

Identification of potential crucial genes associated with the pathogenesis and prognosis of pancreatic adenocarcinoma

LAN-ER SHI, XIN SHANG, KE-CHAO NIE, QIANG XU, NA-BEI CHEN and ZHANG-ZHI ZHU

Department of Endocrinology, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong 510405, P.R. China

Received February 4, 2020; Accepted June 22, 2020

DOI: 10.3892/ol.2020.11921

Abstract. Pancreatic adenocarcinoma (PAAD) is a type of malignant tumor with the highest mortality rate among all neoplasms worldwide, and its exact pathogenesis is still poorly understood. Timely diagnosis and treatment are of great importance in order to decrease the mortality rate of PAAD. Therefore, identifying new biomarkers for diagnosis and prognosis is essential to enable early detection of PAAD and to improve the overall survival (OS) rate. In order to screen and integrate differentially expressed genes (DEGs) between PAAD and normal tissues, a total of seven datasets were downloaded from the Gene Expression Omnibus database and the 'limma' and 'robustRankG' packages in R software were used. The Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analysis of the DEGs was performed using the Database for Annotation, Visualization and Integrated Discovery website, and the protein-protein interaction network analysis was performed using the Search Tool for the Retrieval of Interacting Genes/Proteins database. A gene prognostic signature was constructed using the Cox regression model. A total of 10 genes (CDK1, CCNB1, CDC20, ASPM, UBE2C, TPX2, TOP2A, NUSAP1, KIF20A and DLGAP5) that may be associated with pancreatic adenocarcinoma were identified. According to the differentially expressed genes in The Cancer Genome Atlas, the present study set up four prognostic signatures (matrix metalloproteinase 12, sodium voltage-gated channel α subunit 11, tetraspanin 1 and SH3 domain and tetratricopeptide repeats-containing 2), which effectively predicted OS. The hub genes that were highly associated with the occurrence, development and prognosis of PAAD were identified, which may be helpful to further understand the molecular basis of pancreatic cancer and guide the synthesis of drugs for PPAD.

Introduction

Pancreatic adenocarcinoma (PAAD) is a type of malignant tumor that primarily originates from ductal adenocarcinomas of the glandular epithelium, and causes 331,000 deaths worldwide every year due to there being few symptoms before the disease reaches an advanced stage. Therefore, PAAD has a poor prognosis and a high relative incidence rate (1). Data between 1974 and 2013 has demonstrated that the overall increase in the incidence of PAAD is associated with risk factors such as smoking, obesity and diabetes in the USA (2). In recent decades the proportion of patients with metastatic PAAD with survival time >1 year has increased significantly, but the proportion of deaths within 2 months is still considerable (50.6%; $P < 0.001$) (3). The median overall survival (OS) is ~16 months after resection in patients with non-metastatic pancreatic ductal adenocarcinoma, but the resectability rate is also low, <10% (4). Despite the common use of surgical resection and adjuvant or palliative chemotherapy, improvements in OS are negligible among all patients, as the majority of patients only receive supportive treatment, and most preoperative chemoradiotherapy is associated with high postoperative morbidity and mortality (5,6). Therefore, novel effective diagnosis and treatment options are still required.

Previous studies have demonstrated that drugs that inhibit specific targets may improve the therapeutic effectiveness and overcome the resistance of pancreatic cancer to the majority of standard therapies (7-9). The expression of RNA binding protein in PAAD cells has been demonstrated to change the expression of mRNA, subsequently altering the entire transcriptome and proteome, which suggests that the gene regulatory mechanism monitors the proto-oncogenic signaling pathway (10,11). As the currently published research is limited by small sample sizes, application of different technology platforms, constant discovery of novel mRNAs and different methods of processing and analyzing data, the common disadvantage in mRNA expression profiling research is a lack of consistency, resulting in non-specific and insensitive biomarkers. Thus, there is an urgent requirement for the identification of differential genes that may provide clues to detect PAAD early and improve the OS time of the patients.

The Cancer Genome Atlas (TCGA; <https://www.cancer.gov/>) and Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/pmc/>) are open access databases, which can

Correspondence to: Professor Zhang-Zhi Zhu, Department of Endocrinology, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, 12 Jichang Road, Guangzhou, Guangdong 510405, P.R. China
E-mail: zzz@gzucm.edu.cn

Key words: pancreatic adenocarcinoma, gene, biomarker, Gene Expression Omnibus, The Cancer Genome Atlas

be used to analyze the whole genome and epigenome of the selected pancreatic cancer types. Using large-scale parallel sequencing technology, it is possible to reveal the previously unclear molecular mechanisms, determine the tissue-specific changes and provide clues for PAAD staging, pathological grading and drug target determination of the disease in order to improve the current understanding of PAAD and enable effective treatment decisions (12-14). However, to the best of our knowledge, previous studies on PAAD diagnostic markers have not been comprehensive or only analyzed the genes associated with a poor prognosis (15,16). Traditional experimental methods can only recognize a single gene or a limited number of genes at the same time; thus, the previous research progress on biomarkers of pancreatic cancer has not yet been translated into a significant improvement in OS rates or a reduction in mortality. For example, previous studies have demonstrated that KRAS is the most common mutated gene in pancreatic ductal adenocarcinoma (17), and KRAS mutation can be used as a marker of pancreatic cancer (18). However, inhibitors targeting KRAS gene have not been successful in treatment (19).

The present study aimed to identify hub genes that were highly associated with PAAD development and prognosis in order to improve the current understanding of the molecular basis of PAAD and direct new therapeutic techniques.

Materials and methods

Microarray data. The present study downloaded seven publicly available gene expression profiles (GSE15471, GSE16515, GSE28735, GSE32676, GSE55643, GSE62165 and GSE62452) (20-26) from the GEO database, which met the following criteria: i) Included human pancreas samples; ii) Contained both pancreatic cancer and normal (or adjacent) samples; iii) The sample size was ≥ 30 ; iv) The sample size of the case and control group was >15 samples/group. Table I presents the details of the seven datasets. In total, 566 samples were analyzed in the present study.

Integration of the microarray data. The present study used the 'limma' package (limma: Data analysis, linear models and differential expression for microarray data. URL <http://bioinf.wehi.edu.au/limma>) of R 3.6.1 software (<http://www.R-project.org/>) to normalize and \log_2 transform the matrix files of each GEO dataset, and identify the GEO-DEGs between normal pancreatic tissue and pancreatic cancer tissue in each GEO dataset. 'RobustRankAggreg' (RRA; <https://CRAN.R-project.org/package=RobustRankAggreg>) was used to integrate the GEO-DEGs from each dataset. \log fold-change (FC) >1 , P-value <0.05 and adjusted P-value <0.05 were used as the thresholds of statistical significance for the GEO-DEGs.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the DEGs. In order to further investigate the roles of GEO-DEGs in the development of PAAD, the present study used the Database for Annotation, Visualization and Integrated Discovery (DAVID; <https://david.ncifcrf.gov/>) to annotate and analyze the functions of GEO-DEGs to determine the biological processes, molecular functions, cellular components and signaling

pathways associated with these GEO-DEGs. A false discovery rate (FDR) <0.05 was considered to indicate a statistically significant difference.

Protein-protein interaction (PPI) and modular analysis. A gene network can be used to analyze the association between proteins and genes, and further clarify the specific association between genes and diseases (27). The present study used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; <https://string-db.org/>) database, which is an online tool for assessing PPI information, to analyze the potential association between proteins encoded by the GEO-DEGs; the GEO-DEGs with a minimum required interaction score >0.9 were selected, and disconnected nodes were removed from the network. Cytoscape 3.7.2 (<https://cytoscape.org/>) is a type of software that can graphically display, analyze and edit the network, as well as add annotation information, and it was used in the present study to complete the visualization of the PPI network and calculate the correlation degrees of DEGs. The top 10 genes with the highest degrees in the PPI network were regarded as key genes. Molecular complex detection (MCODE) is a plug-in of Cytoscape 3.7.2 that uses the inherent associations between proteins in the network in order to identify gene clusters (highly interconnected regions); module analysis of the PPI network was performed using MCODE (degree cutoff=2; node score cutoff=0.2; k-core=2 and Max. Depth=100). The functional enrichment analysis of each module was performed using DAVID. The expression of the 10 hub genes in PAAD and adjacent normal pancreatic tissues can be downloaded from GEPIA database (<http://gepia.cancer-pku.cn>), which integrates the relevant information of TCGA database and GTEx database (<http://commonfund.nih.gov/GTEx/>).

Prognostic gene signature construction. The survival time, OS data and status of patients with PAAD were obtained from TCGA. In the repository group of TCGA data portal, the following steps were carried out: selection of 'Pancreatic Ductal Adenocarcinoma'; choosing the TCGA-PAAD project and 'transcriptome profiling' in the 'data category'; 'gene expression quantification' in the 'data type' and 'HTSeq-counts' in the 'workflow type'. Data were downloaded on 18th December 2019. The data of 171 patients with PAAD were used to build a prognostic signature by integrating gene expression and survival information. TCGA-PAAD dataset was normalized and analyzed with the 'edgeR' package (<https://bioconductor.org/packages/edgeR/>) of the R software. FDR <0.05 and \log FC >2 were used as the criteria to screen TCGA-DEGs. The expression values of DEGs in TCGA were analyzed by univariate Cox regression, and the genes associated with OS were determined. To further evaluate the relative contribution of these prognostic gene markers to the survival prediction of patients, multivariate Cox regression analysis was constructed with the top 12 genes with P <0.05 in the univariate analysis. A risk score model was constructed by linear combination of the prognostic gene expression markers and their regression coefficients (β) from the multivariate Cox proportional hazards regression analysis as previously described (28). According to the median risk score (1.069), patients were divided into high-risk (median risk score ≥ 1.069) and a low-risk (median risk score <1.069) groups. The 'survival'

Table I. Details of pancreatic cancer studies and associated microarray data sets from the Gene Expression Omnibus database.

Author, year	Dataset	Platform	Samples, n (tumor/control)	(Refs.)
Badea <i>et al.</i> , 2008	GSE15471	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	78 (39/39)	(20)
Pei <i>et al.</i> , 2009	GSE16515	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	52 (16/36)	(21)
Zhang <i>et al.</i> , 2012	GSE28735	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	90 (45/45)	(22)
Donahue <i>et al.</i> , 2012	GSE32676	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	32 (25/7)	(23)
Lunardi <i>et al.</i> , 2014	GSE55643	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F	53 (45/8)	(24)
Janky <i>et al.</i> , 2016	GSE62165	[HG-U219] Affymetrix Human Genome U219 Array	131 (118/13)	(25)
Yang <i>et al.</i> 2016	GSE62452	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array	130 (69/61)	(26)

package (<https://www.rdocumentation.org/packages/survival/>) of R software was used to analyze the survival of the patients in the high- and low-risk groups. A time-dependent receiver operating characteristic (ROC) curve was constructed using the 'survivalROC' package (<https://cran.r-project.org/web/packages/survivalROC/index.html>) to analyze the predictive accuracy of patient OS obtained using the risk score model.

Statistical analysis. Data are presented as mean \pm SD unless otherwise shown. The associations between gene expression levels and clinicopathological features were analyzed by two-sided Pearson's χ^2 test using IBM SPSS version 20.0 (IBM, Corp.). The regression analysis of univariate and multivariate Cox proportional hazards analysis was completed using the 'survival' package of R software. The Kaplan-Meier method was used to calculate the survival rates of the patients in the low- and high-risk groups. The P-value between two groups was obtained by log-rank test and $P < 0.05$ was considered to indicate a statistically significant difference.

Results

Identification of DEGs. The present study downloaded the raw data and platform information from the GEO database and reannotated them, then normalized. The raw and normalized data are presented in Fig. 1. Under the criteria of $\log_2FC > 1$, P -value < 0.05 and adjusted P -value < 0.05 , the 'limma' package was used to screen GEO-DEGs, and the DEGs in each dataset were obtained. The number of DEGs identified from each dataset are presented in Fig. 2; a total of 622 integrated GEO-DEGs were obtained through RRA rank analysis, including 387 upregulated genes and 235 downregulated genes (Table SI). The top 20 upregulated genes and the top 20 downregulated genes are presented in Fig. 3.

Functional enrichment analysis of the DEGs. In order to investigate the potential biological functions of the GEO-DEGs,

GO term and KEGG pathway analysis was performed (Fig. 4). The results revealed that the 387 upregulated genes were primarily associated with extracellular structure and composition ('extracellular space', 'proteinaceous extracellular matrix', 'extracellular region', 'extracellular exosome', 'extracellular matrix organization', 'extracellular matrix' and 'extracellular matrix disassembly') and collagen ('collagen catabolic process', 'collagen fibril organization', 'collagen binding and collagen trimer'; Fig. 4A). The 235 downregulated genes were primarily associated with extracellular structure and composition ('extracellular space', 'extracellular exosome' and 'extracellular region') and the membrane ('transport vesicle membrane', 'integral component of plasma membrane' and 'basolateral plasma membrane'; Fig. 4B). All GO items of the GEO-DEGs are presented in Table SII. The KEGG pathway analysis revealed that the upregulated GEO-DEGs were mainly enriched in 'ECM-receptor interaction', 'Focal adhesion', 'Pathways in cancer', 'PI3K-Akt signaling pathway' and 'p53 signaling pathway' (Fig. 4C), whereas the downregulated GEO-DEGs were associated with absorption and metabolism ('Pancreatic secretion', 'Protein digestion and absorption', 'Glycine, serine and threonine metabolism', 'Fat digestion and absorption'; Fig. 4D and Table SIII).

PPI network analysis. Using the STRING database, the present study analyzed the PPI of the GEO-DEGs, and visualized the PPI network using Cytoscape software. The PPI network included 291 nodes and 986 edges (Fig. 5A), and the top 10 nodes by degree were cyclin-dependent kinase 1 (CDK1), cyclin B1 (CCNB1), cell division cycle 20 homolog (CDC20), abnormal spindle microtubule assembly (ASPM), ubiquitin-conjugating enzyme E2 C (UBE2C), TPX2 microtubule nucleation factor (TPX2), DNA topoisomerase II α (TOP2A), nucleolar and spindle-associated protein 1 (NUSAP1), kinesin family member 20A (KIF20A) and discs large homolog-associated protein 5 (DLGAP5), which were considered to be hub genes. Subsequently, the present study

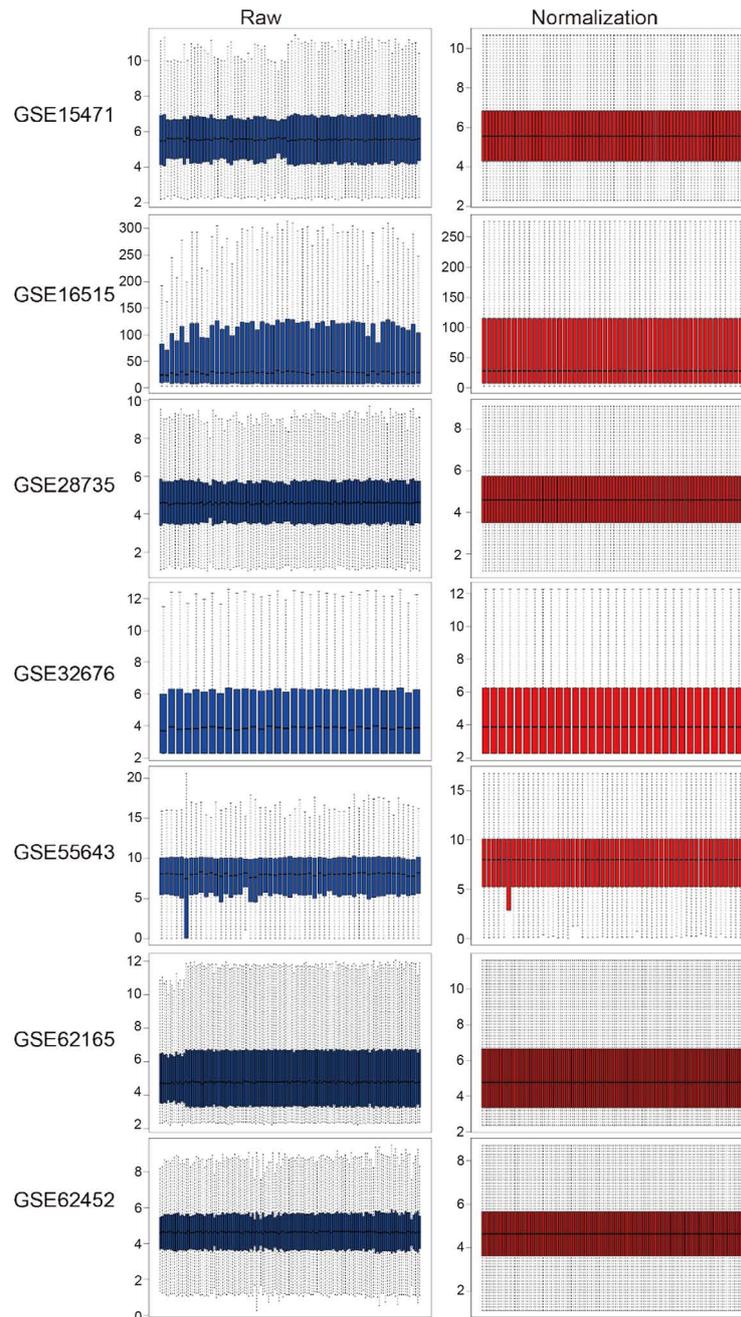


Figure 1. Normalization of raw data in the Gene Expression Omnibus database.

identified the two top-ranking modules with scores of 20.875 and 14.000 in MCODE. Module 1 contained 22 nodes and 219 edges, and module 2 contained 14 nodes and 91 edges (Fig. 5B and C). As presented in Fig. 6, compared with normal tissues, the expression levels of the 10 hub genes were significantly increased in PAAD compared with normal tissues in TCGA cohort. Of note, all the hub genes were in module 1, which suggested that module 1 may serve an important role in the PPI network. The KEGG analysis demonstrated that module 1 was mainly associated with the ‘Cell cycle’ and ‘p53 signaling pathway’ (Table SIV).

Prognostic gene signature. The present study identified 446 TCGA-DEGs from TCGA dataset, including 26 upregulated genes and 420 downregulated genes (Table SV). A total of

281 genes identified using the univariate Cox regression model were significantly associated with survival time ($P < 0.05$; Table SVI). In addition, a prognostic gene characteristic comprising six genes was detected through multivariate Cox regression analysis, containing matrix metalloproteinase 12 (MMP12), sodium voltage-gated channel α subunit 11 (SCN11A), tetraspanin 1 (TSPAN1) and SH3 domain and tetratricopeptide repeats-containing 2 (SH3TC2). Among them, MMP12, TSPAN1 and SH3TC2 with hazard ratios (HRs) > 1 were identified as risk prognostic genes, whereas SCN1A with an HR < 1 was considered as a protective prognostic gene (Table II). According to the risk score model, 86 patients were assigned into the high-risk group, and 85 patients were assigned into the low-risk group (Fig. 7A-C). The survival analysis demonstrated that the OS rate of the high-risk group was

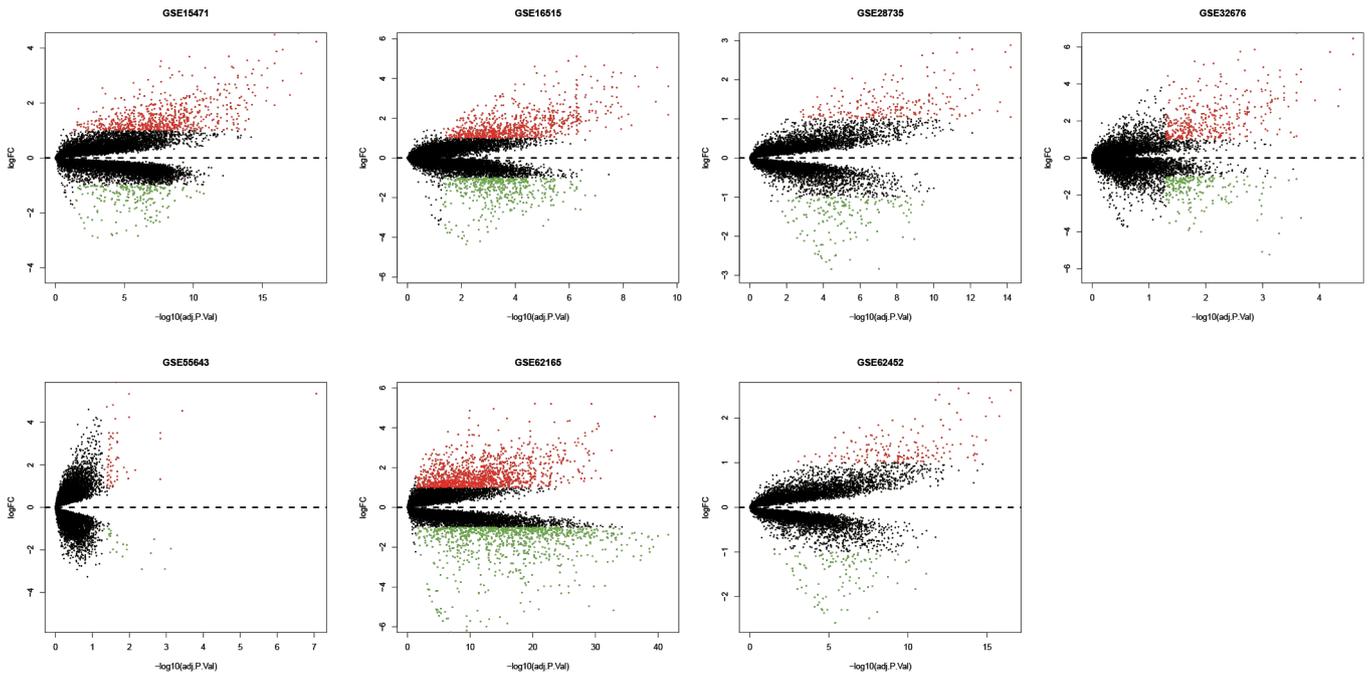


Figure 2. Volcano plots of differentially expressed genes in the Gene Expression Omnibus database. The red dots represent the upregulated genes, the green dots represent the downregulated genes and the black spots represent genes with no significant difference in expression. FC, fold-change; adj P Val, adjusted P-value.

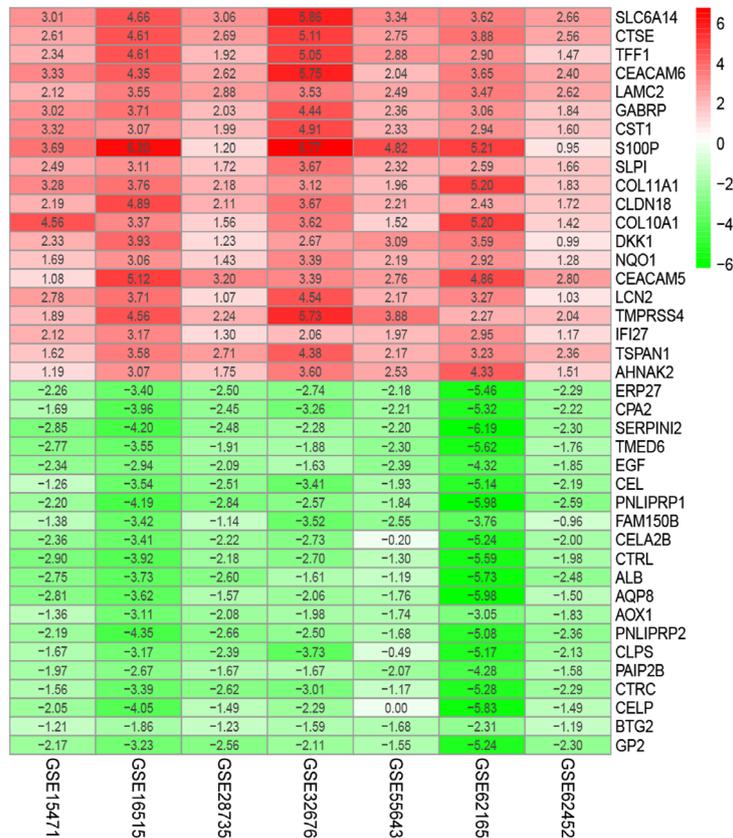


Figure 3. The heat map of top 20 down- and upregulated differentially expressed genes in the integrated microarray analysis. Each column represents one dataset and each row represents one gene. The number in each rectangle represents the value of $\log_2(FC)$. The red column represents the upregulated genes and the green column represents the downregulated genes.

significantly lower compared with that of the low-risk group ($P=6.335 \times 10^{-5}$; Fig. 7D). In addition, the 1-, 3- and 5-year OS

rates in the high-risk group were significantly lower than those in the low-risk group (Table III). Time-dependent ROC analysis

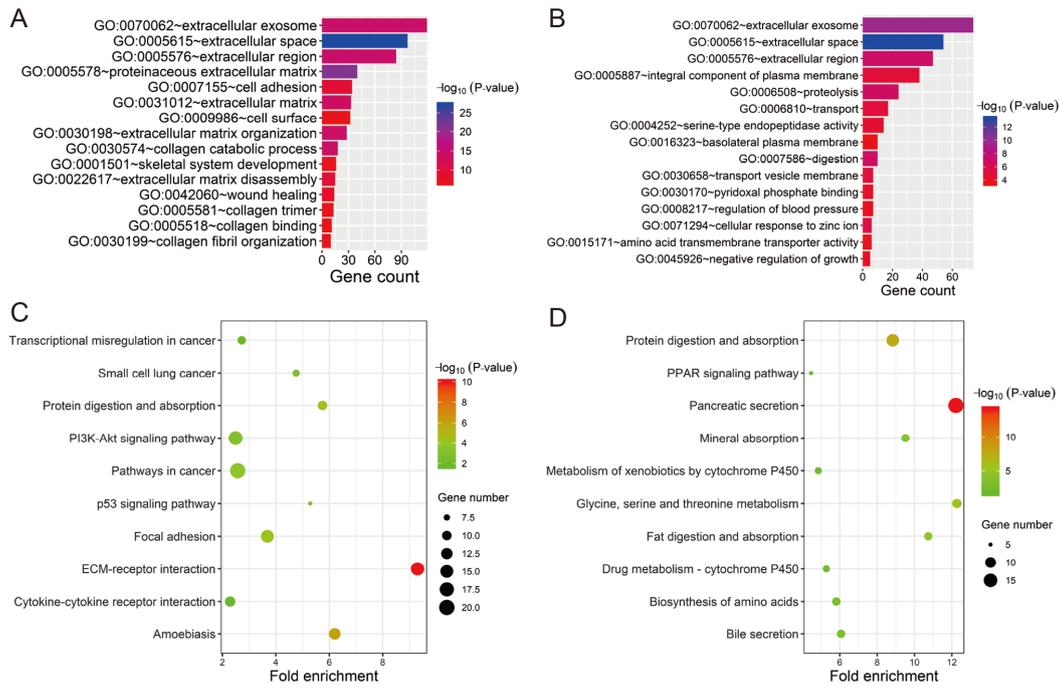


Figure 4. Functional enrichment analysis of the differentially expressed genes. (A) GO enrichment analysis of the overlapping upregulated DEGs. (B) GO enrichment analysis of the overlapping downregulated DEGs. (C) KEGG pathway enrichment analysis of the overlapping upregulated DEGs. (D) KEGG pathway enrichment analysis of the overlapping downregulated DEGs. GO, Gene Ontology; ECM, extracellular matrix; KEGG, Kyoto Encyclopaedia of Genes and Genomes; DEGs, differentially expressed genes.

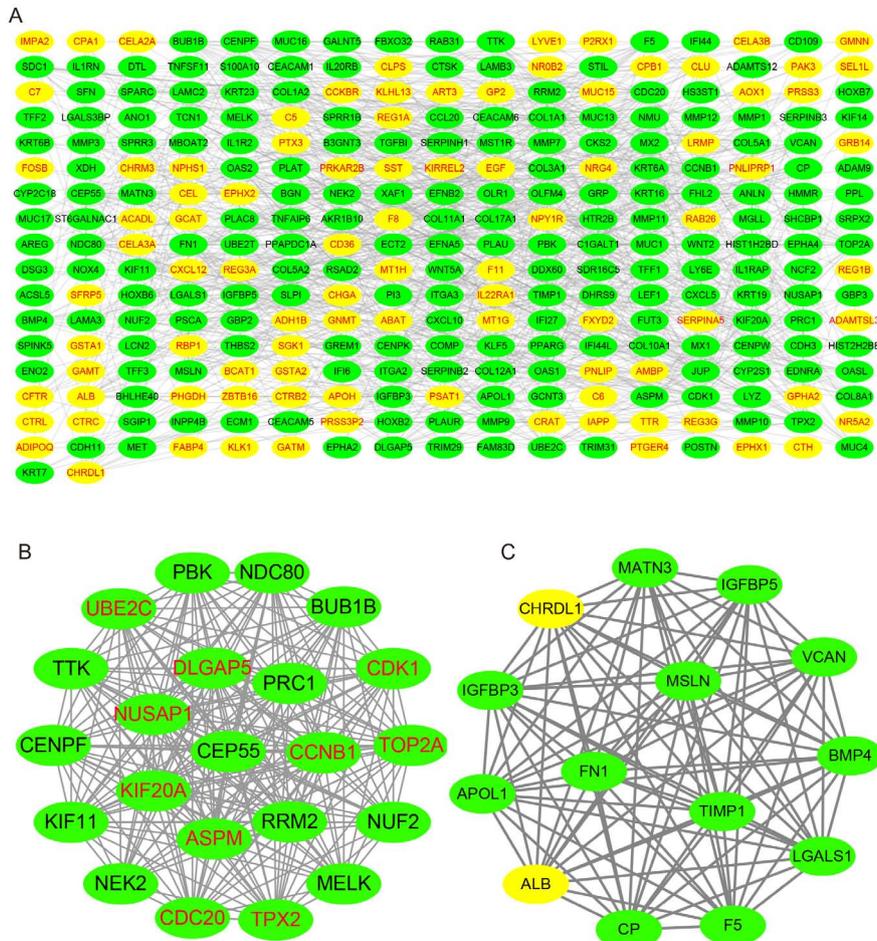


Figure 5. PPI network and the top two clustering modules. (A) PPI network of overlapping DEGs. (B) Module 1. (C) Module 2. Yellow circles represent downregulated genes and green circles represent upregulated genes. PPI, protein-protein interaction; DEGs, differentially expressed genes.

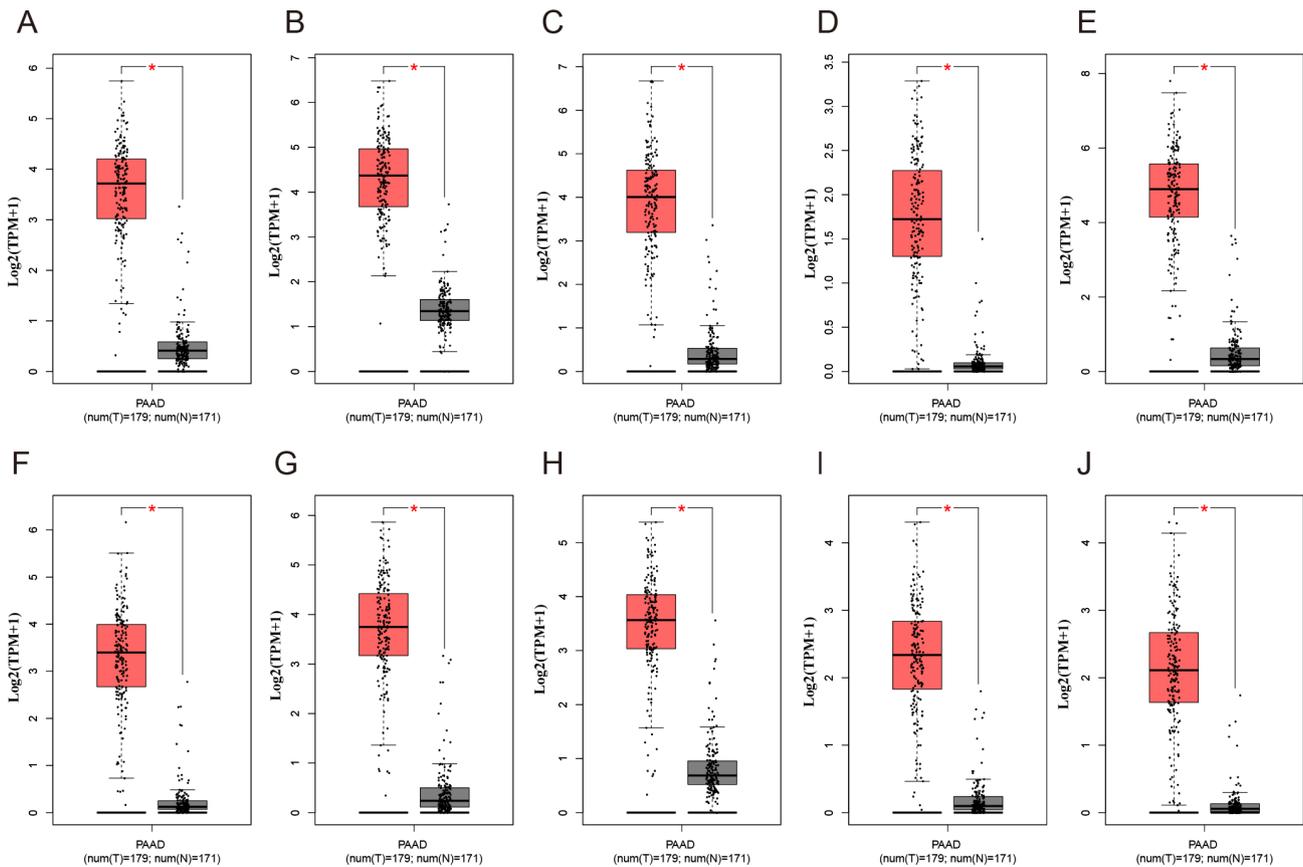


Figure 6. Expression of the ten differentially expressed hub genes in PAAD and normal pancreatic tissues from The Cancer Genome Atlas and Genotype-Tissue Expression datasets. Expression values of genes are log₂-transformed. Expression of (A) CDK1, (B) CCNB1, (C) CDC20, (D) ASPM, (E) UBE2C, (F) TPX2, (G) TOP2A, (H) NUSAP1, (I) KIF20A and (J) DLGAP5. PAAD, pancreatic adenocarcinoma; T, tumour; N, normal. The red boxplot represents pancreatic tumour tissue, and the black boxplot represents normal pancreatic tissue. *P<0.05.

based on the risk score model demonstrated good efficiency in predicting patient survival (area under the curve, 0.718; Fig. 7E).

Discussion

The present study obtained 622 GEO-DEGs (387 upregulated and 235 downregulated) from seven datasets on GEO and identified 10 hub genes from the PPI network, including CDK1, CCNB1, CDC20, ASPM, UBE2C, TPX2, TOP2A, NUSAP1, KIF20A and DLGAP5. However, as the GEO datasets did not provide survival data, these genes were not incorporated into the prognostic risk signature.

CDK1, also termed cell division control protein 2, which is highly expressed in cancer cells, exhibits a vital function in the transition from G₂ stage in mitosis (29). It is one of the potential radiosensitization targets to inhibit the cell cycle-dependent sensitization of PAAD cells, but it may also aggravate the toxicity of normal tissues; thus, the mechanism of CDK1 in cancer cells requires further study (29,30). CCNB1/CDK1-mediated phosphorylation of the mitochondrial substrate provides effective biological energy for cell G₂/M transformation and upregulates mitochondrial respiration to promote successful cell cycle progression (31). CDK1 plays an important role in cell cycle progression and antitumor activity. Its dysfunction or hyperactivity leads to cell different transformation, tumor invasion and other pathological states (31,32).

CDC20 is a cell cycle regulator that coordinates the mitotic process by promoting the orderly degradation of mitotic the anaphase-promoting complex/cyclosome substrates (33). The average level of CDC20 in PDAC was 20 times higher than that in normal pancreas and pancreatitis. The high expression of CDC20 was associated with poor differentiation, and the high expression of CDC20 significantly reduced the 5-year recurrence-free survival rate, and had the trend of shortening the total survival period (34).

ASPM is a novel Wnt and stemness regulatory factor in PAAD (35). Mechanistically, its protein subtype ASPM-iI, which colocalizes with disheveled-2 and active β-catenin as well as the stemness marker aldehyde dehydrogenase-1, is indispensable for the Wnt activity, stemness and the tumorigenicity of PAAD cells (36). Therefore, ASPM-iI staining as a novel Wnt-associated marker of cancer stemness can not only predict the outcome and survival time of patients with resected PAAD, but also may guide future targeted therapies (36).

UBE2C is involved in tumorigenesis by regulating cell cycle, apoptosis, metastasis and transcription (37). UBE2C gene knockdown downregulated the expression of vimentin, an mesenchymal marker, and up-regulated the expression of E-cadherin, an epithelial marker, to promote EMT in lung cancer cells (38). In addition, mouse experiments have demonstrated that UBE2C gene knockout can significantly inhibit tumor growth *in vivo* (39).

Table II. Prognostic values of the four genes in patients with pancreatic adenocarcinoma in The Cancer Genome Atlas cohort.

Gene symbol	Univariate analysis		Multivariate analysis		
	HR (95% CI)	P-value	HR (95% CI)	P-value	Coefficient
MMP12	1.180 (1.062-1.310)	0.00199	1.134 (1.014-1.270)	0.028	0.126
SCN11A	0.703 (0.586-0.845)	0.00017	0.776 (0.638-0.945)	0.011	-0.253
TSPAN1	1.346 (1.140-1.589)	0.00046	1.189 (1.002-1.412)	0.048	0.173
SH3TC2	1.299 (1.106-1.524)	0.00137	1.177 (0.981-1.411)	0.079	0.163

HR, hazard ratio; CI, confidence interval; MMP12, matrix metalloproteinase 12; SCN11A, sodium voltage-gated channel α subunit 11; TSPAN1, tetraspanin 1; SH3TC2, SH3 domain and tetratricopeptide repeats-containing 2.

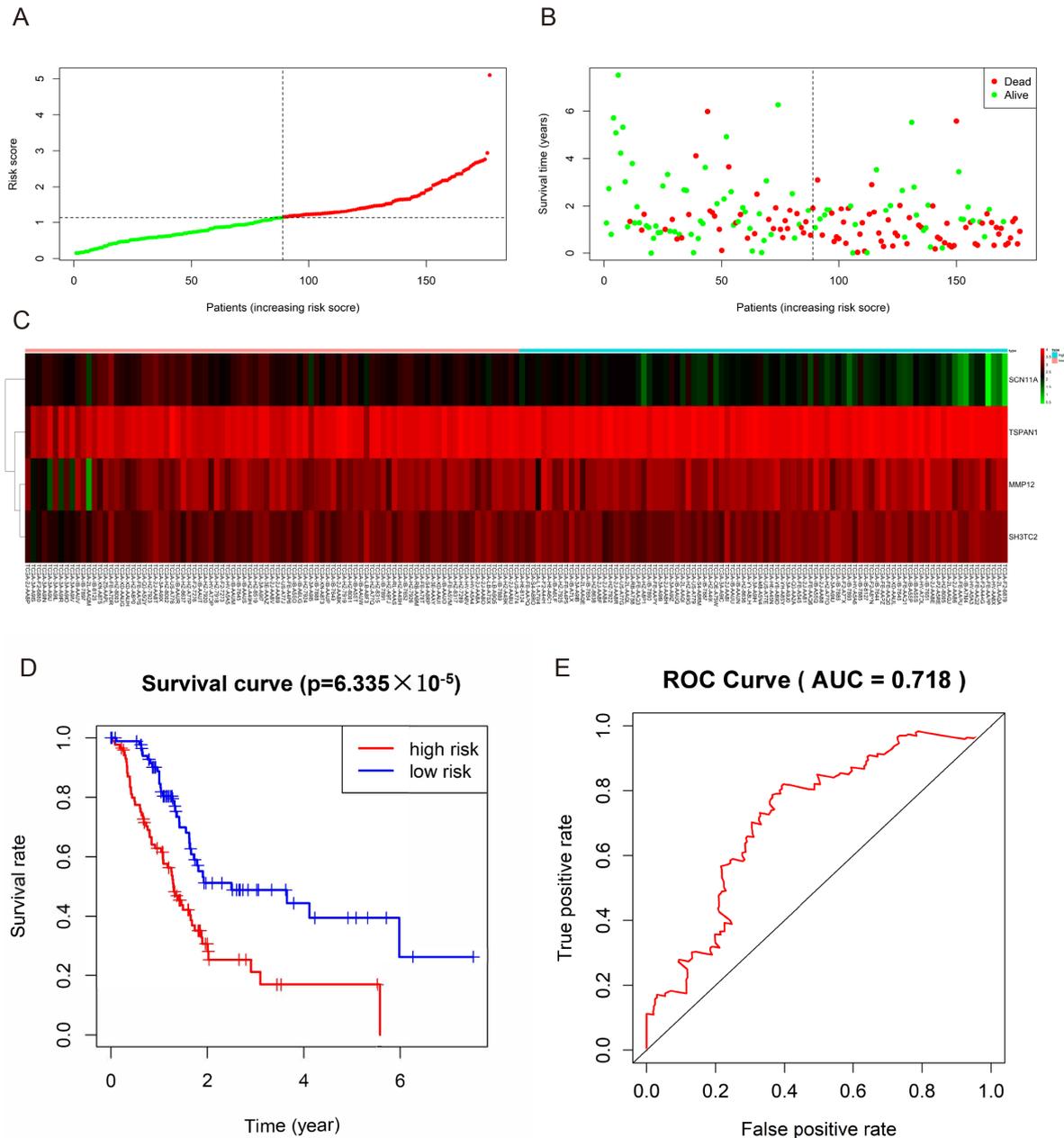


Figure 7. Prognostic gene signature of the four genes in patients with pancreatic adenocarcinoma from The Cancer Genome Atlas dataset. (A) distribution of risk scores in low-risk and high-risk groups, (B) Survival status distribution, (C) The heatmap of the four genes for low- and high-risk group, (D) Kaplan-Meier curves for low- and high-risk groups. (E) ROC curve of OS in PAAD patients was predicted according to risk score. ROC, receiver operating characteristic; AUC, area under the curve; OS, overall survival.

Table III. The OS rates in the high-risk group and low-risk group.

Years	OS (95% CI)	
	High-risk group (%)	Low-risk group (%)
1	62.9 (53.3-74.2)	88.7 (82.0-96.0)
3	21.1 (11.8-37.7)	48.9 (37.4-63.8)
5	16.9 (8.2-35.0)	39.5 (26.6-58.9)

OS, overall survival; CI, confidence interval.

A previous study has demonstrated that TPX2 expression in pancreatic tumors is higher than that in their normal counterparts (40). Targeting TPX2 using small interfering (si)RNAs can effectively inhibit the proliferation of pancreatic cancer cells in tissue culture, induce apoptosis, and inhibit the growth of pancreatic tumors in nude mice (41). TPX2 gene knockout also increases the sensitivity of pancreatic cancer cells to paclitaxel therapy (41).

A high amplification rate of TOP2A is present in multiple different types of malignancy, including PAAD (42). As a co-activator of β -catenin, TOP2A can activate the EMT process and directly target microRNA-139 to drive the malignant progress of pancreatic cancer (43).

Nucleolar and spindle associated protein 1(NUSAP1) is located in dynamic spindle microtubules at metaphase and anaphase of mitosis with a unique chromosomal central pattern. It interacts with SUMO E3 ligase complex in the process of chromosome separation near overlapping microtubules (44). NUSAP1 siRNA (L-004754-00) inhibited the drug resistance of human prostate cancer cells induced by NUSAP1, which suggested that NUSAP1 may be used as a biomarker of the antitumor activity of galiellalactone (45); however, to the best of our knowledge, there are currently few studies that focus on NUSAP1 in PAAD.

KIF20A, which is highly expressed in pancreatic cancer and other malignant tumors, but not expressed in non-cancerous tissues, is a tumor-associated antigen and a potential target for tumor immunotherapy (46,47). It has been reported that kif20a-66 is well-tolerated as an immunotherapy for advanced pancreatic cancer and can effectively induce cytotoxic T lymphocytes (48).

DLGAP5 has the unique function of stabilizing spindle formation and surviving microtubule attack caused by docetaxel in androgen-regulated prostate cancer cell (LNCaP) cycle system (49). Thus, DLGAP5 may be involved in spindle stability in other malignant tumors.

KEGG analysis in the present study revealed that these 10 genes were mainly associated with the 'Cell cycle' and 'p53 signaling pathway'. P53-protein tyrosine phosphatase non-receptor type 14 (PTPN14)-Yap pathway is a tumor suppressor mechanism mediated by p53. The p53 transcription activation domain 2 mutant is a 'super tumor suppressor', with an enhanced ability to restrain pancreatic cancer cell proliferation and to transactivate select p53 target genes (including *ptpn14*) (50). With the successive inactivation of tumor suppressor genes, the proliferation of tumor cells is increased, and remains at high levels in metastatic tumors, which may

be caused by cell cycle regulatory gene variants. This leads to cell cycle disorder, which is characteristic of various cancer subtypes (51). Various target genes affect the cycle, senescence and apoptosis of pancreatic cancer cells through the p53 pathway (52,53), which is consistent with the results of the present study in the biological function analysis.

The aforementioned results indicated that these 10 hub genes may serve a role in PAAD development. Focusing on these 10 hub genes may provide ideas and directions for revealing the molecular mechanism of pancreatic cancer and developing the corresponding therapeutic drugs.

The present study also identified four genes associated with PAAD prognosis (MMP12, SCN11A, TSPAN1 and SH3TC2), which were used to construct a prognostic gene signature. It has been reported that SCN11A is upregulated in breast and prostate cancer compared with adjacent normal tissues (54,55), but its molecular nature and association with pancreatic cancer function have remained elusive. In the present study, SCN1A was considered to be a protective prognostic gene; thus, its role in pancreatic cancer requires further investigation. The other three prognostic genes MMP12, TSPAN1 and SH3TC2 were considered to be risk prognostic genes, implying malignant phenotypes. MMPs are associated with the invasion and metastasis of tumor cells. It promotes tumor cells to degrade the components of the extracellular matrix, separate from the primary site, migrate to the distal site and invade the surrounding tissue to induce metastasis (56). When MMP12, a member of MMP family, is inhibited, the invasion and metastasis of human pancreatic cancer is inhibited, which prolongs the survival period and exhibits antimetastatic effects *in situ* in a mouse model (57). However, to the best of our knowledge, the physiological function of MMP12 has not yet been described completely.

A previous study demonstrated that the increased TSPAN1 in pancreatic cancer tissues was associated with the clinicopathological features and survival rate of patients with PAAD (52). siRNA targeting TSPAN1 significantly inhibits the proliferation of PAAD cells, increases apoptosis, and decreases cell migration and invasion, therefore this may be a potential strategy for the treatment of human PAAD (58,59). Previous studies have suggested that an SH3TC2 variant allele is associated with the cause of Charcot-Marie-Tooth neuropathy (60,61). Although the expression and functions of SH3TC2 in cancer have rarely been described, it remains reasonable to identify it as a prognostic biomarker due to its significance in the present signature model.

The results of the present study demonstrated that 10 hub genes may be involved in the occurrence and progression of PAAD, and these 10 hub genes were highly expressed in PAAD tissues. Therefore, further research may focus on the reasons for the high expression of these genes to develop corresponding drugs to reduce or inhibit the expression of these genes, which may improve the treatment of PAAD. In addition, according to the survival information of patients with PAAD in TCGA database, four prognostic signature genes were identified. By detecting the expression of these four genes, the risk of PAAD could be predicted in advance. These results may provide clues for further investigating the pathogenesis of PAAD and to establish a new risk classification and prognosis assessment model. However, there are limitations to the present study, as it was performed based on data analysis, experimental results

of this prediction in PAAD are required. As the present study identified hub genes in a public database, further experimental research is required to demonstrate the molecular pathogenesis and signal transduction mechanism of these genes in PAAD.

In conclusion, using the datasets of multi-gene expression profiles and the comprehensive bioinformatics analysis, the present study identified 10 hub genes that may be responsible for the pathogenesis of PAAD. In addition, the present study also constructed a four-gene prediction model that performed well in predicting 1-, 3- and 5-year OS, and thus may be used as a prognosis marker for patients with PAAD. The present study may be helpful to improve the current understanding of the potential carcinogenesis or progress of PAAD, as well as for the prognostic prediction and molecular targeted management of PAAD.

Acknowledgements

None.

Funding

This work was supported by the National Natural Science Foundation of China (grant no. 81873190).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Authors' contributions

LES, XS and ZZZ advance the research direction and method. LES, XS, QX and NBC conceived and designed the study. LES, KCN and XS drafted the manuscript, analyzed the data, developed the algorithm and interpreted the results. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D and Bray F: Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 136: E359-E386, 2015.
2. Gordon-Dseagu VL, Devesa SS, Goggins M and Stolzenberg-Solomon R: Pancreatic cancer incidence trends: Evidence from the surveillance, epidemiology and end results (SEER) population-based data. *Int J Epidemiol* 47: 427-439, 2018.
3. Golan T, Sella T, Margalit O, Amit U, Halpern N, Aderka D, Shacham-Shmueli E, Urban D and Lawrence YR: Short- and long-term survival in metastatic pancreatic adenocarcinoma, 1993-2013. *J Natl Compr Canc Netw* 15: 1022-1027, 2017.
4. Mackay MT, van Erning FN, van der Geest LG, Koerkamp BG, van Laarhoven MH, Bonsing BA, Wilmink JW, van Santvoort HC, de Vos-Geelen Jd, van Eijck CH, *et al*: Association of the location of pancreatic ductal adenocarcinoma (head, body, tail) with tumor stage, treatment, and survival: A population-based analysis. *Pancreatology* 18 (Suppl): S132, 2018.
5. Latenstein AEJ, van der Geest LGM, Bonsing BA, Groot Koerkamp B, Haj Mohammad N, de Hingh IHJT, de Meijer VE, Molenaar IQ, van Santvoort HC, van Tienhoven G, *et al*: Nationwide trends in incidence, treatment and survival of pancreatic ductal adenocarcinoma. *Eur J Cancer* 125: 83-93, 2020.
6. Mokdad AA, Minter RM, Yopp AC, Porembka MR, Wang SC, Zhu H, Augustine MM, Mansour JC, Choti MA and Polanco PM: Comparison of overall survival between preoperative chemotherapy and chemoradiotherapy for resectable pancreatic adenocarcinoma. *J Natl Compr Canc Netw* 16: 1468-1475, 2018.
7. Moore MJ and Stathis A: Advanced pancreatic carcinoma: Current treatment and future challenges. *Nat Rev Clin Oncol* 7: 163-172, 2010.
8. Jun I, Park HS, Piao H, Han JW, An MJ, Yun BG, Zhang X, Cha YH, Shin YK, Yook JI, *et al*: ANO9/TMEM16J promotes tumorigenesis via EGFR and is a novel therapeutic target for pancreatic cancer. *Br J Cancer* 117: 1798-1809, 2017.
9. Chio IIC, Jafarnejad SM, Ponz-Sarvise M, Park Y, Rivera K, Palm W, Wilson J, Sangar V, Hao Y, Öhlund D, *et al*: NRF2 promotes tumor maintenance by modulating mrna translation in pancreatic cancer. *Cell* 166: 963-976, 2016.
10. Jain A, Brown SZ, Thomsett HL, Londin E and Brody JR: Evaluation of post-transcriptional gene regulation in pancreatic cancer cells: Studying RNA binding proteins and their mRNA targets. *Methods Mol Biol* 1882: 239-252, 2019.
11. Uchida S, Kinoh H, Ishii T, Matsui A, Tockary TA, Takeda KM, Uchida H, Osada K, Itaka K and Kataoka K: Systemic delivery of messenger RNA for the treatment of pancreatic cancer using polyplex nanomicelles with a cholesterol moiety. *Biomaterials* 82: 221-228, 2016.
12. Hutter C and Zenklusen JC: The cancer genome atlas: Creating lasting value beyond its data. *Cell* 173: 283-285, 2018.
13. Ho J, Li X, Zhang L, Liang Y, Hu W, Yau JC, Chan H, Gin T, Chan MT, Tse G and Wu WK: Translational genomics in pancreatic ductal adenocarcinoma: A review with re-analysis of TCGA dataset. *Semin Cancer Biol* 55: 70-77, 2019.
14. Barrett T: NCBI GEO: Mining millions of expression profiles-database and tools. *Nucleic Acids Res* 33: D562-D566, 2004.
15. Li C, Zeng X, Yu H, Gu Y and Zhang W: Identification of hub genes with diagnostic values in pancreatic cancer by bioinformatics analyses and supervised learning methods. *World J Surg Oncol* 16: 223, 2018.
16. Tang H, Wei P, Chang P, Li Y, Yan D, Liu C, Hassan M and Li D: Genetic polymorphisms associated with pancreatic cancer survival: A genome-wide association study. *Int J Cancer* 141: 678-686, 2017.
17. Sivakumar S, de Santiago I, Chlon L and Markowitz F: Master regulators of oncogenic KRAS response in pancreatic cancer: An integrative network biology analysis. *PLoS Med* 14: e1002223, 2017.
18. Muzumdar MD, Chen PY, Dorans KJ, Chung KM, Bhutkar A, Hong E, Noll EM, Sprick MR, Trumpp A and Jacks T: Survival of pancreatic cancer cells lacking KRAS function. *Nat Commun* 8: 1090, 2017.
19. Wolfgang CL, Herman JM, Laheru DA, Klein AP, Erdek MA, Fishman EK and Hruban RH: Recent progress in pancreatic cancer. *CA Cancer J Clin* 63: 318-348, 2013.
20. Badea L, Herlea V, Dima SO, Dumitrascu T and Popescu I: Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology* 55: 2016-2027, 2008.
21. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, Petersen G, Lou Z and Wang L: FKBP51 affects cancer cell response to chemotherapy by negatively regulating akt. *Cancer Cell* 16: 259-266, 2009.
22. Zhang G, Schetter A, He P, Funamizu N, Gaedcke J, Ghadimi BM, Ried T, Hassan R, Yfantis HG, Lee DH, *et al*: DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One* 7: e31507, 2012.
23. Donahue TR, Tran LM, Hill R, Li Y, Kovoichich A, Calvopina JH, Patel SG, Wu N, Hindoyan A, Farrell JJ, *et al*: Integrative survival-based molecular profiling of human pancreatic cancer. *Clin Cancer Res* 18: 1352-1363, 2012.

24. Lunardi S, Jamieson NB, Lim SY, Griffiths KL, Carvalho-Gaspar M, Al-Assar O, Yameen S, Carter KC, McKay CJ, Spoletini G, *et al*: IP-10/CXCL10 induction in human pancreatic cancer stroma influences lymphocytes recruitment and correlates with poor survival. *Oncotarget* 5: 11064-11080, 2014.
25. Janky R, Binda MM, Allemeersch J, Van den Broeck A, Govaere O, Swinnen JV, Roskams T, Aerts S and Topal B: Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. *BMC Cancer* 16: 632, 2016.
26. Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, Yanaga K, Uwagawa T, Satoskar AR, Gaedcke J, *et al*: A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. *Cancer Res* 76: 3838-3850, 2016.
27. Nie K, Shi L, Wen Y, Pan J, Li P, Zheng Z and Liu F: Identification of hub genes correlated with the pathogenesis and prognosis of gastric cancer via bioinformatics methods. *Minerva Med* 111: 213-225, 2020.
28. Liu L, Lin J and He H: Identification of potential crucial genes associated with the pathogenesis and prognosis of endometrial cancer. *Front Genet* 10: 373, 2019.
29. Prevo R, Pirovano G, Puliyadi R, Herbert KJ, Rodriguez-Berriguete G, O'Docherty A, Greaves W, McKenna WG and Higgins GS: CDK1 inhibition sensitizes normal cells to DNA damage in a cell cycle dependent manner. *Cell Cycle* 17: 1513-1523, 2018.
30. Wei D, Parsels LA, Karnak D, Davis MA, Parsels JD, Marsh AC, Zhao L, Maybaum J, Lawrence TS, Sun Y and Morgan MA: Inhibition of protein phosphatase 2A radiosensitizes pancreatic cancers by modulating CDC25C/CDK1 and homologous recombination repair. *Clin Cancer Res* 19: 4422-4432, 2013.
31. Wang Z, Fan M, Candas D, Zhang TQ, Qin L, Eldridge A, Wachsmann-Hogiu S, Ahmed KM, Chromy BA, Nantajit D, *et al*: Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. *Dev Cell* 29: 217-232, 2014.
32. Levasseur MD, Thomas C, Davies OR, Higgins JM and Madgwick S: Aneuploidy in oocytes is prevented by sustained CDK1 activity through degron masking in cyclin B1. *Dev Cell* 48: 672-684, 2019.
33. Yu H: Cdc20: A WD40 activator for a cell cycle degradation machine. *Mol Cell* 27: 3-16, 2007.
34. Chang DZ, Ma Y, Ji B, Liu Y, Hwu P, Abbruzzese JL, Logsdon C and Wang H: Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression. *J Hematol Oncol* 5: 15, 2012.
35. Wang WY, Hsu CC, Wang TY, Li CR, Hou YC, Chu JM, Lee CT, Liu MS, Su JJ, Jian KY, *et al*: A gene expression signature of epithelial tubulogenesis and a role for ASPM in pancreatic tumor progression. *Gastroenterology* 145: 1110-1120, 2013.
36. Hsu CC, Liao WY, Chan TS, Chen WY, Lee CT, Shan YS, Huang PJ, Hou YC, Li CR and Tsai KK: The differential distributions of ASPM isoforms and their roles in Wnt signaling, cell cycle progression, and pancreatic cancer prognosis. *J Pathol* 249: 498-508, 2019.
37. Liu G, Zhao J, Pan B, Ma G and Liu L: UBE2C overexpression in melanoma and its essential role in G2/M transition. *J Cancer* 10: 2176-2184, 2019.
38. Jin D, Guo J, Wu Y, Du J, Wang X, An J, Hu B, Kong L, Di W and Wang W: UBE2C, directly targeted by miR-548e-5p, increases the cellular growth and invasive abilities of cancer cells interacting with the EMT marker protein zinc finger E-box binding homeobox 1/2 in NSCLC. *Theranostics* 9: 2036-2055, 2019.
39. Wang X, Yin L, Yang L, Zheng Y, Liu S, Yang J, Cui H and Wang H: Silencing ubiquitin-conjugating enzyme 2C inhibits proliferation and epithelial-mesenchymal transition in pancreatic ductal adenocarcinoma. *FEBS J* 286: 4889-4909, 2019.
40. Gomes-Filho SM, Dos Santos EO, Bertoldi ER, Scalabrini LC, Heidrich V, Dazzani B, Levantini E, Reis EM and Basseres DS: Aurora A kinase and its activator TPX2 are potential therapeutic targets in KRAS-induced pancreatic cancer. *Cell Oncol (Dordr)* 43: 445-460, 2020.
41. Warner SL, Stephens BJ, Nwokenkwo S, Hostetter G, Sugeng A, Hidalgo M, Trent JM, Han H and Von Hoff DD: Validation of TPX2 as a potential therapeutic target in pancreatic cancer cells. *Clin Cancer Res* 15: 6519-6528, 2009.
42. Heestand GM, Schwaederle M, Gatalica Z, Arguello D and Kurzrock R: Topoisomerase expression and amplification in solid tumours: Analysis of 24,262 patients. *Eur J Cancer* 83: 80-87, 2017.
43. Pei YF, Yin XM and Liu XQ: TOP2A induces malignant character of pancreatic cancer through activating β -catenin signaling pathway. *Biochim Biophys Acta Mol Basis Dis* 1864: 197-207, 2018.
44. Mills CA, Suzuki A, Arceci A, Mo JY, Duncan A, Salmon ED and Emanuele MJ: Nucleolar and spindle-associated protein 1 (NUSAP1) interacts with a SUMO E3 ligase complex during chromosome segregation. *J Biol Chem* 292: 17178-17189, 2017.
45. Garrido-Rodríguez M, Ortea I, Calzado MA, Muñoz E and García V: SWATH proteomic profiling of prostate cancer cells identifies NUSAP1 as a potential molecular target for galiellalactone. *J Proteomics* 193: 217-229, 2019.
46. Taniuchi K, Furihata M and Saibara T: KIF20A-mediated RNA granule transport system promotes the invasiveness of pancreatic cancer cells. *Neoplasia* 16: 1082-1093, 2014.
47. Imai K, Hirata S, Irie A, Senju S, Ikuta Y, Yokomine K, Harao M, Inoue M, Tomita Y, Tsunoda T, *et al*: Identification of HLA-A2-restricted CTL epitopes of a novel tumour-associated antigen, KIF20A, overexpressed in pancreatic cancer. *Br J Cancer* 104: 300-307, 2011.
48. Asahara S, Takeda K, Yamao K, Maguchi H and Yamaue H: Phase I/II clinical trial using HLA-A24-restricted peptide vaccine derived from KIF20A for patients with advanced pancreatic cancer. *J Transl Med* 11: 291, 2013.
49. Hewit K, Sandilands E, Martinez RS, James D, Leung HY, Bryant DM, Shanks E and Markert EK: A functional genomics screen reveals a strong synergistic effect between docetaxel and the mitotic gene DLGAP5 that is mediated by the androgen receptor. *Cell Death Dis* 9: 1069, 2018.
50. Mello SS, Valente LJ, Raj N, Seoane JA, Flowers BM, McClendon J, Biegging-Rolett KT, Lee J, Ivanochko D, Kozak MM, *et al*: A p53 super-tumor suppressor reveals a tumor suppressive p53-Ptpn14-yap axis in pancreatic cancer. *Cancer Cell* 32: 460-473, 2017.
51. Connor AA, Denroche RE, Jang GH, Lemire M, Zhang A, Chan-Seng-Yue M, Wilson G, Grant RC, Merico D, Lungu I, *et al*: Integration of genomic and transcriptional features in pancreatic cancer reveals increased cell cycle progression in metastases. *Cancer Cell* 35: 267-282, 2019.
52. Jiang W, Zhao S, Jiang X, Zhang E, Hu G, Hu B, Zheng P, Xiao J, Lu Z, Lu Y, *et al*: The circadian clock gene Bmal1 acts as a potential anti-oncogene in pancreatic cancer by activating the p53 tumor suppressor pathway. *Cancer Lett* 371: 314-325, 2016.
53. Zhang H, Zhang X, Li X, Meng WB, Bai ZT, Rui SZ, Wang ZF, Zhou WC and Jin XD: Effect of CCNB1 silencing on cell cycle, senescence, and apoptosis through the p53 signaling pathway in pancreatic cancer. *J Cell Physiol* 234: 619-631, 2019.
54. Fraser SP, Diss JK, Chioni A, Mycielska ME, Pan H, Yamaci RF, Pani F, Siwy Z, Krasowska M, Grzywna Z, *et al*: Voltage-Gated sodium channel expression and potentiation of human breast cancer metastasis. *Clin Cancer Res* 11: 5381-5389, 2005.
55. Diss JK, Archer SN, Hirano J, Fraser SP and Djamgoz MB: Expression profiles of voltage-gated Na(+) channel alpha-subunit genes in rat and human prostate cancer cell lines. *Prostate* 48: 165-178, 2001.
56. Stellas D and Patsavoudi E: Inhibiting matrix metalloproteinases, an old story with new potentials for cancer treatment. *Anticancer Agents Med Chem* 12: 707-717, 2012.
57. Fujisawa T, Rubin B, Suzuki A, Patel PS, Gahl WA, Joshi BH and Puri RK: Cysteamine suppresses invasion, metastasis and prolongs survival by inhibiting matrix metalloproteinases in a mouse model of human pancreatic cancer. *PLoS One* 7: e34437, 2012.
58. Tian J, Zhang R, Piao H, Li X, Sheng W, Zhou J, Dong M, Zhang X, Yan X, Shang W, *et al*: Silencing Tspan1 inhibits migration and invasion, and induces the apoptosis of human pancreatic cancer cells. *Mol Med Rep* 18: 3280-3288, 2018.
59. Hou FQ, Lei XF, Yao JL, Wang YJ and Zhang W: Tetraspanin 1 is involved in survival, proliferation and carcinogenesis of pancreatic cancer. *Oncol Rep* 34: 3068-3076, 2015.
60. Lupski JR, Gonzaga-Jauregui C, Yang Y, Bainbridge MN, Jhangiani S, Buhay CJ, Kovar CL, Wang M, Hawes AC, Reid JG, *et al*: Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing charcot-marie-tooth neuropathy. *Genome Med* 5: 57, 2013.
61. Stendel C, Roos A, Kleine H, Arnaud E, Özçelik M, Sidiropoulos PN, Zenker J, Schüpfer F, Lehmann U, Sobota RM, *et al*: SH3TC2, a protein mutant in charcot-marie-tooth neuropathy, links peripheral nerve myelination to endosomal recycling. *Brain* 133: 2462-2474, 2010.

