RNA-Sequencing for transcriptional profiling of whole blood in early-stage and metastatic pancreatic cancer patients

Sima Kalantari  
Shahid Beheshti University of Medical Sciences

Bahram Kazemi  
Shahid Beheshti University of Medical Sciences

Raheleh Roudi  
University of Minnesota Medical School

Hakimeh Zali  
Shahid Beheshti University of Medical Science

Alberto D'Angelo  
University of Bath

Ashraf Mohamadkhani  
Tehran University of Medical Sciences

Zahra Madjd  
Iran University of Medical Sciences

Akram Pourshams (akrampourshams@gmail.com)  
Tehran University of Medical Sciences

Research Article

Keywords: pancreatic cancer, early stage, metastasis, RNA-sequencing, blood profiling

Posted Date: April 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1574223/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

We investigated the transcriptional profile of whole blood in early and metastatic stages of PaC patients to identify potential diagnostic factors for early diagnosis. Blood samples from 18 participants (6 healthy individuals, 6 patients in early stage (I/II) PaC and 6 patients in metastatic PaC) were analyzed by RNA-sequencing. The expression levels of identified genes were subsequently compared with their expression in pancreatic tumor tissues based on TCGA data reported in UALCAN and GEPIA2 databases. Overall, 331 and 724 genes were identified as differentially expressed genes in early and metastatic stages, respectively. Of these, 146 genes were shared by early and metastatic stages. Upregulation of PTCD3 and UBA52 genes and downregulation of A2M and ARID1B genes in PaC patients were observed from early stage to metastasis. TCGA database showed increasing trend in expression levels of these genes from stage I to IV in pancreatic tumor tissue. Finally, we found that low expression of PTCD3, A2M, and ARID1B genes and high expression of UBA52 gene were positively correlated with PaC patients survival. We identified a four-gene set (PTCD3, UBA52, A2M, and ARID1B) expressed in peripheral blood of early stage and metastatic PaC patients that may be useful for PaC early diagnosis.

Introduction

Pancreatic cancer (PaC), a highly fatal malignancy with a 5-year survival rate of only 7–8%, is the seventh leading cause of global cancer death (1). The high mortality rate of PaC is generally due to a delayed diagnosis at an advanced stage (2) when approximately 80% of patients have local spread or metastatic disease (3). The stomach, spleen, colon, and left adrenal gland are potential sites of extra-pancreatic local disease invasion. In more advanced stages of PaC, liver, lymph nodes, and peritoneum are common sites of distant metastases (4). There are different types of PaC although the term “pancreatic cancer” is usually referred to pancreatic adenocarcinoma. Indeed, more than 95% of pancreas malignant neoplasms involves the exocrine portion of the pancreas (duct system and acinar cells) with adenocarcinoma features (5).

Surgery, chemotherapy, and radiation therapy are treatment options for PaC patients with surgery the only potentially curative treatment (3). Based on cancer stage, PaC patients are usually classified into four distinct groups (stages I-IV), which determine their treatment options. Patients in stage I have resectable tumors while stage II includes patients with “borderline resectable” tumors. In stage III, patients have locally advanced unresectable disease, with no metastasis. The final stage (stage IV) includes patients with metastatic disease, who are not eligible for surgical resection (6, 7). Unfortunately, the absence of specific symptoms in early stage of disease and the ability of pancreatic cancer to early metastasize limit the effectiveness of surgical resection, with only 15–20% of patients eligible for curative pancreatectomy at the time of the diagnosis (3).

Early diagnosis of PaC within the potentially resectable tumor window would represent a significant increase of patients’ survival and a decrease in the mortality rate. Challenges to the early PaC diagnosis include the high cost of imaging investigations - namely endoscopic ultrasound and cross-sectional
imaging - and the lack of specific biomarkers. Currently, CA19-9 (carbohydrate antigen 19–9) is the only blood-based biomarker for pancreatic cancer diagnosis with 75% sensitivity and 90% specificity (8) which, due to the high rate of false-positive results (high values of CA19-9 in the absence of malignancy) (9) it is not widely considered as an accurate test for PaC early diagnosis.

Previously, in a prospective study on 461 PaC cases in the Iranian population, we analyzed the clinical, pathological as well as therapeutic and survival features of these patients to provide reliable information of Iranian PaC patients. We observed that among 461 studied cases of PaC only 43 (9.5%) individuals were classified as stage I (with resectable tumor) at the time of diagnosis, and the 5-year survival of Iranian PaC patients was lower than the developed countries. We concluded that socioeconomic aspects alongside the stage and pathologic features of PaC may affect the probability of receiving timely treatment and, consequently, patient survival (10). In another study, published in 2021, we focused on the potential association between diabetes mellitus (DM) and PaC, to identify common molecular factors between these two conditions with the ultimate goal of PaC early diagnosis. In that study we proposed the potential role of SPI1 and YY1 genes to investigate the relationship between DM and PaC and provide opportunities for PaC screening using blood tests (11). As further perspective of our investigation on pancreatic cancer, in the current project we focused on gene expression profiling of blood samples for early stage and metastatic PaC patients. The main goal of the study was the identification of differentially expressed genes (DEGs) in PaC patients compared to normal individuals with an emphasis on those shared DEGs between early and metastatic stages, to evaluate what molecular factors are suitable for the development of a blood test that could detect PaC in its early and therefore resectable stage.

**Results**

**Demographics**

The whole blood samples collected from 18 participants (12 patients with histologically confirmed PaC and 6 healthy individuals) were divided into three groups as follows: early stage (I/II) PaC, metastatic PaC and normal control. For each group, 6 blood samples were pooled and RNA-sequencing was performed on these pooled samples. The demographic information and clinical characteristics for each group are detailed in Table 1.
Table 1
Demographic and clinical information of patients and healthy individuals

<table>
<thead>
<tr>
<th>Groups</th>
<th>Gender (Male/Female)</th>
<th>Age range (years)</th>
<th>Tumor stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy control</td>
<td>3/3</td>
<td>51–79 (Median: 64)</td>
<td>—</td>
</tr>
<tr>
<td>Pancreatic cancer (early stage)</td>
<td>3/3</td>
<td>56–77 (Median: 67.6)</td>
<td>I / II</td>
</tr>
<tr>
<td>Pancreatic cancer (metastasis)</td>
<td>3/3</td>
<td>50–65 (Median: 56.3)</td>
<td>IV</td>
</tr>
</tbody>
</table>

**RNA-sequencing analysis**

The RNA-sequencing data analysis and reference genome mapping for DEGs of each group was carried out. Compared to normal control, the number of DEGs identified in early stage and metastatic PaC patients were 13,309 and 19,030, respectively. After the exclusion of those DEGs with not significant logFcs (~expression values near to zero), 331 genes of the early stage group and 724 genes of the metastatic group were considered statistically significant DEGs based on logFcs and P-value parameters. Tables with genes ID, logFcs and P-values for each experimental group are reported in Supplementary File S1 and S2 online. As we aimed to identify the shared DEGs between early stage and metastatic PaC patients, using the Venn diagram tool we identified 146 genes shared between these two experimental groups. The result is shown in Fig. 1. These 146 genes are categorized as up and down regulated genes in Supplementary File S3 online. Considering logFcs ≥ ±1.5 and P-value ≤ 0.05 (as the cut-off criteria), we identified *PTCD3*, *A2M*, *UBA52*, and *ARID1B* as statistically significant shared DEGs between early and metastatic stages. Hence, we performed further bioinformatics analysis on these identified genes.

**Pan-cancer view and survival map of Shared DEGs**

Using GEPIA2 dataset analysis, we compared the transcription levels of *PTCD3*, *A2M*, *UBA52* and *ARID1B* genes between different types of cancers and normal tissue. The resulting heatmap showed the different expression levels of these four shared DEGs in almost all types of cancers compared to normal tissue. The survival map showed the impact of four shared DEGs on the overall survival rate of patients in different types of cancers, with *UBA52* gene showing the most significant role in the overall survival of patients with adrenocortical carcinoma (ACC), kidney chromophobe (KICH), and colon adenocarcinoma (COAD). The heatmap and survival map of the four shared DEGs are represented in Fig. 2.

**Shared DEGs analysis based on TCGA data**
UALCAN dataset analysis, which is based on TCGA data, revealed a decrease in expression levels of PTCD3, A2M, and ARID1B genes in pancreatic cancer tissue compared to normal pancreatic tissue. The differential expression was not considerable for the UBA52 gene between tumor and normal pancreas. The box plot in Fig. 3 represents the different transcription levels of these four shared DEGs.

We subsequently investigated PTCD3, A2M, UBA52 and ARID1B gene expression levels in different PaC stages. The stage plot of these shared DEGs reported a decrease of gene expression levels for PTCD3, A2M, and ARID1B genes in PaC stages I-IV when compared to healthy individuals but UBA52 gene did not show a significant differential expression. The box plot in Fig. 4 shows the expression levels of these four shared DEGs based on individual pancreatic cancer stages. Based on pancreatic tumour grade, TCGA data analysis showed low expression of PTCD3, A2M, and ARID1B genes although no significant expression of UBA52 gene in grades 1-4 of pancreatic tumour compared to normal samples. Moreover, we investigated the expression levels of these four shared DEGs in nodal metastasis and we observed that expression levels of PTCD3, A2M, and ARID1B for both metastatic and no-metastatic regional lymph nodes were lower than normal samples. However, expression levels were almost equal for metastatic and no-metastatic regional lymph node samples. In this analysis, UBA52 expression was not significant. TCGA data analysis of four shared DEGs based on pancreatic tumour grade and nodal metastasis are represented in Fig. 5 and 6, respectively.

Using the Kaplan–Meier plot, we compared the overall survival rate of PaC patients according to the high and low expression of PTCD3, A2M, UBA52, and ARID1B genes. The results indicated that low expression of PTCD3, A2M, and ARID1B genes has a positive effect on patient survival while UBA52 gene high expression level was positively correlated with survival of PaC patients. The Kaplan–Meier plot of four shared DEGs is represented in Fig. 7.

**Discussion**

It is commonly agreed that early diagnosis of pancreatic cancer will significantly improve the success of treatment and, consequently, increase the overall survival rate of patients. However, early diagnosis of PaC is not straightforward and current biomarkers have limited utility to detect PaC in the early stage (9). At present, biomarkers identified as potential compounds for the early detection of PaC include proteins (12–17), miRNAs (18–20), metabolites (21–23), aberrantly glycosylated antigens (CA19-9 and Tn antigens) (24–26), autoantibodies (27, 28), and exosomes (29, 30). Some studies have also used the microarray technique to investigate the gene expression profile of peripheral blood mononuclear cells (PBMCs) in PaC patients (31–33). In contact with tumour tissue PBMCs can alter their gene expression profile and induce a complex transcriptional response (34–36). Therefore, studies on blood gene expression profiling may contribute to identify reliable cancer biomarkers.

In the current project, a transcriptome study of differentially expressed genes in blood samples of PaC patients was carried out. We aimed to identify differentially expressed genes from early stage to metastatic PaC, considering the following three inclusion criteria for patients selection: whole blood
sample, the participants’ age between 50-80 years old and patient’s samples collected before surgery or any other medical intervention. Using RNA-sequencing technique, we first identified differentially expressed genes in early stage PaC and metastatic PaC cases compared to normal individuals (significant DEGs of early stage group = 331, metastatic group = 724). Then using the Venn diagram tool, 146 genes were identified as shared DEGs between early stage and metastatic PaC patients; 103 genes were upregulated during cancer progression (from early stage to metastasis) while remaining genes were found downregulated. Based on log2–1.5 and P-value ≤ 0.05 as the cut-off criteria, we finally identified PTCD3 and UBA52 genes upregulate while A2M and ARID1B genes were downregulated (see Supplementary File S3 online).

The AT-rich interaction domain-containing protein 1B (ARID1B) is a member of the human switch/sucrose non-fermentable (SWI/SNF) chromatin remodelling complex (37). SWI/SNF is known as a tumour suppressor and mutations of this complex are widespread across a variety of human cancers (38, 39). The loss of ARID1B has been reported in breast (39), colorectal (40), and pancreatic (41) cancers as well as malignant melanoma and hepatocellular carcinoma (39). Khursheed et al carried out a study on the tumour-suppressor activity of ARID1B gene and reported its reduced expression level in advanced-stage pancreatic tumours when compared with normal pancreas (42). In line with this report, we observed a decreased expression level of the ARID1B gene in blood samples of PaC patients compared to normal sample and decreased expression level pattern of the ARID1B gene from early stage to metastatic PaC (log2: early stage = -4.50163, metastatic = -6.18771).

Low expression levels of A2M were also observed in blood samples of PaC cases compared to normal samples with decreased pattern from early stage to metastatic PaC (log2: early stage = -3.13408, metastasis = -4.01882). Alpha-2-macroglobulin (A2M) plays a key role as a universal proteinase inhibitor (43) by covalently binding to different proteinase, such as carboxyl, thiol, serine, and metalloprotease (44). This glycoprotein can also bind to various growth factors, cytokines and other non-proteolytic molecules (45). Due to its unique molecular structure, A2M can bind to proteinases and plays its anti-inflammatory function, especially for proteinase released from granulocytes and other inflammation-related cells (44). Overexpression matrix metalloproteinase 9 (MMP9) in blood has been correlated with cell-growth, invasion, and tumour metastasis in breast, colorectal, and pancreatic cancers (46–48). In contrast, inhibition of the MMP9 activity has been associated to primary tumour mass shrinkage and metastasis probability (46). Huang et al reported that MMP9 increased level in PBMCs might support PaC development (31). At early cancer stages, macrophages are recruited into the tumour microenvironment for their anti-tumour immune activity but can differentiate into tumour-associated macrophages due to specific signals from the tumour microenvironment, leading to the endothelial to mesenchymal transition (EMT) and metastasis in tumour cells (49, 50). Tekin et al study showed that, protease-activated receptor 1 (PAR1) plays a pivotal role in driving macrophage recruitment into the pancreatic tumour microenvironment. This recruited macrophage can secretes MMP9, undergoing tumor cells to EMT process. They finally speculated that the recruited naive macrophages within the tumour microenvironment can differentiate into tumour-associated macrophages, thus supporting MMP9-PAR1-
EMT dependent process and facilitating tumour progression (51). These studies support our results of downregulated $A2M$ (as proteinase inhibitor) gene expression levels in the blood of PaC patients compared to normal individuals, as well as its decreased gene expression pattern from early to metastatic stage.

Our study indicated that pentatricopeptide repeat domain protein 3 ($PTCD3$) was downregulated in blood samples of PaC patients compared to normal individuals whereas its expression level increased from early to metastatic stage ($\log Fc$: early stage $\approx -5.73478$, metastasis $\approx -4.13016$). The oncogenic role of $PTCD3$ was reported in breast and prostate cancers as well as lymphoma (52), and can corroborate our results given its upregulation during the tumour progression. High $UBA52$ (Ubiquitin A-52) expression level was also observed in our experimental groups compared to normal sample and its expression level was found upregulated during tumor progression ($\log Fc$: early stage $\approx 2.99128$, metastasis $= 5.04053$). Noteworthy, $UBA52$ activates $RPL40$-$MDM2$-$p53$ pathway and, consequently, induces cell cycle arrest and apoptosis in tumour tissue (e.g., colorectal cancer, as reported by Zhou et al. 2019) (53). We can therefore speculate that $UBA52$ plays a key role in tumour inhibition, although its upregulation in our study was in contrast with this report. Hence, we suppose that $UBA52$, when expressed in blood cells, may be involved in other pathways resulting in its upregulation during tumour progression.

Lastly, we compared the expression levels of $PTCD3$ $UBA52$, $A2M$, and $ARID1B$ in blood and pancreatic tumour tissue according to individual cancer stages, pancreatic tumor grade, and nodal metastasis, reported in TCGA database. Similarly, we observed low expression of $PTCD3$, $A2M$, and $ARID1B$ genes in blood and pancreas tumour tissue compared to normal individuals. But $UBA52$ showed high expression in blood samples although its expression was not significant in tumour tissue compared to normal control (see Fig. 3 and Supplementary File S3 online). Similar result was observed for grades 1–4 of pancreatic tumours and in metastatic and no-metastatic regional lymph nodes (see Figs. 5 and 6). The stage plot illustrated sinusoidal changes in expression levels of $PTCD3$, $A2M$, and $ARID1B$ genes from stage I to IV although their increasing trend was observed during cancer progression. While $UBA52$ expression showed a slight increase from stage I to IV (see Fig. 4). The sinusoidal changes in expression levels of $PTCD3$, $A2M$, and $ARID1B$ were also observed in grade 1 to 4 of pancreatic tumours. The $PTCD3$ and $ARID1B$ genes showed a decreasing trend, but $A2M$ gene showed an increasing trend and $UBA52$ showed no significant expression during tumour progression (see Fig. 5). Additionally, we observed that $UBA52$ expression levels were not significant in metastatic and no-metastatic regional lymph nodes tumours compared to normal control. While expression levels of $PTCD3$, $A2M$, and $ARID1B$ genes in metastatic and no-metastatic regional lymph nodes were lower than normal samples, although these expression levels were almost equal in metastatic and no-metastatic regional lymph node tumours (see Fig. 6). Using Kaplan–Meier plot we finally found that low expression of $PTCD3$, $A2M$, and $ARID1B$ genes in PaC patients have positive effect on PaC patients overall survival while high expression level of $UBA52$ gene was in positively correlated with survival of PaC patients (see Fig. 7).

**Conclusion And Future Perspective**
We hypothesized that the identification of genes differentially expressed from early stage to metastatic PaC patients in peripheral blood could be helpful to design non-invasive and cost-effective diagnostic tests for early diagnosis of pancreatic tumor in respectable stages. Using RNA-sequencing analysis, we identified four DEGs in the blood transcriptome profile of pancreatic cancer patients that were shared between early and metastatic stages and we further investigated their expression trend during cancer progression. Overall, 146 DEGs were identified although we subsequently focused on those statistically significant DEGs according to jointly agreed cut-off criteria. The small number of samples was the main limitation of the current study. We believe that study including more samples could draw more accurate conclusions. Of note, there is growing evidence demonstrating that platelets might have an important role in the development and progression of cancer and their mRNA profile is currently emerging as a new potential source in cancer biomarker research. Unlike other studies - that mainly focus on peripheral blood mononuclear cells- we studied whole blood samples of PaC cases. We therefore suggest investigating on our identified DEGs, with particular emphasis on their expression levels in platelets. Finally, as extended investigation, we propose validation of our four DEGs on RNA and protein level as well as the investigation on their regulatory networks, in order to shed light on important molecular pathways of blood cells in pancreatic cancer progression.

**Methods**

**Study design summary**

The current project was performed in three main steps including:

- Specimen collection
- RNA isolation and sequencing
- Bioinformatics analysis

**Specimen collection**

*Study population.* Patients recruited for the current study were selected among those who were doubtful to have a pancreatic mass and subsequently referred to the tertiary referral hospital (Shariati hospital, Tehran, Iran) during 2019. First, informed written consent was obtained from every participant and a validated and reliable questionnaire was used to collect detailed demographic information, lifestyle as well as different exposures. The endoscopic ultrasonography was performed on every participant and in case of mass or a metastatic lesion identification, the patient underwent fine-needle aspiration (FNA). If the diagnosis of pancreatic cancer could not be achieved using hematoxylin/eosin staining, the samples were evaluated by immunohistochemistry staining. The collected pancreatic tissue samples were then reviewed by an expert pathologist who was blinded to the questionnaire data. Finally, the pancreatic mass and tumor staging were reconfirmed by an expert gastroenterologist. The pancreatic cancer cases at early (I/II) and metastatic (IV) stages were included in the study and blood samples were collected from all enrolled patients before surgery or any other medical intervention.
Ethical statement: Sample collection procedures performed in this study involving human participants were conducted according to the principles expressed in the institutional and/or national research committee of the 2013 Helsinki declaration. This study was approved by the Institutional Review Board and Ethics Committees of the Digestive Disease Research Institute of Tehran University of Medical Sciences and the National Institute for Medical Research Development (approval number: IR NIMAD REC. I398.180).

RNA isolation and sequencing

Total RNA extraction, purification, and quantification. A volume of 600 μl peripheral whole blood (for each group) was collected into an EDTA tube and total RNA extraction was performed using QIAzol Lysis Reagent (Qiagen, Germany) according to the manufactures' protocol. Consequently, DNase I, RNase-free Kit (Thermo Scientific, USA) was used to remove potential DNA contamination from extracted RNA. Concentration, RNA integrity number (RIN), 28S/18S ratio and size of RNA were assessed using Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit, Agilent Technologies, USA) and samples with RIN > 6 were used for sequencing. The purity of the samples was tested by NanoDrop ND-100 spectrophotometer (NanoDrop Technologies, USA).

Library construction and RNA sequencing. Library construction was carried out according to the following steps:

a. mRNA isolation: 200ng total RNA was purified using oligo-dT beads, followed by TruSeq RNA Sample Prep Kit v2 (Illumina, USA) protocol for Library construction.

b. Fragmentation: The RNA was fragmented into small pieces with Elute, Prime, Fragment Mix.

c. cDNA synthesis: First-strand cDNA was generated by First Strand Master Mix and Super Script II (Invitrogen, USA) reverse transcription, under reaction conditions of 25°C for 10 min, 42°C for 50 min, and 70°C for 15 min. The product was purified by Agencourt RNA Clean XP Beads (AGENCOURT, USA) and then Second Strand Master Mix and dATP, dCTP, dGTP, dUTP mix were added to synthesize the second-strand cDNA (16°C for 1h).

d. End repair and A-Tailing: The purified fragmented cDNA was combined with End Repair Mix and incubated at 30°C for 30 min. The end-repaired DNA was purified with Ampure XP Beads (AGENCOURT, USA). Then A-Tailing Mix was added, mixed, and incubated at 37°C for 30 min.

e. Adapter Ligation: The Adenylate 3’Ends DNA, RNA Index Adapter and Ligation Mix were combined, mixed and incubated at 30°C for 10 min. Then end-repaired DNA was purified with Ampure XP Beads (AGENCOURT, USA).

f. UNG digestion: The Uracil-N-Glycosylase enzyme was added into the samples containing the DNA and incubated at 37°C for 10 min. The product underwent a second purification step with AGENCOURT Ampure XP Beads.

g. RCR: The PCR Primer Cocktail and PCR Master Mix were used to perform 15 cycles of PCR amplification to enrich the cDNA fragments. Finally, the PCR products were purified by Ampure XP
Beads (AGENCOURT, USA).

The constructed libraries were assessed by checking the distribution of the fragments size using the Agilent 2100 bioanalyzer (Agilent DNA 1000 Reagents, Agilent Technologies, USA). The libraries then were amplified on cBot to generate the cluster on the flow cell (TruSeq PE Cluster Kit V3–cBot–HS Illumina, USA), and therefore the amplified flow cell was sequenced using pair-end strategy (100 bp paired-end read) on the HiSeq 4000 System (TruSeq SBS KIT-HS V3 Illumina, USA). Quality control of FASTQ files was assessed using FastQC:Read QC and Trimmomatic tool was employed to check quality of trimming and adapter removal. The trimmed raw data were mapped to the reference genome (Ensembl GRCh38/ hg19 assembly) via Hisat2 tool for transcripts alignment and quantification. Differentially expressed genes were identified using Cuffdiff (Cufflinks package version 2.1.1) tool and those with log fold change (logFcs) ≥ ±1.5 and Pvalue ≤ 0.05 were considered for further bioinformatics analyses. Additionally, the Venn diagram online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) was used to show the shared DEGs between experimental groups.

**Bioinformatics analysis**

**GEPIA2 Dataset.** GEPIA2 (Gene Expression Profiling Interactive Analysis2) (http://gepia2.cancer-pku.cn/index.html) is a newly developed interactive platform to analyze the RNA sequencing expression data of tumors and normal samples from the TCGA (The Cancer Genome Atlas) and the GTEx (Genotype-TissueExpression) projects, based on a standard processing pipeline(54). We used GEPIA2 dataset for pan-cancer view and survival map of identified DEGs obtained from our RNA sequencing analysis.

**UALCAN Dataset.** UALCAN (http://ualcan.path.uab.edu/) is another interactive online platform for cancer OMICS data analysis. This platform allows researchers to perform in silico validation of potential genes of interest, identification of biomarkers, generation of graphs and plots of expression profile and patient survival information, with easy access to public available TCGA, CPTAC, and MET500 cancer OMICS data (55). We used UALCAN dataset to show expression level of identified DEGs in pancreatic cancer tissue compared to normal tissue, as well as expression levels of DEGs according to pancreatic tumor grade, individual cancer stages, and nodal metastasis. We also used UALCAN dataset to generate Kaplan-Meier survival curves to compare the high and low expression of DEGs in pancreatic cancer and correlation of DEGs with patients overall survival.

**Data Availability:** All data generated or analyzed during this study are included in this published article and its Supplementary Information files.

**Abbreviations**

ACC: Adrenocortical carcinoma  BLCA: Bladder Urothelial Carcinoma  BRCA: Breast invasive carcinoma  CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma  CHOL: Cholangiocarcinoma  COAD: Colon adenocarcinoma  DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma  ESCA:
Esophageal carcinoma **GBM**: Glioblastoma multiforme **HNSC**: Head and Neck squamous cell carcinoma **KICH**: Kidney Chromophobe **KIRC**: Kidney renal clear cell carcinoma **KIRP**: Kidney renal papillary cell carcinoma **LAML**: Acute Myeloid Leukemia **LGG**: Brain Lower Grade Glioma **LIHC**: Liver hepatocellular carcinoma **LUAD**: Lung adenocarcinoma **LUSC**: Lung squamous cell carcinoma **MESO**: Mesothelioma **OV**: Ovarian serous cystadenocarcinoma **PAAD**: Pancreatic adenocarcinoma **PCPG**: Pheochromocytoma and Paraganglioma **PRAD**: Prostate adenocarcinoma **READ**: Rectum adenocarcinoma **SARC**: Sarcoma **SKCM**: Skin Cutaneous Melanoma **STAD**: Stomach adenocarcinoma **TGCT**: Testicular Germ Cell Tumors **THCA**: Thyroid carcinoma **THYM**: Thymoma **UCEC**: Uterine Corpus Endometrial Carcinoma **UCS**: Uterine Carcosarcoma **UVM**: Uveal Melanoma

**Declarations**

**Acknowledgements**

We thank all the patients and healthy donors who participated in the present study. We are grateful to Dr. Mojgan Bandehpour for RNA isolation procedure consulting, Beijing Genomics Institute (BGI.Group) for RNA-sequencing, Dr. Babak Arefnejad for RNA-sequencing analysis, and all those who helped in this project.

**Author contributions**

S.K. designed the study, performed the experiments, and wrote the manuscript; B.K., R.R., and H.Z. analyzed the data; A.D. A. participated in manuscript editing and discussions; A.M. collected samples and obtained clinical data; Z.M. Acquisition of data; A.P. Study conception and design and clinical consulting.

**Additional information**

**Funding**: This project was funded by the National Institute for Medical Research Development (grant number: 982950, grant recipient: Akram Pourshams).

**Competing interests**: The authors have declared that no competing interests exist.

**References**


**Figures**
Figure 1

Venn diagram illustrates the number of DEGs in experimental groups (early stage and metastatic PaC) compared to normal control, obtained from RNA-sequencing analysis. The number of shared DEGs between experimental groups is also demonstrated in this diagram.
Figure 2

The pan-cancer view and survival map of four shared DEGs (shared between experimental groups) (the full name of the cancers has been mentioned in the abbreviation section). A. The heatmap plot represents the transcript expression levels of four shared DEGs in different types of cancers compared to normal samples. T: tumor, N: normal. B. The survival map represents the impact of four obtained DEGs on overall survival rate of patient in different types of cancers. HR: hazardous ratio. (by GEPIA2 dataset analysis)
The box plot represents the different transcription levels of four shared DEGs in pancreatic cancer and normal samples. Blue box: normal tissue (n= 4), Red box: tumor tissue (n= 178) (by UALCAN dataset analysis).
Figure 4

The stage plot represents the expression level of four shared DEGs in pancreatic cancer based on individual cancer stages 1, 2, 3, and 4 compared to normal individuals. Normal (n = 4), Stage1 (n = 6), Stage2 (n = 146), Stage3 (n = 4), and Stage4 (n = 4) (by UALCAN dataset analysis).
Figure 5

The stage plot represents the expression level of four shared DEGs in pancreatic cancer based on tumor grades 1, 2, 3, and 4 compared to normal samples. Normal (n = 4), Grade1 (n = 31), Grade2 (n = 95), Grade3 (n = 48), and Grade4 (n = 2) (by UALCAN dataset analysis).
Figure 6

The stage plot represents the expression level of four shared DEGs in pancreatic cancer based on nodal metastasis status compared to normal samples. Normal (n = 4), N0: No metastatic regional lymph nodes (n = 49), and N1: metastatic regional lymph nodes (n = 124) (by UALCAN dataset analysis).
Figure 7

The Kaplan–Meier plot represents the impact of four shared DEGs on overall survival rate of patients with pancreatic cancer. In all graphs, blue and red lines demonstrate low and high expression levels of transcript, respectively. High expression level \( n \geq 45 \), low expression level \( n \geq 132 \) (by UALCAN dataset analysis).