Identification of prognostic risk factors for pancreatic cancer using bioinformatics analysis

Dandan Jin [1,2], **Yujie Jiao** [1,2], **Jie Ji** [1,2], **Wei Jiang** [3], **Wenkai Ni** [1], **Yingcheng Wu** [2], **Runzhou Ni** [1], **Cuihua Lu** [1], **Lishuai Qu** [1], **Hongbing Ni** [1], **Jinxia Liu** [1], **Weisong Xu** [Corresp., 4], **MingBing Xiao** [Corresp. 1, 5]

[1] Department of Gastroenterology, Affiliated Hospital of Nantong University, Nantong, China

[2] Clinical Medicine, Medical College, Nantong University, Nantong, China

[3] Department of Emergency, Affiliated Hospital of Nantong University, Nantong, China

[4] Department of Gastroenterology, Second People's Hospital of Nantong, Nantong, China

[5] Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China

Corresponding Authors: Weisong Xu, MingBing Xiao
Email address: xws71@sina.com, xmb73@163.com

Background: Pancreatic cancer is one of the most common malignant cancers worldwide. Currently, the pathogenesis of pancreatic cancer remains unclear; thus, it is necessary to explore its precise molecular mechanisms. **Methods:** To identify candidate genes involved in the tumorigenesis and proliferation of pancreatic cancer, the microarray datasets GSE32676, GSE15471 and GSE71989 were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) between Pancreatic ductal adenocarcinoma (PDAC) and nonmalignant samples were screened by GEO2R. The Database for Annotation Visualization and Integrated Discovery (DAVID) online tool was used to obtain a synthetic set of functional annotation information for the DEGs. A PPI network of the DEGs was established using the Search Tool for the Retrieval of Interacting Genes (STRING) database, and a combination of more than 0.4 was considered statistically significant for the PPI. Subsequently, we visualized the PPI network using Cytoscape. Functional module analysis was then performed using Molecular Complex Detection (MCODE). Genes with a degree ≥10 were chosen as hub genes, and pathways of the hub genes were visualized using ClueGO and CluePedia. Additionally, GenCLiP 2.0 was used to explore interactions of hub genes. The Literature Mining Gene Networks module was applied to explore the cocitation of hub genes. The Cytoscape plugin iRegulon was employed to analyze transcription factors regulating the hub genes. Furthermore, the expression levels of the 13 hub genes in pancreatic cancer tissues and normal samples were validated using the Gene Expression Profiling Interactive Analysis (GEPIA) platform. Moreover, overall survival and disease-free survival analyses according to the expression of hub genes were performed using Kaplan-Meier curve analysis in the cBioPortal online platform. The relationship between expression level and tumor grade was analyzed using

the online database Oncomine. Lastly, the eight snap-frozen tumorous and adjacent noncancerous adjacent tissues of pancreatic cancer patients used to detect the CDK1 and CEP55 protein levels by western blot. **Conclusions:** Altogether, the DEGs and hub genes identified in this work can help uncover the molecular mechanisms underlying the tumorigenesis of pancreatic cancer and provide potential targets for the diagnosis and treatment of this disease.

1   **Identification of prognostic risk factors for pancreatic cancer using**

2   **bioinformatics analysis**

3   **Dandan Jin[1,3],Yujie Jiao[1,3], Jie Ji[1,3], Wei Jiang[4], Wenkai Ni[1], Yingcheng Wu[3], Runzhou Ni[1], Cuihua Lu[1],**

4   **Lishuai Qu[1], Hongbing Ni[1], Jinxia Liu[1], Weisong Xu[5,*], Mingbing Xiao[1,2,*]**

5   [1] Department of Gastroenterology, Affiliated Hospital of Nantong University, Nantong, China

6   [2] Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China

7   [3] Clinical Medicine, Medical College, Nantong University, Nantong, China

8   [4] Department of Emergency, Affiliated Hospital of Nantong University, Nantong, China

9   [5] Department of Gastroenterology, Second People's Hospital of Nantong, Nantong, China

10   □ The First Corresponding autho

11   Mingbing Xiao

12   xmb73@163.com (M.Xiao).

13   □ The second Corresponding author

14   Weisong Xu

15   ws71@sina.com (W.Xu).

17
18
19
20
21
22
23
24
25
26
27

Abstract

Background: Pancreatic cancer is one of the most common malignant cancers worldwide. Currently, the pathogenesis of pancreatic cancer remains unclear; thus, it is necessary to explore its precise molecular mechanisms.

Methods: To identify candidate genes involved in the tumorigenesis and proliferation of pancreatic cancer, the microarray datasets GSE32676, GSE15471 and GSE71989 were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) between Pancreatic ductal adenocarcinoma (PDAC) and nonmalignant samples were screened by GEO2R. The Database for Annotation Visualization and Integrated Discovery (DAVID) online tool was used to obtain a synthetic set of functional annotation information for the DEGs. A PPI network of the DEGs was established using the Search Tool for the Retrieval of Interacting Genes (STRING) database, and a combination of more than 0.4 was considered statistically significant for the PPI. Subsequently, we visualized the PPI network using Cytoscape. Functional module analysis was then performed using Molecular Complex Detection (MCODE). Genes with a degree $\geqslant$10 were chosen as hub genes, and pathways of the hub genes were visualized using ClueGO and CluePedia. Additionally, GenCLiP 2.0 was used to explore interactions of hub genes. The Literature Mining Gene Networks module was applied to explore the cocitation of hub genes. The Cytoscape plugin iRegulon was employed to analyze transcription factors regulating the hub genes. Furthermore, the expression levels of the 13 hub genes in pancreatic cancer tissues and normal samples were validated using the Gene Expression Profiling Interactive Analysis (GEPIA) platform. Moreover, overall survival and disease-free survival analyses according to the expression of hub genes were performed using Kaplan-Meier curve analysis in the cBioPortal online platform. The relationship between expression level and tumor grade was analyzed using the online database Oncomine. Lastly, the eight snap-frozen tumorous and adjacent noncancerous adjacent tissues of pancreatic cancer patients used to detect the CDK1 and CEP55 protein levels by western blot.

54 **Conclusions:** Altogether, the DEGs and hub genes identified in this work can help uncover the molecular

55 mechanisms underlying the tumorigenesis of pancreatic cancer and provide potential targets for the diagnosis

56 and treatment of this disease.

57

Introduction

59 Pancreatic cancer is one of the most common lethal tumors worldwide, and its overall 5-year survival rate

60 is less than 5% in the United States. PDAC accounts for 95% of all pancreatic cancers (Siegel et al. 2018).

61 Moreover, it has been reported that only 10-15% of pancreatic cancer patients are eligible for tumor resection, a

62 trend that is attributed to the lack of early diagnostic markers and advanced metastasis (Becker et al. 2016;

63 Caruso Bavisotto et al. 2017). Mounting evidence indicates that abnormal expression and gene variants are

64 related to the tumorigenesis and progression of pancreatic cancer. A study by Caldas C showed that K-ras

65 activation was involved in early events in N-nitroso-bis (2-oxopropyl) amine-induced pancreatic carcinogenesis

66 in hamsters (Caldas et al. 1994). In addition, it has been reported that inactivation of SMAD family member 4

67 (SMAD4) and cyclin dependent kinase inhibitor 2A (CDKN2A) is related to the development of pancreatic

68 cancer (Pihlak et al. 2018). Obviously, mutations in tumor suppressor genes are linked to the progression of

69 pancreatic cancer. Undoubtedly, an early diagnosis is beneficial for patients. Therefore, accurate knowledge of

70 the molecular mechanisms involved in the tumorigenesis and proliferation of pancreatic cancer is vital.

71 Microarray technology and bioinformatic analysis have been extensively applied for screening the

72 expression of genes, miRNAs, lncRNAs, and DNA methylation, helping to identifyDEGs and functional

73 pathways relevant to the tumorigenesis and progression of pancreatic cancer. In this work, we aimed to explore

74 the pathogenesis of pancreatic cancer by a computational bioinformatics analysis of gene expression. Three

75 mRNA microarray datasets from the GEO were extracted and analyzed to identify DEGs between PDAC tissues

76 and noncancerous tissues. Subsequently, the functions of these DEGs were evaluated using Gene Ontology (GO)

77 and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. A protein-protein

78 interaction (PPI) network was visualized using Cytoscape, and the Literature Mining Gene Networks module in

79 GenCLiP 2.0 showed that cyclin-dependent kinase 1 (CDK1) has strong interactions with other hub genes.

80 Enrichment analysis of GenCLiP 2.0 suggested that cell division may be involved in the pathogenesis of

81   pancreatic cancer. In the GEPIA database, the mRNA levels of all hub genes were higher in PAAD (pancreatic

82   adenocarcinoma) tumor tissues than in normal tissues. Additionally, survival analysis indicated that abnormal

83   spindle microtubule assembly (ASPM), CDK1, centromere protein F (CENPF), centrosomal protein 55

84   (CEP55), denticleless E3 ubiquitin protein ligase homolog (DTL), epithelial cell transforming 2 (ECT2), NIMA

85   related kinase 2 (NEK2) and protein regulator of cytokinesis 1 (PRC1) may be associated with the tumorigenesis

86   and development of pancreatic cancer. Overall, 210 DEGs and 13 hub genes that may be candidate biomarkers

87   for pancreatic cancer were identified.

Materials and Methods

Microarray datasets and data processing

90   Three human pancreatic cancer mRNA expression datasets [GSE32676 (Li et al. 2018a), GSE15471 (Lu &

91   Li 2018) and GSE71989 (Li et al. 2018a)] were downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo)

92   (Edgar et al. 2002), a public functional genomic database containing high-throughput gene expression data, chips

93   and microarrays. They all used the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0

94   Array; human PDAC tumors and nonmalignant pancreas samples snap-frozen at the time of surgery were chosen.

95   The GSE32676 dataset contains 25 PDAC tissue samples and 7 nonmalignant pancreas samples, GSE15471

96   contains 39 PDAC tissue samples and 39 nonmalignant pancreas samples, and GSE71989 contains 13 PDAC

97   samples and 8 noncancerous samples.

Identification of DEGs

99   DEGs between PDAC and nonmalignant samples were screened by GEO2R

100 (http://www.ncbi.nlm.nih.gov/geo/geo2r) (Shao et al. 2018), which is an online tool that can be used to compare

101 two or more datasets in a GEO series to identify DEGs according to experimental conditions. Adjusted P-values

102 (adj. P) and Benjamini and Hochberg false discovery rates were employed as criteria for statistically significant

103 genes and to limit false positives. Probe sets with no corresponding gene symbols or genes with multiple gene

104 probe sets were removed or averaged. Log FC (fold change) >1 or <-1 and adj. P <0.01 was considered

105 statistically significant. An online tool (http://bioinformatics. psb.ugent.be/webtools/Venn/) was applied to draw

106 Venn diagrams of the DEGs.

KEGG and GO enrichment analyses of DEGs

107

108   The Database for DAVID; (david.ncifcrf.gov/) online tool was used to obtain a synthetic set of functional

109   annotation information for the DEGs (Huang et al. 2007; Le et al. 2018; Le et al. 2017; Meel et al. 2019). **P<0.01**

110   **was considered statistically significant.**

PPI network construction and module analysis

111

112   A PPI network of the DEGs was established using the STRING (http://string-db.org, version 10.0) database

113   (Szklarczyk et al. 2015), and a combination of more than 0.4 was considered statistically significant for the PPI.

114   Subsequently, we visualized the PPI network using Cytoscape, which is an open-source bioinformatics software

115   platform (Shannon et al. 2003). Functional module analysis was then performed using MCODE, which is an app

116   for Cytoscape that is used to cluster a given network to a densely connected area based on topology. The standard

117   for selection was set as follows: MCODE scores >5, degree cut-off=2, node score cut-off=0.2, max depth=100

118   and k-score=2 (Li et al. 2017a).

Hub gene selection and analysis

119

120   Genes with a degree ≥10 were chosen as hub genes, and pathways of the hub genes were visualized using

121   ClueGO and CluePedia, which are two plugins of Cytoscape (Bindea et al. 2013). Additionally, GenCLiP 2.0

122   (http://ci.smu.edu.cn) (Wang et al. 2014), which facilitates functional annotation and molecular network

123   construction of genes depending on the literature, was used to explore interactions of hub genes. The Gene

124   Cluster with Literature Profiles modules were used to generate statistically overrepresented keywords to annotate

125   genes based on the occurrence of free terms in the literature for a given gene. $P \leq 1 \times 10^6$ and hits ≥6 were

126   considered statistically significant (Li et al. 2017b). We selected keyword annotation to obtain a cluster analysis

127   heatmap of 13 hub genes. The Literature Mining Gene Networks module was applied to explore the cocitation

128   of hub genes. Furthermore, the expression levels of the 13 hub genes in pancreatic cancer tissues and normal

129   samples were validated using the GEPIA platform, which is a free online database (http://gepia.cancer-pku.cn/)

130   (Hauptman et al. 2019). Moreover, overall survival and disease-free survival analyses according to the

131   expression of hub genes were conducted using Kaplan-Meier curve analysis in the cBioPortal

132   (http://www.cbioportal.org) online platform (Gao et al. 2013). The relationship between expression level and

133   tumor grade was analyzed using the online database Oncomine (http://www.oncomine.com) (Rhodes et al.

134   2004).

135   ## Transcription factor analysis

136   The Cytoscape plugin iRegulon (Janky et al. 2014) was employed to analyze transcription factors regulating

137   the hub genes. The iRegulon plugin can identify regulons using motifs and track discovery in an existing network

138   or in a set of coregulated genes. Transcription factor information is obtained from databases such as Transfac,

139   Jaspar, Encode, Swissregulon and Homer, which use genome-wide ranking and recovery to detect enriched

140   transcription factor motifs and optimal sets of their direct targets. The cutoff criteria were as follows: enrichment

141   score threshold=5.0, ROC threshold for AUC calculation=0.03, rank threshold=5000, minimum identity between

142   orthologous genes=0.05, FDR=0.001 and the normalized enrichment score (NES)>10 (Li et al. 2017b).

143   ## Patients and tissue specimens

144   The eight snap-frozen tumorous and adjacent noncancerous adjacent tissues of pancreatic cancer patients

145   used to detect the CDK1 and CEP55 protein levels in this study were provided by the Affiliated Hospital of

146   Nantong University. Our experimental protocols have been subjected to approval by the Institutional Review

147   Board of Affiliated Hospital of Nantong University; all participating patients fully understood the protocols and

148   subscribed informed consent. The cohort of patients included 6 female patients and 2 male patients who

149   underwent surgical resection without chemotherapy or radiotherapy.

150   ## Western blotting

151   Western blotting was performed as previously described (Jiao et al. 2019). After blotting, the membranes

152   were incubated at 4°C overnight with anti-CDK1 (ab18, diluted 1:1000, Abcam, MA, USA), anti-CEP55

153   (ab170414, diluted 1:1000, Abcam, MA), anti-PCNA (ab29, diluted 1:1000, Abcam, MA) and anti-GAPDH

154   (#5174; 1:5000; Cell Signaling Technology, MA, USA) antibodies. This was followed by incubation with anti-

155   mouse IgG (#D110087; 1:2500; Sangon Biotech, Shanghai, China) or anti-rabbit IgG (#D110058; 1:2500;

156   Sangon Biotech, Shanghai, China) secondary antibodies. Finally, an enhanced chemiluminescence (ECL) kit

157   was used to visualize the bands, and the Molecular Imager ChemiDoc XRS System (Bio-Rad Laboratories, CA,

158   USA) was used to analyze and quantify the bands.

Statistical analysis

160   All data were analyzed using the GraphPad Prism 5.0 software. The results were presented as the mean ±

161   standard error of mean of at least three independent experiments. The comparison between multiple groups used

162   single-factor analysis of variance, and the comparison of data between separate groups was performed using t

163   test. $P < 0.05$ indicates statistical significance.

Results

Identification of DEGs in PDAC. In total, 775, 1793 and 3952 DEGs were identified when

166   comparing PDAC tissue samples and normal tissue samples in the GSE32676, GSE15471 and GSE71989

167   datasets, respectively. Additionally, 210 (186 upregulated and 24 downregulated) genes were common to all

168   three datasets (Fig. 1A).

PPI network construction. The PPI network of DEGs was constructed using the STRING database

170   and included 133 genes (125 upregulated and 8 downregulated) with combined scores > 0.4 (Fig. 1B).As the

171   number of downregulated genes was too small for GO and KEGG enrichment analyses, we only performed this

172   analysis for the upregulated genes.

GO and KEGG pathway analyses of the PPI network. GO and KEGG pathway analyses

174   were conducted to explore the potential functions and pathways of the upregulated DEGs using DAVID.

175   According to GO analysis, the upregulated DEGs were enriched in cell migration, cell-cell adhesion and cell

176   adhesion biological process (BP) categories (Table 1). The upregulated DEGs were primarily enriched in the

177   extracellular exosome and cytoplasm cell component (CC) categories (Table 1); DEGs were mainly enriched in

178   the cadherin binding involved in cell-cell adhesion and protein homodimerization activity molecular function

179   (MF) categories (Table 1). Additionally, KEGG pathway analysis indicated that the upregulated DEGs were

180   primarily enriched in the ECM-receptor interaction and pathways in cancer (Table 2).

Hub gene selection and analysis. Thirteen genes were considered to be hub genes with a degree

182   ≥10 (Fig. 1C). Detailed information about these hub genes is presented in Table 3. The pathways of the hub

183   genes were visualized using ClueGO and CluePedia, which are two plugins of Cytoscape (Fig. 2). Subsequently,

184   literature mining was performed in GenCLiP 2.0 to explore hub gene interactions. The cocitation network of

185   these 13 hub genes in the published literature is displayed in Fig. 3A. The results showed that 5 genes, DNA

186   topoisomerase II alpha (TOP2A), PRC1, ECT2, RRM2 and CEP55 interact with CDK1. TOP2A was the top

187   gene because it has been mentioned in 23 published literature sources as interacting with CDK1. The detailed

188   results of previous studies on these genes are shown in Table 4. Fig. 3B illustrates the significant results from

189   the enrichment analysis of the 13 hub genes, with cell division being the most related biological function at

190   76.9% enrichment.

191   **Transcription factor analysis of hub genes.** Transcription factor analysis of the 13 hub genes

192   was conducted using iRegulon, a Cytoscape plugin, and a normalized enrichment score (NES)>10 was

193   considered to be significant. The transcriptional regulation network of these hub genes is shown in Fig. 3C. The

194   transcription factors with NES>10 were NFYC (NES=15.131, targets=11), TFDP3 (NES=11.922, targets=11),

195   NFYA (NES=11.193, targets=9), and E2F4 (NES=10.426, targets=13).

196   **mRNA expression levels of the 13 hub genes in PAAD.** To confirm the expression levels

197   of the 13 identified hub genes, related published data were obtained from TCGA datasets and analyzed using the

198   GEPIA platform. As expected, the results showed that all hub genes were more highly expressed in tumor tissues

199   than in normal tissue samples (Fig. 4).

200   **Survival analyses of hub genes.** Kaplan-Meier curve analysis was used to analyze correlations

201   between overall survival and the hub genes. PDAC patients with alterations in anillin actin-binding protein

202   (ANLN), ASPM, CDK1, CENPF, CEP55, DTL, ETC2, NEK2, TOP2A and PRC1 exhibited poor overall

203   survival (Fig. 5). PDAC patients with ASPM, CDK1, CENPF, CEP55, DTL, ETC2, NEK2, PRC1,

204   ribonucleotide reductase regulatory subunit M2 (RRM2), TOP2A, ZW10 interacting kinetochore protein

205   (ZWINT), and ANLN alterations exhibited poor disease-free survival (Fig. 6). Among these genes, we selected

206   CDK1 and CEP55 for further study with regard to overall survival and disease-free survival times.

207   **Oncomine analysis of CDK1 and CEP55 in cancer vs. normal tissue.** Oncomine

208   analysis of cancer vs. normal tissue indicated that CDK1 and CEP55 were significantly overexpressed in

209   pancreatic cancer in different datasets (Fig. 7A and B). In the Grutzmann Pancreas dataset, CDK1 and CEP55

210   mRNA expression was higher in pancreatic cancer tissues than in normal pancreatic tissues (Fig. 7C and D).

211   Additionally, higher mRNA levels of CDK1 and CEP55 were associated with tumor grade (Fig. 7E and F).

212   **The expression of CDK1 and CEP55 in clinical specimens.** To validate the above

213   results, we detected CDK1 and CEP55 protein levels in clinical specimens. In the clinical specimens, CDK1

214   (Fig. 8A and C) and CEP55 (Fig. 8B and D) protein levels were significantly elevated in pancreatic cancer tissue

215   samples compared with adjacent nontumor tissues.

216   **Discussion**

217   In recent years, microarray technology has been extensively applied to reveal genetic alterations in tumors.

218   Therefore, microarray analysis is a tool for revealing biomarkers for the diagnosis, treatment, and prognosis of

219   pancreatic cancer. In this study, our results demonstrated that the upregulated DEGs were obviously enriched in

220   the BP categories cell migration and cell-cell adhesion. Many studies have shown that migration and invasion

221   are basic characteristics of pancreatic cancer (Zhuo et al. 2018). Moreover, cell-cell adhesion might be involved

222   in cell stretching and movement, which are the molecular bases of the important physiological and pathological

223   processes of tumor invasion and metastasis (Nobes & Hall 1999; Serrill et al. 2018). For the CC category, the

224   upregulated DEGs were primarily enriched in extracellular exosomes, which are nanoscale membrane vesicles

225   with diameters ranging from 40 to 100 nm. A growing number of studies have shown that tumor-derived

226   exosomes are associated with tumor development, metastasis, and drug resistance mechanisms (Jiao et al. 2018).

227   Furthermore, in the MF category, upregulated DEGs were mainly enriched in cadherin binding involved in cell-

228   cell adhesion and protein homodimerization activity, which are associated with invasion and metastasis.

229   According to our KEGG pathway enrichment results, the upregulated DEGs were primarily enriched in the

230   ECM-receptor interaction. During the development of tumors, the ECM experiences a remodeling process that

231   is similar to the process of embryonic development. The most important feature of this remodeling is the change

232   in the molecular composition of the ECM, whereby the reconstituted ECM creates a loose microenvironment for

233   the proliferation and differentiation of tumor cells, leading to high rates of proliferation, poor differentiation,

234   invasion and metastasis (Jin & Liu 2018).

235   In our study, we showed that TOP2A, PRC1, ECT2, RRM2 and CEP55 can interact with CDK1 and TOP2A

236   is the gene most closely related to CDK1. Previous studies have provided evidence about the expression of

237   CDK1 and TOP2A in pancreatic cancer, but the correlation between them in pancreatic cancer has not been fully

238   elucidated (Kalimutho et al. 2018; Li et al. 2017a; Shi et al. 2015; Xu et al. 2016). Cyclin-dependent kinases

239   (CDKs) are important driving factors of the human cell cycle (Zhao et al. 2013). Topoisomerases, including

240   topoisomerase 1 (TOP1) and topoisomerase 2 (TOP2), are key ribozymes that mainly participate in cell growth

241   by breaking and reconnecting DNA strands to change DNA topology (Liu et al. 2018). We predict that the cyclin

242   A2-CDK1-TOP2A axis plays an important role in tumor development. Furthermore, CDK1 might phosphorylate

243   TOP2A to promote S phase transition and influence the progression of pancreatic cancer, but follow-up research

244   is needed to determine the specific mechanism of the interaction (Kalimutho et al. 2018). Our enrichment

245   analysis of 13 hub genes further verified the relationship between metastasis and the malignant progression of

246   pancreatic cancer.

247   The transcription factors NFYC, TFDP3, POLE3 and E2F4 are closely linked to hub genes in pancreatic

248   cancer. NFYC is a histone-fold domain-containing transcription factor engaged in chromatin remodeling,

249   establishing permissive chromatin modifications at CCAAT motifs in promoters (Bieniossek et al. 2013).

250   Deletion of NFYC halts cell cycle progression, predominantly by causing G2/M arrest, and concurrent gain of

251   NFYC may serve to model an aberrant epigenome that promotes a proliferative and relatively undifferentiated

252   state (Benatti et al. 2011). Most E2Fs are localized in the nucleus, but E2F4 shows cell cycle-specific localization

253   and can be found in the nucleus of cycling cells in G0, early G1 and G2 phases (Lindeman et al. 1997). When

254   present in the nucleus, E2F4 has functions necessary for the induction of mitosis; interestingly, E2F4 also appears

255   to have a role in the cytoplasm during multiciliogenesis (van Amerongen et al. 2010). TFDP3 is expressed in

256   most cancer tissues and potentially plays a role in cell differentiation and proliferation. TFDP3 is a novel negative

257   regulator of E2F that can enhance both the DNA binding and transcriptional activity of E2F through the

258   formation of heterodimers; it also potentially plays an important role in the process of tumor development

259   independent of pRb (Ma et al. 2014). POLE3 is similar to the first subclass of core histones with respect to

260   regulation, and POLE3 expression is upregulated at the onset of S phase. E2F4 is often associated with promoters

261   in G0, is minimally bound to the POLE3 gene in starved cells and maximally in cells expressing POLE3 at high

262   levels. These findings indicate that these transcription factors may regulate the progression of pancreatic cancer,

263   which provides direction for our future research (Bolognese et al. 2006).

264   Expression analysis of PAAD tissues and normal tissues based on the TCGA database also indicated higher

265   levels of TOP2A (Jiao et al. 2019), CDK1 (Jing et al. 2019), RRM2 (Zhao et al. 2019), PRC1 (Mao et al. 2019),

266   NEK2 (Deng et al. 2019), ZWINT (Obuse et al. 2004), DTL(Cui et al. 2019), MELK (Meel et al. 2019), CENPF

267   (Chen et al. 2019), CEP55 (Hauptman et al. 2019), ANLN (Wang et al. 2019), ASPM (Hsu et al. 2019), and

268   ECT2 (Daulat et al. 2019) in tumor tissues. Previous studies have provided abundant evidence about the function

269   of the thirteen identified hub genes in cancer development. Mutations in ANLN, ASPM, CDK1, CENPF, CEP55,

270   DTL, ETC2, NEK2, TOP2A and PRC1 affect overall survival and disease-free survival in pancreatic cancer,

271   whereas mutations in RRM2 and ZWINT affect disease-free survival. Hence, we reasoned that ANLN, ASPM,

272   CDK1, CENPF, CEP55, DTL, ETC2, NEK2, TOP2A and PRC1 might be preferable prognostic factors that are

273   positively related to pancreatic cancer. Additionally, Wenzong Lu et al analyzed the microarray datasets

274   GSE32676, GSE15471, GSE71989 and GSE19650 to identify five upregulated hub genes, including MELK,

275   MET, THBS1, TOP2A and SDC1 (Lu et al. 2019); Jun Li et al analyzed the microarray datasets GSE71989,

276   GSE15471, GSE16515, GSE32676, GSE41368 and GSE28735 and found that five genes (BIRC5, CKS2,

277   ITGA3, ITGA6 and RALA) were significantly associated with survival time in patients with pancreatic duct

278   adenocarcinoma (Li et al. 2018a); Tao Zhu et al studied the five GEO datasets (GSE15471, GSE16515,

279   GSE18670, GSE32676, GSE71989) and reported that GJB2 and ERO1LB dysregulation was associated with

280   tumorigenesis in pancreatic adenocarcinoma (Zhu et al. 2017). Within addition to these hub genes, we supplied

281   some new hub genes that may be associated with the tumorigenesis and development of pancreatic cancer.

282   Our results showed that among these hub genes, TOP2A is the most closely related gene to CDK1, TOP2A

283   and CDK1, and it has been previously reported as a biomarker for pancreatic cancer (Kokkinakis et al. 2005);

284   however, to date, little attention has been paid to CEP55 and its possible relationship with CDK1. Indeed, only

285   one study reported that CEP55 can inhibit CDK1 phosphorylation and proteolysis mediated by the anaphase-

286   promoting complex to induce anaphase I in oocytes (Zhou et al. 2019). Nonetheless, the relationship between

287   CEP55 and CDK1 and the mechanism involved have not been studied in pancreatic cancer cells. It is well known

288   that pancreatic cancer has a high degree of malignancy and a short disease course, which is largely attributed to

289   the fact that it readily metastasizes to and invades adjacent organs. CEP55 belongs to the centrosomal family of

290   proteins, which plays an important role in critical cell functions. Mounting evidence shows that CEP55 is highly

291   expressed in multiple cancers, such as colon cancer, hepatocellular carcinoma and bladder cancer (Gao & Wang

292   2015; Li et al. 2018b; Singh et al. 2015). Moreover, overexpression of CEP55 accelerates the cell cycle transition

293   in gastric cancer, and low expression of CEP55 inhibits cell growth in breast cancer and gastric cancer

294   (Kalimutho et al. 2018; Tao et al. 2014). These results demonstrate that CEP55 may serve an oncogenic role and

295   a potential target for tumor treatment. CEP55 can also promote aggressive behavior in pancreatic cancer cells

296   by activating the NF-κB pathway (Peng et al. 2017a). Hence, the underlying mechanism of the proinvasive and

297   prometastatic effects of CEP55 in pancreatic cancer needs to be further studied. Last, we preliminarily validated

298   the ChIP analyses by detecting CDK1 and CEP55 protein levels in eight snap-frozen tumorous and adjacent

299   noncancerous adjacent tissues of pancreatic cancer patients. The results showed that CDK1 and CEP55 were

300   significantly overexpressed in pancreatic tumorous tissues compared to normal tissues, which is consistent with

301   the results in different datasets. However, our sample size was relatively small, so we could not verify the link

302   between clinical samples and prognosis. Fortunately, Piao Junjie et al validated that high expression of CDK1

303   was correlated with the short survival of pancreatic cancer patients by analyzing 99 cases of surgically resected

304   pancreatic cancer samples and 71 cases of normal pancreases (Piao et al. 2019). Tao Peng et al reported that

305   CEP55 expression was an independent prognostic factor of patient outcome and that CEP55 protein expression

306   levels in pancreatic cancer specimens were inversely correlated with survival time by analyzing 126 archived

307   paraffin-embedded pancreatic cancer specimens with immunohistochemical staining using an antibody against

308   human CEP55 (Peng et al. 2017b). The results were in accordance with the Kaplan-Meier curve analysis online.

309   **Conclusion**

310   In summary, by analyzing multiple datasets from the GEO database and validating the results with the

311   TCGA and Oncomine databases, our present work identifies dominant genes, their interaction network and

312   possible transcription factors involved during the progression and metastasis of pancreatic cancer. Some

313   relationships between hub genes and transcription factors have never been reported to influence the progression

314   of pancreatic cancer and may serve as potential targets for pancreatic cancer therapy. However, due to the low

315   quantity of gene probes in our selected datasets, the number of discovered DEGs was strikingly limited. More

316   genes and noncoding RNAs should be detected to enrich the network for a more comprehensive and integrated

317   understanding of pancreatic cancer development. In addition, we did not analyze the expression of hub genes,

318   such as CDK1, in pancreatic patients with or without lymphatic metastasis or use ROC analysis to explain the

319   prognosis and diagnostic value of the genes. We will collect these data in our future research to complete our

320   analysis of the prognosis and diagnostic value of these genes.

Data Availability

322   The datasets generated and analyzed during the current study are available in the GEO repository.
323   GSE32676: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32676
324   GSE15471: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15471
325   GSE71989: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71989

References

327   Becker A, Thakur BK, Weiss JM, Kim HS, Peinado H, and Lyden D. 2016. Extracellular Vesicles in Cancer:
328   Cell-to-Cell Mediators of Metastasis. Cancer Cell 30:836-848. 10.1016/j.ccell.2016.10.009
329   Benatti P, Dolfini D, Vigano A, Ravo M, Weisz A, and Imbriano C. 2011. Specific inhibition of NF-Y subunits
330   triggers different cell proliferation defects. Nucleic Acids Res 39:5356-5368. 10.1093/nar/gkr128
331   Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, Scheer E, Papadopoulos P, Tora L, Schultz P, and
332   Berger I. 2013. The architecture of human general transcription factor TFIID core complex. Nature
333   493:699-702. 10.1038/nature11791
334   Bindea G, Galon J, and Mlecnik B. 2013. CluePedia Cytoscape plugin: pathway insights using integrated
335   experimental and in silico data. Bioinformatics 29:661-663. 10.1093/bioinformatics/btt019
336   Bolognese F, Forni C, Caretti G, Frontini M, Minuzzo M, and Mantovani R. 2006. The Pole3 bidirectional unit
337   is regulated by MYC and E2Fs. Gene 366:109-116. 10.1016/j.gene.2005.07.046
338   Caldas C, Hahn SA, Hruban RH, Redston MS, Yeo CJ, and Kern SE. 1994. Detection of K-ras mutations in the
339   stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. Cancer Res
340   54:3568-3573.
341   Caruso Bavisotto C, Cappello F, Macario AJL, Conway de Macario E, Logozzi M, Fais S, and Campanella C.
342   2017. Exosomal HSP60: a potentially useful biomarker for diagnosis, assessing prognosis, and
343   monitoring   response   to   treatment.   Expert   Rev   Mol   Diagn   17:815-822.
344   10.1080/14737159.2017.1356230
345   Chen EB, Qin X, Peng K, Li Q, Tang C, Wei YC, Yu S, Gan L, and Liu TS. 2019. HnRNPR-CCNB1/CENPF
346   axis contributes to gastric cancer proliferation and metastasis. Aging (Albany NY) 11.
347   10.18632/aging.102254
348   Cui H, Wang Q, Lei Z, Feng M, Zhao Z, Wang Y, and Wei G. 2019. DTL promotes cancer progression by
349   PDCD4 ubiquitin-dependent degradation. J Exp Clin Cancer Res 38:350. 10.1186/s13046-019-1358-x
350   Daulat AM, Finetti P, Revinski D, Silveira Wagner M, Camoin L, Audebert S, Birnbaum D, Kodjabachian L,

351   Borg JP, and Bertucci F. 2019. ECT2 associated to PRICKLE1 are poor-prognosis markers in triple-
352   negative breast cancer. Br J Cancer 120:931-940. 10.1038/s41416-019-0448-z

353   Deng L, Sun J, Chen X, Liu L, and Wu D. 2019. Nek2 augments sorafenib resistance by regulating the
354   ubiquitination and localization of beta-catenin in hepatocellular carcinoma. J Exp Clin Cancer Res
355   38:316. 10.1186/s13046-019-1311-z

356   Edgar R, Domrachev M, and Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and
357   hybridization array data repository. Nucleic Acids Res 30:207-210.

358   Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E,
359   Cerami E, Sander C, and Schultz N. 2013. Integrative analysis of complex cancer genomics and clinical
360   profiles using the cBioPortal. Sci Signal 6:pl1. 10.1126/scisignal.2004088

361   Gao XY, and Wang XL. 2015. An adoptive T cell immunotherapy targeting cancer stem cells in a colon cancer
362   model. J buon 20:1456-1463.

363   Hauptman N, Jevsinek Skok D, Spasovska E, Bostjancic E, and Glavac D. 2019. Genes CEP55, FOXD3,
364   FOXF2, GNAO1, GRIA4, and KCNA5 as potential diagnostic biomarkers in colorectal cancer. BMC
365   Med Genomics 12:54. 10.1186/s12920-019-0501-z

366   Hsu CC, Liao WY, Chan TS, Chen WY, Lee CT, Shan YS, Huang PJ, Hou YC, Li CR, and Tsai KK. 2019. The
367   differential distributions of ASPM isoforms and their roles in Wnt signaling, cell cycle progression, and
368   pancreatic cancer prognosis. J Pathol. 10.1002/path.5341

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

373   Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K,
374   Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, Marine JC, and Aerts S. 2014.
375   iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS
376   Comput Biol 10:e1003731. 10.1371/journal.pcbi.1003731

377   Jiao YJ, Jin DD, Jiang F, Liu JX, Qu LS, Ni WK, Liu ZX, Lu CH, Ni RZ, Zhu J, and Xiao MB. 2018.
378   Characterization and proteomic profiling of pancreatic cancer-derived serum exosomes.
379   10.1002/jcb.27465

380   Jiao YJ, Jin DD, Jiang F, Liu JX, Qu LS, Ni WK, Liu ZX, Lu CH, Ni RZ, Zhu J, and Xiao MB. 2019.
381   Characterization and proteomic profiling of pancreatic cancer-derived serum exosomes. J Cell Biochem
382   120:988-999. 10.1002/jcb.27465

383   Jin Q, and Liu G. 2018. Decellularized breast matrix as bioactive microenvironment for in vitro three-
384   dimensional cancer culture. 10.1002/jcp.26782

385   Jing X, Wang XJ, Zhang T, Zhu W, Fang Y, Wu H, Liu X, Ma D, Ji X, Jiang Y, Liu K, Chen X, Shi Y, Zhang
386   Y, Shi M, Qiu W, and Zhao R. 2019. Cell-Cycle-Dependent Phosphorylation of PRPS1 Fuels
387   Nucleotide Synthesis and Promotes Tumorigenesis. Cancer Res 79:4650-4664. 10.1158/0008-
388   5472.can-18-2486

389   Kalimutho M, Sinha D, Jeffery J, Nones K, Srihari S, Fernando WC, Duijf PH, Vennin C, Raninga P,
390   Nanayakkara D, Mittal D, Saunus JM, Lakhani SR, Lopez JA, Spring KJ, Timpson P, Gabrielli B,
391   Waddell N, and Khanna KK. 2018. CEP55 is a determinant of cell fate during perturbed mitosis in

392 breast cancer. EMBO Mol Med 10. 10.15252/emmm.201708566

393 Kokkinakis DM, Liu X, and Neuner RD. 2005. Modulation of cell cycle and gene expression in pancreatic tumor
394   cell lines by methionine deprivation (methionine stress): implications to the therapy of pancreatic
395   adenocarcinoma. Mol Cancer Ther 4:1338-1348. 10.1158/1535-7163.mct-05-0141

396 Le NQ, Ho QT, and Ou YY. 2018. Classifying the molecular functions of Rab GTPases in membrane trafficking
397   using deep convolutional neural networks. Anal Biochem 555:33-41. 10.1016/j.ab.2018.06.011

398 Le NQ, Nguyen TT, and Ou YY. 2017. Identifying the molecular functions of electron transport proteins using
399   radial basis function networks and biochemical properties. J Mol Graph Model 73:166-178.
400   10.1016/j.jmgm.2017.01.003

401 Li J, Tan W, Peng L, Zhang J, Huang X, Cui Q, Zheng J, Tan W, Wu C, and Lin D. 2018a. Integrative analysis
402   of gene expression profiles reveals specific signaling pathways associated with pancreatic duct
403   adenocarcinoma. Cancer Commun (Lond) 38:13. 10.1186/s40880-018-0289-9

404 Li L, Lei Q, Zhang S, Kong L, and Qin B. 2017a. Screening and identification of key biomarkers in
405   hepatocellular carcinoma: Evidence from bioinformatic analysis. Oncol Rep 38:2607-2618.
406   10.3892/or.2017.5946

407 Li L, Wang G, Li N, Yu H, Si J, and Wang J. 2017b. Identification of key genes and pathways associated with
408   obesity in children. Exp Ther Med 14:1065-1073. 10.3892/etm.2017.4597

409 Li M, Gao J, Li D, and Yin Y. 2018b. CEP55 Promotes Cell Motility via JAK2(-)STAT3(-)MMPs Cascade in
410   Hepatocellular Carcinoma. Cells 7. 10.3390/cells7080099

411 Lindeman GJ, Gaubatz S, Livingston DM, and Ginsberg D. 1997. The subcellular localization of E2F-4 is cell-
412   cycle dependent. Proc Natl Acad Sci U S A 94:5095-5100. 10.1073/pnas.94.10.5095

413 Liu LM, Xiong DD, Lin P, Yang H, Dang YW, and Chen G. 2018. DNA topoisomerase 1 and 2A function as
414   oncogenes in liver cancer and may be direct targets of nitidine chloride. Int J Oncol 53:1897-1912.
415   10.3892/ijo.2018.4531

416 Lu W, Li N, and Liao F. 2019. Identification of Key Genes and Pathways in Pancreatic Cancer Gene Expression
417   Profile by Integrative Analysis. Genes (Basel) 10. 10.3390/genes10080612

418 Lu Y, and Li C. 2018. Identification of hub genes and analysis of prognostic values in pancreatic ductal
419   adenocarcinoma by integrated bioinformatics methods. 10.1007/s11033-018-4325-2

420 Ma Y, Xin Y, Li R, Wang Z, Yue Q, Xiao F, and Hao X. 2014. TFDP3 was expressed in coordination with E2F1
421   to inhibit E2F1-mediated apoptosis in prostate cancer. Gene 537:253-259. 10.1016/j.gene.2013.12.051

422 Mao J, Tian Y, Wang C, Jiang K, Li R, Yao Y, Zhang R, Sun D, Liang R, Gao Z, Wang Q, and Wang L. 2019.
423   CBX2 Regulates Proliferation and Apoptosis via the Phosphorylation of YAP in Hepatocellular
424   Carcinoma. J Cancer 10:2706-2719. 10.7150/jca.31845

425 Meel MH, Guillen Navarro M, de Gooijer MC, Metselaar DS, Waranecki P, Breur M, Lagerweij T, Wedekind
426   LE, Koster J, van de Wetering MD, Schouten-van Meeteren N, Aronica E, van Tellingen O, Bugiani
427   M, Phoenix TN, Kaspers GJL, and Hulleman E. 2019. MEK/MELK inhibition and blood-brain barrier-
428   deficiencies in atypical teratoid/rhabdoid tumors. Neuro Oncol. 10.1093/neuonc/noz151

429 Nobes CD, and Hall A. 1999. Rho GTPases control polarity, protrusion, and adhesion during cell movement. J
430   Cell Biol 144:1235-1244.

431 Obuse C, Iwasaki O, Kiyomitsu T, Goshima G, Toyoda Y, and Yanagida M. 2004. A conserved Mis12
432   centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. Nat Cell

433   Biol 6:1135-1141. 10.1038/ncb1187

434 Peng T, Zhou W, Guo F, Wu HS, Wang CY, Wang L, and Yang ZY. 2017a. Centrosomal protein 55 activates
435   NF-kappaB signalling and promotes pancreatic cancer cells aggressiveness. Sci Rep 7:5925.
436   10.1038/s41598-017-06132-z

437 Peng T, Zhou W, Guo F, Wu HS, Wang CY, Wang L, and Yang ZY. 2017b. Centrosomal protein 55 activates
438   NF-κB signalling and promotes pancreatic cancer cells aggressiveness. Sci Rep 7:5925.
439   10.1038/s41598-017-06132-z

440 Piao J, Zhu L, Sun J, Li N, Dong B, Yang Y, and Chen L. 2019. High expression of CDK1 and BUB1 predicts
441   poor prognosis of pancreatic ductal adenocarcinoma. Gene 701:15-22. 10.1016/j.gene.2019.02.081

442 Pihlak R, Weaver JMJ, Valle JW, and McNamara MG. 2018. Advances in Molecular Profiling and
443   Categorisation of Pancreatic Adenocarcinoma and the Implications for Therapy. Cancers (Basel) 10.
444   10.3390/cancers10010017

445 Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, and Chinnaiyan
446   AM. 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia
447   6:1-6.

448 Serrill JD, Sander M, and Shih HP. 2018. Pancreatic Exocrine Tissue Architecture and Integrity are Maintained
449   by E-cadherin During Postnatal Development.   8:13451. 10.1038/s41598-018-31603-2

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

453 Shao N, Yuan K, Zhang Y, Yun Cheang T, Li J, and Lin Y. 2018. Identification of key candidate genes, pathways
454   and related prognostic values in ER-negative/HER2-negative breast cancer by bioinformatics analysis.
455   J buon 23:891-901.

456 Shi Z, Chiang CI, Labhart P, Zhao Y, Yang J, Mistretta TA, Henning SJ, Maity SN, and Mori-Akiyama Y. 2015.
457   Context-specific role of SOX9 in NF-Y mediated gene regulation in colorectal cancer cells. Nucleic
458   Acids Res 43:6257-6269. 10.1093/nar/gkv568

459 Siegel RL, Miller KD, and Jemal A. 2018. Cancer statistics, 2018.   68:7-30. 10.3322/caac.21442

460 Singh PK, Srivastava AK, Rath SK, Dalela D, Goel MM, and Bhatt ML. 2015. Expression and clinical
461   significance of Centrosomal protein 55 (CEP55) in human urinary bladder transitional cell carcinoma.
462   Immunobiology 220:103-108. 10.1016/j.imbio.2014.08.014

463 Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos
464   A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, and von Mering C. 2015. STRING v10: protein-protein
465   interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447-452.
466   10.1093/nar/gku1003

467 Tao J, Zhi X, Tian Y, Li Z, Zhu Y, Wang W, Xie K, Tang J, Zhang X, Wang L, and Xu Z. 2014. CEP55
468   contributes to human gastric carcinoma by regulating cell proliferation. Tumour Biol 35:4389-4399.
469   10.1007/s13277-013-1578-1

470 van Amerongen MJ, Diehl F, Novoyatleva T, Patra C, and Engel FB. 2010. E2F4 is required for cardiomyocyte
471   proliferation. Cardiovasc Res 86:92-102. 10.1093/cvr/cvp383

472 Wang A, Dai H, Gong Y, Zhang C, Shu J, Luo Y, Jiang Y, Liu W, and Bie P. 2019. ANLN-induced EZH2
473   upregulation promotes pancreatic cancer progression by mediating miR-218-5p/LASP1 signaling axis.

474   J Exp Clin Cancer Res 38:347. 10.1186/s13046-019-1340-7

475   Wang JH, Zhao LF, Lin P, Su XR, Chen SJ, Huang LQ, Wang HF, Zhang H, Hu ZF, Yao KT, and Huang ZX.

476   2014. GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular

477   networks based on free terms. Bioinformatics 30:2534-2536. 10.1093/bioinformatics/btu241

478   Xu X, Zhou Y, Miao R, Chen W, Qu K, Pang Q, and Liu C. 2016. Transcriptional modules related to

479   hepatocellular carcinoma survival: coexpression network analysis. Front Med 10:183-190.

480   10.1007/s11684-016-0440-4

481   Zhao H, Zheng GH, Li GC, Xin L, Wang YS, Chen Y, and Zheng XM. 2019. Long noncoding RNA LINC00958

482   regulates cell sensitivity to radiotherapy through RRM2 by binding to microRNA-5095 in cervical

483   cancer. J Cell Physiol 234:23349-23359. 10.1002/jcp.28902

484   Zhao J, Han SX, Ma JL, Ying X, Liu P, Li J, Wang L, Zhang Y, Ma J, Zhang L, and Zhu Q. 2013. The role of

485   CDK1 in apoptin-induced apoptosis in hepatocellular carcinoma cells. Oncol Rep 30:253-259.

486   10.3892/or.2013.2426

487   Zhou C, Hancock JL, Khanna KK, and Homer HA. 2019. First meiotic anaphase requires Cep55-dependent

488   inhibitory Cdk1 phosphorylation. J Cell Sci. 10.1242/jcs.233379

489   Zhu T, Gao YF, Chen YX, Wang ZB, Yin JY, Mao XY, Li X, Zhang W, Zhou HH, and Liu ZQ. 2017. Genome-

490   scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer.

491   Oncotarget 8:21281-21289. 10.18632/oncotarget.15068

492   Zhuo M, Yuan C, Han T, Cui J, Jiao F, and Wang L. 2018. A novel feedback loop between high MALAT-1 and

493   low miR-200c-3p promotes cell migration and invasion in pancreatic ductal adenocarcinoma and is

494   predictive of poor prognosis. BMC Cancer 18:1032. 10.1186/s12885-018-4954-9

495

Figures

497   **Figure 1. Venn diagram, PPI network and the most significant modules of DEG.** (A) Venn diagram

498   analyzing the DEGs of the GSE32676, GSE15471 and GSE71989 datasets; an overlap of 210 genes was

499   identified. (B) The PPI network of DEGs was constructed by Cytoscape. Upregulated genes are marked in red,

500   and the depth of color represents the degree of gene interaction with other genes; downregulated genes are

501   marked in green. (C) Interaction network of 13 hub genes.

502   **Figure 2. GO and KEGG pathway analysis of hub genes using ClueGO and CluePedia.**

503   **Figure 3. Interactions, enrichment and transcription factor analysis of 13 hub genes.** (A). The cocitation

504   network of 13 hub genes; the number on the line shows the number of studies cocited. (B). Heatmap of

505   enrichment analysis of 13 hub genes; the depth of green color represents the degree of enrichment. (C)

506   Transcription factor analysis of hub genes; the red nodes represent hub genes, and the purple nodes represent

507   transcription factors.

508 **Figure 4. The mRNA expression levels of 13 hub genes in PAAD.** Published online data of gene mRNA

509 expression levels were analyzed using the GEPIA platform. mRNA levels of these 13 hub genes, including

510 TOP2A, CDK1, RRM2, PRC1, NEK2, ZWINT, DTL, MELK, CENPF, CEP55, ANLN, ASPM, and ECT2,

511 were all higher in PAAD tissues than in normal tissues. *P<0.05.

512 **Figure 5. Overall survival analyses of hub genes performed using the cBioPortal online platform.** A log

513 rank test P<0.05 was considered statistically significant.

514 **Figure 6. Disease-free survival analyses of hub genes performed using the cBioPortal online platform.** A

515 log rank test P <0.05 was considered statistically significant.

516 **Figure 7. Oncomine analysis of (A) CDK1 and (B) CEP55 in cancer vs. normal tissue.** Heat maps of CDK1

517 and CEP55 gene expression in clinical pancreatic cancer samples vs. normal tissues. 1. Pancreatic Ductal

518 Adenocarcinoma Epithelia vs. Normal Grutzmann Pancreas, Neoplasia, 2004. 2. Pancreatic Carcinoma vs.

519 Normal Pei Pancreas, Cancer Cell, 2009. 3. Pancreatic Carcinoma vs. Normal Segara Pancreas, Clin Cancer Res,

520 2005. (C) CDK1 mRNA expression and (D) CEP55 mRNA expression in pancreatic cancer compared with that

521 in normal pancreatic tissues in the Grutzmann Pancreas dataset. Association between the expression of (E) CDK1

522 and (F) CEP55 and tumor grade in the Grutzmann Pancreas dataset. P <0.05 was considered statistically

523 significant.

524 **Figure 8. The expression of CDK1 and CEP55 in clinical specimens.** (A, B) Western blotting detection of

525 CDK1 and CEP55 protein in eight snap-frozen tumorous and tissues of pancreatic cancer patients. (C, D) The

526 bar chart reveled that the expression ratio of CDK1 and CEP55 to GAPDH by densitometry. Data are represented

527 as the mean ± SEM (*P < 0.05, tumor tissues compared with the non-cancerous adjacent tissues).

528

529

530

531

532

GO analysis of upregulated DEGs in Pancreatic ductal adenocarcinoma

1   Table 1. GO analysis of upregulated DEGs in Pancreatic ductal adenocarcinoma

GO ID	Description	Gene count	P-value
GO-BP Terms			
GO:0016477	cell migration	13	1.83E-07
GO:0098609	cell-cell adhesion	13	2.08E-05
GO:0007160	cell-matrix adhesion	8	3.55E-05
GO:0022617	extracellular matrix disassembly	7	1.20E-04
GO:0048333	mesodermal cell differentiation	4	1.58E-04
GO:0031581	hemidesmosome assembly	4	2.10E-04
GO:0090004	positive regulation of establishment of protein localization to plasma membrane	5	2.26E-04
GO:0007155	cell adhesion	15	2.45E-04
GO:0007229	integrin-mediated signaling pathway	7	5.08E-04
GO-CC Terms			
GO:0070062	extracellular exosome	63	1.50E-10
GO:0005925	focal adhesion	16	7.65E-06
GO:0005913	cell-cell adherens junction	14	1.91E-05
GO:0005615	extracellular space	28	3.05E-04
GO:0005886	plasma membrane	61	5.23E-04
GO:0005737	cytoplasm	72	9.52E-04
GO:0009986	cell surface	15	9.53E-04
GO-MF Terms			
GO:0098641	cadherin binding involved in cell-cell adhesion	13	3.05E-05
GO:0042803	protein homodimerization activity	18	8.70E-04
GO:0005509	calcium ion binding	16	0.004808
GO:0005515	protein binding	102	0.013495

GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function.

KEGG pathway enrichment analysis of upregulated DEGs in Pancreatic ductal adenocarcinoma

1   Table 2. KEGG pathway enrichment analysis of upregulated DEGs in Pancreatic ductal adenocarcinoma

ID	Description	Gene count	P-value
hsa04512	ECM-receptor interaction	9	1.82E-06
hsa05200	Pathways in cancer	12	0.001424
hsa05205	Proteoglycans in cancer	8	0.003332
hsa04510	Focal adhesion	8	0.003924
hsa04115	p53 signaling pathway	5	0.004128
hsa04151	PI3K-Akt signaling pathway	9	0.019238

KEGG, Kyoto Encyclopedia of Genes and Genomes.

Description of 13 hub genes of Pancreatic ductal adenocarcinoma

1 Table 3. Description of 13 hub genes of Pancreatic ductal adenocarcinoma

NO.	Gene symbol	Full name	Function
1	TOP2A	DNA topoisomerase II alpha	Controlling and altering the topologic states of DNA during transcription. Involving in chromosome condensation, chromatid separation, and the relief of torsional stress.
2	CDK1	cyclin dependent kinase 1	Essential for G1/S and G2/M phase transitions of eukaryotic cell cycle and mitotic cyclins.
3	RRM2	ribonucleotide reductase regulatory subunit M2	Catalyzing the formation of deoxyribonucleotides from ribonucleotides.
4	PRC1	protein regulator of cytokinesis 1	A substrate of several CDKs and necessary for polarizing parallel microtubules and concentrating the factors responsible for contractile ring assembly.
5	NEK2	NIMA related kinase 2	A serine/threonine-protein kinase that is involved in mitotic regulation.
6	ZWINT	ZW10 interacting kinetochore protein	Clearly involved in kinetochore function.
7	DTL	denticleless E3 ubiquitin protein ligase homolog	Cellular response to DNA damage stimulus.
8	MELK	maternal embryonic leucine zipper kinase	G2/M transition of mitotic cell cycle apoptotic process.
9	CENPF	centromere protein F	Associating with the centromere-kinetochore complex and playing a role in chromosome segregation during mitotis.
10	CEP55	centrosomal protein 55	Cranial skeletal system development and establishment of protein localization.
11	ANLN	anillin actin binding protein	An actin-binding protein that plays a role in cell growth and migration, and in cytokinesis.
12	ASPM	abnormal spindle microtubule assembly	Essential for normal mitotic spindle function in embryonic neuroblasts and regulating neurogenesis.
13	ECT2	epithelial cell transforming 2	A guanine nucleotide exchange factor and transforming protein that is related to Rho-specific exchange factors and yeast cell cycle regulators.

CDKs, cyclin-dependent kinases.

Hub genes identified by the present study using Genclip 2.0

1   Table 4. Hub genes identified by the present study using Genclip 2.0

Gene	Co-genes	Co-cite	Total
CDK1	5	53	6822
TOP2A	1	23	6370
RRM2	1	7	395
CEP55	1	1	58
PRC1	1	15	334
ECT2	1	11	171

Figure 1

Venn diagram, PPI network and the most significant modules of DEG.

(A) Venn diagram analyzing the DEGs of GSE32676, GSE15471 and GSE71989 datasets, and an overlap of 210 genes was identified. (B) The PPI network of DEGs was constructed by Cytoscape. Upregulated genes are marked in red, and the depth of color represents the gene interaction degree with other genes; downregulated genes are marked in green. (C) Interaction network of 13 hub genes.

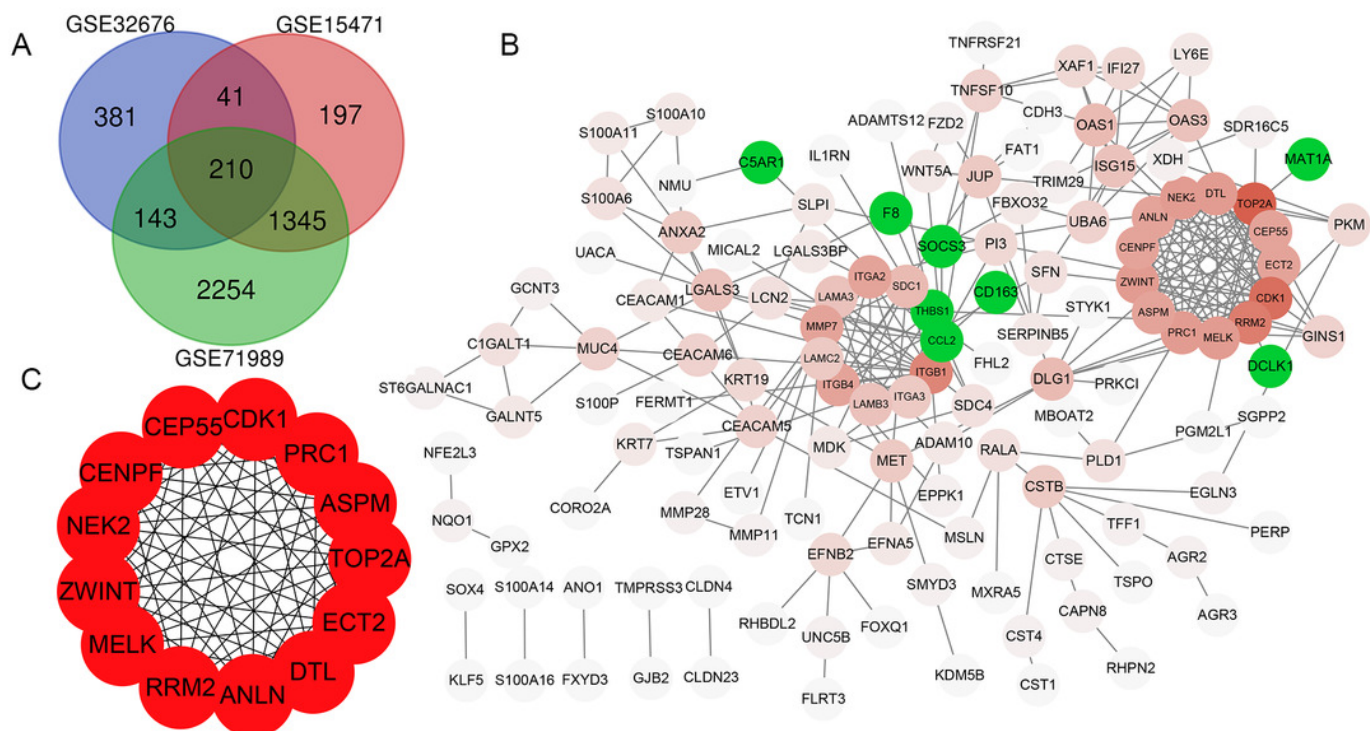



Figure 2

GO and KEGG pathway analysis of hub genes using ClueGO and CluePedia.

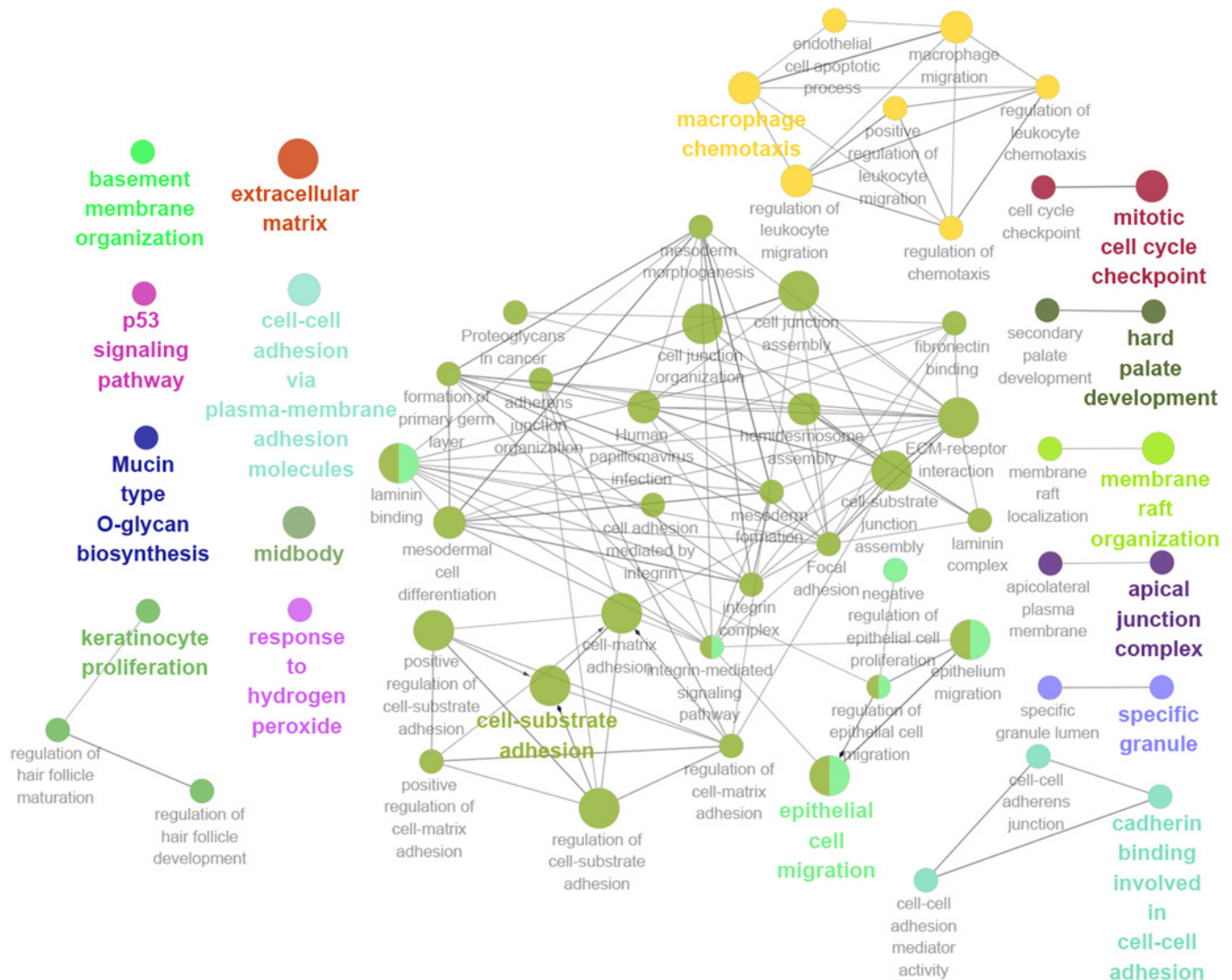



Figure 3

Interactions, enrichment and transcription factor analysis of 13 hubgenes.

(A). The cocitation network of 13 hub genes; the number on the line shows the number of studies cocited. (B). Heatmap of enrichment analysis of 13 hub genes; the depth of green color represents the degree of enrichment. (C) Transcription factor analysis of hub genes; the red nodes represent hub genes, and the purple nodes represent transcription factors.

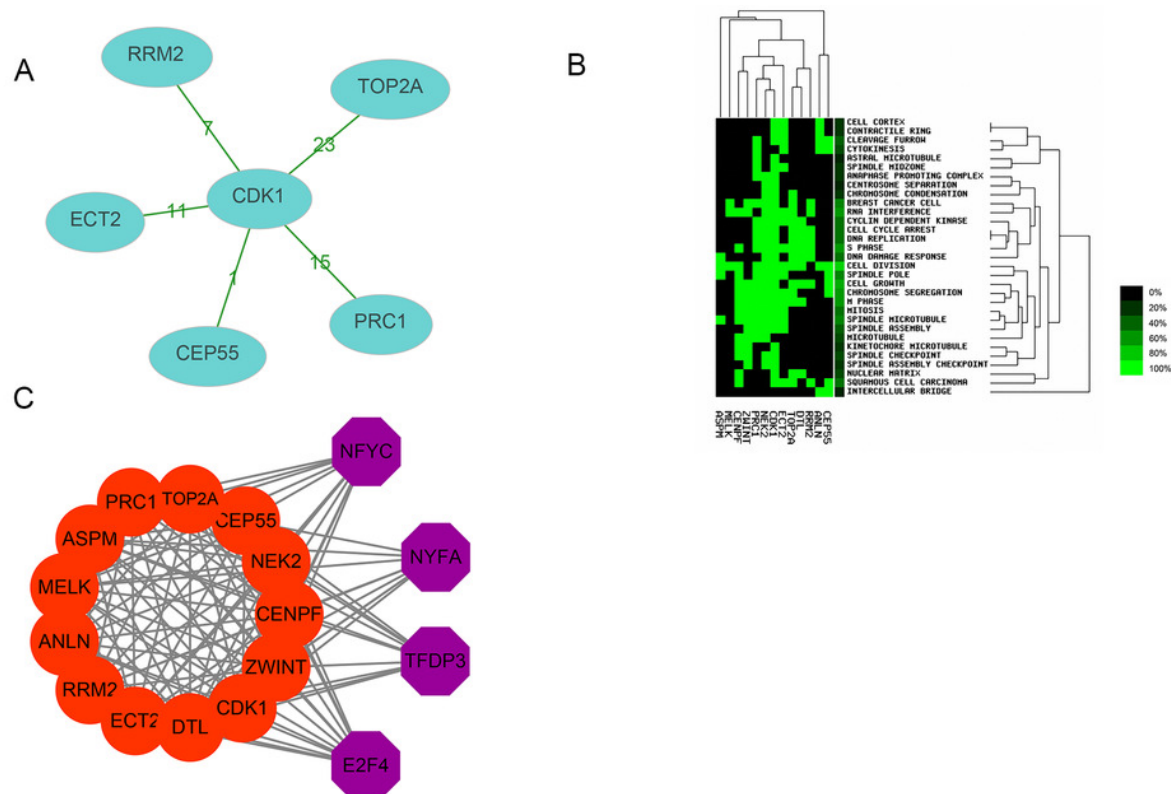



Figure 4

The mRNA expression levels of 13 hubgenes in PAAD.

The mRNA expression levels of 13 hubgenes in PAAD. The published online data of gene mRNA expression level were analyzed by GEPIA platform. These 13 hubgenes were all higher than in PAAD tissues, compared with those in normal tissues, including TOP2A, CDK1, RRM2, PRC1, NEK2, ZWINT, DTL, MELK, CENPF, CEP55, ANLN, ASPM, ECT2. *P<0.05.

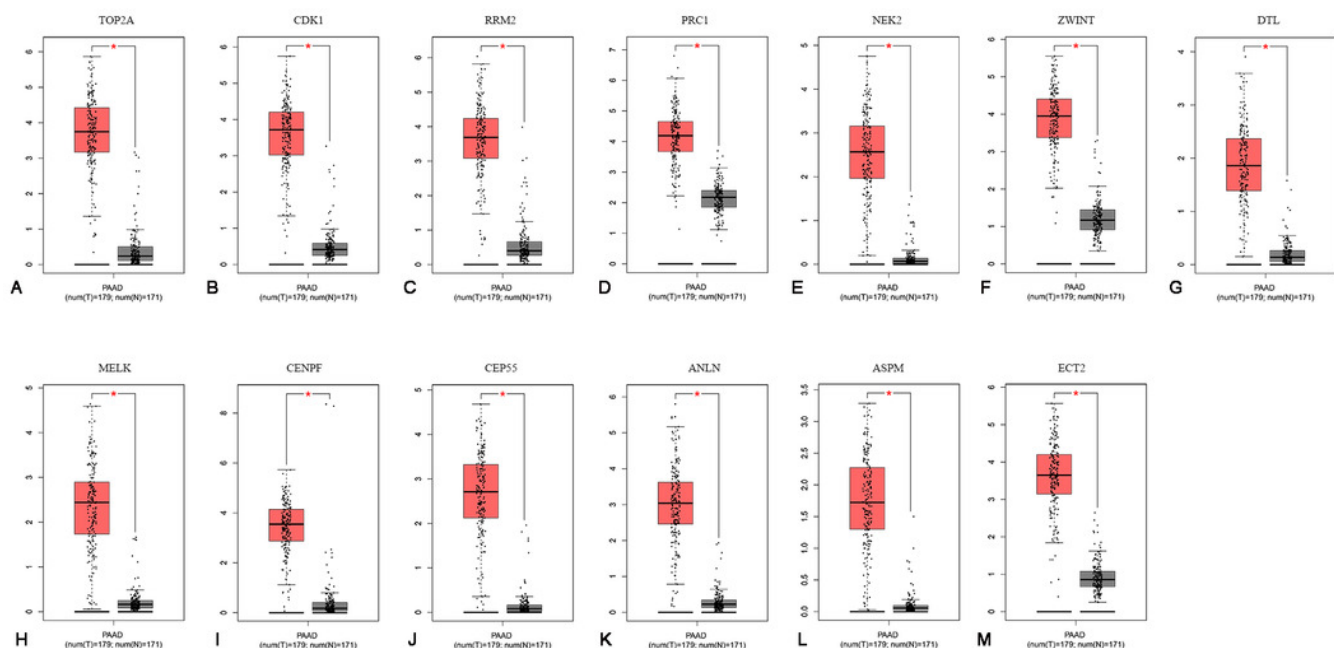



Figure 5

A log rank test P<0.05 was considered statistically significant.

Log rank Test P<0.05 was considered statistically significant.

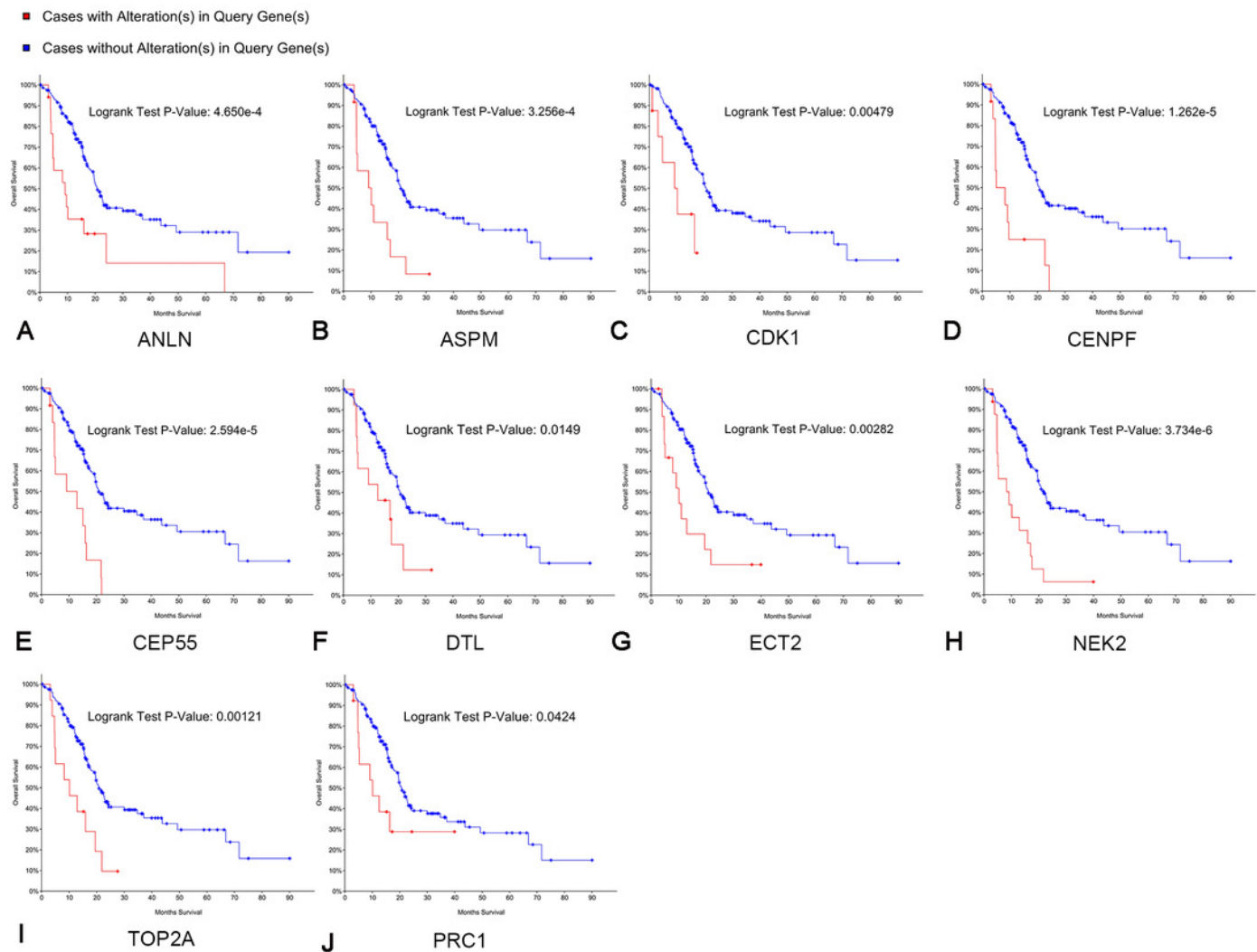



Figure 6

Disease-free survival analyses of hubgenes were performed using the cBioPortal online platform.

A log rank test P <0.05 was considered statistically significant.

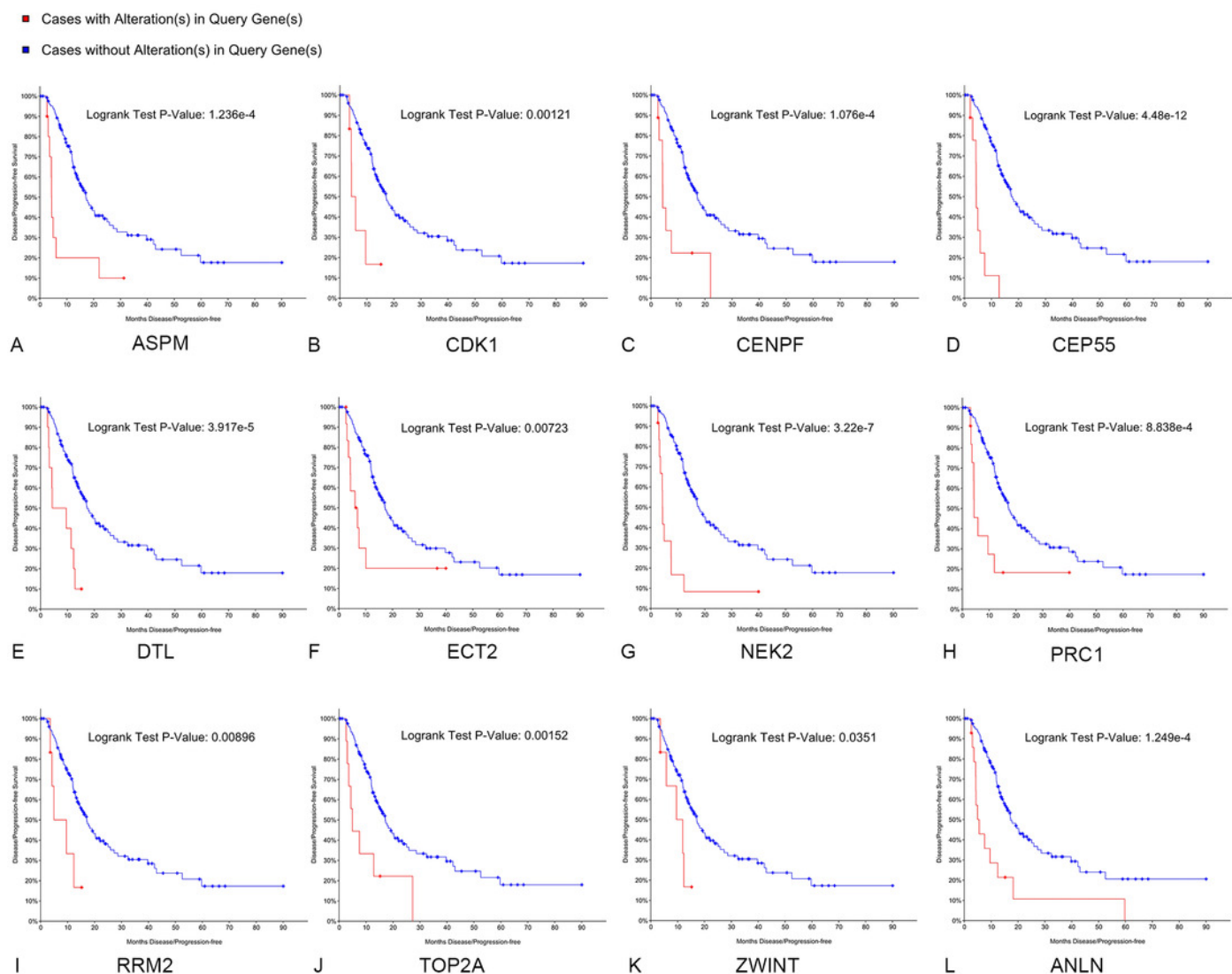



Figure 7

Oncomine analysis of (A) CDK1 and (B) CEP55 in cancer vs. normal tissue.

Heat maps of CDK1 and CEP55 gene expression in clinical pancreatic cancer samples vs. normal tissues. 1. Pancreatic Ductal Adenocarcinoma Epithelia vs. Normal Grutzmann Pancreas, Neoplasia, 2004. 2. Pancreatic Carcinoma vs. Normal Pei Pancreas, Cancer Cell, 2009. 3. Pancreatic Carcinoma vs. Normal Segara Pancreas, Clin Cancer Res, 2005. (C) CDK1 mRNA expression and (D) CEP55 mRNA expression in pancreatic cancer compared with normal pancreatic tissues in the Grutzmann Pancreas dataset. Association between the expression of (E) CDK1 and (F) CEP55 and tumor grade in the Grutzmann Pancreas dataset. P <0.05 was considered statistically significant.

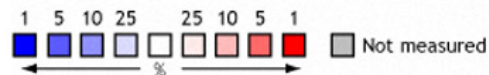
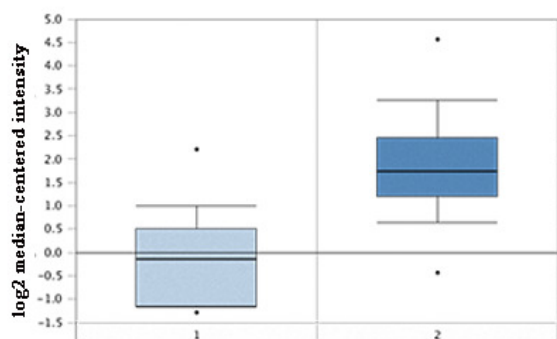
A

Median Rank	p-Value	Gene
337.0	1.37E-7	CDK1

1 2 3

1. P-value = 1.37E-7   fold change = 3.888

2. P-value = 0.006   fold change = 4.942

3. P-value = 0.001   fold change = 1.457

B

Median Rank	p-Value	Gene
87.0	6.13E-10	CEP55

1 2 3

1. P-value = 6.13E-10   fold change = 4.102

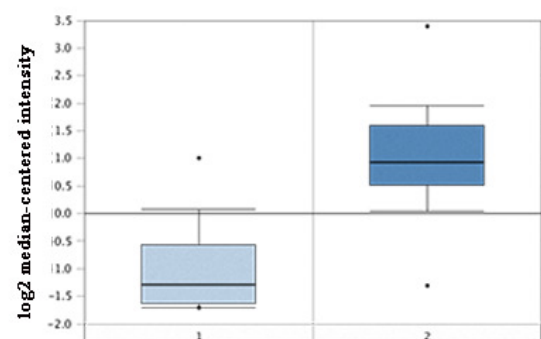
2. P-value = 3.61E-4   fold change = 20.553

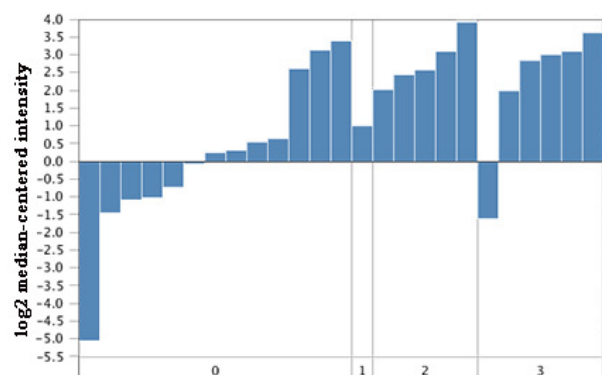
3. P-value = 0.004   fold change = 1.494

1 5 10 25   25 10 5 1   Not measured
%

C   CDK1

D   CEP55

Legend
1. Pancreas (16)
2. Pancreatic Carcinoma (36)

Legend
1. Pancreas (16)
2. Pancreatic Carcinoma (36)

E   CDK1 grade

F   CEP55 grade

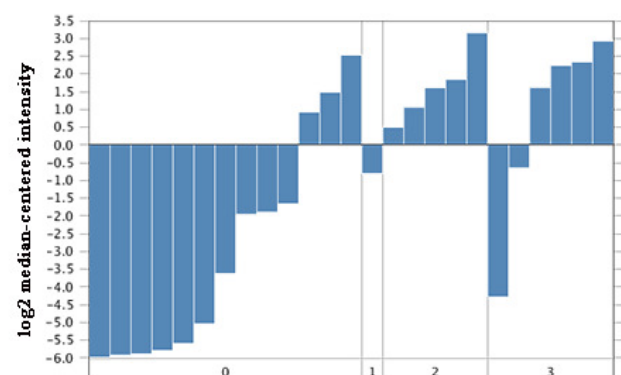
Legend
0. No value (13)   1.Grade 1 (1)
2.Grade 2 (5)   3.Grade 3 (6)

Legend
0. No value (13)   1.Grade 1 (1)
2.Grade 2 (5)   3.Grade 3 (6)

Figure 8

The expression of CDK1 and CEP55 in clinical specimens.

(A, B) Western blotting detection of CDK1 and CEP55 protein in eight snap-frozen tumorous and tissues of pancreatic cancer patients. (C, D) The bar chart reveled that the expression ratio of CDK1 and CEP55 to GAPDH by densitometry. Data are represented as the mean ± SEM (*P < 0.05, tumor tissues compared with the non-cancerous adjacent tissues).