Identifying *MMP14* and *COL12A1* as a potential combination of prognostic biomarkers in pancreatic ductal adenocarcinoma using integrated bioinformatics analysis

Jingyi Ding [1], **yanxi liu** [2], **Yu Lai** [Corresp. 3]

[1] Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China

[2] University of California, Los Angeles, Los Angeles, United States

[3] Chengdu University of Traditional Chinese Medicine, Chengdu, China

Corresponding Author: Yu Lai
Email address: archimedean@rocketmail.com

Background. Pancreatic ductal adenocarcinoma (PDAC) is a fatal malignant neoplasm. It is necessary to improve the understanding of the underlying molecular mechanisms and identify the key genes and signaling pathways involved in PDAC. **Methods**. The microarray datasets GSE28735, GSE62165, and GSE91035 were downloaded from the Gene Expression Omnibus. Differentially expressed genes (DEGs) were identified by integrated bioinformatics analysis, including protein–protein interaction (PPI) network, Gene Ontology (GO) enrichment, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The PPI network was established using the Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape software. GO functional annotation and KEGG pathway analyses were performed using the Database for Annotation, Visualization, and Integrated Discovery. Hub genes were validated via the Gene Expression Profiling Interactive Analysis tool (GEPIA) and the Human Protein Atlas (HPA) website. **Results**. A total of 263 DEGs (167 upregulated and 96 downregulated) were common to the three datasets. We used STRING and Cytoscape software to establish the PPI network and then identified key modules. From the PPI network, 225 nodes and 803 edges were selected. The most significant module, which comprised 11 DEGs, was identified using the Molecular Complex Detection plugin. The top 20 hub genes, which were filtered by the CytoHubba plugin, comprised FN1, COL1A1, COL3A1, BGN, POSTN, FBN1, COL5A2, COL12A1, THBS2, COL6A3, VCAN, CDH11, MMP14, LTBP1, IGFBP5, ALB, CXCL12, FAP, MATN3, and COL8A1. These genes were validated using The Cancer Genome Atlas (TCGA) and Genotype–Tissue Expression (GTEx) databases, and the encoded proteins were subsequently validated using the HPA website. The GO analysis results showed that the most significantly enriched biological process, cellular component, and molecular function terms among the 20 hub genes were cell adhesion, proteinaceous extracellular

matrix, and calcium ion binding, respectively. The KEGG pathway analysis showed that the 20 hub genes were mainly enriched in ECM–receptor interaction, focal adhesion, PI3K-Akt signaling pathway, protein digestion and absorption, and platelet activation. These findings indicated that FBN1 and COL8A1 appear to be involved in the progression of PDAC. Moreover, patient survival analysis performed via the GEPIA using TCGA and GTEx databases demonstrated that the expression levels of COL12A1 and MMP14 were correlated with a poor prognosis in PDAC patients ($p < 0.05$). **Conclusions**. The results demonstrated that Upregulation of MMP14 and COL12A1 is associated with poor overall survival, and these might be a combination of prognostic biomarkers in PDAC. In addition, FBN1 and COL8A1 are probably novel biomarkers of PDAC.

Manuscript Title
Identifying *MMP14* and *COL12A1* as a potential combination of prognostic biomarkers in pancreatic ductal adenocarcinoma using integrated bioinformatics analysis

Jingyi Ding[1], Yanxi Liu[2], Yu Lai[3]

[1] Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China
[2] University of California, Los Angeles, Los Angeles, California, USA
[3] School of Basic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China

Corresponding Author:
Yu Lai[3]
No.1166 Liutai Avenue, Chengdu, Sichuan, 611137, China
Email address: archimedean@rocketmail.com

Abstract

Background. Pancreatic ductal adenocarcinoma (PDAC) is a fatal malignant neoplasm. It is necessary to improve the understanding of the underlying molecular mechanisms and identify the key genes and signaling pathways involved in PDAC.

Methods. The microarray datasets GSE28735, GSE62165, and GSE91035 were downloaded from the Gene Expression Omnibus. Differentially expressed genes (DEGs) were identified by integrated bioinformatics analysis, including protein–protein interaction (PPI) network, Gene Ontology (GO) enrichment, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The PPI network was established using the Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape software. GO functional annotation and KEGG pathway analyses were performed using the Database for Annotation, Visualization, and Integrated Discovery. Hub genes were validated via the Gene Expression Profiling Interactive Analysis tool (GEPIA) and the Human Protein Atlas (HPA) website.

Results. A total of 263 DEGs (167 upregulated and 96 downregulated) were common to the three datasets. We used STRING and Cytoscape software to establish the PPI network and then identified key modules. From the PPI network, 225 nodes and 803 edges were selected. The most significant module, which comprised 11 DEGs, was identified using the Molecular Complex Detection plugin. The top 20 hub genes, which were filtered by the CytoHubba plugin, comprised *FN1*, *COL1A1*, *COL3A1*, *BGN*, *POSTN*, *FBN1*, *COL5A2*, *COL12A1*, *THBS2*, *COL6A3*, *VCAN*, *CDH11*, *MMP14*, *LTBP1*, *IGFBP5*, *ALB*, *CXCL12*, *FAP*, *MATN3*, and *COL8A1*. These genes were validated using The Cancer Genome Atlas (TCGA) and Genotype–Tissue Expression (GTEx) databases, and the encoded proteins were subsequently validated using the HPA website. The GO analysis results showed that the most significantly enriched biological process, cellular component, and molecular function terms among the 20 hub genes were cell adhesion, proteinaceous extracellular matrix, and calcium ion binding, respectively. The KEGG pathway analysis showed that the 20 hub genes were mainly enriched in ECM–receptor interaction, focal adhesion, PI3K-Akt signaling pathway, protein digestion and absorption, and platelet activation. These findings indicated that *FBN1* and *COL8A1* appear to be involved in the progression of PDAC. Moreover, patient survival analysis performed via the GEPIA using TCGA and GTEx databases demonstrated that the expression levels of *COL12A1* and *MMP14* were correlated with a poor prognosis in PDAC patients ($p < 0.05$).

Conclusions. The results demonstrated that Upregulation of *MMP14* and *COL12A1* is associated with poor overall survival, and these might be a combination of prognostic biomarkers in PDAC. In addition, *FBN1* and *COL8A1* are probably novel biomarkers of PDAC.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common malignant tumor of the pancreas and is a lethal malignancy with poor prognosis, which is in part due to its rapid progression and the lack of diagnostic and therapeutic targets. In 2018, pancreatic cancer (PC) ranked 11th among the most common cancers, with 458,918 new cases and 432,242 deaths due to PC

worldwide(Bray et al. 2018). Recent work suggests that alcohol is a risk factor for PC(Go et al. 2005), while both genetic and environmental factors also play a role in the development and progression of PC(Piepoli et al. 2006).

Understanding genetic alterations in the context of biological pathways can help identify specific novel biomarkers of PDAC. Previous studies identified several cancer-associated genes implicated in PDAC, including *KRAS*(Waters & Der 2018), *MYC*(Witkiewicz et al. 2015), and *CDKN2A*(Sikdar et al. 2018). It is widely accepted that the formation of stroma contributes to tumor proliferation, invasion, and metastasis(von Ahrens et al. 2017). Particularly pathognomonic for PDAC is a stromal reaction that occurs during tumor progression and extensively involves fibroblasts and the extracellular matrix (ECM)(Mahadevan & Von Hoff 2007). Nevertheless, the precise etiology and pathogenetic mechanism of PDAC remain unclear.

Microarray technology provides high-throughput methods for quantitatively measuring the expression levels of thousands of genes simultaneously, and microarray-based gene expression profiling can filter differentially expressed genes (DEGs) and biological pathways linked to various malignant tumors. Therefore, microarray techniques are promising and efficient ways to identify candidate biomarkers involved in the pathogenesis of PDAC. The purpose of our study was to determine significant DEGs and pathways implicated in PDAC by integrated bioinformatics analysis and to provide novel insights into the progression, diagnosis, and therapeutic targets of PDAC.

Materials & Methods

2.1 Screening database

The Gene Expression Omnibus (GEO: https://www.ncbi.nlm.nih.gov/geo/) is a public repository of high-throughput gene expression genomics datasets(Clough & Barrett 2016). In this study, we downloaded three microarray datasets, namely, GSE28735, GSE62165, and GSE91035, from the NCBI-GEO database. The array data in GSE28735 consist of 45 matching pairs corresponding to PDAC and adjacent non-tumor tissues(Zhang et al. 2013; Zhang et al. 2012). GSE62165 includes data for 118 whole-tumor tissue and 13 control samples(Janky et al. 2016). GSE91035 incorporates data for 8 normal pancreatic and 25 PDAC tissues(Sutaria et al. 2017). Altogether, data for 188 PDAC tissues and 66 non-tumor tissues were available.

2.2 Screening of DEGs

GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) is an online analysis tool that is based on the R programming language and can be used to identify DEGs that differentiate between cancer and normal samples in a GEO series(Yao & Liu 2018). Using GEO2R, we analyzed DEGs that differentiate between PDAC and non-tumor tissue samples. An adjusted p-value of <0.05 and $|\log FC| > 1$ were employed as the cutoff criteria representing a significant difference. Using a data processing standard, we filtered DEGs via the Venn diagram tool at http://bioinformatics.psb.ugent.be/webtools/Venn/. A total of 263 DEGs were selected, which consisted of 167 upregulated genes and 96 downregulated genes.

2.3 Establishment of the protein–protein interaction (PPI) network

The Search Tool for the Retrieval of Interacting Genes (STRING: http://string-db.org/) is an online application that can be used to assess DEG-encoded proteins and protein–protein interaction (PPI) networks(Szklarczyk et al. 2015). A combined score of >0.4 was set as the threshold.

Cytoscape software(Shannon et al. 2003) was utilized to visualize the PPI network, which established a new way to find potential key candidate genes and core proteins. We utilized cluster analysis via the Molecular Complex Detection (MCODE) plugin with degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max depth = 100, which detected significant modules in the PPI network. To identify the hub genes, we also utilized the CytoHubba plugin, which provided a novel method of exploring significant nodes in PPI networks. These tools yield new insights into normal cellular processes, the underlying mechanisms of disease pathology, and clinical treatment.

2.4 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs

The Gene Ontology (GO) is used to perform enrichment analysis, which covers the cellular component (CC), biological process (BP), and molecular function (MF), of the selected genes(Young et al. 2010). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database that helps to illustrate the functionalities and pathways of the selected genes(Altermann & Klaenhammer 2005). The Database for Annotation, Visualization, and Integrated Discovery (DAVID: http://david.ncifcrf.gov/) is a public online bioinformatics database(Dennis et al. 2003) that contains information on biological functional annotations for genes and proteins. The cutoff criteria were selected on the basis of $p < 0.05$. We performed enrichment of the GO terms and KEGG pathways for the candidate DEGs using DAVID.

2.5 Survival analysis of the candidate genes and validation of DEGs using TCGA and GTEx databases

Based on data for 9,736 tumors and 8,587 normal samples from The Cancer Genome Atlas (TCGA) database and the Genotype–Tissue Expression (GTEx) database, the Gene Expression Profiling Interactive Analysis tool (GEPIA: http://gepia.cancer-pku.cn/) is used to perform functions such as survival analysis, the detection of similar genes, and correlation analysis to clarify the relationships between diseases and DEGs(Tang et al. 2017).

The GEPIA was also utilized for validating and visualizing the selected DEGs using TCGA and GTEx databases(Tang et al. 2017).

2.6 Validation of expression of candidate gene-encoded proteins

The expression of proteins encoded by the PDAC candidate genes was validated using the Human Protein Atlas (HPA: https://www.proteinatlas.org/) website on the basis of spatial proteomics data and quantitative transcriptomics data (RNA-Seq) obtained from immunohistochemical analysis of tissue microarrays.

Results

3.1 Identification of DEGs

A total of 263 DEGs were identified from GSE28735, GSE62165, and GSE91035. There were 167 upregulated genes and 96 downregulated genes in PDAC tissues in comparison with non-tumor tissues (Fig.1) (Table 1).

3.2 Establishment of the PPI network

Using the STRING application and Cytoscape software, 225 nodes and 803 edges were mapped in the PPI network (Fig.2a). In association with these nodes, the whole PPI network was analyzed using the MCODE plugin, and one significant module was identified with average MCODE score = 8.6, nodes = 11, and edges = 43 (Fig.2b). This significant module comprised 11 DEGs, namely, *COL6A3*, *COL3A1*, *VCAN*, *COL5A2*, *COL12A1*, *THBS2*, *FBN1*, *POSTN*, *LTBP1*, *MMP14*, and *CDH11*. Enrichment of KEGG pathways indicated that this module was mainly associated with ECM–receptor interactions and focal adhesion. From the PPI network, the top 20 hub genes were filtered by the CytoHubba plugin using the maximal clique centrality method. Their order of sequence was as follows: *FN1*, *COL1A1*, *COL3A1*, *BGN*, *POSTN*, *FBN1*, *COL5A2*, *COL12A1*, *THBS2*, *COL6A3*, *VCAN*, *CDH11*, *MMP14*, *LTBP1*, *IGFBP5*, *ALB*, *CXCL12*, *FAP*, *MATN3*, and *COL8A1* (Fig.2c). Via data mining, we found that the significant module and hub genes mainly consisted of upregulated genes.

3.3 GO and KEGG pathway analysis of DEGs

Functional and pathway enrichment analyses were accomplished using DAVID. GO analysis showed that the most significant module was mainly enriched in cell adhesion, extracellular matrix structural constituent, and proteinaceous extracellular matrix (Fig.3) (Table 2). Moreover, the 20 hub genes were mainly enriched in cell adhesion, proteinaceous extracellular matrix, and calcium ion binding (Fig.4) (Table 3). In addition, KEGG pathway enrichment analysis demonstrated that the DEGs in the most significant module were enriched in ECM–receptor interaction and focal adhesion (Fig.3) (Table 2) and the hub genes were mainly enriched in ECM–receptor interaction, focal adhesion, protein digestion and absorption, and PI3K-Akt signaling pathway (Fig.4) (Table 3).

3.4 Overall survival analysis of the top 20 hub genes

Patient survival analysis performed via the GEPIA using TCGA and GTEx databases demonstrated that the expression levels of *COL12A1* and *MMP14* were correlated with an unfavorable prognosis in PDAC patients ($p < 0.05$) (Fig.5).

3.5 Validation of DEGs using TCGA and GTEx databases

To ensure the reliability of the identification of the top 20 hub genes, we validated these via the GEPIA using TCGA and GTEx databases. Boxplots of the hub genes associated with PDAC were downloaded from the GEPIA. The results demonstrated that *FN1*, *COL1A1*, *COL3A1*, *BGN*, *POSTN*, *FBN1*, *COL5A2*, *COL12A1*, *THBS2*, *COL6A3*, *VCAN*, *CDH11*, *MMP14*, *LTBP1*, *IGFBP5*, *FAP*, *MATN3*, and *COL8A1* were significantly overexpressed in PDAC tissues in comparison with normal pancreatic tissues, whereas *ALB* was underexpressed in PDAC tissues ($p < 0.05$) (Fig.6).

3.6 Validation of expression of candidate gene-encoded proteins

We obtained the expression levels of proteins encoded by the 20 hub genes associated with PDAC from the HPA website. No data for proteins encoded by *COL5A2*, *IGFBP5*, and *MATN3* are reported on the HPA website, and expression profiles of the other 17 genes in PDAC clinical specimens are shown in Figure 7.

The protein expressions of *FN1*, *MMP14*, *COL12A1*, *COL3A1*, *COL1A1*, *POSTN*, *VCAN*, *LTBP1*, *FBN1*, and *FAP* were upregulated in PDAC tissues in comparison with normal tissues, with only *ALB* being downregulated in PDAC tissues. *COL6A3*, *COL8A1*, *CDH11*, and *CXCL12* were not expressed in either PDAC tissues or normal tissues, and *BGN* and *THBS2* were overexpressed in both cancer and normal tissues.

Discussion

GO analysis showed that the most significantly enriched BP, CC, and MF terms among the 20 hub genes were cell adhesion, proteinaceous extracellular matrix, and calcium ion binding, respectively. Cell adhesion is the attachment of a cell either to another cell or to an underlying substrate. The proteinaceous extracellular matrix provides structural support and biochemical or biomechanical cues for cells or tissues and is a structure located external to one or more cells. The ECM is a crucial factor in both promoting the progression of PDAC and inhibiting the delivery of antitumor therapy(Weniger et al. 2018).

According to the analysis of the MF terms among the hub genes, *MMP14*, *THBS2*, *CDH11*, *FBN1*, *LTBP1*, *MATN3*, and *VCAN* were jointly involved in calcium ion binding, which is defined as selective and non-covalent interactions with calcium ions (Ca^{2+}). Ca^{2+} is a ubiquitous and versatile second messenger involved in the regulation of numerous cellular functions, including gene transcription, vesicular trafficking, and cytoskeletal rearrangements(Nunes-Hasler et al. 2020). Ca^{2+} and Ca^{2+}-regulating proteins contribute to a large number of processes that are key to cancer cells, including proliferation, invasion, and cell death(Monteith et al. 2017; Prevarskaya et al. 2011). A high serum Ca^{2+} level is associated with a poor prognosis in PDAC(Dong et al. 2014). It is thus reasonable that the seven abovementioned genes might regulate calcium ion binding and affect the development of PDAC. Furthermore, our study suggests that the interaction of MMP14 with Ca^{2+} is a promising biomarker for survival in PDAC.

Matrix metalloproteinases (MMPs) are a family of calcium- and zinc-dependent membrane-anchored or secreted endopeptidases that are overexpressed in various diseases, including breast cancer(Min et al. 2015). MMP14 is located in neoplastic epithelium. It is speculated that the overexpression of *MMP14* alone may be sufficient to induce the development of PDAC(Shields et al. 2012). Moreover, *MMP14* is overexpressed in the epithelium in invasive PC(Iacobuzio-Donahue et al. 2002; Shields et al. 2012), and MMP14, as an endopeptidase, can degrade various components of the ECM such as collagen, which possibly leads to metastasis of tumors(Golubkov et al. 2010). Type I collagen can induce the expression of *MMP14* and *TGF-β1* in pancreatic ductal epithelial cells(Ottaviano et al. 2006), and *COL1A1* encodes the major component of type I collagen. The expression of *MMP14* in PDAC cells stimulates pancreatic stellate cells (PSCs) and enhances the production of type I collagen by increasing transforming

growth factor-β signaling(Krantz et al. 2011). Ottaviano et al. found that fibrosis and the expression of *MMP14* in tumor specimens increased in comparison with those in normal pancreatic tissue(Ottaviano et al. 2006). These findings suggest the key role of interactions between *MMP14* and type I collagen in the progression of PDAC and support *MMP14* as a potential target for inhibiting fibrosis, preventing metastasis, and treating PDAC.

The KEGG pathway analysis showed that six hub genes, namely, *COL1A1*, *COL3A1*, *COL5A2*, *COL6A3*, *FN1*, and *THBS2*, were significantly associated with ECM–receptor interactions, focal adhesion, and the phosphatidylinositol-3-kinase–protein kinase B (PI3K-Akt) signaling pathway. In addition, collagen-encoding genes, including *COL1A1*, *COL3A1*, and *COL5A2*, were also enriched in protein digestion and absorption and platelet activation.

ECM–receptor interactions play important roles in the processes of cell shedding, adhesion, degradation, migration, differentiation, hyperplasia, and apoptosis(Bao et al. 2019). PSCs secrete several ECM proteins, including collagen, fibronectin, fibulin-2, and laminin, as well as hyaluronan (Hall et al. 2019). Moreover, *COL1A1* and *COL3A1* were significantly downregulated in PC ($p < 0.0001$) after treatment with gemcitabine in combination with EC359(Hall et al. 2019). The gene *COL1A1* encodes the pro-alpha 1 chain of type I collagen, which is closely associated with *MMP14*. *COL3A1* was found to encode a major structural component of hollow organs such as large blood vessels, the uterus and bowel, and tissues that must withstand stretching(Kuivaniemi & Tromp 2019). As an important molecule, *COL5A2* is associated with remodeling of the ECM and is differentially expressed between in situ ductal carcinoma and invasive ductal carcinoma(Vargas et al. 2012). The alpha 3 chain of type VI collagen is mainly present in the desmoplastic stroma in PDAC, with large deposits between the sites of stromal fatty infiltration and around the malignant ducts(Arafat et al. 2011), and the circulating form of this protein has potential clinical significance in the diagnosis of pancreatic malignancy(Kang et al. 2014). *FN1* encodes a collagen-associated protein that has been identified as a potential biomarker of an unfavorable prognosis in PDAC(Hu et al. 2018). *THBS2* appears in the early stages of PDAC and hence has great potential for the diagnosis of PDAC, with 98% specificity(2017).

At points of ECM–cell contact, specialized structures are formed, which are termed focal adhesions. Some components of focal adhesions contribute to cell migration in PDAC and participate in structural links between the actin cytoskeleton and membrane receptors, whereas others are signaling molecules(Manoli et al. 2019).

The PI3K-Akt signaling pathway regulates fundamental cellular functions, including transcription, translation, proliferation, growth, and survival. Accumulating evidence has implied that the PI3K-Akt signaling pathway promotes malignant processes of PDAC cells, including proliferation, angiogenesis, metastasis, suppression of apoptosis, and chemoresistance, and targeting the PI3K-Akt signaling pathway has been a potential therapeutic strategy for the treatment of PC(Ebrahimi et al. 2017).

In PC, both exocrine and endocrine functions are abnormal, which profoundly influences the secretion of proteases, and hence protein digestion and absorption is a prominent metabolic

change(Gilliland et al. 2017). Platelet activation facilitates the P-selectin- and integrin-dependent accumulation of cancer cell microparticles and promotes tumor growth and metastasis(Mezouar et al. 2015). However, the effect of collagen-mediated platelet activation on the progression of PDAC needs further investigation.

Collagens are centrally involved in the formation of fibrillar and microfibrillar networks of the ECM and basement membranes, as well as other structures of the ECM(Gelse et al. 2003). We further found that the collagen family is closely associated with PDAC. Interestingly, Wang and Li also found that the collagen family and *FN1* have an influence on PC via data mining using a different gene set (GSE15471)(Wang & Li 2015). As we have done, they suggested that *FN1*, together with *COL1A1*, *COL3A1*, and *COL5A2*, may be key molecules in the development and progression of PDAC owing to their involvement in ECM–receptor interactions and focal adhesion pathways. These DEGs were also identified in our study. Furthermore, we found that *COL12A1* and *COL6A3* are probably also key DEGs that influence PDAC, which differs from the results of Wang and Li. Although the specific relationship between *COL12A1* and PDAC has not been reported, our findings also suggest that *COL12A1* is a potential prognostic biomarker in patients with PDAC.

We also found that *FBN1* and *COL8A1* appear to be involved in the progression of PDAC. *FBN1* encodes a structural component of the microfibrils of the ECM that have diameters of 10–12 nm, which impart both regulatory and structural properties to load-bearing connective tissues(Lee et al. 2004). The silencing of *FBN1* inhibits the proliferative, migratory, and invasive activities of gastric cancer cells, whereas upregulation of the expression of *FBN1* has the opposite effect(Yang et al. 2017). *COL8A1* encodes a macromolecular component of the subendothelium(Xu et al. 2001). It is suggested that *COL8A1* may be associated with malignant processes of hepatocarcinoma(Zhao et al. 2009) and the progression and prognosis of human colon adenocarcinoma(Shang et al. 2018).

Conclusions

In conclusion, we screened the top 20 hub genes (*FN1*, *COL1A1*, *COL3A1*, *BGN*, *POSTN*, *FBN1*, *COL5A2*, *COL12A1*, *THBS2*, *COL6A3*, *VCAN*, *CDH11*, *MMP14*, *LTBP1*, *IGFBP5*, *ALB*, *CXCL12*, *FAP*, *MATN3*, and *COL8A1*) and the related enriched functions or pathways, which regulate the progression and metastatic invasion of PDAC, as well as overall survival. The results demonstrate that the upregulation of *MMP14* and *COL12A1* in PDAC is closely associated with poor overall survival, that these might be a potential combination of prognostic biomarkers in patients with PDAC, and that *FBN1* and *COL8A1* might be biomarkers of PDAC. In brief, our study increases the understanding of the potential critical genes and related pathways that participate in the pathogenesis of PDAC.

Acknowledgements

There are no acknowledgements.

References

2017. New Biomarker Identified for PDAC. *Cancer Discov* 7:OF3. 10.1158/2159-8290.CD-NB2017-107
Altermann E, and Klaenhammer TR. 2005. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* 6:60. 10.1186/1471-2164-6-60
Arafat H, Lazar M, Salem K, Chipitsyna G, Gong Q, Pan TC, Zhang RZ, Yeo CJ, and Chu ML. 2011. Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer. *Surgery* 150:306-315. 10.1016/j.surg.2011.05.011
Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, Chen C, Ren Y, and Jia Y. 2019. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell Mol Biol Lett* 24:38. 10.1186/s11658-019-0162-0
Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394-424. 10.3322/caac.21492
Clough E, and Barrett T. 2016. The Gene Expression Omnibus Database. *Methods Mol Biol* 1418:93-110. 10.1007/978-1-4939-3578-9_5
Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, and Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4:P3.
Dong Q, Zhang Y, Yang XH, Jing W, Zheng LQ, Liu YP, Qu XJ, and Li Z. 2014. Serum calcium level used as a prognostic predictor in patients with resectable pancreatic ductal adenocarcinoma. *Clin Res Hepatol Gastroenterol* 38:639-648. 10.1016/j.clinre.2014.01.012
Ebrahimi S, Hosseini M, Shahidsales S, Maftouh M, Ferns GA, Ghayour-Mobarhan M, Hassanian SM, and Avan A. 2017. Targeting the Akt/PI3K Signaling Pathway as a Potential Therapeutic Strategy for the Treatment of Pancreatic Cancer. *Curr Med Chem* 24:1321-1331. 10.2174/0929867324666170206142658
Gelse K, Poschl E, and Aigner T. 2003. Collagens--structure, function, and biosynthesis. *Adv Drug Deliv Rev* 55:1531-1546. 10.1016/j.addr.2003.08.002
Gilliland TM, Villafane-Ferriol N, Shah KP, Shah RM, Tran Cao HS, Massarweh NN, Silberfein EJ, Choi EA, Hsu C, McElhany AL, Barakat O, Fisher W, and Van Buren G. 2017. Nutritional and Metabolic Derangements in Pancreatic Cancer and Pancreatic Resection. *Nutrients* 9. 10.3390/nu9030243
Go VL, Gukovskaya A, and Pandol SJ. 2005. Alcohol and pancreatic cancer. *Alcohol* 35:205-211. 10.1016/j.alcohol.2005.03.010
Golubkov VS, Chekanov AV, Cieplak P, Aleshin AE, Chernov AV, Zhu W, Radichev IA, Zhang D, Dong PD, and Strongin AY. 2010. The Wnt/planar cell polarity protein-tyrosine kinase-7 (PTK7) is a highly efficient proteolytic target of membrane type-1 matrix metalloproteinase: implications in cancer and embryogenesis. *J Biol Chem* 285:35740-35749. 10.1074/jbc.M110.165159
Hall BR, Cannon A, Thompson C, Santhamma B, Chavez-Riveros A, Bhatia R, Nair HB, Nickisch K, Batra SK, and Kumar S. 2019. Utilizing cell line-derived organoids to evaluate the efficacy of a novel LIFR-inhibitor, EC359 in targeting pancreatic tumor stroma. *Genes Cancer* 10:1-10. 10.18632/genesandcancer.184
Hu D, Ansari D, Pawlowski K, Zhou Q, Sasor A, Welinder C, Kristl T, Bauden M, Rezeli M, Jiang Y, Marko-Varga G, and Andersson R. 2018. Proteomic analyses identify prognostic biomarkers for pancreatic ductal adenocarcinoma. *Oncotarget* 9:9789-9807. 10.18632/oncotarget.23929
Iacobuzio-Donahue CA, Ryu B, Hruban RH, and Kern SE. 2002. Exploring the host desmoplastic response to pancreatic carcinoma: gene expression of stromal and

neoplastic cells at the site of primary invasion. *Am J Pathol* 160:91-99. 10.1016/S0002-9440(10)64353-2

Janky R, Binda MM, Allemeersch J, Van den Broeck A, Govaere O, Swinnen JV, Roskams T, Aerts S, and Topal B. 2016. Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. *BMC Cancer* 16:632. 10.1186/s12885-016-2540-6

Kang CY, Wang J, Axell-House D, Soni P, Chu ML, Chipitsyna G, Sarosiek K, Sendecki J, Hyslop T, Al-Zoubi M, Yeo CJ, and Arafat HA. 2014. Clinical significance of serum COL6A3 in pancreatic ductal adenocarcinoma. *J Gastrointest Surg* 18:7-15. 10.1007/s11605-013-2326-y

Krantz SB, Shields MA, Dangi-Garimella S, Cheon EC, Barron MR, Hwang RF, Rao MS, Grippo PJ, Bentrem DJ, and Munshi HG. 2011. MT1-MMP cooperates with Kras(G12D) to promote pancreatic fibrosis through increased TGF-beta signaling. *Mol Cancer Res* 9:1294-1304. 10.1158/1541-7786.MCR-11-0023

Kuivaniemi H, and Tromp G. 2019. Type III collagen (COL3A1): Gene and protein structure, tissue distribution, and associated diseases. *Gene* 707:151-171. 10.1016/j.gene.2019.05.003

Lee SS, Knott V, Jovanovic J, Harlos K, Grimes JM, Choulier L, Mardon HJ, Stuart DI, and Handford PA. 2004. Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization. *Structure* 12:717-729. 10.1016/j.str.2004.02.023

Mahadevan D, and Von Hoff DD. 2007. Tumor-stroma interactions in pancreatic ductal adenocarcinoma. *Mol Cancer Ther* 6:1186-1197. 10.1158/1535-7163.MCT-06-0686

Manoli S, Coppola S, Duranti C, Lulli M, Magni L, Kuppalu N, Nielsen N, Schmidt T, Schwab A, Becchetti A, and Arcangeli A. 2019. The Activity of Kv 11.1 Potassium Channel Modulates F-Actin Organization During Cell Migration of Pancreatic Ductal Adenocarcinoma Cells. *Cancers (Basel)* 11. 10.3390/cancers11020135

Mezouar S, Darbousset R, Dignat-George F, Panicot-Dubois L, and Dubois C. 2015. Inhibition of platelet activation prevents the P-selectin and integrin-dependent accumulation of cancer cell microparticles and reduces tumor growth and metastasis in vivo. *Int J Cancer* 136:462-475. 10.1002/ijc.28997

Min K, Ji B, Zhao M, Ji T, Chen B, Fang X, and Ma Q. 2015. Development of a Radiolabeled Peptide-Based Probe Targeting MT1-MMP for Breast Cancer Detection. *PLoS One* 10:e0139471. 10.1371/journal.pone.0139471

Monteith GR, Prevarskaya N, and Roberts-Thomson SJ. 2017. The calcium-cancer signalling nexus. *Nat Rev Cancer* 17:367-380. 10.1038/nrc.2017.18

Nunes-Hasler P, Kaba M, and Demaurex N. 2020. Molecular Mechanisms of Calcium Signaling During Phagocytosis. *Adv Exp Med Biol* 1246:103-128. 10.1007/978-3-030-40406-2_7

Ottaviano AJ, Sun L, Ananthanarayanan V, and Munshi HG. 2006. Extracellular matrix-mediated membrane-type 1 matrix metalloproteinase expression in pancreatic ductal cells is regulated by transforming growth factor-beta1. *Cancer Res* 66:7032-7040. 10.1158/0008-5472.CAN-05-4421

Piepoli A, Gentile A, Valvano MR, Barana D, Oliani C, Cotugno R, Quitadamo M, Andriulli A, and Perri F. 2006. Lack of association between UGT1A7, UGT1A9, ARP, SPINK1 and CFTR gene polymorphisms and pancreatic cancer in Italian patients. *World J Gastroenterol* 12:6343-6348. 10.3748/wjg.v12.i39.6343

Prevarskaya N, Skryma R, and Shuba Y. 2011. Calcium in tumour metastasis: new roles for known actors. *Nat Rev Cancer* 11:609-618. 10.1038/nrc3105

Shang J, Wang F, Chen P, Wang X, Ding F, Liu S, and Zhao Q. 2018. Co-expression Network Analysis Identified COL8A1 Is Associated with the Progression and Prognosis in Human Colon Adenocarcinoma. *Dig Dis Sci* 63:1219-1228. 10.1007/s10620-018-4996-5

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

Shields MA, Dangi-Garimella S, Redig AJ, and Munshi HG. 2012. Biochemical role of the collagen-rich tumour microenvironment in pancreatic cancer progression. *Biochem J* 441:541-552. 10.1042/BJ20111240

Sikdar N, Saha G, Dutta A, Ghosh S, Shrikhande SV, and Banerjee S. 2018. Genetic Alterations of Periampullary and Pancreatic Ductal Adenocarcinoma: An Overview. *Curr Genomics* 19:444-463. 10.2174/1389202919666180221160753

Sutaria DS, Jiang J, Azevedo-Pouly ACP, Lee EJ, Lerner MR, Brackett DJ, Vandesompele J, Mestdagh P, and Schmittgen TD. 2017. Expression Profiling Identifies the Noncoding Processed Transcript of HNRNPU with Proliferative Properties in Pancreatic Ductal Adenocarcinoma. *Noncoding RNA* 3. 10.3390/ncrna3030024

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

Tang Z, Li C, Kang B, Gao G, Li C, and Zhang Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 45:W98-W102. 10.1093/nar/gkx247

Vargas AC, McCart Reed AE, Waddell N, Lane A, Reid LE, Smart CE, Cocciardi S, da Silva L, Song S, Chenevix-Trench G, Simpson PT, and Lakhani SR. 2012. Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. *Breast Cancer Res Treat* 135:153-165. 10.1007/s10549-012-2123-4

von Ahrens D, Bhagat TD, Nagrath D, Maitra A, and Verma A. 2017. The role of stromal cancer-associated fibroblasts in pancreatic cancer. *J Hematol Oncol* 10:76. 10.1186/s13045-017-0448-5

Wang Y, and Li Y. 2015. Analysis of molecular pathways in pancreatic ductal adenocarcinomas with a bioinformatics approach. *Asian Pac J Cancer Prev* 16:2561-2567. 10.7314/apjcp.2015.16.6.2561

Waters AM, and Der CJ. 2018. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. *Cold Spring Harb Perspect Med* 8. 10.1101/cshperspect.a031435

Weniger M, Honselmann KC, and Liss AS. 2018. The Extracellular Matrix and Pancreatic Cancer: A Complex Relationship. *Cancers (Basel)* 10. 10.3390/cancers10090316

Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, Mollaee M, Wagner KU, Koduru P, Yopp A, Choti MA, Yeo CJ, McCue P, White MA, and Knudsen ES. 2015. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat Commun* 6:6744. 10.1038/ncomms7744

Xu R, Yao ZY, Xin L, Zhang Q, Li TP, and Gan RB. 2001. NC1 domain of human type VIII collagen (alpha 1) inhibits bovine aortic endothelial cell proliferation and causes cell apoptosis. *Biochem Biophys Res Commun* 289:264-268. 10.1006/bbrc.2001.5970

Yang D, Zhao D, and Chen X. 2017. MiR-133b inhibits proliferation and invasion of gastric cancer cells by up-regulating FBN1 expression. *Cancer Biomark* 19:425-436. 10.3233/CBM-160421

Yao S, and Liu T. 2018. Analysis of differential gene expression caused by cervical intraepithelial neoplasia based on GEO database. *Oncol Lett* 15:8319-8324. 10.3892/ol.2018.8403

Young MD, Wakefield MJ, Smyth GK, and Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:R14. 10.1186/gb-2010-11-2-r14

Zhang G, He P, Tan H, Budhu A, Gaedcke J, Ghadimi BM, Ried T, Yfantis HG, Lee DH, Maitra A, Hanna N, Alexander HR, and Hussain SP. 2013. Integration of metabolomics and

transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin Cancer Res* 19:4983-4993. 10.1158/1078-0432.CCR-13-0209

Zhang G, Schetter A, He P, Funamizu N, Gaedcke J, Ghadimi BM, Ried T, Hassan R, Yfantis HG, Lee DH, Lacy C, Maitra A, Hanna N, Alexander HR, and Hussain SP. 2012. DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One* 7:e31507. 10.1371/journal.pone.0031507

Zhao Y, Jia L, Mao X, Xu H, Wang B, and Liu Y. 2009. siRNA-targeted COL8A1 inhibits proliferation, reduces invasion and enhances sensitivity to D-limonence treatment in hepatocarcinoma cells. *IUBMB Life* 61:74-79. 10.1002/iub.151

Figure 1

Venn diagram

Identification of differentially expressed genes (DEGs) from GSE28735, GSE62165, and GSE91035. The different colored areas represent the different datasets, and a total of 263 DEGs were common to all three datasets.

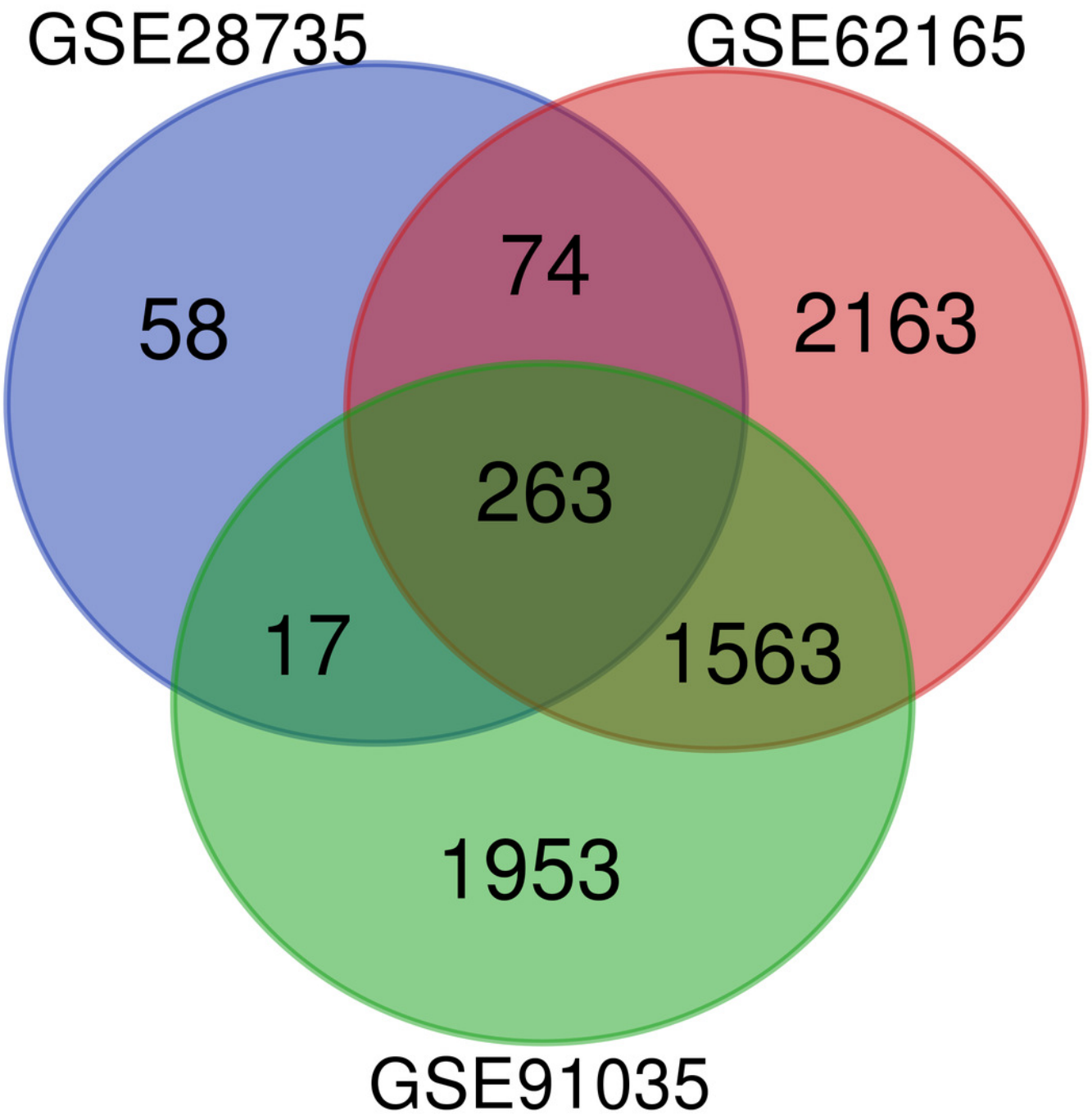
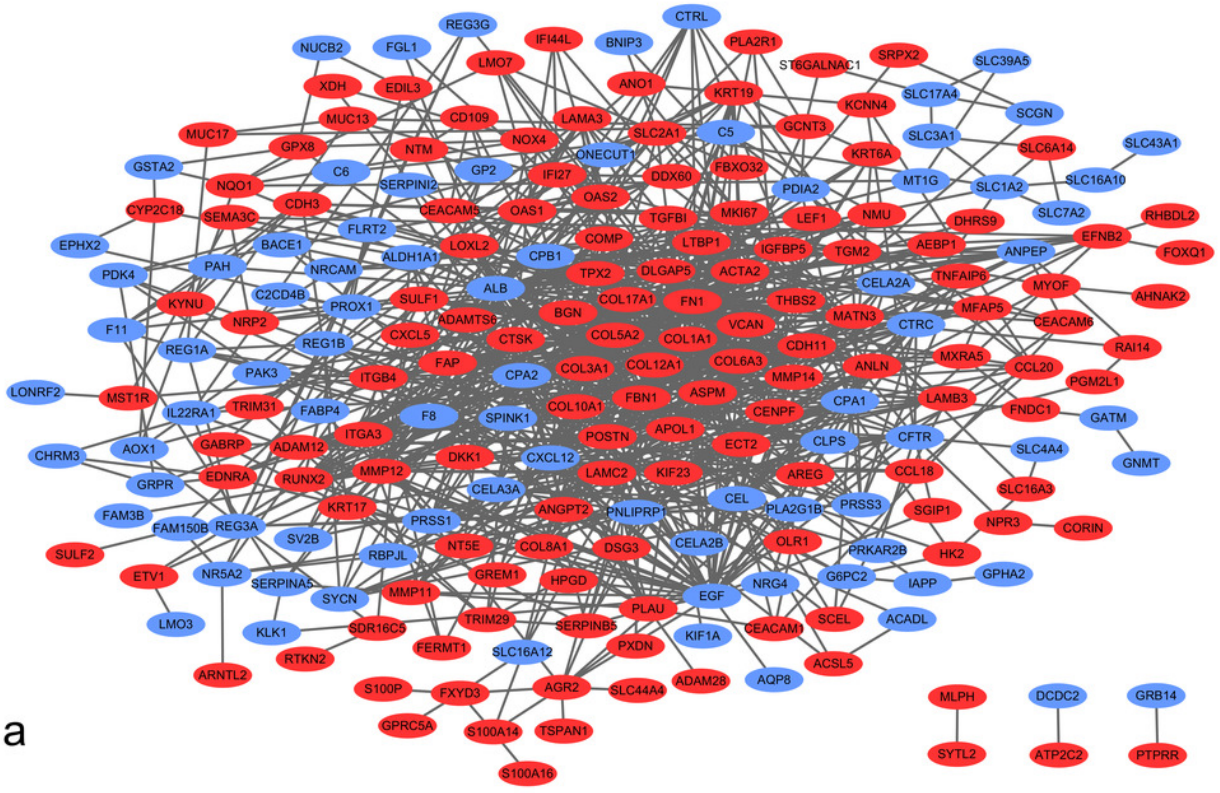
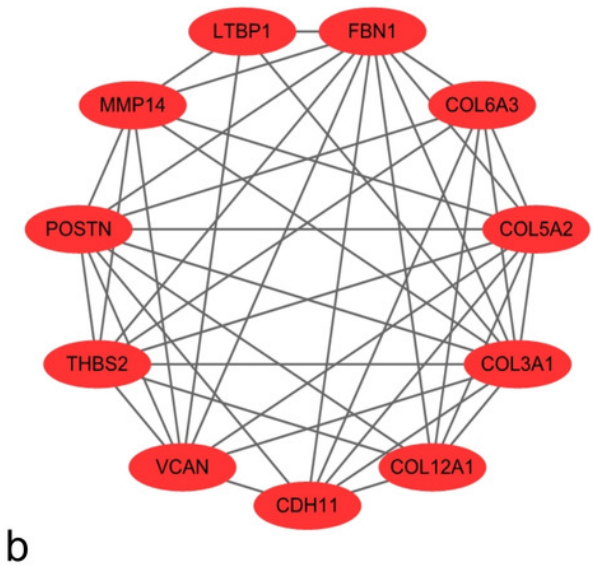

GSE28735
GSE62165
58
74
2163
263
17
1563
1953
GSE91035

Figure 2

Protein–protein interaction (PPI) network of DEGs.

(A) PPI network of 263 DEGs in PDAC tissues. Red nodes represent upregulated genes, whereas blue nodes represent downregulated genes. (B) Significant module identified from PPI network via the Molecular Complex Detection plugin. This module consisted of upregulated genes. (C) Top 20 hub genes filtered using CytoHubba plugin. Nodes in darker colors were found to have higher significance.


CTRL
REG3G
NUCB2
FGL1
IFI44L
BNIP3
PLA2R1
ST6GALNAC1
SRPX2
SLC39A5
LMO7
XDH
EDIL3
ANO1
KRT19
SLC17A4
KCNN4
SCGN
MUC13
CD109
LAMA3
SLC2A1
C5
GCNT3
SLC3A1
SLC43A1
MUC17
GPX8
NTM
NOX4
ONECUT1
KRT6A
SLC6A14
GSTA2
C6
GP2
IFI27
DDX60
FBXO32
MT1G
SLC1A2
SLC16A10
NQO1
SERPINI2
OAS2
PDIA2
CYP2C18
SEMA3C
CDH3
CEACAM5
OAS1
TGFBI
MKI67
LEF1
NMU
DHRS9
SLC7A2
RHBDL2
EPHX2
FLRT2
BACE1
LOXL2
COMP
LTBP1
IGFBP5
TGM2
AEBP1
ANPEP
EFNB2
FOXQ1
PDK4
PAH
ALDH1A1
CPB1
TPX2
DLGAP5
ACTA2
CELA2A
TNFAIP6
MYOF
AHNAK2
NRCAM
KYNU
C2CD4B
PROX1
SULF1
ALB
COL17A1
FN1
THBS2
MATN3
MFAP5
F11
NRP2
ADAMTS6
BGN
VCAN
CTRC
CEACAM6
CXCL5
CTSK
COL5A2
COL1A1
CDH11
REG1A
REG1B
FAP
COL3A1
COL12A1
COL6A3
ANLN
MXRA5
RAI14
PAK3
ITGB4
CPA2
MMP14
CCL20
LONRF2
MST1R
FBN1
ASPM
CPA1
LAMB3
PGM2L1
IL22RA1
TRIM31
FABP4
F8
SPINK1
COL10A1
CENPF
FNDC1
GATM
ITGA3
POSTN
APOL1
CLPS
CFTR
CHRM3
AOX1
GABRP
ADAM12
DKK1
CXCL12
ECT2
SLC4A4
GNMT
EDNRA
MMP12
LAMC2
KIF23
AREG
CCL18
GRPR
RUNX2
CELA3A
CEL
SLC16A3
FAM3B
KRT17
ANGPT2
PNLIPRP1
PLA2G1B
PRSS3
SGIP1
FAM150B
REG3A
PRSS1
NPR3
CORIN
SULF2
SV2B
NT5E
COL8A1
DSG3
CELA2B
OLR1
PRKAR2B
HK2
RBPJL
ETV1
NR5A2
SERPINA5
GREM1
HPGD
NRG4
G6PC2
IAPP
GPHA2
SYCN
MMP11
EGF
PLAU
SCEL
ACADL
LMO3
KLK1
SDR16C5
TRIM29
SERPINB5
CEACAM1
KIF1A
FERMT1
PXDN
SLC16A12
ACSL5
RTKN2
ARNTL2
S100P
FXYD3
AGR2
SLC44A4
ADAM28
AQP8
GPRC5A
S100A14
TSPAN1
S100A16
MLPH
DCDC2
GRB14
SYTL2
ATP2C2
PTPRR
a


LTBP1
FBN1
MMP14
COL6A3
POSTN
COL5A2
THBS2
COL3A1
VCAN
COL12A1
CDH11
b

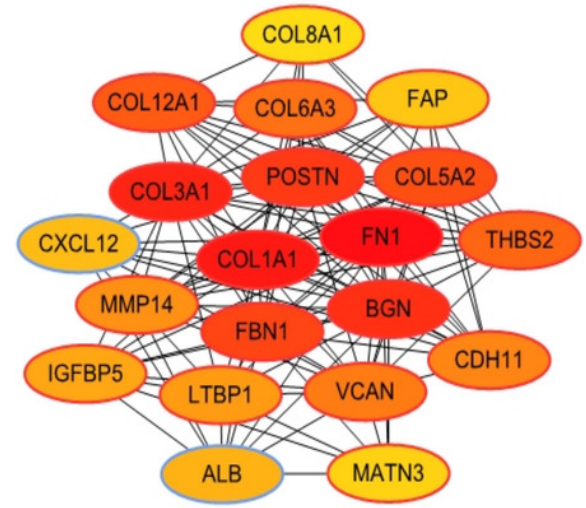

COL8A1
COL12A1
COL6A3
FAP
COL3A1
POSTN
COL5A2
CXCL12
FN1
THBS2
COL1A1
MMP14
BGN
FBN1
CDH11
IGFBP5
LTBP1
VCAN
ALB
MATN3
c

Figure 3

Results of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of the most significant module.

The blue color represents biological process (BP), the gray color represents molecular function (MF), the green color represents cellular component (CC), and the orange color represents KEGG pathways.

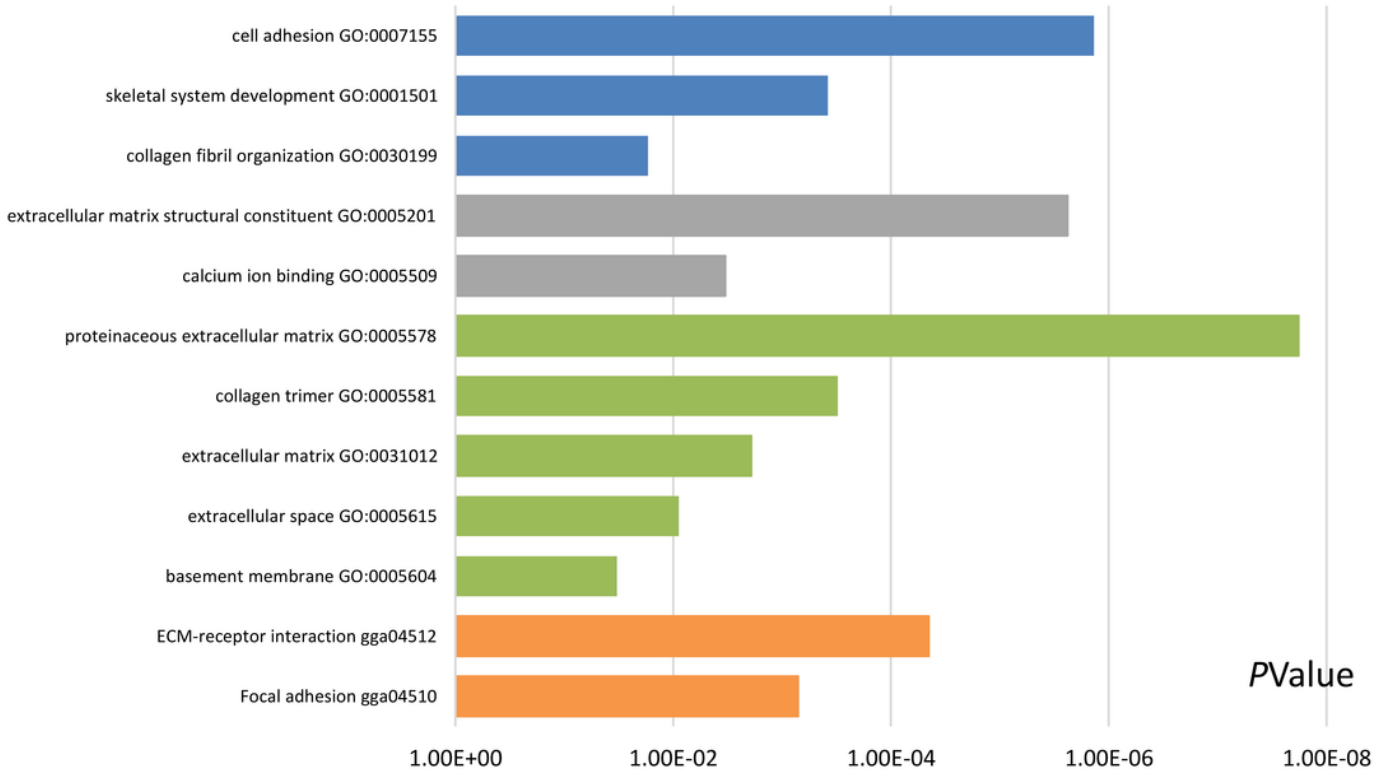


Figure 4

Results of GO and KEGG pathway analyses of 20 hub genes.

The blue color represents BP, the green color represents CC, the yellow color represents MF, and the orange color represents KEGG pathways.

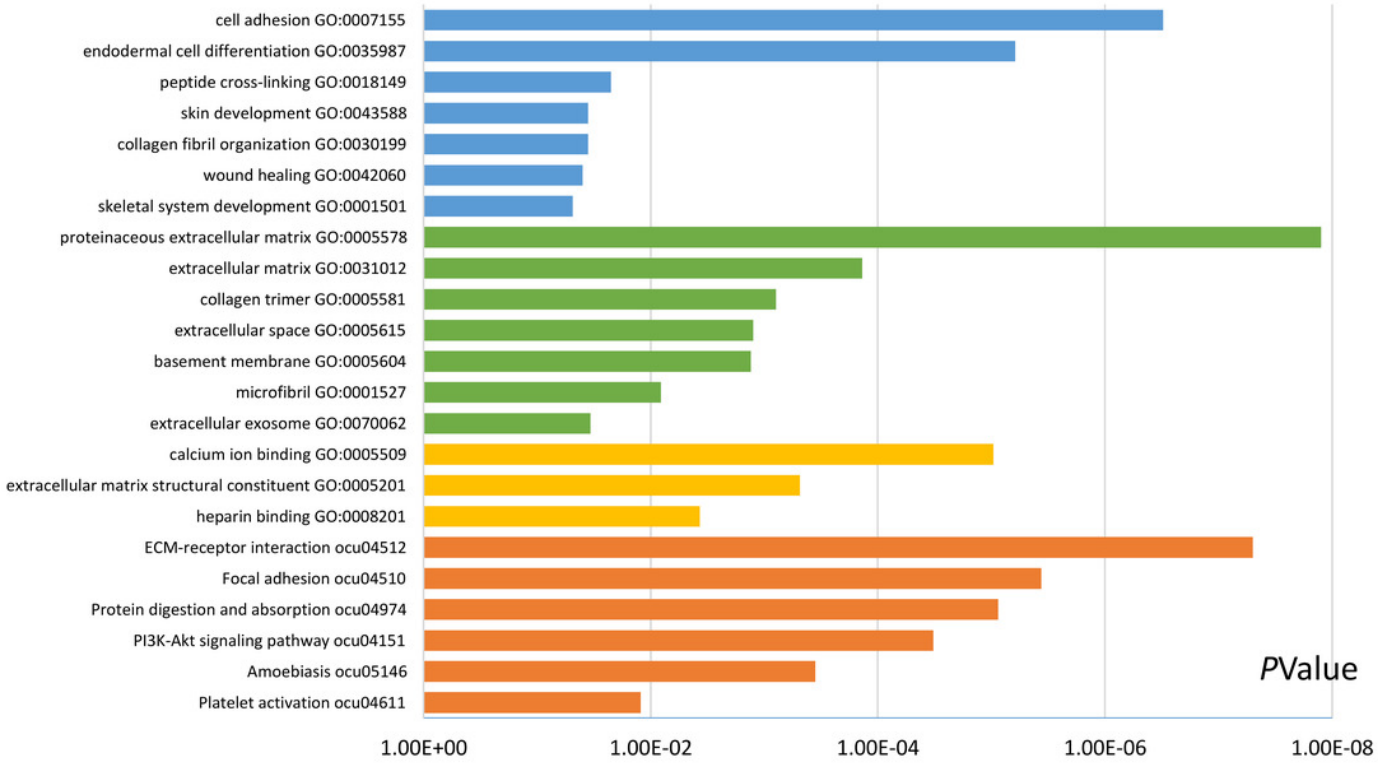


Figure 5

Overall survival analysis.

Overall survival curves for (A) *COL12A1* and (B) *MMP14* expression in PDAC patients in comparison with a high-risk group and a low-risk group. A value of $p < 0.05$ was regarded as statistically significant. HR = hazards ratio.

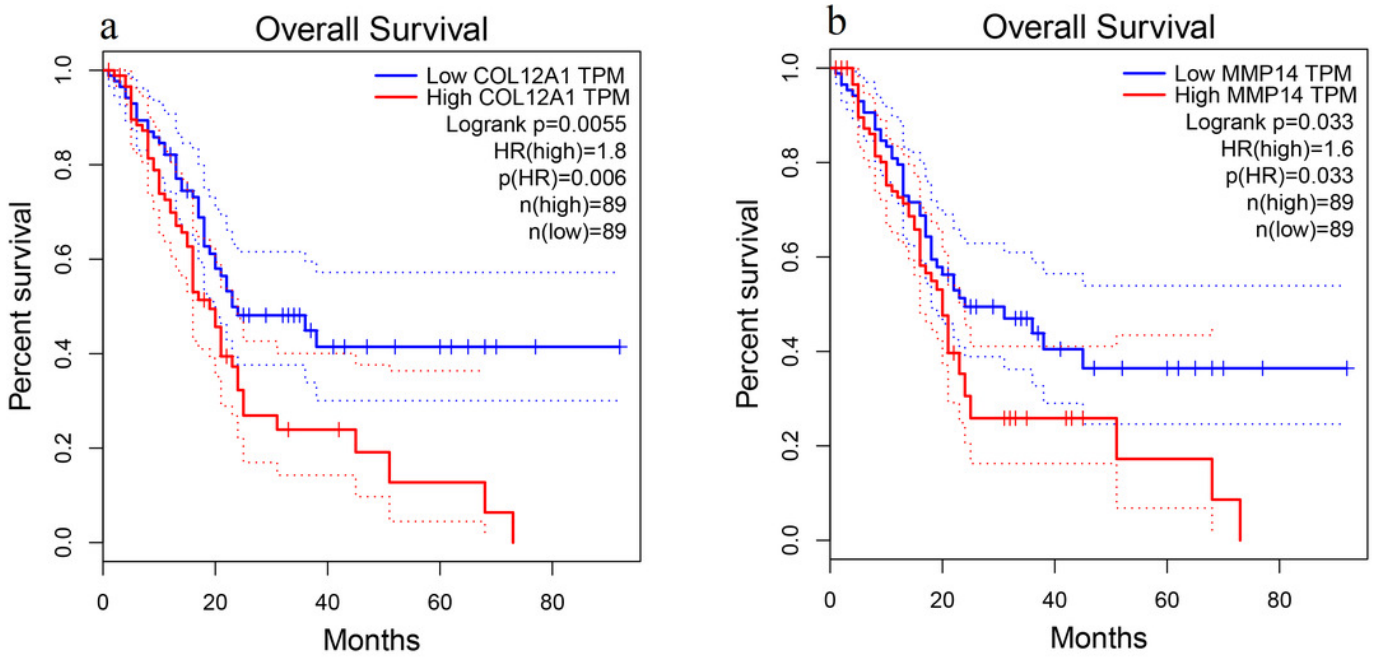


Figure 6

Validation of DEGs using The Cancer Genome Atlas and Genotype–Tissue Expression databases.

The boxplots were downloaded from the Gene Expression Profiling Interactive Analysis tool and are arranged in the following order: (A) *FN1*, (B) *COL1A1*, (C) *COL3A1*, (D) *BGN*, (E) *POSTN*, (F) *FBN1*, (G) *COL5A2*, (H) *COL12A1*, (I) *THBS2*, (J) *COL6A3*, (K) *VCAN*, (L) *CDH11*, (M) *MMP14*, (N) *LTBP1*, (O) *IGFBP5*, (P) *FAP*, (Q) *MATN3*, (R) *COL8A1*, and *(S) ALB*. A value of $p < 0.05$ was regarded as statistically significant. PAAD = pancreatic adenocarcinoma.

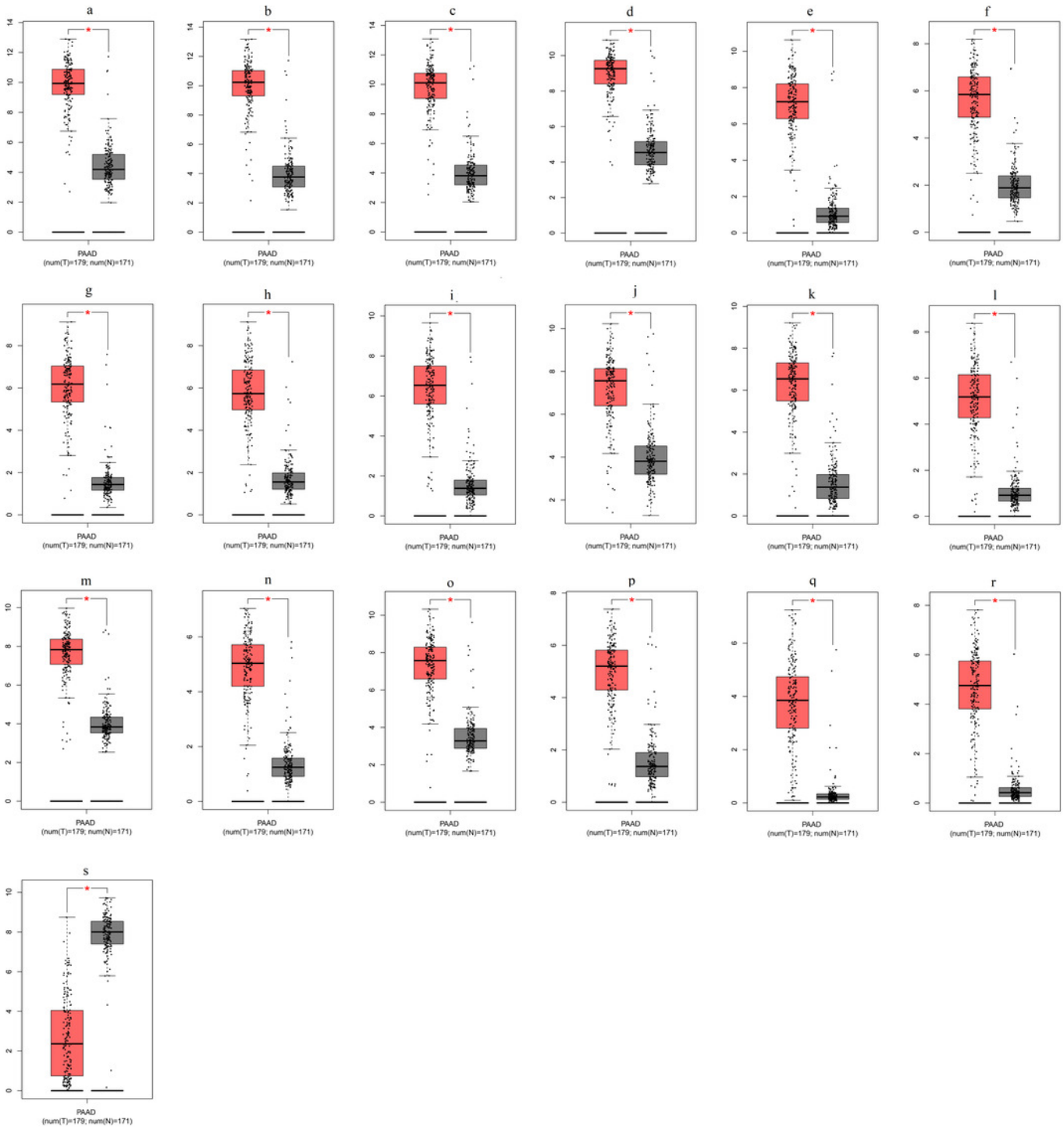

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
PAAD
(num(T)=179; num(N)=171)

Figure 7

Expression of 20 candidate DEGs in human pancreatic cancer specimens.

The immunohistochemical data were obtained from the Human Protein Atlas. Except for *COL5A2*, *IGFBP5*, and *MATN3*, expression profiles of the other 17 genes in PDAC clinical specimens are shown. Staining demonstrated that the protein expressions of (A) *FN1*, (B) *MMP14*, (C) *COL12A1*, (D) *COL3A1*, (I) *COL1A1*, (L) *FAP,* (M) *FBN1*, (N) *LTBP1*, (O) *POSTN*, and (Q) *VCAN* were higher in PDAC tissues than in normal pancreatic tissues, with only (F) *ALB* being downregulated in PDAC tissues. (E) *COL6A3*, (H) *CDH11*, (J) *COL8A1*, and (K) *CXCL12* were not expressed, whereas (G) *BGN* and (P) *THBS2* were overexpressed in both PDAC tissues and normal tissues.

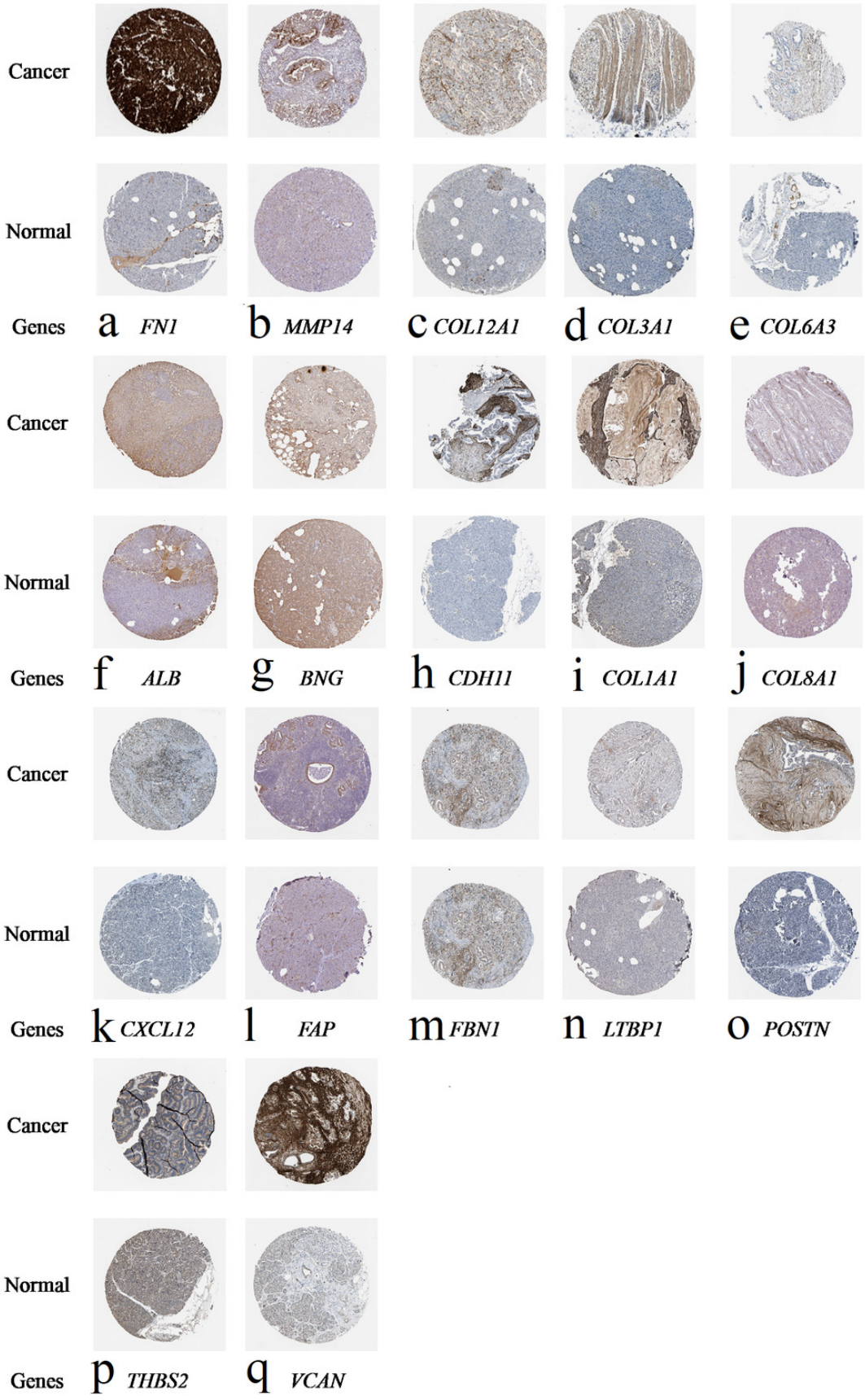

Cancer
Normal
Genes
a FN1
b MMP14
c COL12A1
d COL3A1
e COL6A3
Cancer
Normal
Genes
f ALB
g BNG
h CDH11
i COL1A1
j COL8A1
Cancer
Normal
Genes
k CXCL12
l FAP
m FBN1
n LTBP1
o POSTN
Cancer
Normal
Genes
p THBS2
q VCAN

Table 1(on next page)

A total of 263 DEGs were identified from the three microarray datasets, which consisted of 167 upregulated genes and 96 downregulated genes present in pancreatic ductal adenocarcinoma (PDAC) tissues in comparison with non-tumor tissues.

1

DEGs	Gene names
Upregulated	*XDH*, *RTKN2*, *PTPRR*, *ADAM12*, *STYK1*, *TPX2*, *PADI1*, *HEPH*, *CEACAM6*, *ITGA3*, *COL1A1*, *ANLN*, *FNDC1*, *PCDH7*, *SLC6A6*, *TRIM29*, *PXDN*, *EDNRA*, *LTBP1*, *MFAP5*, *PLA2R1*, *FN1*, *KRT17*, *PGM2L1*, *IFI27*, *ASAP2*, *LAMB3*, *TNFAIP6*, *HOXB5*, *OAS1*, *NTM*, *COL5A2*, *OSBPL3*, *TMPRSS4*, *ANTXR1*, *SDR16C5*, *OLR1*, *NT5E*, *CTSK*, *SULF2*, *MXRA5*, *APOL1*, *CDH11*, *AREG*, *MALL*, *S100A16*, *BGN*, *LAMA3*, *COL8A1*, *IGFBP5*, *MMP12*, *ADAMTS6*, *SLC2A1*, *CD109*, *ECT2*, *KIF23*, *MMP11*, *CDH3*, *LMO7*, *CCL18*, *ATP2C2*, *POSTN*, *MMP14*, *ADAM28*, *SRPX2*, *CEACAM5*, *TMC5*, *OAS2*, *MUC17*, *GABRP*, *COMP*, *SYTL2*, *GPX8*, *RUNX2*, *DLGAP5*, *KRT19*, *VCAN*, *MKI67*, *SULF1*, *LAMC2*, *GCNT3*, *NMU*, *MUC13*, *CEACAM1*, *ETV1*, *COL12A1*, *AGR2*, *ST6GALNAC1*, *SLC44A4*, *PLAU*, *S100P*, *SERPINB5*, *FOXQ1*, *TGM2*, *ITGB4*, *DCBLD2*, *TRIM31*, *RAI14*, *NRP2*, *SGIP1*, *CST1*, *ARNTL2*, *LEF1*, *MYOF*, *ANO1*, *S100A14*, *DDX60*, *KYNU*, *CAPG*, *CCL20*, *MATN3*, *NPR3*, *GPRC5A*, *NOX4*, *IL1RAP*, *ACSL5*, *HPGD*, *GREM1*, *SCEL*, *FBN1*, *IGFL2*, *SLC6A14*, *KRT6A*, *DHRS9*, *ANGPT2*, *MST1R*, *COL3A1*, *TMEM45B*, *EDIL3*, *ASPM*, *FAP*, *INPP4B*, *LOXL2*, *NQO1*, *CYP2C18*, *IFI44L*, *HK2*, *EFNB2*, *AEBP1*, *SLC16A3*, *CORIN*, *THBS2*, *BCAS1*, *DSG3*, *DKK1*, *RHBDL2*, *COL17A1*, *TSPAN1*, *FERMT1*, *CXCL5*, *COL6A3*, *COL10A1*, *ACTA2*, *PLAC8*, *AHNAK2*, *MLPH*, *FBXO32*, *TGFBI*, *KCNN4*, *CLDN18*, *FGD6*, *MTMR11*, *FXYD3*, *MBOAT2*, *SEMA3C*, *DPYSL3*, *CENPF*
Downregulated	*EPB41L4B*, *GSTA2*, *KIAA1324*, *CELA3A*, *ACADL*, *CEL*, *SLC39A5*, *LONRF2*, *SLC3A1*, *NRG4*, *MT1G*, *PROX1*, *G6PC2*, *C5*, *EGF*, *FAM3B*, *AQP8*, *CLPS*, *SLC17A4*, *CPB1*, *GP2*, *PDK4*, *RBPJL*, *PDIA2*, *PM20D1*, *CTRC*, *IAPP*, *PLA2G1B*, *ERP27*, *CELA2B*, *GRPR*, *REG1A*, *KIF1A*, *GUCA1C*, *CTRL*, *SYCN*, *CHRM3*, *TMED6*, *ALB*, *KCNJ16*, *REG3A*, *SLC4A4*, *AOX1*, *SERPINA5*, *CELA2A*, *SPINK1*, *FAM129A*, *FAM150B*, *SLC16A12*, *F11*, *CPA2*, *SV2B*, *BNIP3*, *C2CD4B*, *SLC1A2*, *REG1B*, *SCGN*, *PAK3*, *PRSS3*, *GRB14*, *REG3G*, *DCDC2*, *F8*, *GPHA2*, *EPHX2*, *PNLIPRP2*, *SLC7A2*, *CPA1*, *PRKAR2B*, *ONECUT1*, *BACE1*, *NUCB2*, *HOMER2*, *CXCL12*, *SLC43A1*, *GNMT*, *NR5A2*, *ALDH1A1*, *IL22RA1*, *BEX1*, *ANPEP*, *CFTR*, *FLRT2*, *LMO3*, *FGL1*, *NRCAM*, *FABP4*, *PNLIPRP1*, *KLK1*, *SERPINI2*, *GATM*, *DPP10*, *C6*, *SLC16A10*, *PRSS1*, *PAH*

2

Table 2(on next page)

Results of GO and KEGG pathway analyses of the most significant module .

1

Pathway ID	Pathway description	Count in gene set	*p*-value	FDR
GO:0007155	cell adhesion	5	1.37E−06	0.001121514
GO:0001501	skeletal system development	3	3.80E−04	0.310820683
GO:0030199	collagen fibril organization	2	0.017003019	13.1195401
GO:0005201	extracellular matrix structural constituent	4	2.39E−06	0.001487027
GO:0005509	calcium ion binding	4	0.003312243	2.045566581
GO:0005578	proteinaceous extracellular matrix	6	1.81E−08	1.29E−05
GO:0005581	collagen trimer	3	3.07E−04	0.217600344
GO:0031012	extracellular matrix	3	0.001912518	1.350246354
GO:0005615	extracellular space	4	0.008881226	6.138544445
GO:0005604	basement membrane	2	0.032936583	21.16653102
gga04512	ECM–receptor interaction	4	4.51E−05	0.019602271
gga04510	Focal adhesion	4	7.12E−04	0.309146908

2 Abbreviation: FDR = false discovery rate.

Table 3(on next page)

Results of GO and KEGG pathway analyses of 20 hub genes.

1

Pathway ID	Pathway description	Count in gene set	*p*-value	FDR
GO:0007155	cell adhesion	6	3.07E−07	3.44E−04
GO:0035987	endodermal cell differentiation	4	6.17E−06	0.006927836
GO:0018149	peptide cross-linking	2	0.02230059	22.35847168
GO:0043588	skin development	2	0.035457234	33.30840338
GO:0030199	collagen fibril organization	2	0.035457234	33.30840338
GO:0042060	wound healing	2	0.039805932	36.60570116
GO:0001501	skeletal system development	2	0.048448486	42.72191792
GO:0005578	proteinaceous extracellular matrix	7	1.25E−08	1.13E−05
GO:0031012	extracellular matrix	4	1.37E−04	0.123843274
GO:0005581	collagen trimer	3	7.87E−04	0.70897958
GO:0005615	extracellular space	6	0.001248571	1.122732142
GO:0005604	basement membrane	3	0.001310663	1.178271997
GO:0001527	microfibril	2	0.008113224	7.097635089
GO:0070062	extracellular exosome	7	0.033863185	26.75315613
GO:0005509	calcium ion binding	7	9.57E−06	0.007173793
GO:0005201	extracellular matrix structural constituent	3	4.84E−04	0.362368483
GO:0008201	heparin binding	3	0.003684116	2.728227814
ocu04512	ECM–receptor interaction	6	5.00E−08	4.15E−05
ocu04510	Focal adhesion	6	3.64E−06	0.003011517
ocu04974	Protein digestion and absorption	5	8.69E−06	0.007197854
ocu04151	PI3K-Akt signaling pathway	6	3.26E−05	0.026969277
ocu05146	Amoebiasis	4	3.54E−04	0.292970736
ocu04611	Platelet activation	3	0.012199259	9.669074187

2