

Specialized DNA Arrays for the Differentiation of Pancreatic Tumors

Malte Buchholz,¹ Hans A. Kestler,^{1,4} Andrea Bauer,⁵ Wolfgang Böck,¹ Bettina Rau,⁶ Gerhard Leder,² Wolfgang Kratzer,¹ Martin Bommer,³ Aldo Scarpa,⁷ Martin K. Schilling,⁶ Guido Adler,¹ Jörg D. Hoheisel,⁵ and Thomas M. Gress¹

Abstract Purpose: Malignant tumors of the pancreas are frequently indistinguishable from inflammatory tumors arising in the context of a chronic pancreatitis with the use of conventional imaging techniques. Thus, cytologic analysis of cells obtained by abdominal ultrasound, computed tomography, or endoscopic ultrasound – guided fine needle aspiration biopsy is required for diagnosis. However, the reliability of cytologic analyses of pancreatic fine needle aspirates remains unsatisfactory, with a diagnostic accuracy of $\leq 80\%$. The purpose of the current study was therefore to develop a novel diagnostic approach based on expression profiling of biopsy material using a specialized diagnostic cDNA array.

Experimental Design: Previous gene expression profiling studies were reevaluated to design a 558-feature diagnostic array. Minimal amounts of residual material from pancreatic cytology samples as well as surgically resected tumor and control tissue specimens were analyzed using the diagnostic array and a newly developed statistical classification system.

Results and Conclusions: Our diagnostic approach resulted in 95% accurate differentiation between ductal adenocarcinomas and nonmalignant tumors of the pancreas. The diagnostic array, in conjunction with conventional diagnostic procedures, is thus suitable to significantly improve the reliability of pancreatic cancer diagnostics and can be expected to become a valuable new tool in the routine workup of suspect masses in the pancreas.

Pancreatic cancer is the fifth leading cause of cancer-related deaths in industrialized countries. With a 5-year survival rate of $<5\%$ and a median survival of <6 months, pancreatic cancer carries the most dismal prognosis of all solid tumors. Chronic pancreatitis is a persistent inflammatory disease of the pancreas most often caused by alcohol abuse. In the course of chronic pancreatitis, inflammatory tumors may develop in the pancreas causing the same signs and symptoms as malignant pancreatic

tumors. Malignant and inflammatory tumors are frequently indistinguishable with the use of conventional imaging modalities such as computed tomography, abdominal, or endoscopic ultrasound, thus requiring cytologic analysis of cells obtained by abdominal ultrasound, computed tomography-, or endoscopic ultrasound-guided fine needle aspiration biopsy (FNAB). However, the reliability of the largely morphology-based cytologic analyses of fine needle aspirates of pancreatic tumors remains unsatisfactory with a diagnostic accuracy between 60% and 80% (1–5). Well-differentiated carcinomas may escape recognition because of the minimal cytologic atypia they display. Conversely, chronic pancreatitis may give rise to atypical cells that can be mistaken for neoplastic cells. For both malignant and benign tumors, diagnosis is extremely difficult when intact cells in the aspirate are rare or completely missing (Fig. 1).

It is well known that the process of carcinogenesis in the pancreas is associated with the accumulation of characteristic genetic changes within the cells of origin. Among the hallmark features of pancreatic ductal adenocarcinoma, which accounts for $>90\%$ of all malignant tumors in the pancreas, are mutations in the *K-ras* and *HER2/neu* oncogenes as well as the *p53*, *p16^{INK4a}*, and *SMAD4/DPC4* tumor suppressor genes (for an overview, see ref. 6). Based on these observations, several attempts have been made to improve the accuracy of preoperative diagnostics by analyzing molecular markers in pancreatic juice (7, 8), brush cytologies (7, 9), or FNAB's (10–12) by means of RNA, DNA, or protein analysis. Most of these studies were aimed at detecting mutant *K-ras* in the biopsy samples because this is the gene most frequently affected by mutations ($>85\%$ of cases) in pancreatic ductal

Authors' Affiliations: Departments of ¹Internal Medicine I, ²Visceral and Transplantation Surgery, ³Internal Medicine III, and ⁴Neuroinformatics, University of Ulm, Ulm, ⁵Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum, Heidelberg, ⁶Department of General, Visceral, and Vascular Surgery, University of the Saarland, Homburg/Saar, Germany, and ⁷Department of Pathology, University of Verona, Verona, Italy
Received 6/14/05; revised 7/28/05; accepted 8/10/05.

Grant support: German Cancer Research Foundation (Deutsche Krebshilfe/Mildred Scheel-Stiftung) grant no. 10-1473-Gr2 (T.M. Gress and J.D. Hoheisel) and supported in part by grants from the Deutsche Forschungsgemeinschaft (SFB 518 project B1, T.M. Gress), NGFN (01GR 0101, J.D. Hoheisel), Stifterverband für die Deutsche Wissenschaft (T.M. Gress and H.A. Kestler), Associazione Italiana Ricerca Cancro (A. Scarpa), Ministero Università e Ricerca (A. Scarpa), Fondazione Giorgio Zanotto (A. Scarpa), and the European Community (QLG1-CT-2002-01196, T.M. Gress and J.D. Hoheisel).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: M. Buchholz and H.A. Kestler contributed equally to this study and should both be considered first authors.

Requests for reprints: T.M. Gress, Abteilung Innere Medizin I, Universität Ulm, Robert-Koch-Str. 8, 89081 Ulm, Germany. Phone: 49-731-500-24385/24311; Fax: 49-731-500-24302; E-mail: thomas.gress@medizin.uni-ulm.de.

©2005 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-05-1274

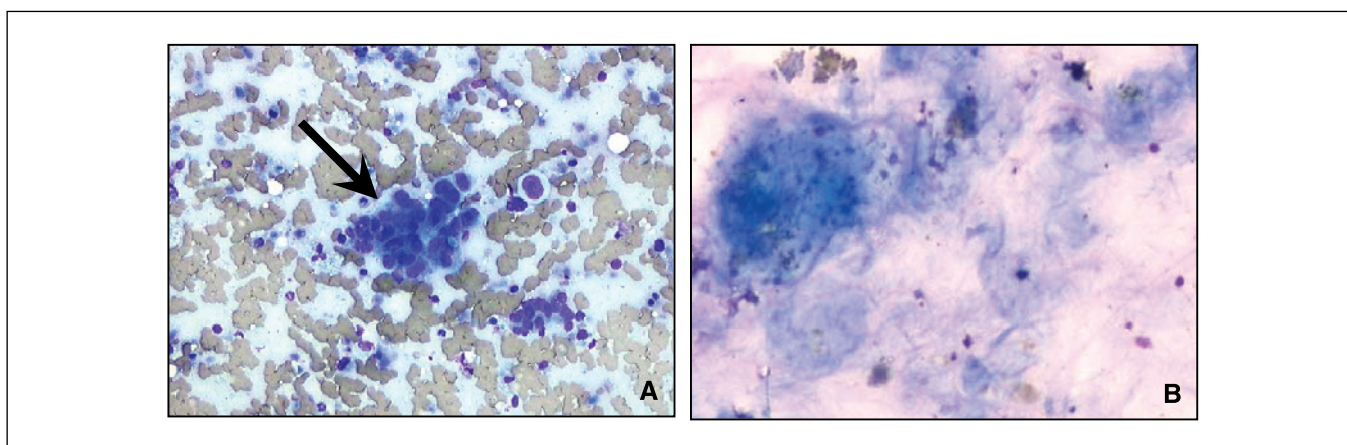


Fig. 1. Typical examples of a clearly malignant (*A*) and a nondiagnostic (*B*) finding illustrating the limitations of cytologic analyses. Both samples were obtained by FNAB of suspect pancreatic masses and analyzed within the course of this study. *A*, sample from "Biopsy_3" (see Table 1): malignant cells (*arrow*) showing nuclear overlapping, irregular contours, high nuclear/cytoplasmic ratios and anisocaryosis are readily detectable. *B*, sample from "Biopsy_13" (see Table 2): the specimen contains large amounts of mucus and cellular debris with few or no intact cells; original magnification, $\times 200$.

adenocarcinoma. However, *K-ras*-mutations were also detected in up to 25% of samples from chronic pancreatitis patients (7, 13), severely compromising the specificity of the test. Analyses of other single markers, including p53, CA19.9, SMAD4/DPC4, or Mucin expression, have likewise shown either low specificity, low sensitivity or, in the case of immunocytologic analyses, dependency on the presence of significant numbers of intact tumor cells.

Analysis of single molecular markers is therefore not sufficient to provide for accurate diagnosis of suspect pancreatic masses. DNA arrays with their potential to survey the expression levels of many genes simultaneously represent ideal tools to circumvent this problem. Several expression profiling analyses using different technological platforms (14–19) have shown the existence of distinct gene expression signatures characteristic of pancreatic cancer. However, the use of large-scale ("whole-genome") arrays is extremely costly and generates vast amounts of data which are difficult to analyze in a routine diagnostic setting. Both drawbacks can be circumvented by designing dedicated arrays with limited numbers of genes specifically selected for diagnostic purposes. The aim of this study was therefore to develop specialized cDNA arrays specifically designed for the differential diagnosis of pancreatic tumors based on expression profiling of fine needle aspiration biopsies. Because >90% of all malignant pancreatic tumors represent ductal adenocarcinomas (20), we focused on this tumor type for the present study.

Materials and Methods

Tissue and biopsy samples. For cytologic analysis of FNAB samples, aspirated material was expelled onto microscope slides and smeared. The slides were air-dried, stained with May-Grunwald-Giemsa stain, and evaluated by an expert cytologist. Samples were classified as "malignant" if cell aggregates showing clear signs of malignancy were detected on at least two different slides from the same biopsy, "benign" if evaluable material without any signs of malignancy were detected, or "nondiagnostic" if the results were inconclusive (e.g., the cellularity of the sample was too low).

Surgically resected pancreatic adenocarcinoma and chronic pancreatitis tissues were provided by the surgery departments at the

Universities of Ulm and Homburg/Saar. Normal pancreas samples were obtained from healthy areas at the borders of chronic pancreatitis resectates. Informed consent was obtained from all patients prior to using tissue or biopsy samples. The study was approved by the local ethics committees at the Universities of Ulm (Germany), Homburg/Saar (Germany), and Verona (Italy).

RNA isolation and linear amplification. Snap-frozen surgical samples were ground on dry ice with a mortar and pestle, suspended in RLT buffer and total RNA were isolated using the RNEasy Mini Kit (Qiagen, Hilden, Germany). Fine needle biopsy in the routine diagnostic workup of pancreatic tumors was done with transabdominal ultrasound or endoscopic ultrasound guidance. FNAB sample material was recovered by flushing the needle and syringe with RLT buffer after material for cytologic analysis had been removed. Total RNA was then isolated using the RNEasy Mini Kit (Qiagen). In both cases, the total RNA was finally dissolved in water and quality-checked on a BioAnalyzer Lab-on-a-Chip system (Agilent, Waldbronn, Germany). In order to obtain sufficient material for hybridization, the complete FNAB RNA samples were subjected to one round of T7 RNA polymerase-based linear amplification using the MessageAmp Kit (Ambion, Huntingdon, Great Britain). To avoid data bias, all surgical samples were treated likewise by linearly amplifying 0.5 μ g of total RNA prior to hybridization.

Array production and hybridization. cDNA fragments were PCR-amplified using vector primers and contact-printed in duplicate on nylon membranes (Nytran N+, Schleicher and Schuell, Germany). For radioactive labeling, the complete amplified RNA samples were labeled with 32 P-dATP using the StripEZ-RT Kit (Ambion) and hybridized overnight to nylon membrane arrays in ULTRArray hybridization buffer (Ambion) at 50°C. Radioactive signals were detected using a STORM phosphorimaging system (Amersham Biosciences, Feiburg, Germany) and quantified with the ArrayVision software (InterFocus, Haverhill, Great Britain). Signal intensities were normalized to the mean signal intensity of all features on an individual array.

Construction of the classifier. All equations used are listed in Panel 1. Details of the analysis and complete data sets are available as part of the Supplementary Data.⁸ The nylon array data set was filtered to include only features (genes) for which normalized intensities exceeded a value of 0.8 in at least 10 samples to remove uniformly low (and thus uninformative) signals. Control spots were excluded as well. The remaining 169 features were used in the construction of a linear classifier based on the analysis of the 42-sample training set (Table 1).

⁸ Supplementary information is accessible at <http://www.informatik.uni-ulm.de/ni/mitarbeiter/HKestler/DiagArray/default.html>.

During the first step of the analysis, a principal component analysis (21) was done on the training data set. The first 30 principal components of the training data set, which represented 99.9% of the total variation within the data, were then used for a linear discriminant analysis to search for combinations of principal components facilitating complete separation of the tumor from the control tissue samples in the training set. All possible combinations of up to 7 out of the 30 principal components were tested for their performance in the linear separation of the diagnostic classes. Evaluation of the feature set combinations was done by measuring the area under the receiver operating characteristic curve (22–24). All combinations producing an area under the receiver operating characteristic curve of ≥ 0.95 were subjected to a stochastic search to add additional discriminative principal components until perfect separation of the diagnostic classes was achieved. Out of all

combinations producing perfect linear separation, we selected the set that resulted in the greatest margin between tumor and control samples when plotting the samples according to their relative distances to the separating hyperplane. The resulting linear classifier was then evaluated using the independent 20-sample test set (Table 2).

Results and Discussion

In order to develop the pancreatic cancer diagnostic cDNA array, we extensively analyzed the results of various studies on differential gene expression in pancreatic cancer done in our own group (25, 26) as well as information obtained from the SAGE (<http://www.ncbi.nlm.nih.gov/SAGE/>) and Digital

Table 1. Composition of the 42-sample training set

| Sample | Diagnosis | Tumor-node-metastasis classification | Cytologic diagnosis | Final diagnosis (follow up) |
|--------------|----------------------|--|---------------------|-----------------------------|
| Biopsy_1 | — | — | malignant | adenocarcinoma |
| Biopsy_3 | — | — | malignant | adenocarcinoma |
| Biopsy_4 | — | — | malignant | adenocarcinoma |
| Biopsy_5 | — | — | malignant | adenocarcinoma |
| Biopsy_6 | — | — | malignant | adenocarcinoma |
| Biopsy_8 | — | — | nondiagnostic | adenocarcinoma |
| Biopsy_10 | — | — | malignant | adenocarcinoma |
| Biopsy_12 | — | — | benign | pseudocyst |
| Biopsy_13 | — | — | nondiagnostic | chronic pancreatitis |
| Biopsy_14 | — | — | nondiagnostic | chronic pancreatitis |
| Biopsy_16 | — | — | benign | pseudocyst |
| Tumor_1 | adenocarcinoma | T ₃ N ₁ M ₁ | — | — |
| Tumor_2 | adenocarcinoma | T ₃ N _x M _x | — | — |
| Tumor_3 | adenocarcinoma | T ₃ N ₁ M _x | — | — |
| Tumor_4 | adenocarcinoma | T ₃ N ₁ M _x | — | — |
| Tumor_5 | adenocarcinoma | T ₃ N ₁ M _x | — | — |
| Tumor_6 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_7 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_8 | adenocarcinoma | T ₃ N ₀ M ₁ | — | — |
| Tumor_9 | adenocarcinoma | T ₂ N ₁ M ₁ | — | — |
| Tumor_19 | adenocarcinoma | T ₂ N ₁ M ₀ | — | — |
| Tumor_20 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_21 | adenocarcinoma | T ₂ N ₀ M _x | — | — |
| Tumor_22 | adenocarcinoma | T ₂ N ₀ M _x | — | — |
| Tumor_23 | adenocarcinoma | T ₃ N ₀ M ₀ | — | — |
| Tumor_24 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_25 | adenocarcinoma | T ₃ N ₁ M ₁ | — | — |
| Tumor_26 | adenocarcinoma | T ₃ N ₀ M _x | — | — |
| Tumor_27 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Infl/Norm_1 | chronic pancreatitis | — | — | — |
| Infl/Norm_3 | chronic pancreatitis | — | — | — |
| Infl/Norm_4 | chronic pancreatitis | — | — | — |
| Infl/Norm_6 | chronic pancreatitis | — | — | — |
| Infl/Norm_7 | chronic pancreatitis | — | — | — |
| Infl/Norm_9 | chronic pancreatitis | — | — | — |
| Infl/Norm_10 | normal pancreas | — | — | — |
| Infl/Norm_12 | chronic pancreatitis | — | — | — |
| Infl/Norm_13 | normal pancreas | — | — | — |
| Infl/Norm_16 | chronic pancreatitis | — | — | — |
| Infl/Norm_17 | normal pancreas | — | — | — |
| Infl/Norm_18 | chronic pancreatitis | — | — | — |
| Infl/Norm_19 | chronic pancreatitis | — | — | — |

Table 2. Composition of the 20-sample test set

| Sample | Diagnosis | Tumor-node-metastasis classification | Cytologic diagnosis | Final diagnosis (follow up) |
|--------------|----------------------|--|---------------------|-----------------------------|
| Biopsy_2 | — | — | malignant | adenocarcinoma |
| Biopsy_7 | — | — | malignant | adenocarcinoma |
| Biopsy_9 | — | — | malignant | adenocarcinoma |
| Biopsy_11 | — | — | benign | pseudocyst |
| Biopsy_15 | — | — | nondiagnostic | chronic pancreatitis |
| Tumor_10 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_11 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_12 | adenocarcinoma | T ₂ N ₀ M ₀ | — | — |
| Tumor_13 | adenocarcinoma | T ₃ N ₀ M ₀ | — | — |
| Tumor_14 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_15 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_16 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Tumor_17 | adenocarcinoma | T ₄ N ₀ M ₀ | — | — |
| Tumor_18 | adenocarcinoma | T ₃ N ₁ M ₀ | — | — |
| Infl/Norm_2 | chronic pancreatitis | — | — | — |
| Infl/Norm_5 | normal pancreas | — | — | — |
| Infl/Norm_8 | chronic pancreatitis | — | — | — |
| Infl/Norm_11 | normal pancreas | — | — | — |
| Infl/Norm_14 | normal pancreas | — | — | — |
| Infl/Norm_15 | chronic pancreatitis | — | — | — |

Differential Display (http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.shtml) gene expression databases and reports from the literature to identify genes with the potential to differentiate between malignant and nonmalignant tumors of the pancreas. In order to allow for robust normalization of the hybridization results, we have designed the array to comprise a sufficiently high total number of features (558), including balanced numbers of up- and down-regulated genes. In addition, important genes were represented by multiple cDNA clones and control spots of mixed cDNA clones were included to facilitate grid alignment (for a complete list of features, see Supplementary Information).

In the present study, 16 FNAB samples of pancreatic adenocarcinoma and benign pancreatic tumors for which clinical patient follow up with a definitive diagnosis was available were analyzed both by conventional cytology and diagnostic array hybridization. Cytologic analysis of the 16 FNAB samples correctly identified a malignant process in 9 out of 10 adenocarcinoma cases (90%) and a benign process in 3 out of 6 chronic pancreatitis and pseudocyst cases (50%). The remaining adenocarcinoma case as well as the three benign cases were nondiagnostic due to the absence of evaluable intact cells. The resulting overall diagnostic accuracy of 75% is well in agreement with the numbers reported in the literature (1–4).

Residual material from the same biopsy samples which were used for cytologic analyses were subjected to expression profiling analyses using the diagnostic arrays. In order to ensure an adequate representation of different tumor stages in the data sets used for the development and evaluation of the classification procedure, we analyzed an additional 27 samples of histopathologically well-defined surgically resected ductal adenocarcinomas as well as 19 surgically resected control samples of chronic pancreatitis or normal pancreas. The samples were arbitrarily divided into a 42-sample training set (Table 1) and a 20-sample test set (Table 2), such that both sets

contained equal proportions of malignant and benign samples as well as FNAB's. The training set was subsequently used to develop the classification system for the distinction between malignant and benign samples (see below), which was independently evaluated using the test sample set. The complete process is schematically outlined in Fig. 2.

Even though the number of genes featured on the diagnostic array was low compared with large-scale arrays, it still far exceeded the number of tissue and biopsy samples available for training of the classifier in the training set. It was therefore of paramount importance to first reduce the number of features used for classification and thus the dimensionality of the data set in order to avoid overadaptation of the classifier to this specific set of data (27). Instead of omitting individual genes from the analysis to achieve this purpose, we opted to apply principal component analysis (21) to the data, resulting in a reduced set of combined features (principal components) representing weighted combinations of all genes in the data set. Tissue or biopsy samples can be mapped to the principal components (or a subset thereof), effectively creating a coordinate system of uncorrelated variables which replace the high dimensional space that individual gene expression values fall into. Principal component analyses thus serves to greatly reduce the dimensionality of the data whereas preserving its general structure, resulting in reduced sensitivity to outliers or hybridization artifacts in individual diagnostic samples.

For the construction of the classifier, we opted to perform linear discriminant analysis using the first 30 principal components of the training data set, which represented 99.9% of the total variation within the data. Because linear discriminant analysis assumes a relatively simple model of sample distribution, it is far less prone to overadaptation to a specific data set than nonlinear methods, again increasing the robustness of the classification procedure. We identified a total of 429,917 different combinations of principal components

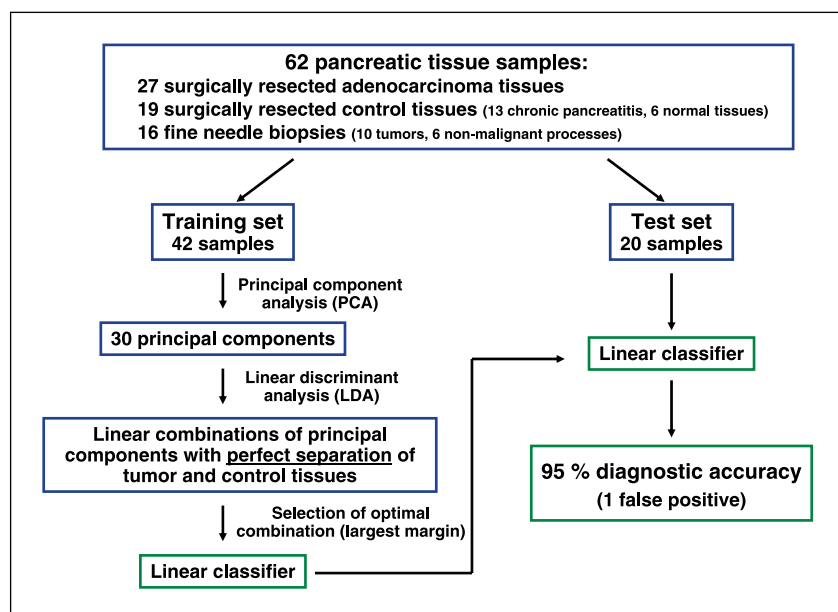


Fig. 2. Flowchart outlining the process of construction and evaluation of the linear classifier using the independent training and test sample sets.

producing perfect linear separation of tumor and control samples in the training data set. Out of these, a set of 23 principal components which provided the maximum margin between tumor and control samples was selected to define the linear classifier (Fig. 3A; see also Supplementary Information).

The predictive performance of the linear classifier was then evaluated by assessing its performance on the independent 20-sample test set. Our system correctly classified 19 out of the total 20 test samples, resulting in an overall diagnostic accuracy of 95% (Fig. 3B). Only one surgically resected chronic pancreatitis sample ("Infl/Norm_15"; Table 2) was

misclassified, producing one false-positive result. Of note, all FNAB samples were correctly classified regardless of the outcome of the cytologic analysis. Moreover, all FNAB samples were placed well clear of the separating cutoff by the classifier (Fig. 3B).

We were thus able to show that expression profiling analyses of surgically resected tumor specimens and FNAB samples using specialized diagnostic arrays with limited numbers of highly selected genes produce reliable, reproducible, and informative results. Supplementing conventional cytologic analyses of aspiration biopsies with diagnostic array profiling

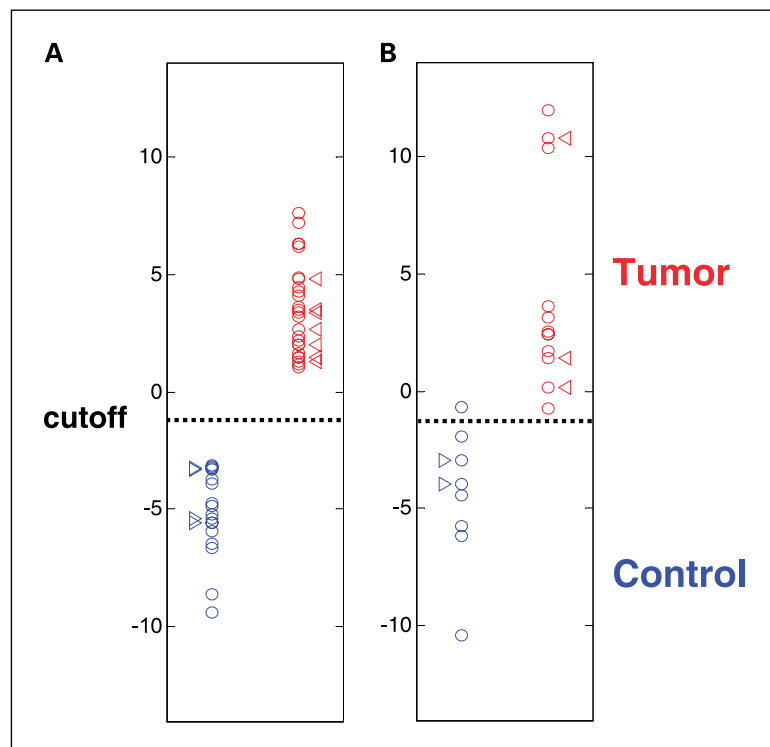


Fig. 3. Separation of the adenocarcinoma (red circles) and control (blue circles) samples in the training and test data sets using the selected combination of 23 principal components. The samples are plotted according to their relative distances to the separating hyperplane. Dotted lines, cutoff calculated by linear discriminant analysis. FNAB samples are identified by triangles next to the corresponding circles. A, perfect linear separation of the diagnostic classes (including surgically resected tissue as well as FNAB samples) in the training set. B, 95% diagnostic accuracy using the linear classifier on the independent test set. One surgically resected chronic pancreatitis sample is misclassified; all FNAB samples are correctly classified.

thus promises to significantly improve the accuracy of preoperative diagnostics of suspect masses in the pancreas. The large number of different combinations of principal components yielding perfect linear separation of the diagnostic classes in the training set, as well as the convincing performance of the classifier on the independent test set, serve to show both the expedience of the diagnostic gene collection as well as the validity of the analytic approach. Our results confirm conclusions drawn from earlier expression profiling studies using large-scale arrays, which have shown that the number of informative genes for the classification of different types and subtypes of cancer is usually <100 (28–30), suggesting that dedicated diagnostic arrays should perform as well as whole-genome arrays in defined diagnostic settings.

Due to the use of residual material from biopsy needles for the analysis of the FNAB samples, the amount of starting material available for expression profiling analysis was extremely limited, so that we initially produced the arrays in the nylon membrane format to take advantage of the superior sensitivity of radioactive labeling and detection. Parallel hybridizations of a subset of samples to diagnostic arrays produced in the glass microarray format (see Supplementary Material), however, showed that the concept and design of the diagnostic array can readily be transferred to the glass microarray/fluorescent labeling platform as well, which may be better suited for routine clinical settings.

In the present study, we have focused on the distinction between pancreatic ductal adenocarcinoma and nonmalignant diseases of the pancreas, because pancreatic ductal adenocarcinoma is by far the most frequent malignant tumor arising in the pancreas and thus poses the clinically most relevant diagnostic problem (20). We are currently in the process of analyzing additional tumor entities, such as acinar and neuroendocrine tumors, using both the diagnostic array as well as large-scale arrays, in order to develop a multiclass classification system for the comprehensive diagnosis of different malignancies in the pancreas. In addition, we expect further development of the array in combination with careful analysis of clinical patient data to result in the recognition of

distinct prognostic gene expression signatures predicting important clinical variables such as stage of disease, response to therapy, or prognosis, thus setting the stage for therapeutic regimens custom tailored to the individual patient.

Appendix. Panel 1 (Equations used in the report)

Given is a set of n d -dimensional samples x_1, \dots, x_n , n_1 in the subset D_1 and n_2 in the subset D_2 .

The sample mean $\hat{\mu}_i$ is estimated by:

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{x \in D_i} x$$

The scatter matrices S_i and S_w are:

$$S_i = \sum_{x \in D_i} (x - \mu_i)(x - \mu_i)^T, S_w = S_1 + S_2$$

The projection vector w for Fisher's linear discriminant is given by:

$$w = S_w^{-1}(\mu_1 - \mu_2)$$

The area under the receiver operating characteristic curve is estimated by:

$$\hat{A} = \sum_{j=1}^{n_2} (\Xi_j - j) / n_1 n_2$$

where n_2 is the number of pancreatic carcinoma patients, and Ξ_j ($j = 1, \dots, n_2$) are the ranks of these cases obtained by ranking all $n = n_1 + n_2$ values of $w^T x$.

The margin m for $\hat{A} = 1$ is calculated by:

$$m = \begin{cases} \min(w^T x_j) - \max(w^T x_i) & \text{if } w^T x_i < w^T x_j, \\ \min(w^T x_i) - \max(w^T x_j) & \text{if } w^T x_i > w^T x_j, \end{cases}$$

where $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, and the threshold is centered in the margin.

References

1. Kehagias D, Smyrniotis V, Kalovidouris A, et al. Cystic tumors of the pancreas: preoperative imaging, diagnosis, and treatment. *Int Surg* 2002;87:171–4.
2. Brandwein SL, Farrell JJ, Centeno BA, Brugge WR. Detection and tumor staging of malignancy in cystic, intraductal, and solid tumors of the pancreas by EUS. *Gastrointest Endosc* 2001;53:722–7.
3. Chhieng DC, Jhala D, Jhala N, et al. Endoscopic ultrasound-guided fine-needle aspiration biopsy: a study of 103 cases. *Cancer* 2002;96:232–9.
4. Voss M, Hammel P, Molas G, et al. Value of endoscopic ultrasound guided fine needle aspiration biopsy in the diagnosis of solid pancreatic masses. *Gut* 2000;46:244–9.
5. Afify AM, al Khafaji BM, Kim B, Scheiman JM. Endoscopic ultrasound-guided fine needle aspiration of the pancreas. Diagnostic utility and accuracy. *Acta Cytol* 2003;47:341–8.
6. Li D, Xie K, Wolff R, Abbruzzese JL. Pancreatic cancer. *Lancet* 2004;363:1049–57.
7. Pugliese V, Pujic N, Saccomanno S, et al. Pancreatic intraductal sampling during ERCP in patients with chronic pancreatitis and pancreatic cancer: cytologic studies and k-ras-2 codon 12 molecular analysis in 47 cases. *Gastrointest Endosc* 2001;54:595–9.
8. Wang Y, Yamaguchi Y, Watanabe H, et al. Detection of p53 gene mutations in the supernatant of pancreatic juice and plasma from patients with pancreatic carcinomas. *Pancreas* 2004;28:13–9.
9. Van Laethem JL, Bourgeois V, Parma J, et al. Relative contribution of Ki-ras gene analysis and brush cytology during ERCP for the diagnosis of biliary and pancreatic diseases. *Gastrointest Endosc* 1998;47:479–85.
10. Urgell E, Puig P, Boadas J, et al. Prospective evaluation of the contribution of K-ras mutational analysis and CA 19.9 measurement to cytological diagnosis in patients with clinical suspicion of pancreatic cancer. *Eur J Cancer* 2000;36:2069–75.
11. Chhieng DC, Benson E, Eltoun I, et al. MUC1 and MUC2 expression in pancreatic ductal carcinoma obtained by fine-needle aspiration. *Cancer* 2003;99:365–71.
12. Pinto MM, Emanuel JR, Chaturvedi V, Costa J. Ki-ras mutations and the carcinoembryonic antigen level in fine needle aspirates of the pancreas. *Acta Cytol* 1997;41:427–34.
13. Lohr M, Muller P, Mora J, et al. p53 and K-ras mutations in pancreatic juice samples from patients with chronic pancreatitis. *Gastrointest Endosc* 2001;53:734–43.
14. Gress TM, Muller-Pillasch F, Geng M, et al. A pancreatic cancer-specific expression profile. *Oncogene* 1996;13:1819–30.
15. Crnogorac-Jurcovic T, Efthimiou E, Nielsen T, et al. Expression profiling of microdissected pancreatic adenocarcinomas. *Oncogene* 2002;21:4587–94.
16. Han H, Bearss DJ, Browne LW, et al. Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. *Cancer Res* 2002;62:2890–6.
17. Nakamura T, Furukawa Y, Nakagawa H, et al. 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;23:2385–400.
18. Iacobuzio-Donahue CA, Ashfaq R, Maitra A, et al. Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from

- three major technologies. *Cancer Res* 2003;63:8614–22.
19. Maitra A, Hansel DE, Argani P, et al. Global expression analysis of well-differentiated pancreatic endocrine neoplasms using oligonucleotide microarrays. *Clin Cancer Res* 2003;9:5988–95.
20. Carriaga MT, Henson DE. Liver, gallbladder, extrahepatic bile ducts, and pancreas. *Cancer* 1995;75:171–90.
21. Jolliffe IT. *Principal component analysis*. 2nd ed. New York: Springer; 2002.
22. Kestler HA. ROC with confidence—a Perl program for receiver operator characteristic curves. *Comput Methods Programs Biomed* 2001;64:133–6.
23. Lang TA, Secic M. *How to report statistics in medicine*. Philadelphia (PA): American College of Physicians; 1997.
24. Swets JA, Pickett RM. *Evaluation of diagnostic systems*. New York: Academic Press; 1982.
25. Gress TM, Wallrapp C, Frohme M, et al. Identification of genes with specific expression in pancreatic cancer by cDNA representational difference analysis. *Genes Chromosomes Cancer* 1997;19:97–103.
26. Geng M, Wallrapp C, Muller-Pillasch F, et al. Isolation of differentially expressed genes by combining representational difference analysis (RDA) and cDNA library arrays. *Biotechniques* 1998;25:434–8.
27. Cover TM. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans Electronic Comput* 1965;14:326–34.
28. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7:673–9.
29. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
30. 'tVeer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.