Supplementary material

Combined Gene Expression Analysis of Whole-Tissue and Microdissected Pancreatic Ductal Adenocarcinoma identifies Genes Specifically Overexpressed in Tumor Epithelia

Liviu Badea, Vlad Herlea, Simona Olimpia Dima, Traian Dumitraşcu, Irinel Popescu

Materials and Methods

Pancreatic tissues. Pairs of normal and tumor tissue samples were obtained at the time of surgery from resected pancreas of 36 pancreatic cancer patients. The tissue samples were snap-frozen in liquid nitrogen immediately upon surgical removal and maintained at −85°C. For histological evaluation, sections were stained with haematoxylin and eosin (H&E). The diagnosis, histological typing, grading and staging were performed by an experienced pathologist (V.H.) from the Pathology Department of the Fundeni Clinical Institute. The resected cancer specimens were not microdissected. Pairs of normal pancreatic tissue and primary adenocarcinoma were also paraffin-embedded. Written informed consent was obtained from all patients. Clinical and pathological data were obtained from the patients' medical records and from pathology investigations (see Supplementary Information). Approval for the study has been obtained from the Ethics Committee of the Fundeni Clinical Institute.

RNA preparation. Total RNA was extracted from the frozen tissue samples in two steps: RNA isolation using the TriReagent (Sigma) protocol, followed by RNA clean-up performed as described in the RNeasy (Qiagen) protocol. Briefly, tissues were homogenized in TriReagent (1 mL of TriReagent per 50-100mg of tissue). Then, RNA was precipitated from the aqueous phase with addition of isopropanol at half the volume of the original TriReagent volume and centrifugated at 12000 × g for 10 min. RNA was cleaned using an RNeasy column according to the manufacturer's instructions (RNeasy Mini Kit, Qiagen) and eluted in 100 µL of RNase-free water. RNA sample quantitation was performed by means of spectrophotometry (Genequant, Pharmacia), while RNA quality (integrity) control was done with microfluidics technology (2100

Bioanalyser, Agilent Technologies). All the RNA extractions used in the study responded to the quality criteria defined by Agilent (Agilent Technologies, Palo Alto, CA).

GeneChip hybridization. Gene expression was analyzed on Affymetrix U133 plus 2.0 high-density oligonucleotide microarrays (Affymetrix, Santa Clara, CA), which provide simultaneous measurements of the expression levels of virtually all human genes (54675 probesets). Target preparation and microarray processing procedures were performed as described in the Affymetrix GeneChip Expression Analysis Manual (Affymetrix, Santa Clara, CA). Briefly, 3 µg of total RNA was used to synthesize double-stranded cDNA with SuperScript II reverse transcriptase (Life Technologies, Inc. Rockville, MD) and a T7-$(dT)_{24}$ primer (Proligo, France). Then, biotinylated cRNA was synthesized from the double-stranded cDNA using the IVT Labeling kit and was purified and fragmented. The fragmented cRNA was hybridized to the oligonucleotide microarray, which was washed and stained with streptavidin-phycoerythrin. Scanning was performed with an Affymetrix Microarray Scanner. For three of the 36 normal-tumor sample pairs we performed replicate microarray hybridizations in order to gauge the technical measurement errors. We thus performed 78 genechip hybridizations in total. After publication, the data will be deposited in the Gene Expression Omnibus (GEO).

Scanning data analysis. Affymetrix GeneChip Operating Software (GCOS) Version 1.4 (Affymetrix Inc., Santa Clara, CA) was used for low-level scanning data processing. The scanned image ('.dat') files and the report ('.rpt') files corresponding to the individual gene chips were analyzed as described in the Affymetrix GeneChip Manual (we checked the uniformity of the signal, the positioning of the grid, the scale factor, background, percent of present probesets, the housekeeping and spike controls). A patient sample pair was excluded from further analysis since one of the samples did not meet the quality controls. The microarray data was subsequently normalized using the Robust Microarray Analysis (RMA) algorithm [33] using the Bioconductor 1.9 package (http://www.bioconductor.org/) running under R version 1.9.1. All further processing was performed in Matlab version 7 (www.mathworks.com).

Statistical analysis. Since the expression data is approximately log-normally distributed, we used the log-transformed data as produced by the RMA algorithm for all subsequent statistical tests. For visualization purposes, we centred the log-transformed expression data by subtracting the average probeset log-expression values. Probesets with relatively low expression (average expression values below 100 Affymetrix units) or with nearly constant expression values (standard deviation below 50) were excluded from further consideration. Of the 54675 probesets on the U133 Plus 2.0 chip, 12209 were thus retained. An unpaired t-test was used to determine the probesets (genes) that are differentially expressed between the normal and the tumor tissue samples. The first 400 probesets with the lowest t-test p-values (corresponding to a p-value cutoff of $9 \cdot 10^{-12}$) were retained for further analysis. We also used a more stringent fold-change constraint that excluded the probesets with log-fold change < 1 (roughly corresponding to a fold change < 2), where the log-fold change of gene g between classes N (‘normal’) and T (‘tumor’) is defined as $\textit{log-fc}(g, N-T) = \overline{\log_2 g(T)} - \overline{\log_2 g(N)}$, with $\overline{\log_2 g(C)} = \frac{\sum_{S_i \in C} \log_2 g(S_i)}{|C|}$ the average log-expression value of gene g in the samples S_i of class C. We also performed a univariate Cox proportional hazards analysis of the correlation of the individual gene expression profiles with the patient post-operative survival. A preliminary functional analysis of the set of differentially expressed genes was performed with the L2L online tool [34], as well as with the TRED Transcription Regulatory Element Database [35]. The gene lists obtained were compared with gene lists from published PDAC studies [1-25]. Lists [1-11] were retrieved from the Pancreatic Expression Database (www.pancreasexpression.org), while lists [12-25] were obtained from the meta-analysis of Brandt et al. [68].

RT-PCR validation. A number of 20 genes were selected for RT-PCR validation: SULF1, WISP1, FN1, HNT, CSPG2, NOX4, SOX4, BGN, BHLHB2, INHBA, PLAU, IGFBP5, PLAT, DKK3, PDGFC, HOP, CAPG, ETV1, LTBP1, CTHRC1. RLP13A and 18S RNA were used as controls. (RLP13A was selected by searching for the gene with the smallest std/mean ratio across all chips in the microarray study. Since 18S was observed to be much more variable in the RT-PCR data, we report measurements normalized with respect to RLP13A.) Total RNA from the tissues samples under study was isolated and transcribed into cDNA as described above. Quantitative RT-PCR was performed using an Applied Biosystems Sequence Detection System 7300 according to the manufacturer’s recommendations and the data was acquired using SDS Software 1.4.

Reprocessing of the Grützmann and Pilarsky studies. For separating the epithelial contribution from the dominant stromal signal in our dataset, we also used a PDAC microdissection dataset described in Grützmann et al (2004) [36] and Pilarsky et al. (2008) [37], which contains Affymetrix U133 A and B measurements for 17 normal epithelia (NE) samples, 20 tumor epithelia (TE), 9 chronic pancreatitis stroma (CPS) and 11 tumor stroma (TS) samples, as well as pancreatic stellate cell and PDAC cell lines. In this paper, we use the following abbreviations: *GP* for the Grützmann and Pilarsky study data [36,37] and *ICF* for our data (Clinical Institute Fundeni). We downloaded the raw expression data (CEL files) from ArrayExpress (http://www.ebi.ac.uk/microarray-as/aer/, E-MEXP-950 and E-MEXP-1121) and reprocessed the raw data using the RMA normalization algorithm (the dChip algorithm with the PM/MM setting as used in the original papers [36,37] occasionally produces negative expression values). The cell line measurements were excluded from further processing since the cell lines have been observed to have significantly different gene expression values than the surgical tissue samples, thereby potentially affecting RMA quantile normalization. Supplementary Figure 1 shows that the relative standard deviations of the Affymetrix control probesets are about two times larger in the microdissection study than in our study. The experimental noise is therefore significantly higher in the microdissection study than in our whole tissue study.

We also observed a pronounced 3' bias for the housekeeping genes (e.g. GAPDH), which is probably due to the several rounds of RNA amplification used in the GP study. However, since the majority of probesets on the U133A/B chips represent the 3' ends of transcripts, this bias may not have a major impact on the quantification of most transcripts.

We excluded samples TPK9, NPK13 and NPD15 from the dataset based on their very large scaling factors (over 39), small percentage of present probesets (< 16%) as well as very frequent outliers for the Affymetrix control probesets. Due to the smaller sample sizes as well as to the significantly higher experimental noise in the microdissection study, we used a more lenient p-value cutoff of 0.01 for all differential expression t-tests on the GP data.

Joint statistical analysis of the ICF and GP datasets. Since the normal and the tumor whole tissue samples from the ICF study contain very different cell compositions, the genes that we find overexpressed in whole tissue tumors may not necessarily be due to the tumor epithelia, but could reflect the abundant stromal proliferation in tumors due to the desmoplastic reaction. In the following, we use the following abbreviations for the various cell types: *NE* – normal epithelia, *TE* – tumor epithelia, *CPS* – chronic pancreatitis stroma, *TS* – tumor stroma. Since both the whole tissue (ICF) and the microdissection study (GP) have their advantages and drawbacks, we combined the two studies. The whole tissue study has a larger number of samples and smaller measurement errors, which enables a more reliable determination of the genes that are differentially expressed in PDAC whole tissue tumors. Unfortunately, many of these genes may be due to the stromal over-proliferation in PDAC tumors rather than the epithelial tumor cells proper. Although the microdissection study is less reliable due to its higher experimental noise and smaller number of samples, it provides crucial information about the localization of the over-expressed genes in the neoplastic epithelia, the tumor stroma, or both.

Although the two datasets (ICF and GP) have been obtained with different microarray chips (U133 Plus 2.0 and U133A/B respectively), the probesets on the U133A and B chips are also present on the U133 Plus 2.0 chip, which establishes a straight-forward correspondence between the two datasets at the probeset level and thereby avoids certain delicate issues related to inter-platform reproducibility and compatibility.

To determine the set of genes overexpressed in ICF that are specifically upregulated in the tumor epithelia (rather than just the tumor stroma), we constructed three subsets of genes as follows (here we used a more lenient p-value cutoff of 10^{-9} for ICF, since the genes overexpressed specifically in *TE* are masked by the *TS*):

TE_fc_epithelial: $p_{\mathrm{ICF}}(N-T)<10^{-9}$ and $log\text{-}fc_{\mathrm{ICF}}(N-T) \geq 1$ and $log\text{-}fc_{\mathrm{GP}}(NE-TE) \geq 1$ and $TE >_{\mathrm{GP}} TS$.

TE_p_epithelial: $p_{\mathrm{ICF}}(N-T)<10^{-9}$ and $log\text{-}fc_{\mathrm{ICF}}(N-T) \geq 1$ and $p_{\mathrm{GP}}(NE-TE) <10^{-2}$ and $p_{\mathrm{GP}}(TE-TS) <10^{-2}$ and $TE >_{\mathrm{GP}} TS$ and $0 < log\text{-}fc_{\mathrm{GP}}(NE-TE) < 1$.

TE_fc_stromal: $p_{\mathrm{ICF}}(N-T)<10^{-9}$ and $log\text{-}fc_{\mathrm{ICF}}(N-T) \geq 1$ and $log\text{-}fc_{\mathrm{GP}}(NE-TE) \geq 1$ and $TE <_{\mathrm{GP}} TS$.

TE_fc_epithelial contains the genes that are (more or less) highly over-expressed (at least two-fold) in *TE* vs. *NE* and which are also over-expressed in *TE* vs. *TS* (otherwise, their expression in *TE* would be masked by the *TS* in the ICF tumor samples).

TE_p_epithelial collects the genes that are only mildly (less than 2-fold) but quite consistently (with a small p-value) over-expressed in *TE* vs. *NE*.

TE_fc_stromal also contains genes highly over-expressed in *TE* vs. *NE*, but their expression is even higher in *TS*, so their increase observed in the whole tissue tumor samples (in the ICF data) is predominantly due to their stromal component.

Comparison with the study of Logsdon et al.

Since microarray analyses of microdissected tissue are inherently yielding more variable results than whole tissue studies due to the low amounts of sample tissue, Logsdon et al. (2003) [17] approached the problem of separating the epithelial from the stromal contribution to gene expression profiles by selecting genes that are overexpressed in whole tumors w.r.t. both normal tissue and chronic pancreatitis (CP). Although that study is one of the most accurate ones, it has certain problems related to an assumption it is forced to make in order to separate the epithelial from the stromal contribution with the available data. More precisely, Logsdon et al. implicitly assume that the stromal gene expression profiles of PDAC and CP are very similar, which implies that the genes that are overexpressed in whole tumor tissue w.r.t. both normal and CP tissues must be overexpressed in the tumor epithelia rather than the tumor stroma (which is assumed to be similar to CP). In our study, we have observed that while the above-mentioned assumption is largely correct, there are some notable cases of genes that are specifically overexpressed in tumor stroma when compared to CP stroma. Moreover, we argue that being able to determine these genes is very important for understanding the specificities of stromal component in the epithelial-stromal interactions in PDAC tumors. One such gene is IGFBP3, which we found specifically overexpressed in tumor stroma (w.r.t. both tumor epithelia and CP stroma) – this is fully consistent with IGFBP3 being overexpressed in whole tumor tissue samples w.r.t. both normal and CP tissues. (See Supplementary Figure 5.) The overexpression of IGFBP3 in tumor stroma was also observed immunohistochemically by Jenkins et al. [40] in colon cancer: “Immunohistochemistry confirmed the expression [of IGFBP3] and showed it to be equally distributed between epithelial and stromal components in normal tissue, but *to be mainly restricted to the stromal component of malignant tissue.* This differential expression was confirmed by RT-PCR of RNA from laser-capture microdissected samples.”

Other genes, like MMP11, are upregulated in the GP data in both tumor stroma and epithelia, not just in the epithelia, as inferred in the Logsdon study.

On the other hand, three of the four genes whose specific localization to neoplastic cells was confirmed by Logsdon et al. using immunocytochemistry are also in our ***TE_fc_epithelial***+***TE_p_epithelial*** lists: S100P, SFN and S100A6 (the fourth gene, ITGB4, is below our thresholds used for the GP data, but is qualitatively consistent with the Logsdon et al. study).

Figure 1. Centred $\log_2$-expression values of the genes overexpressed in tumor epithelia in the Grützmann-Pilarsky (GP) study

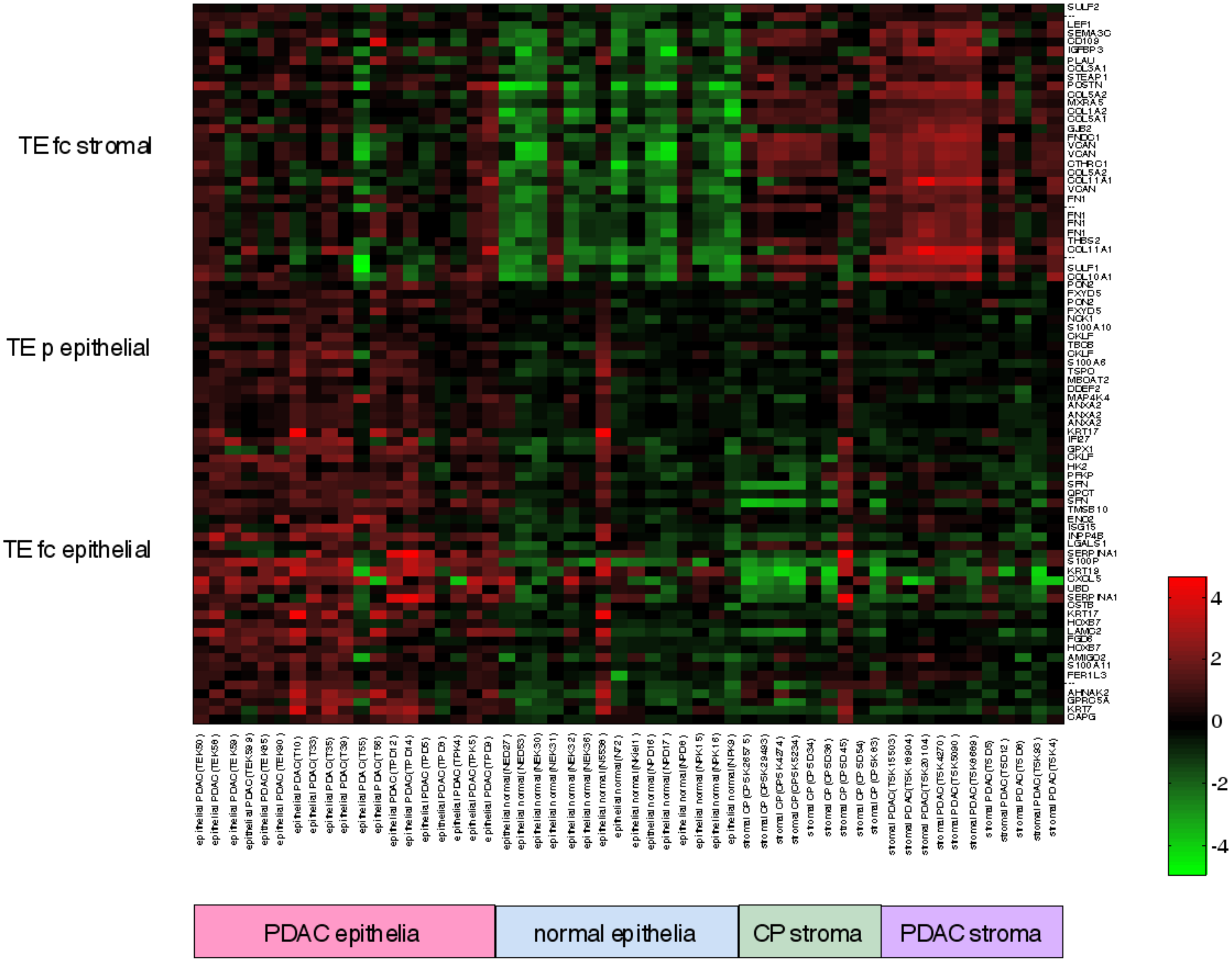
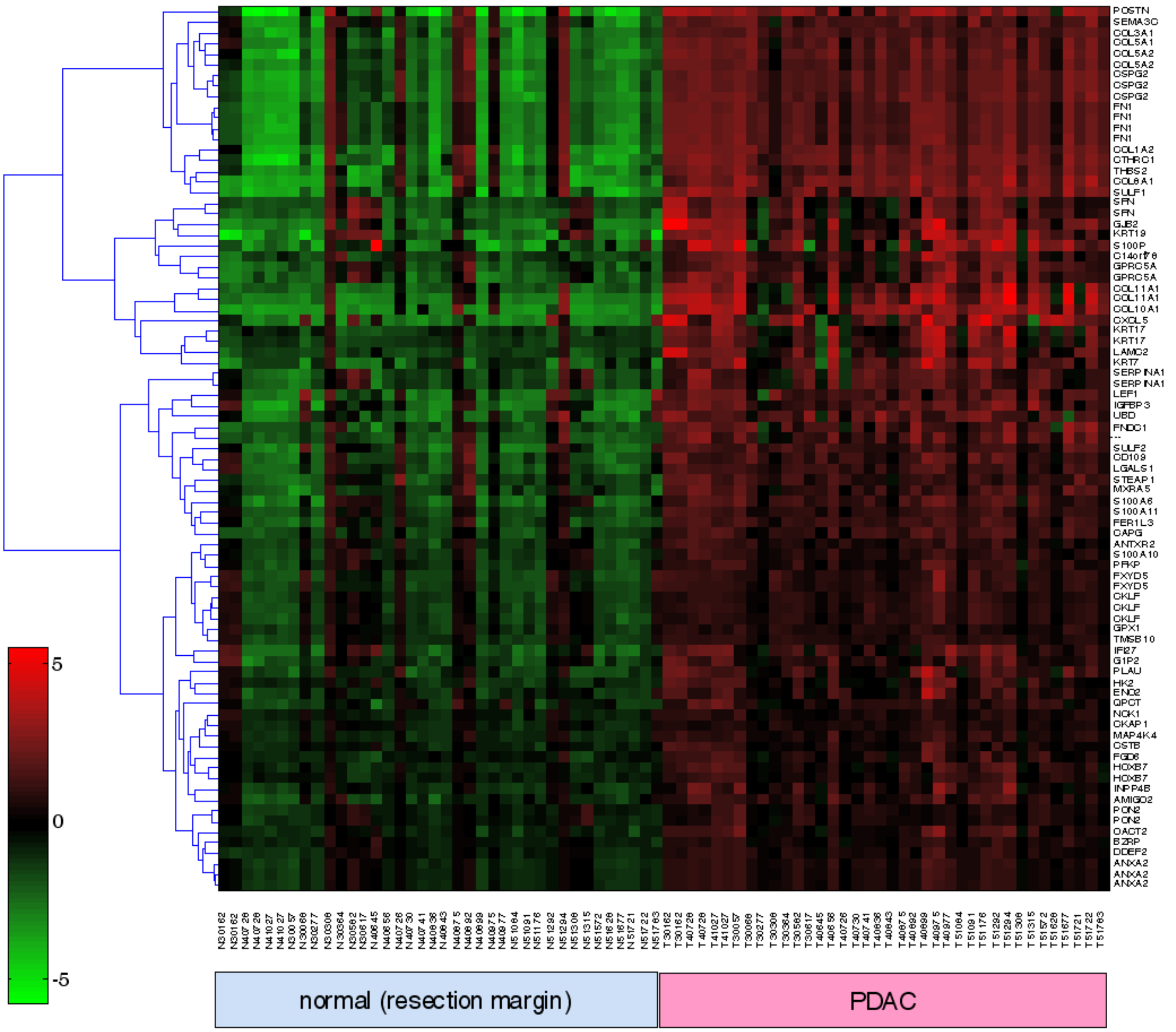
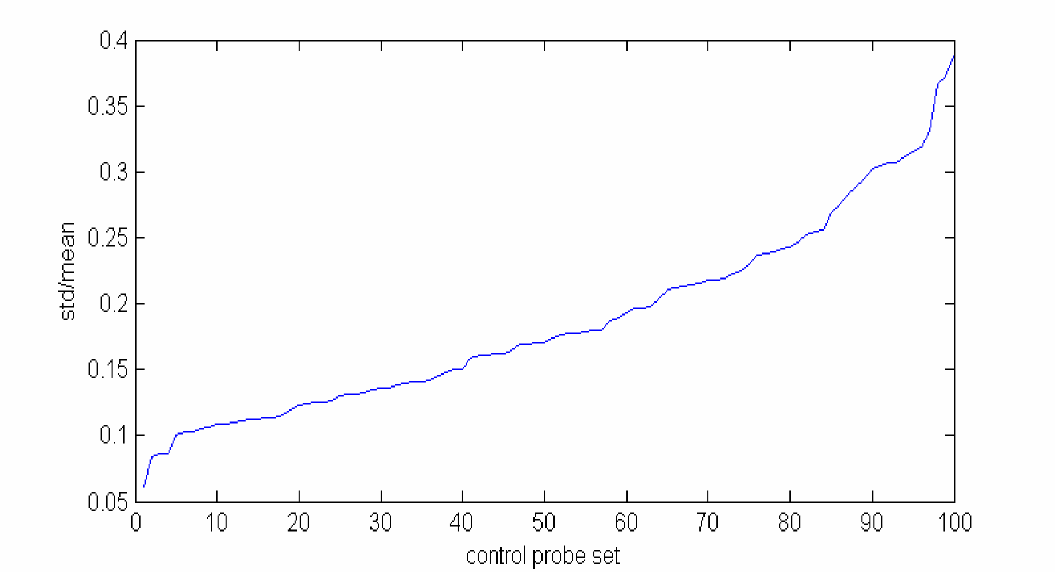


Figure 2. Centred log_2-expression values of the genes overexpressed in tumor epithelia in our whole-tissue study (ICF). Note the anomalous "normal samples" (red columns in the green area)

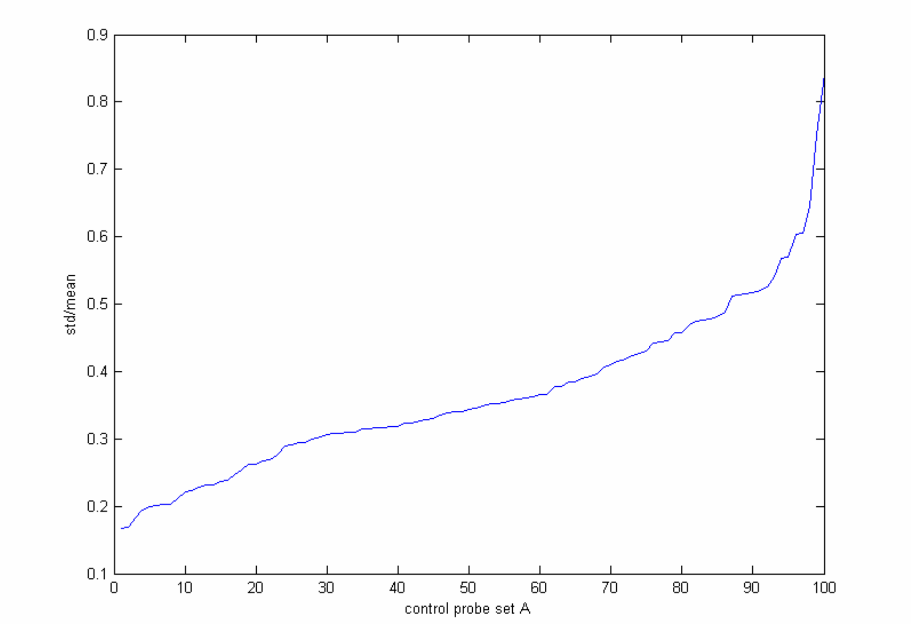


Supplementary Figure 1. The relative standard deviations (std/mean) for the control probesets of the ICF dataset (a) are two times lower than for the GP dataset (b and c)

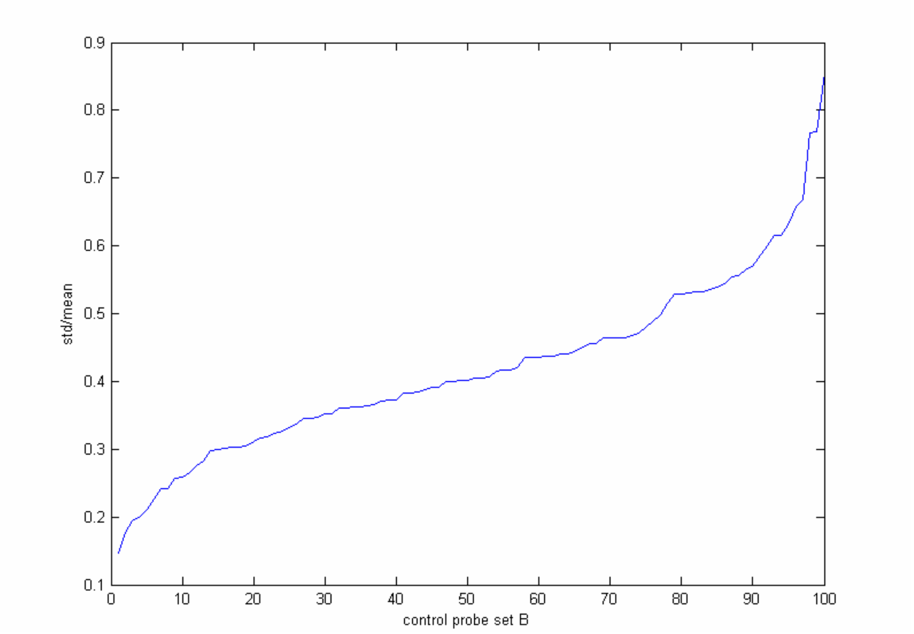
a. ICF dataset (U133 Plus 2.0 chip)



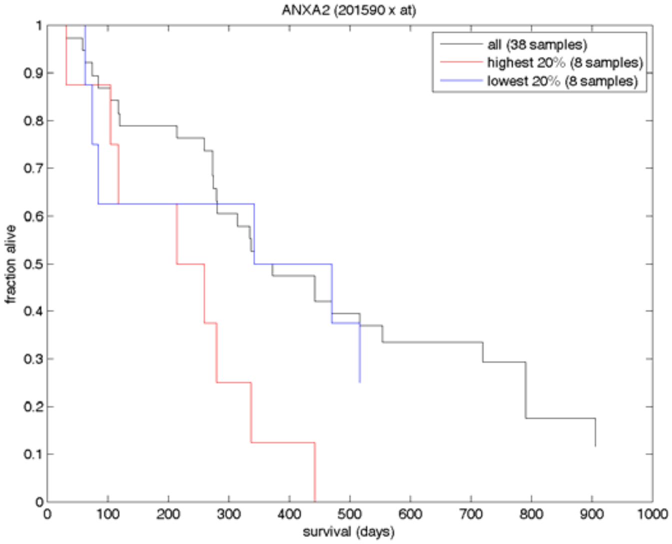
b. GP dataset (U133A chip)

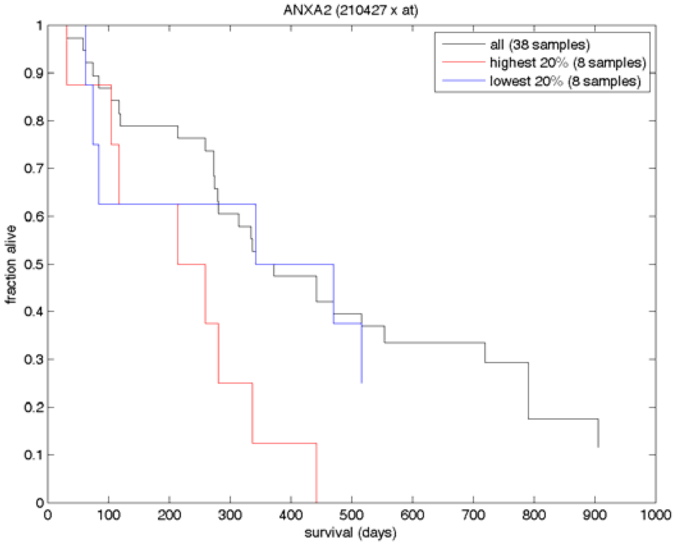


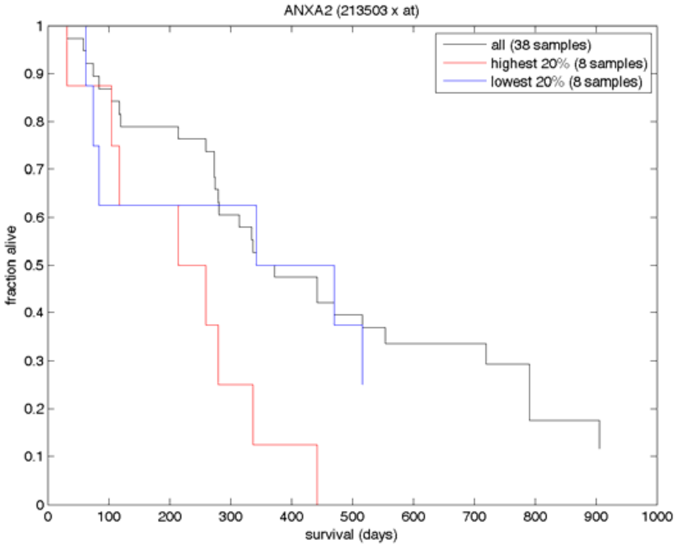
c. GP dataset (U133B chip)

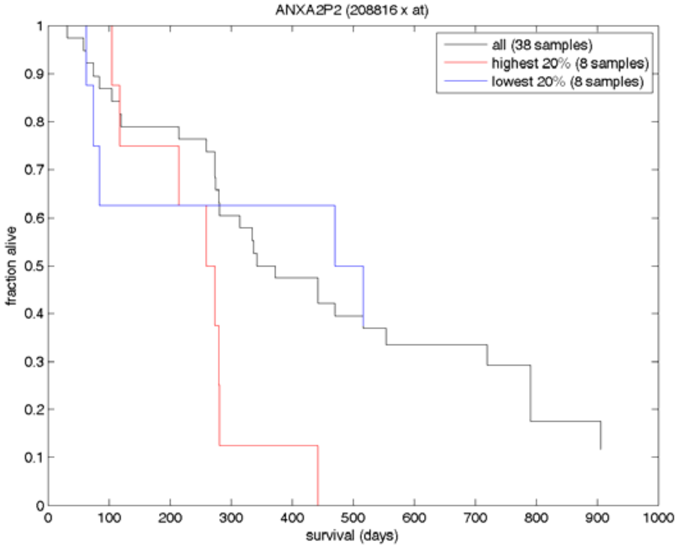


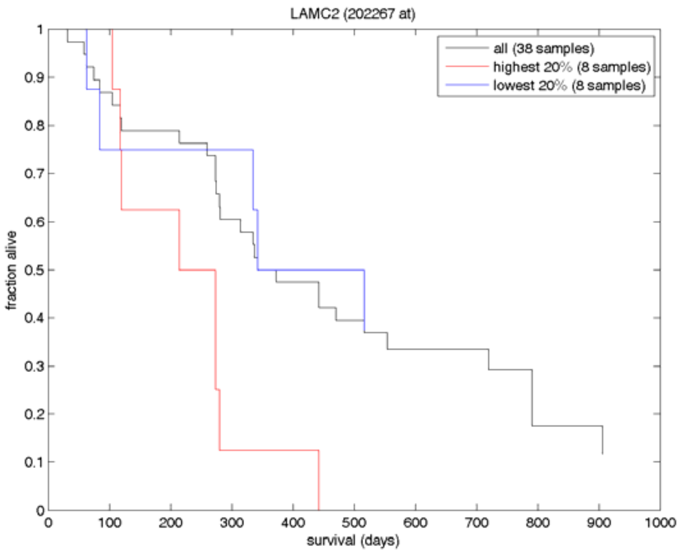
Supplementary Figure 2. Kaplan-Meier curves for the patients with high levels of expression of the differentially expressed genes found marginally correlated with survival (red curves) compared to those for all patients (black curves)

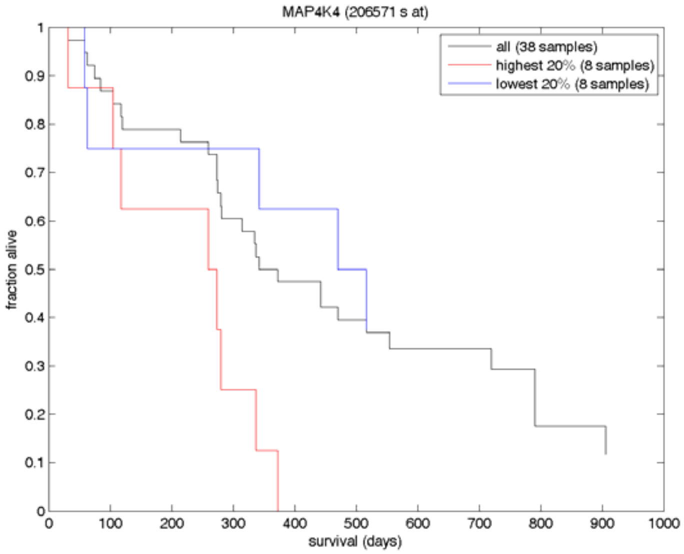


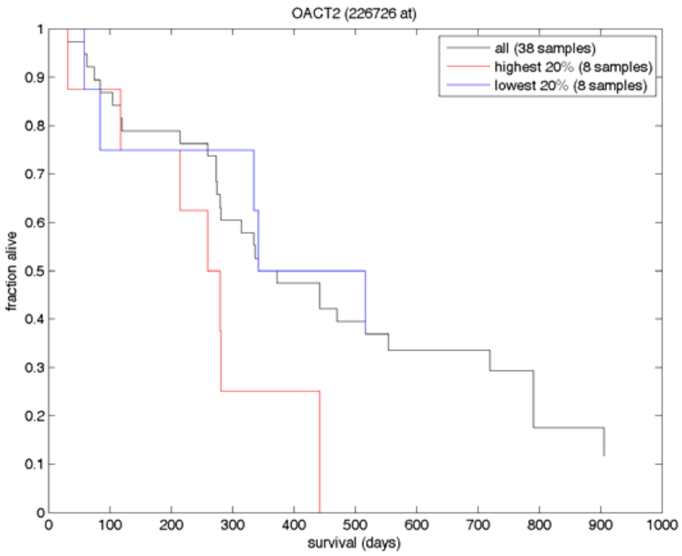


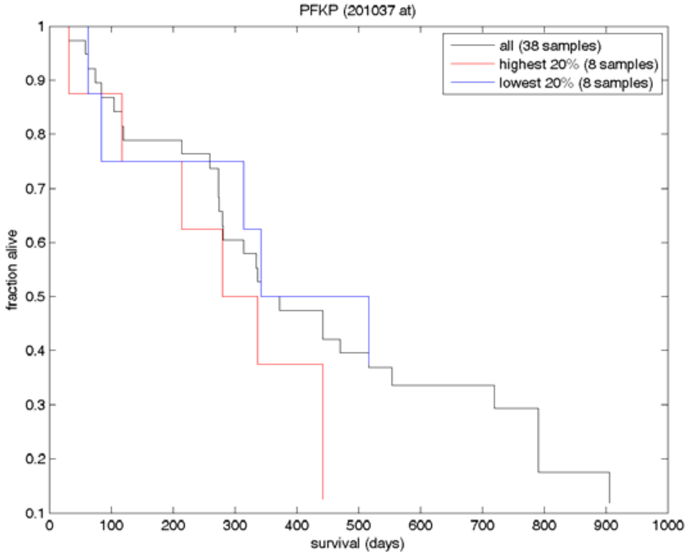

ANXA2 (213503 x at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000
survival (days)

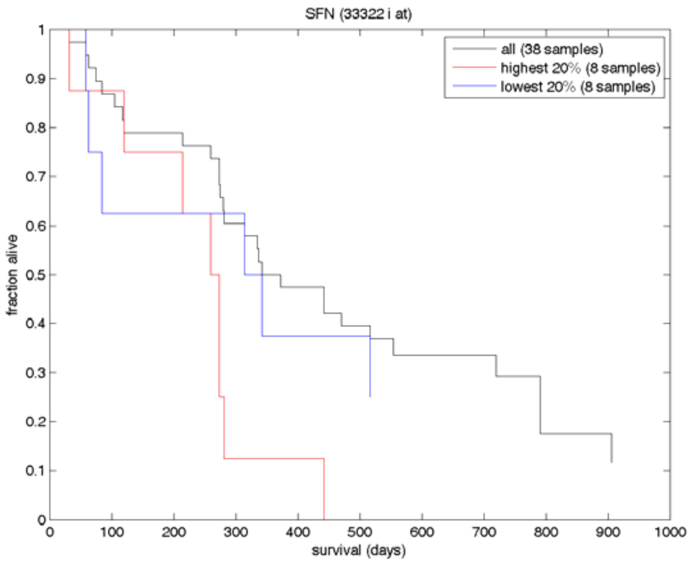

ANXA2P2 (208816 x at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000
survival (days)

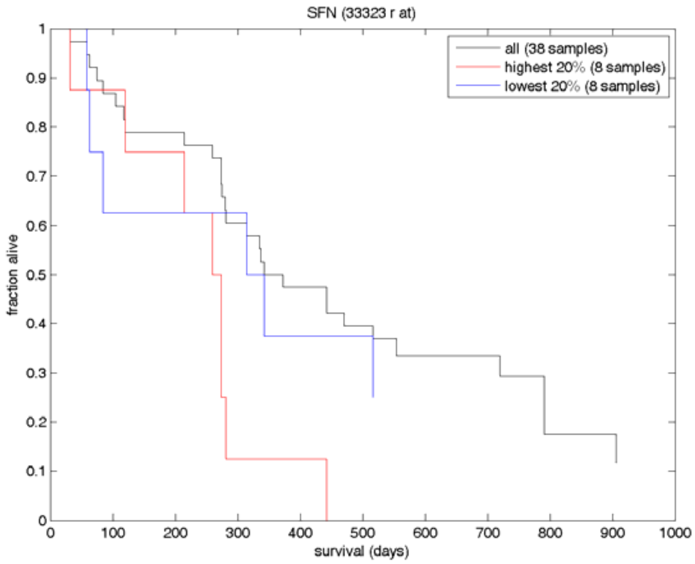

LAMC2 (202267 at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
survival (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000

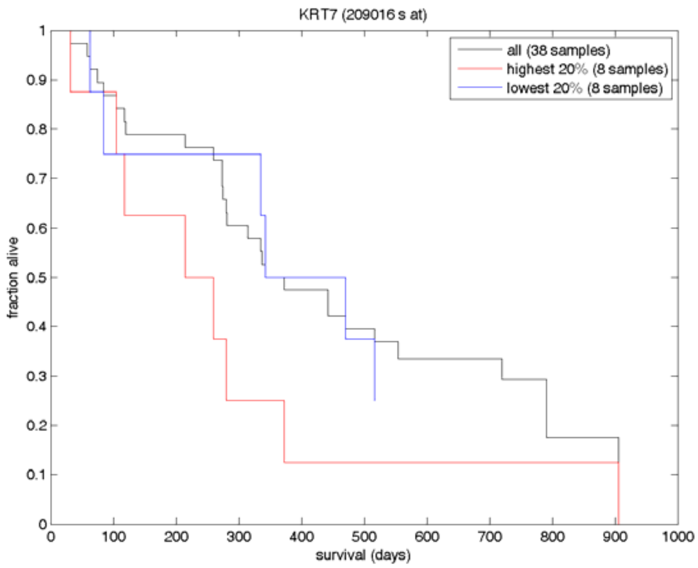

MAP4K4 (206571 s at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
survival (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000

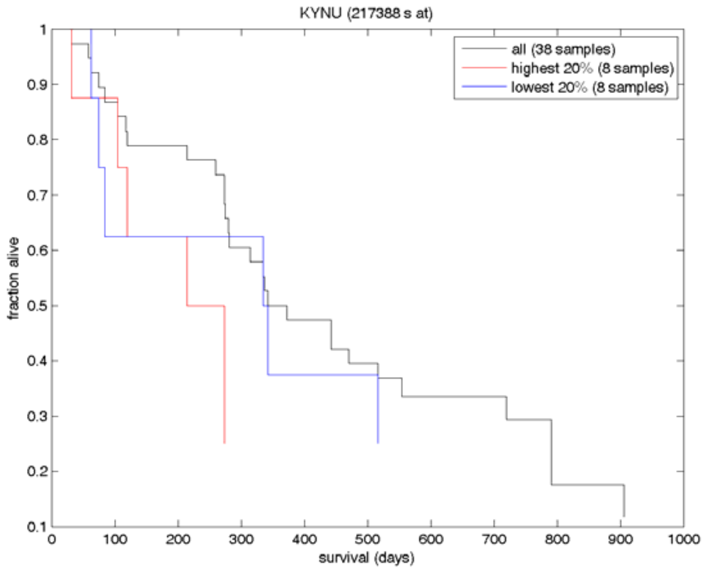

OACT2 (226726 at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
700
800
900
1000
survival (days)

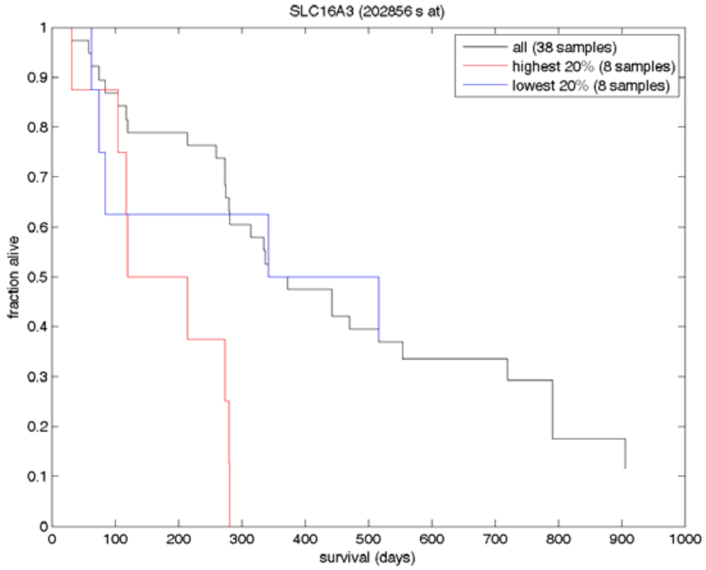

PFKP (201037 at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
200
300
400
500
600
700
800
900
1000
survival (days)

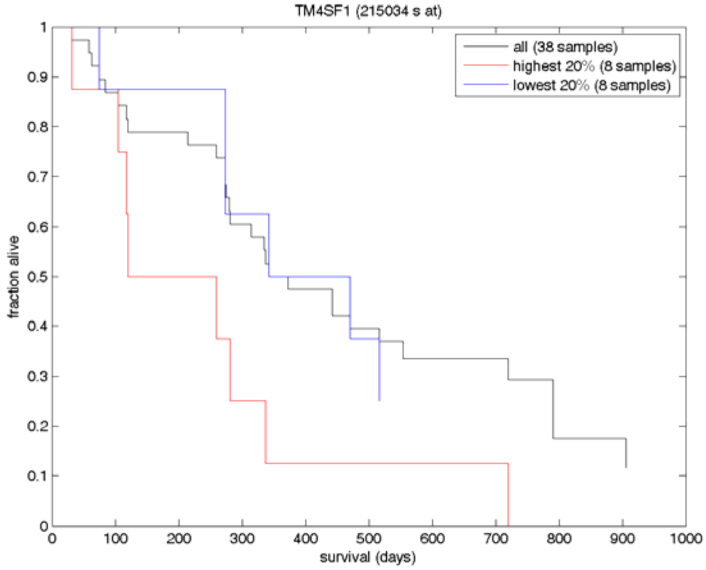

SFN (33322 i at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
survival (days)
0
100
200
300
400
500
600
700
800
900
1000

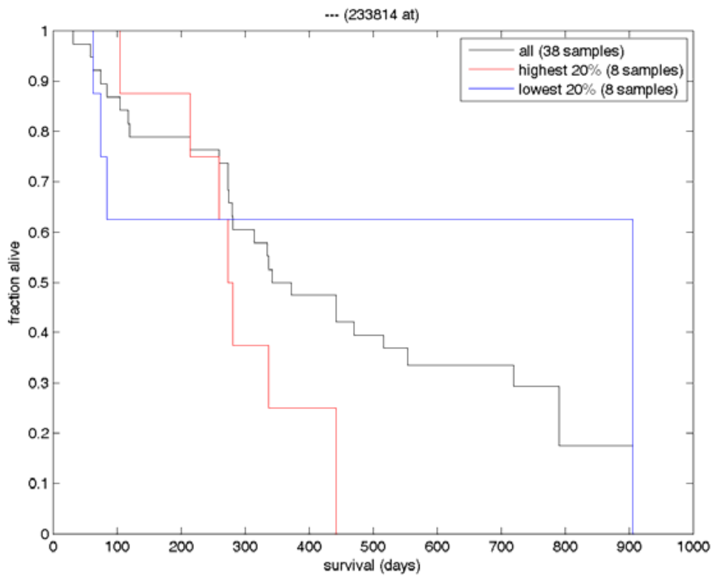

SFN (33323 r at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
survival (days)
0
100
200
300
400
500
600
700
800
900
1000


KRT7 (209016 s at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
survival (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000


KYNU (217388 s at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
survival (days)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000


SLC16A3 (202856 s at)


TM4SF1 (215034 s at)


--- (233814 at)
all (38 samples)
highest 20% (8 samples)
lowest 20% (8 samples)
fraction alive
survival (days)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000

Supplementary Table 1. The genes overexpressed in whole tumor tissue vs. normal pancreatic tissue ($p < 9 \cdot 10^{-12}$ and *log-fc* ≥ 1). Genes marginally correlated with survival are in bold font. Genes involved in the TGF-beta pathway or influenced by it transcriptionally are indicated with an X in the last column.

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
INHBA	Inhibin, beta A (activin A, activin AB alpha polypeptide)	227140_at	<2.2E-16	5.15	0.6706	X
COL10A1	collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)	217428_s_at	<2.2E-16	5	0.9609	X
SULF1	sulfatase 1	212353_at	<2.2E-16	4.63	0.8061	
COL8A1	Collagen, type VIII, alpha 1	226237_at	<2.2E-16	4.6	0.6288	X
SULF1	sulfatase 1	212354_at	<2.2E-16	4.38	0.8108	
COL10A1	collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)	205941_s_at	<2.2E-16	4.1	0.8388	X
INHBA	inhibin, beta A (activin A, activin AB alpha polypeptide)	210511_s_at	<2.2E-16	3.91	0.5268	X
HNT	neurotrimin	227566_at	<2.2E-16	3.9	0.9306	
SULF1	sulfatase 1	212344_at	<2.2E-16	3.68	0.7177	
NOX4	NADPH oxidase 4	219773_at	<2.2E-16	3.02	0.5473	X
SRPX2	sushi-repeat-containing protein, X-linked 2	205499_at	<2.2E-16	2.44	0.6478	
---	CDNA FLJ10196 fis, clone HEMBA1004776	226997_at	<2.2E-16	2.41	0.5994	
CAPG	capping protein (actin filament), gelsolin-like	201850_at	<2.2E-16	2.25	0.9151	
TMEPAI	transmembrane, prostate androgen induced RNA	222449_at	<2.2E-16	2.19	0.0187	X
LTBP1	latent transforming growth factor beta binding protein 1	202729_s_at	<2.2E-16	1.81	0.8294	X
ITGB5	integrin, beta 5	201125_s_at	<2.2E-16	1.41	0.1833	X
RASAL2	RAS protein activator like 2	222810_s_at	<2.2E-16	1.23	0.3589	
LOC162073	Hypothetical protein LOC162073	1568619_s_at	<2.2E-16	1.03	0.5727	
COL11A1	collagen, type XI, alpha 1	37892_at	2.2E-16	4.4	0.772	X
THBS2	thrombospondin 2	203083_at	2.2E-16	3.97	0.8929	X
FN1	fibronectin 1	212464_s_at	2.2E-16	3.69	0.2846	X
FN1	fibronectin 1	216442_x_at	2.2E-16	3.61	0.3046	X
FN1	fibronectin 1	210495_x_at	2.2E-16	3.56	0.2774	X
---	---	238617_at	2.2E-16	2.87	0.598	
AEBP1	AE binding protein 1	201792_at	2.2E-16	2.68	0.8476	X
DCBLD1	discoidin, CUB and LCCL domain containing 1	226609_at	2.2E-16	1.66	0.974	
TMEM16A	transmembrane protein 16A	218804_at	4.4E-16	2.9	0.5386	
TPBG	trophoblast glycoprotein	203476_at	4.4E-16	1.63	0.0945	
COMP	cartilage oligomeric matrix protein	205713_s_at	6.7E-16	3.53	0.4253	X
SOX4	SRY (sex determining region Y)-box 4	201416_at	8.9E-16	1.76	0.2043	X
FLJ12442	hypothetical protein FLJ12442	218051_s_at	8.9E-16	1.6	0.4759	
LTBP1	latent transforming growth factor beta binding protein 1	202728_s_at	8.9E-16	1.31	0.9404	X
FN1	fibronectin 1 /// fibronectin 1	211719_x_at	1.1E-15	3.68	0.2842	X
TREM2	triggering receptor expressed on myeloid cells 2	219725_at	1.1E-15	1.62	0.1062	
CSPG2	chondroitin sulfate proteoglycan 2 (versican)	204619_s_at	1.3E-15	3.87	0.9914	X
COL1A2	Collagen, type I, alpha 2	229218_at	1.3E-15	3.16	0.6324	X
SLPI	secretory leukocyte peptidase inhibitor	203021_at	1.3E-15	2.65	0.0642	X
MARVELD1	MARVEL domain containing 1	223095_at	1.6E-15	1.2	0.8907	
FN1	Fibronectin 1	235629_at	2.0E-15	2.89	0.4181	X

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
ITGA2	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	227314_at	2.0E-15	2.58	0.2731	X
PGM2L1	phosphoglucomutase 2-like 1	229553_at	2.0E-15	1.63	0.5814	
ANTXR1	anthrax toxin receptor 1	224694_at	2.2E-15	2.82	0.9509	
OLR1	oxidised low density lipoprotein (lectin-like) receptor 1	210004_at	2.4E-15	3.03	0.4339	X
MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	212473_s_at	2.7E-15	1.52	0.2504	
MFAP2	microfibrillar-associated protein 2	203417_at	2.9E-15	1.97	0.9108	
ANXA2	annexin A2	210427_x_at	2.9E-15	1.31	**0.0073**	X
KRT7	keratin 7	209016_s_at	3.1E-15	3.13	**0.0051**	X
MXRA8	matrix-remodelling associated 8	213422_s_at	3.1E-15	1.69	0.8417	
PDGFC	platelet derived growth factor C	218718_at	3.3E-15	1.36	0.7998	X
POSTN	periostin, osteoblast specific factor	1555778_a_at	3.6E-15	4.92	0.4624	X
IRS1	insulin receptor substrate 1	204686_at	3.6E-15	1.24	0.0553	
COX7A1	Cytochrome c oxidase subunit VIIa polypeptide 1 (muscle)	228481_at	3.8E-15	2.77	0.725	
ADAMTS2	MRNA; cDNA DKFZp686F12218 (from clone DKFZp686F12218) /// ADAM metallopeptidase with thrombospondin type 1 motif, 2	226311_at	3.8E-15	2.76	0.5509	X
SPON2	spondin 2, extracellular matrix protein	218638_s_at	3.8E-15	1.8	0.6711	
FBXO32	F-box protein 32	225803_at	3.8E-15	1.69	0.6703	
FAP	fibroblast activation protein, alpha	209955_s_at	4.2E-15	3.37	0.9252	
LY6E	lymphocyte antigen 6 complex, locus E	202145_at	4.2E-15	1.53	0.0856	
RASAL2	RAS protein activator like 2	227036_at	4.2E-15	1.44	0.4376	
LOXL1	lysyl oxidase-like 1	203570_at	4.4E-15	2.3	0.845	X
SH3MD1	SH3 multiple domains 1	224817_at	4.4E-15	1.59	0.3483	
MFGE8	milk fat globule-EGF factor 8 protein	210605_s_at	4.7E-15	1.21	0.1058	
SLC24A3	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	57588_at	4.9E-15	1.49	0.514	
COL11A1	collagen, type XI, alpha 1	204320_at	5.1E-15	3.51	0.7601	X
C5orf13	chromosome 5 open reading frame 13	201310_s_at	5.1E-15	2.07	0.9133	X
S100A11	S100 calcium binding protein A11 (calgizzarin)	208540_x_at	5.1E-15	1.65	0.3704	X
ANXA2	annexin A2	201590_x_at	5.1E-15	1.28	**0.0087**	X
CSPG2	chondroitin sulfate proteoglycan 2 (versican)	211571_s_at	5.6E-15	3.83	0.3866	X
EDNRA	endothelin receptor type A	216235_s_at	5.6E-15	1.82	0.8131	X
COL5A2	collagen, type V, alpha 2	221729_at	5.8E-15	3.38	0.8486	X
PLXDC2	plexin domain containing 2	227276_at	6.0E-15	2.16	0.3087	
NUAK1	NUAK family, SNF1-like kinase, 1	204589_at	6.0E-15	1.67	0.8995	
ANXA2	annexin A2	213503_x_at	6.7E-15	1.38	**0.008**	X
DACT1	dapper, antagonist of beta-catenin, homolog 1 (Xenopus laevis)	219179_at	7.1E-15	2.36	0.8479	
IER5L	immediate early response 5-like	226552_at	7.1E-15	1.31	0.5595	
CTHRC1	collagen triple helix repeat containing 1	225681_at	7.3E-15	4.38	0.9419	X
SOX4	SRY (sex determining region Y)-box 4	201417_at	7.5E-15	1.45	0.2886	X
CSPG2	chondroitin sulfate proteoglycan 2 (versican)	204620_s_at	8.2E-15	3.7	0.5497	X
ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	211945_s_at	8.4E-15	1.12	0.0335	X
COL1A1	collagen, type I, alpha 1	202311_s_at	8.7E-15	4.12	0.8808	X
COL16A1	collagen, type XVI, alpha 1	204345_at	9.1E-15	1.5	0.3062	X
WISP1	WNT1 inducible signaling pathway protein 1	229802_at	9.5E-15	3.8	0.686	X

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
IGFBP5	insulin-like growth factor binding protein 5	211958_at	1.1E-14	1.89	0.4299	X
CSPG2	chondroitin sulfate proteoglycan 2 (versican)	221731_x_at	1.1E-14	3.73	0.4459	X
SLC24A3	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	219090_at	1.1E-14	1.75	0.6179	
COL5A1	collagen, type V, alpha 1	212488_at	1.2E-14	3.46	0.7485	X
PGM2L1	phosphoglucomutase 2-like 1	229256_at	1.2E-14	1.57	0.3207	
FNDC1	fibronectin type III domain containing 1	226930_at	1.3E-14	2.71	0.2714	X
GPRC5A	G protein-coupled receptor, family C, group 5, member A	203108_at	1.3E-14	2.97	0.0545	
---	MRNA; cDNA DKFZp686M0856 (from clone DKFZp686M0856)	232458_at	1.3E-14	3.09	0.9233	
ETV1	ets variant gene 1	221911_at	1.3E-14	2.24	0.6177	
CSPG2	chondroitin sulfate proteoglycan 2 (versican) /// chondroitin sulfate proteoglycan 2 (versican)	215646_s_at	1.3E-14	4.11	0.4669	X
KIAA1223	KIAA1223 protein	225731_at	1.3E-14	1.19	0.8215	
ITGAV	integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)	202351_at	1.4E-14	1.13	0.7412	X
THY1	Thy-1 cell surface antigen	213869_x_at	1.4E-14	1.96	0.3948	
DKK3	dickkopf homolog 3 (Xenopus laevis)	214247_s_at	1.6E-14	1.85	0.8741	
DIO2	Deiodinase, iodothyronine, type II	231240_at	1.6E-14	2.04	0.9998	
---	Transcribed locus, strongly similar to NP_002137.3 homeo box B3; homeo box 2G; homeobox protein Hox-B3 [Homo sapiens]	228904_at	1.6E-14	2.13	0.9869	
ITGBL1	Integrin, beta-like 1 (with EGF-like repeat domains)	214927_at	1.7E-14	2.86	0.168	
C14orf78	chromosome 14 open reading frame 78	212992_at	1.7E-14	2.49	0.0179	
ANXA2P2	annexin A2 pseudogene 2	208816_x_at	1.7E-14	1.22	**0.0031**	
COL8A2	collagen, type VIII, alpha 2	221900_at	1.8E-14	1.95	0.5375	
FER1L3	fer-1-like 3, myoferlin (C. elegans)	211864_s_at	1.8E-14	2.27	0.457	
COL1A1	collagen, type I, alpha 1	202310_s_at	1.9E-14	4.21	0.8216	X
GPRC5A	G protein-coupled receptor, family C, group 5, member A	212444_at	1.9E-14	2.67	0.1102	
GJB2	gap junction protein, beta 2, 26kDa (connexin 26)	223278_at	2.0E-14	3.67	0.0301	
FER1L3	fer-1-like 3, myoferlin (C. elegans)	201798_s_at	2.0E-14	2.32	0.3854	
ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	1553678_a_at	2.1E-14	1.57	0.026	X
SPOCK	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican)	202363_at	2.2E-14	2.21	0.6925	
NUDT4	Nudix (nucleoside diphosphate linked moiety X)-type motif 4 pseudogene 2	228141_at	2.2E-14	2.12	0.7493	
S100A11	S100 calcium binding protein A11 (calgizzarin)	200660_at	2.2E-14	2.03	0.5039	X
ISLR	immunoglobulin superfamily containing leucine-rich repeat	207191_s_at	2.4E-14	2.42	0.7081	
CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)	207172_s_at	2.4E-14	2.33	0.1356	X
COL5A1	collagen, type V, alpha 1	212489_at	2.5E-14	3.42	0.9691	X
SLC6A6	solute carrier family 6 (neurotransmitter transporter, taurine), member 6	228754_at	2.5E-14	2.62	0.1819	
COL1A2	collagen, type I, alpha 2	202404_s_at	2.7E-14	4.17	0.6707	X
EDNRA	endothelin receptor type A	204463_s_at	2.8E-14	3.27	0.9798	X
COL5A1	collagen, type V, alpha 1	203325_s_at	2.8E-14	2.87	0.9703	X
MXRA5	matrix-remodelling associated 5	209596_at	3.0E-14	2.34	0.5318	
TNFRSF21	tumor necrosis factor receptor superfamily, member 21	218856_at	3.5E-14	1.68	0.0131	
COL5A2	collagen, type V, alpha 2	221730_at	3.9E-14	3.69	0.8346	X
TMEPAI	transmembrane, prostate androgen induced RNA	217875_s_at	3.9E-14	1.57	0.0581	X
CALU	calumenin	200755_s_at	4.1E-14	1.15	0.4584	

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
DKFZP564D166	putative ankyrin-repeat containing protein	224952_at	4.5E-14	1.38	0.3593	
ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	1553530_a_at	5.4E-14	1.43	0.0178	X
ACTN1	actinin, alpha 1	208637_x_at	5.5E-14	1.46	0.6115	X
THY1	Thy-1 cell surface antigen	208850_s_at	5.8E-14	2.33	0.2879	
ITGBL1	integrin, beta-like 1 (with EGF-like repeat domains)	205422_s_at	5.9E-14	3.38	0.1453	
OLFML2B	olfactomedin-like 2B	213125_at	6.0E-14	2.95	0.2051	
COL1A2	collagen, type I, alpha 2	202403_s_at	6.0E-14	2.98	0.6657	X
TGIF	TGFB-induced factor (TALE family homeobox)	203313_s_at	6.0E-14	1.19	0.0651	X
TIMP1	TIMP metallopeptidase inhibitor 1	201666_at	6.4E-14	3.06	0.2548	X
DKK3	dickkopf homolog 3 (Xenopus laevis)	202196_s_at	6.4E-14	1.75	0.9177	
PKM2	pyruvate kinase, muscle	201251_at	6.8E-14	1.49	0.0164	X
POSTN	periostin, osteoblast specific factor	210809_s_at	7.3E-14	5.13	0.3573	X
BGN	biglycan	201261_x_at	7.4E-14	2.29	0.9406	X
COL6A3	collagen, type VI, alpha 3	201438_at	7.8E-14	2.69	0.7181	
---	MRNA; cDNA DKFZp564O0862 (from clone DKFZp564O0862)	214807_at	8.0E-14	1.94	0.3478	
KIAA1217	KIAA1217	231807_at	8.4E-14	1.38	0.8033	
C20orf103	chromosome 20 open reading frame 103	219463_at	8.8E-14	2.23	0.3975	
ACTN1	Actinin, alpha 1	208636_at	9.9E-14	1.48	0.2278	X
MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4	206571_s_at	1.0E-13	1.35	**0.0079**	
IGFBP5	insulin-like growth factor binding protein 5	211959_at	1.0E-13	2.59	0.6833	X
NOTCH3	Notch homolog 3 (Drosophila)	203238_s_at	1.1E-13	1.55	0.6556	
C5orf13	chromosome 5 open reading frame 13	230424_at	1.1E-13	1.51	0.89	X
AMIGO2	adhesion molecule with Ig-like domain 2	222108_at	1.2E-13	1.79	0.1869	X
HOXB7	homeo box B7	216973_s_at	1.3E-13	1.23	0.522	
MOXD1	monooxygenase, DBH-like 1	1554474_a_at	1.5E-13	2.14	0.0219	
FGD6	FYVE, RhoGEF and PH domain containing 6	219901_at	1.5E-13	1.42	0.0717	
THY1	Thy-1 cell surface antigen	208851_s_at	1.5E-13	1.68	0.3714	
SQLE	squalene epoxidase	209218_at	1.5E-13	1.53	0.4757	
PLXDC1	plexin domain containing 1	219700_at	1.8E-13	1.86	0.578	
TWSG1	twisted gastrulation homolog 1 (Drosophila)	225406_at	1.8E-13	1.12	0.4667	
TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	201042_at	1.9E-13	2.16	0.1017	X
SYTL4	Synaptotagmin-like 4 (granuphilin-a)	229991_s_at	2.0E-13	1	0.904	
ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	205032_at	2.3E-13	1.38	0.1971	X
FKBP7	FK506 binding protein 7	224002_s_at	2.5E-13	1.6	0.8696	
DDEF2	development and differentiation enhancing factor 2	206414_s_at	2.7E-13	1.22	0.0483	
SMYD3	SET and MYND domain containing 3	218788_s_at	2.7E-13	1.14	0.3364	
GLIS2	GLIS family zinc finger 2	223378_at	2.8E-13	1.6	0.3992	
LUM	Lumican	229554_at	2.9E-13	2.65	0.4957	
LAMC2	laminin, gamma 2	202267_at	2.9E-13	2.76	**0.0097**	X
C1QTNF5	C1q and tumor necrosis factor related protein 5	223499_at	3.0E-13	1.22	0.2264	
BGN /// SDCCAG33	biglycan /// serologically defined colon cancer antigen 33	213905_x_at	3.1E-13	2.43	0.8944	X
DCBLD2	discoidin, CUB and LCCL domain containing 2	224911_s_at	3.2E-13	1.62	0.0593	X
MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	201069_at	3.3E-13	2.48	0.8054	X

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
TNFRSF21	tumor necrosis factor receptor superfamily, member 21	214581_x_at	3.3E-13	1.81	0.0106	
HOXB2	homeo box B2	205453_at	3.3E-13	1.64	0.8694	
HOXB7	homeo box B7	204779_s_at	3.4E-13	1.53	0.3354	
PDGFRB	platelet-derived growth factor receptor, beta polypeptide	202273_at	3.4E-13	1.78	0.9237	X
MOXD1	monooxygenase, DBH-like 1	209708_at	3.5E-13	2.25	0.0198	
C16orf30	chromosome 16 open reading frame 30	219315_s_at	3.5E-13	1.51	0.0884	
ITGBL1	Integrin, beta-like 1 (with EGF-like repeat domains)	1557080_s_at	3.5E-13	2.8	0.2006	
---	---	209446_s_at	3.7E-13	1.82	0.0292	
SERPINH1	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	207714_s_at	3.9E-13	2	0.9403	X
RUNX1	runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)	209360_s_at	3.9E-13	2.24	0.3775	X
KAL1	Kallmann syndrome 1 sequence	205206_at	3.9E-13	1.97	0.5365	X
C5orf13	chromosome 5 open reading frame 13	201309_x_at	4.1E-13	1.97	0.3906	X
JARID1B	Jumonji, AT rich interactive domain 1B (RBP2-like)	201548_s_at	4.2E-13	1.02	0.4165	
COL3A1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	201852_x_at	4.2E-13	3.23	0.942	X
OACT2	O-acyltransferase (membrane bound) domain containing 2	226726_at	4.3E-13	1.52	**0.0025**	
KRT17	keratin 17	212236_x_at	4.5E-13	2.49	0.0122	X
HTRA1	HtrA serine peptidase 1	201185_at	4.5E-13	1.52	0.5381	X
ACTN1	actinin, alpha 1	211160_x_at	4.6E-13	1.13	0.7762	X
ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	209935_at	4.8E-13	1.04	0.1142	
C8orf70	chromosome 8 open reading frame 70	205308_at	4.9E-13	1.2	0.3014	
LOC283480	Hypothetical protein LOC283480	231993_at	5.2E-13	2.55	0.2055	
PHLDA2	pleckstrin homology-like domain, family A, member 2	209803_s_at	5.2E-13	2.81	0.2145	X
LOC493869	similar to RIKEN cDNA 2310016C16	227628_at	5.7E-13	1.71	0.6221	
LTBP3	latent transforming growth factor beta binding protein 3	219922_s_at	5.9E-13	1.32	0.8045	X
ACVR1	activin A receptor, type I	203935_at	5.9E-13	1.44	0.8095	X
IGFBP3	insulin-like growth factor binding protein 3	212143_s_at	6.3E-13	2.42	0.937	X
GGTLA1	gamma-glutamyltransferase-like activity 1	205582_s_at	6.3E-13	1.12	0.2472	
CTSK	cathepsin K (pycnodysostosis)	202450_s_at	6.5E-13	2.54	0.7279	X
ADAM9	ADAM metallopeptidase domain 9 (meltrin gamma)	202381_at	6.8E-13	1.74	0.4298	X
TM4SF1	transmembrane 4 L six family member 1	209386_at	6.9E-13	1.43	0.0161	
PDLIM7	PDZ and LIM domain 7 (enigma)	203370_s_at	7.0E-13	1.28	0.474	X
ITGB6	integrin, beta 6	208083_s_at	7.0E-13	2.63	0.0243	X
FUT1	Fucosyltransferase 1 (galactoside 2-alpha-L-fucosyltransferase, H blood group)	225776_at	7.3E-13	1.01	0.6399	
INPP5F	inositol polyphosphate-5-phosphatase F	203607_at	7.4E-13	1.14	0.6646	
PTGFRN	prostaglandin F2 receptor negative regulator	224937_at	7.5E-13	1.31	0.4616	
---	Transcribed locus	226885_at	7.6E-13	1.25	0.4335	
RUNX1	Runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)	236114_at	7.6E-13	2.18	0.5631	X
EDNRA	endothelin receptor type A	204464_s_at	7.7E-13	2.8	0.6945	X
COL3A1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	215076_s_at	7.7E-13	2.76	0.5728	X
HSPG2	heparan sulfate proteoglycan 2 (perlecan)	201655_s_at	7.9E-13	1.07	0.8472	X
BHLHB2	basic helix-loop-helix domain containing, class B, 2	201170_s_at	7.9E-13	1.61	0.1384	X
SMURF1	SMAD specific E3 ubiquitin protein ligase 1	212666_at	8.2E-13	1.05	0.5539	X
LAMB1	laminin, beta 1 /// laminin, beta 1	211651_s_at	8.2E-13	1.43	0.7498	X

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
STEAP1	six transmembrane epithelial antigen of the prostate 1	205542_at	8.3E-13	2.17	0.4539	
CSTB	cystatin B (stefin B)	201201_at	8.4E-13	1.21	0.0831	
CD9	CD9 antigen (p24)	201005_at	8.7E-13	1.41	0.4408	
ITGB5	integrin, beta 5	201124_at	8.8E-13	1.1	0.564	X
YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	201020_at	9.0E-13	1	0.9451	
KIAA1295	KIAA1295	231823_s_at	9.3E-13	1.29	0.9483	
HOP	homeodomain-only protein /// homeodomain-only protein	211597_s_at	9.5E-13	3.04	0.6314	
BZRP	benzodiazapine receptor (peripheral)	202096_s_at	9.9E-13	1.01	0.2967	
LRRC32	leucine rich repeat containing 32	203835_at	1.0E-12	1.38	0.1463	
MGC24103	hypothetical protein MGC24103	232568_at	1.1E-12	2.33	0.7664	
UNC5B	unc-5 homolog B (C. elegans)	226899_at	1.1E-12	1.41	0.5662	
SESTD1	SEC14 and spectrin domains 1	226763_at	1.1E-12	1.04	0.0211	
CST1	cystatin SN	206224_at	1.2E-12	3.3	0.9436	
SESTD1	SEC14 and spectrin domains 1	227041_at	1.2E-12	1.29	0.0627	
CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)	207173_x_at	1.2E-12	3.57	0.1425	X
PLAU	plasminogen activator, urokinase /// plasminogen activator, urokinase	211668_s_at	1.2E-12	2.36	0.0947	X
SLC16A3	solute carrier family 16 (monocarboxylic acid transporters), member 3	202856_s_at	1.3E-12	2.63	**0.003**	
FBXO32	F-box protein 32	225328_at	1.4E-12	1.86	0.4178	
TMEPAI	transmembrane, prostate androgen induced RNA	222450_at	1.4E-12	1.71	0.7096	X
FLJ20481	hypothetical protein FLJ20481	227889_at	1.5E-12	1.41	0.1668	
ENC1	ectodermal-neural cortex (with BTB-like domain)	201341_at	1.5E-12	1.15	0.1224	X
TRIO	triple functional domain (PTPRF interacting)	209012_at	1.5E-12	1.25	0.9234	
CLEC11A	C-type lectin domain family 11, member A /// C-type lectin domain family 11, member A	211709_s_at	1.5E-12	1.65	0.9755	
RAB31	RAB31, member RAS oncogene family	217763_s_at	1.6E-12	2.73	0.8449	
PDGFC	platelet derived growth factor C	222719_s_at	1.6E-12	1.37	0.6574	X
COL12A1	collagen, type XII, alpha 1	225664_at	1.6E-12	2.57	0.7426	
RAB31	RAB31, member RAS oncogene family	217762_s_at	1.7E-12	2.84	0.9975	
SFRP2	secreted frizzled-related protein 2	223121_s_at	1.7E-12	3.29	0.4674	
MYO1E	myosin IE	203072_at	1.7E-12	1.2	0.1968	X
ANKRD10	Ankyrin repeat domain 10	235008_at	1.8E-12	1.18	0.9283	
PLEC1	plectin 1, intermediate filament binding protein 500kDa	216971_s_at	1.8E-12	1.14	0.0849	
ARHGAP26	Rho GTPase activating protein 26	205068_s_at	1.9E-12	1.15	0.337	
ASPN	asporin (LRR class 1)	219087_at	1.9E-12	2.21	0.3364	
COL3A1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	211161_s_at	1.9E-12	3.23	0.6954	X
FAM20C	family with sequence similarity 20, member C	226722_at	2.0E-12	1.16	0.9639	
SPARC	secreted protein, acidic, cysteine-rich (osteonectin) /// secreted protein, acidic, cysteine-rich (osteonectin)	200665_s_at	2.0E-12	2.94	0.8426	X
OLR1	Oxidised low density lipoprotein (lectin-like) receptor 1	242397_at	2.0E-12	2.51	0.848	X
ECM1	extracellular matrix protein 1	209365_s_at	2.1E-12	1.15	0.3669	
---	MRNA; cDNA DKFZp564O0862 (from clone DKFZp564O0862)	226865_at	2.2E-12	2.41	0.1459	
MRC2	mannose receptor, C type 2	37408_at	2.2E-12	1.38	0.7727	
FN1	fibronectin 1	214701_s_at	2.2E-12	1.57	0.0267	X
BGN	biglycan	201262_s_at	2.3E-12	1.6	0.888	X
SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	211429_s_at	2.3E-12	2.31	0.9265	

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
PLAU	plasminogen activator, urokinase	205479_s_at	2.4E-12	2.04	0.0967	X
TM4SF1	transmembrane 4 L six family member 1	215034_s_at	2.5E-12	1.66	**0.0085**	
MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	212472_at	2.5E-12	1.11	0.496	
CDH11	Cadherin 11, type 2, OB-cadherin (osteoblast)	236179_at	2.5E-12	2.59	0.7495	X
UBD	ubiquitin D	205890_s_at	2.6E-12	2.9	0.2384	
IGFBP3	insulin-like growth factor binding protein 3	210095_s_at	2.6E-12	3.01	0.9378	X
COL12A1	collagen, type XII, alpha 1	231766_s_at	2.7E-12	2.93	0.387	
LTBP2	latent transforming growth factor beta binding protein 2	204682_at	2.8E-12	1.68	0.427	
---	MRNA; cDNA DKFZp564O0862 (from clone DKFZp564O0862)	227995_at	3.0E-12	1.83	0.1256	
SERPING1	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	200986_at	3.0E-12	1.47	0.6369	
MMP11	matrix metallopeptidase 11 (stromelysin 3)	203878_s_at	3.2E-12	2.1	0.3425	
SGCB	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)	205120_s_at	3.2E-12	1.12	0.0289	
RAB31	RAB31, member RAS oncogene family	217764_s_at	3.2E-12	2.67	0.9571	
COL1A1	collagen, type I, alpha 1	1556499_s_at	3.2E-12	2.54	0.5279	X
SFRP4	secreted frizzled-related protein 4	204051_s_at	3.2E-12	3.38	0.0287	
CALD1	Caldesmon 1	235834_at	3.2E-12	1.26	0.805	
MMP14	matrix metallopeptidase 14 (membrane-inserted)	202827_s_at	3.3E-12	1.22	0.0771	
RSAD2	radical S-adenosyl methionine domain containing 2	242625_at	3.4E-12	1.85	0.5598	
FZD1	frizzled homolog 1 (Drosophila)	204451_at	3.4E-12	1.11	0.5405	
CXCL5	chemokine (C-X-C motif) ligand 5	214974_x_at	3.5E-12	3.73	0.0835	
LEPRE1	leucine proline-enriched proteoglycan (leprecan) 1	220750_s_at	3.6E-12	1.1	0.8589	
GLT8D2	glycosyltransferase 8 domain containing 2	227070_at	3.6E-12	2.39	0.6342	
CKLFSF3	chemokine-like factor superfamily 3	224733_at	3.8E-12	1.66	0.7115	
RBM9	RNA binding motif protein 9	212104_s_at	3.8E-12	1.03	0.823	
CKLFSF3	chemokine-like factor superfamily 3	1555705_a_at	4.1E-12	1.18	0.901	
IGFBP7	insulin-like growth factor binding protein 7	201162_at	4.2E-12	1.63	0.1371	
GPNMB	glycoprotein (transmembrane) nmb	201141_at	4.3E-12	2.39	0.3595	
SFRP4	secreted frizzled-related protein 4	204052_s_at	4.4E-12	3.15	0.0907	
ASPN	asporin (LRR class 1) /// asporin (LRR class 1)	224396_s_at	4.4E-12	2.36	0.5048	
DBN1	drebrin 1	202806_at	4.4E-12	1.11	0.1214	
LXN	latexin	218729_at	4.5E-12	1.96	0.4117	
PLAUR	plasminogen activator, urokinase receptor	210845_s_at	4.6E-12	1.9	0.0182	
ADAM9	ADAM metallopeptidase domain 9 (meltrin gamma)	1555326_a_at	4.9E-12	1.7	0.4045	X
CD109	CD109 antigen (Gov platelet alloantigens)	226545_at	4.9E-12	2.35	0.0133	
---	---	214806_at	4.9E-12	1.36	0.9353	
RARRES1	retinoic acid receptor responder (tazarotene induced) 1	206392_s_at	4.9E-12	3	0.6491	
C1RL	complement component 1, r subcomponent-like	218983_at	5.2E-12	1.11	0.2993	
PPAPDC1A	phosphatidic acid phosphatase type 2 domain containing 1A	236044_at	5.4E-12	2.2	0.1356	
ZNF532	zinc finger protein 532	220617_s_at	5.5E-12	1.42	0.5204	
ADAM12	ADAM metallopeptidase domain 12 (meltrin alpha)	226777_at	5.5E-12	3.07	0.9313	
PRKCI	protein kinase C, iota	213518_at	5.7E-12	1.14	0.4	
LAMB1	laminin, beta 1	201505_at	5.7E-12	1.56	0.9641	X
GALNACT-2	chondroitin sulfate GalNAcT-2	222235_s_at	5.9E-12	1.7	0.9932	

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	survival (Cox p)	TGF-beta signature
TIMP2	TIMP metallopeptidase inhibitor 2	224560_at	6.0E-12	1.53	0.4772	
MGP	matrix Gla protein	238481_at	6.1E-12	2.47	0.0275	
TNFAIP6	tumor necrosis factor, alpha-induced protein 6	206026_s_at	6.2E-12	2.56	0.1748	
F2R	coagulation factor II (thrombin) receptor	203989_x_at	6.6E-12	2.23	0.243	
MITF	microphthalmia-associated transcription factor	207233_s_at	6.7E-12	1.11	0.5767	
KYNU	kynureninase (L-kynurenine hydrolase)	217388_s_at	6.9E-12	2.72	**0.0029**	
PRRX1	paired related homeobox 1	226695_at	7.0E-12	2.64	0.7754	
LTBP2	latent transforming growth factor beta binding protein 2	223690_at	7.0E-12	1.58	0.3105	
RARRES1	retinoic acid receptor responder (tazarotene induced) 1	221872_at	7.0E-12	3.25	0.5892	
TGFBR1	transforming growth factor, beta receptor I (activin A receptor type II-like kinase, 53kDa)	224793_s_at	7.2E-12	1.39	0.4777	
FLJ14001	Hypothetical protein FLJ14001	212985_at	7.2E-12	1.39	0.9163	
---	---	226535_at	7.3E-12	3.07	0.0511	
HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1	205404_at	7.4E-12	1.73	0.1011	
UACA	uveal autoantigen with coiled-coil domains and ankyrin repeats	223279_s_at	7.5E-12	1.22	0.6757	
KRT19	keratin 19	201650_at	7.6E-12	3.7	0.0131	
---	CDNA: FLJ22256 fis, clone HRC02860	233814_at	7.7E-12	1.81	**0.0072**	
LOC133308	hypothetical protein BC009732	229491_at	7.9E-12	1.28	0.5763	
PPM2C	protein phosphatase 2C, magnesium-dependent, catalytic subunit	218273_s_at	8.1E-12	1.1	0.012	
LOX	lysyl oxidase	215446_s_at	8.1E-12	3.04	0.3003	
RARRES1	retinoic acid receptor responder (tazarotene induced) 1	206391_at	8.4E-12	2.49	0.8118	
TIMP2	TIMP metallopeptidase inhibitor 2	231579_s_at	8.7E-12	1.45	0.5122	
PLXDC1	plexin domain containing 1	214081_at	8.7E-12	1.26	0.4276	
COL4A1	collagen, type IV, alpha 1	211981_at	8.8E-12	2.29	0.7164	
S100P	S100 calcium binding protein P	204351_at	8.9E-12	3.66	0.6669	
HIP1	Huntingtin interacting protein 1	226364_at	9.0E-12	1.6	0.9789	

Supplementary Table 2. The genes underexpressed in whole tumor tissue vs. normal pancreatic tissue
($p < 9\cdot10^{-12}$ and $\mathit{log\text{-}fc} < 0$)

Gene Symbol	Gene Title	ProbeSetID	p	log-fc	TGF-beta signature
EIF4E3	eukaryotic translation initiation factor 4E member 3	225941_at	1.6E-12	-0.6	
SFXN2	sideroflexin 2	227560_at	1.6E-12	-0.76	
TOPORS	Topoisomerase I binding, arginine/serine-rich	221979_at	6.5E-12	-0.75	

Supplementary Table 3. Comparison with other microarray studies of pancreatic cancer
T: gene over-expressed in tumors, N: gene over-expressed in normal samples (i.e. *under*-expressed in tumors)

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
S100P	8.90e-012 (T)	15	T	T	T			T	T				T			T		T	T	T	T	T	T	T		T	
FN1	2.22e-016 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	2.22e-016 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	2.22e-016 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	1.11e-015 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	2.00e-015 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	2.24e-012 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
LUM	2.94e-013 (T)	9					N		T				T	T			T				T				T	T	X
CAPG	<2.22e-016 (T)	9					T	T	T					T		T		T		T		T	T				
CSPG2	1.33e-015 (T)	9	T		T				T		T			T		T	T								T	T	
CSPG2	5.55e-015 (T)	9	T		T				T		T			T		T	T								T	T	
CSPG2	8.22e-015 (T)	9	T		T				T		T			T		T	T								T	T	
COL1A1	8.66e-015 (T)	9	T		T				T					T		T					T			T		T	X
CSPG2	1.13e-014 (T)	9	T		T				T		T			T		T	T								T	T	
CSPG2	1.33e-014 (T)	9	T		T				T		T			T		T	T								T	T	
COL1A1	1.89e-014 (T)	9	T		T				T					T		T					T			T		T	X
COL1A1	3.24e-012 (T)	9	T		T				T					T		T					T			T		T	X
COL1A2	1.33e-015 (T)	9	T		T				T	T				T		T	T								T		X
COL1A2	2.71e-014 (T)	9	T		T				T	T				T		T	T								T		X
COL1A2	6.00e-014 (T)	9	T		T				T	T				T		T	T								T		X
THBS2	2.22e-016 (T)	8	T		T				T					T		T							T	N		T	
POSTN	3.55e-015 (T)	8	T				T		T					T		T				T						T	X
S100A11	5.11e-015 (T)	8			T				T						T			T	T			T	T			T	
S100A11	2.22e-014 (T)	8			T				T						T			T	T			T	T			T	
PKM2	6.77e-014 (T)	8					T		T					T					T		T			N	T	T	
POSTN	7.35e-014 (T)	8	T				T		T					T		T				T						T	X
COL3A1	4.23e-013 (T)	8	T		T				T		T		T	T											T		X
COL3A1	7.72e-013 (T)	8	T		T				T		T		T	T											T		X

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
COL3A1	1.92e-012 (T)	8	T		T				T		T		T	T											T		X
PLAUR	4.62e-012 (T)	8		T					T						T	T		T	T				T			T	
IGFBP3	6.31e-013 (T)	7				T			T									T	T			T	T				X
IGFBP3	2.61e-012 (T)	7				T			T									T	T			T	T				X
SERPINH1	3.90e-013 (T)	7	T				T		T					T		T							T			T	
KRT17	4.49e-013 (T)	7	T						T					T				T	T				T	N			
SPARC	2.02e-012 (T)	7	T		T	T			T					T			T									T	
CALD1	3.24e-012 (T)	7			T		T		T			T		T											T	T	
KRT19	7.63e-012 (T)	7					T		T								T	T	T			T	T				
LAMC2	2.94e-013 (T)	7	T						T							T		T	T			T	T				
IGFBP5	1.09e-014 (T)	6							T				N	T										T		T	X
IGFBP5	1.04e-013 (T)	6							T				N	T										T		T	X
SLPI	1.33e-015 (T)	6	T						T									T				T	T			T	
GPRC5A	1.29e-014 (T)	6			T								T			T		T		T							X
GPRC5A	1.91e-014 (T)	6			T								T			T		T		T							X
CD9	8.71e-013 (T)	6							T		T							T				T			T	T	
PLAU	1.24e-012 (T)	6	T													T					T		T			T	X
PLAU	2.39e-012 (T)	6	T													T					T		T			T	X
MMP11	3.16e-012 (T)	6	T		T															T			T			T	X
COMP	6.66e-016 (T)	5	T						T					T									T				X
KRT7	3.11e-015 (T)	5	T				T			T						T			T								
HSPG2	7.86e-013 (T)	5	T						T					T					T			T					
COL10A1	<2.22e-016 (T)	5	T						T													T	T				X
COL10A1	<2.22e-016 (T)	5	T						T													T	T				X
FER1L3	1.84e-014 (T)	5	T												T	T		T					T				
FER1L3	2.04e-014 (T)	5	T												T	T		T					T				
CDH11	2.40e-014 (T)	5	T						T					T		T										T	
PHLDA2	5.20e-013 (T)	5							T				T			T			T			T					
CDH11	1.23e-012 (T)	5	T						T					T		T										T	
CDH11	2.53e-012 (T)	5	T						T					T		T										T	
COL6A3	7.82e-014 (T)	4	T						T					T													X
TGM2	1.94e-013 (T)	4		T			T									T		T									

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
ADAM9	6.83e-013 (T)	4							T									T				T				T	
ADAM9	4.85e-012 (T)	4							T									T				T				T	
ITGA2	2.00e-015 (T)	4										T						T	T							T	
FAP	4.22e-015 (T)	4	T						T							T										T	
LOXL1	4.44e-015 (T)	4	T						T					T												N	
COL5A2	5.77e-015 (T)	4	T						T																T	T	
GJB2	2.00e-014 (T)	4							T							T						T	T				
COL5A2	3.93e-014 (T)	4	T						T																T	T	
ITGA2	2.28e-013 (T)	4										T						T	T							T	
RUNX1	3.92e-013 (T)	4	T						T							T							T				
RUNX1	7.63e-013 (T)	4	T						T							T							T				
RAB31	1.58e-012 (T)	4							T								T					T				T	
RAB31	1.67e-012 (T)	4							T								T					T				T	
RAB31	3.22e-012 (T)	4							T								T					T				T	
INHBA	<2.22e-016 (T)	3	T																T				T				
INHBA	<2.22e-016 (T)	3	T																T				T				
TIMP1	6.37e-014 (T)	3								T								T							T		
PLEC1	1.84e-012 (T)	3											T										T			T	
ECM1	2.05e-012 (T)	3							T																	T	X
GPNMB	4.35e-012 (T)	3							T					T													X
SULF1	<2.22e-016 (T)	3	T										T			T											
SULF1	<2.22e-016 (T)	3	T										T			T											
SULF1	<2.22e-016 (T)	3	T										T			T											
TMEPAI	<2.22e-016 (T)	3														T							T				X
COL11A1	2.22e-016 (T)	3	T						T							T											
TMEM16A	4.44e-016 (T)	3	T													T											X
COL11A1	5.11e-015 (T)	3	T						T							T											
C5orf13	5.11e-015 (T)	3							T					T												T	
ITGBL1	1.67e-014 (T)	3											T			T							T				
TMEPAI	3.93e-014 (T)	3														T							T				X
ITGBL1	5.86e-014 (T)	3											T			T							T				
C5orf13	1.12e-013 (T)	3							T					T												T	

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
ITGBL1	3.54e-013 (T)	3											T			T							T				
C5orf13	4.10e-013 (T)	3							T					T												T	
ACVR1	5.94e-013 (T)	3							T					T												T	
CTSK	6.50e-013 (T)	3							T																T	T	
TM4SF1	6.89e-013 (T)	3	T													T											X
PDLIM7	6.96e-013 (T)	3	T						T										T								
FUT1	7.25e-013 (T)	3			N				N					N													
TMEPAI	1.43e-012 (T)	3														T							T				X
SFRP2	1.72e-012 (T)	3									N						T										X
TM4SF1	2.48e-012 (T)	3	T													T											X
MMP14	3.30e-012 (T)	3	T													T		T									
CXCL5	3.48e-012 (T)	3	T																				T				X
SERPINA1	2.32e-012 (T)	2																						T		T	
ANXA2	2.89e-015 (T)	2					T					T															
ANXA2	5.11e-015 (T)	2					T					T															
ANXA2	6.66e-015 (T)	2					T					T															
ACTN1	5.46e-014 (T)	2	T									T															
ACTN1	9.88e-014 (T)	2	T									T															
ACTN1	4.57e-013 (T)	2	T									T															
SERPING1	3.05e-012 (T)	2			N																						X
COL4A1	8.77e-012 (T)	2																							T		X
EDNRA	5.55e-015 (T)	2	T						T																		
THY1	1.44e-014 (T)	2							T					T													
ISLR	2.35e-014 (T)	2			T																						X
EDNRA	2.80e-014 (T)	2	T						T																		
CALU	4.06e-014 (T)	2																		T							X
THY1	5.80e-014 (T)	2							T					T													
BGN	7.44e-014 (T)	2	T													T											
THY1	1.49e-013 (T)	2							T					T													
PDGFRB	3.37e-013 (T)	2														T										T	
EDNRA	7.72e-013 (T)	2	T						T																		
BGN	2.26e-012 (T)	2	T													T											

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
RARRES1	4.93e-012 (T)	2		T										T													
RARRES1	7.01e-012 (T)	2		T										T													
RARRES1	8.37e-012 (T)	2		T										T													
ITGB5	<2.22e-016 (T)	2															T									T	
RASAL2	<2.22e-016 (T)	2														T							T				
AEBP1	2.22e-016 (T)	2	T																							N	
TPBG	4.44e-016 (T)	2							T					T													
SOX4	8.88e-016 (T)	2	T										T														
MICAL2	2.66e-015 (T)	2			T	T																					
MFAP2	2.89e-015 (T)	2	T																				T				
LY6E	4.22e-015 (T)	2																	T								X
RASAL2	4.22e-015 (T)	2														T							T				
SOX4	7.55e-015 (T)	2	T										T														
COL16A1	9.10e-015 (T)	2							T				N														
C14orf78	1.67e-014 (T)	2																		T							X
MXRA5	3.00e-014 (T)	2	T		T																						
MAP4K4	1.00e-013 (T)	2														T							T				
FGD6	1.47e-013 (T)	2														T											X
DDEF2	2.68e-013 (T)	2											T					T									
DCBLD2	3.16e-013 (T)	2											N										T				
HOXB2	3.34e-013 (T)	2		T										N													
GGTLA1	6.31e-013 (T)	2							T																	T	
ITGB5	8.79e-013 (T)	2															T									T	
CST1	1.20e-012 (T)	2	T						T																		
SLC16A3	1.29e-012 (T)	2														T							T				
MICAL2	2.49e-012 (T)	2			T	T																					
UBD	2.61e-012 (T)	2	T													T											
DBN1	4.45e-012 (T)	2														T							T				
LTBP1	<2.22e-016 (T)	1																								T	
LTBP1	8.88e-016 (T)	1																								T	
NOTCH3	1.05e-013 (T)	1							T																		
MMP2	3.28e-013 (T)	1																								T	

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
CSTB	8.43e-013 (T)	1																								T	
SFRP4	3.24e-012 (T)	1	T																								
IGFBP7	4.19e-012 (T)	1											T														
SFRP4	4.35e-012 (T)	1	T																								
COL8A1	<2.22e-016 (T)	1											T														
SRPX2	<2.22e-016 (T)	1																					T				
IRS1	3.55e-015 (T)	1																								T	
SPON2	3.77e-015 (T)	1																									X
FBXO32	3.77e-015 (T)	1																									X
CTHRC1	7.33e-015 (T)	1	T																								
ITGB1	8.44e-015 (T)	1										T															
COL5A1	1.24e-014 (T)	1												T													
ETV1	1.31e-014 (T)	1	T																								
DKK3	1.55e-014 (T)	1																									X
ITGB1	2.07e-014 (T)	1										T															
COL5A1	2.46e-014 (T)	1												T													
COL5A1	2.80e-014 (T)	1												T													
ITGB1	5.44e-014 (T)	1										T															
DKK3	6.39e-014 (T)	1																									X
HOXB7	1.30e-013 (T)	1			T																						
SQLE	1.50e-013 (T)	1											T														
SYTL4	2.01e-013 (T)	1																									X
FKBP7	2.47e-013 (T)	1											N														
HOXB7	3.36e-013 (T)	1			T																						
HTRA1	4.49e-013 (T)	1			T																						
LOC493869	5.66e-013 (T)	1														T											
INPP5F	7.38e-013 (T)	1				N																					
LAMB1	8.24e-013 (T)	1		T																							
YWHAH	9.02e-013 (T)	1																								T	
BZRP	9.92e-013 (T)	1																								T	
FBXO32	1.35e-012 (T)	1																									X
ENC1	1.49e-012 (T)	1			T																						

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
TRIO	1.50e-012 (T)	1											N														
ANKRD10	1.78e-012 (T)	1											T														
ARHGAP26	1.88e-012 (T)	1																T									
ASPN	1.90e-012 (T)	1	T																								
ASPN	4.38e-012 (T)	1	T																								
ZNF532	5.47e-012 (T)	1	T																								
ADAM12	5.48e-012 (T)	1	T																								
LAMB1	5.72e-012 (T)	1		T																							
MGP	6.10e-012 (T)	1																								T	
TNFAIP6	6.17e-012 (T)	1																								T	
F2R	6.62e-012 (T)	1							T																		
MITF	6.66e-012 (T)	1											T														
KYNU	6.86e-012 (T)	1																								T	
PPM2C	8.11e-012 (T)	1			N																						
HIP1	8.98e-012 (T)	1							N																		
	203	750	85	9	40	12	13	2	87	5	10	12	35	56	5	61	23	31	19	9	14	20	42	16	26	70	48

Supplementary Table 4. Comparison of the list of genes specifically over-expressed in tumor epithelia with other microarray studies of pancreatic cancer

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
S100P	8.90e-012 (T)	15	T	T	T			T	T				T			T		T	T	T	T	T	T	T		T	
FN1	2.22e-016 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	2.22e-016 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	2.22e-016 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
FN1	1.11e-015 (T)	12	T		T	T			T				T	T			T	T			T			T	T	T	
CAPG	0.00e+000 (T)	9					T	T	T					T		T		T		T		T	T				
SFN	1.31e-010 (T)	9			T				T					T				T	T			T	T			T	X
SFN	1.90e-010 (T)	9			T				T					T				T	T			T	T			T	X
CSPG2	1.33e-015 (T)	9	T		T				T		T			T		T	T								T	T	
CSPG2	8.22e-015 (T)	9	T		T				T		T			T		T	T								T	T	
CSPG2	1.13e-014 (T)	9	T		T				T		T			T		T	T								T	T	
COL1A2	2.71e-014 (T)	9	T		T				T	T				T		T	T								T		X
S100A11	2.22e-014 (T)	8			T				T						T			T	T			T	T			T	
THBS2	2.22e-016 (T)	8	T		T				T					T		T							T	N		T	
POSTN	7.35e-014 (T)	8	T				T		T					T		T				T						T	X
COL3A1	1.92e-012 (T)	8	T		T				T		T		T	T											T		X
LAMC2	2.94e-013 (T)	7	T						T							T		T	T			T	T				
KRT17	4.49e-013 (T)	7	T						T					T				T	T				T	N			
KRT19	7.63e-012 (T)	7					T		T								T	T	T			T	T				
KRT17	8.21e-010 (T)	7	T						T					T				T	T				T	N			
IGFBP3	2.61e-012 (T)	7				T			T									T	T			T	T				X
GPRC5A	1.29e-014 (T)	6			T								T			T		T		T							X
GPRC5A	1.91e-014 (T)	6			T								T			T		T		T							X
PLAU	2.39e-012 (T)	6	T													T					T		T			T	X
KRT7	3.11e-015 (T)	5	T				T			T						T			T								
FER1L3	2.04e-014 (T)	5	T												T	T		T					T				
COL10A1	0.00e+000 (T)	5	T						T													T	T				X
LGALS1	1.92e-011 (T)	4							T					T									T				X
QPCT	1.41e-010 (T)	4							T				T									T		T			
IFI27	7.04e-010 (T)	4			T				T				T						T								
S100A6	9.35e-012 (T)	4											T						T		T						X
COL5A2	5.77e-015 (T)	4	T						T																T	T	
GJB2	2.00e-014 (T)	4							T							T						T	T				
COL5A2	3.93e-014 (T)	4	T						T																T	T	
CXCL5	3.48e-012 (T)	3	T																				T				X
TMSB10	1.12e-010 (T)	3			T												T									T	

Gene Symbol	p value (T/N)	nr.of studies	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]
SULF1	0.00e+000 (T)	3	T										T			T											
COL11A1	2.22e-016 (T)	3	T						T							T											
COL11A1	5.11e-015 (T)	3	T						T							T											
SEMA3C	3.04e-011 (T)	3							T				T														X
C14orf78	1.67e-014 (T)	2																		T							X
FGD6	1.47e-013 (T)	2														T											X
SERPINA1	2.32e-012 (T)	2																						T		T	
UBD	2.61e-012 (T)	2	T													T											
SERPINA1	1.18e-011 (T)	2																						T		T	
G1P2	3.04e-011 (T)	2																T								N	
PFKP	2.14e-010 (T)	2							T										T								
GPX1	5.89e-010 (T)	2											T													T	
ANXA2	2.89e-015 (T)	2					T					T															
ANXA2	5.11e-015 (T)	2					T					T															
ANXA2	6.66e-015 (T)	2					T					T															
MAP4K4	1.00e-013 (T)	2														T							T				
DDEF2	2.68e-013 (T)	2											T					T									
S100A10	3.20e-011 (T)	2																T								T	
MXRA5	3.00e-014 (T)	2	T		T																						
HOXB7	1.30e-013 (T)	1			T																						
HOXB7	3.36e-013 (T)	1			T																						
CSTB	8.43e-013 (T)	1																								T	
ENO2	5.94e-011 (T)	1							T																		
HK2	2.21e-010 (T)	1												N													
BZRP	9.92e-013 (T)	1																								T	
CKAP1	2.24e-011 (T)	1																								T	
FXYD5	1.03e-010 (T)	1														T											
FXYD5	3.26e-010 (T)	1														T											
COL8A1	0.00e+000 (T)	1											T														
CTHRC1	7.33e-015 (T)	1	T																								
COL5A1	2.46e-014 (T)	1												T													
LEF1	1.91e-010 (T)	1	T																								
SULF2	4.03e-010 (T)	1											T														
	69	253	29	1	21	5	7	2	33	2	4	3	17	19	2	23	10	20	13	6	7	11	18	11	11	25	16

Microarray studies of PDAC referenced in Table 2 and Supplementary Tables 3 and 4

Index	Short reference	Reference
1	Van Heek CB04	Van Heek NT, Maitra A, Koopmann J, Fedarko N, Jain A, Rahman A, Iacobuzio-Donahue CA, Adsay V, Ashfaq R, Yeo CJ, Cameron JL, Offerhaus JA, Hruban RH, Berg KD, Goggins M. Gene expression profiling identifies markers of ampullary adenocarcinoma. Cancer Biol Ther. 2004 Jul;3(7):651-6. PMID: 15136764
2	Segara CCR05	Segara D, Biankin AV, Kench JG, Langusch CC, Dawson AC, Skalicky DA, Gotley DC, Coleman MJ, Sutherland RL, Henshall SM. Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia. Clin Cancer Res. 2005 May 1;11(9):3587-96. PMID: 15867264
3	Nakamura OG04	Nakamura T, Furukawa Y, Nakagawa H, Tsunoda T, Ohigashi H, Murata K, Ishikawa O, Ohgaki K, Kashimura N, Miyamoto M, Hirano S, Kondo S, Katoh H, Nakamura Y, Katagiri T. Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection.Oncogene. 2004 Mar 25;23(13):2385-400. PMID: 14767473
4	Maitra CCR03	Maitra A, Hansel DE, Argani P, Ashfaq R, Rahman A, Naji A, Deng S, Geradts J, Hawthorne L, House MG, Yeo CJ. Global expression analysis of well-differentiated pancreatic endocrine neoplasms using oligonucleotide microarrays. Clin Cancer Res. 2003 Dec 1;9(16 Pt 1):5988-95. PMID: 14676124
5	Lu P04	Lu Z, Hu L, Evers S, Chen J, Shen Y. Differential expression profiling of human pancreatic adenocarcinoma and healthy pancreatic tissue.Proteomics. 2004 Dec;4(12):3975-88. PMID: 15526344
6	Grützmann N04	Grützmann R, Pilarsky C, Ammerpohl O, Lüttges J, Böhme A, Sipos B, Foerder M, Alldinger I, Jahnke B, Schackert HK, Kalthoff H, Kremer B, Klöppel G, Saeger HD. Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays.Neoplasia. 2004 Sep-Oct;6(5):611-22. PMID: 15548371
7	Friess CMLS03	Friess H, Ding J, Kleeff J, Fenkell L, Rosinski JA, Guweidhi A, Reidhaar-Olson JF, Korc M, Hammer J, Büchler MW. Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer. Cell Mol Life Sci. 2003 Jun;60(6):1180-99. PMID: 12861384 (retrieved from Pancreatic Expression Database www.pancreasexpression.org)
8	Crnogorac-Jurcevic OG02	Crnogorac-Jurcevic T, Efthimiou E, Nielsen T, Loader J, Terris B, Stamp G, Baron A, Scarpa A, Lemoine NR. Expression profiling of microdissected pancreatic adenocarcinomas.Oncogene. 2002 Jul 4;21(29):4587-94. PMID: 12085237
9	Crnogorac-Jurcevic OG01	Crnogorac-Jurcevic T, Efthimiou E, Capelli P, Blaveri E, Baron A, Terris B, Jones M, Tyson K, Bassi C, Scarpa A, Lemoine NR. Gene expression profiles of pancreatic cancer and stromal desmoplasia.Oncogene. 2001 Nov 1;20(50):7437-46. PMID: 11704875
10	Crnogorac-Jurcevic GE05	Crnogorac-Jurcevic T, Gangeswaran R, Bhakta V, Capurso G, Lattimore S, Akada M, Sunamura M, Prime W, Campbell F, Brentnall TA, Costello E, Neoptolemos J, Lemoine NR. Proteomic analysis of chronic pancreatitis and pancreatic adenocarcinoma. Gastroenterology. 2005 Nov;129(5):1454-63. PMID: 16285947
11	Buchholz OG05	Buchholz M, Braun M, Heidenblut A, Kestler HA, Klöppel G, Schmiegel W, Hahn SA, Lüttges J, Gress TM. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions.Oncogene. 2005 Oct 6;24(44):6626-36. PMID: 16103885
12	Friess AS01	Friess H, Ding J, Kleeff J, Liao Q, Berberat PO, Hammer J, Büchler MW. Identification of disease-specific genes in chronic pancreatitis using DNA array technology. Ann Surg. 2001 Dec;234(6):769-78; discussion 778-9. PMID: 11729383
13	Han CR02	Han H, Bearss DJ, Browne LW, Calaluce R, Nagle RB, Von Hoff DD. Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. Cancer Res. 2002 May 15;62(10):2890-6. Erratum in: Cancer Res 2002 Aug;62(15):4532. PMID: 12019169
14	Iacobutio-Donahue AJP02	Iacobuzio-Donahue CA, Maitra A, Shen-Ong GL, van Heek T, Ashfaq R, Meyer R, Walter K, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH. Discovery of novel tumor markers of pancreatic cancer using global gene expression technology. Am J Pathol. 2002 Apr;160(4):1239-49. PMID: 11943709
15	Tan WJG03	Tan ZJ, Hu XG, Cao GS, Tang Y. Analysis of gene expression profile of pancreatic carcinoma using cDNA microarray.World J Gastroenterol. 2003 Apr;9(4):818-23. PMID: 12679940
16	Iacobutio-Donahue AJP03	Iacobuzio-Donahue CA, Maitra A, Olsen M, Lowe AW, van Heek NT, Rosty C, Walter K, Sato N, Parker A, Ashfaq R, Jaffee E, Ryu B, Jones J, Eshleman JR, Yeo CJ, Cameron JL, Kern SE, Hruban RH, Brown PO, Goggins M. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. Am J Pathol. 2003 Apr;162(4):1151-62. PMID: 12651607
17	Logsdon CR03	Logsdon CD, Simeone DM, Binkley C, Arumugam T, Greenson JK, Giordano TJ, Misek DE, Kuick R, Hanash S. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. Cancer Res. 2003 May 15;63(10):2649-57. Erratum in: Cancer Res. 2003 Jun 15;63(12):3445. PMID: 12750293
18	Grützmann VA03	Grützmann R, Foerder M, Alldinger I, Staub E, Brümmendorf T, Röpcke S, Li X, Kristiansen G, Jesnowski R, Sipos B, Löhr M, Lüttges J, Ockert D, Klöppel G, Saeger HD, Pilarsky C. Gene expression profiles of microdissected pancreatic ductal adenocarcinoma. Virchows Arch. 2003

		Oct;443(4):508-17. PMID: 12942322
19	Cmogorac-Jurcevic JP03	Crnogorac-Jurcevic T, Missiaglia E, Blaveri E, Gangeswaran R, Jones M, Terris B, Costello E, Neoptolemos JP, Lemoine NR. Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent.J Pathol. 2003 Sep;201(1):63-74. PMID: 12950018
20	Friess CMLS03	Friess H, Ding J, Kleeff J, Fenkell L, Rosinski JA, Guweidhi A, Reidhaar-Olson JF, Korc M, Hammer J, Büchler MW. Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer. Cell Mol Life Sci. 2003 Jun;60(6):1180-99. PMID: 12861384 (retrieved from metaanalysis by Brandt et al.}
21	Iacobuzio-Donahue CR03	Iacobuzio-Donahue CA, Ashfaq R, Maitra A, Adsay NV, Shen-Ong GL, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH. Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. Cancer Res. 2003 Dec 15;63(24):8614-22. PMID: 14695172
22	Terris AJP02	Terris B, Blaveri E, Crnogorac-Jurcevic T, Jones M, Missiaglia E, Ruszniewski P, Sauvanet A, Lemoine NR. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol. 2002 May;160(5):1745-54. PMID: 12000726
23	Jin WJG05	Jin G, Hu XG, Ying K, Tang Y, Liu R, Zhang YJ, Jing ZP, Xie Y, Mao YM. Discovery and analysis of pancreatic adenocarcinoma genes using cDNA microarrays.World J Gastroenterol. 2005 Nov 7;11(41):6543-8. PMID: 16425432
24	metaanalysis Grützmann 05	Grützmann R, Boriss H, Ammerpohl O, Lüttges J, Kalthoff H, Schackert HK, Klöppel G, Saeger HD, Pilarsky C. Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes.Oncogene. 2005 Jul 28;24(32):5079-88. PMID: 15897887
25	Hustinx CBT04	Hustinx SR, Cao D, Maitra A, Sato N, Martin ST, Sudhir D, Iacobuzio-Donahue C, Cameron JL, Yeo CJ, Kern SE, Goggins M, Mollenhauer J, Pandey A, Hruban RH. Differentially expressed genes in pancreatic ductal adenocarcinomas identified through serial analysis of gene expression. Cancer Biol Ther. 2004 Dec;3(12):1254-61. PMID: 15477757

Supplementary Table 5. Comparison of the RT-PCR measurements with the microarray measurements (correlation and p-values computed on $\log_2$-transformed data for a few genes only for a subset of samples)

Gene	Pearson correlation (RT-PCR, microarray)	p-value(RT-PCR)	p-value(microarray)
BGN	0.931283	1.85E-09	1.05E-11
BHLHB2	0.948764	6.47E-08	7.68E-09
CAPG	0.905623	5.96E-09	1.77E-12
CSPG2	0.977309	3.23E-11	1.47E-11
CTHRC1	0.932596	2.60E-09	2.56E-11
DKK3	0.769944	2.70E-09	6.47E-11
ETV1	0.868632	4.35E-08	6.34E-12
FN1	0.976707	9.48E-11	2.99E-12
HNT	0.958995	2.12E-09	8.24E-13
HOP	0.965494	8.59E-09	6.70E-09
IGFBP5	0.962842	6.53E-10	1.49E-09
INHBA	0.985245	1.56E-10	4.16E-12
LTBP1	0.859544	1.27E-06	3.45E-12
NOX4	0.957538	5.49E-11	3.69E-14
PDGFC	0.871077	9.37E-07	3.38E-11
PLAT	0.952203	2.47E-06	6.91E-08
PLAU	0.962242	1.05E-10	1.74E-10
SOX4	0.902353	2.69E-08	2.47E-12
SULF1	0.980882	1.72E-12	9.77E-14
WISP1	0.931112	1.96E-08	1.50E-10

Supplementary Table 6. The anomalous “normal” samples

Sample	**Histopatologic analysis**
N30308	Chronic pancreatitis, significant fibrosis, 10% tumor contamination
N40726	Chronic pancreatitis, very much fibrotic tissue
N51294	Perhaps increased fibrosis
N40875	More fibrotic tissue than normal pancreatic tissue
N40892	Reduced normal pancreatic parenchyma (fasciae, fibrotic tissue, vascularization)

Supplementary Table 7. A. Gene-set enrichment of the genes overexpressed in whole-tissue PDAC samples (ICF dataset). **B.** Enrichment in Gene Ontology biological process annotations of the genes overexpressed in the ICF data. **C.** Mentions in the literature of the genes from Table 3 in connection with pancreatic cancer or other cancers (only one reference per gene is shown).

A.

MSigDB gene set	**Description**	**p-value**
Tgfbeta_all_up	Upregulated by TGF-beta treatment of skin fibroblasts, at any timepoint	$4.31 \cdot 10^{-29}$
Tgfbeta_early_up	Upregulated by TGF-beta treatment of skin fibroblasts at 30 min (clusters 1-3)	$3.18 \cdot 10^{-27}$
adip_vs_fibro_dn	Downregulated following 7-day differentiation of murine 3T3-L1 fibroblasts into adipocytes	$4.64 \cdot 10^{-22}$
vegf_hmmec_6hrs_up	Up-regulated at 6hrs follwing VEGF treatment of human myometrial microvascular endothelial cells	$2.04 \cdot 10^{-15}$
Esr_fibroblast_up	Up-regulated in the environmental stress response in human fibroblasts (regulated similarly by gamma and UV rediation and 4-NQO)	$4.26 \cdot 10^{-9}$
emt_up	Up-regulated during the TGFbeta-induced epithelial-to-mesenchymal transition (EMT) of Ras-transformed mouse mammary epithelial (EpH4) cells (EMT is representative of late-stage tumor progression and metastasis)	$1.76 \cdot 10^{-5}$

B.

GO Biological Process	**p-value**
cell adhesion	$8.02 \cdot 10^{-23}$
response to wounding	$5.08 \cdot 10^{-8}$
integrin-mediated signaling pathway	$5.80 \cdot 10^{-8}$
transforming growth factor beta receptor signaling pathway	$3.28 \cdot 10^{-7}$
cell migration	$3.45 \cdot 10^{-6}$
focal adhesion formation	$4.68 \cdot 10^{-6}$
inflammatory response	$7.16 \cdot 10^{-6}$
defense response	$2.61 \cdot 10^{-5}$

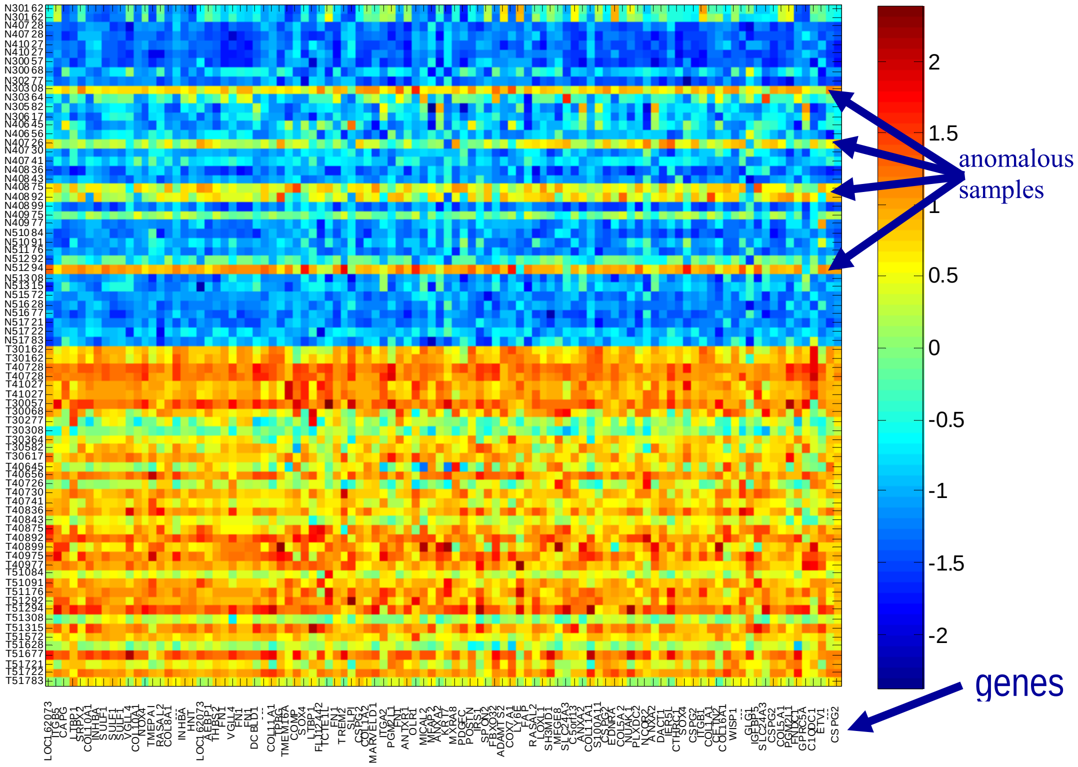
C.

Gene	**pancreatic cancer**	**other cancers**
TE_fc_epithelial, TE_p_epithelial		
CAPG	PMID:16847067	
KRT7	PMID:16362976	
GPRC5A	PMID:15659406	
C14orf78		

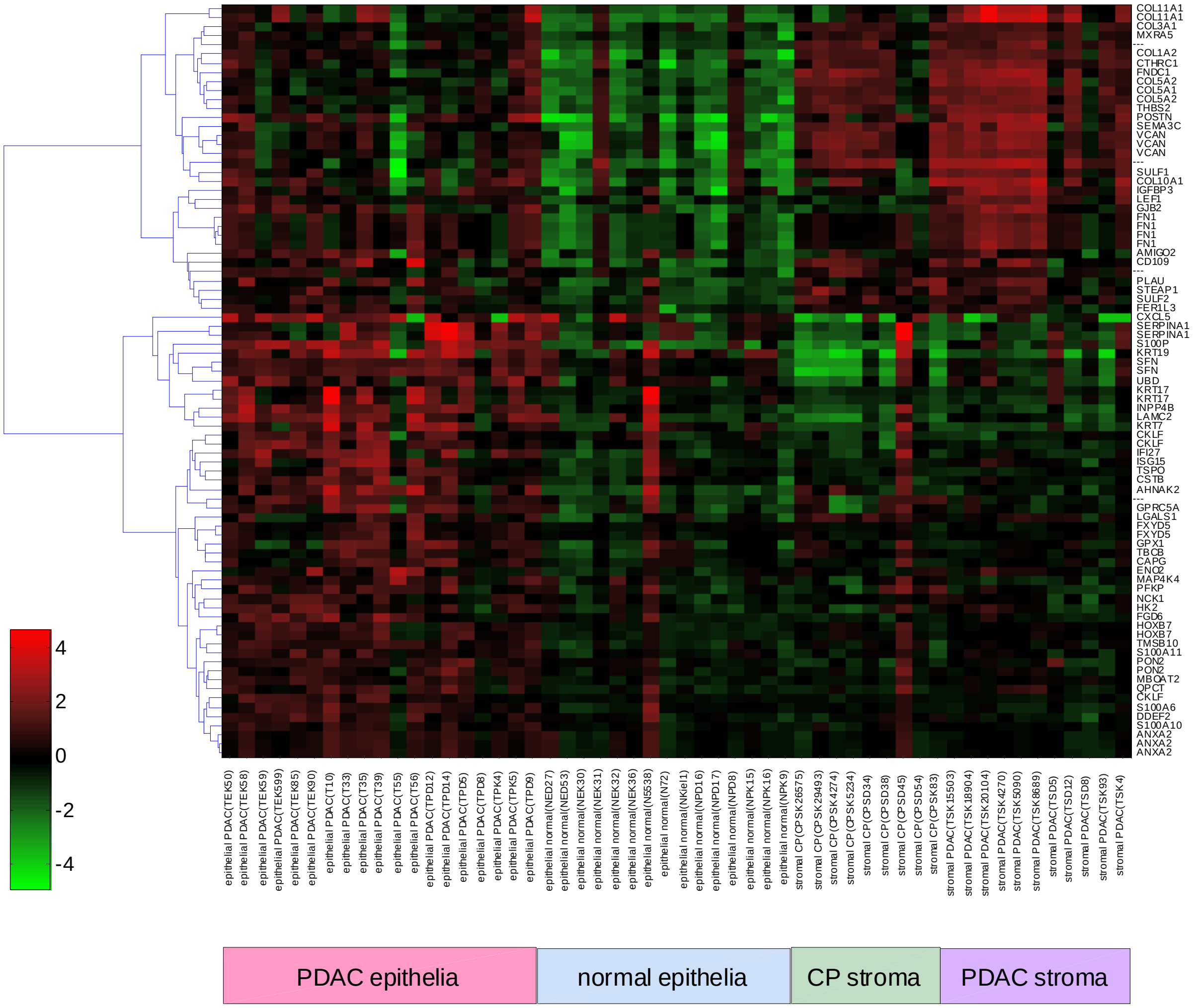
Gene	**pancreatic cancer**	**other cancers**
FER1L3		
S100A11	PMID:17000675	
AMIGO2		PMID:15107827
HOXB7		PMID:17018609
FGD6		
LAMC2	PMID:15922869	
KRT17	PMID:7686885	
CSTB	PMID:3594445	
SERPINA1	PMID:18218118	
UBD		PMID:12730673
CXCL5		PMID:17479287
KRT19	PMID:17226905	
S100P	PMID:16061848	
LGALS1	PMID:16038798	
INPP4B		
G1P2		PMID:17097911
ENO2	PMID:8950756	
TMSB10	PMID:15983444	
SFN	PMID:15073049	
QPCT		
PFKP		PMID:2973647
HK2	PMID:17530485	
CKLF		
GPX1		PMID:17465855
IFI27		PMID:15086558
ANXA2	PMID:8269629	
MAP4K4		PMID:16537454
DDEF2		
OACT2		
BZRP		PMID:17126818
S100A6	PMID:15833853	
CKAP1		
S100A10	PMID:12962548	
NCK1		
FXYD5	PMID:12586803	
PON2		

TE_fc_stromal		
COL10A1		
SULF1	PMID:17460759	
COL8A1		
COL11A1		PMID:11375892
THBS2		PMID:9012463
FN1	PMID:8286197	
CSPG2		PMID:17065588
COL5A2		PMID:11375892
CTHRC1	PMID:16778098	
FNDC1		
GJB2	PMID:15502644	
COL5A1		
COL1A2		
MXRA5		
POSTN	PMID:17408641	
STEAP1		
COL3A1		
PLAU	PMID:15937032	
IGFBP3		
CD109		PMID:17922683
SEMA3C		
LEF1		PMID:16142310
ANTXR2		
SULF2	PMID:17460759	

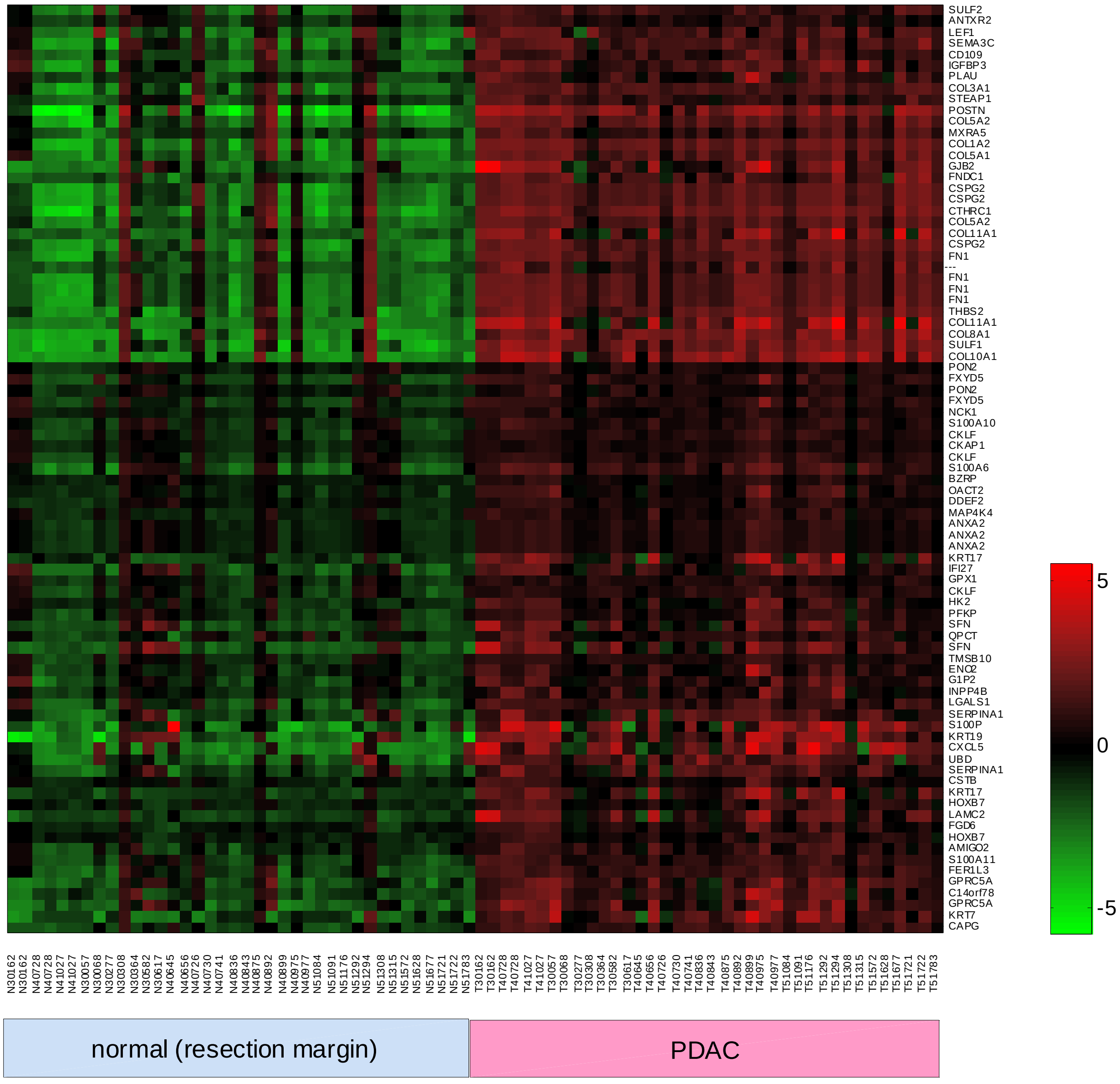
Supplementary Figure 2. Gene expression profiles of the best 100 differentially expressed genes showing the anomalous samples



Supplementary Figure 3. Clustergram of the genes over-expressed in tumor epithelia (centred $\log_2$-expression values in the GP study) (clustergram_1D_Pilarsky_TE)

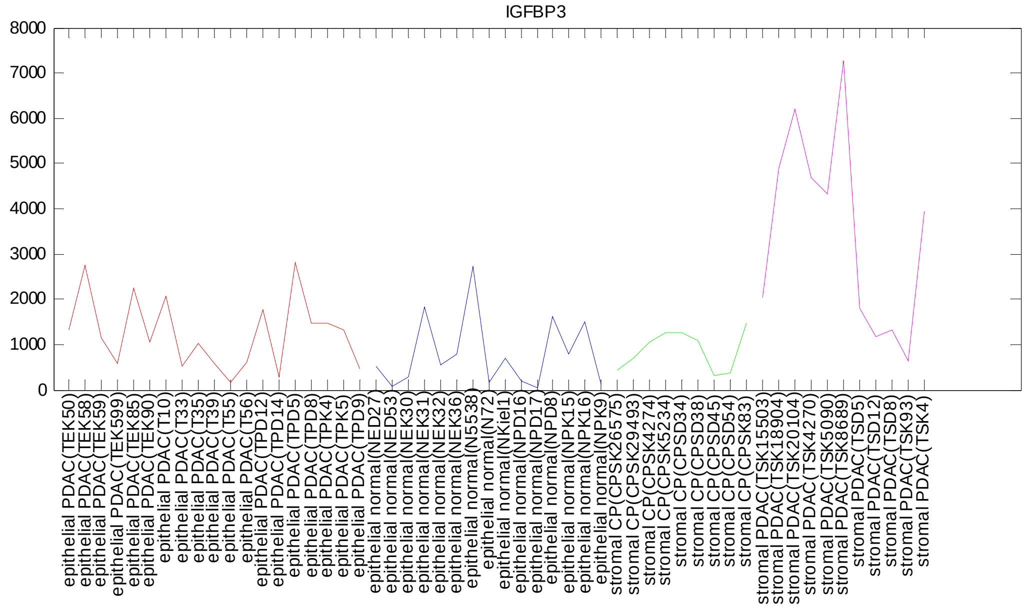


Supplementary Figure 4. Centred $\log_2$-expression values in our whole-tissue study (ICF) of the genes overexpressed in tumor epithelia. Note the anomalous "normal samples" (red columns in the green area) (image_ICF_Pilarsky_TE)

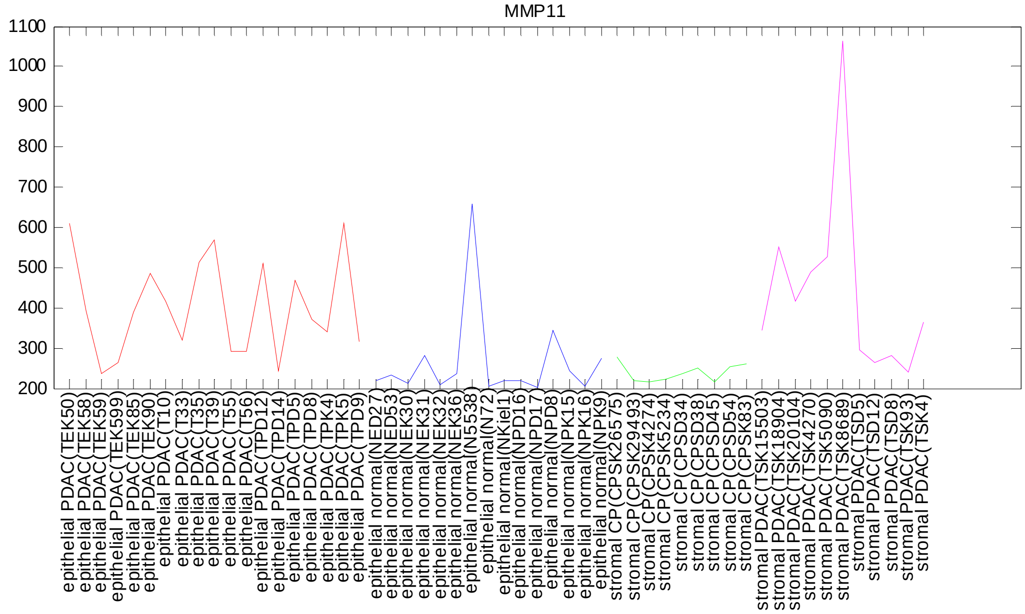


Supplementary Figure 5. (A) IGFBP3 was selected in the Logsdon et al. study based on its overexpression in whole tissue tumor w.r.t. both normal and CP tissues. The expression values from the GP microdissection study depicted in the Figure show that IGFBP3 is actually overexpressed in tumor stroma (rather than tumor epithelia). Color codes: tumor epithelia: red, normal epithelia: blue, chronic pancreatitis stroma: green, tumor stroma: pink. (B) In the GP data, MMP11 is increased in both tumor stroma and epithelia, not just in the epithelia, as inferred in the Logsdon study.

(A)



(B)



References

1 Van Heek NT, Maitra A, Koopmann J, Fedarko N, Jain A, Rahman A, Iacobuzio-Donahue CA, Adsay V, Ashfaq R, Yeo CJ, Cameron JL, Offerhaus JA, Hruban RH, Berg KD, Goggins M. Gene expression profiling identifies markers of ampullary adenocarcinoma. Cancer Biol Ther. 2004 Jul;3(7):651-6. PMID: 15136764

2 Segara D, Biankin AV, Kench JG, Langusch CC, Dawson AC, Skalicky DA, Gotley DC, Coleman MJ, Sutherland RL, Henshall SM. Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia. Clin Cancer Res. 2005 May 1;11(9):3587-96. PMID: 15867264

3 Nakamura T, Furukawa Y, Nakagawa H, Tsunoda T, Ohigashi H, Murata K, Ishikawa O, Ohgaki K, Kashimura N, Miyamoto M, Hirano S, Kondo S, Katoh H, Nakamura Y, Katagiri T. Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection. Oncogene. 2004 Mar 25;23(13):2385-400. PMID: 14767473

4 Maitra A, Hansel DE, Argani P, Ashfaq R, Rahman A, Naji A, Deng S, Geradts J, Hawthorne L, House MG, Yeo CJ. Global expression analysis of well-differentiated pancreatic endocrine neoplasms using oligonucleotide microarrays. Clin Cancer Res. 2003 Dec 1;9(16 Pt 1):5988-95. PMID: 14676124

5 Lu Z, Hu L, Evers S, Chen J, Shen Y. Differential expression profiling of human pancreatic adenocarcinoma and healthy pancreatic tissue.Proteomics. 2004 Dec;4(12):3975-88. PMID: 15526344

6 Grützmann R, Pilarsky C, Ammerpohl O, Lüttges J, Böhme A, Sipos B, Foerder M, Alldinger I, Jahnke B, Schackert HK, Kalthoff H, Kremer B, Klöppel G, Saeger HD. Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays.Neoplasia. 2004 Sep-Oct;6(5):611-22. PMID: 15548371

7 Friess H, Ding J, Kleeff J, Fenkell L, Rosinski JA, Guweidhi A, Reidhaar-Olson JF, Korc M, Hammer J, Büchler MW. Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer. Cell Mol Life Sci. 2003 Jun;60(6):1180-99. PMID: 12861384 (corresponding gene lists retrieved from Pancreatic Expression Database www.pancreasexpression.org)

8 Crnogorac-Jurcevic T, Efthimiou E, Nielsen T, Loader J, Terris B, Stamp G, Baron A, Scarpa A, Lemoine NR. Expression profiling of microdissected pancreatic adenocarcinomas.Oncogene. 2002 Jul 4;21(29):4587-94. PMID: 12085237

9 Crnogorac-Jurcevic T, Efthimiou E, Capelli P, Blaveri E, Baron A, Terris B, Jones M, Tyson K, Bassi C, Scarpa A, Lemoine NR. Gene expression profiles of pancreatic cancer and stromal desmoplasia.Oncogene. 2001 Nov 1;20(50):7437-46. PMID: 11704875

10 Crnogorac-Jurcevic T, Gangeswaran R, Bhakta V, Capurso G, Lattimore S, Akada M, Sunamura M, Prime W, Campbell F, Brentnall TA, Costello E, Neoptolemos J, Lemoine NR. Proteomic analysis of chronic pancreatitis and pancreatic adenocarcinoma. Gastroenterology. 2005 Nov;129(5):1454-63. PMID: 16285947

11 Buchholz M, Braun M, Heidenblut A, Kestler HA, Klöppel G, Schmiegel W, Hahn SA, Lüttges J, Gress TM. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions.Oncogene. 2005 Oct 6;24(44):6626-36. PMID: 16103885

12 Friess H, Ding J, Kleeff J, Liao Q, Berberat PO, Hammer J, Büchler MW. Identification of disease-specific genes in chronic pancreatitis using DNA array technology. Ann Surg. 2001 Dec;234(6):769-78; discussion 778-9. PMID: 11729383

13 Han H, Bearss DJ, Browne LW, Calaluce R, Nagle RB, Von Hoff DD. Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. Cancer Res. 2002 May 15;62(10):2890-6. Erratum in: Cancer Res 2002 Aug;62(15):4532. PMID: 12019169

14 Iacobuzio-Donahue CA, Maitra A, Shen-Ong GL, van Heek T, Ashfaq R, Meyer R, Walter K, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH. Discovery of novel tumor markers of pancreatic cancer using global gene expression technology. Am J Pathol. 2002 Apr;160(4):1239-49. PMID: 11943709

15 Tan ZJ, Hu XG, Cao GS, Tang Y. Analysis of gene expression profile of pancreatic carcinoma using cDNA microarray.World J Gastroenterol. 2003 Apr;9(4):818-23. PMID: 12679940

16 Iacobuzio-Donahue CA, Maitra A, Olsen M, Lowe AW, van Heek NT, Rosty C, Walter K, Sato N, Parker A, Ashfaq R, Jaffee E, Ryu B, Jones J, Eshleman JR, Yeo CJ, Cameron JL, Kern SE, Hruban RH, Brown PO, Goggins M. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. Am J Pathol. 2003 Apr;162(4):1151-62. PMID: 12651607

17 Logsdon CD, Simeone DM, Binkley C, Arumugam T, Greenson JK, Giordano TJ, Misek DE, Kuick R, Hanash S. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. Cancer Res. 2003 May 15;63(10):2649-57. Erratum in: Cancer Res. 2003 Jun 15;63(12):3445. PMID: 12750293

18 Grützmann R, Foerder M, Alldinger I, Staub E, Brümmendorf T, Röpcke S, Li X, Kristiansen G, Jesnowski R, Sipos B, Löhr M, Lüttges J, Ockert D, Klöppel G, Saeger HD, Pilarsky C. Gene expression profiles of microdissected pancreatic ductal adenocarcinoma. Virchows Arch. 2003 Oct;443(4):508-17. PMID: 12942322

19 Crnogorac-Jurcevic T, Missiaglia E, Blaveri E, Gangeswaran R, Jones M, Terris B, Costello E, Neoptolemos JP, Lemoine NR. Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent. J Pathol. 2003 Sep;201(1):63-74. PMID: 12950018

20 Friess H, Ding J, Kleeff J, Fenkell L, Rosinski JA, Guweidhi A, Reidhaar-Olson JF, Korc M, Hammer J, Büchler MW. Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer.

Cell Mol Life Sci. 2003 Jun;60(6):1180-99. PMID: 12861384 (corresponding gene lists retrieved from metaanalysis by Brandt et al. [68])

21 Iacobuzio-Donahue CA, Ashfaq R, Maitra A, Adsay NV, Shen-Ong GL, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH. Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. Cancer Res. 2003 Dec 15;63(24):8614-22. PMID: 14695172

22 Terris B, Blaveri E, Crnogorac-Jurcevic T, Jones M, Missiaglia E, Ruszniewski P, Sauvanet A, Lemoine NR. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol. 2002 May;160(5):1745-54. PMID: 12000726

23 Jin G, Hu XG, Ying K, Tang Y, Liu R, Zhang YJ, Jing ZP, Xie Y, Mao YM. Discovery and analysis of pancreatic adenocarcinoma genes using cDNA microarrays.World J Gastroenterol. 2005 Nov 7;11(41):6543-8. PMID: 16425432

24 Grützmann R, Boriss H, Ammerpohl O, Lüttges J, Kalthoff H, Schackert HK, Klöppel G, Saeger HD, Pilarsky C. Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes.Oncogene. 2005 Jul 28;24(32):5079-88. PMID: 15897887

25 Hustinx SR, Cao D, Maitra A, Sato N, Martin ST, Sudhir D, Iacobuzio-Donahue C, Cameron JL, Yeo CJ, Kern SE, Goggins M, Mollenhauer J, Pandey A, Hruban RH. Differentially expressed genes in pancreatic ductal adenocarcinomas identified through serial analysis of gene expression. Cancer Biol Ther. 2004 Dec;3(12):1254-61. PMID: 15477757

26 Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA. Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev. 2006 May 15;20(10):1218-49. PMID: 16702400

27 Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, Bruns CJ, Heeschen C. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell. 2007 Sep 13;1(3):313-23. PMID: 18371365

28 Bashyam MD et al. Array-based comparative genomic hybridization identifies localized DNA amplifications and homozygous deletions in pancreatic cancer. Neoplasia. 2005 Jun;7(6):556-62. PMID: 16036106

29 Heidenblad M et al. Genome-wide array-based comparative genomic hybridization reveals multiple amplification targets and novel homozygous deletions in pancreatic carcinoma cell lines. Cancer Res. 2004 64(9):3052-9. PMID: 15126341

30 Heidenblad M et al. Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: implica-tions for the interpretation of genomic amplifications. Oncogene. 2005 Mar 3;24(10):1794-801. PMID: 15688027

31 Chu GC, Kimmelman AC, Hezel AF, DePinho RA. Stromal biology of pancreatic cancer. J Cell Biochem. 2007 Jul 1;101(4):887-907. PMID: 17266048; Mahadevan D, Von Hoff DD. Tumor-stroma interactions in pancreatic ductal adenocarcinoma. Mol Cancer Ther. 2007 Apr;6(4):1186-97. PMID: 17406031

32 Mahadevan D, Von Hoff DD. Tumor-stroma interactions in pancreatic ductal adenocarcinoma. Mol Cancer Ther. 2007 Apr;6(4):1186-97. PMID: 17406031.

33 Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003 Feb 15;31(4):e15. PMID: 12582260

34 L2L. L2L Microarray Analysis Tool http://depts.washington.edu/l2l/

35 TRED. Transcriptional Regulatory Element Database. http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=home.

36 Grützmann R, Pilarsky C, Ammerpohl O, Lüttges J, Böhme A, Sipos B, Foerder M, Alldinger I, Jahnke B, Schackert HK, Kalthoff H, Kremer B, Klöppel G, Saeger HD. Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays. Neoplasia. 2004 Sep-Oct;6(5):611-22. PMID: 15548371

37 Pilarsky C, Ammerpohl O, Sipos B, Dahl E, Hartmann A, Wellmann A, Braunschweig T, Löhr M, Jesnowski R, Friess H, Wente MN, Kristiansen G, Jahnke B, Denz A, Rückert F, Schackert HK, Klöppel G, Kalthoff H, Saeger HD, Grützmann R. Activation of Wnt signalling in stroma from pancreatic cancer identified by gene expression profiling. J Cell Mol Med. 2008 Feb 24. PMID: 18298655.

38 Truty MJ, Urrutia R. Basics of TGF-beta and pancreatic cancer. Pancreatology. 2007; 7(5-6):423-35. PMID: 17898532.

39 Kleeff J, Ishiwata T, Friess H, Büchler MW, Korc M. Concomitant over-expression of activin/inhibin beta subunits and their receptors in human pancreatic cancer.Int J Cancer. 1998 Sep 11;77(6):860-8. PMID: 9714055

40 Jenkins PJ, Khalaf S, Ogunkolade W, McCarthy K, David T, Hands RE, Davies D, Bustin SA. Differential expression of IGF-binding protein-3 in normal and malignant colon and its influence on apoptosis. Endocr Relat Cancer. 2005 Dec;12(4):891-901. PMID: 16322329

41 Katayama M, Sanzen N, Funakoshi A, Sekiguchi K. Laminin gamma2-chain fragment in the circulation: a prognostic indicator of epithelial tumor invasion. Cancer Res. 2003 Jan 1;63(1):222-9. PMID: 12517801

42 Wu X, Chen H, Parker B, Rubin E, Zhu T, Lee JS, Argani P, Sukumar S. HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial-mesenchymal transition.Cancer Res. 2006 Oct 1;66(19):9527-34. PMID: 17018609

43 Collins CS, Hong J, Sapinoso L, Zhou Y, Liu Z, Micklash K, Schultz PG, Hampton GM. A small interfering RNA screen for modulators of tumor cell motility identifies MAP4K4 as a promigratory kinase. Proc Natl Acad Sci U S A. 2006 Mar 7;103(10):3775-80. PMID: 16537454

44 Nawshad A, Hay ED. TGFbeta3 signaling activates transcription of the LEF1 gene to induce epithelial mesenchymal transformation during mouse palate development. J Cell Biol. 2003 Dec 22;163(6):1291-301. PMID: 14691138

45 Thompson CC, Ashcroft FJ, Patel S, Saraga G, Vimalachandran D, Prime W, Campbell F, Dodson A, Jenkins RE, Lemoine NR, Crnogorac-Jurcevic T, Yin HL, Costello E. Pancreatic cancer cells overexpress gelsolin family-capping proteins, which contribute to their cell motility. Gut. 2007 Jan;56(1):95-106. PMID: 16847067

46 Wu Q, Ding W, Mirza A, Van Arsdale T, Wei I, Bishop WR, Basso A, McClanahan T, Luo L, Kirschmeier P, Gustafson E, Hernandez M, Liu S. Integrative genomics revealed RAI3 is a cell growth-promoting gene and a novel P53 transcriptional target. J Biol Chem. 2005 Apr 1;280(13):12935-43. PMID: 15659406

47 Arumugam T, Simeone DM, Van Golen K, Logsdon CD. S100P promotes pancreatic cancer growth, survival, and invasion. Clin Cancer Res. 2005 Aug 1;11(15):5356-64. PMID: 16061848

48 Sakaguchi M, Sonegawa H, Murata H, Kitazoe M, Futami J, Kataoka K, Yamada H, Huh NH. S100A11, an Dual Mediator for Growth Regulation of Human Keratinocytes.Mol Biol Cell. 2008 Jan;19(1):78-85. PMID: 17978094

49 Kwon M, MacLeod TJ, Zhang Y, Waisman DM. S100A10, annexin A2, and annexin a2 heterotetramer as candidate plasminogen receptors. ront Biosci. 2005 Jan 1;10:300-25. Print 2005 Jan 1. Review. PMID: 15574370

50 Benaud C, Gentil BJ, Assard N, Court M, Garin J, Delphin C, Baudier J. AHNAK interaction with the annexin 2/S100A10 complex regulates cell membrane cytoarchitecture. J Cell Biol. 2004 Jan 5;164(1):133-44. PMID: 14699089

51 Vimalachandran D, Greenhalf W, Thompson C, Lüttges J, Prime W, Campbell F, Dodson A, Watson R, Crnogorac-Jurcevic T, Lemoine N, Neoptolemos J, Costello E. High nuclear S100A6 (Calcyclin) is significantly associated with poor survival in pancreatic cancer patients.Cancer Res. 2005 Apr 15;65(8):3218-25. PMID: 15833853

52 Dimberg J, Dienus O, Löfgren S, Hugander A, Wågsäter D. Expression and gene polymorphisms of the chemokine CXCL5 in colorectal cancer patients. Int J Oncol. 2007 Jul;31(1):97-102. PMID: 17549409

53 Park JY, Park KH, Bang S, Kim MH, Lee JE, Gang J, Koh SS, Song SY. CXCL5 overexpression is associated with late stage gastric cancer. J Cancer Res Clin Oncol. 2007 Nov;133(11):835-40. PMID: 17479287

54 Fitzner B, Walzel H, Sparmann G, Emmrich J, Liebe S, Jaster R. Galectin-1 is an inductor of pancreatic stellate cell activation. Cell Signal. 2005 Oct;17(10):1240-7. PMID: 16038798

55 Jung EJ, Moon HG, Cho BI, Jeong CY, Joo YT, Lee YJ, Hong SC, Choi SK, Ha WS, Kim JW, Lee CW, Lee JS, Park ST. Galectin-1 expression in cancer-associated stromal cells correlates tumor invasiveness and tumor progression in breast cancer. Int J Cancer. 2007 Jun 1;120(11):2331-8. PMID: 17304502

56 Shimamura T, Sakamoto M, Ino Y, Sato Y, Shimada K, Kosuge T, Sekihara H, Hirohashi S. Dysadherin overexpression in pancreatic ductal adenocarcinoma reflects tumor aggressiveness: relationship to e-cadherin expression. J Clin Oncol. 2003 Feb 15;21(4):659-67. PMID: 12586803

57 Shimamura T, Yasuda J, Ino Y, Gotoh M, Tsuchiya A, Nakajima A, Sakamoto M, Kanai Y, Hirohashi S. Dysadherin expression facilitates cell motility and metastatic potential of human pancreatic cancer cells. Cancer Res. 2004 Oct 1;64(19):6989-95. PMID: 15466191

58 Fischer H, Stenling R, Rubio C, Lindblom A. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. Carcinogenesis. 2001 Jun;22(6):875-8. PMID: 11375892

59 Neuschwander-Tetri BA, Talkad V, Otis Stephen F. Induced thrombospondin expression in the mouse pancreas during pancreatic injury. Int J Biochem Cell Biol. 2006 Jan;38(1):102-9. PMID: 16181801

60 Erkan M, Kleeff J, Gorbachevski A, Reiser C, Mitkus T, Esposito I, Giese T, Büchler MW, Giese NA, Friess H. Periostin creates a tumor-supportive microenvironment in the pancreas by sustaining fibrogenic stellate cell activity.Gastroenterology. 2007 Apr;132(4):1447-64. PMID: 17408641

61 Baril P, Gangeswaran R, Mahon PC, Caulee K, Kocher HM, Harada T, Zhu M, Kalthoff H, Crnogorac-Jurcevic T, Lemoine NR. Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the beta4 integrin and the PI3k pathway. Oncogene. 2007 Mar 29;26(14):2082-94. PMID: 17043657

62 Yan W, Shao R. Transduction of a Mesenchyme-specific Gene Periostin into 293T Cells Induces Cell Invasive Activity through Epithelial-Mesenchymal Transformation. J. Biol. Chem., Vol. 281, Issue 28, 19700-19708, July 14, 2006, PMID: 16702213

63 Bao S, Ouyang G, Bai X, Huang Z, Ma C, Liu M, Shao R, Anderson RM, Rich JN, Wang XF. Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. Cancer Cell. 2004 Apr;5(4):329-39. PMID: 15093540

64 Abiatari I, Kleeff J, Li J, Felix K, Büchler MW, Friess H. Hsulf-1 regulates growth and invasion of pancreatic cancer cells. J Clin Pathol. 2006 Oct;59(10):1052-8. PMID: 16603650

65 Bardeesy N, Cheng KH, Berger JH, Chu GC, Pahler J, Olson P, Hezel AF, Horner J, Lauwers GY, Hanahan D, DePinho RA. Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. Genes Dev. 2006 Nov 15;20(22):3130-46. PMID: 17114584

66 Ijichi H, Chytil A, Gorska AE, Aakre ME, Fujitani Y, Fujitani S, Wright CV, Moses HL. Aggressive pancreatic ductal adenocarcinoma in mice caused by pancreas-specific blockade of transforming growth factor-beta signaling in cooperation with active Kras expression. Genes Dev. 2006 Nov 15;20(22):3147-60. PMID: 17114585

67 Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genet. 25:25-29, 2000. PMID: 10802651

68 Brandt R, Grützmann R, Bauer A, Jesnowski R, Ringel J, Löhr M, Pilarsky C, Hoheisel JD. DNA microarray analysis of pancreatic malignancies.Pancreatology. 2004;4(6):587-97. PMID: 15557762