

Gene Expression Profiling of Pancreatic Cancer Reveals a Significant Deregulation of the TGF- β Pathway and the Discovery of Genes for Prognosis

A. Klein^{*1} and H. Pospisil²

¹Charité Medical University Berlin, Institute of Biochemistry, Charitéplatz 1 / Sitz: Virchowweg 6, 10117 Berlin, Germany

²Technical University of Applied Sciences Wildau, High Performance Computing in Life Sciences / Bioinformatics, Hochschulring 1, 15745 Wildau, Germany

Abstract: We have re-analyzed previously published gene expression data from ninety-four pancreatic ductal adenocarcinomas (PDAC) samples. We determined the gene expression profile of genes differentially expressed in PDAC compared to non-malignant pancreatic tissue. Using the 100 top-ranked genes, we were able to discriminate between PDAC and non-malignant pancreatic tissue. A hierarchical cluster analysis revealed only a 6 % false discovery rate. The prognostic strength of these discriminative genes was underscored by a SVM classification and 3-fold cross validation with an 89 % correct class assignment. The annotation of the 100 top-ranked genes revealed that most of the genes were involved in the processes of signal transduction, cell adhesion, extracellular matrix organization and cell migration. The most greatly affected signal cascade was the transforming growth factor β receptor signaling pathway, which was significantly enriched in the top-ranked genes. Furthermore, we identified eleven genes that were associated with good prognosis.

Keywords: Pancreatic ductal adenocarcinoma, gene expression profiling, TGF- β .

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is a very aggressive malignancy and one of the most lethal types of human cancer [1-3]. PDAC in Germany shows an increased incidence, with 15.630 cases in 2009 to 16.080 in 2010, and up to an estimated 17.400 cases in 2014. The number of deaths in 2010 by this disease is, with 15.487 cases, almost as high as the incidence. The 5-year survival rate of PDAC in Germany is 7% [1]. Tumor resection is today the best therapeutic option and the only therapy to achieve long-term survival. However, most patients were diagnosed in an advanced stage of the disease, namely metastatic (50%) or locally advanced cancer (30%) [4], where standard therapy is limited to conventional chemotherapy, at least in the case of metastatic PDAC. Standard therapy has been and continues to be a systemic chemotherapy with gemcitabine [5]. In the last years, it has been shown that three therapies, namely gemcitabine-erlotinib, FOLFIRINOX, and gemcitabine-nab-paclitaxel, prolonged overall survival (OS) in PDAC patients, when compared with gemcitabine monotherapy [6]. Although the survival benefit for the combination of the EGFR tyrosine-kinase-inhibitor erlotinib and gemcitabine was significant compared to gemcitabine monotherapy, this regime leads only to an

improvement from 5.91 months to 6.24 months in the OS [7]. Up to today, clinical trials with targeted therapies failed to improve OS in PDAC [8]. For example, the humanized monoclonal antibody against vascular endothelial growth factor A (VEGFA), bevacizumab, did not change the overall survival when compared to gemcitabine and erlotinib [9]. Only the FOLFIRINOX regime and Nab-paclitaxel in combination with gemcitabine show an improved response rate (32% and 23%) and a better OS (11.1 and 8.5 month) compared to gemcitabine monotherapy (response rate 9 and 7% and an OS of 6.8 or 6.7 month, respectively), but at a cost of increased toxicity [10-11]. Recently, some promising results came from the use of trabedersen, a transforming growth factor- β (TGF- β) pathway inhibiting substance. Trabedersen is a synthetic 18mer phosphorothioate antisense oligonucleotide that binds specifically to human TGFB2 mRNA and inhibits TGFB2 protein synthesis [12]. In a phase I/II study using this inhibitor, Oettle and colleagues found an OS of 13.4 months in pancreatic cancer patients [13]. The TGF- β pathway is one of the core signaling pathways affected in pancreatic cancer and has a major impact on the pathogenesis of PDAC [14-15]. The importance of TGF- β in pancreatic cancer was also confirmed in an elegant gene expression study by Badea *et al.* in 2008 [16].

The TGF- β superfamily of cytokines consists of more than 30 ligands, namely: TGF- β s, activins,

*Address correspondence to this author at the Charité Medical University Berlin, Institute of Biochemistry, Charitéplatz 1 / Sitz: Virchowweg 6, 10117 Berlin, Germany; Tel: +49 30 450 528087; Fax: +49 30 450 52945; E-mail: andreas.klein@charite.de

inhibins, bone morphogenetic proteins (BMPs), growth and differentiation factors (GDFs), Nodal and anti-Müllerian hormone (AMH). The three TGF- β isoforms, TGF- β 1, TGF- β 2 and TGF- β 3 were synthesized as a precursor. A homodimer of each form interacts with the latency-associated peptide (LAP) and a latent TGF- β -binding protein (LTBP) in order to form a larger complex called the large latent complex (LLC). To be activated, the LLC has to be released from the extracellular matrix, which is then followed by a proteolytic cleavage of LAP to release an active TGF- β isoform to its receptor. TGF- β receptor binding leads to the phosphorylation of receptor type I by the constitutively active type II receptor. Receptor type I in turn phosphorylates and activates the receptor-specific SMADs (R-SMADs). By binding SMAD4, the R-SMADs translocate to the nucleus in order to regulate their target genes [reviewed in 12,17,18].

In non-malignant tissue, TGF- β is a multifunctional regulator of cell communication, proliferation, apoptosis, differentiation, immune modulation, angiogenesis and extracellular matrix production, and has tumor-suppressor properties. Perturbations of the TGF- β -pathway during tumorigenesis, promotes migration, invasion and metastasis through epithelial-mesenchymal transition as well as stimulation of extracellular matrix deposition, angiogenesis and escape from immunosurveillance. It is well known that TGF- β has tumor-suppressive functions in early tumor stages and tumor-promoting functions in later tumor stages. In this study, we have re-analyzed gene expression data published on PDAC. We determined the 100 top-ranked genes that were able to discriminate between PDAC and non-malignant pancreatic tissue. About one third of these genes were associated with the TGF- β pathway, and some of these genes were known to be involved in the cancer switch of TGF- β from a tumor-suppressor to a tumor-promoter.

MATERIALS AND METHODS

We reanalyzed 134 recently published gene expression data sets that were produced using the HG-U133 Plus 2.0 microarray platform. The 94 tumor probes comprise 25 PDAC samples from Donahue *et al.* [19] (GEO [20] Series GSE32688: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32688>), 30 samples from Winter *et al.* [21] (available in the ArrayExpress database [22] under accession number E-MEXP-2780: <http://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-2780/>) and 39 samples from Badea *et al.* [16] (GEO Series GSE1547: [\[ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15471\]\(http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15471\)\). 39 samples from Badea *et al.* \[16\] \(GEO Series GSE15471: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15471>\) derived from normal tissue as well as one sample from Roth \(GSM175950 from GEO Series GSE7307: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7307>\).](http://www.</p></div><div data-bbox=)

Additionally, the 30 samples from the Winter data set [21] and the Badea set [16] comprising 39 tumor samples included survival information that was used for the identification of prognostic genes.

For all 134 samples, background correction and normalization was performed, and gene expression level was summarized on probes using the RMA method [23]. A variance shrinkage method was employed [24] and revealed the top 100 differentially expressed genes based on the *fdr* corrected *p*-values. All gene expressions were log₂-transformed, and the signal intensity changes between tumor and normal samples are expressed as log₂-foldchange (logFC). The *fdr* corrected *p*-values and the log₂-ratio between tumor and normal samples are shown in supplementary file 1.

Both samples and probe sets were unsupervised clustered using euclidean distance and a complete agglomeration function.

Classification was performed using kernlab's implementation of support vector machines [25] using a linear kernel. The models were validated with a performance estimation using a *k*-fold cross validation (*k*=3) approach.

For identification of prognostic genes, we used the above-mentioned, two data sets and classified those samples that survived more than 900 days as having good prognosis, and classified all other samples as having bad prognosis. For the identification of prognostic genes, we performed background correction, normalization and gene expression level summarization on probes using the RMA method [23], followed by a variance shrinkage method [24] that revealed the top differentially expressed genes based on the *p*-values. The samples were unsupervised clustered using euclidean distance and a complete agglomeration function.

All analyses were performed using the software R [26] associated with packages of the Bioconductor project [27].

For the generation of Figure 2, CorrXpression Software was used [28-29]. Only the 121 top-ranked transcripts were used for the correspondence plot.

Gene annotation was performed with the help of the data bases: PubMed [30], Entrez Gene [31], UniProt [32] and iHOP [33].

For an enrichment analysis, a list of TGF- β pathway associated genes published by Ranganathan *et al.* [34] was further processed. The authors characterized 2.041 genes that are regulated by TGF- β (1.757 genes are found exclusively in A549 cells, 733 genes are

exclusively regulated in HPL1D cells, and 267 genes are commonly regulated in both cell-lines). The 2.041 genes in this list were further approved to get only those genes with a known HGNC symbol. We used the symbol checker (http://www.genenames.org/cgi-bin/symbol_checker) and removed all non-approved gene symbols. Furthermore, the synonyms were checked to avoid doubled gene identifiers in the resulting list of TGF- β pathway associated genes.

RESULTS

In this study, we have re-analyzed previously published gene expression data from 94 pancreatic

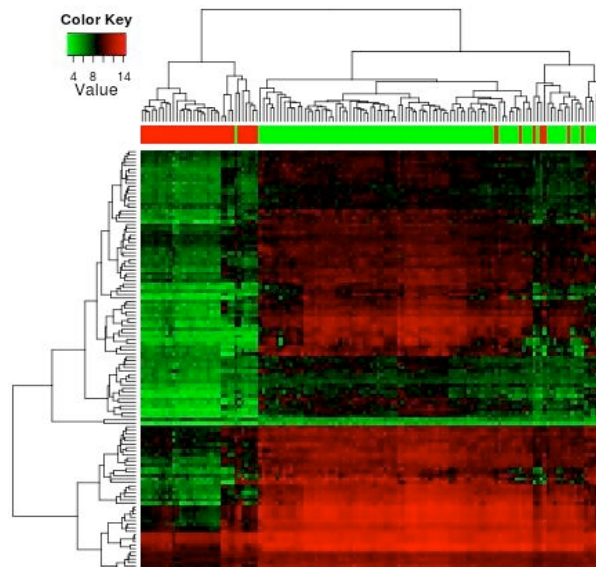


Figure 1: Two-dimensional clustering of all samples (in columns) and the selected 121 probe sets comprising 100 top differentially expressed genes (in rows). The 94 PDAC samples are marked in green in the color bar below the horizontal dendrogram; all 40 normal samples are marked in red. The log₂-transformed gene expression values are indicated by a green to red color gradient.

- Expression profiles of 121 top-ranked transcripts
- 33 non-malignant samples (correctly classified)
- 7 non-malignant samples (falsely classified)
- 93 pancreatic cancer samples (correctly classified)
- 1 pancreatic cancer sample (falsely classified)

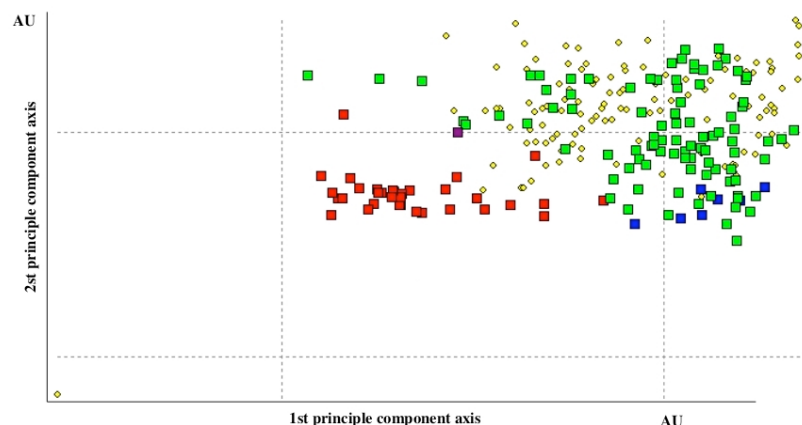


Figure 2: Illustration of the results of a principal component analysis for the top-ranked transcripts with CorrXpression. The transcripts are shown as yellow circles, the samples as colored squares that indicate the correctly- and falsely- classified specimens. AU stands for arbitrary unit.

ductal adenocarcinoma (PDAC) samples and 40 non-malignant pancreatic tissue samples. Data on the PDAC samples were taken from three different studies: Badea *et al.* 2008 (GSE 15471), Donahue *et al.* 2012 (GSE32688) and Winter *et al.* 2012 (E-MEXP-2780). Data on 39 non-malignant samples were taken from Badea *et al.* 2008 (GSE 15471) and one sample was taken from a healthy pancreas from Roth 2007 (GSE7307). We determined the gene expression profile of genes differentially expressed in PDAC compared to non-malignant pancreatic tissue. Using the 100 top-ranked genes, we were able to discriminate between PDAC and non-malignant pancreatic tissue (Figure 1).

Gene Expression Analysis and Classification

Based on the *fdr* corrected *p*-value, the best 100 differentially expressed genes (121 probe sets: see supplementary file 1 for detailed information) were selected. The gene wise depiction of the gene expression values for all 100 selected genes is given in supplementary file 2 and is exemplarily shown for fibronectin 1 (FN1) in Figure 3.

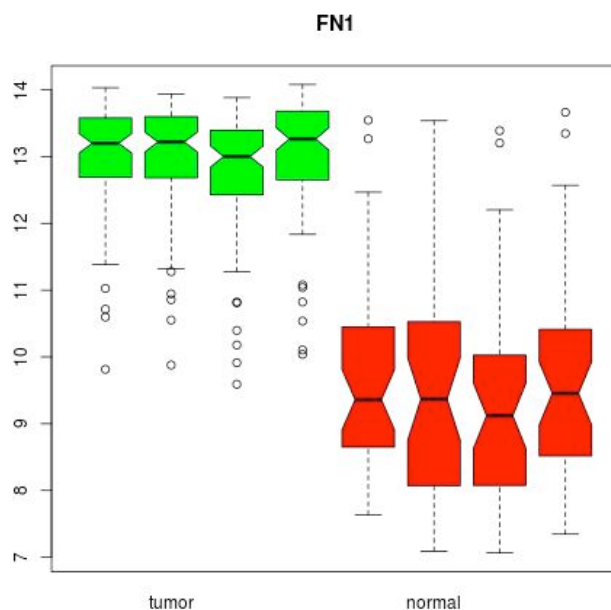


Figure 3: Box-Whisker-Plot for the four gene expression sets 210495_x_at, 211719_x_at, 212464_s_at and 216442_x_at representing fibronectin-1 (FN1) for the tumor samples (left, shown in green) and the normal samples (right, red).

An unsupervised clustering (Figure 1) reveals 6% falsely assigned samples (one tumor and seven normal probes out of 134 samples). 90.3% of all samples were classified correctly as normal or tumor sample using a *k*-fold (*k*=3) cross validation.

In order to verify these results, we performed a principal component analysis (PCA) for the top-ranked transcripts with the help of CorrXpression [28-29]. The PCA was used to examine the relationship between transcripts and experiments. As shown in Figure 2, the non-malignant and the PDAC samples cluster mainly in two distinct groups. The only healthy pancreatic sample lies in-between the non-malignant samples (not shown). The seven falsely classified samples cluster also into the tumor samples. However, in this analysis, the falsely classified tumor sample is located in the PDAC cluster.

The 121 transcripts corresponding to 100 genes were further annotated with the help of the data bases: PubMed, Entrez Gene, UniProt and iHOP. From the 100 top-ranked genes, 82 genes are known to be involved in cancer pathways. For 68 protein-coding genes a tumor-promoting property was described, of which 23 were already known PDAC cancer genes (red symbols in Table 1). For seven genes, tumor-promoting and tumor-suppressor properties were described, and eight further genes have known tumor-suppressor functions (Table 1).

Table 1: The 100 Top-Ranked Genes

68 genes were involved in cancer promotion (genes associated with pancreatic cancer were shown in red): ACTN1, ACVR1, ADAM9, AFAP1, ANO1, ANTXR1, ANXA2, CAPG, CD109, CD47, COL10A1, COL11A1, COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A3, COL8A1, CTHRC1, ECT2, EDNRA, FN1, GJB2, GPRC5A, HOXB3, INHBA, ITGA2, ITGB5, ITGB6, KDM5B, KIF26B, KRT19, KRT7, LAMC2, LBH, LTBP1, MAP4, MAP4K4, MICAL2, MYOF, NOX4, NREP, PLXDC2, PMEPA1, POSTN, PRKCI, PTGER4, PTGFRN, RAB31, RCN1, RUNX1, S100P, SLC39A10, SLPI, SOX4, SPARC, STEAP1, SULF1, TGIF1, THBS2, THY1, TM4SF1, TNFRSF21, TPBG, TRIO, VCAN, WISP1
4 genes were cancer related: ARHGAP1, HOXA3, MXRA5, ZNF532
8 tumor-suppressor genes: ADAMTS12, ASAP2, DKK3, PHLDA2, PIAS3, RASAL2, STK3, VGLL4
2 tumor-suppressor candidates: MORC2, TTL5
7 genes with tumor-promoting and tumor-suppressing properties: ACVR1, COL8A1, RUNX1, SULF1, SLPI, TGIF1, THBS2
For 18 genes no involvement in tumorigenesis was found: AHNK2, ANXA2P2, ATP2C1, CAMSAP2, CEP170, CKLF, CYB5R3, DCBLD1, IDS, ITPRIPL2, KIAA1217, NHS, NTM, PGM2L1, SESTD1, SLC6A6, TMEM87A, LOC283508

The most affected signaling pathways were signal transduction (33 genes), cell adhesion (28 genes), extracellular matrix organization (20 genes), cell migration (20 genes; 5 genes were associated with leucocyte migration), cell apoptosis (18 genes; 8

inhibitors of apoptosis, 9 inducers of apoptosis, and 1 gene where both functions have been described), cell proliferation (17 genes), regulation of transcription (15 genes), and angiogenesis (14 genes). Furthermore, 15 genes were stem cell-associated genes, and 11 genes were known to code for proteins involved in epithelial-mesenchymal transition (EMT). For more detailed

information, see supplementary files 1 and 3.

TGF- β Pathway Associated Genes

Interestingly, 31 out of the selected genes were identified as associated with the TGF- β pathway. These genes were described in Table 2, including their association to tumor progression found in literature.

Table 2: Genes Associated with the TGF-Beta Pathway

Affymetrix ID	Gene-Symbol	TGF- β Signaling Involvement	Relevance in Cancer Promotion	Relevance in PDAC	Tumor-Suppressor Function
208636_at	ACTN1	TGF- β responsive gene (up)	+		
203935_at	ACVR1	TGF- β responsive gene (up), belongs to TGF- β superfamily	+		
202381_at	ADAM9	TGF- β responsive gene (down/up: depending on cell type)	+		
226545_at	CD109	TGF- β negative regulation (TGF- β co-receptor, CD109, promotes internalization and degradation of TGF- β receptors)	+		
205941_s_at, 217428_s_at	COL10A1	TGF- β responsive gene (up)	+	+	
204320_at, 37892_at	COL11A1	TGF- β responsive gene (up)	+	+	
202311_s_at	COL1A1	TGF- β responsive gene (up)	+		
202403_s_at, 202404_s_at	COL1A2	TGF- β responsive gene (up)	+		
211161_s_at, 215076_s_at	COL3A1	TGF- β responsive gene (up)	+		
203325_s_at, 212488_at, 212489_at	COL5A1	TGF- β responsive gene (up)	+		
221729_at, 221730_at	COL5A2	TGF- β responsive gene (up)	+		
201438_at	COL6A3	TGF- β responsive gene (up)	+		
226237_at	COL8A1	TGF- β modulating	+		+
225681_at	CTHRC1	TGF- β responsive gene (up) and inhibition of TGF- β signaling	+	+	
214247_s_at	DKK3	TGF- β negative-regulation (limiting TGF- β /Smad signaling)			+
210495_x_at, 211719_x_at, 212464_s_at, 216442_x_at	FN1	TGF- β responsive gene (up) and enhancer of endogenous TGF- β effects	+	+	
210511_s_at, 227140_at	INHBA	Member of the TGF- β superfamily	+	+	
201125_s_at, 214020_x_at	ITGB5	TGF- β responsive gene (up), activation of TGF- β	+		
226535_at	ITGB6	TGF- β responsive gene (up), activation of TGF- β	+	+	
202267_at	LAMC2	TGF- β responsive gene (up)	+	+	
202728_s_at, 202729_s_at	LTBP1	TGF- β responsive gene (up) and activation of TGF- β	+		

Table 2 continue ...

Affymetrix ID	Gene-Symbol	TGF- β Signaling Involvement	Relevance in Cancer Promotion	Relevance in PDAC	Tumor-Suppressor Function
219773_at	NOX4	TGF- β responsive gene (up) and mediates TGF- β -induced profibrotic response	+	+	
201310_s_at	NREP	Cooperating effect with TGF- β and positively regulates TGF- β translation	+		
217875_s_at, 222449_at	PMEPA1	TGF- β responsive gene (up) and molecular switch that converts TGF- β from a tumor-suppressor to a tumor-promoter,	+		
209360_s_at	RUNX1	Interaction with SMAD proteins	+		+
201416_at, 201417_at	SOX4	TGF- β responsive gene (up)	+		
200665_s_at	SPARC	TGF- β responsive gene (up) and activator of TGF- β 1	+	+	
212353_at, 212354_at	SULF1	TGF- β responsive gene (up) and negative regulator of TGF- β 1	+	+	+
203313_s_at	TGIF1	TGF- β responsive gene (up) and Co-repressor of TGF- β signaling	+		+
203083_at	THBS2	TGF- β responsive gene (up), THBS2 inhibits activation of TGF- β by THBS1	+		+
204619_s_at, 221731_x_at	VCAN	TGF- β responsive gene (up)	+	+	

To determine whether the mentioned 31 TGF- β genes are randomly selected or significantly enriched in the list of differentially expressed genes, we used a list of approved genes associated with the TGF- β pathway. Ranganathan *et al.* published such a list with 2,041 genes [34]. We checked these genes for approved HGNC symbols and for duplications due to synonymous naming using a symbol checker (http://www.genenames.org/cgi-bin/symbol_checker). At the end of this process, we obtained a list of 1,655 genes (see supplementary file 1). Comparing 1,655 TGF- β associated genes out of 20,805 human genes (Ensembl [35] version 21.2.14) to 31 TGF- β genes within our list of 100 revealed a clear enrichment (3.9-fold) of TGF- β genes associated with pancreatic tumor probes.

Prognostic Genes / Survival Analysis

For the prediction of prognostic genes, 69 samples with survival data were taken. We selected two distinct groups: bad prognosis with survival time < 900 days, and good prognosis with survival time > 900 days. The threshold of 900 days was used to obtain two groups of nearly equal size. We performed a differential expression analysis and selected eleven genes that were able to discriminate between patients with good and bad prognoses. The number of genes in the prognosis list was estimated in such a way as to

minimize the false-discovery rate. The genes are summarized in Table 3. The unsupervised clustering of the samples based on the 11 selected genes is shown in Figure 4. It can be seen that 64 samples (out of 69) were classified correctly (92.8%).

Table 3: Genes Identified by Survival Analysis

	GeneSymbol	logFC	P.Value
226301_at	SLC18B1	0.5470565	3.799938e-05
225368_at	HIPK2	0.7001166	3.822944e-05
202913_at	ARHGEF11	0.3462814	5.880808e-05
243403_x_at	CPM	0.4778296	5.984464e-05
234552_at	EPPK1	0.4601844	1.401751e-04
222817_at	HSD3B7	0.4451037	1.761632e-04
210975_x_at	FASTK	0.2395951	1.823211e-04
202519_at	MLXIP	0.4185767	2.016941e-04
232181_at	PPARGC1B	0.8450016	2.518462e-04
232336_at	ZSWIM5	0.6791138	2.553374e-04
219771_at	TBC1D8B	0.4204538	2.731243e-04

DISCUSSION

A genetic progression model for pancreatic cancer was recently established [36]. In their model, metastatic PDAC arises from PanIN (pancreatic intraepithelial neoplasia) through multiple genetic and epigenetic

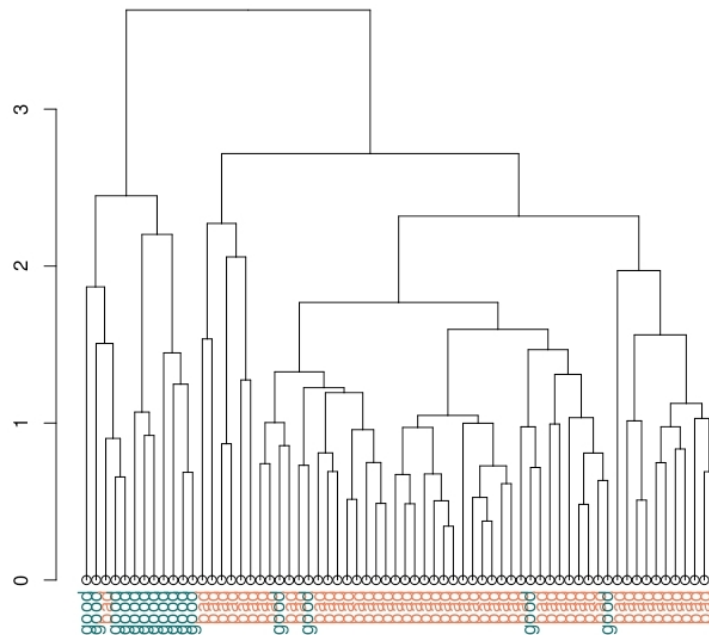


Figure 4: Unsupervised hierarchical clustering of the 69 samples with survival information. Samples associated with good prognosis (>900 days' survival) are marked in green and the label 'good'; the others are marked as 'bad' (in red).

events. Early events in tumorigenesis were found to be activating mutations in KRAS in PanIN-1, mutations or epigenetic silencing of CDKN2A in PanIN-2, and inactivating mutations of TP53 and of SMAD4 in PanIN-3. Approximately 55% of pancreatic cancers bear deletions or mutations in SMAD4 [37-38]. Beside mutations in SMAD4, genetic alterations in genes relevant to the TGF- β pathway occur mainly in TGFBR2, BMPR2 and SMAD3. In a study by Jones and colleagues published in 2008, it was shown that 100% of the pancreatic cancers had genetic alterations in the TGF- β pathway. Therefore, the TGF- β pathway belongs to the core signaling pathways in PDAC [14]. Badea *et al.* revealed, in a gene expression study that tumor tissues contain a TGF-beta signature as a result, mainly due to the high amounts of stromal cells in tumor samples [16].

In this work, we have re-analyzed gene expression data from earlier [16] and from more recently published gene expression studies [19,21]. We were able to determine the 100 top-ranked genes which discriminate best between PDAC and non-malignant pancreatic tissue. Interestingly, about one third of these genes were associated with the TGF- β pathway.

Pancreas cancer progression is a cooperate action between tumor cells and stromal cells, namely pancreatic stellate cells (PSC) [39]. The pancreatic tumor cells induce via TGF- β , amongst other growth

factors, the PSC to build up the extracellular matrix, which itself promotes tumor cell migration, invasion and survival. Thus, we found an overexpression of protein-coding genes like: COL11A1, INHBA and THBS2 that were produced from PSC. These genes were associated with an invasion-associated variant of a desmoplastic reaction and belong to a metastasis-associated gene expression signature [40]. Furthermore, stromal cells of the tumoral pancreas express the reactive oxygen species (ROS)-producing NOX4. It was shown by Barcellos-Hoff and Dix that ROS formation triggers conversion of the latent form of TGF- β to its active form, which is one of the most important mechanisms that regulate activation of the TGF- β signaling [41]. However, TGF- β activation can also be effected by integrins $\alpha\beta 5$ and $\alpha\beta 6$ [42]. Integrins $\beta 5$ and $\beta 6$ were expressed in pancreatic cancer cell lines (GSE36139 [43]); for integrin $\beta 6$, a strong protein expression in the pancreatic cell lines CAPAN2 and HPAC were shown [44]. These integrins were not normally expressed in epithelial cells and enhance the migratory and invasive behavior of carcinoma cells [42]. Additionally, it was shown that TMEPAI functions in breast cancer cells as a molecular switch that converts TGF-beta from a tumor-suppressor to a tumor-promoter [45]. TMEPAI is also highly expressed in the pancreatic cancer cell lines Panc 04.03, Panc 05.04 and SU.86.86 [43]. In return, TMEPAI was also found in a stromal signature of esophageal adenocarcinomas and patients overexp-

ressing this gene had a significantly poorer outcome [46]. Thus, the activation of the TGF- β pathway is also a cooperative effect of PSC and tumor cells.

Since the TGF- β pathway is one of the core pathways in PDAC leading to desmoplasia and to highly invasive and metastatic tumor cells, inhibiting this pathway might be a promising therapeutic option. Through this inhibition the extracellular matrix might be reduced, which would lead to a better accessibility for chemotherapeutic drugs.

The earliest data to support the hypothesis of the current study stems from a phase I/II study (NCT00844064) using trabectedin – a complementary sequence to the TGFB2 mRNA. Survival analysis revealed a median overall survival of 13.4 months [13]. A further clinical phase II study applying LY2157299, an orally active transforming growth factor beta receptor kinase inhibitor with or without gemcitabine, is still ongoing (NCT01373164).

The search for a prognostic gene set in pancreatic cancer is still challenging since it is difficult to discriminate pancreatic cancers with a good or bad prognosis. Pancreatic cancer surgery is still a complex operation with morbidity rates of 30% – 60% [47], and most pancreatic cancers were diagnosed in an advanced stage of disease. Further, the median age of onset of PDAC is 71 years [48], so that many patients suffer from co-morbidity. Another drawback is a lack of freely available clinical data for such a difficult task. In our analysis, we used the survival time of 900 days to select two distinct groups with good and bad prognosis. We performed a differential expression analysis and selected eleven genes that were able to discriminate between patients with a good and bad prognosis. The clustering reveals only a 7.2% misclassification rate. However, these results have to be approved in a larger clinical data set to further evaluate the gene set.

ACKNOWLEDGMENTS

We thank Roland Bell for editing the English version of this manuscript.

SUPPLEMENTARY MATERIAL

Supplementary file 1: Complete list of the 100 top-ranked genes (sheet 1) and complete list of the 1.655 TGF- β associated genes based on the publication of Ranganathan *et al.* 2007 (sheet 2). Sheet 1: Beside Affymetrix ID, adj.P.Val, log-ratio, GeneSymbol and Gene Title, the GO biological process term, relevance

in cancer and in the TGF- β pathway is shown. References are given supporting relevance for cancer and the TGF- β pathway.

Supplementary file 2: Box-Whisker-Plots for all 100 top-ranked gene expression sets. The tumor samples are shown on the left side in green and the normal samples are shown on the right side in red.

Supplementary file 3: Annotation of the 100 top-ranked genes based on the data bases PubMed, Entrez Gene, UniProt and iHOP.

REFERENCES

- [1] Robert Koch Institute German Centre for Cancer Registry Data. Krebs in Deutschland 2009/2010; http://www.krebsdaten.de/Krebs/EN/Content/Cancer_sites/Pancreatic_cancer/pancreatic_cancer_node.html.
- [2] Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA: a cancer journal for clinicians 2014; 64: 9-29. <http://dx.doi.org/10.3322/caac.21208>
- [3] Ma J, Siegel R, Jemal A. Pancreatic cancer death rates by race among US men and women, 1970-2009. Journal of the National Cancer Institute 2013; 105(22): 1694-1700. <http://dx.doi.org/10.1093/jnci/djt292>
- [4] Werner J, Combs SE, Springfield C, Hartwig W, Hackert T, Büchler MW. Advanced-stage pancreatic cancer: therapy options. Nature reviews Clinical oncology 2013; 10(6): 323-333. <http://dx.doi.org/10.1038/nrclinonc.2013.66>
- [5] Burris HA, Moore MJ, Andersen J, Green MR, Rothenberg ML, Modiano MR, *et al.* Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 1997; 15(6): 2403-2413.
- [6] Chan SL, Chan ST, Chan EH, He ZX. Systemic treatment for inoperable pancreatic adenocarcinoma: review and update. Chinese Journal of Cancer 2014.
- [7] Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, *et al.* Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. 2007. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 2007; 25(15): 1960-1966. <http://dx.doi.org/10.1200/JCO.2006.07.9525>
- [8] Preis M, Korc M. Kinase signaling pathways as targets for intervention in pancreatic cancer. Cancer Biology & Therapy 2010; 9(10): 754-763. <http://dx.doi.org/10.4161/cbt.9.10.11534>
- [9] Van Cutsem E, Vervenne WL, Bennouna J, Humblet Y, Gill S, Van Laethem JL, *et al.* Phase III trial of bevacizumab in combination with gemcitabine and erlotinib in patients with metastatic pancreatic cancer. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology 2009; 27(13): 2231-2237. <http://dx.doi.org/10.1200/JCO.2008.20.0238>
- [10] Conroy T, Desseigne F, Ychou M, Bouché O, Guimbaud R, Bécouarn Y, *et al.* FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. The New England Journal of Medicine 2011; 364(19): 1817-1825. <http://dx.doi.org/10.1056/NEJMoa1011923>
- [11] Von Hoff DD, Ervin T, Arena FP, Chiorean EG, Infante J, Moore M, *et al.* Increased survival in pancreatic cancer with

- nab-paclitaxel plus gemcitabine. *The New England Journal of Medicine* 2013; 369(18): 1691-1703.
<http://dx.doi.org/10.1056/NEJMoa1304369>
- [12] Akhurst R, Hata A. Targeting the TGF signalling pathway in disease. *Nature Reviews Drug Discovery* 2012; 11(10): 790-811.
<http://dx.doi.org/10.1038/nrd3810>
- [13] Oettle H, Seufferlein T, Luger T, Schmid RM, von Wichert G, Endlicher E, *et al.* Core signaling pathways in human pancreatic cancers, malignant melanoma, and colorectal carcinoma with trabedersen. *J Clin Oncol* 2012; 30(suppl; abstr 4034).
- [14] Jones S, Zhang X, Parsons DW, Lin JCH, Leary RJ, Angenendt P, *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008; 321(5897): 1801-1806.
<http://dx.doi.org/10.1126/science.1164368>
- [15] Truty MJ, Urrutia R. Basics of TGF-beta and pancreatic cancer. *Pancreatology* 2007; 7(5-6): 423-435.
<http://dx.doi.org/10.1159/000108959>
- [16] Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepato-gastroenterology* 2008; 55(88): 2016-2027.
- [17] Wakefield LM, Hill CS. Beyond TGFβ: roles of other TGFβ superfamily members in cancer. *Nat Rev Cancer* 2013; 13(5): 328-41.
<http://dx.doi.org/10.1038/nrc3500>
- [18] Leask A, Abraham DJ. TGF-beta signaling and the fibrotic response. *FASEB J* 2004; 18(7): 816-27.
<http://dx.doi.org/10.1096/fj.03-1273rev>
- [19] Donahue TR, Tran LM, Hill R, Li Y, Kovochich A, Calvopina JH, *et al.* Integrative survival-based molecular profiling of human pancreatic cancer. *Clinical Cancer Research* 2012; 18(5): 1352-1363.
<http://dx.doi.org/10.1158/1078-0432.CCR-11-1539>
- [20] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* 2002; 30(1): 207-210.
<http://dx.doi.org/10.1093/nar/30.1.207>
- [21] Winter C, Kristiansen G, Kersting S, Roy J, Aust D, Knösel T, *et al.* Google goes cancer: Improving outcome prediction for cancer patients by network-based ranking of marker genes. *PLoS Computational Biology* 2012; 8(5).
<http://dx.doi.org/10.1371/journal.pcbi.1002511>
- [22] Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, *et al.* ArrayExpress update-trends in database growth and links to data analysis tools. *Nucleic acids research* 2013; 41(Database issue): D987-D990.
<http://dx.doi.org/10.1093/nar/gks1174>
- [23] Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)* 2003; 4(2): 249-264.
<http://dx.doi.org/10.1093/biostatistics/4.2.249>
- [24] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* 2004; 3: Article3.
- [25] Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software* 2004; 11(9): 1-20.
- [26] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013 <http://www.R-project.org/>.
- [27] Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* 2004; 5(10): R80.
<http://dx.doi.org/10.1186/gb-2004-5-10-r80>
- [28] Klein A, Olendrowitz C, Schmutzler R, Hampl J, Schlag PM, Maass N, *et al.* Identification of brain- and bone-specific breast cancer metastasis genes. *Cancer letters* 2009; 276(2): 212-220.
<http://dx.doi.org/10.1016/j.canlet.2008.11.017>
- [29] Wessel R, Foos V, Aspelmeier A, Jürgens M, Graessmann A, Klein A. CorrXpression - identification of significant groups of genes and experiments by means of correspondence analysis and ratio analysis. *In Silico Biology* 2006; 6(1): 61-70.
- [30] NCBI Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Research* 2013; 41: D8-D20.
- [31] Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez gene: gene-centered information at NCBI. *Nucleic Acids Research* 2005; 33: D54-D58.
- [32] Bairoch A, Bougueleret L, Altairac S, Amendolia V, Auchincloss A, Argoud-Puy G, *et al.* The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Research* 2009; 37(Database issue): D169-D174.
- [33] Fernández JM, Hoffmann R, Valencia A. iHOP web services. *Nucleic Acids Research* 2007; 35(Web Server issue): W21-W26.
- [34] Ranganathan P, Agrawal A, Bhushan R, Chavalmane AK, Kalathur RKR, Takahashi T, *et al.* Expression profiling of genes regulated by TGFbeta: differential regulation in normal and tumour cells. *BMC genomics* 2007; 8: 98.
<http://dx.doi.org/10.1186/1471-2164-8-98>
- [35] Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, *et al.* Ensembl 2014. *Nucleic Acids Research* 2014; 42(1): D749-D755.
<http://dx.doi.org/10.1093/nar/gkt1196>
- [36] Yachida S, Iacobuzio-Donahue CA. Evolution and dynamics of pancreatic cancer progression. *Oncogene* 2013; 32(45): 5253-5260.
<http://dx.doi.org/10.1038/onc.2013.29>
- [37] Hahn SA, Schutte M, Hoque ATMS, Moskaluk C, da Costa LT, Rozenblum E *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* 1996; 271: 350-353.
<http://dx.doi.org/10.1126/science.271.5247.350>
- [38] Liu F. SMAD4/DPC4 and pancreatic cancer survival: Commentary to: M. Mascilar *et al.*, the SMAD4 protein and prognosis of pancreatic ductal adenocarcinoma. *Clin. Cancer Res.*, 7: 4115-4121, 2001. *Clinical Cancer Research* 2001; 7(12): 3853-3856.
- [39] Lunardi S, Muschel RJ, Brunner TB. The stromal compartments in pancreatic cancer: Are there any therapeutic targets? *Cancer Letters* 2014; 343(2): 147-155.
<http://dx.doi.org/10.1016/j.canlet.2013.09.039>
- [40] Kim H, Watkinson J, Varadan V, Anastassiou D. Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. *BMC Medical Genomics* 2010; 3(1): 51.
<http://dx.doi.org/10.1186/1755-8794-3-51>
- [41] Barcellos-Hoff MH, Dix TA. Redox-mediated activation of latent transforming growth factor-beta 1. *Mol Endocrinol* 1996; 10(9): 1077-1083.
- [42] Margadant C, Sonnenberg A. Integrin-TGF-beta crosstalk in fibrosis, cancer and wound healing. *EMBO Rep* 2010; 11(2): 97-105.
<http://dx.doi.org/10.1038/embor.2009.276>
- [43] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug

- sensitivity. *Nature* 2012; 483(7391): 603-607.
<http://dx.doi.org/10.1038/nature11003>
- [44] Lee CN, Heidbrink JL, McKinnon K, Bushman V, Olsen H, FitzHugh W, *et al.* RNA Interference Characterization of Proteins Discovered by Proteomic Analysis of Pancreatic Cancer Reveals Function in Cell Growth and Survival. *Pancreas* 2012; 41(1): 84-94.
<http://dx.doi.org/10.1097/MPA.0b013e3182236385>
- [45] Singha PK, Yeh IT, Venkatachalam MA, Saikumar P. Transforming growth factor- β (TGF- β)-inducible gene TMEPAI converts TGF- β from a tumor suppressor to a tumor promoter in breast cancer. *Cancer Research* 2010; 70(15): 6377-6383.
<http://dx.doi.org/10.1158/0008-5472.CAN-10-1180>
- [46] Saadi A, Shannon NB, Lao-Sirieix P, O'Donovan M, Walker E, Clemons NJ *et al.* Stromal genes discriminate preinvasive from invasive disease, predict outcome, and highlight inflammatory pathways in digestive cancers. *Proceedings of the National Academy of Sciences* 2010; 107(5): 2177-2182.
<http://dx.doi.org/10.1073/pnas.0909797107>
- [47] Ho C, Kleeff J, Friess H, Büchler MW. Complications of pancreatic surgery. *HPB (Oxford)* 2005; 7(2): 99-108.
<http://dx.doi.org/10.1080/13651820510028936>
- [48] Yadav D, Lowenfels AB. The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology* 2013; 144(6): 1252-1261.
<http://dx.doi.org/10.1053/j.gastro.2013.01.068>

Received on 28-02-2014

Accepted on 06-03-2014

Published on 25-03-2014

DOI: <http://dx.doi.org/10.14205/2311-0309.2014.02.01.3>

© 2014 Klein and Pospisil; Licensee Pharma Publisher.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.