RESEARCH ARTICLE

Detection of pancreatic cancer using antibody microarray-based serum protein profiling

Johan Ingvarsson[1, 2], Christer Wingren[1, 2], Anders Carlsson[1, 2], Peter Ellmark[1, 2], Britta Wahren[3], Gunnel Engström[3], Ulrika Harmenberg[4], Morten Krogh[5], Carsten Peterson[5] and Carl A. K. Borrebaeck[1, 2]

[1] Department of Immunotechnology, Lund University, Lund, Sweden
[2] CREATE Health, Lund University, Lund, Sweden
[3] Department of Virology, Swedish Institute for Infectious Disease Control, Karolinska Institute, Stockholm, Sweden
[4] Radiumhemmet, Karolinska Hospital, Stockholm, Sweden
[5] Department of Theoretical Physics, Lund University, Lund, Sweden

The driving force behind oncoproteomics is to identify protein signatures that are associated with a particular malignancy. Here, we have used a recombinant scFv antibody microarray in an attempt to classify sera derived from pancreatic adenocarcinoma patients *versus* healthy subjects. Based on analysis of nonfractionated, directly labeled, whole human serum proteomes we have identified a protein signature based on 19 nonredundant analytes, that discriminates between cancer patients and healthy subjects. Furthermore, a potential protein signature, consisting of 21 protein analytes, could be defined that was shown to be associated with cancer patients having a life expectancy of <12 months. Taken together, the data suggest that antibody microarray analysis of complex proteomes will be a useful tool to define disease associated protein signatures.

Received: December 18, 2007
Revised: February 4, 2008
Accepted: February 12, 2008

Keywords:
Cancer-associated signature / Expression profiling proteome analysis / Oncoproteomics / Pancreatic cancer / Recombinant antibody microarray

1 Introduction

Recent progress in proteomics has opened up improved cancer-associated biomarker discovery [1, 2], although high-throughput proteomic approaches targeting complex biological fluids, such as serum, have proven to be challenging [2–4]. Antibody-based microarrays is a rapidly emerging affinity-proteomic technology that is likely to play an increasing role within oncoproteomics [3] and that has made significant progress in recent years, for review see [5, 6]. This approach is capable of simultaneously profiling numerous low-abundant protein analytes in complex proteomes, while consuming only microliter amounts of samples [7–10]. Translational proteomics is one immediate application for antibody microarrays, where comparative protein expression profiling of cancer *versus* normal proteomes might yield tentative predictive biomarker signatures.

In oncology, one major challenge is the ability to stratify patients, relating to their probability to experience tumor relapse or drug treatment resistance, or to their survival expectancy. Although gene expression profiling of cancer has, in a few cases, demonstrated the ability to predict time of survival [11, 12], no serum signature, *i.e.*, a combination of serum proteins, has so far been associated with any of the

Correspondence: Professor Carl A. K. Borrebaeck, Department of Immunotechnology, BMC D13, SE-221 84 Lund, Sweden
E-mail: carl.borrebaeck@immun.lth.se
Fax: +46-46-222-4200

Abbreviations: IL, interleukin; **MCP,** monocyte chemotactic protein; **PA,** pancreatic adenocarcinoma; **ROC,** receiver operator characteristic; **scFv,** single-chain fragment variable; **SVM,** support vector machine

above clinical parameters, although single inflammatory analytes such as C-reactive protein (CRP) have been shown to correlate well with patient survival [13]. Antibody microarrays [3] have been suggested as the technology platform that eventually could deliver a defined protein biomarker signature that distinguish cancer from normal patients. Furthermore, a serum sample analysis that can predict survival time would allow for improved individualized cancer therapy. This has been emphasized for *e.g.*, pancreatic adenocarcinomas (PAs), where no tumor-specific markers exist, and although most patients will have an elevated cancer antigen 19-9 (CA 19-9) at the time of diagnosis, individual prognostic markers have shown to be inconclusive [14, 15]. Furthermore, noninvasive approaches, such as computed tomography, are not sufficiently sensitivity to detect small tumors, whereas *e.g.*, endoscopic ultrasonography can be used to survey at-risk individuals for pancreatic lesions [16]. Pancreatic ductal adenocarcinoma is also the most lethal malignancy by anatomic site, with >30 000 new cases and deaths annually in the United States alone, and with a 5 year survival of 3–5%. This extreme mortality is due to the lack of effective early diagnostic methods and to poor efficacy of existing therapies for advanced disease [16]. Even the patients (10–20%) diagnosed with a surgically resectable tumor, ultimately die of recurrent and metastatic disease.

Consequently, an increased ability to detect and predict cancer is crucial for individual patient management, and to increase the analytical resolution we need to adopt novel high-throughput technologies. The present study describes an affinity proteomic attempt to explore differences in serum protein content in cancer patients *versus* healthy subjects, based on a recombinant antibody microarray, containing array adapted single-chain fragment variable (scFv) fragments [17, 18]. The results of this proof-of-concept study demonstrated that an array of antibody fragments, specific for immunoregulatory proteins, could discriminate between human serum proteomes derived from either PA patients or healthy individuals, based on a signature of 19 non-redundant serum proteins.

2 Materials and methods

2.1 Production and purification of scFv

One hundred twenty-nine human recombinant scFv antibody fragments against sixty different proteins mainly involved in immunoregulation, were stringently selected from the *n*-CoDeR library [18] and kindly provided by BioInvent International AB (Lund, Sweden). Thus, each antigen was recognized by up to four different scFv fragments. All scFv antibodies were produced in 100 mL *Escherichia coli* cultures and purified from expression supernatants, using affinity chromatography on Ni-NTA agarose (Qiagen, Hilden, Germany). Bound molecules were eluted with 250 mM imidazole, extensively dialyzed against PBS, and stored at 4°C, until further use. The protein concentration was determined by measuring the absorbance at 280 nm (average concentration 210 μg/mL, range 60–1090 μg/mL). The purity was evaluated by 10% SDS-PAGE (Invitrogen, Carlsbad, CA, USA).

2.2 Serum samples

In total, 44 serum samples were collected, using standard procedures, at Stockholm South General Hospital (Sweden) and Lund University Hospital (Lund, Sweden). Serum samples (24, PA1–PA30) were collected from patients with pancreatic ductal adenocarcinoma before initiation of therapy, whereas 20 serum samples (N1–N20) (no clinical symptoms) were collected from healthy subjects. Patient demographics are shown in Table 1. All samples were aliquoted and stored at −80°C.

Table 1. Patients demographics

Class	No.	Gender	Age	
			Mean (SD)	Range
PA*	10	M	74 (8)	60–85
	14	F	69 (14)	31–82
Normal	18	M	49 (23)	22–85
	2	F	28 (1)	27–29
All	44	M/F	61 (21)	22–85

PA* = pancreatic adenocarcinoma.

2.3 Labeling of serum samples

The serum samples were labeled using previously optimized labeling protocols for serum proteomes [7, 10, 19]. All serum samples were biotinylated using the EZ-Link Sulfo-NHS-LC-Biotin (Pierce, Rockford, IL, USA). Serum aliquots (50 μL) were centrifuged at 16 000 × *g* for 20 min at 4°C and diluted 1:45 in PBS, resulting in a concentration of about 2 mg/mL. The samples were then biotinylated by adding sulfo-NHS-biotin to a final concentration of 0.6 mM for 2 h on ice, with careful Vortexing every 20 min. Unreacted biotin was removed by dialysis against PBS for 72 h, using a 3.5 kDa MW dialysis membrane (Spectrum Laboratories, Rancho Dominguez, CA, USA). The samples were aliquoted and stored at −20°C.

2.4 ELISA

The serum concentration of four protein analytes monocyte chemotactic protein (MCP-3, interleukin (IL-4, IL-5, and IL-13)) were measured in all samples, using commercial ELISA kits (Quantikine, R&D Systems, Minneapolis, MN, USA). The measurements were performed according to the recommendations provided by the supplier.

2.5 Fabrication and processing of antibody microarrays

For production of the antibody microarrays, we used a set-up previously optimized and validated [7, 10, 17, 19]. Briefly, the scFv microarrays were fabricated, using a noncontact printer (Biochip Arrayer, PerkinElmer Life & Analytical Sciences), which deposits approximately 330 pL/drop, using piezo technology. The scFv antibodies were arrayed by spotting two drops at each position and the first drop was allowed to dry out before the second drop was dispensed. The antibodies were spotted onto black polymer MaxiSorb microarray slides (NUNC A/S, Roskilde, Denmark), resulting in an average of 5 fmol scFv *per* spot (range 1.5–25 fmol). Eight replicates of each scFv clone were arrayed to ensure adequate statistics. The on-chip performances of the antibodies have been previously validated *e.g.*, [7, 10, 17, 20], for review see [6, 21]. In total, 160 antibodies and controls were printed *per* slide orientated in two columns with 8×80 antibodies *per* column. To assist the alignment of the grid during the quantification a row containing Cy5 conjugated streptavidin (2 μg/mL) was spotted for every tenth row. A hydrophobic pen (DakoCytomation Pen, DakoCytomation, Glostrup, Denmark) was used to draw a hydrophobic barrier around the arrays. The arrays were blocked with 500 μL 5% w/v fat-free milk powder (Semper AB, Sundbyberg, Sweden) in PBS overnight. All incubations were conducted in a humidity chamber at room temperature. The arrays were then washed four times with 400 μL 0.05% Tween-20 in PBS (PBS-T), and incubated with 350 μL biotinylated serum, diluted 1:10 (resulting in a total serum dilution of 1:450) in 1% w/v fat-free milk powder, and 1% Tween in PBS (PBS-MT) for 1 h. Next, the arrays were washed four times with 400 μL PBS-T and incubated with 350 μL, 1 μg/mL Alexa-647 conjugated streptavidin, diluted in PBS-MT for 1 h. Finally, the arrays were washed four times with 400 μL PBS-T, dried immediately under a stream of nitrogen gas and scanned with a confocal microarray scanner (ScanArray Express, PerkinElmer Life and Analytical Sciences) at 5 μm resolution, using six different scanner settings. The ScanArray Express software V2.0 (PerkinElmer Life and Analytical Sciences) was used to quantify the intensity of each spot, using the fixed circle method. The local background was subtracted and to compensate for possible local defects, the two highest, and the two lowest replicates were automatically excluded and each data point represents the mean value of the remaining four replicates. The coefficient of correlation for intraassays was >0.99 and for interassays >0.96, respectively.

2.6 Data normalization

Only nonsaturated spots were used for further analysis of the data. Chip-to-chip normalization of the datasets was performed, using a semiglobal normalization approach, conceptually similar to the normalization developed for DNA microarrays. Thus, the CV was first calculated for each scFv and ranked. Fifteen percentage of the scFv antibodies that displayed the lowest CV-values over all samples were identified, corresponding to 21 scFvs, and used to calculate a chip-to-chip normalization factor. The normalization factor N_i was calculated by the formula $N_i = S_i/\mu$, where S_i is the sum of the signal intensities for the 21 scFvs for each sample and μ is the sum of the signal intensities for the 21 scFvs averaged over all samples. Each dataset generated from one sample was divided with the normalization factor N_i. For the intensities, log2 values were used in the analysis.

2.7 Data analysis

All statistics and data analysis was performed in the statistical language *R* [22]. The Sammon map, which is an unsupervised analysis, was performed using Euclidean distance in the space of all 129 analytes. Supervized classification was done with a support vector machine (SVM) using a linear kernel (Chih-chung, C., Chih-Jen, L., LIBSVM: a library for support vector machines. 2007 http//:www.csie.ntu.edu.tw/cjlin/libsvm) [23, 24]. A SVM attempts to find a hyperplane that separates the two groups in the trainng set. A test sample is then classified depending on which side of the hyperplane the sample is located. The cost of constraints violation (the parameter C in the SVM) was fixed to one, which is the default value in the *R* function SVM, and no attempt was done to tune it. This absence of parameter tuning was chosen to avoid overfitting and to make the classification procedure more transparent. The output of the SVM on a test sample is a SVM decision value, which is the signed distance of the sample to the hyperplane. In Figs. 1C and 2C, the split into training and test set was done randomly once and kept fixed from thereon. In Fig. 2A, a leave-one-out cross-validation procedure is used. For every number K between 1 and 129 the following procedure was carried out. For a training set, *i.e.*, all samples except one, the K highest ranked analytes with a Wilcoxon test were chosen, and a SVM was trained with those K analytes. A SVM decision value was then calculated for the left out sample with this classifier. As is common practice, this was done for all samples in the leave-one-out crossvalidation. Both variants, which are common practice, the fixed test set and leave-one-out cross-validation procedures, ensure that the resulting performance of the classifier can be regarded as blinded tests.

A receiver operating characteristics (ROC) curve was constructed using the SVM decision values and the area under the curve (AUC) was found. The ROC area was plotted as a function of K.

3 Results

One approach to improved diagnosis of pancreatic cancer would be to identify a set of biomarkers that is associated with the malignancy. In an attempt to identify such a protein signature linked to pancreatic cancer, we have designed the

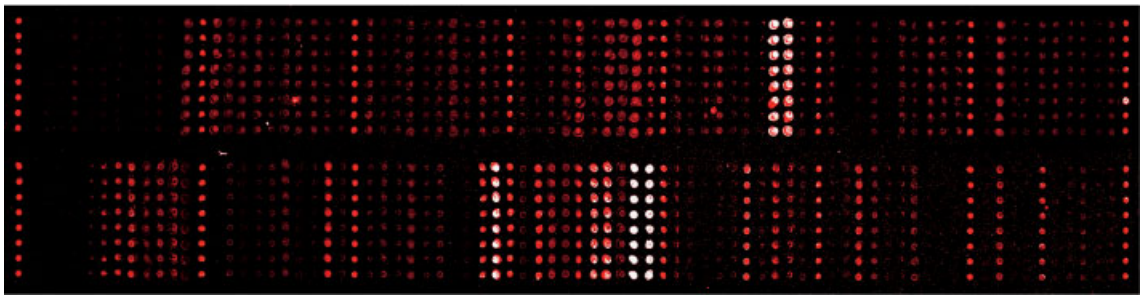

Figure 1. Scanned microarray image for a PA serum sample analyzed using recombinant antibody microarrays. The microarray is composed of 129 recombinant scFv antibodies dispensed in eight replicates, each array providing 1280 data points (including controls).

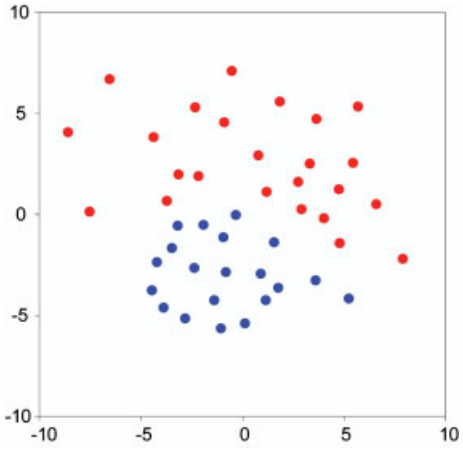

Figure 2. Detection of PAs by serum protein expression analysis using recombinant antibody microarrays. A multidimensional analysis represented as an unsupervised Sammon plot based on all 129 antibodies, directed against 60 different serum proteins, where cancer patients (red, $n = 24$) are shown to be completely separated from healthy subjects (blue, $n = 20$).

Table 2. The different scFv specificities used for the antibody microarray

Antigen (no. of clones)	Antigen (no. of clones)
IL-1α (3)	GLP-1 (1)
IL-1β (3)	GLP-1R (1)
IL-1-ra (3)	C1q (1)
IL-2 (3)	C1s (1)
IL-3 (3)	C3 (2)
IL-4 (4)	C4 (1)
IL-5 (3)	C5 (2)
IL-6 (4)	Factor B (1)
IL-7 (2)	B (1)
IL-8 (3)	Properdin (1)
IL-9 (3)	C1-INH (1)
IL-10 (3)	CD40L (1)
IL-11 (3)	PSA (1)
IL-12 (4)	Leptin (1)
IL-13 (3)	LDL (2)
IL-16 (3)	Integrin α10 (1)
IL-18 (3)	Integrin α11 (1)
TGF-β1 (3)	Procathepsin (1)
TNF-β (3)	Tyrosine-protein kinase BTK (1)
TNF-β (4)	Tyrosine-protein kinase JAK3 (1)
INF-γ (3)	B-lactamase (1)
VEGF (4)	$Lewis^x$ (2)
Angiomotin (2)	$Lewis^y$ (1)
MCP-1 (3)	B cell lymphoma ag (1)
MCP-3 (3)	Sialo $Lewis^x$ (1)
MCP-4 (3)	MUC-1 (1)
Eotaxin (3)	Streptavidin (control) (1)
RANTES (3)	Digoxin (control) (1)
GM-CSF (3)	FITC (control) (1)
CD40 (4)	TAT (control) (2)

first large-scale microarray (Fig. 1) based on 129 recombinant antibody fragments [7, 10, 17], directed against 60 serum proteins, mainly of immunoregulatory nature (Table 2). In this study, directly labeled sera from 24 pancreatic cancer patients and 20 healthy patients were each incubated on antibody microarrays. Bound analytes were subsequently quantified, using a confocal scanner and fluorescence as mode of detection. The statistical analysis was performed in two steps. First, to test our ability to detect cancer, the microarray data was displayed in an unsupervised Sammon plot based on all 129 antibodies, directed against the 60 different serum proteins, and two distinct populations could be distinguished (Fig. 2). This indicated the existence of a clear difference between the cancer and the normal serum proteomes, based on the serum proteins analyzed by the microarray. Secondly, we ran a leave-one-out crossvalidation, with an SVM, and collected the decision values for each sample. The decision value is the output of the predictor, and samples with a prediction value above a threshold are predicted to be pancreatic carcinomas. The threshold parameterizes the trade-off between sensitivity and specificity and is often, but not always, set to zero. The 24 pancreatic carcinoma samples obtained decision values in the interval from 0.30 to 1.93, and the healthy samples in the interval from −1.84 to −0.30. Thus, with a threshold value of zero or any other value between −0.30 and 0.30, the cancer *versus* healthy samples were correctly classified with an ROC area (AUC) of 1.

In a further attempt to illustrate the clear separation between the cancer and the normal group and also to identify the individual serum biomarkers, we randomly selected a training set from the total of 44 samples. The training set consisted of 28 samples, 18 cancer and 10 normal samples. A condensed set of biomarkers consisting of 19 nonredundant serum proteins, that differed significantly ($p<0.05$) between cancer and normal samples could be selected from the training set (Fig. 3A). These differentially expressed proteins were subsequently used to construct a dendrogram of the 28 training samples as well as of the 16 samples used as an independent test set. As seen in Fig. 3B, the test samples were all correctly classified.

Importantly, the protein signature, defined by the training and used for classification of the independent test samples, seemed to be associated with PAs, since it differed from serum signatures found by our microarray setup in other cancers such as gastric [17] and breast adenocarcinomas [20]. The expression of four of the serum biomarkers in the signature, discriminating between cancer and normal samples, were also confirmed using conventional ELISA. The data for IL-13, as a representative example, is shown in Fig. 4.

While an early detection of cancer has its merits, especially in pancreatic cancer, serum protein profiling has also been suggested as the approach to define signatures that might be associated with clinical parameters [24]. Predictions of expected survival time would be of high relevance for the therapeutic regimes assigned to each patient. Consequently, in an attempt to further interrogate our recombinant antibody microarray platform, we compared two cohorts of cancer patients consisting of 18 short survivors (<12 months) and five long survivors (>24 months), respectively. Leave-one-out crossvalidation was used to assess the prediction accuracy for every number of analytes between 1 and 129. This allowed us to estimate the size of a signature, *i.e.*, the number of analytes, without considering the nature of an individual analyte. For every fixed number of analytes, every sample was left out of the training set once and the number of analytes separating the short and long survivors in the training set were chosen. An SVM was trained using those analytes, and the decision value for the left out sample was collected. The decision values, one for each of the 23 samples, were then used to construct ROC curves and the AUC. Figure 5A shows the AUC as a function of the number of analytes in the signature.

It was evident that the two different cohorts can be discriminated, with an ROC area (AUC) of >0.80. Figure 5A also shows that protein signatures consisting of less then approximately 26 analytes yielded more variability and less robust signatures. Consequently, to obtain less variability and still keep the number of analytes to a minimum, we chose to work with a signature size of 29 analytes for further analysis. The ROC curve for 29 analytes had an AUC value of 0.86 (Fig. 5B).

To illustrate the predictive ability of a signature consisting of 29 proteins, the cancer patients ($n = 23$) comprising short- and long-term survivors, were randomly split into a training set of 13 patients and a test set of 10 patients. The 29 most significant analytes in this training set (Wilcoxon test) were then used to train an SVM and the prediction values were plotted against each patient. Figure 6A shows that the training set classified the samples correctly. The test set of

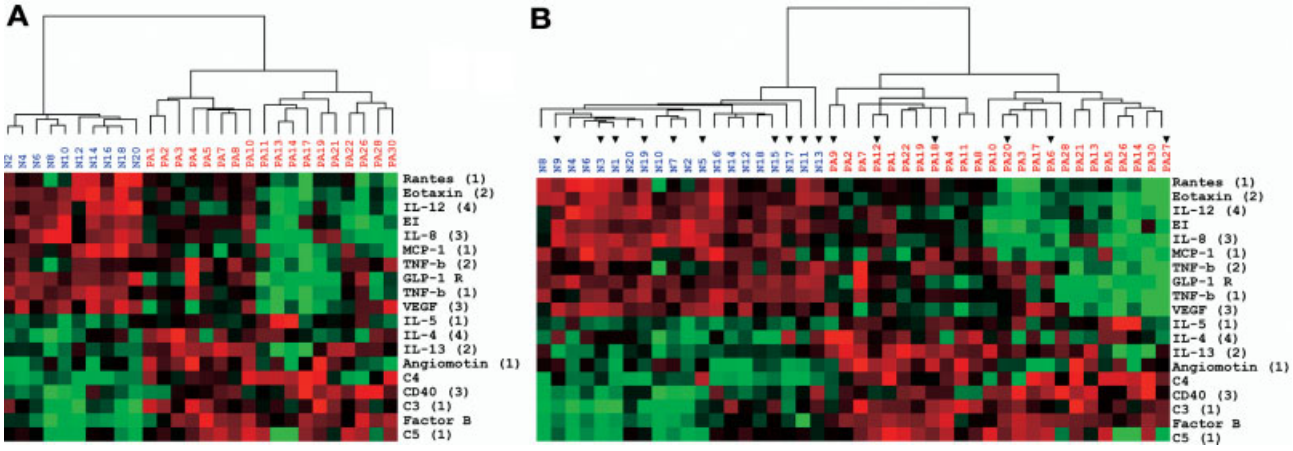

Figure 3. Comparison of protein expression serum signatures from pancreatic cancer patients *versus* healthy subjects. (A) Two-way hierarchial cluster analysis based on the 19 nonredundant serum protein analytes that were found to be differentially expressed ($p<0.05$) in cancer patients *versus* healthy controls, using a training set composed of 28 samples (18 cancer and 10 controls). Columns represent donors, where red is cancer patients (PA) and blue is healthy controls (N). Each row represents a serum biomarker, as denoted on the right hand side, where each pixel demonstrates the expression level of that particular biomarker in each donor (over-expression (red), under-expression (green), or no change (black) in pancreatic cancer sera *vs.* normal sera). (B) Two-way hierarchial cluster analysis of a test set (marked by arrows) composed of 16 serum samples (six cancer and ten controls) using the biomarker signature identified in (A). For visualization, the training set is also included in the dendrogram.

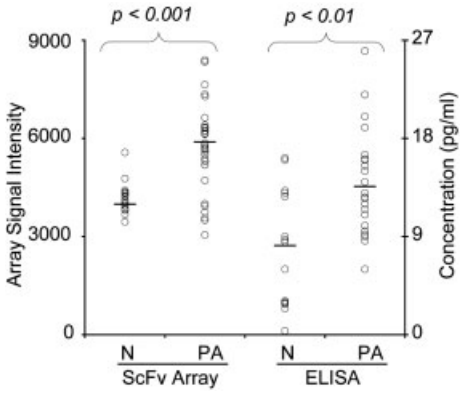

Figure 4. Validation of the antibody microaray data using ELISA. The serum levels of several of the candidate biomarkers, such as IL-4, IL-5, IL-13, and MCP-1, were confirmed using ELISA. A representative result, obtained for IL-13 is shown.

ten different patient samples was then classified, using this signature. The prediction values of the test samples are shown in Fig. 6B. All patients surviving <12 months were correctly classified, using an SVM prediction value threshold of 0. One long-term survivor was misclassified. The 29 most significant analytes separating long- and short-term survivors were subsequently identified as 21 nonredundant serum proteins (eight out of the 29 analytes were duplicates but defined by different antibody clones). The differential expression levels of the 21 nonredundant proteins in all patients are shown as a heat map in Fig. 7, where the short and long survivors are grouped. However, a study with more than 18 short survivors and five long survivors is needed to establish a survival classifying protein profile, although the data indicates the possibility for the existence of such a profile. When analyzing the individual proteins there was no strict consensus patter among the serum proteins, although it was evident that cytokines, such as IL-1α, IL-3, IL-8, and IL-11 were upregulated in short term survivors, while Rantes, IL-16, IL-4, and eotaxin were mostly upregulated in long term survivors. The significance of this remains to be validated but it could indicate a more active T-cell compartment in the latter population.

4 Discussion

Antibody microarrays have evolved over the last several years from a promising tool, in affinity proteomics, to an approach that is starting to deliver promising results, in particular in oncoproteomics [3, 9, 17, 25]. The main focus of these endeavors has been to detect cancer at an early stage, to predict tumor relapse and treatment resistance, or to select patients for a particular treatment regime [3]. This is, in particular, important for cancers with poor prognosis, which is also intrinsic to pancreatic cancer since it rapidly metastasizes to, *e.g.*, lymph nodes, lungs, and peritoneum [14] and is difficult to diagnose at an early stage. However, the ability of a biomarker signature to distinguish between different carcinomas or between cancer and inflammation has so far been difficult to achieve [26], for review see [3]. The reason for the observed distinction between cancer and normal serum proteomes in this study is most likely dependent on the range of antibody specificities on the microarray, which is also recently supported by the rationally designed array,

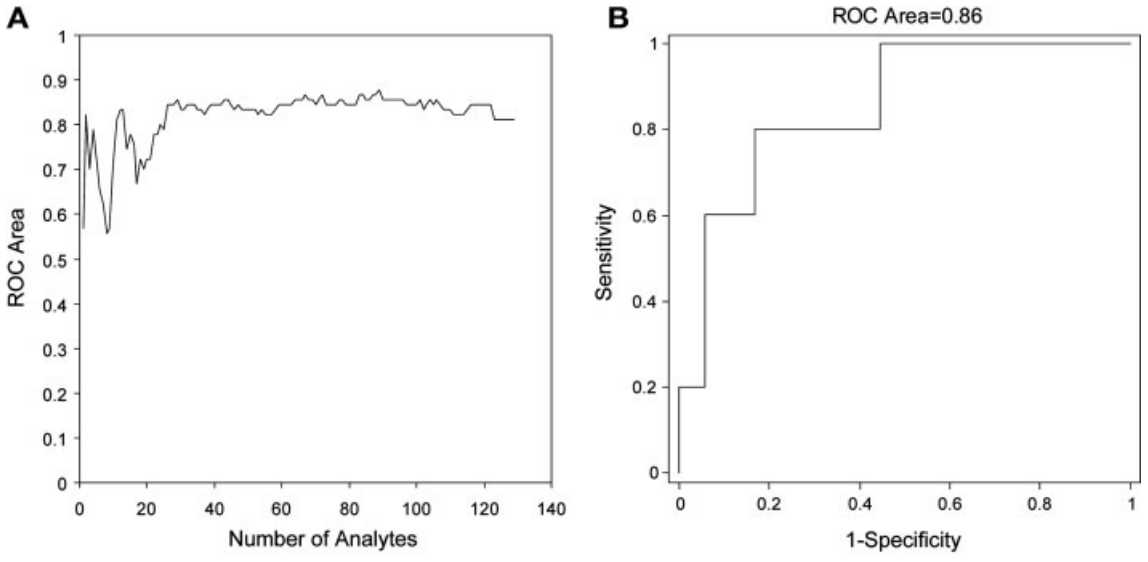

Figure 5. Influence of signature size on the discrimination between short (<12 months) *versus* long (>24 months) surviving cancer patients. (A) ROC area as a function of the number of analytes, from 1 to 129, included in a tentative predictor signature, based on a leave-one-out crossvalidation, using all 23 cancer patients. (B) The ROC area of a serum biomarker signature, based on 29 identified analytes.

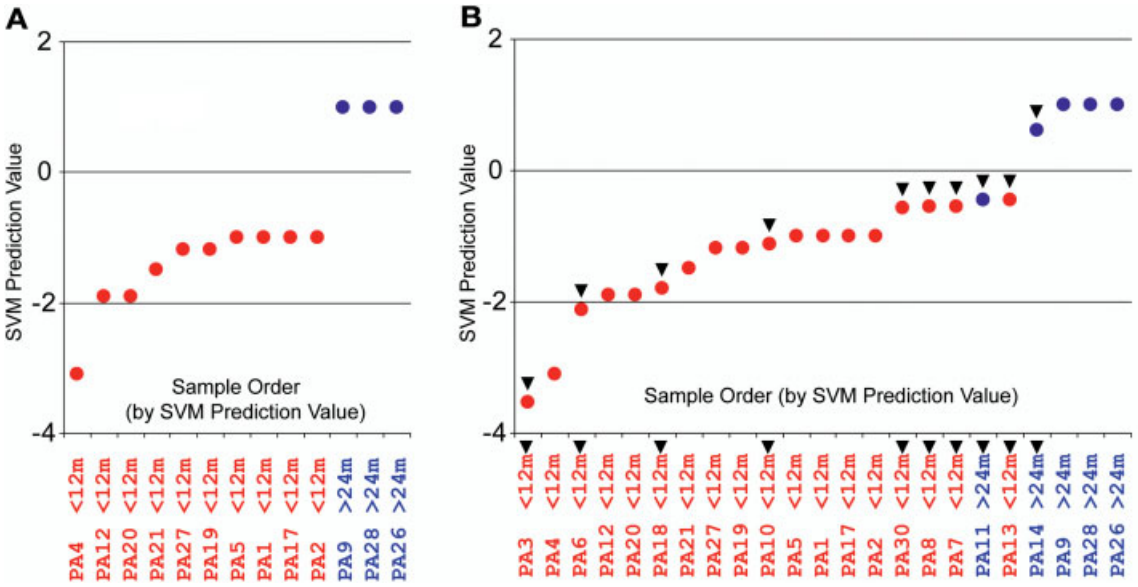

Figure 6. Identification of a signature discriminating between short (<12 months) *versus* long (>24 months) surviving cancer patients. (A) Identification of the 29 most significantly expressed analytes, between short (red) and long (blue) survivors, based on a training set composed of ten short survivors and three long survivors. This corresponds to 21 nonredundant serum proteins. (B) Classification of a test set (marked by arrows) composed of ten samples (eight short survivors and two long survivors) into short survivors (red) and long survivors (blue), using the signature identified in (A). For visualization, the training set is also included in the plot.

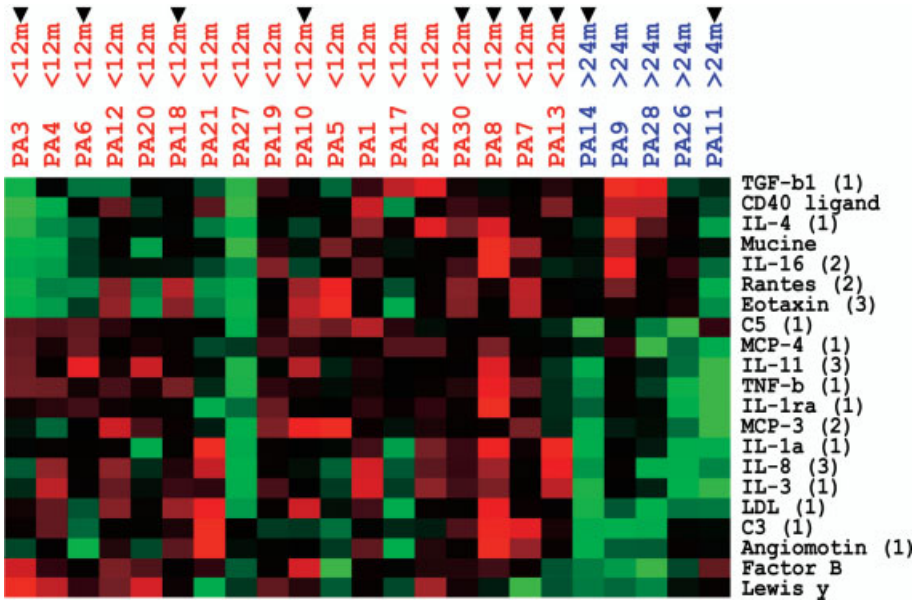

Figure 7. A heatmap of individual analytes from short survivors *versus* long survivors. The columns represent cancer patients, where blue is long (>24 months) survivors and red is short (<12 months) survivors. See legend to Fig. 3A for color coding.

reported by Sanchez-Carbayo *et al.* [9]. These investigators could stratify patients with bladder tumors on the basis of their overall survival, using antibodies generated against differentially expressed gene products. During the last years, we have developed a high-performing, recombinant antibody microarray platform for complex proteome analysis [6, 7, 10, 17, 21], by evaluating and optimizing key technological parameters [6] such as probe and substrate design [7, 18, 21, 27], array/assay design [10, 19], and sample format [7, 10, 19]. This has allowed us to perform differential protein expression profiling of the human plasma proteome, using the optimized scFv microarrays targeting mainly immunoregulatory proteins. In agreement with previous result, this antibody microarray displayed sensitivities in the picomolar to femtomolar range, allowing us to detect low-abundant analytes, such as cytokines. Furthermore, we maintained an

assay reproducibility with a coefficient of correlation in the range of 0.96–0.99, which is a key feature of multiplexed analysis and which compares well with previous reports [8, 10, 17].

Patients with pancreatic cancer are often diagnosed late, resulting in a poor prognosis. Due to low incidence it is fairly difficult to gather large sample numbers, especially for long-term survivors, *i.e.*, >24 months. We had access to 24 patients for this study, which using a rigorous statistical evaluation still supported an attempt to classify cancer and normal proteomes. We have employed an SVM for supervised classification. Very similar results were obtained with a naive Bayesian classifier (data not shown). The SVM separated the two groups by finding a hyperplane in space of all analytes, from which the prediction or decision values were found (Fig. 6B). The hyperplane and, thus, the classification of groups, was determined by the training set. The performance of the classifier was then estimated by using an independent test set. Very importantly no overlap between the training and test set was allowed. However, a dataset can randomly be split into different training and an independent test set. The final result then depends on the split into training and test set. Consequently, we used crossvalidation as the procedure of making several splits of our dataset and used the average performance of the test sets as a measure of the accuracy of data classification. Thus, in the leave-one-out crossvalidation that was performed, the test set contains one sample and the training set contains the remainder. The performance of the SVM as measured by the ROC curve was very good. The normal and pancreatic carcinoma samples were remarkably well separated, with a gap between the two groups.

In this study, we did not compare the cancer patients with a cohort of patients with chronic pancreatitis. This could have been a relevant comparison, although the differential diagnosis of cancer and pancreatitis is normally not a problem, due to improved magnetic resonance cholangiopancreatography (MRCP) to visualize the biliary tract and pancreatic ducts. However, since chronic pancreatitis most probably would have displayed an inflammatory signature, we compared the present signature associated with pancreatic cancer with several other signatures associated with inflammation. The present pancreatic cancer associated biomarker signature only had eotaxin, IL-5, and IL-13 in common with 14 biomarkers found as a result of a *Helicobacter pylori*-based infection, associated with gastrointestinal cancer [17], indicating that our pancreatic signature was not related to general inflammation. Furthermore, our signature was not similar to serum biomarkers found in systemic lupus erythematosus, an autoimmune disorder with a significant inflammatory component (Wingren *et al.*, manuscript in preparation). The signature was also completely different from what Orchekowski *et al.* [8] reported, when profiling pancreatic cancer serum samples, using a microarray based on monoclonal and polyclonal antibodies. They analyzed serum proteins, such as albumin, transferrin, and hemoglobin, as well as more common inflammation markers, such as CRP, serum amyloid A, and Igs together with eight different cytokines. Furthermore, our present cancer signature contained a number of overexpressed Th2 cytokines (IL-4, -5, and -13), whereas classical Th1 cytokines (IL-12 and TNF-β) were downregulated. This was also in agreement with the study of Bellone *et al.* [28], who showed that blood-derived monocytes from pancreatic cancer patients were primed to develop a TH2-like response, rather than a TH1-like response, with increased expression of IL-4, and decreased expression of IL-12 [28]. Taken together, the proposed serum signature associated with pancreatic cancer does not resemble any of the reported inflammatory signatures. In the comparison between cancer *versus* normal subjects, there was a different gender, and age distribution between normal and pancreatic patients, which opens the possibility that our classifier is affected by age or gender instead of cancer. As can be seen in Table 1, there is a skewed age distribution between the cancer and normal samples. The cancer patients were on average older than the normal persons, which opens up the possibility that the signature proteins distinguished age instead of cancer, *i.e.*, that age is a confounding factor. However, if that was the case, the normal samples from older subjects would have been classified as cancer, and the cancer samples from younger subjects would have been classified as normal, which did not happen. All samples were correctly classified as normal or cancer by both the hierarchical clustering (Fig. 3B) and the SVM. The same argument applies to gender, *i.e.*, if the classifier was predicting females instead of cancer, the ten males with cancer and the two normal females would have been misclassified.

In conclusion, using a recombinant antibody microarray against immunoregulatory proteins, we have been able to detect PAs and discriminate between cancer *versus* normal serum proteomes. The first attempt to define a signature capable of predicting survival of cancer patients was also made. Taken together, the result indicates the power of affinity oncoproteomics, in particular based on antibody microarrays, for future clinical decision making.

Serum samples were kindly supplied by Dr. Stefan Linder (Södersjukhuset). This study was supported by grants from the Lund Institute of Technology, Åke Wiberg Foundation, Swedish National Research Council (VR-NT), and the Foundation of Strategic Research (Strategic Center for Translational Cancer Research - CREATE Health (www.createhealth.lth.se).

The authors have declared no conflict of interest.

5 References

[1] Cho, W. C., Contribution of oncoproteomics to cancer biomarker discovery. *Mol. Cancer* 2007, *6,* 25.

[2] Hanash, S., Disease proteomics. *Nature* 2003, *422,* 226–232.

[3] Borrebaeck, C. A., Antibody microarray-based oncoproteomics. *Expert Opin. Biol. Ther.* 2006, *6*, 833–838.

[4] Hu, S., Loo, J. A., Wong, D. T., Human body fluid proteome analysis. *Proteomics* 2006, *6*, 6326–6353.

[5] Kingsmore, S. F., Multiplexed protein measurement: Technologies and applications of protein and antibody arrays. *Nat. Rev. Drug Discov.* 2006, *5*, 310–320.

[6] Wingren, C., Borrebaeck, C. A. K., High-throughput proteomic using antibody microarrays – an update. *Exp. Rev. Mol. Diagn.* 2007, *7*, 673–686.

[7] Ingvarsson, J., Larsson, A., Sjoholm, A. G., Truedsson, L. *et al.*, Design of recombinant antibody microarrays for serum protein profiling: Targeting of complement proteins. *J. Proteome Res.* 2007, *6*, 3527–3536.

[8] Orchekowski, R., Hamelinck, D., Li, L., Gliwa, E. *et al.*, Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer Res.* 2005, *65*, 11193–11202.

[9] Sanchez-Carbayo, M., Socci, N. D., Lozano, J. J., Haab, B. B., Cordon-Cardo, C., Profiling bladder cancer using targeted antibody arrays. *Am. J. Pathol.* 2006, *168*, 93–103.

[10] Wingren, C., Ingvarsson, J., Dexlin, L., Szul, D., Borrebaeck, C. A. K., Design of recombinant antibody microarrays for complex proteome analysis: Choice of sample labeling-tag and solid support. *Proteomics* 2007, *7*, 3055–3065.

[11] Rosenwald, A., Wright, G., Wiestner, A., Chan, W. C. *et al.*, The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* 2003, *3*, 185–197.

[12] van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H. *et al.*, A gene-expression signature as a predictor of survival in breast cancer. *N. Eng. J. Med.* 2002, *347*, 1999–2009.

[13] Tingstedt, B., Johansson, P., Andersson, B., Andersson, R., Predictive factors in pancreatic ductal adenocarcinoma: Role of the inflammatory response. *Scand. J. Gastroenetrol.* 2007, *42*, 754–759.

[14] Garcea, G., Neal, C. P., Pattenden, C. J., Steward, W. P., Berry, D. P., Molecular prognostic markers in pancreatic cancer: A systematic review. *Eur. J. Cancer* 2005, *41*, 2213–2236.

[15] Rustgi, A. K., Pancreatic cancer: Novel approaches to diagnosis and therapy. *Gastroenterology* 2005, *129*, 1344–1347.

[16] Yeo, T. P., Hruban, R. H., Leach, S. D., Wilentz, R. E. *et al.*, Pancreatic cancer. *Curr. Probl. Cancer* 2002, *26*, 176–275.

[17] Ellmark, P., Ingvarsson, J., Carlsson, A., Lundin, B. S. *et al.*, Identification of protein expression signatures associated with Helicobacter pylori infection and gastric adenocarcinoma using recombinant antibody microarrays. *Mol. Cell. Proteomics* 2006, *5*, 1638–1646.

[18] Soderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N. *et al.*, Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat. Biotechnol.* 2000, *18*, 852–856.

[19] Wingren, C., Steinhauer, C., Ingvarsson, J., Persson, E. *et al.*, Microarrays based on affinity-tagged single-chain Fv antibodies: Sensitive detection of analyte in complex proteomes. *Proteomics* 2005, *5*, 1281–1291.

[20] Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P. *et al.*, Serum protein profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur. J. Cancer* 2008, *44*, 472–480.

[21] Wingren, C., Borrebaeck, C. A. K., Antibody microarrays: Current status and key technological advances. *OMICS* 2006, *10*, 411–427.

[22] Ihaka, R., Gentleman, R., A language for data analysis and graphics. *J. Comp. Graph. Stat.* 1996, *5*, 299–314.

[23] Cristianini, N., Shawe-Taylor, J., *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge 2000.

[24] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14863–14868.

[25] Shafer, M. W., Mangold, L., Partin, A. W., Haab, B. B., Antibody array profiling reveals serum TSP-1 as a marker to distinguish benign from malignant prostatic disease. *Prostate* 2007, *67*, 255–267.

[26] Mor, G., Visintin, I., Lai, Y., Zhao, H. *et al.*, Serum protein markers for early detection of ovarian cancer. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 7677–7682.

[27] Steinhauer, C., Ressine, A., Marko-Varga, G., Laurell, T. *et al.*, Biocompatibility of surfaces for antibody microarrays: Design of macroporous silicon substrates. *Anal. Biochem.* 2005, *341*, 204–213.

[28] Bellone, G., Turletti, A., Artusio, E., Mareschi, K. *et al.*, Tumor-associated transforming growth factor-beta and interleukin-10 contribute to a systemic Th2 immune phenotype in pancreatic carcinoma patients. *Am. J. Pathol.* 1999, *155*, 537–547.

