

Tasneem H. Patwa<sup>1</sup>  
 Chen Li<sup>1</sup>  
 Laila M. Poisson<sup>2</sup>  
 Hye-Yeung Kim<sup>1</sup>  
 Manoj Pal<sup>1</sup>  
 Debashis Ghosh<sup>3,4</sup>  
 Diane M. Simeone<sup>5,6</sup>  
 David M. Lubman<sup>1,5,7</sup>

<sup>1</sup>Department of Chemistry,  
 University of Michigan, Ann  
 Arbor, MI, USA

<sup>2</sup>Department of Biostatistics,  
 University of Michigan, Ann  
 Arbor, MI, USA

<sup>3</sup>Department of Statistics,  
 Pennsylvania State University,  
 University Park, PA, USA

<sup>4</sup>Department of Public Health  
 Sciences, Penn State, University  
 Park, PA, USA

<sup>5</sup>Department of Surgery,  
 University of Michigan, Ann  
 Arbor, MI, USA

<sup>6</sup>Department of Molecular and  
 Integrative Physiology,  
 University of Michigan, Ann  
 Arbor, MI, USA

<sup>7</sup>Comprehensive Cancer Center,  
 University of Michigan, Ann  
 Arbor, MI, USA

Received December 30, 2008

Revised February 17, 2009

Accepted March 11, 2009

## Research Article

# The identification of phosphoglycerate kinase-1 and histone H4 autoantibodies in pancreatic cancer patient serum using a natural protein microarray

Protein microarrays have been used to explore whether a humoral response to pancreatic cancer-specific tumor antigens has utility as a biomarker of pancreatic cancer. To determine if such arrays can be used to identify novel autoantibodies in the sera from pancreatic cancer patients, proteins from a pancreatic adenocarcinoma cell line (MIAPACA) were resolved by 2-D liquid-based separations, and then arrayed on nitrocellulose slides. The slides were probed with serum from a set of patients diagnosed with pancreatic cancer and compared with age- and sex-matched normal subjects. To account for patient-to-patient variability, we used a rank-based non-parametric statistical testing approach in which proteins eliciting significant differences in the humoral response in cancer compared with control samples were identified. The prediction analysis for microarrays classification algorithm was used to explore the classification power of the proteins found to be differentially expressed in cancer and control sera. The generalization error of the classification analysis was estimated using leave-one-out cross-validation. A serum diagnosis of pancreatic cancer in this set was predicted with 86.7% accuracy, with a sensitivity and specificity of 93.3 and 80%, respectively. Candidate autoantibody biomarkers identified using this approach were studied for their classification power by performing a humoral response experiment on recombinant proteins using an independent sample set of 238 serum samples. Phosphoglycerate kinase-1 and histone H4 were noted to elicit a significant differential humoral response in cancer sera compared with age- and sex-matched sera from normal patients and patients with chronic pancreatitis and diabetes. This work demonstrates the use of natural protein arrays to study the humoral response as a means to search for the potential markers of cancer in serum.

### Keywords:

Cancer / Liquid separations / Microarrays

DOI 10.1002/elps.200800857



## 1 Introduction

Advances in the treatment of pancreatic cancer will be greatly aided by early detection so as to diagnose and treat

cancer while it is in an early, curable state. Unfortunately, for pancreatic adenocarcinoma (PDAC), the fourth leading cause of cancer death in the United States [1], effective early detection and screening are not currently available and tumors are typically diagnosed at a late stage, frequently after metastasis. PDAC is generally considered to be largely incurable by available treatment modalities, with a 5-year survival rate of less than 4%. Existing biomarkers for this disease are inadequate [2]. CA19-9 has been tested for its utility as an early detection marker in PDAC [2–5]; however, the sensitivity and specificity of this biomarker are not high, and serum levels are significantly increased in inflammatory diseases of the pancreas and biliary tract. Therefore, CA19-9 is not useful for early diagnosis, mass screening, distinguishing between PDAC and chronic pancreatitis or the targeting of therapeutics. Thus, there is a great need for new

**Correspondence:** Dr. David M. Lubman, Department of Surgery, University of Michigan, 1150 West Medical Center Drive, Room A510B MSRB I, Ann Arbor, MI 48109-0656, USA  
**E-mail:** dmlubman@umich.edu  
**Fax:** +1-734-615-2088

**Abbreviations:** CF, chromatofocusing; LOOCV, leave-one-out cross-validation; PAM, prediction analysis for microarrays; PDAC, pancreatic adenocarcinoma; PGK-1, phosphoglycerate kinase; PPV, positive predictive value; ROC, receiver operator characteristic; NPS, non-porous silica

biomarkers for PDAC. In the absence of good biomarkers, 80–90% of PDAC cases are diagnosed too late in the disease process for surgical resection to be an effective option. Even among the 10–20% of PDAC cases where surgical resection is an option, most patients ultimately die of recurrent or metastatic disease [6]. Identification of novel biomarkers for PDAC may have utility for the detection of this malignancy.

A humoral response to cancer in humans has been well demonstrated by the identification of autoantibodies to a number of different intracellular and surface antigens in patients with various tumor types [7–14]. Tumor-specific humoral responses directed against oncoproteins [15, 16], mutated proteins such as p53 [17–19] or other aberrantly expressed proteins have all been described. Although it is currently unknown whether the occurrence of such antibodies is beneficial to the patient, knowledge of potential tumor antigens that can evoke tumor-specific immune responses may have utility in cancer diagnosis, in establishing prognosis and in targeted immunotherapy against the disease. In PDAC, autoimmunity has been shown against a number of cellular proteins (or protein isoforms), including MUC1 [20, 21], p53 [18, 19], Rad51 [22] and DEAD-box protein 48 [23]. However, in most cases, autoantibodies to specific proteins occur in only a small percent (10–20%) of patient's sera and therefore, they may not be effective individually for the early detection of PDAC, but rather may have utility as part of a panel [24].

In the current study a strategy using liquid-based multi-dimensional procedures to separate proteins allows distinct protein-containing fractions to be arrayed and interrogated using various types of probes. We have utilized methodology that first employs the separation of cell and/or tissue lysates by chromatofocusing (CF), followed by liquid-phase separation by non-porous silica (NPS) RP HPLC (according to hydrophobicity) [25]. This methodology allows a large number of proteins to be resolved using a liquid-based system, without the need for cumbersome 2-D gel analysis. Importantly, liquid-based protein separations are well suited for fractionation of lysates into individual protein fractions or for purification of individual proteins. Additionally, the separated proteins are maintained in solution, thus facilitating intact protein identification by MS and the spotting of individual fractions on protein microarrays with a robotic arrayer. Also, these natural microarrays express PTMs as expressed in the disease state, which may not be true of recombinant methods for producing microarrays [26, 27]. Such protein microarrays have been utilized to assess the binding characteristics of multiple samples (probes) simultaneously [25, 28].

## 2 Materials and methods

### 2.1 Sample preparation

Serum samples were obtained from patients with a confirmed diagnosis of PDAC who were seen in the Multidisciplinary Pancreatic Tumor Clinic at the University

of Michigan Comprehensive Cancer Center. Sera from the pancreatic cancer patients were randomly selected from a clinic population that sees, on average, at the time of initial diagnosis, 15% of PDAC patients presenting with early stage (*i.e.* stage I/II) disease and 85% presenting with advanced stage (*i.e.* stage III/IV). All sera samples selected for this study were stages III/IV since sufficient numbers of early stage pancreatic cancer sera could not be obtained. Inclusion criteria for the study included patients with a confirmed diagnosis of pancreatic cancer, the ability to provide written, informed consent and the ability to provide 40 mL of blood. Exclusion criteria included the inability to provide informed consent, patients' actively undergoing chemotherapy or radiation therapy for pancreatic cancer and patients with other malignancies diagnosed or treated within the last 5 years. The mean age of the tumor group was 65.4 years (range 54–74 years). The sera from the normal subject group were age- and sex-matched to the tumor group. The criteria for the healthy controls included no family history of pancreatic cancer, no personal history of acute or chronic pancreatitis, no prior history of any malignancy except non-melanoma skin cancers for the past 10 years, no personal history of diabetes, no concurrent abdominal pain and no concurrent unexplained weight loss. The chronic pancreatitis group was sampled when there were no symptoms of acute flare of their disease. The diabetes samples used in the pre-validation experiments were patients with a history of type 2 diabetes mellitus for 10 years or more. In addition, these patients did not have a family history of pancreatic cancer, no personal history of acute or chronic pancreatitis, no prior history of any malignancy, no concurrent abdominal pain and no concurrent unexplained weight loss. Chronic pancreatitis sera were chosen to control for the inflammatory response typically seen in pancreatic cancers, and diabetes sera to control for the presence of type 2 diabetes that develops in many of these patients. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 min (and a maximum of 60 min) to allow the clot to form in the red top tubes, and then centrifuged at  $1300 \times g$  at 4°C for 20 min. The serum was removed, transferred to a polypropylene, capped tube in 1 mL aliquots and frozen. The frozen samples were stored at –80°C until assayed. All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. None of the samples were thawed more than twice before analysis.

### 2.2 Cell culture

The cells used in this work were from the pancreatic cancer cell line, MIAPACA. The cells were cultured at 37°C in a 5% CO<sub>2</sub>-humidified incubator in DMEM growth medium supplemented with 10% FBS and 1% penicillin/streptomycin (Invitrogen, Carlsbad, CA, USA). When the cells reached ~90% confluence, they were harvested with a cell scraper and lysed in lysis buffer containing 7 M urea, 2 M thiourea,

100 mM DTT, 2% n-octyl- $\beta$ -glucopyranoside (OG), 10% glycerol, 10 mM sodium orthovanadate, 10 mM sodium fluoride (all from Sigma, St. Louis, MO, USA), 0.5% Biolyte ampholyte (Bio-Rad, Hercules, CA, USA) and protease inhibitor cocktail (Roche Diagnostics, GmbH, Mannheim, Germany). Cell lysates were centrifuged at 35 000 rpm for 1 h and were then buffer exchanged into start buffer (6 M urea, 25 mM Bis-Tris and 0.2% n-octyl- $\beta$ -glucopyranoside) using a PD-10 G-25 column (Amersham Biosciences, Piscataway, NJ, USA) and stored at  $-80^{\circ}\text{C}$  until further use.

### 2.3 CF

Prior to CF the extracted protein content from the cell line was assayed using a Bradford protein assay kit, using BSA (Bio-Rad) as the standard protein. CF was performed using a Beckman System Gold model 127 pump and 166 UV detector module (Beckman Coulter, Fullerton, CA, USA) as described previously [25]. Fractions were collected at 0.2 pH unit intervals. The pH was monitored with a post-detector on-line pH-flow cell (Lazar Research Laboratories, Los Angeles, CA, USA). After the CF gradient run was completed, the column was flushed with a 1 M NaCl solution, followed by deionized water. Finally, the column was flushed with isopropanol and stored with the same until further use. The collected fractions were stored at  $-80^{\circ}\text{C}$  until further use.

### 2.4 NPS-RP-HPLC

Each fraction from the first dimension CF was further separated in the second dimension by NPS-RP-HPLC, according to protein hydrophobicity. An ODSIII-E ( $8 \times 33$  mm) column (Eprogen, Darien, IL, USA) packed with  $1.5 \mu\text{m}$  NPS was used to achieve high-separation efficiency. A total of 0.1% TFA with water (A) and 0.08% TFA with acetonitrile (B) gradient were used in the separation. The following gradient was applied at a flow rate of 1 mL/min and fractions were collected by peak using an automated fraction collector (model SC 100; Beckman Coulter) in 96-well plates with a gradient as described in the previous work [25]. All separations were performed at  $60^{\circ}\text{C}$  and were monitored at 214 nm. All 2-D fractions were stored at  $-80^{\circ}\text{C}$  until further use.

### 2.5 Microarray printing

Approximately 30% of each protein sample was transferred to 96-well printing plates (Bio-Rad) and were completely dried using a speedvac concentrator at  $60^{\circ}\text{C}$ . The remaining portion of each sample was used for protein identification. The fractions were then resuspended in printing buffer (62.5 mM Tris-HCl (pH 6.8), 1% w/v SDS, 5% w/v DTT and 1% glycerol in  $1 \times$  PBS) and were left to shake overnight at  $4^{\circ}\text{C}$ . Slides were printed by transferring each fraction from

the plate onto nitrocellulose slides (GenTel Bioscience) using a non-contact piezoelectric printer (Nanoplotter 2, GeSiM). Each spot resulted from deposition of five spotting events of 500 pL each, such that a total volume of 2.5 nL of each fraction was spotted. Each spot was found to be  $\sim 450 \mu\text{m}$  in diameter, with the distance between spots maintained at  $600 \mu\text{m}$ . Printed slides were left on the printer deck overnight to dry and were then stored in a desiccator at  $4^{\circ}\text{C}$  until further use.

### 2.6 Hybridization of slides

The printed arrays were rehydrated in  $1 \times$  PBS with 0.1% Tween-20 (PBS-T), and were then blocked overnight in a solution of 1% BSA in PBS-T. Each serum sample was diluted 1:400 in probe buffer (5 mM magnesium chloride, 0.5 mM DTT, 0.05% Triton X-100, 5% glycerol and 1% BSA in  $1 \times$  PBS) to make a total solution of 4 mL and kept on ice. The slides were hybridized in diluted serum for 2 h (one serum sample *per* slide). Hybridization was done at  $4^{\circ}\text{C}$  in heat-sealable pouches with agitation, using a mini-rotator. The slides were then washed five times with probe buffer (5 min each), and were then hybridized with 4 mL goat anti-human IgG conjugated with Alexafluor647 (Invitrogen) (at  $1 \mu\text{g}/\text{mL}$  in probe buffer), for 1 h at  $4^{\circ}\text{C}$ . After secondary incubation all slides were washed in probe buffer five times, for 5 min each, and were then dried by centrifugation for 10 min. The sample hybridization was totally randomized in no specific order to prevent bias. All processed slides were immediately scanned using an Axon 4000B microarray scanner (Axon Instruments, Foster City, CA, USA).

### 2.7 Data acquisition and analysis

GenePix 6.0 software was used to grid all spots, to determine the median Cy5 single-channel intensities and median local background intensities for each spot. A spot was considered positive if the foreground measure was at least  $2 \times$  the background intensity measure. We used foreground data alone as well as the background-subtracted data for analysis. To account for the variation between arrays, each array was median-centered and scaled by its interquartile range. After standardization the replicate arrays were averaged. To assess the differences between humoral response in cancer and normal sera, the non-parametric Wilcoxon rank-sum test was employed. Additionally z-score statistics were used on the foreground data to look for subtle differences between the two sera groups. Finally, a classifier was built from the differential proteins found by these methods.

#### 2.7.1 Non-parametric method

A two-sample Wilcoxon rank-sum test between cancer and benign sera was run for each spot on the array. Each pH/fraction combination was tested and the *p*-values were visualized in a grid plot to highlight the regions of spots that

exhibited differential response between normal and cancer sera. A *p*-value threshold of 0.05 was used to determine differential proteins for further study.

### 2.7.2 z-Score method

The standardized data were log-transformed after adding a small constant value to each point to eliminate negative values. z-score measures were constructed for each spot by subtracting the mean and dividing by the standard deviation of only the control serum samples for that spot. Resulting z-scores were then on the scale of standard deviations from the mean of the control samples. Proteins that had z-scores of  $>2$  (or  $<-2$ ) in 20% of the cancer serum samples were determined to be differential and considered for further study.

### 2.7.3 Prediction analysis for microarrays (PAM) classification algorithm and leave-one-out cross-validation (LOOCV)

The PAM classification algorithm [29], as implemented in R, was used to explore the classificatory power of the proteins found to be differential between control and cancer sera using either the Wilcoxon rank-sum test or the z-score method. From PAM the smallest subset of proteins that gave the lowest error rate was chosen to be used as a classifier. An receiver operator characteristic (ROC) curve was drawn to illustrate the selection of this “best” subset of proteins and the area under the curve (AUC) was estimated.

The generalizability of the PAM analysis was estimated using LOOCV in which each sample was left out of the data set in turn and classified using the remaining samples. Specifically, using only the 29 remaining samples, the same analysis scheme as performed above for the full set of 30 samples was repeated, including reselection of differential proteins using Wilcoxon tests and z-scores from the 29 samples (The median number of differential proteins across the 30 leave-one-out data sets was 96 (range = [65, 109]).) and classifier selection using PAM. The resulting classifier was then used to predict the diagnosis of the excluded sample. Each of the 30 samples was predicted in this way and error rates were estimated. The ROC curves were drawn to illustrate the selection of the “best” protein subset for the classification in each of the 30 leave-one-out cycles and the AUC was estimated.

### 2.7.4 Heatmaps

Heatmaps were drawn using Cluster and TreeView software [30]. Spots were median-centered across samples and average linkage clustering was used.

## 3 Protein identification – nano-LC-ESI-MS/MS

Individual protein fractions were first dried down to  $\sim 10\ \mu\text{L}$ , and then mixed with  $40\ \mu\text{L}$  100 mM ammonium bicarbo-

nate,  $10\ \mu\text{L}$  20 mM DTT and  $0.5\ \mu\text{L}$  of sequence grade modified trypsin (Promega). The mixture was allowed to incubate at  $37^\circ\text{C}$  overnight with agitation after which the digest was stopped by addition of  $1\ \mu\text{L}$  TFA (Baker and Baker). The digested sample was separated using a gradient that was started at 3% B, ramped to 35% B in 25 min, 60% B in 15 min, 90% in 1 min, maintained at 90% B for 1 min and finally ramped back down to 3% in another 1 min. The eluting peptides were analyzed on a linear ion-trap based mass spectrometer (LTQ, Thermo, San Jose, CA, USA) with an nano-ESI platform (Michrom Biosciences). The capillary temperature was set at  $200^\circ\text{C}$ , the spray voltage was 2.6 kV and the capillary voltage was 20 V. The normalized collision energy was set at 35% for MS/MS. The top five peaks were selected for CID. MS/MS spectra were interrogated using the SEQUEST algorithm in Bioworks software (Thermo) against the Swiss-Prot human protein database. Two missed cleavages were allowed during the database search. The search threshold was set to 1000 and tolerances were set at 1.4 and 0.00 for peptide and fragment ion tolerances, respectively. Protein identification was considered positive when a peptide showed an  $X_{\text{corr}} \geq 3.0$ , 2.5 and 1.9 for triply, doubly and singly charged ions, respectively. Only proteins with greater than 10% coverage were considered in the analysis and a minimum of three good-scoring peptides were required for positive identification. In the event that more than one protein were found in a fraction, the data were filtered. If the spot of interest was unique and did not lie between adjacent reactive spots then only the highest scoring protein that was not found in adjacent fractions in the separation profile was considered true hits, since the adjacent fractions did not elicit a humoral response. If the spot of interest was part of a group of spots eliciting a positive response within a separation profile, the common protein identified in all these spots was considered a true hit.

## 4 Antibody array pre-validation with serum hybridization

The recombinant protein phosphoglycerate kinase (PGK-1) was purchased from Abcam, (Cambridge, MA, USA) and histone H4 from New England Biolabs (Ipswich, MA, USA). The concentration of each recombinant protein was  $10\ \mu\text{g}/\text{mL}$ . A piezoelectric non-contact printer (Nano Plotter, GESIM) was used to print all the recombinant protein arrays on ultra-thin nitrocellulose slides (PATH slides, GenTel Bioscience). An aliquot of  $2.5\ \text{nL}$  of each concentrated fraction was spotted. Each recombinant protein was printed in triplicate and 14 identical blocks were printed on each slide. The slides were washed three times with 0.1% Tween in PBS buffer (PBST 0.1) and then blocked with 1% BSA (Roche Diagnostics) in PBST 0.1 for 1 h. The blocked slides were dried by centrifugation and inserted into a SIMplex (GenTel Bioscience) multi-array device, which divides each of the slides by 16 wells. The wells separate

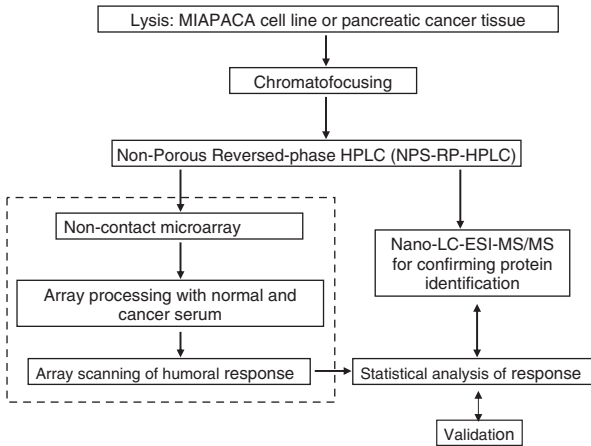
the neighboring blocks and prevent cross-contamination. Serum samples were diluted 10 times with PBST 0.1 containing 0.1% Brij. One hundred microliter of each diluted sample was applied to the recombinant protein array and the hybridization was performed in a humidified chamber for 1 h. The slides were then rinsed three times to remove the unbound proteins. An aliquot of 1  $\mu\text{g}/\text{mL}$  goat anti-human IgG conjugated with Alexafluor647 (Invitrogen)

solution was made and used for detection. After a second 1 h hybridization with antihuman IgG, the slides were washed and dried again, then scanned with a microarray scanner (Axon 4000A). The program Genepix Pro 6.0 was used to extract the numerical data.

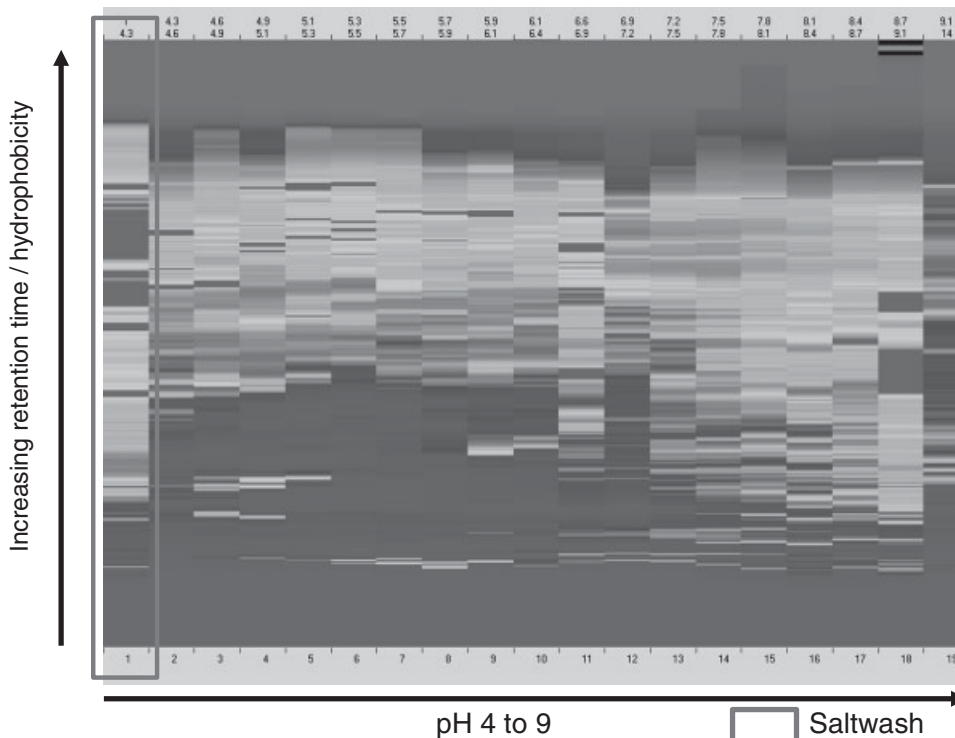
## 5 Results and discussion

### 5.1 Experimental scheme

We spotted native proteins derived from the MIAPACA pancreatic cancer cell line on protein microarrays to characterize a pancreatic cancer-specific humoral response. Such a study has potential utility in identification of novel potential candidate autoantibodies that may serve as markers of pancreatic cancer. Figure 1 illustrates schematically the methodology employed in this study. Proteins from the MIAPACA PDAC cell line were first solubilized, and then separated using 2-D liquid-based separation that employs CF (separation according to protein pI) in the first dimension and non-porous RP HPLC (separation according to protein hydrophobicity) in the second dimension. The MIAPACA cell line was chosen because it is a well-characterized pancreatic cancer cell line that has a gene expression profile that mimics that of primary pancreatic cancers. The separated proteins were then arrayed on nitrocellulose slides using non-contact piezoelectric printing. Following printing, slides were hybridized with serum from patients diagnosed with pancreatic cancer or normal subjects. Spots on the slides were statistically evaluated



**Figure 1.** Humoral response experimental overview. Proteins are first extracted from cell line and separated in two orthogonal dimensions. Separated fractions are spotted by non-contact means on nitrocellulose slides, which are then probed with serum from normal or cancer sera. Antibody–antigen response is detected using antihuman IgG conjugated to a fluorophore. Following statistical analysis, proteins of interest are identified by MS/MS.



**Figure 2.** 2-D UV chromatogram of separated MIAPACA cell lysate. On the horizontal axis are fractions from CF starting from the lowest pH to the highest pH. On the vertical axis is increasing retention time or hydrophobicity of the separated protein.

using non-parametric statistical methods to identify proteins that elicited a pancreatic cancer-specific humoral response. Proteins that elicited a statistically significant humoral response difference were subjected to classification analysis to obtain a panel of classifiers which were subsequently identified by nano-LC-linear ion trap MS. For the selection of identified proteins, a pre-validation study using a second, independent set of serum samples was performed where the recombinant protein was arrayed on nitrocellulose slides and probed with serum from a separate cohort of normal, pancreatitis, diabetic and pancreatic cancer patients.

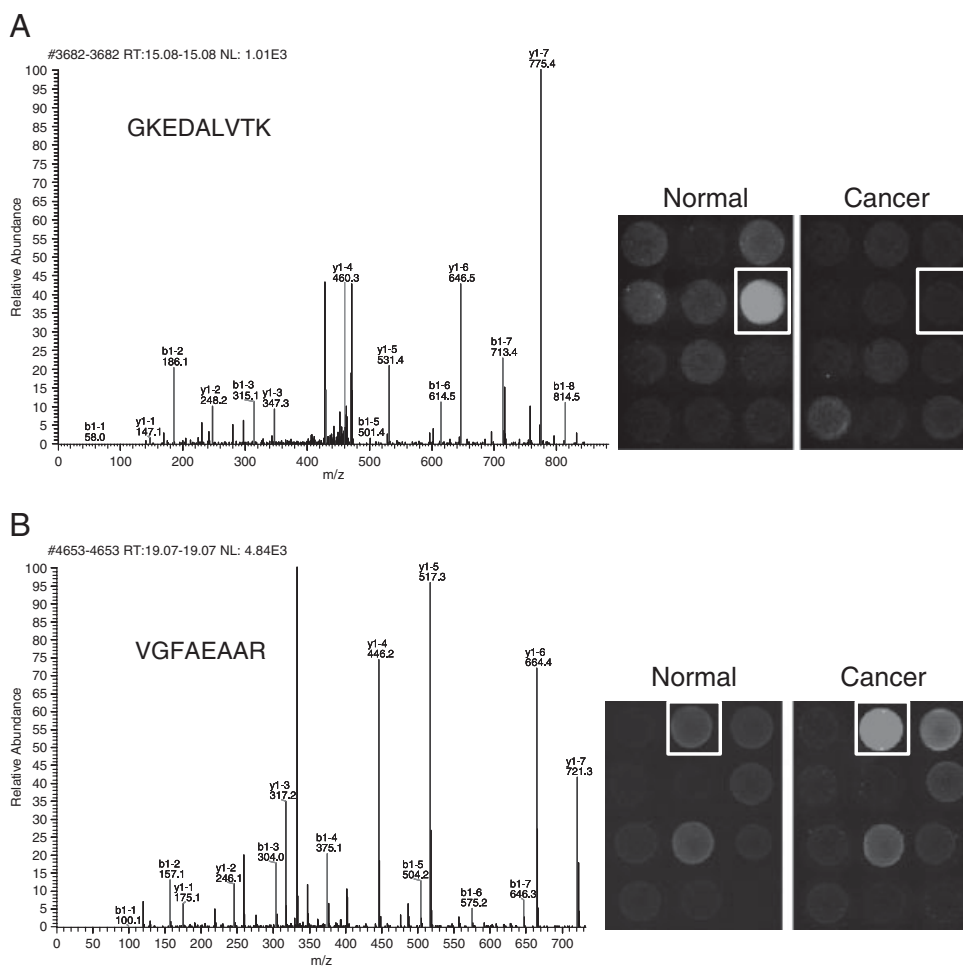
## 5.2 2-D liquid separation

MIAPACA proteins were separated by CF from pH 9.2 to 3.9, and each CF fraction was subsequently further separated by NPS-RP-HPLC. Figure 2 represents the 2-D UV chromatogram from these separations. The horizontal axis depicts all the fractions from the first dimension from lower *pI* to higher *pI*. The vertical axis represents retention times from the second dimension separations and the horizontal axis corresponds to increased hydrophobicity. A

typical 2-D separation across the pH range above results in about 1300 total fractions which are subsequently printed. A majority of these fractions contain one major protein since manual collection by peak is performed and only the center 6 s of the peak is collected. However, there are instances when more than one major protein may be present in the peak particularly for more highly abundant proteins that elute over a longer time.

## 5.3 Microarray printing and processing

The separated proteins were printed on nitrocellulose slides and probed with sera from normal individuals and patients diagnosed with pancreatic cancer. The immune response in the sera was visualized using an antihuman-IgG-Alexaflor647 conjugate. Figure 3 illustrates portions of the arrays printed on nitrocellulose slides to indicate the typical appearance of slides and spot quality, with specific examples of differential humoral responses. Tandem mass spectra are also shown to indicate the protein identity present in the spot of interest. It can be seen that spot intensities appear homogeneous throughout the spot. However, it was found that some fractions from the separated MIAPACA lysates were not printed on all the



**Figure 3.** Selected microarray shots of differential humoral response as well as selected tandem mass spectrum for sequence confirmation of (A) fibrinogen and (B) cathepsin D.



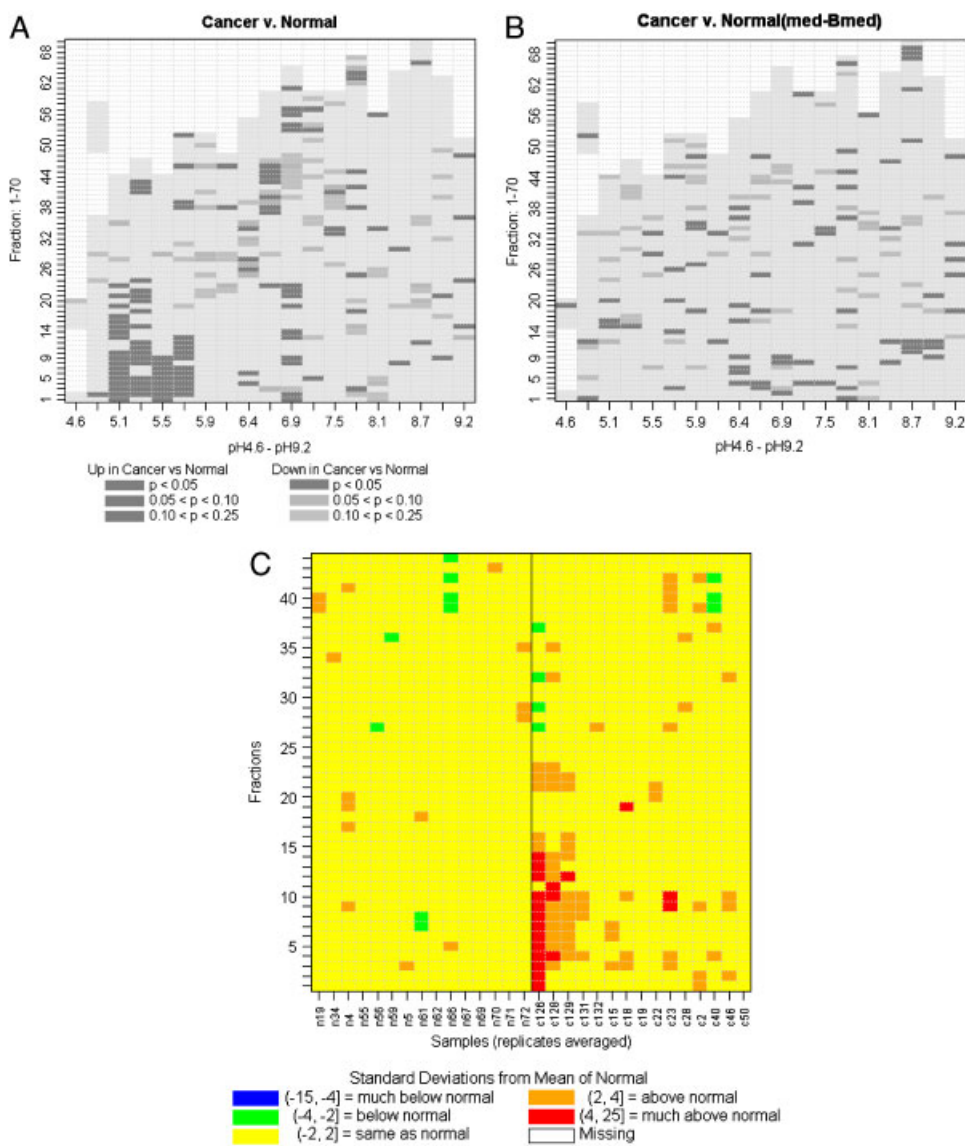
nitrocellulose pads due to incorrect calibration of the printing surface and printing errors that occurred during the print run for the lower pH fractions. Thus, all subsequent data representations indicate these missed spots, which were not considered further in the statistical analysis.

#### 5.4 Statistical analysis of immune response

Although looking at the humoral response from one normal and one cancer serum sample may indicate a difference as shown in Fig. 3, it is critical to assess this response in a larger set to identify if the difference is indeed statistically significant. Two analysis approaches were used to analyze the humoral response differences in a test set of 15 control and 15 pancreatic cancer sera. The first approach utilized a non-parametric Wilcoxon rank-sum test which was repeated using both the locally derived background-subtracted

median spot intensities as well as foreground median intensities without background subtraction. In the analysis of gene expression microarray data there is an ongoing debate as to whether background corrections are too severe [31]. In the spirit of hypothesis generation we chose to forego background subtraction since we saw a large amount of signal are washed away by background correction (Fig. 4B) compared with the analysis of the foreground data alone (Fig. 4A). Given the method of fractionation, it was expected that in some cases neighboring fractions containing higher abundant reactive proteins would be correlated, thus producing hot (cold) regions. These hot spots (cold spots) – found only in the foreground  $p$ -value grid – are missing in the background-subtracted  $p$ -value grid, supporting our choice to focus on uncorrected measures.

In the pancreatic cancer data set, uniform increases or decreases across all cancer samples were not expected. We



**Figure 4.** A grid of  $p$ -values from Wilcoxon rank-sum tests between cancer and normal sera *per* separated fraction in (A) foreground only and (B) background-subtracted data. The grid is arranged according to the 2-D fractionation of the whole cell lysate and colored according to the level of significance of the direction of the difference between cancer and normal sera where gray indicates no evidence of change. (C) z-score plot for proteins separated from pH fraction 5.1–4.9. On the vertical axis are all fractions by increasing retention time and on the horizontal axis are each of the serum samples with which samples were probed. Red and yellow blocks represent responses significantly higher than the mean of the normal sample ( $4 < z < 25$  and  $2 < z < 4$ , respectively), whereas blue and green blocks represent responses significantly lower than the mean of the normal sample ( $-25 < z < -4$  and  $-4 < z < -2$ , respectively).

sought to identify those proteins that elicited a pancreatic cancer-specific humoral response in at least 20% of the samples. This is generally considered a relatively significant response in these autoantibody experiments. For an alternate view of the changes in the immune response between healthy and pancreatic cancer diagnosed patients, z-score plots of each studied pH range were also generated in which z-scores were calculated, *per* spot, using the mean and standard deviation of the normal samples only. Resulting z-scores were thus on a scale of standard deviations from the mean of the normal samples. Thus, if a fraction had a high-z-score it had well above the average normal reactivity at that spot. Likewise, a low-z-score indicated that the fraction had well below the normal reactivity. When plotted in grids of spot by sample, patterns could be easily discerned in cancer samples. An example of such a z-score grid is illustrated in Fig. 4C, where the multiple orange/red fractions across the cancer samples but not control samples are indicative of a protein of interest. Increases or decreases that persist across at least 20% of the cancer samples were pursued for further study.

### 5.5 Classification and cross-validation

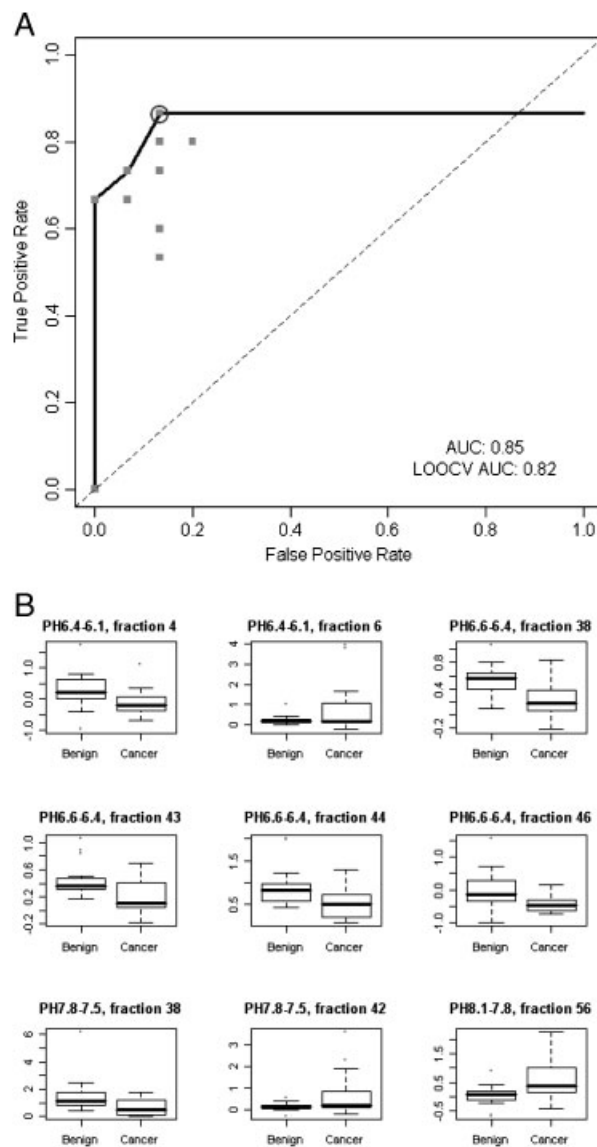
The PAM classification algorithm [29] was used to explore the classificatory power of the proteins found to be differentially expressed between control and pancreatic cancer sera. Differential proteins were selected as having (i) a Wilcoxon *p*-value of 0.05 or less or (ii) having over 20% of the cancer samples with a z-score  $> 2$  (or  $< -2$ ). The PAM algorithm selects the most predictive subset of proteins for classification. The best classifier, resulting in the smallest error using the fewest proteins, used nine proteins, chosen from 98 differential proteins, and only misclassified four samples (Fig. 5A). In an effort to estimate the generalizability of the classification analysis, an LOOCV was used. For the 30 leave-one-out cycles, the median size of the protein subset chosen for the classifier was 12 proteins (range = [4, 83]) which resulted in a median error rate of 4 (range = [2, 6]) and an average AUC of 0.82 (range = [0.63, 0.96]) for classifier selection. This is comparable to what we found when using all 30 samples.

From predictions of the left out sample, it was found that if generalized to a new population our classification analysis should predict the serum diagnosis with 86.7% accuracy (four misclassified samples). Among these four misclassified samples, three were false positives and only one was a false negative. This gives an expected sensitivity of 93.3% and an expected specificity of 80%.

To assess the stability of the classifier, we examined how frequently each protein was selected as an important predictor across the 30 LOOCV classifiers built. Two proteins (pH 6.6–6.4, fraction 44 and pH 8.1–7.8, fraction 56) were selected in all 30 LOOCV classifiers. Four other proteins were selected 22 times (pH 6.6–6.4, fraction 38, pH 6.6–6.4, fraction 43, pH 6.6–6.4, fraction 46 and pH 7.8–7.5, fraction 42). It is interesting to note that the nine protein

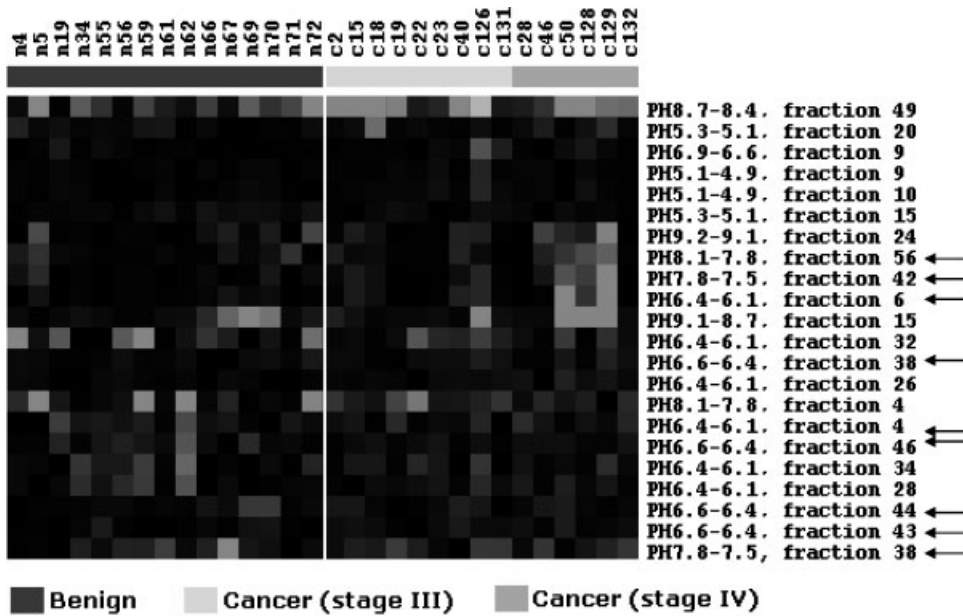
spots selected initially are among the most common proteins used in the LOOCV classifiers; see Supporting Information Table 1 column 1. Figure 5B illustrates the response of all serum groups to these nine proteins.

Figure 6 shows the scaled humoral response distribution across all serum samples considered to be differential on a scale of dark to light (lowest response to highest response), based on data from the Wilcoxon tests and z-score plots combined. The nine proteins spots that comprised the best classifier are indicated by arrows. The



**Figure 5.** (A) This ROC curve shows error rates attributed to each of the 30 threshold values considered by PAM, corresponding to 30 subsets of the 98 proteins. The circle highlights the chosen threshold that uses only nine proteins for classification, which provides the lowest error using the smallest number of proteins. The AUC for this ROC curve was estimated to be 0.85. (B) Boxplots of the intensities *per* diagnosis for each of the nine proteins included in the classifier panel built using all 30 samples.





**Figure 6.** Heatmap showing median-centered responses of all serum samples to selected proteins of interest. The scale from dark to light represents lower response to higher response on a scale of  $-2$  to  $2$ . The arrows indicate the protein spots that formed the panel of nine potential markers with highest sensitivity and specificity.

protein IDs of these are detailed in Supporting Information Table 1. In addition, the percentage of cancer samples in which the panel was able to distinguish from normal sera is also indicated.

### 5.6 Selected targets that respond in cancer sera

A panel of three proteins was found to discriminate pancreatic cancer sera with high sensitivity and specificity from normal sera by generating a higher response in cancer samples. Our efforts identified two of these three proteins. PGK-1 is a glycolytic enzyme but is also known to be active as a primer recognition protein. PGK-1 is known to show antigen activity in other types of cancers [32]. Histone H4 is a nuclear protein that maintains DNA in its proper configuration. As mentioned earlier, certain variants of histones have been implicated in the DNA repair process. The presence of antibodies against histone H4 in cancer sera but not in normal sera may serve as an important indicator of improper DNA regulatory mechanisms in cancer patients.

Although none of the proteins discussed were individually able to discriminate clearly between the two clinical groups, used together, as a nine-protein panel, they showed high specificity, sensitivity and selectivity, and may have potential diagnostic utility in the identification of patients with pancreatic cancer. We have focused on studying a pre-validation set using a larger, independent cohort of sera from patients with pancreatic cancer *versus* a set of control sera samples from normal patients and patients with chronic pancreatitis and type 2 diabetes. These are important controls since it is often difficult to discriminate between pancreatic cancer and pancreatitis and some pancreatic cancer patients develop type 2 diabetes.

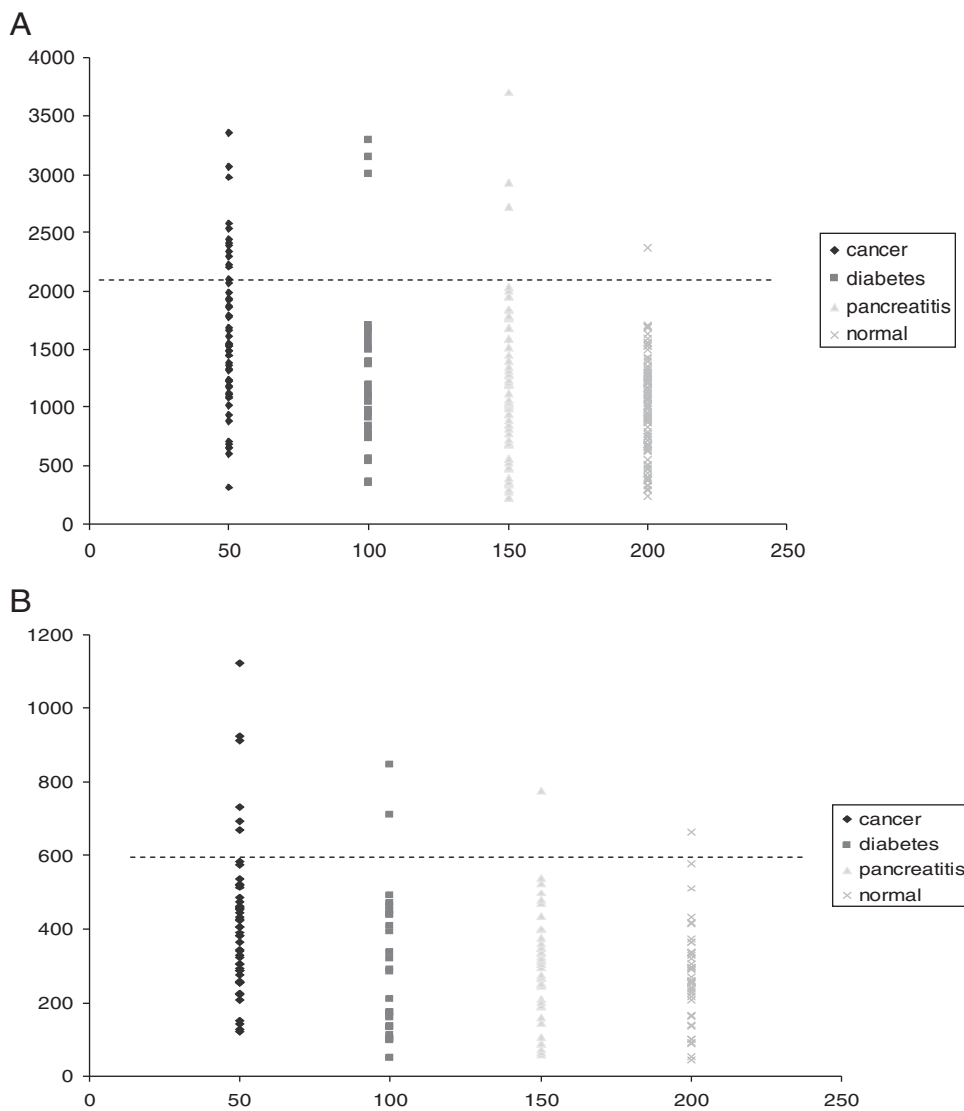
### 5.7 Pre-validation using recombinant proteins

In order to pre-validate proteins eliciting a humoral response, we initially tried using proteins fractionated from the Miapaca cell line, but we were not able to obtain enough material to spot a sufficient number of spots for the number of samples used in these studies. Alternatively, we were able to obtain recombinant proteins for further validation studies using a separate set of samples. The four recombinant proteins were chosen because they were available and the molecular weight was similar to that expected from the databases. Other commercial recombinant proteins with much larger molecular weight values than expected were not used since they may contain tags that interfere with the response. Chronic pancreatitis and diabetes serum samples were also used in these experiments to determine if the potential markers that were identified were unique for pancreatic cancer or more characteristic of an inflammatory or non-specific metabolic response. In order to measure the concentration of the autoantibody that is reactive against the recombinant proteins, the serum must be diluted properly such that the amount of available autoantibody in the serum should be lower than the binding capacity of the specific recombinant protein to avoid saturation while still providing good signal. Therefore, a saturation curve was made using different dilutions of serum to hybridize against identical blocks of the recombinant proteins. The result of the saturation test showed that with 400-fold dilution, the recombinant proteins were not saturated and all of them yielded a signal/background ratio of five or higher, which was felt to be optimal for the planned experiments. All the data of the pre-validation experiment were normalized with the control blocks to eliminate slide-to-slide variation. On each slide 16 identical blocks of recombinant proteins were printed in a format of two columns and seven rows. A block

in the middle of the each slide was used as a control block and hybridized with the same sample. The averaged intensity of the spots in the control block was considered to be a fixed number as  $A$ . The intensity for each of the recombinant protein spots was calculated as  $S \times A/B$  where  $B$  is averaged intensity of the spots in the control block on the specific slide.

Four proteins were used in these pre-validation experiments based on microarray discovery experiments using MIAPACA cells. These included PGK-1, histone H4, HSP27 and pterin-4- $\alpha$ -carbinolamine dehydratase, where commercial recombinant proteins were readily available for each of these. Recombinant proteins were arrayed on nitrocellulose slides and the slides were then probed with sera samples of normal, chronic pancreatitis, type 2 diabetes and pancreatic cancer sera. In the case of histone H4, 54 pancreatic cancer, 30 diabetes, 60 pancreatitis and 94 normal samples were analyzed (Fig. 7A). The signal in histone H4 case was strong and a response was observed for every

sample. A differential response was observed above the baseline as shown in Fig. 7A where there are clearly a much larger number of samples (14 samples, 25.9%) that respond to cancer as compared with diabetes (3 samples, 10%) or pancreatitis (3 samples, 5%). The response of cancer is much stronger than that of the normal samples and there is only one outlier point out of the 94 normal samples tested (1.1%). Also, among the cancer samples, none of the samples that responded above the baseline shown were diabetic. Overall, histone H4 gives a 25.9% sensitivity with a 66.7% positive predictive value (PPV) and a 96.2% specificity for detecting pancreatic cancer. Analysis of the presence of PGK-1 auto-antibodies in 49 pancreatic cancer samples, 30 diabetes, 42 pancreatitis and 43 normal sera samples (Fig. 7B) that were hybridized against the recombinant protein demonstrated a clear difference between pancreatic cancer and normal samples, where there is only one outlier for the normal samples. The response observed was also able to distinguish cancer from pancreatitis or diabetes. Overall, PGK-1 gives



**Figure 7.** Scatterplot illustrating the differential humoral response in recombinant human (A) histone H4 and (B) PGK-1 used for validating initial experimental results. Code: cancer, diamond; diabetes, square; pancreatitis, triangle; normal, x.

only a 12.2% sensitivity but retains a 60% PPV and a 96.5% specificity. It is interesting to note that in comparing the results from the response of histone 4 and PGK-1 that only two of the samples in the cancer set that responded above the baseline in the case of PGK-1 also were above the line among the 14 samples that responded in cancer for histone 4, suggesting that the autoantibodies developed toward these two proteins function in a complementary manner in identifying patients with pancreatic cancer. In the case of the diabetes and pancreatitis samples none of the samples that responded above the line were the same in either PGK-1 or histone 4. Using the two proteins jointly we can achieve a sensitivity of 33.0% while retaining a PPV of 62.1% and specificity of 94%.

Pterin carbinolamine dehydratase and HSP27 autoantibodies did not show a differential humoral response similar to the test set where both normal and pancreatic cancer samples responded. One possible reason for this lack of differential response could be the nature of the recombinant proteins that were arrayed. It is quite possible that the recombinant protein synthesized in bacteria did not possess key modifications responsible for the antigenicity of the endogenous proteins. Alternatively, a eukaryotic system could be used to produce these proteins, but these would not contain the modifications as expressed in the disease state. Thus, so far we have not been able to identify any PTMs in these two proteins extracted from the liquid separations method. Another issue could be biases introduced in the relatively small discovery set used in this work. It should be noted that histone H4 and PGK-1 were found to have a positive response in cancer compared with normals in our discovery microarray analysis, whereas Pterin carbinolamine dehydratase and HSP27 were found to have a stronger response in normals compared with cancer. The latter two proteins may not be real markers since autoantibodies result from new proteins that are secreted from disease cells and it is unlikely to find such responses that are on all the time from normal cells. The pre-validation set eliminates such proteins that may be falsely identified.

The pre-validation studies showed that PGK-1 and histone H4 do in fact differentiate normal and chronic pancreatitis and pancreatic cancer sera. This differential humoral response is present at an overall higher degree in the cancer serum compared with the pancreatitis serum. However, it is unlikely that this differential humoral response was in fact due to inflammation alone since it was not entirely unique to the pancreatitis cancer serum group. The response for the diabetes samples was present so that this factor could potentially interfere with analysis. However, none of the cancer samples that responded above the baseline were diabetic. Also, the response is not “absent in normal and present in pancreatitis or cancer” and there are always only a limited number of samples that respond to any marker with this method; hence, these proteins are not suitable as single markers for diagnostic purposes. However, their ability to distinguish normal *versus* cancer sera *versus* pancreatitis provides important potential markers

for pancreatic cancer that may be used as part of a larger panel of markers for the autoantibody response.

## 6 Concluding remarks

A humoral response to tumor proteins may have utility for the detection of pancreatic cancer. We have used 2-D liquid separation and protein microarrays to study the humoral response in pancreatic cancer. Several different statistical treatments of results were used to highlight proteins that elicited a differential humoral response pattern between the different clinical groups. Rank-based statistics (Wilcoxon rank-sum tests) highlighted differences between the two clinical groups. Significant variability existed between the measurements obtained with the cancer sera and z-score statistics was utilized as a complementary statistical tool to further analyze the differences between the cancer and control samples.

The PAM classification algorithm and LOOCV highlighted a panel of nine spots that was able to classify groups with high sensitivity and specificity. Furthermore, an independent pre-validation study using available human recombinant proteins was able to substantiate results obtained with LOOCV for PGK-1 and histone H4. This validation study was pursued on an independent cohort of samples where for the case of histone H4 54 pancreatic cancer, 30 type 2 diabetes, 60 pancreatitis and 94 normal sera samples were used against the recombinant protein. The study indicated that a strong differential humoral response was observed for pancreatic cancer sera compared with chronic pancreatitis and type 2 diabetes control samples. The response of pancreatic cancer was much stronger based on the number of samples that responded above the baseline compared with the large number of normal sera samples used in these experiments. The response to PGK-1 also provided a stronger response of cancer compared with pancreatitis and diabetes and a much stronger response compared with normals. Jointly the two proteins achieve a 33.0% sensitivity with 62.1% PPV and 94.0% specificity for the detection of pancreatic cancer. These two proteins at present are not clinically useful in themselves but may serve as part of a larger panel to detect pancreatic cancer. This still needs to reach a sensitivity of 85% and a specificity of 85% to compare with CA19-9, which has been used as a marker for pancreatic cancer [33].

Other proteins did not provide a differential response where it is possible that all recombinant proteins used did not provide optimal results because they were not in their active form, *i.e.* the correct isoform or PTM was absent or simply the limited number of samples in the test set resulted in biased results. A study comparing the printed protein in the initial study with the recombinant protein to verify this hypothesis could not be performed because of insufficient sample from the initial study. Proteins that have been previously implicated in cancer progression as well as

