RESEARCH ARTICLE

Analysis of Molecular Pathways in Pancreatic Ductal Adenocarcinomas with a Bioinformatics Approach

Yan Wang, Yan Li*

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a leading cause of cancer death worldwide. Our study aimed to reveal molecular mechanisms. Microarray data of GSE15471 (including 39 matching pairs of pancreatic tumor tissues and patient-matched normal tissues) was downloaded from Gene Expression Omnibus (GEO) database. We identified differentially expressed genes (DEGs) in PDAC tissues compared with normal tissues by limma package in R language. Then GO and KEGG pathway enrichment analyses were conducted with online DAVID. In addition, principal component analysis was performed and a protein-protein interaction network was constructed to study relationships between the DEGs through database STRING. A total of 532 DEGs were identified in the 38 PDAC tissues compared with 33 normal tissues. The results of principal component analysis of the top 20 DEGs could differentiate the PDAC tissues from normal tissues directly. In the PPI network, 8 of the 20 DEGs were all key genes of the collagen family. Additionally, FN1 (fibronectin 1) was also a hub node in the network. The genes of the collagen family as well as FN1 were significantly enriched in complement and coagulation cascades, ECM-receptor interaction and focal adhesion pathways. Our results suggest that genes of collagen family and FN1 may play an important role in PDAC progression. Meanwhile, these DEGs and enriched pathways, such as complement and coagulation cascades, ECM-receptor interaction and focal adhesion may be important molecular mechanisms involved in the development and progression of PDAC.

Keywords: Pancreatic ductal adenocarcinoma - functional enrichment analysis - principal component analysis

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common pancreatic neoplasm that arising within the ductal of the pancreas, comprising over 85% of all pancreatic tumor cases (Li et al., 2004; Hezel et al., 2006). Given the few early indicators of illness, it is diagnosed late in the natural history of the disease and lack of screening tests (Michaud, 2004). At present, the standard care for the 20% of patients with localized disease is surgery followed by chemotherapy, and radiation in some cases (Neuhaus et al., 2008). Until now, treatment has not improved substantially and has little effect on prolonging survival time (Michaud, 2004). Additionally, the genetic and biology of PDAC remain unclear until now. Hence, exploring the molecular mechanism may contribute to the development of effective therapeutic approaches for this disease.

PDAC is a multifactorial disease influenced by a variety of risk factors, including smoking, chronic pancreatitis and other systemic inflammatory factors (Pandol et al., 2012; Malats, 2014). Meanwhile, considerable advances have been made in understanding the genetic alterations involved in the progression PDAC (Ostmanand Augsten, 2009; Jiao et al., 2011). Several cancer-associated genes

have been identified in PDAC, among which the most frequently affected are *DPC4* (SMAD family member 4), K-*ras* (Kirsten rat sarcoma viral oncogene homolog) and *p53* (Dempe et al., 2010; Morton et al., 2010; Navas et al., 2012), which all appear to play an important role for the development of PDAC. Additionally, it has been reported that germline mutations have been linked to familial PDAC, including the DNA mismatch repair gene *MLH1* and the cationic trypsinogen gene PRSS1 (protease, serine, 1 (trypsin 1)), and as well those targeting the tumor suppressor genes INK4A (cyclin-dependent kinase 4), *BRCA2* (breast cancer 2, early onset), and LKB1 (serine/threonine kinase 11) (Whitcomb et al., 1996; Jaffee et al., 2002). Besides, accumulating evidence demonstrates that the aggressive nature of PDAC can be enhanced via interactions between the epithelial and stromal compartments, and extracellular matrix (ECM) proteins may play a key role in PDAC progression (Grippo et al., 2012). These evidences all suggest that the occurrence and development of PDAC are a complex process. Even though progresses have been made, the molecular mechanisms underlying its development are still unclear.

Microarray analysis is increasingly being used to identify genes or gene signatures that are strongly

*Department of Gastroenterology, Shengjing Hospital of China Medical University, Shenyang, China *For correspondence: in_liyan@163.com*

Yan Wang and Yan Li

associated with pancreatic cancer (Campagna et al., 2008; Bournet et al., 2012). In previous study, Shi et al. screened differentially expressed genes (DEGs) and constructed a co-expression network to identify dysregulated pathways mainly involved in specific cellular functions, such as immune response, homeostasis and cell adhesion, in which microarray data GSE15471 was used (Shi et al., 2014). In addition, microarray data GSE15471 was also utilized to screen the Master Regulators (MRs) of transcription involved in PDAC disease and highlighted the potential value of Tubby-like protein 3 (*TULP3*) as a clinical prognostic biomarker (Sartor et al., 2014). In contrast to these findings, we downloaded microarray data GSE15471 in the present study to identify DEGs in PDAC tissues compared with patient-matched normal pancreas tissues. Then dimensionality of the top 20 DEGs was analyzed by principal component analysis. Besides, functional enrichment analysis was performed and protein-protein interaction (PPI) network was constructed. Our study aimed to improve the understanding to the underlying molecular mechanisms of PDAC and provide novel insights for the early diagnosis and medication control of PDAC.

Materials and Methods

Affymetrix microarray data

The transcriptional profile of GSE15471 deposited by Badea et al. (Badea et al., 2008) was downloaded from Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo/) in the National Center for Biotechnology Information (NCBI). Total 78 samples, including 39 PDAC tissues and patient-matched normal pancreas tissues from 39 PDAC patients, were available for microarray analysis based on the GPL6244 Platform (Affymetrix Human Gene 1.0 ST Array) (Affymetrix Inc., Santa Clara, California, USA).

Data preprocessing

The GSE14905 raw data in CEL files were converted into probe expression matrix by the ReadAffy of affy package in R language (Gautier et al., 2004). The expression values of all probes were normalized by the robust multiarray average algorithm (RMA) (Irizarry et al., 2003). Then the probe number was converted as gene symbol using the R/Bioconductor annotation package or based on the annotation information of Affymetrix Human Gene 1.0 ST Array. If multiple probes corresponded to the same gene, the mean expression value was calculated as the gene value.

Identification of DEGs and hierarchical clustering analysis

For the GSE15471 dataset, the paired t-test of samr package in R language (SAM method) (Tusher et al., 2001) was used to identify DEGs in PDAC tissues compared to patient-matched normal pancreas tissues. When the Del value was set at 4.0, only DEGs with the false discovery rate (FDR) < 0.01 and fold change (FC) value >2 or < 1/2 were selected. In addition, we performed hierarchical

clustering analysis to better visualize the gene expression values with drawing dendrogram. In the hierarchical clustering analysis, the gene expression value of each sample was clustered by Pearson coefficient (Derrick et al., 1994) and Spearman coefficient of statistics respectively, and the unreasonable clustered samples were filtered. In order to reduce systematic errors of the sample grouping that could induce interference on the subsequent analysis, we filtered the gene expression values of unreasonable samples. Then DEGs in PDAC tissues compared to patient-matched normal pancreas tissues were screened again using the limma package (Smyth, 2004). The adjusted p value < 0.01 and |log2 FC| > 1.5 were used as the cut-off criteria.

Functional enrichment analysis of DEGs after sample filtration

The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) is a functional annotation tool for investigators to understand biological meaning behind large list of genes (Huang et al., 2008). Gene Ontology (GO) (Ashburner et al., 2000) is a widely used for functional studies of large-scale genomic data, which mainly contains biological process (BP), molecular function (MF), and cellular component (CC). Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisaand Goto, 2000) is a pathway-related database for providing the information how molecules or genes are networked. In this study, we performed GO-BP and KEGG pathway enrichment analysis for the screened up- and down-regulated DEGs by online DAVID tool. The overrepresent GO-BP terms were screened only with the adjusted p value less than 1e-5, and the significant enriched KEGG pathways were identified with FDR less than 0.05.

Principal component analysis of top 20 DEGs

To distinguish PDAC tissues from normal tissues, 20 DEGs with the minimum significant p value were identified to conduct principal component analysis. Principal component analysis (Raychaudhuri et al., 2000) is a statistical technique for determining the main variables in a multidimensional data set that stand for the differences in the observations, which can simplify the analysis and visualize the direction of multidimensional data sets. Therefore, we can just use several key variables rather than thousands of variables to classify the samples by means of principal component analysis.

PPI network construction of the top 20 DEGs

The protein-protein interactions (PPIs) was used to discover the interactions between proteins and to reveal the functions of proteins at the molecular level (Giot et al., 2003). Search Tool for the Retrieval of Interacting Genes (STRING) (Jensen et al., 2009) database collected the information of predicted and experimental PPIs in a given cell. Thus the top 20 DEGs were mapped STRING database to construct PPI network of them. The combined score of protein pairs >0.4 was considered as the cutoff value.

Results

Data preprocessing

We obtained 19944 gene expression values of the paired PDAC tissues and patient-matched normal pancreas tissues from 39 patients after data preprocessing. As shown in Figure 1, all the median of gene expression values of the 78 samples were in a straight line after being normalized, indicating that the data had been normalized and could be used for further study.

Identification of DEGs and hierarchical clustering analysis of the paired PDAC tissues and normal tissues

A total of 727 DEGs, including 643 up-regulated genes and 84 down-regulated ones, were identified in the PDAC tissues compared to patient-matched normal tissues. Total 6 normal tissue samples (N40875, N40892, N40726, N40975, N51294, and N30308) and 1 PDAC tissue sample

T30308 were found gathered in PDAC tissue sample clusters and normal tissue sample clusters respectively (Figure 2), which were unreasonable. Notably, part of the patient-matched normal pancreas tissues can be grouped into two clusters through hierarchical clustering, the distance from one of the clusters to the PDAC tissues closer than between the normal tissues. After sample filtration, total 532 DEGs in 38 PDAC tissue samples compared to 33 normal controls were identified again by using the limma package, including 450 up-regulated genes and 82 down-regulated ones. The number of up-regulated genes was significantly more than down-regulated ones. The top20 DEGs was shown in Table 1.

Functional enrichment analysis of DEGs after sample filtration

A total of 19 overrepresent GO-BP terms (Table 2A) related to immune response and trauma dealing, and only 3 signaling pathways (Table 2B), including Complement and coagulation cascades, ECM-receptor interaction and Focal adhesion, were significantly enriched by the up-regulated genes using DAVID tool. However, no significant GO-BP terms and pathways were enriched by down-regulated genes.

Principal component analysis of the top 20 DEGs

The results of principal component analysis of the top 20 DEGs could differentiate the PDAC tissues from normal tissues directly (Figure 3). The first constituted principal component explained 94.72% of the variance of the 20 variables, the second principal component explained 1.84% of the variance, and the cumulative variance that explained is 96.56%. Moreover, variable pointing of the 20 DEGs had more great effect on PDAC tissues. Notably, the expression levels of these 20 genes in PDAC tissues were significantly higher than the normal tissues.
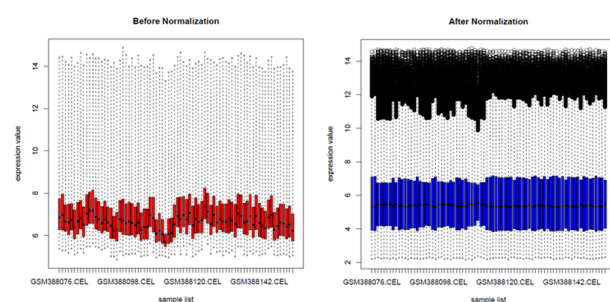

Figure 1. Expression Data Box Plots of all Samples before and After Normalized. The horizontal axis represents the sample name, and the ordinate axis represents the expression value. The black line in box represents the median of each sample; we can determine the degree of standardization by the distribution, the black line in box in almost the same straight line, indicating a good degree of normalization. The left is box plot of expression values of various samples before normalized; the right is box plot of expression values of various samples after normalized

Table 1. The 20 Most Differentially Expressed Genes in PDAC Tissues

Gene	logFC	t	adj.P.Value	Gene description
VCAN	4.612069	25.46883	2.41E-33	versican
SULF1	4.867507	25.10537	3.05E-33	sulfatase 1
COL8A1	2.651996	24.73784	5.26E-33	collagen, type VIII, alpha 1
FAP	4.041172	24.45816	8.18E-33	fibroblast activation protein, alpha
COL1A1	2.812473	23.65893	4.55E-32	collagen, type I, alpha 1
INHBA	3.595737	23.67401	4.55E-32	inhibin, beta A
THBS2	4.68946	23.52303	5.63E-32	thrombospondin 2
CTHRC1	5.201109	21.9905	3.45E-30	collagen triple helix repeat containing 1
COL1A2	4.113832	21.87112	4.30E-30	collagen, type I, alpha 2
COL5A2	4.257605	21.37353	1.62E-29	collagen, type V, alpha 2
COL6A3	3.293194	21.01047	4.27E-29	collagen, type VI, alpha 3
FN1	3.39725	20.58908	1.36E-28	fibronectin 1
COL10A1	5.185537	20.43247	2.01E-28	collagen, type X, alpha 1
COL3A1	3.333812	20.15564	4.30E-28	collagen, type III, alpha 1
PLXDC2	2.564791	20.06903	5.22E-28	plexin domain containing 2
TIMP1	3.606368	19.84014	9.83E-28	TIMP metallopeptidase inhibitor 1
LOXL1	2.7619	19.80988	1.02E-27	lysyl oxidase-like 1
DACT1	2.845461	19.77039	1.08E-27	dapper, antagonist of beta-catenin, homolog 1 (Xenopus laevis)
AEBP1	3.171337	19.53532	2.12E-27	AE binding protein 1
COL5A1	3.907796	19.2491	4.90E-27	collagen, type V, alpha 1

Table 2. Functional Enrichment Analysis of the Differentially Expressed Genes

GO-BP terms and KEGG pathways	Count	Fold Enrichment	FDR
A: The enriched GO-BP terms			
GO:0009611 response to wounding	65	4.753852	1.03E-22
GO:0006955 immune response	67	3.763864	7.59E-18
GO:0006952 defense response	57	3.592592	1.89E-13
GO:0006954 inflammatory response	41	4.889998	1.89E-13
GO:0007155 cell adhesion	59	3.267098	2.48E-12
GO:0022610 biological adhesion	59	3.262437	2.48E-12
GO:0030198 extracellular matrix organization	21	7.826978	2.86E-09
GO:0002541 activation of plasma proteins involved in acute inflammatory response	14	12.62024	5.39E-08
GO:0019724 Bcell mediated immunity	15	10.38273	1.60E-07
GO:0006959 humoral immune response	17	8.341228	2.21E-07
GO:0043062 extracellular structure organization	23	5.46951	3.27E-07
GO:0006956 complement activation	13	11.99782	6.58E-07
GO:0002526 acute inflammatory response	18	7.119584	7.55E-07
GO:0016064 immunoglobulin mediated immune response	14	10.04945	1.27E-06
GO:0030199 collagen fibril organization	11	14.70289	2.48E-06
GO:0006958 complement activation, classical pathway	11	14.70289	2.48E-06
GO:0002252 immune effector process	20	5.7854	2.49E-06
GO:0002449 lymphocyte mediated immunity	15	8.306181	3.85E-06
GO:0002455 humoral immune response mediated by circulating immunoglobulin	11	13.75432	5.24E-06
B: The enriched KEGG pathways			
hsa04610 Complement and coagulation cascades	22	10.39298	2.44E-13
hsa04512 ECM-receptor interaction	21	8.149038	3.15E-10
hsa04510 Focal adhesion	25	4.054248	5.93E-06

*BP is the abbreviation of biological process; FDR is the abbreviation of false discovery rate
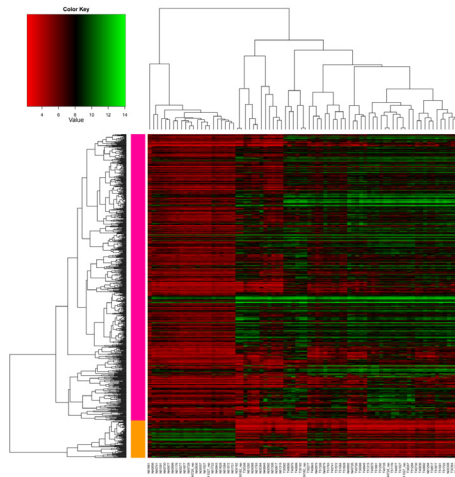

Figure 2. Hierarchical Clustering Diagram of Differences in Pancreatic Ductal Adenocaicinoma. The horizontal axis represents the sample name (N indicates the patient-matched normal tissues; T indicates the pancreatic ductal adenocarcinoma tissues). The right ordinate axis represents the clustering condition of gene; the upper horizontal axis represents the clustering situation of the sample. Red indicates the down regulation of the gene, green indicates the up regulation. Pancreatic ductal adenocarcinoma tissues and normal tissues are mainly grouped into 3 clusters. Total 6 normal tissue samples (N40875, N40892, N40726, N40975, N51294, and N30308) and 1 PDAC tissue sample (T30308) were found gathered in PDAC tissue sample clusters and normal tissue sample clusters respectively, which are unreasonable

Figure 3. Principal Component Analysis of the 20 Differentially Expressed Genes that Most Significantly Expressed. The horizontal axis represents the score of the first principal component of each sample; the ordinate axis represents the score of the second principal component. Of which the first principal component explains 94.72% of the variance of the 20 variables, the second principal component explains 1.84% of the variance, and the cumulative variance that explained is 96.56%. N represents the patient-matched normal tissue samples (total of 33), T represents the carcinoma tissue samples (total of 38), the red lines indicate the influencing size of the 20 gene variables to pancreatic ductal adenocarcinoma tissues, and these 20 genes are all highly expressed in pancreatic ductal adenocarcinoma tissues

PPI network construction of the top 20 DEGs

Based on the information of STRING database, PPI network of the most significant 20 genes in PDAC tissues were constructed. As shown in the Figure 4, no interaction of SULF1 (sulfatase 1), PLXDC2 (plexin domain containing 2) and DACT1 (dishevelled-binding antagonist
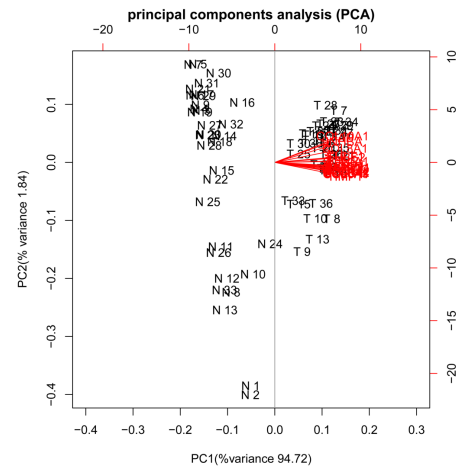
of beta-catenin 1) with other proteins was found presently, while the rest 17 constituted 72 PPI pairs. Strikingly, among the rest 17 DEGs, 8 DEGs, such as COL3A1 (collagen, type III, alpha 1), COL1A2 (collagen, type I, alpha 2), COL1A1 (collagen, type I, alpha 1), COL5A2 (collagen, type V, alpha 2), and COL5A1 (collagen, type
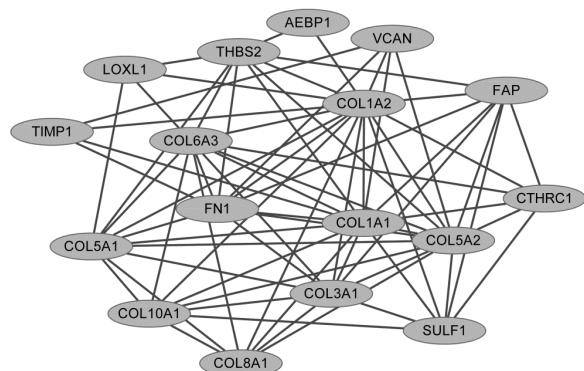
Figure 4. PPI Network of the top 20 Differentially Expressed Genes. The nodes represent 17 genes of the 20; lines represent the protein-protein interaction relationships that corresponding to the genes

V, alpha 1), were found to be members of collagen family. Additionally, FN1 (fibronectin 1) was also hub nodes in the network. Besides, these proteins of collagen family as well as FN1 are all key molecules involved in the above 3 enriched pathways.

Discussion

PDAC is the most lethal type of cancer and therapeutic interventions currently available for it are ineffective (Ishiguro et al., 2014). The identification of highly expressed genes in PDAC is critical to the development of novel strategies to detect and treat this highly lethal cancer (Iacobuzio-Donahue et al., 2003). In this study, we investigated the underlying molecular mechanism of PDAC development by using bioinformatics methods. The results of principal component analysis showed that the top 20 DEGs could differentiate the PDAC tissues from normal tissues directly. Strikingly, 8 of the top 20 highly expressed DEGs were key proteins of collagen family. Additionally, FN1 was also hub nodes in the network and could interact with these proteins of collagen family directly, suggesting these DEGs may be responsible for PDAC development.

The members of collagen family are the most abundant proteins in extracellular matrix and is a major component of epithelial basement membranes, which contribute to the structural integrity of epithelial organs via providing tensile strength to the interstitial matrix (Gelse et al., 2003). Previous study demonstrates that loss of basement membrane integrity is one key and prognostic event in spreading pancreatic cancer (Apte et al., 2004). In addition, collagens exert important functions in the storage and release of cellular mediators, including growth factors (GelsePöschl et al., 2003). Very recent studies confirm that the multifunctional growth factor midkine promotes cellular functions leading to increased proliferation, migration, and survival in PDAC (Rawnaq et al., 2014). The chronic stimulation of epidermal growth factor is found to cause Smad4-dependent signaling bypass in PDAC (Moz et al., 2014). Besides, previous study has revealed that the genes of collagen family are differentially expressed in PDAC tissue and significantly involved in

pathways of focal adhesion (Lin et al., 2014b), which are in line with our results. Therefore, our results suggest that genes of collagen family may responsible for progression of PDAC.

In addition to collagen family proteins, FN1 was also hub nodes in the network and could interact with these proteins of collagen family proteins directly. Fibronetin (FN) is a major component of the extracellular matrix (ECM) and functions in cell adhesion, cell spreading and cell migration (Stenzel et al., 2011). Moreover, the inhibition of FN-dependent PDAC cell proliferation may be a crucial mechanism of significant antitumor effects of EMAP II against PDAC (Schwarz et al., 2010). Studies have discovered that FN stimulates invasion and adhesion and markedly inhibits cell death of pancreatic cancer cells (Vaquero et al., 2003; Edderkaoui et al., 2007). Besides, FN receptors, β1 and β3 integrins, have been shown to be up-regulated in PDAC cells (Linder, 2001). Therefore, our findings suggest that FN1 may be a key molecule involved in the development and progression of PDAC and could be used as a specific therapeutic molecular target in the treatment of this disease.

Strikingly, our results showed that these proteins of collagen family as well as FN1 were significantly enriched in complement and coagulation cascades, ECM-receptor interaction and focal adhesion. The work of Zhao et al. confirmed that the above 3 enriched pathways were significantly enriched and were regarded to be key mechanisms of pancreatic cancer (Zhao et al., 2014). ECM receptor interaction combined with complement and coagulation cascades are identified in multiple cancers suggesting an essential role of these pathways in cancer biology (Krupp et al., 2011). Focal adhesion is likely linked to pancreatic cancer, and the extracellular matrix regulates pancreatic cancer stem cell function via focal adhesion kinase signaling (Furuyama et al., 2006; Rasheed et al., 2012). In addition, genes of collagen family in PDAC tissue are involved in pathways of ECM-receptor interaction and focal adhesion (Lin et al., 2014a). Collagens, such as COL1A1, COL1A2 and COL3A1, and FN1 are found to be involved in ECM-receptor interaction (Cassar-Malek et al., 2007). Thus, we speculate that the above 3 enriched pathways are more likely to be the crucial mechanisms involved in PDAC. Further, the results suggest that these proteins of collagen family as well as FN1 may play important roles in PDAC progression via involving in these pathways.

In conclusion, we screened the DEGs and investigated the underlying molecular mechanism of PDAC by using bioinformatics methods. The genes of collagen family and FN1 may play an important role in PDAC progression. Meanwhile, complement and coagulation cascades, ECM-receptor interaction and focal adhesion are significantly enriched pathways involved in PDAC. These DEGs and enriched pathways may be important molecular mechanisms for promoting the development and progression of PDAC. Results from this study will provide the groundwork for the understanding of PDAC. However, further experiments are still needed to be explored and researched.

Yan Wang and Yan Li

References

Apte M, Park S, Phillips P, et al (2004). Desmoplastic reaction in pancreatic cancer: role of pancreatic stellate cells. *Pancreas*, **29**, 179-87.

Ashburner M, Ball CA, Blake JA, et al (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25-9.

Badea L, Herlea V, Dima SO, et al (2008). Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia-the authors reported a combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepato-gastroenterol*, **55**, 2016.

Bournet B, Pointreau A, Souque A, et al (2012). Gene expression signature of advanced pancreatic ductal adenocarcinoma using low density array on endoscopic ultrasound-guided fine needle aspiration samples. *Pancreatol*, **12**, 27-34.

Campagna D, Cope L, Lakkur SS, et al (2008). Gene expression profiles associated with advanced pancreatic cancer. *Int J Clin Exp Pathol*, **1**, 32-43.

Cassar-Malek I, Passelaigue F, Bernard C, et al (2007). Target genes of myostatin loss-of-function in muscles of late bovine fetuses. *BMC genomics*, **8**, 63.

Dempe S, Stroh-Dege AY, Schwarz E, et al (2010). SMAD4: a predictive marker of PDAC cell permissiveness for oncolytic infection with parvovirus H-1PV. *Int J Cancer*, **126**, 2914-27.

Derrick TR, Bates BT, Dufek JS (1994). Evaluation of time-series data sets using the Pearson product-moment correlation coefficient. *Med Sci Sports Exercise*, **26**, 919-28.

Edderkaoui M, Hong P, Lee JK, et al (2007). Insulin-like growth factor-I receptor mediates the prosurvival effect of fibronectin. *J Biol Chem*, **282**, 26646-55.

Furuyama K, Doi R, Mori T, et al (2006). Clinical significance of focal adhesion kinase in resectable pancreatic cancer. *World J Surg*, **30**, 219-26.

Gautier L, Cope L, Bolstad BM , Irizarry RA (2004). Affy-analysis of affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307-15.

Gelse K, Poschl E , Aigner T (2003). Collagens-structure, function, and biosynthesis. *Adv Drug Delivery Reviews*, **55**, 1531-46.

Giot L, Bader J S, Brouwer C, et al (2003). A protein interaction map of *Drosophila melanogaster. Science*, **302**, 1727-36.

Grippo PJ, Munshi HG, Rasheed ZA, et al (2012). Pathology of pancreatic stroma in PDAC. **JOURNAL DETAILS**

Hezel AF, Kimmelman AC, Stanger BZ, et al (2006). Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev*, **20**, 1218-49.

Huang DW, Sherman BT , Lempicki RA (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**, 44-57.

Iacobuzio-Donahue CA, Ashfaq R, Maitra A, et al (2003). Highly expressed genes in pancreatic ductal adenocarcinomas A comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. *Cancer Res*, **63**, 8614-22.

Irizarry R, Hobbs B, Collin F, et al (2003). Biostatistics. Vol. 4. **DETAILS** Oxford, England.

Ishiguro S, Yoshimura K, Takao S, et al (2014). Involvement of an angiotensin II type 2 receptor (AT2R) signalling in human pancreatic ductal adenocarcinoma (PDAC): a novel AT2R agonist effectively attenuates growth of PDAC grafts in mice. *Cancer Res*, **74**, 4606-.

Jaffee E M, Hruban R H, Canto M , Kern S E (2002). Focus on pancreas cancer. *Cancer Cell*, 2, 25-8.

Jensen LJ, Kuhn M, Stark M, et al (2009). STRING 8-a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, **37**, 412-6.

Jiao Y, Shi C, Edil BH, et al (2011). DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*, **331**, 1199-203.

Kanehisa M, Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, **28**, 27-30.

Krupp M, Maass T, Marquardt JU, et al (2011). The functional cancer map: a systems-level synopsis of genetic deregulation in cancer. *BMC Medical Genomics*, **4**, 53.

Li D, Xie K, Wolff R , Abbruzzese JL (2004). Pancreatic cancer. *Lancet*, **363**, 1049-57.

Lin Y, Jin Y, Lin LJ, et al (2014a). Candidate agents for pancreatic ductal adenocarcinoma identified by a sub-pathway based method. *Gene*, **540**, 232-7.

Linder S (2001). Immunohistochemical expression of extracellular matrix proteins and adhesion molecules in pancreatic carcinoma, **48**, 1321-7.

Malats N (2014). Pancreatic cancer: risk factors. *Pancreatol*, **14**, 2.

Michaud D (2004). Epidemiology of pancreatic cancer. *Minerva Chirurgica*, **59**, 99-111.

Morton JP, Timpson P, Karim SA, et al (2010). Mutant *p53* drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc Natl Acad Sci USA*, **107**, 246-51.

Moz S, Bozzato D, Galozzi P, et al (2014). Epidermal growth factor (EGF) chronic stimulation causes Smad4-dependent signaling bypass in PDAC. *Pancreatol*, **14**, 69.

Navas C, Hernandez-Porras I, Schuhmacher A J, et al (2012). EGF receptor signaling is essential for K Ras oncogene-driven pancreatic ductal adenocarcinoma. *Cancer cell*, **22**, 318-30.

Neuhaus P, Riess H, Post S, et al (2008). CONKO-001: Final results of the randomized, prospective, multicenter phase III trial of adjuvant chemotherapy with gemcitabine versus observation in patients with resected pancreatic cancer (PC). *J Clin Oncol*, **26**, 15.

Ostman A , Augsten M (2009). Cancer-associated fibroblasts and tumor growth-bystanders turning into key players. *Current Opinion Genetics Dev*, **19**, 67-73.

Pandol SJ, Apte MV, Wilson JS, et al (2012). The burning question: why is smoking a risk factor for pancreatic cancer? *Pancreatol*, **12**, 344-9.

Rasheed ZA, Jung C, Huang A, et al (2012). Abstract A70: The extracellular matrix regulates pancreatic cancer stem cell function via focal adhesion kinase signaling. *Cancer Res*, **72**, 70.

Rawnaq T, Dietrich L, Wolters-Eisfeld G, et al (2014). The multifunctional growth factor midkine promotes proliferation and migration in pancreatic cancer. *Molecular Cancer Res*, **12**, 670-80.

Raychaudhuri S, Stuart JM , Altman RB (2000). Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput*, **2000**, 455-66.

Sartor I, Zeidán-Chuliá F, Albanus R, et al (2014). Computational analyses reveal a prognostic impact of *TULP3* as a transcriptional master regulator in pancreatic ductal adenocarcinoma. *Molecular BioSystems*, **10**, 1461-8.

Schwarz R E, Awasthi N, Konduri S, et al (2010). Antitumor effects of EMAP II against pancreatic cancer through inhibition of fibronectin-dependent proliferation. *Cancer Biology Therapy*, **9**, 632-9.

Shi B, Wang X, Han X, et al (2014). Functional modules analysis based on coexpression network in pancreatic ductal adenocarcinoma. *Pathol Oncol Res*, **20**, 293-9.

Smyth G K (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, **3**, 3.

Stenzel D, Lundkvist A, Sauvaget D, et al (2011). Integrin-dependent and -independent functions of astrocytic fibronectin in retinal angiogenesis. *Dev*, **138**, 4451-63.

Tusher V G, Tibshirani R , Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA*, **98**, 5116-21.

Vaquero EC, Edderkaoui M, Nam KJ, et al (2003). Extracellular matrix proteins protect pancreatic cancer cells from death via mitochondrial and nonmitochondrial pathways. *Gastroenterol*, **125**, 1188-202.

Whitcomb DC, Gorry MC, Preston RA, et al (1996). Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nature Genetics*, **14**, 141-5.

Zhao LL, Zhang T, Liu BR, et al (2014). Construction of pancreatic cancer double-factor regulatory network based on chip data on the transcriptional level. *Molecular biol Rep*, **41**, 2875-83.