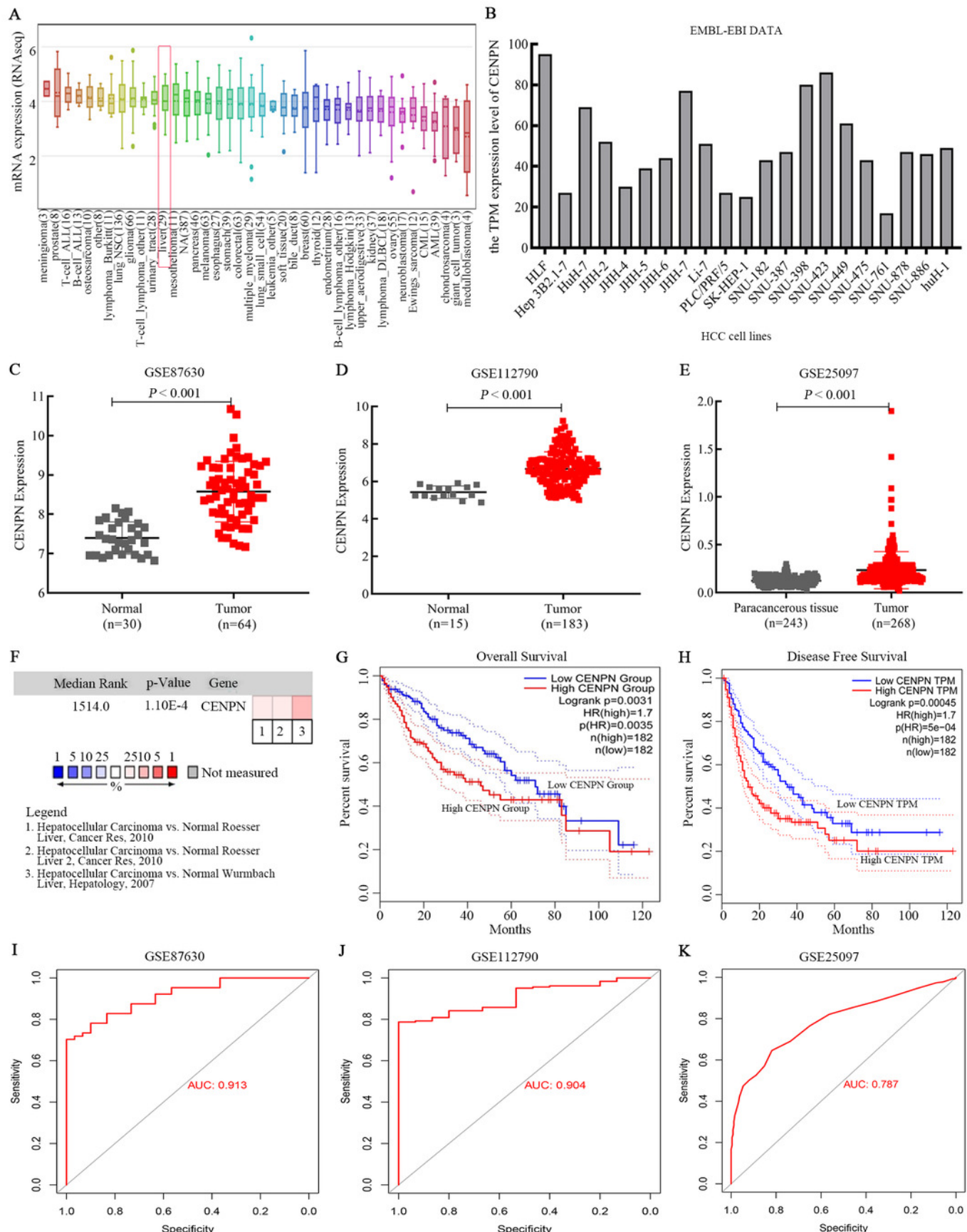



Figure 5

Verification of CENPN expression in HCC patients and its correlation with survival and clinicopathological stage using the TCGA database.

(A) The differential expression of CENPN in normal and tumor tissues. All normal and tumor samples were analyzed by the Wilcoxon rank-sum test ($p < 0.001$). (B) Paired differentiation analysis of CENPN expression in normal and tumor samples derived from the same patient ($p < 0.001$ by the Wilcoxon rank-sum test). (C) Survival analysis of LIHC patients with different CENPN expression levels. Patients were classified as having high or low expression according to the optimal cutoff, 3.006547, which was calculated with the survminer package. $P=0.004$ by the log-rank test. (D-F) Relationship between CENPN expression and clinicopathological stage. The Wilcoxon rank-sum test or the Kruskal-Wallis rank-sum test was used for statistical processing.

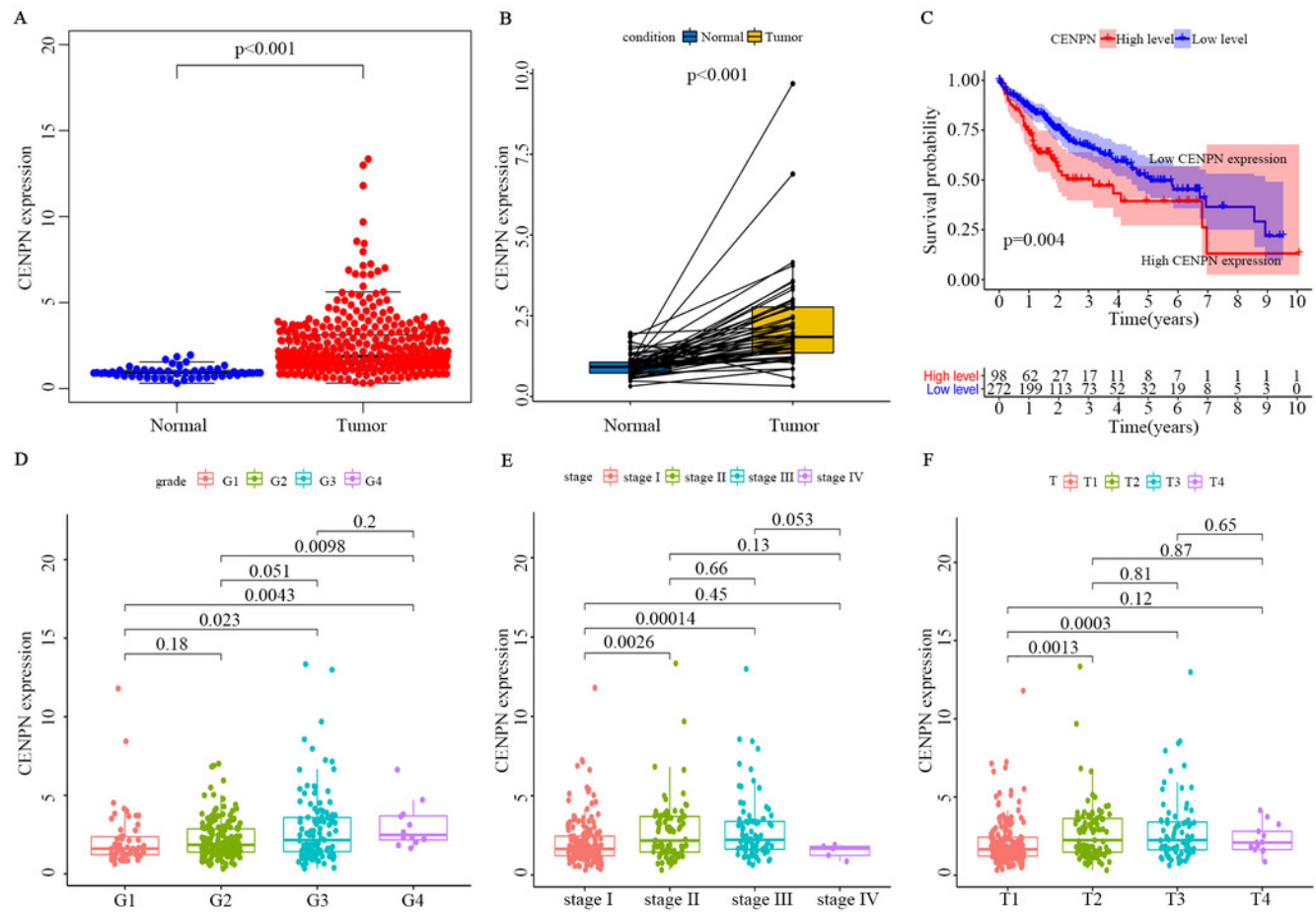



Figure 6

K-M analysis of clinicopathological parameters.

(A-C) The effect of stage, M stage and T stage on survival according to the K-M curve. The log-rank test was used for statistical processing. Abbreviations: K-M, Kaplan-Meier; M, migration; T, tumor size.

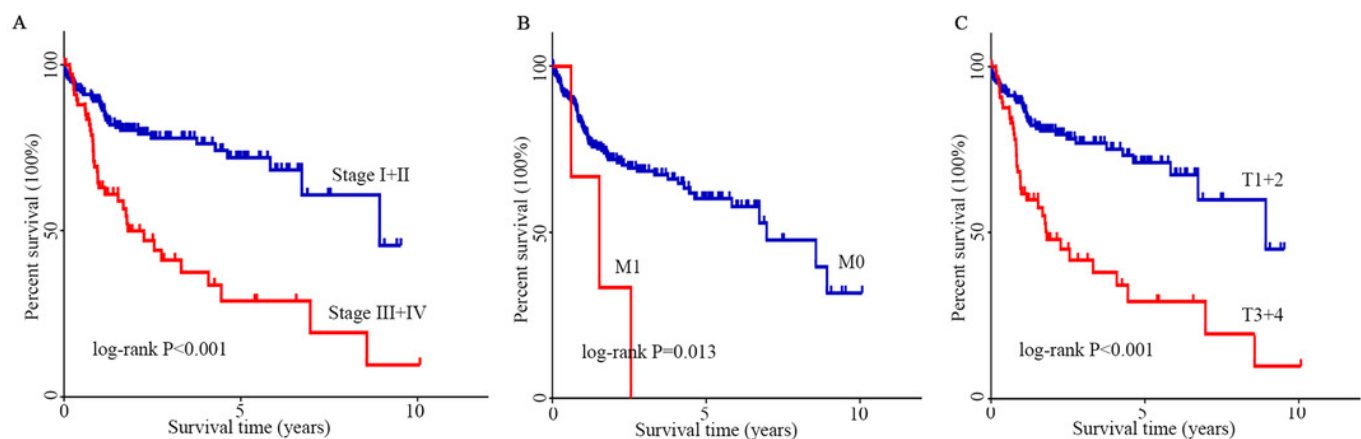



Figure 7

Interfering with CENPN expression inhibits the proliferation of HCC cells and the G1/S transition.

(A-H) GSEA of CENPN in HCC. (I, J) qRT-PCR and western blot assays of the interference efficiency of CENPN. (K, L) CCK-8 assay of the proliferation ability of cells with CENPN knockdown. (M, O) Immunofluorescence was used to detect the efficiency of siRNA targeting CENPN. (N, P) Immunofluorescence was used to detect the fluorescence intensity of Ki67 in cells with CENPN knockdown. **(**Q-X) Flow cytometry was used to detect changes in the cell cycle distribution after CENPN knockdown. Each data point represents the mean ± SD from three independent experiments. $^{*}p < 0.05$. $^{**}p < 0.001$. $^{***}p < 0.0001$. Scale bars, 50 μm.

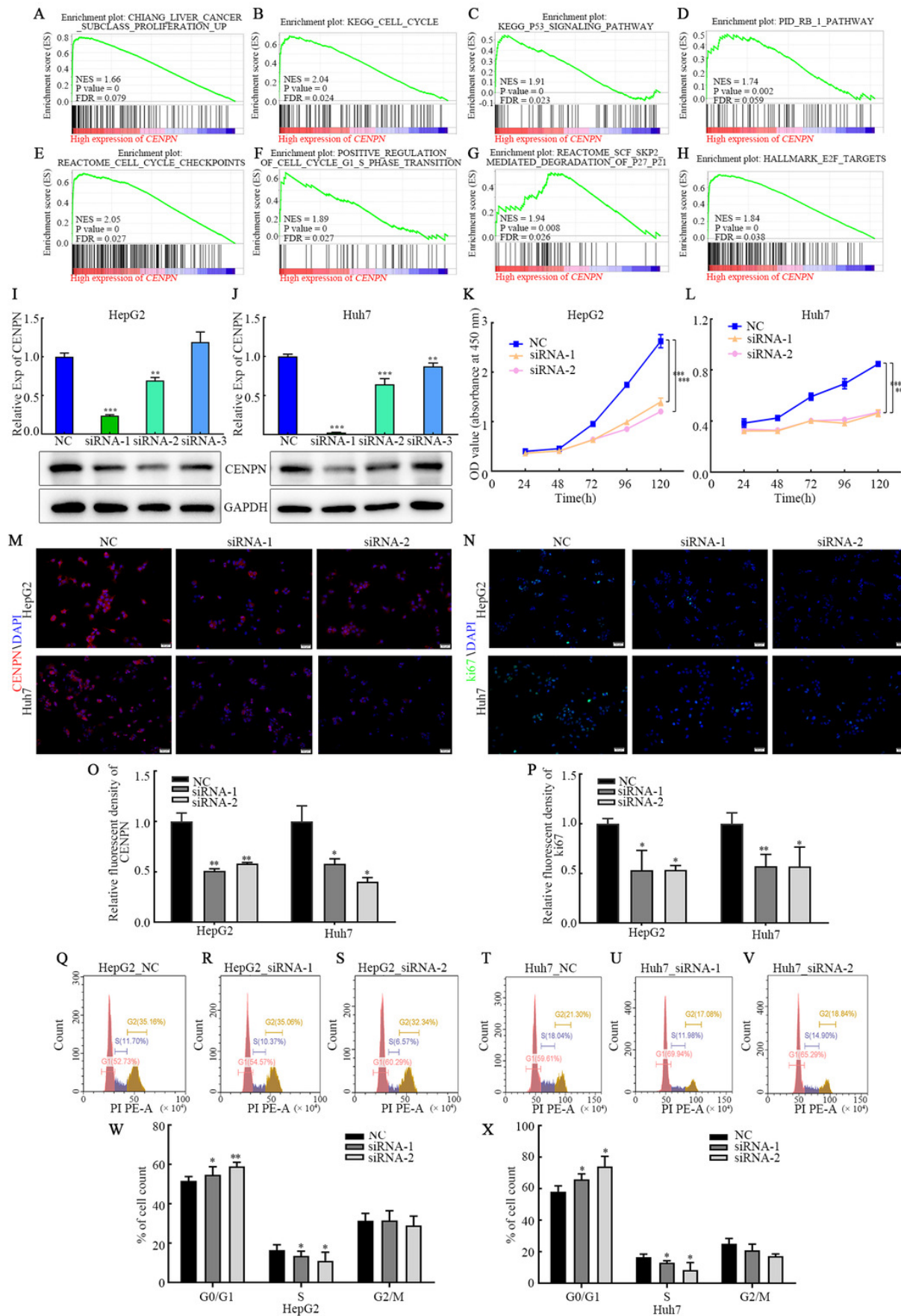



Figure 8

Exploring the mechanism of CENPN in HCC cells.

(A) GSEA of CENPN in HCC. (B, C) Immunofluorescence detection of γ-H2AX foci. (D, E) Colony formation assays were used to detect the number of clones in HCC cells after CENPN knockdown alone or combined with 2 Gy X-rays. (F) Western blot assays were performed to detect the expression of G1/S phase-related checkpoint proteins and corresponding upstream and downstream markers in cells with CENPN knockdown. (G) Mechanistic diagram of the biological function of CENPN in HCC cells. Each data point represents three independent experiments. Scale bars, 10 μm.

A

Enrichment Score

- GO_DNA_DAMAGE_RESPONSE_DETECTION_OF_DNA_DAMAGE
- GO_DNA_DAMAGE_RESPONSE_SIGNAL_TRANSDUCTION_BY_P53_CLASS_MEDIATOR
- GO_DNA_REPAIR
- GO_SIGNAL_TRANSDUCTION_IN_RESPNSE_TO_DNA_DAMAGE
- HALLMARK_DNA_REPAIR
- REACTOME_DNA_DAMAGE_TELOMERE_STRESS_INDUCED_SENESCENCE
- REACTOME_G1_S_DNA_DAMAGE_CHECKPOINTS
- WP_DNA_DAMAGE_RESPONSE
- WP_DNA_IRDAMAGE_AND_CELLULAR_RESONSE
- WP_MIRNA_REGULATION_OF_DNA_DAMAGE_RESPONSE

high expression <----------> low expression

B

| | NC | siRNA-1 | siRNA-2 |

HepG2 — 0Gy, 10Gy

C

Huh7 — 0Gy, 10Gy

D

| | NC | siRNA-1 | siRNA-2 |

HepG2 — 0Gy, 2Gy

E

| | NC | siRNA-1 | siRNA-2 |

Huh7 — 0Gy, 2Gy

F

HepG2 / Huh7

NC siRNA-1 siRNA-2 NC siRNA-1 siRNA-2

CENPN
P53
P27
P21
CDK4
cyclin D1
CDK2
cyclin E
pRb(ser780)
Rb
E2F1
c-myc
β-tubulin

G

Hub genes for DEGs ranked in the cytoHubba plugin of Cytoscape

Table 1. Hub genes for DEGs ranked in cytoHubba plugin of Cytoscape

Catelogy	Rank methods in cytoHubba	
	MCC	DMNC
Gene symbol top 30	*CCNA2*	**CDKN3**
	CDKN3	**KIF20A**
	KIF20A	**PRC1**
	PRC1	**RACGAP1**
	RACGAP1	**KIF2C**
	KIF2C	*CDCA5*
	UBE2C	**CDCA8**
	CDCA8	**ASPM**
	CDC20	**HJURP**
	TPX2	**OIP5**
	AURKA	*MCM7*
	ASPM	*MCM6*
	HJURP	**GINS2**
	TOP2A	**CENPN**
	OIP5	**CENPM**
	GINS2	*MCM4*
	CENPN	**HMMR**
	CENPM	*NCAPD2*
	HMMR	**NCAPG**
	NCAPG	**NUSAP1**
	AURKB	**MELK**
	NUSAP1	**TRIP13**
	MELK	**FOXM1**
	CCNB2	**MCM2**
	TRIP13	*MCM5*
	FOXM1	**KIF4A**
	MCM2	*FEN1*
	KIF4A	**PTTG1**
	PTTG1	**CENPF**
	CENPF	*UHRF1*

Bold gene symbols were the overlap hub genes in top 30 by two ranked methods respectively in cytoHubba.MCC Maximal clique centrality, DMNC Density of Maximum Neighborhood Component.

Cox regression analysis of overall survival of HCC

Table 2. Cox regression analysis of overall survival of HCC

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI of HR	*P*-value	HR	95% CI of HR	*P*-value
Age (≤65 vs >65)	0.989	0.606-1.614	0.964	0.875	0.521-1.470	0.614
Gender (female vs male)	1.285	0.804-2.054	0.294	1.141	0.691-1.883	0.606
Grade (grade 1+2 vs grade 3+4)	0.934	0.592-1.474	0.769	0.839	0.524-1.344	0.466
Stage (stageI+II vs stageIII+IV)	0.324	0.206-0.512	**0.000**	3.568	0.190-66.959	0.395
T (T1+T2 vs T3+T4)	0.322	0.204-0.509	**0.000**	0.098	0.005-1.772	0.116
M (M0 vs M1)	0.255	0.080-0.813	**0.021**	0.411	0.118-1.435	0.163
N (N1+N2 vs N3+N4)	0.484	0.118-1.982	0.313	0.213	0.028-1.634	0.137
CENPN expression (low vs high)	0.593	0.368-0.955	**0.032**	0.608	0.372-0.995	**0.048**