Pancreatology 15 (2015) 423–431

Contents lists available at ScienceDirect

Pancreatology

journal homepage: www.elsevier.com/locate/pan

Original article

Stratification of pancreatic tissue samples for molecular studies: RNA-based cellular annotation procedure

Anette Heller [a, *], Matthias M. Gaida [b], D. Männle [a], Thomas Giese [c], Aldo Scarpa [d], John P. Neoptolemos [e], Thilo Hackert [a], Oliver Strobel [a], Jörg D. Hoheisel [f], Nathalia A. Giese [a, 1], Andrea S. Bauer [f, 1]

[a] Department of Surgery, University Hospital Heidelberg, Heidelberg, Germany
[b] Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
[c] Institute of Immunology, University Hospital Heidelberg, Heidelberg, Germany
[d] Department of Pathology and ARC-NET Research Centre, University of Verona, Verona, Italy
[e] National Institute for Health Research, Pancreas Biomedical Research Unit and the Liverpool Experimental Cancer Medicine Centre, Liverpool, UK
[f] Functional Genome Analysis, German Research Cancer Center, Heidelberg, Germany

ARTICLE INFO

Article history:
Available online 15 June 2015

Keywords:
Pancreatic cancer
Tissue sample quality
Cellular annotation
qRT-PCR
Microarray analysis
Survival analysis

ABSTRACT

Background/objectives: Meaningful profiling of pancreatic cancer samples is particularly challenging due to their complex cellular composition. Beyond tumor cells, surgical biopsies contain desmoplastic stroma with infiltrating inflammatory cells, adjacent normal parenchyma, and "non-pancreatic tissues". The risk of misinterpretation rises when the heterogeneous cancer tissues are sub-divided into smaller fragments for multiple analytic procedures. Pre-analytic histological evaluation is the best option to characterize pancreatic tissue samples. Our aim was to develop a complement or alternative procedure to determine the cellular composition of pancreatic cancerous biopsies, basing on intra-analytic molecular annotation. A standard process for sample stratification at a molecular level does not yet exist. Particularly in the case of retrospective or data depository-based studies, when hematoxylin-eosin stained sections are not available, it supports the correct interpretation of expression profiles.
Methods: A five-gene transcriptional signature (*RNA*CellStrat) was defined that allows cell type-specific stratification of pancreatic tissues. Testing biopsy material from biobanks with this procedure demonstrated high correspondence of molecular (qRT-PCR and microarray) and histologic (hematoxylin-eosin stain) evaluations.
Results: Notably, about a quarter of randomly selected samples (tissue fragments) were exposed as inappropriate for subsequent clinico-pathological interpretation.
Conclusions: Via immediate intra-analytical procedure, our RNA-based stratification *RNA*CellStrat increases the accuracy and reliability of the conclusions drawn from diagnostic and prognostic molecular information.

Introduction

A major challenge in the differential analysis of pancreatic tissues (healthy vs. inflamed vs. malignant pancreas) is the highly heterogeneous tissue composition. Normal pancreatic tissue contains exocrine (80%) and endocrine cells, a distinct ductal system and discrete connective tissue carrying lymphoid and blood vessels as well as nerves [1]. In pancreatic ductal adenocarcinoma (PDAC), this architecture is markedly changed. A varying number of tumor cells become surrounded by desmoplastic stroma, generated by myofibroblastic stellate cells. It contains hypertrophic vessels and nerves, diverse immune cell infiltrates and necrotic areas [2–6]. Also, malignant pancreatic tumors very often infiltrate peripancreatic fat, lymph nodes, the biliary duct or the duodenum, and – when progressed – the gall bladder, stomach or colon, so that

* Corresponding author. European Pancreas Centre, Department of Surgery, University Hospital Heidelberg, INF 116, 69120 Heidelberg, Germany. Tel.: +49 6221 56 39489; fax: +49 6221 56 4830.
E-mail addresses: anette.heller@med.uni-heidelberg.de (A. Heller), nathalia.giese@med.uni-heidelberg.de (N.A. Giese).
[1] Equally contributing authors.

http://dx.doi.org/10.1016/j.pan.2015.05.480

other tissue types may be included in tissue samples obtained from pancreatic cancer patients during collection of the biopsies.

Molecular studies require pre-analytic quantitative characterization of the cell types contained in a complex tumor. Usually, the tissue specimen is evaluated by a pathologist, followed by the diagnosis and visual annotation of the cellular composition based on routine H&E tissue staining. However, patient material, which is provided for non-diagnostically research purposes, is frequently cut into several small fragments and further processed in the form of consecutive sections with changing architecture. This increases the risk of having varying non-tumorous/non-pancreatic proportions in the respective cancer samples. Those samples have to be reevaluated before analysis. Also, in studies using stored tissue homogenates or RNA, the tissue stratification by H&E-staining and visual evaluation is not feasible. Furthermore, for analyses based on already existing RNA-profiles, such as publically available data depositories, it is frequently impossible to procure the cellular composition of the analyzed specimen and ensure accuracy of the diagnostic (i.e. overexpression of tumor-cell specific molecules) or prognostic assumptions.

Our aim was to develop a simple and reliable molecular strategy for a cellular annotation of pancreatic samples, allowing intra-analytic adjustment and generation of comparable sets of tissue samples. In a previous study [7], we observed that the expression of the *PNLIPRP2* gene correlated with the percentage of normal acinar parenchyma contained in the analyzed pancreata. This observation enabled an intra-analytical stratification of the samples and consequently a superior interpretation of the data, showing immunogenicity of the pancreatic molecules. These data allowed us to avoid overstatements in regard to *PNLIPRP2* in a study dealing with epigenetic regulation in pancreatic cancer [8], because emergence of *PNLIPRP2* among top epigenetically-silenced pancreatic genes might rather reflect replacement of *PNLIPRP2*-positive acini (epithelium) with *PNLIPRP2*-negative stroma (mesenchyme), i.e. represent a lineage-specific but not tumor-specific phenomenon. This encouraged us to develop a stratification procedure by employing a multi-gene marker algorithm for cellular annotation of RNA-profiled data (*RNA*CellStrat). Our approach enables a simple, intra-analytic classification of each particular specimen, beyond routine H&E staining, and consequently a reliable interpretation of the molecular data resulting from biomedical studies.

Material and methods

Patients and specimens

The study was performed with tissue samples obtained from patients admitted to the Department of General, Visceral and Transplantation Surgery, University of Heidelberg and National Institute for Health Research, Liverpool. The samples were deposited into PancoBank (Prof. Dr. M.W. Büchler) supported by Heidelberger Stiftung Chirurgie/HSC and by Biomaterial Bank Heidelberg/BMBH (Prof. Dr. P. Schirmacher; BMBF grant 01EY1101). The Pancobank started sample collection in 2001 and consists of currently about 10,000 tissue specimens. The study was approved by the ethical committee of the University of Heidelberg (case number 301/2001) and conducted in accordance with the Helsinki Declaration; written informed consent was obtained from all patients. Diagnoses were established by a pathologist according to World Health Organization classification 2010 [9]. Pancreatic biopsies were immediately snap-frozen in liquid nitrogen and stored at −80 °C. Table 1 provides an overview of the patient cohorts used in the study.

Samples preparation

To achieve most accurate comparability of histological and molecular evaluations (H&E-stained tissues' images and RNA expression profiles, respectively), the frozen tissue samples were cut into 10 μm sections with a cryotom (LEICA Biosystems, Wetzlar, Germany). After 10 serial cuts, one tissue section was transferred to a glass slide and stained with H&E. Subsequently, a pathologist visually quantified the proportion of acinar, stromal fat, endocrine and immune compartments in the specimen. A percentage value was assigned to each compartment. The remaining sections were collected for RNA isolation and transcriptional profiling by qRT-PCR- or microarray-based technologies. Candidate annotation genes are given in Tables 1 and 2.

Quantitative mRNA expression analyses by qRT-PCR

For qRT-PCR-based profiling, tissue sections have been processed to mRNA using MagNA Pure system (Roche Applied Science, Mannheim, Germany). Following cDNA synthesis, LightCycler®-based real-time PCR was performed using the FastStart DNA SYBR Green kit and primers obtained from Search-LC (Heidelberg, Germany) as described previously [4,5,10]. The expression of each specific gene was normalized to housekeeping gene Cyclophilin B and presented as the number of transcripts per 10,000 copies of Cyclophilin B (10 kCPB).

Microarray RNA expression analysis

For microarray-based profiling, tissue sections have been processed to total RNA using AllPrep Isolation kit (Qiagen, Hilden, Germany). To synthesize first and second strand cDNA and for amplifying biotinylated cRNA, the Illumina Totalprep RNA Amplification Kit (Illumina) was used [11]. Genome-wide expression profiling was performed using Sentrix Human-6v3 Whole Genome Expression BeadChips (Illumina San Diego, CA, USA). Hybridization to the BeadChip microarrays was performed

Table 1
Patient information.

		Female	Male	Diagnosis
Quantitative PCR analysis cohort	Gender (n=)	13	21	34 pancreatic ductal adenocarcinoma
	Age (median; IQR)	57.7 (54.1–67.2)	60.9 (56.1–64.1)	
Microarray analysis cohort	Gender (n=)	11	21	24 pancreatic ductal adenocarcinoma,
	Age (median; IQR)	69.3 (58.6–71.8)	62.8 (54.2–73.6)	1 chronic pancreatitis, 1 pancreas donor, 2 cystic adenoma, 1 cystic adenocarcinoma, 3 endocrine tumors
Survival analysis cohort	Gender (n=)	37	61	98 pancreatic ductal adenocarcinoma
	Age (median; IQR)	61.4 (57.9–70.3)	60.9 (55.6–67.1)	

IQR: interquartile range.

Table 2
Correlation of molecular (qRT-PCR) and histological cellular annotation.

Celltype marker genes (qRT-PCR analysis)[a]	Cell/tissue type at histological validation[b]	Correlation coefficient[c]
Amylase2a	Normal pancreas parenchyma	0.846**
PNLIPRP2	Normal pancreas parenchyma	0.899**
Insulin	Islet cells	0.709**
CK19	Ductal/tumor cells	0.710**
MSLN	Ductal/tumor cells	0.698**
Collagen1A1	Stroma	0.493*
Smacta2	Stroma	0.376*
Leptin	Fat	0.802**
CD2	Inflammatoric infiltrates	0.683**
CD8	Inflammatoric infiltrates	0.608**
CD20	Inflammatoric infiltrates	0.587**
FOXp3	Inflammatoric infiltrates	0.667**
CD45	Inflammatoric infiltrates	0.716**

[a] The mRNA copy number of cell-type specific expression markers was analyzed by qRT-PCR.
[b] The areas covered by the respective cell types as percent of the total area upon visual quantification of HE-stained tissue slices by a pathologist.
[c] The correlation coefficient of expression and cellular composition is given (Spearman-Rho coefficient; *$p < 0.05$; **$p < 0.01$).

as recommended by the manufacturer: 10 μl cRNA was mixed with 20 μl GEX-HYB hybridization solution, heated, dispensed onto the large sample port of each array and incubated at 58 °C for 18 h. Following hybridization, the samples were washed according to the standard protocol and scanned with a BeadArray Reader. Raw data was exported from the Beadstudio software to R [12]. The data was quantile normalized and log2 transformed. The threshold for background absorption was set to 4.0 and subtracted from all values. The potential cell-specific tissue marker probes are shown in Supplementary Table S1. Clustering analyses and generation of the heat maps were performed using Chipster Software [13].

Statistical analyses

Boxplots (individual gene expression values, median, interquartile range) and statistical analysis of patients' survival (Kaplan–Meier curves and Log-rank test) was performed with GraphPad Prism 5 Software (La Jolla, CA, USA). Correlation of all expression values with all area-% values provided by a pathologist upon visual evaluation of HE-stained tissue sections was performed via Spearman test (two-sided) using SPSS Statistics 21 software (IBM, New York, USA). Correlation coefficients from 0 to 0.25 indicate little or no relationship between the two parameters; 0.25 to 0.50 indicate a fair degree of correlation; 0.50 to 0.75 show a moderate to good relationship; values above 0.75 represent a very good to excellent relationship [14].

Results

Definition of potential molecular marker genes for annotation of cellular composition in pancreatic biopsies

A literature search was performed to identify cell-type specific genes, which may be capable of representing the different compartments in pancreatic tumor samples. Under consideration were endocrine, acinar, ductal, tumor, stromal, fat, and immune markers. Expression of β-cell-produced insulin (*INS*) marked Langerhans islets - a major component of the endocrine compartment [15,16]. The genes encoding enzymes produced by pancreatic acinar cells, amylase-A2-alpha (*AMY2A*) [17–19] and pancreatic lipase related protein 2 (*PNLIPRP2*) [20,21], were chosen as markers of normal exocrine parenchyma. Cytokeratin-19 (*CK19*) served as a marker for ductal epithelial cells, both normal and transformed [22,23]. Differential overexpression of mesothelin (*MSLN*) was assumed to mark only malignant ductal epithelium [24,25]. A desmoplastic reaction is characteristic of both inflammatory and cancerous pancreata. Generally, pancreatic stellate cells are activated in these tissues, with increased smooth muscle actin alpha (*ACTA2/SMACTA2*) expression as an indicator [2,26,27]. These cells produce excessive amounts of extracellular matrix proteins such as collagen type I (*COL1A1*) and fibronectin (*FN1*) [27,28]. These three markers were therefore used for the identification of stellate cells and the extracellular matrix. Leptin (*LEP*) is secreted by adipocytes and serves as a marker for adipose tissue (fat) [29,30]. To identify immune cell accumulation in tissue samples, diverse markers representing different types of immune cells were selected: *CD45* antigen *(PTPRC)* is a receptor type protein tyrosine phosphatase necessary in the regulation of signal transduction in T- and B-cells. Therefore, *CD45* is a suitable candidate to represent the collectivity of immune cells in the tissue sample or tertiary lymphatic inclusions. We then further characterized the immune cell compartment according to *CD2*, *CD8*, *FOXp3* and *CD20* expression. Infiltration of these cells into chronic pancreatitis (CP) and PDAC has been shown [31]. CD2 is a surface antigen expressed on all peripheral blood T-cells [32–34]; CD8 antigen is found in cancer patients on cytotoxic T-cells, mediating responses against cancer antigens [35]. Forkhead box protein 3 (*FOXp3*) serves as marker for regulatory T-cells [36–38]. CD20 is a marker for B-cells [39,40].

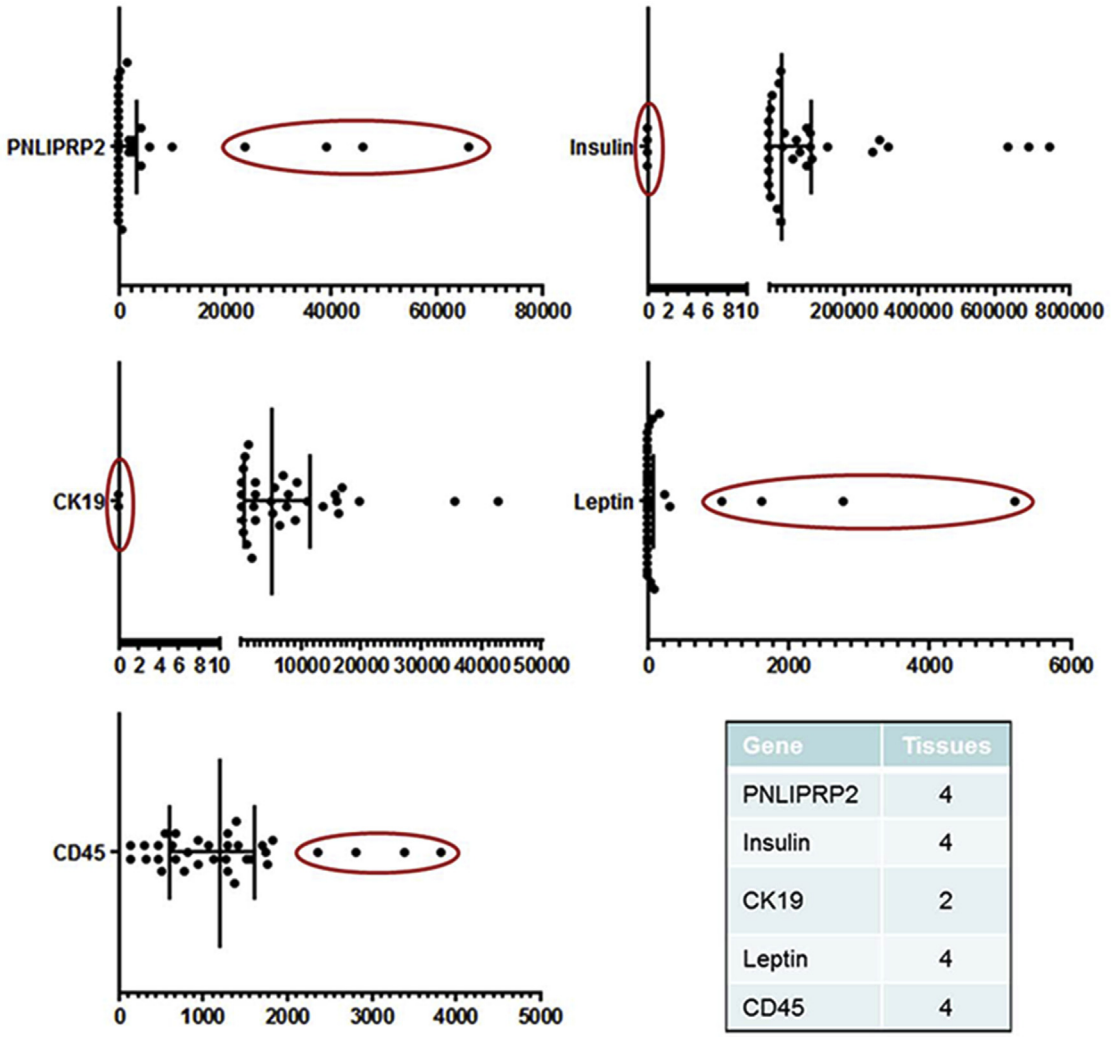
Analysis of specimens by quantitative reverse transcription (qRT)-PCR

Frozen pancreatic tumor tissue (34 samples) from PDAC patients was randomly chosen from an existing Pancobank of about 10,000 pancreatic specimens. The gene expression (number of mRNA copies per 10 kCPB) of the potential marker genes was correlated with the area-percentage values of the different cellular compartments on the slices via Spearman correlation (Table 2, Supplementary Table S2; Supplementary Table S3). Overall, there was good correlation between RNA-based and histological annotation. Eleven of the 13 marker genes analyzed correlated with the appropriate cell types. Only the stroma content did not fit to the expression of *COL1A1* or *SMACTA2*.

Based on these results, the best five molecular markers were defined to as *RNA*CellStrat panel and studied in more detail: *PNLIPRP2* for normal acinar parenchyma; *INS* for islets; *CK19* for ductal and PDAC epithelium; *LEP* for fat tissue; and *CD45* for immune infiltrates. Visualization via a boxplot indicated the presence of outliers. Gene expression values of *PNLIPRP2*, *LEP* and *CD45* were defined as outliers that were distinctly separable. The values of *INS* and *CK19* were regarded as outliers if expression was missing (<5 mRNA copies). Four outliers were detected each for *PNLIPRP2*, *INS*, *LEP* and *CD45*, and two for *CK19* (Fig. 1A). Strong expression of *PNLIPRP2*, *LEP* and *CD45* suggested a sizeable amount of either normal pancreatic parenchyma, fat or immune infiltrates in the tissue samples. The absence of insulin expression in four samples should reflect an absence of β-cells of islets; consequently, the pancreatic nature of these four tissue samples was not certain. Since *CK19* should be usually expressed in ductal tumor cells, the two samples lacking *CK19* expression were conspicuous as well, assuming that no tumor cells are included in the tumor sample. In total, 18 abnormal expression values were noted for 11 tissue samples.

We scrutinized histological data to elucidate why the 11 tissues had such distinct profiles (Fig. 1B). The four samples with high

a



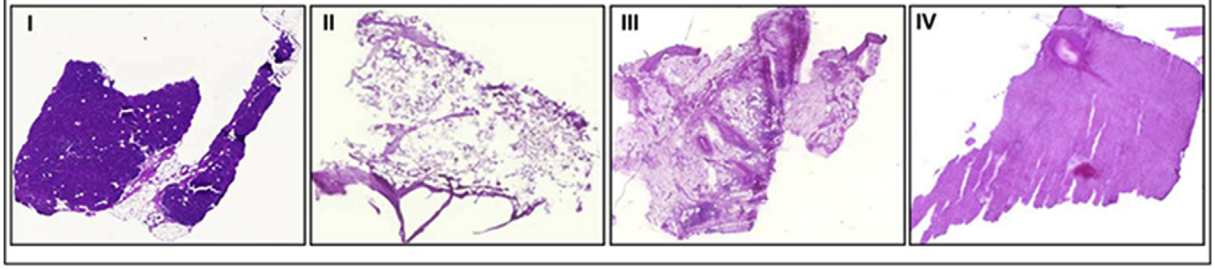
Gene	Tissues
PNLIPRP2	4
Insulin	4
CK19	2
Leptin	4
CD45	4

b

	Acini[1] [% area]	PNLIPRP2[2] [mRNA copies]	Islet cells[1] [% area]	Insulin[2] [mRNA copies]	Ductal/Tumor cells[1] [% area]	CK19[2] [mRNA copies]	Fat[1] [% area]	Leptin[2] [mRNA copies]	inflammatoric infiltrates[1] [% area]	CD45[2] [mRNA copies]	Exculsion criterium
I	78	66195	2	276230	0	657	10	0	1	137	high PNLIPRP2
	62	46270	2	26443	0	481	2	0	3	512	high PNLIPRP2
	87	39148	1	154289	0	2005	5	0	2	336	high PNLIPRP2
II	34	23784	1	62624	11	5097	0	0	6	666	high PNLIPRP2
	0	0	0	0	0	0	82	5202	3	2818	high leptin, high CD45, no CK19, no insulin
	0	0	0	0	0	151	82	2776	3	626	high leptin, no insulin
	0	0	0	0	0	0	74	1620	7	1699	high leptin, no insulin, no CK19
III	0	0	0	109761	5	15673	68	1057	9	3392	high leptin, high CD45
	2	29	0	252	0	195	70	328	13	3820	high CD45
IV	3	0	3	109125	9	2472	0	0	19	2363	high CD45
	0	0	0	0	93	67	0	0	2	140	no insulin

[1]Histological validation of HE-stained tissue section [2]qRT-PCR

c



expression of *PNLIPRP2* contained 34%–78% of normal pancreatic parenchyma (Fig. 1C, I). The high *LEP* levels coincided with 68%–82% inclusions of peripancreatic fat (Fig. 1C, II). The strong *CD45* expression correlated with a large amount of desmoplastic stroma infiltrated by immune cells (Fig. 1C, III). The two samples lacking *CK19* expression were the same ones that exhibited high *LEP* expression. Three other samples also have extremely low *CK19* level. Three of the four tissues, which showed no expression of *INS*, had already been defined as outliers because of their high fat content. The source of the fourth sample was confirmed histologically as being pancreatic (Fig. 1C, IV). The lack of *INS* expression was actually due to the absence of normal tissue areas, because the sample consisted primarily of pancreatic tumor cells. We would like to point out that only this one tissue sample had a tumor cell percentage of over 90% and, paradoxically, did not show high *CK19* expression. All other 33 samples obtained from tumor patients contained between 0% and 40% of ductal tumor cells embedded in desmoplastic stroma.

We noted that the five samples with low stromal content (16 ± 1% area) but high *SMACTA2 and COL1A1* values (15,272 ± 5421 and 5452 ± 2807 copies/10 kCPB) displayed high fat content (75 ± 3% area and *LEP*-expression of 2197 ± 851 copies/10 kCPB; Table S2) Exclusion of LEP^{high} samples increased correspondence of mesenchymal histological and molecular annotations (Rho = 0.568 and 0.663 with $p < 0.001$). Thus, $SMACTA2/COL1A1^{high}$-mesencynal cells might occupy adipose tissue suggesting that prognostic interpretation of stroma-related molecular findings requires prior exclusion of LEP^{high} samples.

In summary, *RNA*CellStrat procedure reduced the number of comparable tissue samples suitable for diagnostic and/or prognostic correlations from 34 to 23; a reduction of 32%. The 11 conspicuous samples had a composition that made them inappropriate for a comparative study since they were highly contaminated with normal parenchyma, fat or stroma.

Independent microarray analysis

For testing the robustness and thus broad applicability of the *RNA*CellStrat annotation, an independent analysis was performed on a different set of samples with a different expression profiling procedure. 32 pancreatic tissue samples were selected from the Pancobank; there were 24 samples from patients with ductal adenocarcinoma, 1 case of chronic pancreatitis, 2 patients with cystic adenoma, 1 sample of cystic adenocarcinoma, 1 healthy pancreas, as well as 3 endocrine pancreas tumors.

Total RNA was isolated and used for a genome-wide expression analysis utilizing a commercial microarray platform. As before, a pathologist examined the composition of the H&E-stained tissue sections. The histological data was then compared to the results of the mRNA expression analysis. The signal intensities obtained from 17 microarray probes representing 10 genes were studied (Table 3, Supplementary Table S4; Supplementary Table S5). The results of six gene probes correlated significantly with the area percentages of corresponding cell types: expression of *AMY2A* and *PNLIPRP2* correlated with amount of normal pancreatic parenchyma (R = 0.670, 0.664); *INS* with the islets (R = 0.568); *CK19* and *MSLN* with ductal tumor cells (R = 0.598, 0.579); and the signal at the *LEP*_2 probe exhibited good correlation (R = 0.535) with the amount of fat in the tissue samples. Correlation of the signals at the microarray oligonucleotide probes CD45_3 and CD45_4 and the volume of inflammatory infiltrates in the H&E-stained tissue section was significant, but the correlation coefficient showed only a fair degree of relationship (R = 0.349, 0.410). As in the PCR-approach, stromal marker molecules (*SMACTA2*, *COL1A1*, *FN1*) did not correlate with the visually assessed areas of that tissue type in the sample. Therefore, *PNLIPRP2*, *INS*, *CK19*, and the *LEP*_2, and *CD45*_4 probes were selected to perform cellular annotation.

Table 3
Correlation of molecular (microarray signal intensity variations) and histological cellular annotation.

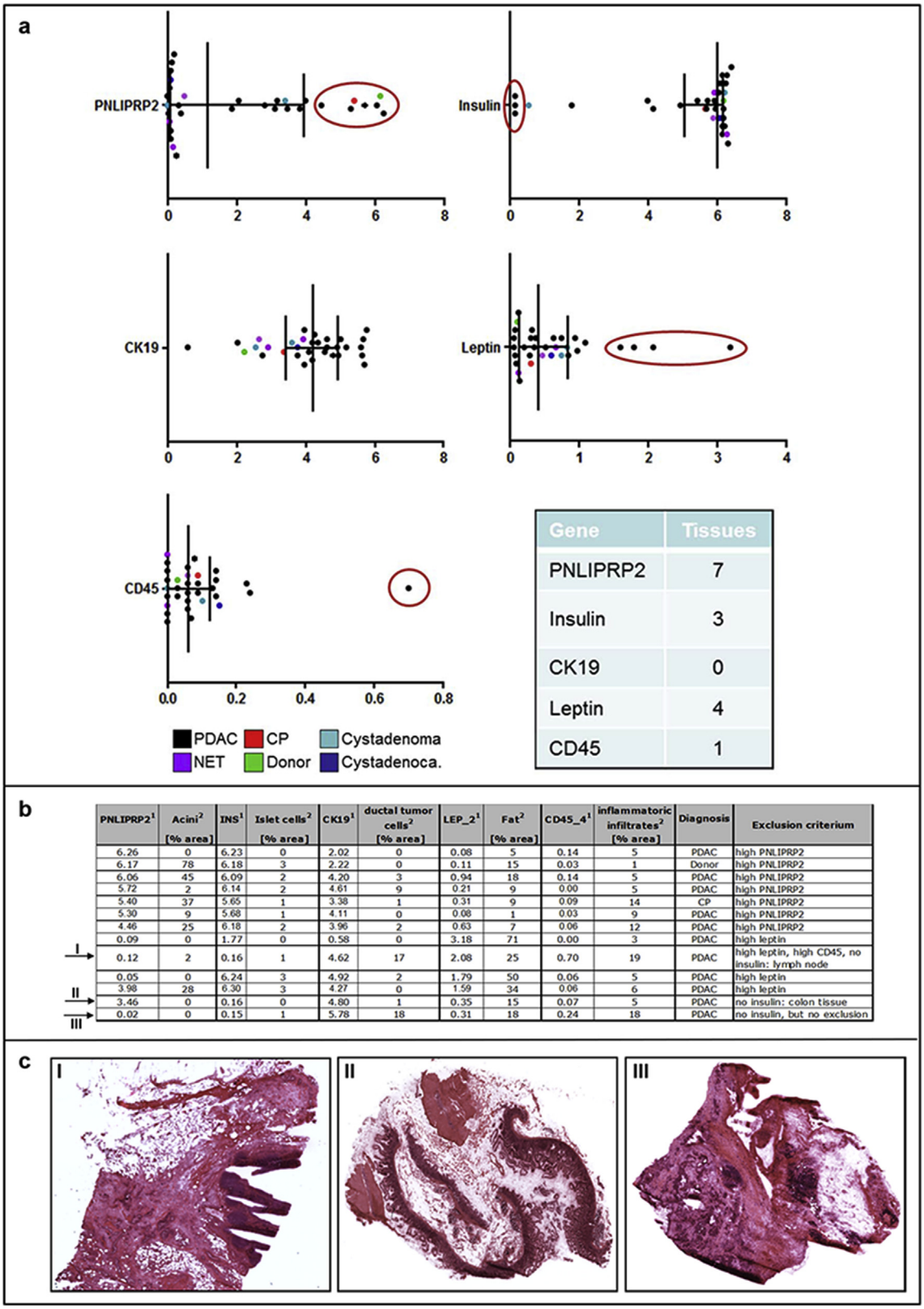
Celltype marker genes (microarray analysis)[a]	Cell/tissue type at histological validation[b]	Correlation coefficient[c]
Amylase2a_1	Normal pancreas parenchyma	−0.173
Amylase2a_2	Normal pancreas parenchyma	0.670**
PNLIPRP2	Normal pancreas parenchyma	0.664**
Insulin	Islet cells	0.568**
CK19	Ductal/tumor cells	0.598**
MSLN	Ductal/tumor cells	0.579**
Collagen1A1	Stroma	0.224
Smacta2	Stroma	0.087
Fibronectin1_1	Stroma	0.173
Fibronectin1_2	Stroma	−0.033
Fibronectin1_3	Stroma	0.216
Leptin_1	Fat	−0.018
Leptin_2	Fat	0.535**
CD45_1	Inflammatoric infiltrates	−0.296
CD45_2	Inflammatoric infiltrates	0.185
CD45_3	Inflammatoric infiltrates	0.349*
CD45_4	Inflammatoric infiltrates	0.410*

[a] The cell-type specific gene expression markers were analyzed on a microarray. For some transcripts, the result of different oligonucleotide probes are shown, indicated by an underscore followed by a number.
[b] Area-percentages as determined by a pathologist evaluating HE-stained tissue section.
[c] The correlation coefficient of expression and cellular composition is given (Spearman-Rho coefficient; *p < 0.05; **p < 0.01).

As before, the expression of the five marker genes in the pancreatic tissue samples was visualized via boxplots in a blinded manner, without presenting the sample annotations; an unblinded version is shown in Fig. 2A. There were 15 outlying values in 13 different samples. In twelve tissues, only one particular marker gene was expressed differently, while three markers were recognized as outliers in one tissue. Gene expression values of *PNLIPRP2*, *LEP* and *CD45*, which are out of range of dispersion, were defined as outliers. The values of *INS* and *CK19* are regarded as outlier if expression is missing (<0.2 AU). Specifically, seven samples showed high *PNLIPRP2* and three samples high *LEP* expression. One sample was conspicuous due to high *LEP*, extremely high *CD45* and missing *INS* expression. Another two tissues did also not express *INS*. For these candidates, the tissue origin had to be further clarified. Actually, all outlier samples showed only minimal *CK19* expression. The tissue with the lowest *CK19* value (=0.58) also had the highest *LEP* expression (=3.18).

Upon un-blinding the diagnoses, it became apparent that most 'normal' pancreas parenchyma was observed in the healthy donor tissue. The highest *CK19* expression was observed in the samples

Fig. 1. (a) Expression values of the five *RNA*CellStrat genes in qRT-PCR analyses of 34 PDAC tissue samples. Expression is displayed in relative mRNA copy numbers along the x-axis; median and interquartile ranges are indicated. For four of the marker genes – *PNLIPRP2*, *LEP*, insulin and *CD45* – four values each were outliers with an extreme gene expression. Two samples without *CK19* expression were conspicuous as well. The 18 abnormal expression values were present in 11 patient tissues (some samples were conspicuous in more than one marker gene). (b + c) Evaluation of the 11 conspicuous PDAC tissues: (b) Comparison of the visually assessed tissue compartments (areas in percent) with RNA expression of the five marker genes analyzed via qRT-PCR. (c) Exemplary, the H&E-staining of cancer samples with a high expression of *PNLIPRP2* (I), *LEP* (II), and *CD45* (III) or missing insulin expression (IV) are shown. Staining confirmed the unusual composition of the tissue samples: a high amount of normal pancreatic parenchyma (I), peripancreatic fat (II) and inflammatory infiltrates (III) were seen, respectively. The source of the pancreatic tissue without insulin expression was confirmed to be pancreatic PDAC tumor (IV).



Gene	Tissues
PNLIPRP2	7
Insulin	3
CK19	0
Leptin	4
CD45	1

b

	PNLIPRP2[1]	Acini[2] [% area]	INS[1]	Islet cells[2] [% area]	CK19[1]	ductal tumor cells[2] [% area]	LEP_2[1]	Fat[2] [% area]	CD45_4[1]	inflammatoric infiltrates[2] [% area]	Diagnosis	Exclusion criterium
	6.26	0	6.23	0	2.02	0	0.08	5	0.14	5	PDAC	high PNLIPRP2
	6.17	78	6.18	3	2.22	0	0.11	15	0.03	1	Donor	high PNLIPRP2
	6.06	45	6.09	2	4.20	3	0.94	18	0.14	5	PDAC	high PNLIPRP2
	5.72	2	6.14	2	4.61	9	0.21	9	0.00	5	PDAC	high PNLIPRP2
	5.40	37	5.65	1	3.38	1	0.31	9	0.09	14	CP	high PNLIPRP2
	5.30	9	5.68	1	4.11	0	0.08	1	0.03	9	PDAC	high PNLIPRP2
	4.46	25	6.18	2	3.96	2	0.63	7	0.06	12	PDAC	high PNLIPRP2
	0.09	0	1.77	0	0.58	0	3.18	71	0.00	3	PDAC	high leptin
I →	0.12	2	0.16	1	4.62	17	2.08	25	0.70	19	PDAC	high leptin, high CD45, no insulin: lymph node
	0.05	0	6.24	3	4.92	2	1.79	50	0.06	5	PDAC	high leptin
II →	3.98	28	6.30	3	4.27	0	1.59	34	0.06	6	PDAC	high leptin
	3.46	0	0.16	0	4.80	1	0.35	15	0.07	5	PDAC	no insulin: colon tissue
III →	0.02	0	0.15	1	5.78	18	0.31	18	0.24	18	PDAC	no insulin, but no exclusion

Fig. 2. (a) Expression results of *RNA*CellStrat markers in microarray experiments of patient samples with different pancreatic diseases. Results are displayed as relative mRNA copy numbers along the x-axis; median and interquartile ranges are shown. In total, there were 15 expression outliers, representing 13 tissues. Twelve of these samples showed a noticeable expression of one marker gene. One sample featured significant variations of three marker genes. The different origin of the samples is indicated by a color code. (b + c) Evaluation of the 13 conspicuous tissues: (b) Comparison of the marker gene expression values and the histological validation (areas in percent). (c) The H&E-staining is shown of a sample with (I) high *LEP*, high *CD45* and no insulin expression, revealing the inclusion of a lymph node. (II) H&E-staining of a sample without insulin expression showed the inclusion of colon tissue. The other pancreas sample without insulin expression (III) was confirmed to be from a pancreas.

from PDAC patients with the highest amount of ductal tumor cells. The endocrine tumor patients did not overexpress *INS*. The clinical pathological report of the three endocrine tumors confirmed the absence of *INS*, proven by immunohistochemistry. A detailed overview about the status of the 13 conspicuous samples is provided in Fig. 2B.

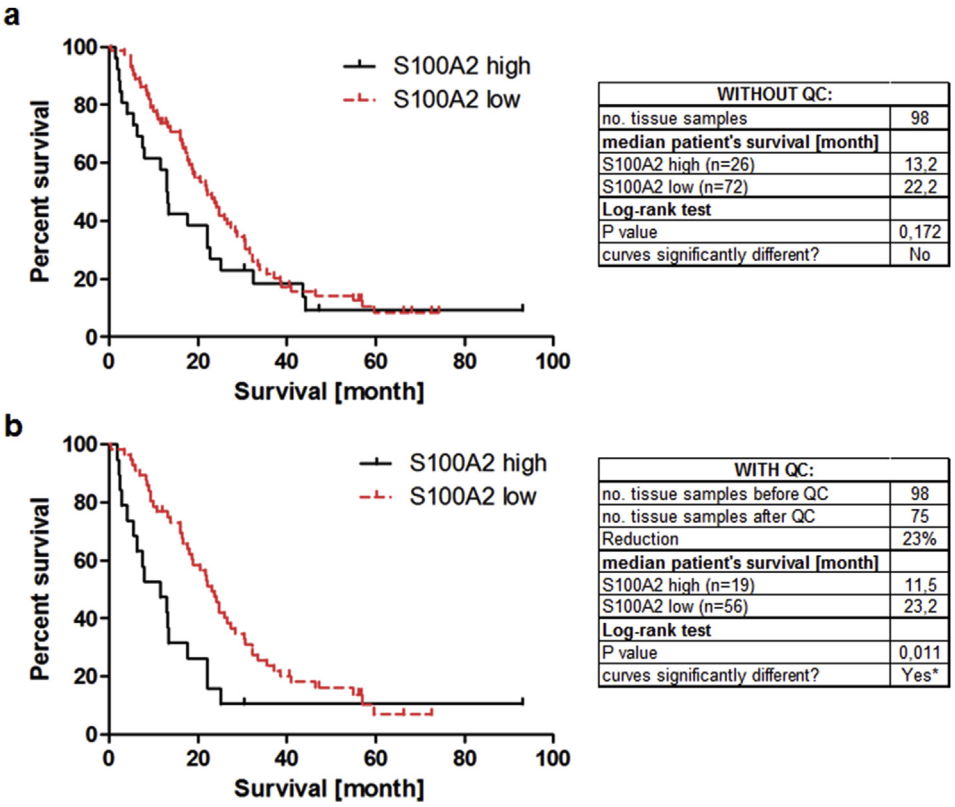
The five tumor samples that exhibited high *PNLIPRP2* expression had areas of normal pancreas parenchyma that varied from 0% to 78%. The four *LEP*-expressing samples represented tissues containing a high amount of fat (34 %–71%). Histological staining of the one PDAC sample with the three remarkable results (high *LEP*, high *CD45*, no *INS*) revealed inclusions of fat and peripancreatic lymph node (Fig. 2C, I). Histological validation of the H&E-stained sections of the remaining two PDAC tissue samples without *INS* expression elucidated that one sample was normal colon tissue (Fig. 2C, II). In the other specimen, the pancreatic tumor type (desmoplastic areas and tumor cells) was confirmed and the sample was not excluded (Fig. 2C, III).

In summary, *RNA*CellStrat procedure reduced the number of tissue specimens from 32 to 21, a decrease of 34%. Utilizing the five markers, it was possible to eliminate samples that were rich in fat or normal pancreatic parenchyma and to identify non-PDAC entities and tissues from non-pancreatic origin such as lymph node or colon.

Influence of sample composition on prognosis

In order to prove the importance of performing the *RNA*CellStrat procedure before evaluating prognostic or diagnostic markers and to show the applicability of the established procedure, a marker-based (S100A2) survival time analysis and signature-based PDAC sybtypes identification (with PDAssigner) was performed on an independent set of Pancobank samples, with and without intra-analytic *RNA*CellStrat annotation. Microarray expression data of 98 PDAC patients were used for this analysis. The expression values observed for the five *RNA*CellStrat marker genes were considered for stratification.

It had been shown that patients with low S100 calcium binding protein A2 (*S100A2)* expression in tumors had a better survival rate [41,42]. In order to confirm prognostic relevance of *S100A2* in our cohort, we assigned patients into two groups, $S100A2^{high}$ and $S100A2^{low}$ (mean value as cut-off), and performed a standard Kaplan–Meier survival analysis with a subsequent Log-rank test. Without *RNA*CellStrat, 26 patients were allocated to the group with high *S100A2* expression, and 72 to the group with low expression, based on the mean expression of the gene (cut-off = 0.859). The median survival of the former group was 13.2 months, that of the latter 22.2 months. The Log-rank test showed that there is no significant difference between the two groups (p-value = 0.172). The hypothesis that low expression of the *S100A2* gene is associated with increased survival could not be confirmed (Fig. 3A). With *RNA*CellStrat, 23 tissue samples were identified as outliers (23%). Those tissue samples were excluded from the survival analysis (Supplementary Fig. S1). The patient number was thus reduced from 98 to 75. Survival of 19 patients with high *S100A2* expression (median survival time of 11.5 months) was compared to that of the other 56 patients (median survival time of 23.2 months). The mean *S100A2* expression value of 0.893 served as the cut-off to divide the



WITHOUT QC:	
no. tissue samples	98
median patient's survival [month]	
S100A2 high (n=26)	13,2
S100A2 low (n=72)	22,2
Log-rank test	
P value	0,172
curves significantly different?	No

WITH QC:	
no. tissue samples before QC	98
no. tissue samples after QC	75
Reduction	23%
median patient's survival [month]	
S100A2 high (n=19)	11,5
S100A2 low (n=56)	23,2
Log-rank test	
P value	0,011
curves significantly different?	Yes*

Fig. 3. Influence of tissue cell composition on the prognostic value of the marker molecule *S100A2*. (a) From a total of 98 randomly selected patients, 26 patients were allocated in the group of high *S100A2* expression (median survival 13.2 months) and 72 in the group with low expression (median survival 22.2 months). The result of a Log-rank test without prior *RNA*CellStrat stratification according to cell composition showed no significant difference between the two groups. (b) After the *RNA*CellStrat was performed, 28 samples were removed from the analysis. Of the remaining patients, 19 with high *S100A2* expression (median survival 11.5 months) were compared to 56 patients with low expression (median survival time 23.2 months). A log-rank test showed a significant difference between the two groups (p-value = 0.011).

patient into two groups. The result was in clear contrast to that of the first calculation about a correlation of survival and *S100A2* expression. The log-rank test showed a highly significant difference between the two groups (p-value = 0.011; Fig. 3B), demonstrating the impact of the *RNA*CellStrat process on prognostic significance.

Collisson et al. developed a 62-gene signature, designated PDAssigner, capable of prognosis-relevant discrimination of the three PDAC subtypes: classical, quasimesenchymal and exocrine-like [43]. Cluster analysis of our microarray data sets showed that this identification by PDAC transcriptional phenotype could be achieved only with (n = 75 patients) but not without (n = 98 patients) *RNA*CellStrat clearance (Heat maps generated using Chipster software [13] are given in Supplementary Fig. S2).

Discussion

Human tissue samples are an essential resource for oncological research. The prevalent sources of such material are biobanks. These facilities strive to provide well-annotated material of high quality, meeting ISO standards requiring an extensive quality control management [44–46]; but not each analyzed tissue sample is derived from an accredited biobank. At the same time, the need of effective networking in multi-centric and multi-national projects is growing with the intention of collecting a sufficiently large number of cases of different origin to assess the usefulness of a marker or drug [46,47]. To such ends, a rigorous quality assessment of tissue samples is prerequisite.

Translational research requires tissue samples that represent a tumor as a whole and are annotated to a high level of accuracy. Often, this is not true for research material that is left over from standard diagnostics. It is therefore essential to define a tissue sample's quality very precisely before utilizing it for scientifically motivated molecular analyses and drawing any conclusions from their results. A precise determination of the abundance of different cell-types in tumor tissue is crucial to assuring comparability of samples, in particular for multicentric studies or for retrospective analyses that are based on data from public databases.

The established pathological method of tissue stratification is the macroscopic evaluation, followed by the evaluation of H&E-stained slices to define the diagnosis. It is essential to know if the resection margins are tumor-free and lymph-nodes are affected, which can be ascertained by the pathological examination. For molecular research, however, the precise amount of the different tissue types is of enormous interest. In the routine, tissue stratification is done with representative tissue slides to fulfill the diagnostic requirements, but the tissues are not necessarily prepared to perform molecular analyses. To save biochemical standards in molecular analysis, an additional approach for quantifying different tissue types, the imprint cytology, was created. It has been shown that this method is effective in eliminating samples with significant necrosis, but not in evaluating the quantity of tumor cells [48].

In our study, we introduced an RNA-based procedure for a quantification of the cell composition of pancreatic tissue samples for molecular analyses, employing either qRT-PCR or microarray data. In 2010, Shen-Orr and colleagues already reported that differential gene expression patterns of diverse cell types correlate with relative cell-type frequencies in rat brain, liver and lung [49]. In a previous study [7], we demonstrated that *PNLIPRP2* gene expression allows to predict the percentage of normal pancreatic parenchyma found in pancreatic tumor samples. This prompted us to create a procedure for the molecular annotation of a tissue cellular composition, based on mRNA-expression profiles of a set of cell-type marker genes (*RNA*CellStrat). This approach is advantageous in that it can be used as an internal control as long as the mRNA-profiles of the five marker genes are part of the relevant studies. Also, an internal validation of tissues can be performed, to guarantee the molecular based tissue composition, which is particularly important for retrospective or data depositories-based studies.

To maximize the number of experiments that can be performed from single specimen and to avoid repeated freeze-thaw cycles, pancreatic biopsies used for research purposes are frequently subdivided into smaller tissue specimens, thus delivering different cell composition of the individual fragments randomly selected for molecular profiling. From in total 164 tissue samples that were randomly selected from our Pancobank of some 10,000 samples, 45 would have been excluded from comparative analyses, equivalent to about 27%. When taking into account only the PDAC-diagnosed cohorts from qRT-PCR and microarray analyses (in total 156 samples), 43 samples which equal 28% were eliminated. *RNA*CellStrat allowed identification of the samples suitable for prognostic associations. Next to stratification according to cellular composition, the established method allowed the identification of samples where the amount of malignant cells is insufficient or contain a high amount of fat or acinar cells, or even other adjacent organs such as colon or lymph nodes. By using only five marker genes, we managed to stratify tumor tissues according to their cellular composition. In consequence, 32% and 21% of the samples were eliminated from qRT-PCR and microarray analyses, respectively. Also, we demonstrated the applicability and importance of the method for improving prognostic accuracy through an exemplary survival analysis of PDAC patients. Usually, a significant reduction of the sample number in a study, here by 23%, leads to worsening of its statistical power. In this case, however, the reduction led to an increase in significance because the suitability of the remaining samples for prognostic correlations was proven by *RNA*CellStrat yielding probes of a comparable composition. Moreover, *RNA*CellStrat precluded wrong prognoses for the patients with excluded samples. Finally, our data suggested that an interpretation of stroma-related molecular findings should be considered prior to exclusion of LEP^{high} (fat) samples because mesenchymal cells in adipose tissues might express similar sets of marker molecules.

In conclusion, the results show that the determination of tumor tissue composition is crucial for the interpretation of expression profiles and other molecular data and influences their prognostic value significantly. The process represents a complement or alternative procedure to pre-analytic histological evaluation. In this way, the procedure increases reliability of the results of biomedical studies and improves prognostic accuracy while reducing the probability of findings' misinterpretation. While the process is optimized for a cellular annotation of pancreatic tumor tissues, the strategy may be transferred and applicable to other diseases and tissue types, displaying a tool to strengthen the molecular tissue analysis.

Acknowledgments

We are grateful to Brunhilde Bentzinger, Melanie Bier, Monika Meinhardt, and Esther Soyka for their excellent technical support. We thank William Greenhalf and Eithne Costello for study support.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.pan.2015.05.480.

References

[1] Evans DBH, von D, Hruban RH. Pancreatic cancer. Ma: Jones & Bartlett Pub; 2005.

[2] Apte MV, Park S, Phillips PA, Santucci N, Goldstein D, Kumar RK, et al. Desmoplastic reaction in pancreatic cancer: role of pancreatic stellate cells. Pancreas 2004;29:179–87.
[3] Bockman DE. Transition to pancreatic cancer in response to carcinogen. Langenbecks Arch Surg 2008;393:557–60.
[4] Ceyhan GO, Bergmann F, Kadihasanoglu M, Altintas B, Demir IE, Hinz U, et al. Pancreatic neuropathy and neuropathic pain–a comprehensive pathomorphological study of 546 cases. Gastroenterology 2009;136:177–186 e171.
[5] Koninger J, Giese T, di Mola FF, Wente MN, Esposito I, Bachem MG, et al. Pancreatic tumor cells influence the composition of the extracellular matrix. Biochem Biophys Res Commun 2004;322:943–9.
[6] Pandol S, Edderkaoui M, Gukovsky I, Lugea A, Gukovskaya A. Desmoplasia of pancreatic ductal adenocarcinoma. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc 2009;7:S44–7.
[7] Heller A, Zornig I, Muller T, Giorgadze K, Frei C, Giese T, et al. Immunogenicity of serex-identified antigens and disease outcome in pancreatic cancer. Cancer Immunol Immunother CII 2010;59:1389–400.
[8] Dutruel C, Bergmann F, Rooman I, Zucknick M, Weichenhan D, Geiselhart L, et al. Early epigenetic downregulation of wnk2 kinase during pancreatic ductal adenocarcinoma development. Oncogene 2014;33:3401–10.
[9] Bosman FT, CF, Hruban RH, Theise ND. Who classification of tumours of the digestive system. 2010.
[10] Bartel M, Hansch GM, Giese T, Penzel R, Ceyhan G, Ketterer K, et al. Abnormal crosstalk between pancreatic acini and macrophages during the clearance of apoptotic cells in chronic pancreatitis. J Pathol 2008;215:195–203.
[11] Bauer AS, Keller A, Costello E, Greenhalf W, Bier M, Borries A, et al. Diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis by measurement of microrna abundance in blood and tissue. PloS One 2012;7:e34151.
[12] Ritchie ME, Dunning MJ, Smith ML, Shi W, Lynch AG. Beadarray expression analysis using bioconductor. PLoS Comput Biol 2011;7:e1002276.
[13] Kallio MA, Tuimala JT, Hupponen T, Klemela P, Gentile M, Scheinin I, et al. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC genomics 2011;12:507.
[14] Dawson BT, Robert G. Basic & clinical biostatistics. Lange Medical Books; 2004.
[15] Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 2006;103:2334–9.
[16] Houssay BA, Foglia VG, Smyth FS, Rietti CT, Houssay AB. The hypophysis and secretion of insulin. J Exp Med 1942;75:547–66.
[17] Gardner JD, Jackson MJ. Regulation of amylase release from dispersed pancreatic acinar cells. J Physiol 1977;270:439–54.
[18] Matthews EK, Petersen OH, Williams JA. Pancreatic acinar cells: acetylcholine-induced membrane depolarization, calcium efflux and amylase release. J Physiol 1973;234:689–701.
[19] Pandol SJ. In: The exocrine pancreas; 2010. San Rafael (CA).
[20] Lowe ME. The triglyceride lipases of the pancreas. J lipid Res 2002;43:2007–16.
[21] Watson RR, Tye JG, McMurray DN, Reyes MA. Pancreatic and salivary amylase activity in undernourished colombian children. Am J Clin Nutr 1977;30:599–604.
[22] Bouwens L. Cytokeratins and cell differentiation in the pancreas. J Pathol 1998;184:234–9.
[23] Cao D, Maitra A, Saavedra JA, Klimstra DS, Adsay NV, Hruban RH. Expression of novel markers of pancreatic ductal adenocarcinoma in pancreatic nonductal neoplasms: additional evidence of different genetic pathways. Mod Pathol Off J U S Can Acad Pathol Inc 2005;18:752–61.
[24] Hassan R, Laszik ZG, Lerner M, Raffeld M, Postier R, Brackett D. Mesothelin is overexpressed in pancreaticobiliary adenocarcinomas but not in normal pancreas and chronic pancreatitis. Am J Clin Pathol 2005;124:838–45.
[25] Hucl T, Brody JR, Gallmeier E, Iacobuzio-Donahue CA, Farrance IK, Kern SE. High cancer-specific expression of mesothelin (msln) is attributable to an upstream enhancer containing a transcription enhancer factor dependent mcat motif. Cancer Res 2007;67:9055–65.
[26] Kloppel G. Pathology of non-endocrine pancreas. In: DiMagno EP, Gardner J, et al., editors. The pancreas: biology, pathobiology and disease. New York: Raven Press; 1993. p. 871–97.
[27] Korc M. Pancreatic cancer-associated stroma production. Am J Surg 2007;194:S84–6.
[28] Iacobuzio-Donahue CA, Maitra A, Olsen M, Lowe AW, van Heek NT, Rosty C, et al. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cdna microarrays. Am J Pathol 2003;162:1151–62.
[29] Chandra R, Liddle RA. Neural and hormonal regulation of pancreatic secretion. Curr Opin Gastroenterol 2009;25:441–6.
[30] MacDougald OA, Hwang CS, Fan H, Lane MD. Regulated expression of the obese gene product (leptin) in white adipose tissue and 3t3-l1 adipocytes. Proc Natl Acad Sci U S A 1995;92:9034–7.
[31] Demir IE, Schorn S, Schremmer-Danninger E, Wang K, Kehl T, Giese NA, et al. Perineural mast cells are specifically enriched in pancreatic neuritis and neuropathic pain in pancreatic cancer and chronic pancreatitis. PloS One 2013;8:e60529.
[32] Huet S, Wakasugi H, Sterkers G, Gilmour J, Tursz T, Boumsell L, et al. T cell activation via cd2 [t, gp50]: the role of accessory cells in activating resting t cells via cd2. J Immunol 1986;137:1420–8.
[33] Sewell WA, Brown MH, Dunne J, Owen MJ, Crumpton MJ. Molecular cloning of the human t-lymphocyte surface cd2 (t11) antigen. Proc Natl Acad Sci U S A 1986;83:8718–22.
[34] Suthanthiran M. Human t cell activation: participation of tll antigen (cd2, t,gp.50) in t cell-accessory cell interactions. Transplant Proc 1987;19:336–7.
[35] Schmitz-Winnenthal FH, Escobedo LV, Beckhove P, Schirrmacher V, Bucur M, Ziouta Y, et al. Specific immune recognition of pancreatic carcinoma by patient-derived cd4 and cd8 t cells and its improvement by interferon-gamma. Int J Oncol 2006;28:1419–28.
[36] Felcht M, Heck M, Weiss C, Becker JC, Dippel E, Muller CS, et al. Expression of the t-cell regulatory marker foxp3 in primary cutaneous large b-cell lymphoma tumour cells. Br J Dermatol 2012;167:348–58.
[37] Sakaguchi S, Miyara M, Costantino CM, Hafler DA. Foxp3+ regulatory t cells in the human immune system. Nat Rev Immunol 2010;10:490–500.
[38] Seidel MG, Ernst U, Printz D, Juergens B, Pichler J, Attarbaschi A, et al. Expression of the putatively regulatory t-cell marker foxp3 by cd4(+)cd25+ t cells after pediatric hematopoietic stem cell transplantation. Haematologica 2006;91:566–9.
[39] Golay JT, Clark EA, Beverley PC. The cd20 (bp35) antigen is involved in activation of b cells from the g0 to the g1 phase of the cell cycle. J Immunol 1985;135:3795–801.
[40] Reff ME, Carner K, Chambers KS, Chinn PC, Leonard JE, Raab R, et al. Depletion of b cells in vivo by a chimeric mouse human monoclonal antibody to cd20. Blood 1994;83:435–45.
[41] Biankin AV, Kench JG, Colvin EK, Segara D, Scarlett CJ, Nguyen NQ, et al. Expression of s100a2 calcium-binding protein predicts response to pancreatectomy for pancreatic cancer. Gastroenterology 2009;137:558–68. 568 e551–511.
[42] Mahon PC, Baril P, Bhakta V, Chelala C, Caulee K, Harada T, et al. S100a4 contributes to the suppression of bnip3 expression, chemoresistance, and inhibition of apoptosis in pancreatic cancer. Cancer Res 2007;67:6786–95.
[43] Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17:500–3.
[44] Herpel E, Rocken C, Manke H, Schirmacher P, Flechtenmacher C. Quality management and accreditation of research tissue banks: experience of the national center for tumor diseases (nct) heidelberg. Virchows Arch Int J Pathol 2010;457:741–7.
[45] Oberlander M, Linnebacher M, Konig A, Bogoevska V, Brodersen C, Kaatz R, et al. The “north german tumor bank of colorectal cancer”: status report after the first 2 years of support by the german cancer aid foundation. Langenbecks Arch Surg 2013;398:251–8.
[46] Shickle D, Griffin M, El-Arifi K. Inter- and intra-biobank networks: classification of biobanks. Pathobiol J Immunopathol Mol Cell Biol 2010;77:181–90.
[47] Morente M. Tissue banks: who decides what is ethical? Eur J cancer 2004;40:5.
[48] Lorand S, Lavoue V, Tas P, Foucher F, Mesbah H, Rouquette S, et al. Intraoperative touch imprint cytology of axillary sentinel nodes for breast cancer: a series of 355 procedures. Breast 2011;20:119–23.
[49] Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, et al. Cell type-specific gene expression differences in complex tissues. Nat methods 2010;7:287–9.