

Identification of key regulators of pancreatic ductal adenocarcinoma using bioinformatics analysis of microarray data

Nan Li, PhD^a, Xin Zhao, PhD^b, Shengyi You, PhD^{a,*}

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal forms of cancer, and its etiology remains largely unknown. This study aimed to screen a panel of key genes and to identify their potential impact on the molecular pathways associated with the development of PDAC. Four gene expression profiles, GSE28735, GSE15471, GSE102238, and GSE43795, were downloaded from the Gene Expression Omnibus (GEO) database. The intersection of the differentially expressed genes (DEGs) in each dataset was obtained using Venn analysis. Gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) analysis were subsequently carried out. To screen for hub genes, a protein–protein interaction (PPI) network was constructed.

The intersection of the DEGs revealed 7 upregulated and 9 downregulated genes. Upon relaxation of the selection criteria, 58 upregulated and 32 downregulated DEGs were identified. The top 5 biological processes identified by GO analysis involved peptide cross-linking, extracellular matrix (ECM) disassembly, regulation of the fibroblast growth factor receptor signaling pathway, mesoderm morphogenesis, and lipid digestion. The results of KEGG analysis revealed that the DEGs were significantly enriched in pathways involved in protein digestion and absorption, ECM-receptor interaction, pancreatic secretion, and fat digestion and absorption. The top ten hub genes were identified based on the PPI network.

In conclusion, the identified hub genes may contribute to the elucidation of the underlying molecular mechanisms of PDAC and serve as promising candidates that can be utilized for the early diagnosis and prognostic prediction of PDAC. However, further experimental validation is required to confirm these results.

Abbreviations: ALB = albumin, BP = biological process, CC = cellular component, CEL = carboxyl ester lipase, CELA3A = chymotrypsin like elastase family member 3A, CELA3B = chymotrypsin like elastase family member 3B, COL1A1 = collagen type I alpha 1, COL1A2 = collagen type I alpha 2 chain, COL3A1 = collagen type III alpha 1 chain, COL5A2 = collagen type V alpha 2 chain, DEGs = differentially expressed genes, DRG = dorsal root ganglia, ECM = extracellular matrix, FGFs = fibroblast growth factors, FN = fibronectin, FN1 = fibronectin 1, GEO = Gene Expression Omnibus, GO = Gene Ontology, ITGA2 = integrin subunit alpha 2, KEGG = Kyoto Encyclopedia of Genes and Genomes, MCODE = molecular complex detection, MF = molecular function, MMP1 = matrix metalloproteinase 1, MODY = maturity-onset diabetes of the young, PDAC = pancreatic ductal adenocarcinoma, PPI = protein–protein interaction, STRING = search tool for the retrieval of interacting genes.

Keywords: bioinformatics analysis, differentially expressed gene, microarray, pancreatic ductal adenocarcinoma

1. Introduction

Pancreatic cancer is currently the third leading cause of cancer death; based on the latest cancer statistics, it could lead to approximately 79,400 deaths in China and 330,400 deaths every year worldwide and is predicted to become the second leading

cause of cancer death by 2020.^[1,2] The median survival of all pancreatic ductal adenocarcinoma (PDAC) patients is less than 6 months, and only 6% of patients are still alive 5 years after diagnosis. Despite the rapid developments in medicine in recent years, the therapies used to treat PDAC are inadequate and its prognosis remains unoptimistic. Moreover, it is almost impossible to detect tumors, as metastasis often take place during the early stages.^[3] Accumulated evidence has demonstrated that multiple genes and cellular pathways are closely linked to the occurrence and development of pancreatic carcinoma. Therefore, the development of strategies for early PDAC detection that allow for a surgical cure and increase understanding of the underlying molecular mechanisms that are involved in PDAC is extremely important and desperately needed.

To date, a lack of knowledge regarding the precise molecular mechanisms underlying pancreatic cancer progression has limited the ability to treat advanced disease. There has been widespread research of the genetic biomarkers for PDAC that has utilized high-throughput platforms to analyze gene expression, such as microarrays, that have been increasingly valued as promising tools to identify the genes at the hub of PDAC progression.^[4] Previous evidence has identified several genetic markers of pancreatic cancer that have prognostic and therapeutic significance.^[5,6] Zhang et al

Editor: Bernhard Schaller.

The authors have no conflicts of interest to disclose.

This work was supported by the Key Program sponsored by the Tianjin Science and Technology Committee, China (12ZCZDSY03500).

NL and XZ contributed equally to this work.

^a Department of General Surgery, Tianjin Medical University General Hospital,

^b Tianjin Medical University, Tianjin, China.

* Correspondence: Shengyi You, Anshan Road 154, Tianjin 300052, China (e-mail: shengyiyu126@126.com).

Copyright © 2019 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medicine (2019) 98:2(e14074)

Received: 8 October 2017 / Received in final form: 29 November 2018 /

Accepted: 18 December 2018

<http://dx.doi.org/10.1097/MD.0000000000014074>

investigated gene-metabolite networks by integrating metabolomics and transcriptomics and demonstrated that impairment of a lipolytic pathway that involves lipases and a unique set of free fatty acids may play an important role in the development and progression of PDAC.^[7] Genes having different expression levels in PDAC have been linked to growth factor receptors, tyrosine kinase inhibitors, mitogen-activated protein kinases, the rapamycin blockade, and the phosphatidylinositol 3-kinase and HER2-neu pathways.^[8] Additionally, the cell cycle and p53 signaling pathways may play significant roles in pancreatic carcinoma, and differentially expressed genes (DEGs), such as cell division cycle 20, the BUB1 mitotic checkpoint serine/threonine kinase B, cyclin B1, and pituitary tumor-transforming 1, may be potential targets that could be used in the diagnosis and treatment of PDAC.^[9] Due to the important role it plays in tumorigenesis, the Ras signaling pathway could also function as an anti-cancer target.^[10] Moreover, it has been verified that many other pathways have important effects on tumorigenesis.^[11,12] However, the mechanisms underlying the effects of DEGs, particularly those involved in the signal pathways in the association networks implicated in PDAC, remain to be elucidated.

To analyze and enhance understanding of the underlying molecular mechanisms of PDAC, we downloaded microarray data for GSE28735, GSE15471, GSE102238, and GSE43795 from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and identified the DEGs in PDAC tissues and adjacent non-tumor tissues. We identified the common and significant DEGs identified in other studies and selected DEGs present in at least 3 of the microarray datasets for inclusion in gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment pathway analysis. Finally, the protein-protein interaction (PPI) networks were constructed. By analyzing their biological functions and associated signaling pathways, this study is likely to shed light on PDAC development at the molecular level and may identify potential candidate biomarkers that may aid in PDAC treatment.

2. Materials and methods

2.1. Microarray data

The microarray gene expression profile GSE28735,^[5,7] GSE15471,^[13] GSE102238, and GSE43795^[14] related to PDAC were downloaded from GEO (<http://www.ncbi.nlm.nih.gov/geo/>), which were based on the platform of GPL6244 Affymetrix Human Gene 1.0 ST Array, GPL570 Affymetrix Human Genome U133 Plus 2.0 Array, GPL19072 Agilent-052909 CBC_lncRNAmRNA_V3, and GPL10558 Illumina HumanHT-12 V4.0 expression beadchip, respectively. GSE28735 dataset contained 45 PDAC tumors and 45 matching normal pancreatic tissue samples; GSE15471 contained 39 PDAC samples and 39 normal controls; GSE102238 contained 22 PDAC samples and 22 normal controls; and GSE43795 contained 6 PDAC samples and 5 normal controls were selected for subsequent analyses. Ethical approval was not necessary in our study because we downloaded the expression profiles from the public database and do not perform any experiments in patients or animals.

2.2. Data preprocessing and DEG identification

The original data were subject to background correction, quantile normalization, and log transformation using the robust multiarray average algorithm.^[15] Preprocessing included background correction, quantile normalization, and the transformation of probe IDs into gene symbols. The probes were mapped to genes in NCBI

Entrez using the Gene ID converter.^[16] If there were multiple probes that corresponded to the same gene, their average value was calculated and used as the final gene expression level.

The raw expression data were screened using the limma package in R/Bioconductor in the GEO2R web tool.^[17,18] The *P*-values, which were used to assess the significance of the differences, were calculated using a paired samples *t* test. The obtained *P*-values were adjusted using the Benjamini and Hochberg method.^[19] The thresholds used were: an adjusted *P* value < .05, *P* < .05, and $|\log_2\text{-fold change}| > 2.0$.

2.3. Acquisition of the intersection of the DEGs

The intersection of the DEGs, which originated from the 4 expression profiles (GSE28735, GSE15471, GSE102238, and GSE43795), was obtained using Venn analysis.^[20] The intersection of the DEGs that originated from the 4 GEO datasets was used in the subsequent analysis.

2.4. Functional enrichment analysis of the DEGs

GO analysis (<http://www.geneontology.org/>) was used to predict the potential functions of the target genes based on molecular function (MF), cellular components (CC), and biological processes (BP).^[21] The KEGG (<http://www.genome.jp/>) is a knowledge base that is used for systematic analysis of gene functions by linking genomic information with higher-level systemic functions.^[22] ClueGO is a plugin of Cytoscape that is used to facilitate biological interpretation and to visualize functionally grouped terms in network form.^[23] In the present study, we performed KEGG pathway enrichment analysis using the ClueGO plugin with a *P*-value < .05 and a kappa score = 0.4. The DEGs that were found to be enriched within various pathways were visualized using the CluePedia plugin^[24] in Cytoscape. For both analyses, a *P*-value of < .05 was considered to denote statistical significance.

2.5. Integration of PPI network analysis

The search tool for the retrieval of interacting genes (STRING) database is an online tool that is designed to evaluate PPI information.^[25] In this study, DEGs with coexpression coefficients greater than 0.4 were extracted from the STRING database. Thereafter, the PPI networks were constructed using Cytoscape software.^[26] The Molecular Complex Detection plugin (MCODE) was used to screen the PPI network modules in Cytoscape; an MCODE score > 5 was used as the selection criterion.

3. Results

3.1. Identification of DEGs from the 4 selected GEO datasets

We aimed to compare the gene expression profiles of PDAC and normal pancreatic tissues. Four GEO datasets (GSE28735, GSE15471, GSE102238, and GSE43795) were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). The main characteristics of these 4 studies are shown in Table 1. We then determined which DEGs were significantly upregulated or downregulated in each study. Compared to the control group, 22 upregulated and 29 downregulated DEGs were identified from GSE28735, 232 upregulated DEGs and 36 downregulated DEGs were identified from GSE15471, 187 upregulated and

Table 1
The main features of 4 studies of gene expression microarray data.

| GEO accession | Platform | Case/control | Sample type | Year |
|---------------|----------|--------------|-------------|------|
| GSE28735 | GPL6244 | 45/45 | Pancreas | 2012 |
| GSE15471 | GPL570 | 39/39 | Pancreas | 2009 |
| GSE102238 | GPL19072 | 22/22 | Pancreas | 2017 |
| GSE43795 | GPL10558 | 6/5 | Pancreas | 2013 |

GEO = Gene Expression Omnibus.

89 downregulated DEGs were identified from GSE102238, and 494 upregulated DEGs and 455 downregulated DEGs were identified from GSE43795. The top 10 DEGs from each dataset are listed in Table 2.

3.2. The intersection of the DEGs

To identify the DEGs that play a role in PDAC, the consistently upregulated or downregulated genes were identified using Venn analysis; a Venn diagram created using Venny2.1.0 is shown in

Table 2
The most significant up- and downregulated DEGs in GSE28735, GSE15471, GSE102238, and GSE43795.

| GSE28735 | | | | GSE15471 | | | |
|----------------------|-------------|--------------|----------|----------------------|-------------|--------------|----------|
| Gene symbol | Fold change | Adj. P value | P value | Gene symbol | Fold change | Adj. P value | P value |
| Upregulated | | | | Upregulated | | | |
| CEACAM5 | 3.181287 | 1.07E-10 | 2.16E-13 | INHBA | 5.184665 | 9.65E-19 | 7.06E-23 |
| SLC6A14 | 3.075768 | 3.82E-12 | 2.65E-15 | POSTN | 5.124611 | 8.21E-15 | 1.05E-17 |
| LAMC2 | 2.901654 | 1.59E-14 | 7.03E-19 | COL10A1 | 5.047442 | 1.33E-18 | 1.21E-22 |
| GALNT5 | 2.789465 | 8.70E-13 | 3.62E-16 | POSTN | 4.898659 | 3.45E-15 | 3.60E-18 |
| TSPAN1 | 2.726684 | 1.63E-14 | 2.49E-18 | SULF1 | 4.648408 | 1.66E-19 | 9.10E-24 |
| CTSE | 2.684101 | 6.93E-12 | 6.00E-15 | COL8A1 | 4.626125 | 3.55E-18 | 4.54E-22 |
| POSTN | 2.629837 | 2.53E-10 | 6.76E-13 | CTHRC1 | 4.430754 | 6.72E-17 | 2.49E-20 |
| CEACAM6 | 2.627122 | 3.78E-10 | 1.06E-12 | COL11A1 | 4.416254 | 4.13E-15 | 4.54E-18 |
| ANXA10 | 2.357937 | 7.58E-09 | 5.23E-11 | SULF1 | 4.397156 | 1.31E-19 | 4.81E-24 |
| LAMB3 | 2.34423 | 1.59E-14 | 1.19E-18 | COL1A1 | 4.250822 | 2.69E-16 | 1.46E-19 |
| Downregulated | | | | Downregulated | | | |
| PNLIPRP1 | -3.02666 | 3.86E-05 | 2.37E-06 | ALB | -3.38302 | 1.42E-04 | 5.75E-05 |
| PNLIPRP2 | -2.82015 | 7.56E-05 | 5.36E-06 | SPX | -3.24613 | 3.66E-07 | 4.23E-08 |
| IAPP | -2.80976 | 1.53E-07 | 2.36E-09 | CTRL | -3.00886 | 7.59E-04 | 3.70E-04 |
| CTRC | -2.79822 | 5.30E-05 | 3.47E-06 | SERPINI2 | -2.97325 | 1.63E-03 | 8.59E-04 |
| GP2 | -2.79381 | 1.12E-04 | 8.79E-06 | TEX11 | -2.91789 | 2.06E-05 | 6.31E-06 |
| CEL | -2.71677 | 1.39E-04 | 1.14E-05 | AQP8 | -2.88069 | 7.06E-05 | 2.60E-05 |
| CPA2 | -2.65023 | 2.47E-04 | 2.38E-05 | TMED6 | -2.76898 | 9.89E-05 | 3.81E-05 |
| ALB | -2.6311 | 4.12E-06 | 1.43E-07 | GNMT | -2.74551 | 1.55E-06 | 2.76E-07 |
| FAM24B-CUZD1 | -2.62588 | 2.01E-04 | 1.84E-05 | TEX11 | -2.71543 | 1.58E-05 | 4.63E-06 |
| ERP27 | -2.58858 | 2.88E-05 | 1.64E-06 | DNASE1 | -2.697 | 2.42E-09 | 4.93E-11 |
| GSE102238 | | | | GSE43795 | | | |
| Gene symbol | Fold change | Adj. P value | P value | Gene symbol | Fold change | Adj. P value | P value |
| Upregulated | | | | Upregulated | | | |
| CST2 | 5.380273 | 6.10E-08 | 2.07E-11 | S100P | 5.93878 | 4.89E-04 | 1.94E-05 |
| CST1 | 5.208182 | 6.10E-08 | 1.31E-11 | PLAC8 | 5.739866 | 5.60E-07 | 1.53E-09 |
| CEACAM5 | 4.875727 | 3.40E-05 | 8.78E-07 | GCNT3 | 5.261017 | 1.81E-03 | 1.21E-04 |
| MSLN | 4.851955 | 1.99E-07 | 2.65E-10 | LCN2 | 5.257575 | 5.61E-05 | 9.42E-07 |
| S100P | 4.684682 | 1.15E-06 | 4.19E-09 | CST1 | 5.121328 | 1.14E-06 | 4.23E-09 |
| TRIM29 | 4.322046 | 6.66E-08 | 3.66E-11 | LAMC2 | 4.99819 | 5.03E-05 | 8.03E-07 |
| KLK6 | 4.315727 | 6.10E-08 | 1.65E-11 | SLC16A3 | 4.997346 | 8.54E-07 | 2.68E-09 |
| CEACAM6 | 4.196818 | 1.09E-05 | 1.57E-07 | CCL20 | 4.988855 | 2.08E-04 | 5.94E-06 |
| CEMIP | 4.159727 | 1.69E-07 | 1.86E-10 | SLPI | 4.782297 | 1.26E-04 | 2.92E-06 |
| COL10A1 | 4.015227 | 6.10E-08 | 1.96E-11 | CEACAM6 | 4.744282 | 1.93E-03 | 1.32E-04 |
| Downregulated | | | | Downregulated | | | |
| ALB | -4.25118 | 1.47E-03 | 2.33E-04 | CTRB1 | -7.42405 | 1.02E-05 | 8.39E-08 |
| SPX | -3.85518 | 2.94E-04 | 2.29E-05 | PRSS3P2 | -7.40334 | 1.38E-07 | 2.49E-10 |
| SERPINI2 | -3.77605 | 3.65E-03 | 8.14E-04 | SERPINI2 | -7.39187 | 2.65E-06 | 1.29E-08 |
| CTRL | -3.65573 | 2.63E-03 | 5.21E-04 | PNLIPRP1 | -7.32253 | 3.79E-05 | 5.34E-07 |
| CELA2A | -3.31768 | 1.19E-02 | 4.01E-03 | CTRB2 | -7.16668 | 1.80E-04 | 4.80E-06 |
| PNLIPRP2 | -3.30573 | 8.98E-03 | 2.75E-03 | CELA2B | -7.12215 | 2.61E-05 | 3.18E-07 |
| ERP27 | -3.30291 | 1.12E-03 | 1.58E-04 | AQP8 | -7.1176 | 2.59E-07 | 5.54E-10 |
| AQP12A | -3.29168 | 4.58E-04 | 4.42E-05 | ALB | -7.01679 | 2.93E-07 | 6.51E-10 |
| PNLIPRP1 | -3.24995 | 1.26E-02 | 4.35E-03 | CTRL | -6.9817 | 3.73E-07 | 9.02E-10 |
| PNLIP | -3.24636 | 9.45E-03 | 2.95E-03 | CPA1 | -6.90112 | 2.35E-03 | 1.74E-04 |

DEGs = differentially expressed gene.

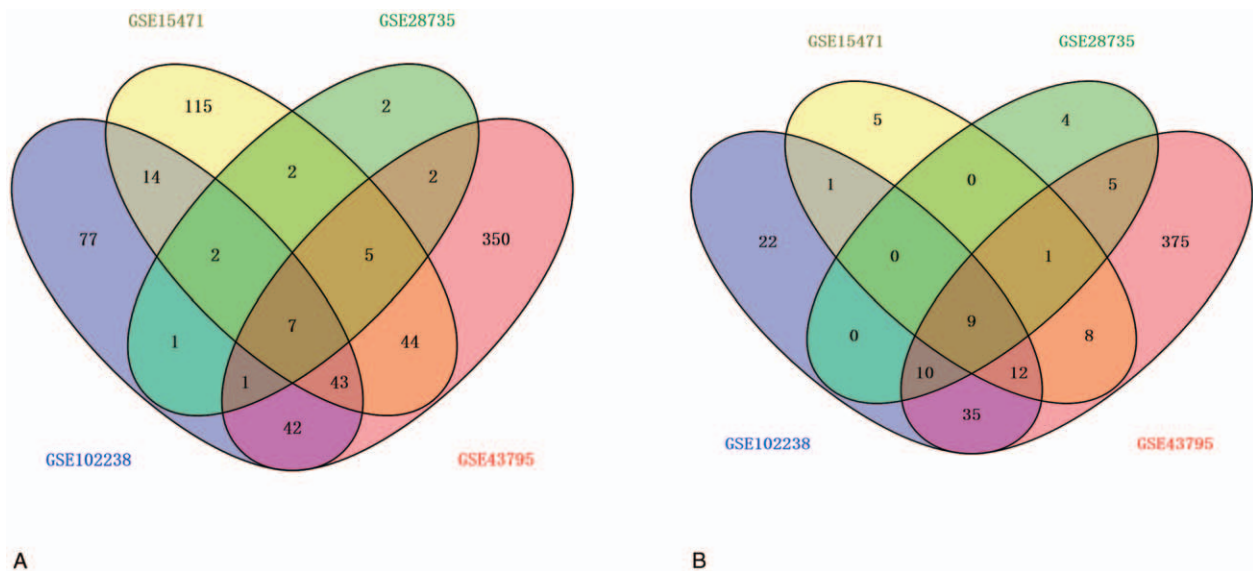


Figure 1. Intersection of the up- and downregulated differentially expressed gene from GSE28735, GSE15471, GSE102238, and GSE43795. (A) Upregulated DEGs. (B) Downregulated DEGs. Each oval represents a study. Numbers in each overlapped area means numbers of differently expression genes in each area. The brown intersection in the middle represents genes which is significantly differentially expressed in 4 microarray datasets consistently. DEGs = differentially expressed gene.

Figure 1. A total of 16 shared DEGs, including 7 upregulated and 9 downregulated DEGs, were found to be consistently and significantly differentially expressed in 4 microarrays.

3.3. GO term enrichment and KEGG pathway analysis

GO analysis of the 16 DEGs that were shared across the 4 datasets indicated that they were significantly enriched in the extracellular region and the extracellular space of the CC. The results of the KEGG analysis revealed that the 16 shared genes were enriched in the pancreatic secretion pathway. Upon relaxing the criterion used for selection, we identified more DEGs.

As shown in Figure 1, 58 overlapping upregulated and 32 overlapping downregulated DEGs in PDAC and control samples were identified in at least 3 of the microarray datasets. Functional enrichment analysis of these DEGs revealed that they were associated with multiple BPs. The GO analysis revealed that the upregulated DEGs were significantly enriched in a number of BP, including peptide cross-linking, extracellular matrix (ECM) disassembly, regulation of the fibroblast growth factor receptor signaling pathway, mesoderm morphogenesis, muscle cell migration, ECM organization, and endodermal cell differentiation; the downregulated DEGs were significantly enriched in BPs involved in lipid digestion. As for CC, the upregulated DEGs were significantly enriched in proteinaceous ECM, the specific granule lumen, ECM components, and collagen trimers. Moreover, GO MF analysis indicated that the upregulated DEGs were significantly enriched in BPs involved in protease binding, collagen binding, and platelet-derivation; the downregulated DEGs were significantly enriched in BPs involved in the triglyceride lipase activity (Fig. 2). The results of the KEGG analysis revealed that the DEGs were significantly enriched in processes involved in protein digestion and absorption, ECM-receptor interactions, pancreatic secretion, and fat digestion and absorption (Table 3).

3.4. Module screening of the PPI network and pathway enrichment analysis

Based on information found in the STRING database, the gene interaction network was shown to contain 89 nodes and 183 edges. The nodes represented the DEGs, and the edges represented the interactions among the DEGs. The Network Analyzer in Cytoscape was used to analyze these genes, (Fig. 3) and the core genes were ranked according to the predicted scores. The top 10 high-degree hub nodes included albumin (ALB), collagen type I alpha 2 chain (COL1A2), epidermal growth factor (EGF), collagen type III alpha 1 chain (COL3A1), fibronectin 1 (FN1), carboxyl ester lipase (CEL), integrin subunit alpha 2 (ITGA2), collagen type V alpha 2 chain (COL5A2), matrix metalloproteinase 1 (MMP1), and chymotrypsin-like elastase family member 3B (CELA3B). The core genes and their corresponding degrees are shown in Table 4. The distribution of the core genes in the interaction network is shown in Figure 4; the black node represents the core gene, the red line represents the fitted line, and the blue line represents the power law. The correlation between the data points and the corresponding points on the line was approximately 0.908. The *R*-squared value was 0.768, which indicated with relatively high confidence that the underlying model was linear. Subsequently, MCODE was applied in Cytoscape to screen the modules of the gene interaction network; the top 2 modules are shown in Figure 5.

4. Discussion

Currently, the early detection of pancreatic cancer remains a challenge for clinical scientists, and understanding of the molecular mechanisms underlying the progression of PDAC is still relatively poor. In this study, we utilized functional genomic data to discover targeted pathways and function-based genes, and a bioinformatics method was used to identify the key regulators involved in the development of PDAC.

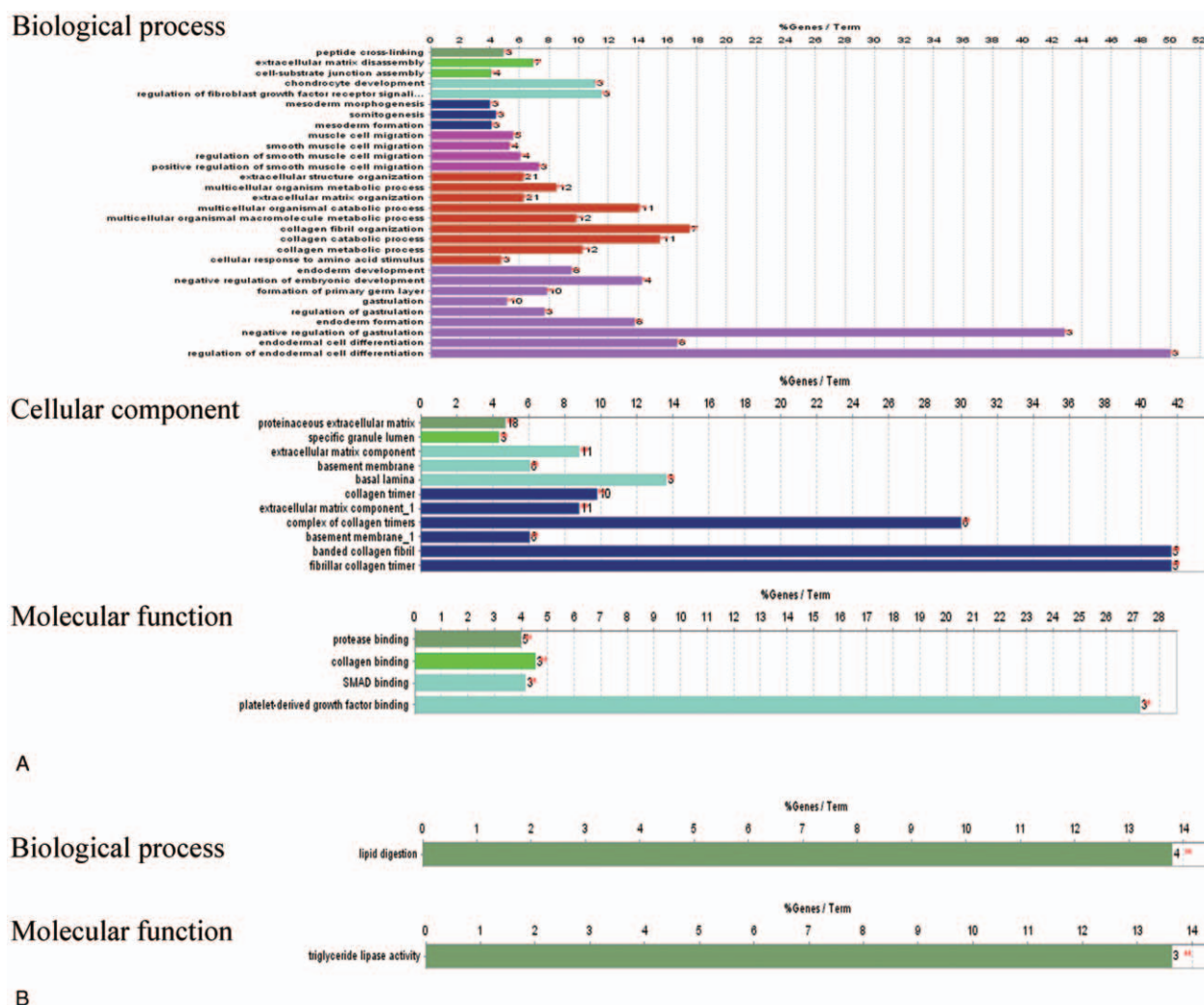


Figure 2. GO enrichment of 4 microarray datasets. GO enrichment analysis of biological process, cellular components and molecular function of upregulated DEGs in at least 3 of microarray datasets. GO enrichment analysis of biological process, cellular components and molecular function of downregulated DEGs in at least 3 of microarray datasets. GO = Gene Ontology.

In the present study, the gene expression profiles GSE28735, GSE15471, GSE102238, and GSE43795 were downloaded and a comprehensive bioinformatics analysis was performed. Using the previously defined cut-off criterion, we identified 16 shared DEGs (7 upregulated and 9 downregulated) in 4 microarrays. We then relaxed the selection criteria, which allowed us to identify more DEGs. GO function and KEGG pathway annotation of the 90

overlapped DEGs identified in at least 3 of the microarray datasets showed that these genes may play vital roles in the occurrence of PDAC. By constructing a PPI network, we also identified several hub genes that are highly correlated with the development of this disease.

GO analysis of the 16 DEGs that were shared across all 4 datasets indicated that these genes were significantly enriched in

Table 3
KEGG pathway analysis of differentially expressed genes.

| ID | Description | P value | Counts | Genes |
|---------------|----------------------------------|---------|--------|--|
| Upregulated | | | | |
| GO:0004974 | Protein digestion and absorption | 7.8E-8 | 7 | COL10A1, COL11A1, COL12A1, COL1A2, COL3A1, COL5A1, COL5A2 |
| GO:0004512 | ECM-receptor interaction | 2.2E-5 | 5 | COL1A2, FN1, ITGA2, LAMA3, LAMC2 |
| Downregulated | | | | |
| GO:0004972 | Pancreatic secretion | 6.2E-14 | 10 | CEL, CELA2A, CELA2B, CELA3B, CPA1, CPA2, CTRL, PLA2G1B, PNLIPRP1, PNLIPRP2 |
| GO:0004974 | Protein digestion and absorption | 2.0E-7 | 6 | CELA2A, CELA2B, CELA3B, CPA1, CPA2, CTRL |
| GO:0004975 | Fat digestion and absorption | 1.1E-7 | 5 | CEL, CLPS, PLA2G1B, PNLIPRP1, PNLIPRP2 |

CEL = carboxyl ester lipase, CELA3B = chymotrypsin-like elastase family member 3B, COL1A2 = collagen type I alpha 2 chain, COL3A1 = collagen type III alpha 1 chain, COL5A2 = collagen type V alpha 2 chain, COL11A1 = collagen type XI alpha 1, ECM = extracellular matrix, FN1 = fibronectin 1, ITGA2 = integrin subunit alpha 2, KEGG = Kyoto Encyclopedia of Genes and Genomes.

| Gene | Degree | Gene | Degree |
|--------|--------|--------|--------|
| ALB | 23 | CEL | 14 |
| COL1A2 | 17 | ITGA2 | 13 |
| EGF | 16 | COL5A2 | 12 |
| COL3A1 | 15 | MMP1 | 11 |
| FN1 | 14 | CELA3B | 11 |

ALB = albumin, CEL = carboxyl ester lipase, CELA3B = chymotrypsin-like elastase family member 3B, COL1A2 = collagen type I alpha 2 chain, COL3A1 = collagen type III alpha 1 chain, COL5A2 = collagen type V alpha 2 chain, EGF = epidermal growth factor, FN1 = fibronectin 1, ITGA2 = integrin subunit alpha 2, MMP1 matrix metalloproteinase 1.

been reported that protein digestion and absorption are implicated in the development of pancreatic neuroendocrine tumors.^[33] An improved understanding of metabolic dysregulation in pancreatic cancer could lead to the discovery of novel therapeutic targets.^[34] Therefore, GO analysis could help to identify the BPs involved in the occurrence of PDAC, as well as the related MFs and CCs. Moreover, it may provide evidence of the impact of the functioning of various genes on metabolism and metabolite networks.

The enriched KEGG pathway comprised of 90 DEGs included DEGs involved in ECM-receptor interaction, pancreatic secretion, and protein and fat digestion and absorption. Cancer cells actively take part in the production of ECM proteins, which are then deposited into the tumor stroma. In recent years, studies focused on the tumor microenvironment have suggested that the remodeling of the tumor ECM increases the invasive and metastatic capabilities of tumor cells.^[35] Maturity-onset diabetes

of the young (MODY) is a genetically defined subgroup of diabetes characterized by an autosomal dominant inheritance and early onset, non-insulin dependent diabetes that is closely related to pancreatic secretory dysfunction.^[36] Patients with CEL mutations develop pancreatic exocrine dysfunction in early childhood and eventually develop MODY.^[37] Interestingly, a study has demonstrated that CEL-MODY patients may be at an increased risk of developing pancreatic cancer.^[38] Collagen type XI alpha 1 (COL11A1) belongs to the collagen family and is a major component of the interstitial extracellular matrix.^[39] According to 1 report, high COL11A1 mRNA levels are significantly associated with poor chemo-responses and clinical outcomes in epithelial ovarian carcinoma.^[40] In a proton magnetic resonance study, it was found that lipids, choline-containing compounds, and fatty acids were decreased in pancreatic cancer tissues compared with normal pancreatic tissue.^[41] Lipids may be protumorigenic, as a high-fat diet was found to promote tumor growth in a murine model of pancreatic cancer.^[42] Indeed, lipid metabolism is likely to be an important source of energy in pancreatic cancer. However, it remains to be determined which fatty acids are cytotoxic for tumor cells and which fatty acids provide tumors with metabolic substrates. Further studies of tumor metabolism could reveal exciting possibilities for therapeutic intervention. Thus, a comprehensive understanding of the involved pathways and their internal links may aid us in elucidating the major mechanisms underlying the development of PDAC.

To obtain more evidence, we analyzed the PPI network and found that ALB, COL1A2, EGF, COL3A1, FN1, CEL, ITGA2, COL5A2, MMP1, and CELA3B were the top 10 hub genes, which may be essential to the molecular mechanisms underlying

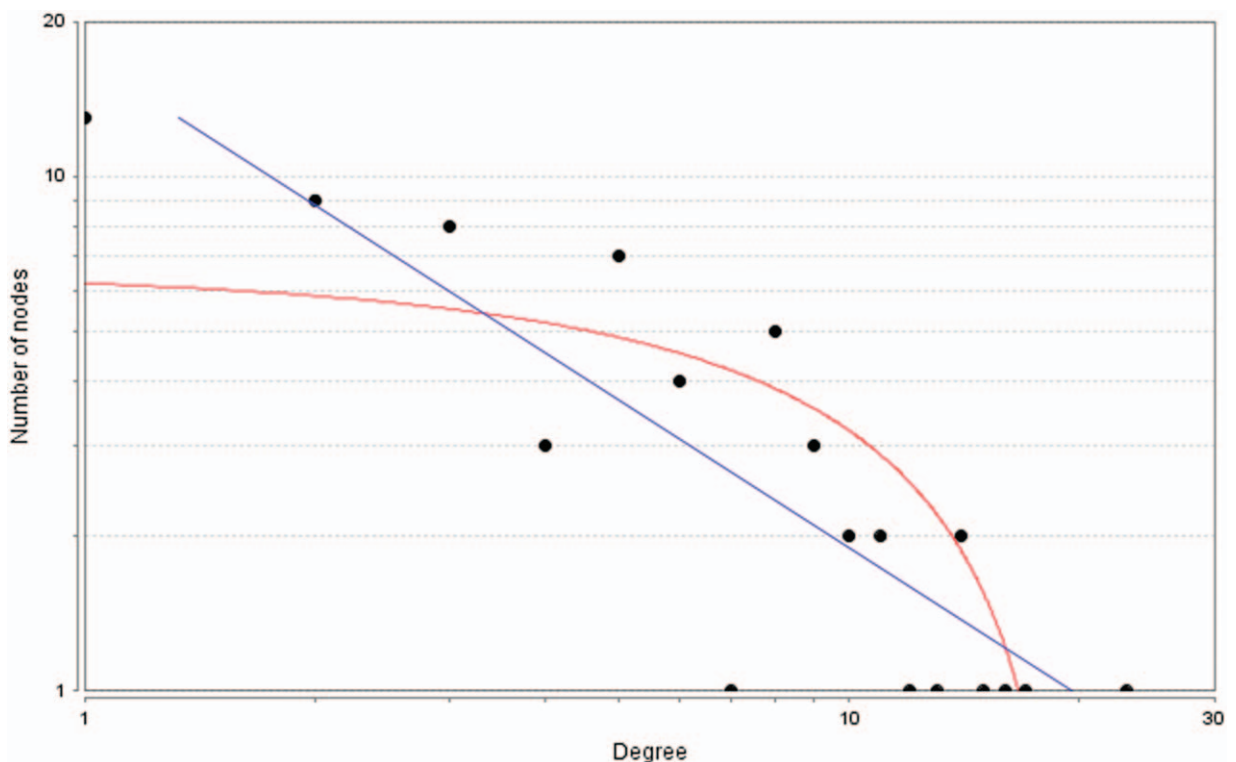


Figure 4. The distribution of core genes in the interaction network. The black node means the core gene. The red line means the fitted line, and the blue line means the power law. The correlation between the data points and corresponding points on the line is approximately 0.908. The R -squared value is 0.768, giving a relatively high confidence that the underlying model is indeed linear.

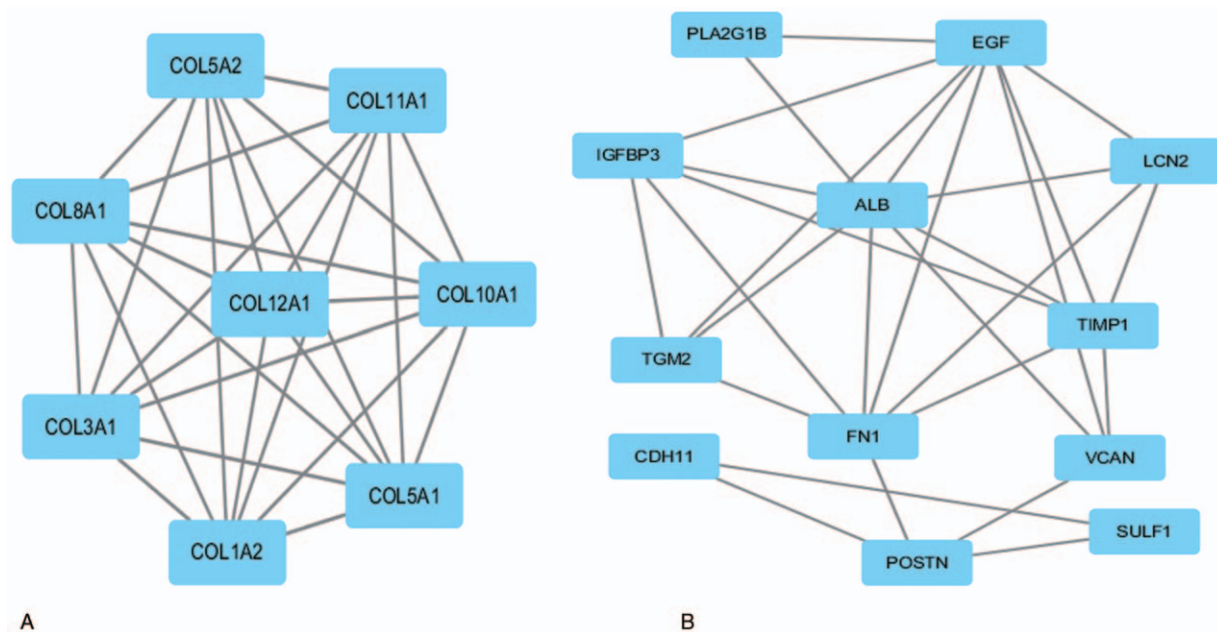


Figure 5. The modules from protein-protein interaction network. The squares represent the DEGs in modules, and the lines show the interaction between the DEGs. DEGs = differentially expressed gene.

the development of PDAC and may therefore serve as potential therapeutic targets. ALB was identified as one of the hub genes with the greatest degree of connectivity. Pant et al^[43] reported that the baseline serum ALB level is a predictive biomarker for patients with advanced PC that are treated with bevacizumab, which may be explained by a pharmacokinetic description of bevacizumab. Researchers found that fibroblast growth factors (FGF) treatment of gynecological cancer cells could promote calpain 2 activity.^[44] This finding is consistent with those of a previous study that reported that EGF increased calpain-2 activity via the activation of extracellular signal-regulated kinase pathways.^[45] The 2 most extensively studied FGFs are FGF-1 and FGF-2.^[46] FGF-2 is the prototypic heparin-binding protein and has growth-related, antiapoptotic and angiogenic activity.^[47] It has been found to be overexpressed in PDAC,^[48] and the FGF-2 receptor has been reported to be involved in tumor metastasis, stage, and retroperitoneal invasion.^[49] Fibronectin (FN), which exists in both plasma and cells, functions in cellular adhesion, spreading, and cell migration.^[50] Studies have demonstrated that FN stimulates invasion and adhesion and markedly inhibits cell death in pancreatic cancer cells.^[51,52] CEL has been demonstrated to be expressed primarily in pancreatic acinar cells and lactating mammary tissue and encodes a digestive enzyme that has been shown to play a role in cholesterol ester digestion.^[53] In addition, Ferraro et al^[54] proved that a decrease in ITGA2 could reduce cell migration in colon cancer. Reportedly, CEL-mutation carriers develop multiple pancreatic cysts during the development of diabetes,^[38] and these cysts have been shown to promote the occurrence of pancreatic cancer.^[55] Genetic analysis has shown that human chymotrypsin-like elastase family member 3A (CELA3A) and CELA3B are involved in complex formation between proelastases and procarboxypeptidases in chronic pancreatitis.^[56] It has been reported CELA3A and CELA3B are downregulated in prostate cancer,^[57] which is consistent with our results. Fibrosis is a consequence of injury that is characterized by the accumulation of excess collagen and other

ECM components that results in the destruction of normal tissue architecture and loss of functioning; COL1A2, COL3A1, and COL5A2 are genes that are involved in the maintenance of the ECM that have been shown to have a close relationship with human dermal fibroblasts.^[58] Matrix metalloproteinases (MMPs) are a group of zinc-dependent proteases involved in ECM and basement membrane degradation.^[59] A recent study showed that MMP1 activated AKT and induced protease-activated receptor-1-expressing dorsal root ganglia (DRG) to release substance P, which, in turn, activated neurokinin 1 receptor-expressing PDAC cells and enhanced cellular migration and invasion.^[60] The results of the studies of the hub genes are consistent with the results of the enriched function and pathway analysis and could provide for effective diagnostic and therapeutic strategies for PDAC. Furthermore, we analyzed the gene interaction network and the top 2 modules using MCODE, and found that COL1A2, COL3A1, and COL5A2 were the core interaction genes for PDAC. This finding was in accordance with our STRING database results.

Collagen is the major component of the interstitial extracellular matrix. The ECM is known to play an active role in numerous BP including those involved in the maintenance of cellular shape, proliferation, migration, differentiation, and apoptosis, as well as carcinogenesis.^[61] A previous study found that nano-targeted relaxin impairs fibrosis and tumor growth in pancreatic cancer, reduces collagen I, desmin (hPSC marker) and CD31 (endothelial marker) expression, and improves the efficacy of gemcitabine.^[62] High levels of COL1A2 expression are associated with the pathogenesis of systemic sclerosis.^[63] Cancer-associated fibroblasts, which are the key effector cells in PDAC, are known to induce tumor growth and progression. It has been reported that COL3A1 is a gene related to adhesion that plays a role in EMT in hepatocellular carcinoma.^[64] A previous study has found that molecules involved in ECM remodeling (eg, COL5A2 and MMP13) were differentially expressed in ductal carcinoma in situ and invasive ductal carcinoma in breast cancer.^[65] Therefore, we

propose that COL1A2, COL3A1, and COL5A2 may play a role in promoting pancreatic fibrosis and epithelial-mesenchymal transition via the ECM-receptor interaction pathway in the early stages of PDAC and may thus serve as indicators of a critical point in disease development. Upon reviewing the relevant literature, no report was found that indicates that these genes play a definitive role in the development of PDAC. Further studies that aim to measure the levels of these genes and confirm their explicit roles in PDAC should be conducted. Improved understanding of the roles of these genes in PDAC would help us to elucidate the crucial mechanisms that underlie tumorigenesis. These genes could potentially serve as targets for the early diagnosis or treatment of this disease.

Although bioinformatics technologies have the ability to identify the potential gene candidates involved in tumorigenesis, several limitations of this study should be noted. First, the sample sizes used for the expression profiling were small and some clinical variables were not available, and thus, we failed to consider factors such as the presence of different cancer states to distinguish driver/casual genes and passenger genes and to determine the availability of diagnostic biomarkers in blood. Second, the results obtained solely by means of bioinformatic analysis are not sufficient, and experimental verification is needed to confirm these results. Therefore, further genetic and experimental studies with larger sample sizes and experimental verification are required.

In conclusion, we identified several core genes and pathways that were closely related to the initiation and progression of PDAC using a series of bioinformatics analyses of DEGs in tumor and normal samples. Identification of the involved genes, COL1A2, COL3A1, and COL5A2, provided greater insight into the specific molecular mechanisms underlying PDAC occurrence and development, especially in terms of the pathways involved in fat metabolism and ECM-receptor interaction. Upon further experimental validation of these results, these genes may serve as potential research targets for therapy and thereby contribute to greater understanding of the molecular mechanisms underlying the development of PDAC.

Author contributions

Formal analysis: Nan Li.

Investigation: Nan Li.

Methodology: Nan Li, Xin Zhao.

Software: Xin Zhao.

Supervision: Shengyi You.

Writing – original draft: Nan Li.

References

- Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115–32.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7–30.
- Martinez-Useros J, Li W, Cabeza-Morales M, et al. Oxidative stress: a new target for pancreatic cancer prognosis and treatment. *J Clin Med* 2017;6:
- Labonte ED, Pfluger PT, Cash JG, et al. Postprandial lysophospholipid suppresses hepatic fatty acid oxidation: the molecular link between group 1B phospholipase A2 and diet-induced obesity. *FASEB J* 2010;24:2516–24.
- Zhang G, Schetter A, He P, et al. DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One* 2012;7:e31507.
- Yeh JJ. Prognostic signature for pancreatic cancer: are we close? *Future Oncol* 2009;5:313–21.
- Zhang G, He P, Tan H, et al. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin Cancer Res* 2013;19:4983–93.
- Mosquera C, Maglic D, Zervos EE. Molecular targeted therapy for pancreatic adenocarcinoma: a review of completed and ongoing late phase clinical trials. *Cancer Genet* 2016;209:567–81.
- Long J, Zhang Z, Liu Z, et al. Identification of genes and pathways associated with pancreatic ductal adenocarcinoma by bioinformatics analyses. *Oncol Lett* 2016;11:1391–7.
- Downward J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer* 2003;3:11–22.
- Morris JPt, Wang SC, Hebrok M. KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. *Nat Rev Cancer* 2010;10:683–95.
- Lauth M, Toftgard R. Hedgehog signaling and pancreatic tumor development. *Adv Cancer Res* 2011;110:1–7.
- Badea L, Herlea V, Dima SO, et al. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepato-gastroenterology* 2008;55:2016–27.
- Park M, Kim M, Hwang D, et al. Characterization of gene expression and activated signaling pathways in solid-pseudopapillary neoplasm of pancreas. *Mod Pathol* 2014;27:580–93.
- Irizarry RA, Bolstad BM, Collin F, et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003;31:e15.
- Alibes A, Yankilevich P, Canada A, et al. IDconverter and IDClight: conversion and annotation of gene and protein IDs. *BMC Bioinformatics* 2007;8:9.
- Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013;41:D991–5.
- Smyth GK. *limma: Linear Models for Microarray Data*. Springer, New York:2005.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 1995;289–300.
- Yu JW, Mai W, Cui YL, et al. Genes and pathways identified in thyroid carcinoma based on bioinformatics analysis. *Neoplasma* 2016;63:559–68.
- Tweedie S, Ashburner M, Falls K, et al. FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res* 2009;37:D555–9.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30.
- Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;25:1091–3.
- Bindea G, Galon J, Mlecnik B. CluePedia cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* 2013;29:661–3.
- Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;39:D561–8.
- Smoot ME, Ono K, Ruscheinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;27:431–2.
- Beck MR, Otey CA, Campbell SL. Structural characterization of the interactions between palladin and alpha-actinin. *J Mol Biol* 2011;413:712–25.
- Singh P, Draviam E, Guo YS, et al. Molecular characterization of bombesin receptors on rat pancreatic acinar AR42J cells. *Am J Physiol* 1990;258:G803–9.
- Steinestel K, Bruderlein S, Lennerz JK, et al. Expression and Y435-phosphorylation of Abelson interactor 1 (Abl1) promotes tumour cell adhesion, extracellular matrix degradation and invasion by colorectal carcinoma cells. *Mol Cancer* 2014;13:145.
- Cho K, Ishiwata T, Uchida E, et al. Enhanced expression of keratinocyte growth factor and its receptor correlates with venous invasion in pancreatic cancer. *Am J Pathol* 2007;170:1964–74.
- Fritz V, Fajas L. Metabolism and proliferation share common regulatory pathways in cancer cells. *Oncogene* 2010;29:4369–77.
- Scott SA, Mathews TP, Ivanova PT, et al. Chemical modulation of glycerolipid signaling and metabolic pathways. *Biochim Biophys Acta* 2014;1841:1060–84.
- Wang DD, Liu ZW, Han MM, et al. Microarray based analysis of gene expression patterns in pancreatic neuroendocrine tumors. *Eur Rev Med Pharmacol Sci* 2015;19:3367–74.
- Svensson RU, Shaw RJ. Cancer metabolism: tumour friend or foe. *Nature* 2012;485:590–1.

- [35] Friedl P, Alexander S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* 2011;147:992–1009.
- [36] Stride A, Hattersley AT. Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Ann Med* 2002;34:207–16.
- [37] Raeder H, Johansson S, Holm PI, et al. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat Genet* 2006;38:54–62.
- [38] Raeder H, McAllister FE, Tjora E, et al. Carboxyl-ester lipase maturity-onset diabetes of the young is associated with development of pancreatic cysts and upregulated MAPK signaling in secretin-stimulated duodenal fluid. *Diabetes* 2014;63:259–69.
- [39] Kuhn K. Basement membrane (type IV) collagen. *Matrix Biol* 1995;14:439–45.
- [40] Wu YH, Chang TH, Huang YF, et al. COL11A1 confers chemoresistance on ovarian cancer cells through the activation of Akt/c/EBPbeta pathway and PDK1 stabilization. *Oncotarget* 2015;6:23748–63.
- [41] Ma X, Zhao X, Ouyang H, et al. The metabolic features of normal pancreas and pancreatic adenocarcinoma: preliminary result of in vivo proton magnetic resonance spectroscopy at 3.0 T. *J Comput Assist Tomogr* 2011;35:539–43.
- [42] Philip B, Roland CL, Daniluk J, et al. A high-fat diet activates oncogenic Kras and COX2 to induce development of pancreatic ductal adenocarcinoma in mice. *Gastroenterology* 2013;145:1449–58.
- [43] Pant S, Martin LK, Geyer S, et al. Baseline serum albumin is a predictive biomarker for patients with advanced pancreatic cancer treated with bevacizumab: a pooled analysis of 7 prospective trials of gemcitabine-based therapy with or without bevacizumab. *Cancer* 2014;120:1780–6.
- [44] Bernaudo S, Khazai S, Honarparvar E, et al. Epidermal growth factor promotes cyclin G2 degradation via calpain-mediated proteolysis in gynaecological cancer cells. *PLoS One* 2017;12:e0179906.
- [45] Glading A, Uberall F, Keyse SM, et al. Membrane proximal ERK signaling is required for M-calpain activation downstream of epidermal growth factor receptor signaling. *J Biol Chem* 2001;276:23341–8.
- [46] Auguste P, Javerzat S, Bikfalvi A. Regulation of vascular development by fibroblast growth factors. *Cell Tissue Res* 2003;314:157–66.
- [47] Powers CJ, McLeskey SW, Wellstein A. Fibroblast growth factors, their receptors and signaling. *Endocr Relat Cancer* 2000;7:165–97.
- [48] Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer* 2010;10:116–29.
- [49] Matsuda Y, Yoshimura H, Suzuki T, et al. Inhibition of fibroblast growth factor receptor 2 attenuates proliferation and invasion of pancreatic cancer. *Cancer Sci* 2014;105:1212–9.
- [50] Stenzel D, Lundkvist A, Sauvaget D, et al. Integrin-dependent and -independent functions of astrocytic fibronectin in retinal angiogenesis. *Development* 2011;138:4451–63.
- [51] Vaquero EC, Edderkaoui M, Nam KJ, et al. Extracellular matrix proteins protect pancreatic cancer cells from death via mitochondrial and nonmitochondrial pathways. *Gastroenterology* 2003;125:1188–202.
- [52] Edderkaoui M, Hong P, Lee JK, et al. Insulin-like growth factor-I receptor mediates the prosurvival effect of fibronectin. *J Biol Chem* 2007;282:26646–55.
- [53] Lombardo D. Bile salt-dependent lipase: its pathophysiological implications. *Biochim Biophys Acta* 2001;1533:1–28.
- [54] Ferraro A, Boni T, Pintzas A. EZH2 regulates cofilin activity and colon cancer cell migration by targeting ITGA2 gene. *PLoS One* 2014;9:e115276.
- [55] Tanaka S, Nakao M, Ioka T, et al. Slight dilatation of the main pancreatic duct and presence of pancreatic cysts as predictive signs of pancreatic cancer: a prospective study. *Radiology* 2010;254:965–72.
- [56] Parniczky A, Hegyi E, Toth AZ, et al. Genetic analysis of human chymotrypsin-like elastases 3A and 3B (CELA3A and CELA3B) to assess the role of complex formation between proelastases and procarboxypeptidases in chronic pancreatitis. *Int J Mol Sci* 2016;17:2148.
- [57] Shaikhbrahim Z, Lindstrot A, Buettner R, et al. Analysis of laser-microdissected prostate cancer tissues reveals potential tumor markers. *Int J Mol Med* 2011;28:605–11.
- [58] Verrecchia F, Rossert J, Mauviel A. Blocking sp1 transcription factor broadly inhibits extracellular matrix gene expression in vitro and in vivo: implications for the treatment of tissue fibrosis. *J Invest Dermatol* 2001;116:755–63.
- [59] Okazaki I, Noro T, Tsutsui N, et al. Fibrogenesis and carcinogenesis in nonalcoholic steatohepatitis (NASH): involvement of matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinase (TIMPs). *Cancers (Basel)* 2014;6:1220–55.
- [60] Huang C, Li Y, Guo Y, et al. MMP1/PAR1/SP/NK1R paracrine loop modulates early perineural invasion of pancreatic cancer cells. *Theranostics* 2018;8:3074–86.
- [61] Kang Y, Roife D, Lee Y, et al. Transforming growth factor-beta limits secretion of lumican by activated stellate cells within primary pancreatic adenocarcinoma tumors. *Clin Cancer Res* 2016;22:4934–46.
- [62] Mardhian DF, Storm G, Bansal R, et al. Nano-targeted relaxin impairs fibrosis and tumor growth in pancreatic cancer and improves the efficacy of gemcitabine in vivo. *J Control Release* 2018;290:1–0.
- [63] Ponticos M, Papaioannou I, Xu S, et al. Failed degradation of JunB contributes to overproduction of type I collagen and development of dermal fibrosis in patients with systemic sclerosis. *Arthritis Rheumatol* 2015;67:243–53.
- [64] Tong R, Yang B, Xiao H, et al. KCTD11 inhibits growth and metastasis of hepatocellular carcinoma through activating Hippo signaling. *Oncotarget* 2017;8:37717–29.
- [65] Vargas AC, McCart Reed AE, Waddell N, et al. Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. *Breast Cancer Res Treat* 2012;135:153–65.