Mathematical Biosciences and Engineering

MBE, 17(1): 910–927.
DOI: 10.3934/mbe.2020048
Received: 20 May 2019
Accepted: 24 September 2019
Published: 06 November 2019

http://www.aimspress.com/journal/MBE

*Research article*

# Screening and validating the core biomarkers in patients with pancreatic ductal adenocarcinoma

**Yan Li, Yuzhang Zhu, Guiping Dai, Dongjuan Wu, Zhenzhen Gao, Lei Zhang and Yaohua Fan***

Department of Oncology, The Second Affiliated Hospital of Jiaxing University, Jiaxing 314000, China

* **Correspondence:** Email: fyh20190517@163.com; Tel: +8613486337597.

**Abstract:** Pancreatic ductal adenocarcinoma (PAAD) is one of the most common malignant tumors in digestive system. To find the new therapeutic targets and explore potential mechanisms underlying PAAD, the bioinformatics has been performed in our study. The PAAD gene expression profile GSE28735 was chosen to analyze the differentially expressed genes (DEGs) between PAAD carcinoma tissues and normal adjacent tissues from 45 patients with PAAD. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed using Database for Annotation, Visualization and Integrated Discovery (DAVID). Moreover, a protein-protein interaction (PPI) network was also constructed to help us screen the top 20 hub genes in this profile and demonstrated the underlying interactions among them. The Gene Expression Profiling Interactive Analysis (GEPIA) was further performed in order to valid the mRNA levels of top5 up-regulated and top5 down-regualted DEGs, apart from exploring their association with survival rate as well as tumor stage. Finally, Q-PCR was further employed to valid the top5 up-regulated and top5 down-regulated genes in patients with PAAD. In our study, there were a total of 444 DEGs captured (271 up-regulated genes and 173 down-regulated genes). Among these DEGs, the top5 up-regulated genes were CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1 while the top5 down-regulated genes were GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1. GO analysis disclosed that the DEGs were predominantly enriched in cell adhesion, lipid metabolism, integrin binding, proteolysis and calcium ion binding. KEGG analysis disclosed that the enriched pathway included pancreatic secretion, protein digestion and absorption, fat digestion and absorption, ECM-receptor interaction, focal adhesion and PI3K-Akt signaling pathway. Survival analysis unveiled that the high expression levels of SLC6A14, GALNT5 and TSPAN1 may correlate with the poor prognosis while high expression levels of IAPP may contribute to a better prognosis in patients with PAAD. Additionally, the levels of CEACAM5, SLC6A14, LAMC2 and GALNT5 were also associated with tumor stage. Furthermore, according to the connectivity degree of these DEGs, we

selected the top20 hub genes, namely ALB, FN1, EGF, MMP9, COL1A1, COL3A1, FBN1, CXCL12, POSTIN, BGN, VCAN, THBS2, KRT19, MET, MMP14, COL5A2, GCG, MUC1, MMP1 and CPB1, which were expected to be promising therapeutic targets in PAAD. Collectively, our bioinformatics analysis showed that DEGs and hub genes may be defined as new biomarkers for diagnosis and for guiding the therapeutic strategies of PAAD.

**Keywords:** pancreatic ductal adenocarcinoma; biomarkers; hub genes; bioinformatics analysis

## 1. Introduction

The pancreas is a complex and unique digestive organ, which is composed of exocrine cells which are responsible for the production of pancreatic juice containing essential digestive enzymes, and endocrine cells which are implicated with synthesis, storage and secretion of insulin as well as glucagon [1,2]. Pancreatic cancer costs countless of lives each year all over the world, having one of the lowest survival rates among various cancers [3]. Pancreatic ductal adenocarcinoma (PAAD), one of the main types of pancreatic cancer, is the fifth cause of cancer-related death and is particularly aggressive and resistant to conventional and targeted therapeutic agents, giving rise to a dismal 5-year survival rate of 5% [4,5]. Currently, imageological examination and pathological biopsy are the conventional diagnostic methods of PAAD [6]. However, the current biomarkers are of little help in addressing diagnosis of PAAD and no specific biomarker for this disease exists [7]. Hence, the identification of specific and sensitive biomarkers are of necessity in order to obtain accurate diagnosis and treatment of PAAD as early as possible.

More recently, high-throughput gene microarray has been regarded as a very crucial strategy for medical research especially in various cancers [8]. Bioinformatics analysis could help to screen and identify the differentially expressed genes (DEGs) between normal samples (tissues or serum) as well as pathological samples, and further explore the potential mechanisms underlying the occurrence and development of diseases by processing high-throughput sequencing data, which has been applied to the extensive research about carcinoma, involving early diagnosis, staging and grading, prognosis and so on [9]. However, the application of the existing microarrys in clinical application is greatly restricted by countless number of genes identified by gene profiling, short of both repeatability and independent verification, apart from complex statistical analyses [10]. Therefore, to put these expression profiles applying in clinic as soon as possible, it is vital to identify an appropriate amount of genes which could serve as biomarkers or therapeutic targets and develop a suitable way that can be done by routine assay.

In this study, we selected the dataset GSE28735 from Gene Expression Omnibus (GEO) and applied bioinformatics analysis to screen the DEGs in PAAD. Meanwhile, we used STRING to construct a protein-protein interaction (PPI) network to identify the hub genes with top20 degree of connectivity in PAAD. Moreover, the Gene ontology (GO) analysis including biological process (BP), molecular function (MF), cellular component (CC) and KEGG pathways analysis were performed based on DAVID. We also analyzed tumor stage and overall survival (OS) of top5 up-regulated and top5 down-regulated DEGs in PAAD. Finally, we further re-identified the mRNA expression levels of the top5 up-regulated and top5 down-regulated DEGs by Q-PCR. Meanwhile, we randomly selected one of the top5 up-regulated genes (TSPAN1) to observe its impact on the

proliferation of pancreatic cancer cells. These genes are promised to act as novel biomarkers and therapeutic targets for the diagnosis and treatment of PAAD in the near future.

## 2. Methods

### *2.1. Screening database*

We searched the relevant datasets about PAAD in PubMed and the dataset GSE28735 was finally selected out. GSE28735 is a dataset based on GPL570 platform, which contains 45 pairs of carcinoma tissues and adjacent non-tumor pancreatic tissue from patients with PAAD. Meanwhile, we also downloaded the raw series matrix file of the GSE28735 dataset.

### *2.2. Screen genes of differential expression*

The analysis of DEGs between PAAD samples and normal pancreatic ductal tissues samples was performed by using GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/), an online analysis method for GEO database based on R language [11]. We regarded DEGs as differentially expressed with logFC < －1 (down-regulated genes with the fold change of 2) or logFC > 1 (up-regulated genes with the fold change of 2), according to a criteria described in previous research [9]. The adjust P value < 0.05 was regarded as statistically significant, aiming to reduce the false positive rate. Subsequently, 444 DEGs were selected out, including 271 up-regulated genes and 173 down-regulated genes. What is more, we also utilized visual hierarchical cluster analysis to get the heatmap and volcano plot of two groups by ImageGP (http://www.ehbio.com/ImageGP/index.php/Home/Imdex/index.html) after the correlative raw data of TXT files was downloaded.

### *2.3. Gene ontology and KEGG pathway analysis of DEGs*

Gene ontology (GO) analysis is in essence a common framework annotating genes and gene products, which includes functions of cellular components (CC), biological pathways (BP) and molecular function (MF) [12]. Kyoto Encyclopedia of Genes and Genomes (KEGG) was a collection of databases which help to analyze genomes, biological pathways, diseases, chemical substances and drugs [13]. The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.ncifcrf.gov) is an online bioinformatics database [14]. It has covered a great many biological data and relevant analysis tools, and provide tools for the biological function annotation information for plenty of genes or proteins. P < 0.05 was considered as the cut-off criterion with significant difference. We could visualize the key biological processes (BP), molecular functions (CC), cellular components (MF) and KEGG pathways of these DEGs by using DAVID online database.

### *2.4. Identify sensitive biomarkers of PAAD*

The Gene Expression Profiling Interactive Analysis (GEPIA), designed by Chenwei Li, Zefang Tang, and Boxi Kang of Zhang Lab, Peking University, is a newly developed interactive web server (http://gepia.cancer-pku.cn/index.html), in order to analyze the RNA sequencing expression data of

9736 tumors and 8587 normal samples from the GTEx and TCGA projects in a standard processing pipeline [15]. In our study, we employed the boxplot to visualize the mRNA expression of top5 up-regulated and top5 down-regulated DEGs in PAAD tissues and normal pancreas tissues.

### *2.5. The overall survival (OS) and tumor stage of the top5 upregulated and top5 downregulated DEGs*

Similarly, we used the GEPIA database to get the overall survival (OS) and tumor stage information of these DEGs. The logrank P value and hazard ratio (HR) with 95% confidence intervals were showed on the plot. P < 0 05 was statistically significant.

### *2.6. Establishment of the PPI network*

Search Tool for the Retrieval of Interacting Genes (STRING) is an online app for evaluating PPI network. In order to investigate the potential correlation of DEGs, we used online app STRING to map the DEGs, subsequently the Cytoscape software was utilized to construct a PPI network. The confidence score $\geq 0.4$ and maximum number of interactors = 0 were set as the criterion.

### *2.7. Quantitative Real-time PCR (Q-PCR)*

A total of 30 male patients who were diagnosed with PAAD by pathology reports in our hospital were enrolled. The samples from patients with PAAD for the validation of these DEGs were supported by the ethics committee in our hospital. To begin with, the paired carcinoma tissues and adjacent tissues were isolated. Then total RNA was isolated using the TRIzol (Invitrogen, Carlsbad, CA, USA) kit, the concentrations and purities of which were quantified by a ultraviolet spectrophotometer (Synergy 4 Microplate Reader, BioTek, USA). The RNA was then reversely transcribed according to the standard operation process [16]. The expression levels of top5 up-regulated genes and top5 down-regulated genes were normalized to GAPDH. Relative mRNA expression levels were analyzed by 2-ΔΔcycle threshold (CT) method.

### *2.8. Cell culture and MTT assay*

The PAAD SW1990 cell line was obtained from Shanghai Cell Bank (Shanghai, China). The cell line were then propagated and cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen, Carlsbad, CA, USA), supplemented with 10% fetal bovine serum (FBS; Sigma, St. Louis, MO, USA), 100 μg/mL penicillin and streptomycin, in a humidified chamber at 37˚C with 5% $CO_2$. SW1990 cells in the logarithmic growth phase were seeded in 6-well plates at a density of $1.5 \times 10^5$ cells per well the day before transfection. To knockdown TSPAN1 expression, SW1990 cells were transiently transfected with siRNA directed against TSPAN1 (50 nM, Santa Cruz Biotechnology) according to the protocol provided by the manufacture. Cells transfected with control (scrambled) siRNAs (50 nM, Santa Cruz Biotechnology) served as a control.

Cell proliferation was evaluated by performing CellTiter 96® $AQ_{ueous}$ One Solution Cell Proliferation Assay (MTS). MTS assay was carried out according to the manufacturer's instruction. The absorbance of each group was detected at 490 nm with the Synergy 4 Microplate Reader (BioTek). Three independent experiments were performed for each experimental condition.

### *2.9. Statistical analysis*

All values were reported as means ± SD. Statistical significance was analyzed by SPSS 19.0 software. Differences were considered significant when $P < 0.05$.

## 3. Results

### *3.1. Identification of DEGs*

In this study, a total of 45 paired carcinoma tissues and adjacent tissues from patients with PAAD were analyzed. We applied the GEO2R online analysis tool with default parameters to screen the DEGs, using adjust P value < 0.05 and logFC ≤ −1 or logFC ≥ 1 as the cut-off criteria. We captured 444 DEGs including 271 up-regulated DEGs and 173 down-regulated DEGs. Subsequently, the DEGs were presented in the form of a heatmap and a volcano plot (Figure 1A,B). In the heatmap, the top25 up-regulated genes and top25 down-regulated genes between carcinoma tissues and adjacent tissues were displayed. Among the 444 DEGs, the top5 up-regulated genes involved CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1 while the top5 down-regulated genes were GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1. The gene tiles and biological functions of top5 up-regulated and top5 down-regulated genes were displayed in Table 1.

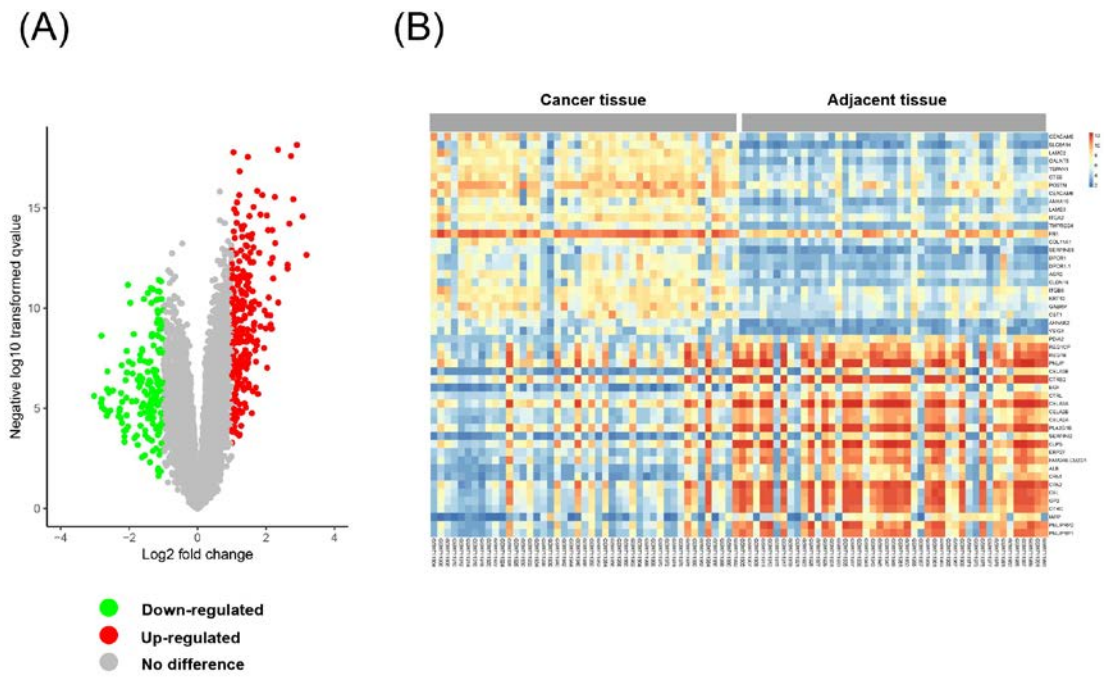


**Figure 1.** Volcano plot and heatmap of the differentially expressed genes (DEGs) between adjacent tissues and carcinoma tissues from patients with pancreatic ductal adenocarcinoma (PAAD). (A) Volcano plot of genes detected in PAAD green means up-regulated DEGs. Red means down-regulated DEGs; gray means no difference. (B) Heatmap of top 25 up-regulated DEGs and top25 down-regulated DEGs between normal and PAAD tissues.

**Table 1.** The top 5 up-regulated and down-regulated differentially expressed genes in patients with pancreatic ductal adenocarcinoma.

| DEGs | Gene title | Gene symbol | LogFC | Biological function |
|---|---|---|---|---|
| Up-regulated | carcinoembryonic antigen related cell adhesion molecule 5 | CEACAM5 | 3.1812869 | Cell adhesion, invasion, and metastasis by antibodies |
| | solute carrier family 6 member 14 | SLC6A14 | 3.075768 | Sodium and chloride dependent neurotransmitter transporters |
| | laminin subunit gamma 2 | LAMC2 | 2.9016536 | Cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis |
| | polypeptide N-acetylgalactosaminyltransferase 5 | GALNT5 | 2.7894651 | Cell migration and metastasis by antibodies |
| | tetraspanin 1 | TSPAN1 | 2.7266844 | Regulating cell development, activation, growth and motility. |
| Down-regulated | glycoprotein 2 | GP2 | - 2.793808 | Innate immune response |
| | chymotrypsin C | CTRC | - 2.7982229 | Serum calcium-decreasing factor |
| | islet amyloid polypeptide | IAPP | - 2.809762 | Inducing apoptotic cell-death |
| | pancreatic lipase related protein 2 (gene/pseudogene) | PNLIPRP2 | - 2.8201451 | Fat digestion |
| | pancreatic lipase related protein 1 | PNLIPRP1 | - 3.0266571 | Fatty acid decomposition |

### *3.2. GO enrichment analysis*

The results (Table 2 and Figure 2A,B) from GO term enrichment analysis varied from expression levels and GO classification of the DEGs. By analyzing GO enrichment of these up-regulated and down-regulated DEGs of PAAD via DAVID, we found that the up-regulated DEGs in BP were mainly enriched in extracellular matrix organization, cell adhesion, collagen catabolic process, extracellular matrix disassembly and skeletal system development while the down-regulated DEGs in BP were mainly concentrated on proteolysis, digestion, lipid digestion, lipid catabolic process and lipid metabolic process. As for CC, the up-regulated DEGs were principally enriched in extracellular space, extracellular matrix, extracellular region, extracellular exosome and proteinaceous extracellular matrix while the down-regulated DEGs were enriched in extracellular space, extracellular region, extracellular exosome, platelet alpha granule lumen and endoplasmic reticulum. Additionally, MF analysis uncovered that the up-regulated DEGs were mainly enriched in integrin binding, extracellular matrix structural constituent, collagen binding, metalloendopeptidase activity and calcium ion binding while the down-regulated DEGs are responsible for serine-type endopeptidase activity, triglyceride lipase activity, lipase activity, acylglycerol lipase activity, protein disulfide isomerase activity.

### *3.3. KEGG pathway analysis*

To obtain a more comprehensive information regarding to the critical pathways of those selected DEGs, KEGG pathways analysis were also performed via DAVID. The results in Table 3 and Figure 3 disclosed the most vital KEGG pathways of the down-regulated and up-regulated

DEGs. The down-regulated DEGs were mainly enriched in pancreatic secretion, protein digestion and absorption, fat digestion and absorption, glycerolipid metabolism and maturity onset diabetes of the young. By contrast, the up-regulated DEGs were mainly responsible for ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway and protein digestion and absorption.

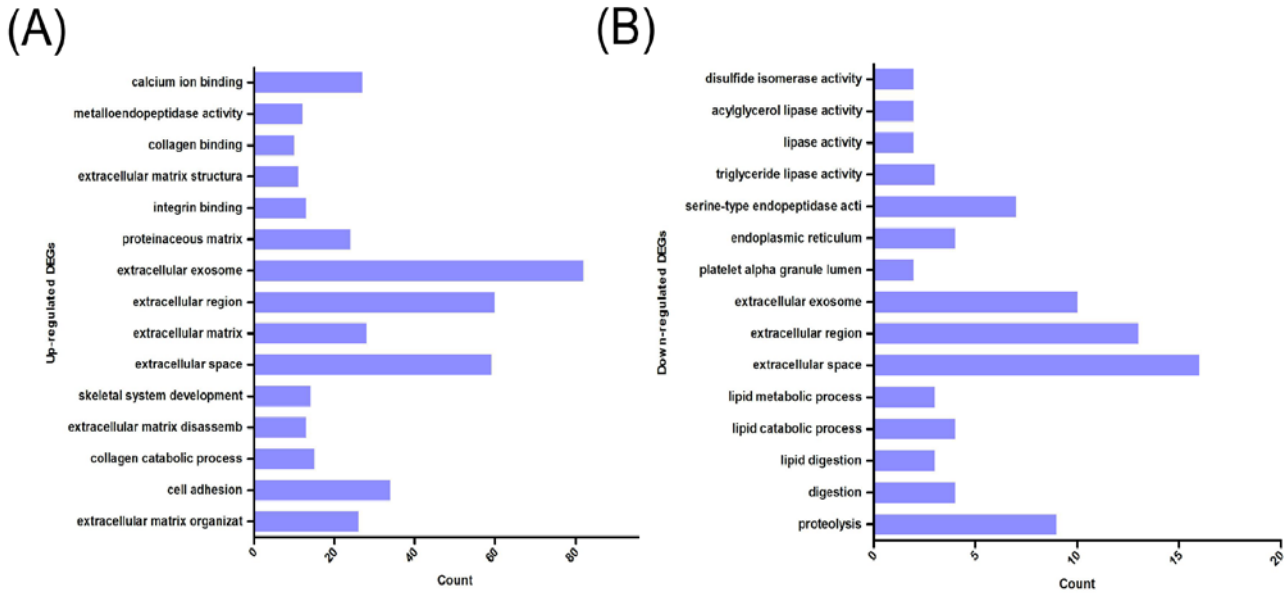


**Figure 2.** GO term enrichment analysis of up-regulated DEGs (A) and down-regulated DEGs (B).

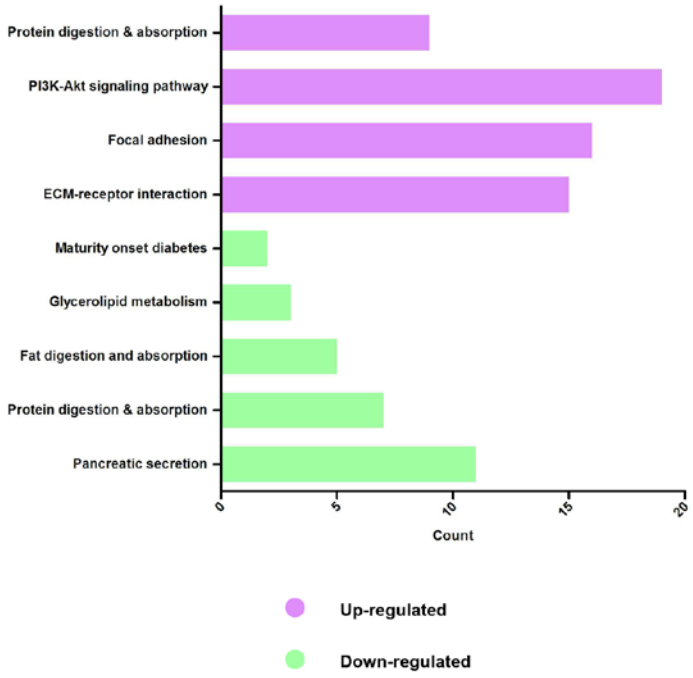


**Figure 3.** The enriched KEGG pathway in PAAD.

**Table 2.** Gene ontology analysis of differentially expressed genes in PBMCs from patients with NICM.

| Expression | Category | Term | Count | P-Value | Genes |
|---|---|---|---|---|---|
| Up-regulated | GOTERM_BP_DIRECT | GO: 0030198 ~ extracellular matrix organization | 26 | $1.77 \times 10^{-17}$ | PXDN, COL3A1, ITGA11, ITGB4, POSTN, LAMB3, ERO1A, COMP, TGFBI, COL6A3, COL8A1 |
| | GOTERM_BP_DIRECT | GO: 0007155 ~ cell adhesion | 34 | $4.10 \times 10^{-15}$ | NRP2, ITGA11, ITGB4, FERMT1, POSTN, EDIL3, CDH3, LAMB3, COMP, FAP, TGFBI, COL6A3 |
| | GOTERM_BP_DIRECT | GO: 0030574 ~ collagen catabolic process | 15 | $1.27 \times 10^{-13}$ | MMP9, COL3A1, MMP7, MMP14, COL5A2, MMP12, MMP1, MMP11, CTSK, COL6A3, COL12A1 |
| | GOTERM_BP_DIRECT | GO: 0022617 ~ extracellular matrix disassembly | 13 | $4.24 \times 10^{-10}$ | CTSK, LAMB3, LAMA3, MMP9, CAPG, FBN1, MMP7, LAMC2, MMP14, MMP12, MMP1, MMP11, FN1 |
| | GOTERM_BP_DIRECT | GO: 0001501 ~ skeletal system development | 14 | $4.58 \times 10^{-8}$ | MATN3, AEBP1, MMP9, COL3A1, FBN1, POSTN, NPR3, COL5A2, COMP, COL12A1, VCAN |
| | GOTERM_CC_DIRECT | GO: 0005615 ~ extracellular space | 59 | $3.44 \times 10^{-16}$ | CTHRC1, AEBP1, PXDN, MMP9, MMP7, POSTN, SEMA7A, FAP, TGFBI, CEACAM6, COL12A1, SEMA3C, |
| | GOTERM_CC_DIRECT | GO: 0031012 ~ extracellular matrix | 28 | $2.23 \times 10^{-15}$ | PXDN, AEBP1, LTBP1, COL3A1, MMP7, POSTN, EDIL3, MMP1, PKM, COMP, TGFBI, COL6A3, TGM2, |
| | GOTERM_CC_DIRECT | GO: 0005576 ~ extracellular region | 60 | $2.05 \times 10^{-13}$ | NRP2, ADGRF1, LTBP1, CORIN, MMP9, MMP7, CDCP1, MMP1, IL1RAP, TGFBI, COL12A1, ANGPT2 |
| | GOTERM_CC_DIRECT | GO: 0070062 ~ extracellular exosome | 82 | $9.00 \times 10^{-13}$ | AEBP1, PXDN, TSPAN1, MLPH, SLC44A4, MMP9, ANO1, MMP7, TSPAN8, EDIL3, GPX2, PKM, SLC2A1 |
| | GOTERM_CC_DIRECT | GO: 0005578 ~ proteinaceous extracellular matrix | 24 | $1.07 \times 10^{-12}$ | CTHRC1, MATN3, PXDN, LTBP1, MMP9, FBN1, MMP7, POSTN, COL5A2, MMP1, MMP |
| | GOTERM_MF_DIRECT | GO: 0005178 ~ integrin binding | 13 | $1.39 \times 10^{-8}$ | FBN1, COL3A1, ITGA2, TSPAN8, ITGA3, EDIL3, ESM1, MMP14, SEMA7A, FAP, TGFBI, ADAM9, FN1 |
| | GOTERM_MF_DIRECT | GO: 0005201 ~ extracellular matrix structural constituent | 11 | $1.70 \times 10^{-8}$ | MATN3, PXDN, BGN, COMP, COL3A1, FBN1, VCAN, COL1A1, COL11A1, MFAP5, COL5A2 |

*Continued on next page*

| Expression | Category | Term | Count | P-Value | Genes |
|---|---|---|---|---|---|
| Up-regulated | GOTERM_MF_DIRECT | GO: 0005518 ~ collagen binding | 10 | $8.59 \times 10^{-8}$ | CTSK, COMP, MMP9, TGFBI, ITGA11, ITGA2, ITGA3, ANTXR1, FN1, ADAM9 |
| | GOTERM_MF_DIRECT | GO: 0004222 ~ metalloendopeptidase activity | 12 | $2.98 \times 10^{-7}$ | ADAM28, ADAMTS6, FAP, MMP9, MMP7, ADAMTS12, ADAM12, MMP14, MMP12, MMP1, MMP11 |
| | GOTERM_MF_DIRECT | GO: 0005509 ~ calcium ion binding | 27 | $3.09 \times 10^{-6}$ | LTBP1, EDIL3, CDH3, MMP1, COMP, PLS1, THBS2, MATN3, S100A16, S100P, FBN1, PCDH7, MMP14 |
| Down-regulated | GOTERM_BP_DIRECT | GO: 0006508 ~ proteolysis | 9 | $1.93 \times 10^{-7}$ | F11, CELA3A, CELA3B, CTRC, CPA2, CELA2B, CELA2A, CPA1, CTRL |
| | GOTERM_BP_DIRECT | GO: 0007586 ~ digestion | 4 | $8.45 \times 10^{-5}$ | CELA3A, CLPS, PNLIPRP2, CTRL |
| | GOTERM_BP_DIRECT | GO: 0044241 ~ lipid digestion | 3 | $1.39 \times 10^{-4}$ | CEL, CLPS, PNLIPRP2 |
| | GOTERM_BP_DIRECT | GO: 0016042 ~ lipid catabolic process | 4 | $2.06 \times 10^{-4}$ | CLPS, PNLIPRP1, PNLIPRP2, PLA2G1B |
| | GOTERM_BP_DIRECT | GO: 0006629 ~ lipid metabolic process | 3 | 0.019323177 | CEL, CLPS, PNLIPRP1 |
| | GOTERM_CC_DIRECT | GO: 0005615 ~ extracellular space | 16 | $6.89 \times 10^{-12}$ | F11, CELA3A, PNLIPRP1, CELA3B, PNLIPRP2, SERPINI2, CEL, ALB, IAPP, PLA2G1B |
| | GOTERM_CC_DIRECT | GO: 0005576 ~ extracellular region | 13 | $2.14 \times 10^{-7}$ | F11, CEL, CLPS, PNLIPRP1, PNLIPRP2, ALB, IAPP, CTRC, GP2, PLA2G1B, CPA2, CELA2B, EGF |
| | GOTERM_CC_DIRECT | GO: 0070062 ~ extracellular exosome | 10 | 0.007042958 | F11, CEL, CLPS, REG1B, ALB, AOX1, GP2, KIAA1324, EGF, SERPINI2 |
| | GOTERM_CC_DIRECT | GO: 0031093 ~ platelet alpha granule lumen | 2 | 0.070015503 | ALB, EGF |
| | GOTERM_CC_DIRECT | GO: 0005783 ~ endoplasmic reticulum | 4 | 0.093223495 | CEL, ALB, ERP27, PDIA2 |
| | GOTERM_MF_DIRECT | GO: 0004252 ~ serine-type endopeptidase activity | 7 | $9.12 \times 10^{-7}$ | F11, CELA3A, CELA3B, CTRC, CELA2B, CELA2A, CTRL |
| | GOTERM_MF_DIRECT | GO: 0004806 ~ triglyceride lipase activity | 3 | $2.39 \times 10^{-4}$ | CEL, PNLIPRP1, PNLIPRP2 |
| | GOTERM_MF_DIRECT | GO: 0016298 ~ lipase activity | 2 | 0.012198565 | CEL, PNLIPRP1 |
| | GOTERM_MF_DIRECT | GO:0047372~acylglycerol lipase activity | 2 | 0.013545141 | CEL, PNLIPRP2 |
| | GOTERM_MF_DIRECT | GO:0003756~protein disulfide isomerase activity | 2 | 0.029567714 | ERP27, PDIA2 |

GO: Gene Ontology.

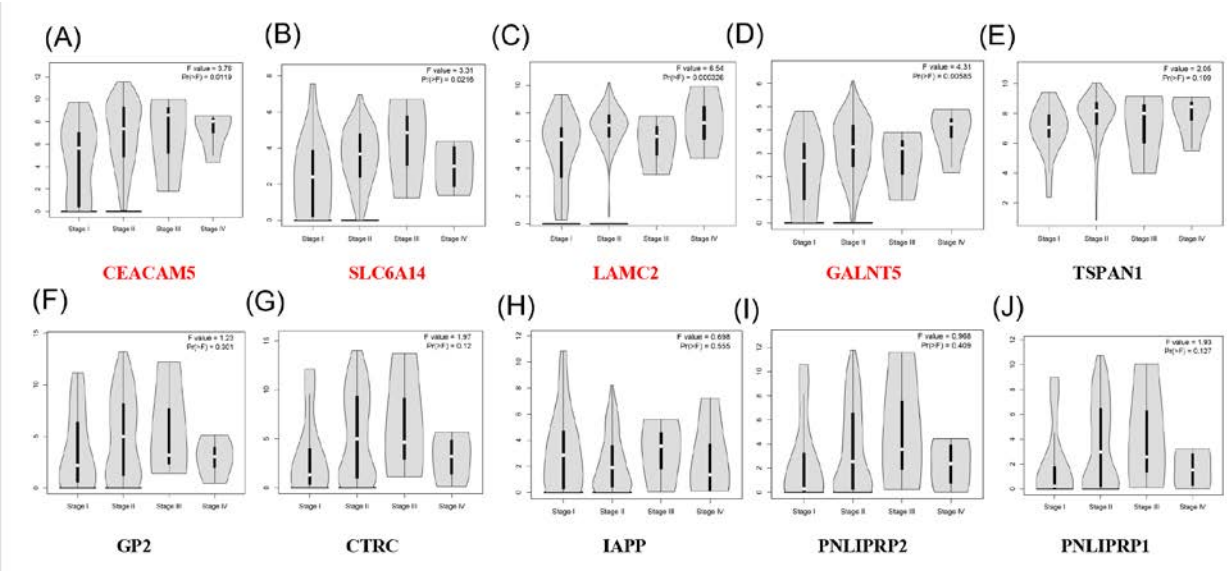


**Figure 4.** Tumor stage of top5 up-regulated and top5 down-regulated genes. (A−J) represents CEACAM5, SLC6A14, LAMC2, GALNT5, TSPAN1, GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1. P < 0 05 was regarded statistically different.

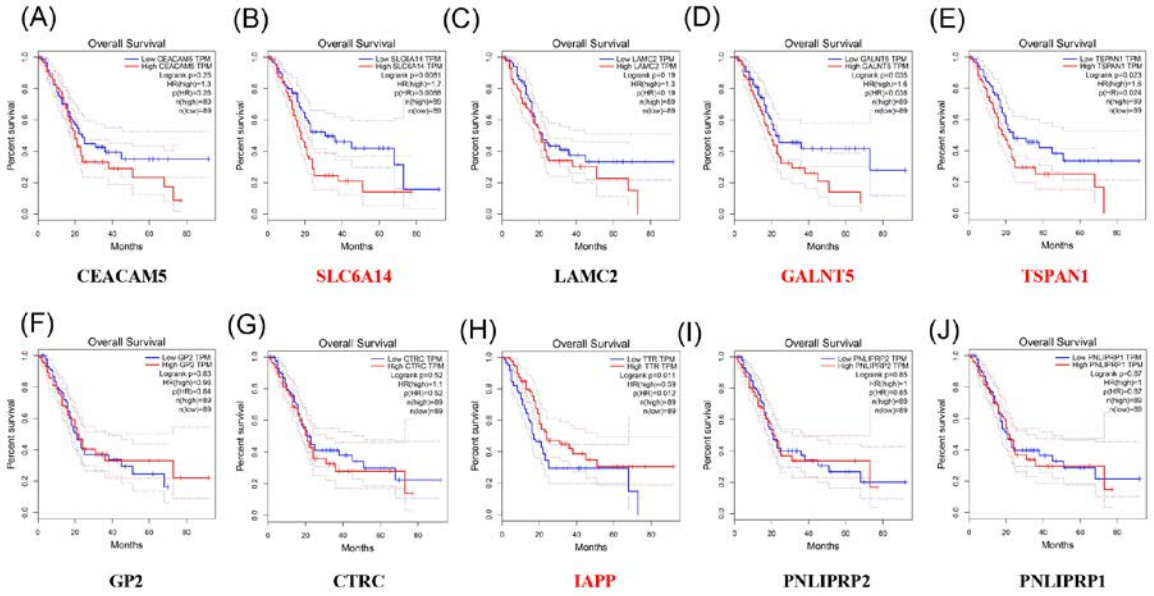


**Figure 5.** Overall survival associated with top5 up-regulated and top5 down-regulated genes base on TCGA database via GEPIA. (A–J) represents CEACAM5, SLC6A14, LAMC2, GALNT5, TSPAN1, GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1. P < 0 05 was regarded statistically different.

### *3.4. Tumor stage and overall survival (OS)*

We further analyzed the potential associations between the expression levels of top5 up-regulated genes as well as top5 down-regulated genes and the tumor stage and OS of patients

with PAAD. The results obtained from GEPIA database demonstrated that the levels of CEACAM5, SLC6A14, LAMC2 and GALNT5 were also associated with tumor stage (Figure 4). To be more specific, the expression levels of the above 4 genes in tumor stage I were significantly lower than that in stage II ($P < 0.05$).The Kaplan-Meier showed that the high expression levels of SLC6A14, GALNT5 and TSPAN1 may correlate with the poor prognosis while high expression levels of IAPP may contribute to a better prognosis in patients with PAAD ($P < 0.05$) (Figure 5).

**Table 3.** KEGG pathway analysis of differentially expressed PBMCs from patients with NICM. (KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False Discovery Rate.)

| Category | Term | Count | P-Value | Genes |
|---|---|---|---|---|
| Down-regulated | hsa04972: Pancreatic secretion | 11 | $9.26 \times 10^{-16}$ | CELA3A, CEL, CELA3B, PNLIPRP1, PNLIPRP2, PLA2G1B, CPA2, CELA2B, CELA2A, CPA1, CTRL |
| | hsa04974: Protein digestion and absorption | 7 | $2.67 \times 10^{-8}$ | CELA3A, CELA3B, CPA2, CELA2B, CELA2A, CPA1, CTRL |
| | hsa04975: Fat digestion and absorption | 5 | $1.53 \times 10^{-6}$ | CEL, CLPS, PNLIPRP1, PNLIPRP2, PLA2G1B |
| | hsa00561: Glycerolipid metabolism | 3 | 0.007771843 | CEL, PNLIPRP1, PNLIPRP2 |
| | hsa04950: Maturity onset diabetes of the young | 2 | 0.058851896 | IAPP, NR5A2 |
| Up-regulated | hsa04512: ECM-receptor interaction | 15 | $2.22 \times 10^{-11}$ | COL3A1, ITGB4, ITGA11, ITGA2, ITGA3, COL5A2, LAMB3, LAMA3, COMP, COL6A3, LAMC2, COL1A1, THBS2, COL11A1, FN1 |
| | hsa04510: Focal adhesion | 16 | $2.94 \times 10^{-7}$ | COL3A1, MET, ITGB4, ITGA11, ITGA2, ITGA3, COL5A2, LAMB3, LAMA3, COMP, COL6A3, LAMC2, COL1A1, THBS2, COL11A1, FN1 |
| | hsa04151: PI3K-Akt signaling pathway | 19 | $2.46 \times 10^{-6}$ | COL3A1, MET, ITGA11, ITGB4, ITGA2, ITGA3, COL5A2, LAMB3, LAMA3, COMP, COL6A3, LAMC2, EFNA5, IL2RG, COL1A1, THBS2, ANGPT2, COL11A1, FN1 |
| | hsa04974: Protein digestion and absorption | 9 | $4.28 \times 10^{-5}$ | KCNN4, COL17A1, COL3A1, COL6A3, COL12A1, COL1A1, COL11A1, COL5A2, COL10A1 |

KEGG: Kyoto Encyclopedia of Genes and Genome.

### *3.5. Validation of DEGs by TCGA*

To ensure the credibility of the microarray of GSE28735 and proceed further credible analysis, we validated the top5 up-regulated genes and top5 down-regulated genes based on TCGA database via GEPIA. The results showed that the mRNA expression levels of CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1 were significantly higher in carcinoma tissues compared to adjacent tissues while the mRNA expression level of GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1 in carcinoma tissues were statistically lower than the adjacent tissues ($P < 0.05$) (Figure 6A–J).

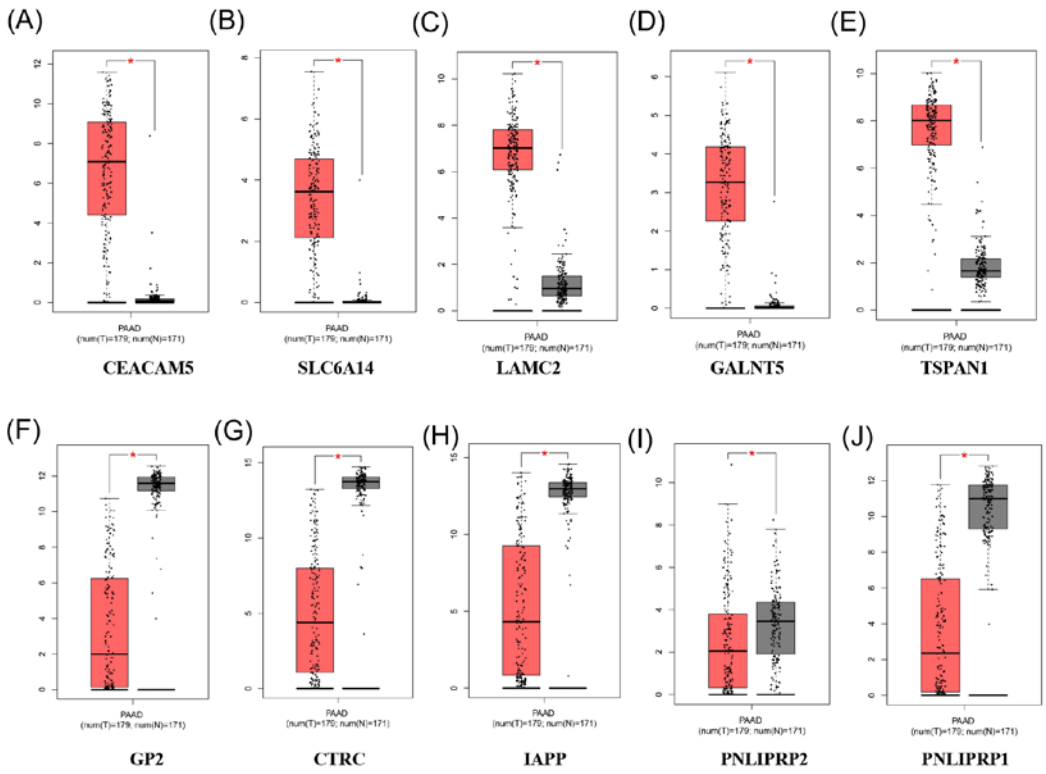


**Figure 6.** Validation of top5 up-regulated and top5 down-regulated DEGs of PAAD based on GEPIA. (A–E) CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1 were significantly up-regulated in carcinoma group. (F–J) GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1 were significantly down-regulated in carcinoma group. * P < 0.05 versus adjacent tissues. P < 0 05 was regarded statistically different.

### *3.6. Re-identification of DEGs by Q-PCR*

Furthermore, we re-identified the top5 up-regulated genes and top5 down-regulated genes in patients with PAAD from our hospital via Q-PCR. Q-PCR (Figure 7A,B) showed that the mRNA expression levels of CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1 were significantly higher in carcinoma tissues compared with adjacent tissues ($P < 0.05$). Meantime, the mRNA expression levels of GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1 were obviously down-regulated in carcinoma tissues from patients with PAAD ($P < 0.05$).

### *3.7. Hub genes with top20 connectivity degree*

Hub genes are considered to be the most critical genes located in the node of the network of DEGs, which may play vital roles [17]. Applying the STRING online tool, 199 nodes with 178 PPI relationships were established, accounting for about 89% of these selected DEGs. According to the degree of connectivity of these DEGs, we constructed the PPI network and selected the top20 hub genes (Figure 8A,B). The top15 hub genes with high degree of connectivity in PAAD are as follows, ALB, FN1, EGF, MMP9, COL1A1, COL3A1, FBN1, CXCL12, POSTIN, BGN, VCAN, THBS2, KRT19, MET, MMP14, COL5A2 GCG, MUC1, MMP1 and CPB1. The 20 hub genes could interact

with 288 genes directly, and ALB acted as the most intensive gene which could interact with 76 up-regulated genes and 56 down-regulated genes. Intriguingly, among these hub genes, there also displayed very strong interactions.

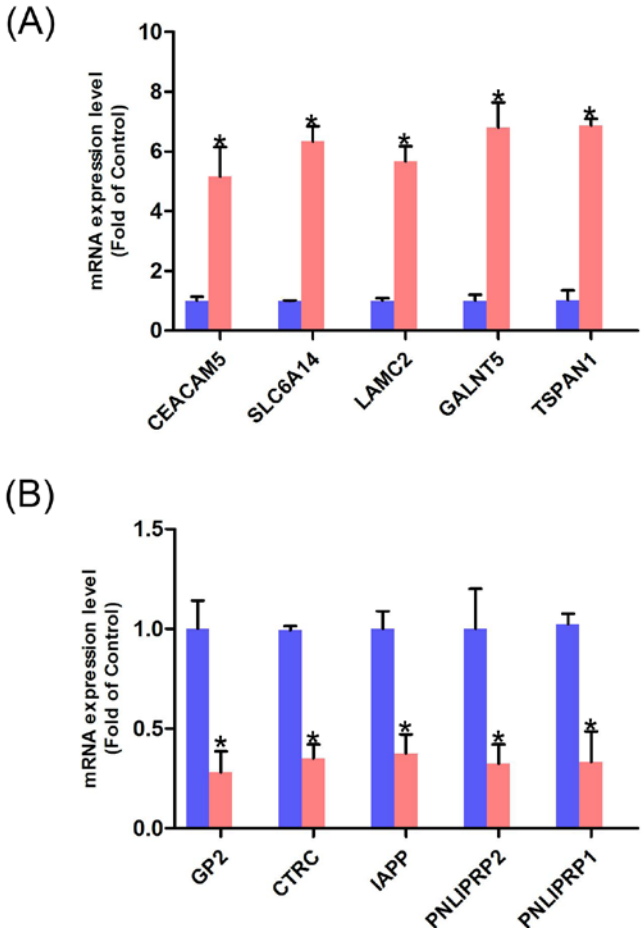


**Figure 7**. Re-identification of top5 up-regulated and top5 down-regulated DEGs in the samples from patients with PAAD using Q-PCR. (A) CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1 were significantly up-regulated in carcinoma group. (B) GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1 were significantly down-regulated in carcinoma tissues. * P < 0.05 versus adjacent tissue group. P < 0 05 was regarded statistically different.

### *3.8. Inhibition of TSPAN1 reduces PAAD cells proliferation*

Finally, we selected 1 of the top5 up-regulated genes (TSPAN1) to observe its impact on the proliferation of PAAD cells. To determine whether TSPAN1 blockade affects SW1990 cell proliferation, TSPAN1 in SW1990 cells were knocked down with TSPAN1 siRNA. MTS assays were performed to evaluate cell proliferation (Figure 9A,B). The result showed that transfection SW1990 cells with TSPAN1 siRNA significantly inhibited the protein expression of TSPAN1 and resulted in an obvious time-dependent reduction in cell proliferation.

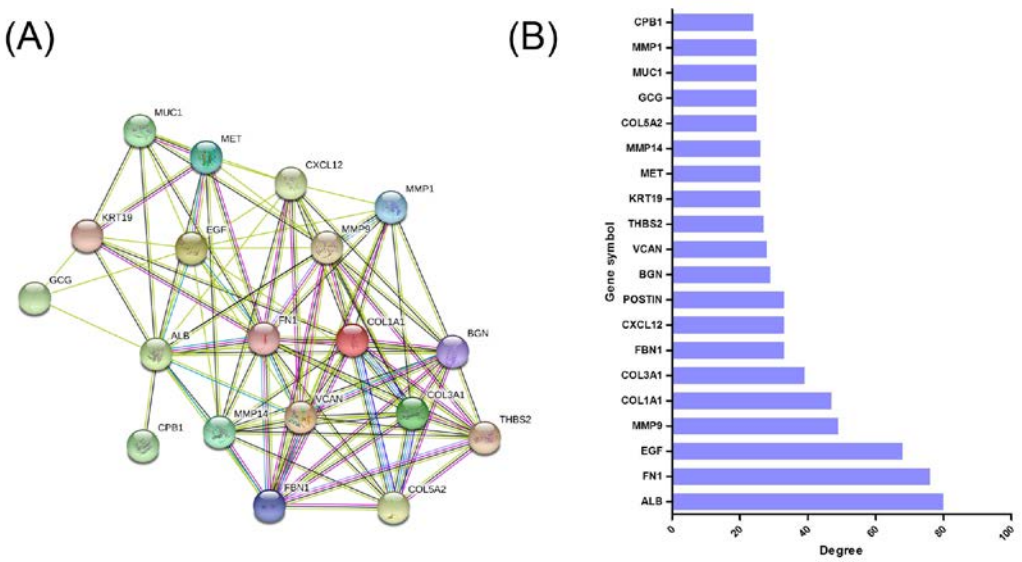


**Figure 8**. Protein-protein interaction (PPI) network. (A) The PPI network of top20 hub genes. (B) The connectivity degree of top20 hub gens.

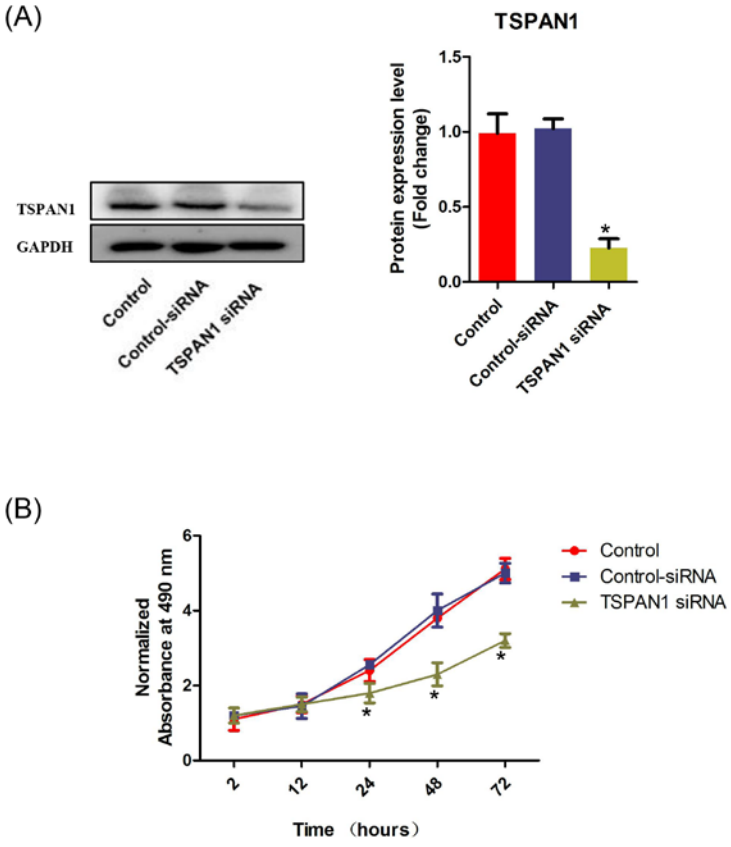


**Figure 9.** Inhibition of TSPAN1 reduces PAAD cells proliferation. (A) Western blotting analysis and quantitative data of TSPAN1 in indicated groups (n = 6). (B) Time course of MTS assay of PAAD cell proliferation. * P < 0.05 versus control group. P < 0 05 was regarded statistically different.

## 4. Discussion

In recent years, the morbidity and mortality of PAAD have been increasing globally. Although the early diagnosis and treatment have greatly developed, the overall survival rates and prognosis of PAAD is still poor, suggesting that sensitive and specific biomarkers for PAAD are urgently necessary [18]. High throughput study can develop the thorough exploration of the vital mechanisms by which lead to PAAD [19]. In our study, we identified DEGs between 45 pairs of carcinoma tissues and normal tissues from patients diagnosed with PAAD in the GEO database. In order to increase the statistical power of these DEGs, we defined that the absolute value of the logarithm (base 2) fold change (logFC) greater than 2 and a total of 444 DEGs were captured including 271 up-regulated genes and 173 down-regulated genes. In order to have an in-depth understanding of these DEGs, we employed GO function, KEGG pathway, PPI network and connectivity analysis of these DEGs, via which we screened PAAD-related genes and pathways attaching great important to cancer initiation and progression.

Our analysis screened 444 DEGs with a double fold change between carcinoma tissues and adjacent tissues from patients with PAAD. A total of 50 DEGs involving top25 up-regulated genes and top25 down-regulated genes were presented in the heatmap. From our perspective, these DEGs could serve as potential candidates for the diagnosis of PAAD in near future. At presently, some of these DEGs, in fact, have been already proved to be novel biomarkers of PAAD. For example, The CEACAM5, also known as CD66e, could code for the protein, CEA, which was first described in 1965 as a gastrointestinal oncofetal antigen, and is now found to be increased in most human carcinomas, such as the respiratory and genitourinary systems [20], gastrointestinal tract [21], as well as breast cancer [22]. As for PAAD, CEACAM 5 expressions were significantly correlated with lymph node metastasis, which gave rise to a shortened disease free survival and overall survival [23]. SLC6A14 (also known as ATB0, +) is another DEG screened by our bioinformatics analysis, which is in essence a $Na^+$ and $Cl^-$ dependent solute transport system of the SLC6 family responsible for the active transport of all neutral and basic amino acids [24]. Alan R. Penheiter et al. [25] found that SLC6A14 was significantly overexpressed at the transcriptional level in patients with PAAD, the staining of which was predominantly cytoplasmic in most resected PAAD tissues. For the reason that protein kinase C alpha (PKCα) activation could induce an obvious translocation of SLC6A14 to the membrane and enhance its transport activity [26], here, we speculated that the loss of blood supply caused by surgical resection could give rise to the inhibition of signals required for membrane localization of SLC6A14. To the best of our knowledge, our study screened the DEGs of PAAD from the angle of bioinformatics for the first time, however, the clinical practice of these DEGs needs to be further investigated.

Meanwhile, we also screened 20 hub genes of PAAD, all of which located in the core nodes in PPI network, indicating that these genes could be critical therapeutic targets to combat PAAD. For example, some evidence has characterized the effects of chemokine CXCL12 and its corresponding receptor CXCR4 in various cancers as well as their effects in local invasion and distant metastasis of PAAD [27,28]. Recent study also disclosed that CXCL12 bound to and activated chemokine receptor CXCR7 [29]. Eileen L Heinrich et al. [27] found that CXCR4 and CXCR7 were frequently co-expressed in both human pancreatic cancer cell lines and tissues in a β-arrestin-2-dependent manner, which subsequently controlled CXCL12 signals to the MAPK pathway. Additionally, CXCL12 also activated both canonical and non-canonical G protein coupled receptor signals in pancreatic cancer cell lines, eventually driveing an increase in pancreatic cancer cells proliferation.

On the other one hand, CXCL12/CXCR4 axis could also trigger sonic hedgehog expression in pancreatic cancer cells by activating extracellular regulated kinase- and Akt kinase-mediated nuclear factor-κB signals [28]. Finally, the protein expression of CXCL12 was also significantly correlated with microvessel density but not significantly correlated with microlymphatic vessel density, while CXCR4 expression was significantly associated with microlymphatic vessel density but not significantly associated with microvessel density [30]. Our study also showed that the expression of CXCL12 was significantly repressed in carcinoma tissues from patients with PAAD, and the connectivity degree of CXCL12 was relatively high among all the DEGs, unveiling its importance in the development of PAAD.

Our GO and KEGG analysis also showed that most of these DEGs were mainly enriched in pathways implicated with pancreatic secretion, protein digestion and absorption, fat digestion and absorption, glycerolipid metabolism and PI3K/Akt signaling pathway. Beyond our expectation, most abnormally expressed genes were associated with pancreatic digestive function but not the oncogenes, suggesting that biomarkers in pancreatic juice may be more promising than those in carcinoma tissues for the diagnosis of PAAD.

Although a majority of the identified connectivity genes have been tried as therapeutic targets, no obvious success was obtained in pancreatic cancer. In particular, genes such as EGF, MMP9, MUC1, MET are not actionable targets in pancreatic cancer. To allow these biomakers and targets to be used more routinely in clinic, we need to perform further research in the future to clarify the underlying mechanisms of these differences in PAAD.

## 5. Conclusion

In conclusion, we provided a comprehensive and novel analysis of gene expression profiles patients with PAAD. The top5 down-regulated genes including GP2, CTRC, IAPP, PNLIPRP2 and PNLIPRP1 and the top5 up-regulated genes including CEACAM5, SLC6A14, LAMC2, GALNT5 and TSPAN1, are expected to novel biomarkers in diagnosis of PAAD. Meanwhile, we also screened the top 20 hub genes involving ALB, FN1, EGF, MMP9, COL1A1, COL3A1, FBN1, CXCL12, POSTIN, BGN, VCAN, THBS2, KRT19, MET, MMP14, COL5A2, GCG, MUC1, MMP1 and CPB1, which could be potential targets of PAAD. Additionally, genes involved in pancreatic digestive function were also significantly altered in patients with PAAD. Anyway, this analysis may offer the powerful evidence and clues for the future genomic individualized treatment of PAAD.

## Conflict of Interest

The authors declare they have no conflicts of interest.

## Reference

1. V. Canzoniert, M. Berretta, A. Buonadonna, et al., Solid pseudopapillary tumour of the pancreas, *Lancet Oncol.*, **4** (2003), 225-6.
2. M. Noë, N. Rezaee, K. Asrani, et al., Immunolabeling of Cleared Human Pancreata Provides Insights into Three-dimensional Pancreatic Anatomy and Pathology, *Am. J. Pathol.*, **188** (2018), 1530–1535.

3. N. Martinez-Bosch, J. Vinaixa and P. Navarro, Immune Evasion in Pancreatic Cancer: From Mechanisms to Therapy, *Cancers,* **10** (2018), 6.
4. C. Napoli, N. Sperandio, R. T. Lawlor, et al., Urine metabolic signature of pancreatic ductal adenocarcinoma by $^{1}$H nuclear magnetic resonance: Identification, mapping, and evolution, *J. Proteome Res*., **11** (2011), 1274–1283.
5. B. J. Raphael, R. H. Hruban, A. J. Aguirre, et al., Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma, *Cancer Cell*, **32** (2017), 185–203.
6. V. P. Groot, N. Rezaee, W. C. Wu, et al., Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma, *Ann. Surg.*, **267** (2018), 936–945.
7. J. Kim, W. R. Bamlet, A. L. Oberg, et al., Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers, *Sci. Trans. Med.*, **9** (2017), eaah5583.
8. P. T. Nelson, D. A. Baldwin, L. M. Scearce, et al., Microarray-based, high-throughput gene expression profiling of microRNAs, *Nat. Methods,* **1** (2004), 155–161.
9. N. Li, L. Li, and Y. Chen, The Identification of Core Gene Expression Signature in Hepatocellular Carcinoma, *Oxid. Med. Cell. Longevity*, **2018** (2018), 1–15.
10. M. A. Fernández, C. Rueda and S. D. Peddada, Identification of a core set of signature cell cycle genes whose relative order of time to peak expression is conserved across species, *Nucleic Acids Res.*, **40** (2012), 2823–2832.
11. S. Yao, and T. Liu, Analysis of differential gene expression caused by cervical intraepithelial neoplasia based on GEO database, *Oncol. Lett.*, **15** (2018), 8319–8324.
12. M. D. Young, M. J. Wakefield, J. K. Smyth, et al., Gene ontology analysis for RNA-seq: Accounting for selection bias, *Genome Biol.*, **11** (2010), R14.
13. E. Altermann, and T. R. Klaenhammer, Pathway Voyager: Pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, *BMC Genomics,* **6** (2005), 60.
14. G. Dennis, B. T. Sherman, D. A. Hosack, et al., DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biol*., **4** (2003), R60.
15. Z. Tang, C. Li, B. Kang, et al., GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses, *Nucleic Acids Res.*, **45** (2017), W98–W102.
16. H. Feng, J. Chen, H. Wang, et al., Roflumilast reverses polymicrobial sepsis-induced liver damage by inhibiting inflammation in mice, *Lab. Invest*., **97** (2017), 1008–1019.
17. Z. Zhou, Y. Cheng, Y. Jiang, et al., Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis, *Int. J. Biol. Sci.*, **14** (2018), 124–136.
18. A. Adamska, A. Domenichini, and M. Falasca, Pancreatic Ductal Adenocarcinoma: Current and Evolving Therapies, *Int. J. Mol. Sci.*, **18** (2017), 1338.
19. E. S. Knudsen, P. Vail, U. Balaji, et al., Stratification of Pancreatic Ductal Adenocarcinoma: Combinatorial Genetic, Stromal, and Immunological Markers, *Clin. Cancer Res.*, **23** (2017), 4429–4440.
20. S. Govindan, T. Cardillo, C. D'Souza, et al., Therapy of human colonic and lung cancer xenografts with SN-38 conjugates of anti-CEACAM5 and anti-EGP-1 humanized monoclonal antibodies, *Mol. Cancer Ther.*, **6** (2007).
21. R. D. Blumenthal, E. Leon, H. J. Hansen, et al., Expression patterns of CEACAM5 and CEACAM6 in primary and metastatic cancers, *BMC Cancer,* **7** (2007), 2.

22. E. Powell, J. Shao, H. M. Picon, et al., A functional genomic screen in vivo identifies CEACAM5 as a clinically relevant driver of breast cancer metastasis, *NPJ. Breast Cancer,* **4** (2018), 9.
23. F. Gebauer, D. Wicklein, J. Horst, et al., Carcinoembryonic antigen-related cell adhesion molecules (CEACAM) 1, 5 and 6 as biomarkers in pancreatic cancer, *Plos One*, **9** (2014), e113023.
24. S. Bröer and U. Gether, The solute carrier 6 family of transporters, *Br. J. Pharmacol.*, **167** (2012), 256–278.
25. A. R. Penheiter, S. Erdogan, S. J. Murphy, et al., Transcriptomic and Immunohistochemical Profiling of SLC6A14 in Pancreatic Ductal Adenocarcinoma, *Biomed. Res. Int.*, **2015** (2015), 593572.
26. L. Samluk, M. Czeredys, K. Skowronek, et al., Protein kinase C regulates amino acid transporter $ATB^{0,+}$, *Biochem. Biophys. Res. Commun.*, **422** (2012), 64–69.
27. A. P. Singh, S. Arora, A. Bhardwaj, et al., CXCL12/CXCR4 protein signaling axis induces sonic hedgehog expression in pancreatic cancer cells via extracellular regulated kinase- and Akt kinase-mediated activation of nuclear factor κB: Implications for bidirectional tumor-stromal interactions, *J. Biol. Chem.*, **287** (2012), 39115–39124.
28. E. L. Heinrich, W. Lee, J. Lu, et al., Chemokine CXCL12 activates dual CXCR4 and CXCR7-mediated signaling pathways in pancreatic cancer cells, *J. Transl. Med.*, **10** (2012), 68.
29. F. Sieero, C.Biben, L. Martínez-Muñoz, et al., Disrupted cardiac development but normal hematopoiesis in mice deficient in the second CXCL12/SDF-1 receptor, CXCR7, *Proc. Natl. Acad. Sci. U. S. A.*, **104** (2007), 14759–14764.
30. K. Cui, W. Zhao, C. Wang, et al., The CXCR4-CXCL12 Pathway Facilitates the Progression of Pancreatic Cancer Via Induction of Angiogenesis and Lymphangiogenesis, *J. Surg. Res.*, **171** (2011), 143–150.


AIMS Press

©2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)