

Published in final edited form as:

Cancer Biomark. 2010 January 1; 7(1): 25–37. doi:10.3233/CBM-2010-0145.

The identification of auto-antibodies in pancreatic cancer patient sera using a naturally fractionated Panc-1 cell line

Chen Li^{a,1}, Hye-Yeung Kim^{b,1}, Huy Vuong^c, Tasneem Patwa^d, Manoj Pal^e, Randall E. Brand^f, Diane M. Simeone^b, and David M. Lubman^{a,b,c,g,*}

^a Department of Chemistry, The University of Michigan, Ann Arbor, MI, USA

^b Department of Surgery, The University of Michigan Medical Center, Ann Arbor, MI, USA

^c Bioinformatics Program, The University of Michigan, Ann Arbor, MI, USA

^d Pfizer Global Research and Development Eastern Point Road Groton, CT, USA

^e Genomics Institute of the Novartis Research Foundation, CA, USA

^f Division of Gastroenterology, Hepatology & Nutrition, The University of Pittsburgh, Pittsburgh, PA, USA

^g Comprehensive Cancer Center, The University of Michigan Medical Center, Ann Arbor, MI, USA

Abstract

The immunogenic nature of cancer can be explored to distinguish pancreatic cancer from related non-cancer conditions. We describe a liquid-based microarray approach followed by statistical analysis and confirmation for discovery of auto-immune biomarkers for pancreatic cancer. Proteins from the Panc-1 pancreatic cancer cell line were fractionated using a 2-D liquid separation method into over 1052 fractions and spotted onto nitrocellulose coated glass slides. The slides were hybridized with 37 pancreatic cancer sera, 24 chronic pancreatitis sera and 23 normal sera to detect elevated levels of reactivity against the proteins in spotted fractions. The response data obtained from protein microarrays was first analyzed by Wilcoxon Rank-Sum Tests to generate two lists of fractions that positively responded to the cancer sera and showed p-values less than 0.02 in the pairwise comparison between cancer specimens and normal and chronic pancreatitis specimens. The top 3 fractions with the lowest correlations were combined in receiver operating characteristic analyses. The area-under-the-curve (AUC) values are 0.813 and 0.792 for cancer vs. normal and cancer vs. pancreatitis respectively. Outlier-Sum statistics were then applied to the microarray data to determine the existence of outliers exclusive in cancer sera. The selected fractions were identified by LC-MS/MS. We further confirmed the occurrence of outliers with three proteins among cancer samples in a confirmation experiment using a separate dataset of 165 serum samples containing 48 cancer sera and 117 non-cancer controls. Phosphoglycerate kinase 1 (PGK1) elicited greater reactivity in 20.9% (10 in 48) of the samples in the cancer group, while no outlier was present in the non-cancer groups.

Keywords

Pancreatic cancer; microarray; humoral response; PGK; Panc-1; Outlier-sum

© 2010 – IOS Press and the authors. All rights reserved

*Corresponding author: David M. Lubman, Ph.D., The University of Michigan, Department of Surgery, 1150 West Medical Center Drive, Building MSRB1, Rm A510B, Ann Arbor, MI 48109, USA. Tel.: +1 734 647 8834; Fax: +1 734 615 2088; dmlubman@umich.edu.

¹Contributed equally to this work.

1. Introduction

Pancreatic adenocarcinoma is the 4th leading cause of cancer-related death in the United States [1]. Pancreatic ductal adenocarcinoma (PDAC) has one of the poorest survival rates of any cancers, where according to the American Cancer Society, for all stages of pancreatic cancer combined, the one-year relative survival rate is 20%, and the five-year rate is 4% [1]. These low survival rates result from the failure to diagnose PDAC at an early stage when the possibility of a curative resection still exists. This is due to a variety of factors including the inaccessible location of the pancreas deep in the abdomen, late-presenting clinical manifestations (e.g., weight loss, or abdominal pain), and the early development of metastasis. Fewer than 10% of patients' tumors are confined to the pancreas when in most cases, diagnosis of 80% to 90% of PDAC cases are too late for surgical procedures to have a positive outcome. Unfortunately there are not any available diagnostic tools that allow for detection of early stage pancreatic cancer. Although there has been an effort to find protein markers in serum, none have shown sufficient sensitivity and specificity for early diagnosis including the commonly used CA 19-9 test [2,3] which may be significantly increased in pancreatitis in addition to pancreatic cancer, and is not reliably elevated in early stage cancer.

There remains a need for the discovery of innovative serological biomarkers that effectively improve diagnosis and prognosis of human cancer. Antibody responses associated with the occurrence and progression of solid tumors have been identified in multiple cancer types [4–6]. The underlying mechanism of the auto-immune response is still not fully understood [7]. However, the known molecular changes that can induce auto-immune response include proteins expressed at an aberrant level, and mutated gene products and isoforms of proteins with abnormal post-translational modifications (PTMs) [8–10]. The immunogenic proteins are often found to be intracellular proteins whose functions are linked to the onset and growth of malignant tumors, such as oncoproteins HER-2/Neu and c-MYC [11–13] and, tumor suppression proteins such as p53 [14].

Although research has suggested strong correlation between the presence of some auto-antibodies and the process of tumorigenesis, the frequency of the appearance of auto-antibodies in cancer patients varies i.e. elevated level of a specific autoantibody is always present in a variable subset of patients [5]. A mutation in the p53 gene elicits an auto-immune response in 4–30% of patients in several types of cancer [14]. Around 30% of patients with lung adenocarcinoma exhibit a humoral response to glycosylated annexins I and/or II whereas none of the noncancerous standards exhibit such a response [15]. In PDAC, auto-antibodies to DEAD-box protein 48 were observed in 33.3% of pancreatic patient sera, while none of the patients with benign disease and healthy controls showed reactivity against the antigen [16]. MUC1 [17,18], p53 [19] and Rad51 [18] have also shown restricted immune reactivity in a subset of cancer samples. The typical frequency of the detection of a particular auto-antibody in a cancer type is 10–20% and may not be sufficient when used as a biomarker individually, but may be combined as a panel for improved performance [20–22].

Adding new cancer specific auto-antigens to the existing biomarker repertoire is the impetus of developing analytical and statistical techniques for auto-immune response studies. There are several approaches currently available for the identification of auto-antigens. One is targeting specific proteins or gene products that are known for their roles in cancer including p53, c-myc, and erB-2. This method only provides limited candidates for biomarkers. Recombinant protein microarrays produced from cDNA expression libraries has been used as a comprehensive antigen substrate to profile the auto immune reactivity, however, it is

unable to profile PTM dependent antibody-antigen interactions [5]. The development of proteomic separation and identification techniques has benefited the discovery of auto-antibody biomarkers, where proteins from tissue or cell lines are fractionated by gel or liquid based multidimensional separations, while maintaining the natural PTMs.

In current work, we have used liquid fractionation methods to produce microarrays for the humoral response experiment against a Panc-1 pancreatic cancer cell line. The methods that are used involve separating intact proteins from cell lysates using two dimensions. A total cell fractionation can be performed using chromatofocusing separation in the first dimension where the proteins are fractionated according to pI. Each fraction is then separated in a second dimension by non-porous silica RP HPLC [23]. Using this method isolated proteins in the liquid phase can be collected for spotting on coated glass slides [24]. The protein spots are probed for their humoral response by exposing them to sera from cancer and chronic pancreatitis patients, and normal individuals. This method offers a means for comprehensive proteomic analysis of proteins from large numbers of purified proteins as expressed in cancer cells while maintaining their PTMs that are often critical to the humoral response [25]. The method can produce arrays with over a thousand spots and can produce large numbers of slides for testing the response against a large number of patients.

In order to account for the fact that a specific autoantibody is more often found in only a subset of the patients than all of them with the corresponding tumor, we attempt to apply two statistical methods, Wilcoxon Rank-Sum Test and Outlier Sum Statistics [26], to find potential markers that show different types of immune reactivity patterns. Based on our results we perform a confirmation study of 5 potential markers for pancreatic cancer against recombinant proteins on a microarray based format against samples from pancreatic cancer, pancreatitis, diabetes type 2 and normal controls. For three of the five proteins, a substantial number of samples from the cancer group show higher reactivity than the non-cancer sample groups.

2. Experimental section

2.1. Chemicals

Methanol, acetonitrile, urea, thiourea, iminodiacetic acid, dithiothreitol (DTT), n-octyl-D-glucopyranoside (OG), glycerol, bis-tris, Trifluoroacetic acid (TFA), and PMSF (Phenylmethanesulfonyl fluoride) were purchased from Sigma (St. Louis, MO). Water was purified using a Milli-Q water filtration system (Millipore Inc., Bedford, MA) and all solvents were HPLC grade unless otherwise specified. Reagents used were in the purest form commercially available. Polybuffer 74 and polybuffer 96 were purchased from GE Healthcare BioSciences Corp. (Piscataway, NJ). 1x PBS and ultra-pure DNase/RNase free distilled water were obtained from Invitrogen (Carlsbad, CA).

2.2. Serum samples

As a discovery set, eighty six serum samples were obtained at the time of diagnosis following informed consent using IRB-approved guidelines. Sera were obtained from patients with a confirmed diagnosis of pancreatic adenocarcinoma in the Multidisciplinary Pancreatic Tumor clinic at the University of Michigan Hospital. 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 inability to provide informed consent, patient's actively undergoing chemotherapy or radiation therapy for pancreatic cancer, and patients with other malignancies diagnosed or treated within the last 5 years. Sera were also obtained from patients with chronic pancreatitis who were seen in the Gastroenterology Clinic at University of Michigan Medical Center and from control healthy individuals collected at the

University of Michigan under the auspices of the Early Detection Research Network (EDRN). Some pancreatic cancer samples were obtained under IRB approval from UPMC and processed similarly following EDRN guidelines. The mean age of the tumor group was 65.4 years (range 54–74 years) and the chronic pancreatitis group was 54 years (range 45–65). The sera from the normal subject group and the tumor group were similar in age and sex. The chronic pancreatitis group was sampled when there were no symptoms of acute flare of their disease. All sera were processed using identical procedures. The samples were permitted to sit at room temperature for a minimum of 30 minutes (and a maximum of 60 minutes) to allow the clot to form in the red top tubes, and then centrifuged at $1,300 \times g$ at 4°C for 20 minutes. The serum was removed, transferred to polypropylene, capped tubes in 1 ml aliquots, and frozen. The frozen samples were stored at -70°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. In addition to the discovery set, another set of samples with no overlap with the discovery set was used in the confirmation experiment. The demographic and clinical information of the samples in the confirmation set are shown in Table 2.

2.3. Cell culture

The Panc-1 PDAC cell line was cultured in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum, 100 units/mL penicillin and 100 units/mL streptomycin (Invitrogen, Carlsbad, CA). Upon reaching 80% confluence, the cells were washed twice in 10mL 1X PBS containing 4 mM Na_3VO_4 , 10 mM NaF and one half of a protease inhibitor cocktail tablet. The sample was then solubilized in 300 ul lysis buffer consisting of 7 M urea, 2 M thiourea, 100 mM DTT, 0.5% biolyte ampholyte 3–10, 2% OG, 4 mM Na_3VO_4 , 10 mM NaF and 1 mM PMSF at room temperature for 30 min, followed by centrifugation at 35000 rpm at 4°C for 1hr. The supernatant was stored at -80°C until further use.

2.4. Chromatofocusing(CF)

Prior to CF, a PD10 column(Amersham Biosciences) was used to exchange the cell lysate from the lysis buffer solution to the CF buffer solution according to the manufacturer's protocols. The start buffer consisted of 6 M Urea, 0.2% OG, 25 mM bis-tris. The elution buffer solution was composed of 6 M urea, 0.2% OG, and a 10 fold dilution of polybuffer 96 and poly-buffer 74 in a ratio of 3:7. The pH of both buffer solutions (7.9, 4.0) was adjusted with saturated iminodiacetic acid. A chromatofocusing column (weak anion exchange HPCF-1D prep column, 250 mm \times 4.6 mm ID, Eprogen, Darien, IL) was pre-equilibrated with the start buffer solution and 13 mg of the cell lysate was injected into the CF column with multiple injections. Fractionation was started after switching elution buffer and a stable base line achieved. The pH fractions were collected in 0.3 pH intervals and pH was monitored using a flow-through on-line pH probe. UV absorption was recorded at 280 nm. When a pH of 4.0 was reached, elution buffer solution was switched to a 1M NaCl solution to wash the column followed by Isopropanol to elute out strongly bound proteins from the column. The collected fractions were stored at -80°C .

2.5. Reverse phase HPLC separation

An ODSI-1 (8 \times 33 mm) column (Eprogen, Inc.) was used to separate the pH fractions of the Panc-1 cell line after CF. Solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in acetonitrile. The gradient was run from 5% to 15% B in 1 min, 15% to 25% in 2 min, 25% to 31% in 2 min, 31% to 45% in 10 min, 41% to 47% in 6 min, 47% to 67% in 4 min, 67% to 100% B in 3 min, and reduced to 5% B in 1min after maintaining 100% B for 1 min. The flow rate was 1 ml/min and the column temperature 65°C . UV absorption was monitored at 214 nm. The fractions were collected in 96 well plates and stored at -80°C .

2.6. Protein microarrays

Approximately 30% of the total sample of the fractionated Panc-1 proteins obtained using 2D separation were transferred into 96-well printing plates (Bio-Rad) and lyophilized to dryness. The fractions were reconstituted in printing buffer which was composed of 62.5 mM Tris-HCl (pH 6.8), 1% w/v sodium dodecyl sulfate (SDS), 5% w/v dithiothreitol (DTT) and 1% glycerol in 1 X PBS. Reconstituted fractions in the printing plate were placed in a shaker overnight at 4°C. The fractions from the printing plate were spotted onto nitrocellulose slides using a non-contact piezoelectric printer (nanoplotter 2 GeSiM). Each spot contained 2.5 nL of liquid of ~450 μm diameter, and the distance between spots was 600 μm . Printed slides were dried on the printer deck overnight and stored in a refrigerator desiccated at 4°C if the slides were not used immediately.

2.7. Hybridization of slides

The printed slides were blocked in a solution of 1% BSA in PBS-T (0.1%) overnight. Each serum sample was diluted 1:400 in probe buffer which consisted of 1% BSA, 0.5 mM DTT, 5 mM magnesium chloride, 0.05% Triton X-100, 5% glycerol in 1 X PBS. The slides were hybridized in diluted serum for 2 hrs using a mini-rotator at 4°C. After hybridization, slides were washed five times using probe buffer for 5 min each time, and then re-hybridized with goat-anti-human IgG conjugated with Alexafluor 647 (1 $\mu\text{g}/\text{mL}$, Invitrogen, Calsbad, CA) for 1 hr at 4°C. The slides were washed five times again with probe buffer for 5 min each and dried. All slides were scanned using an Axon 4000B microarray scanner (Axon Instruments Inc., Foster City, CA).

3. Data acquisition and analysis

3.1. LC-MS/MS

The residual two-thirds of the sample in 96 well plates which was not used in microarray experiments were dried down to approximately 10 μL and mixed with 10% (v/v) ammonium bicarbonate, 10% (v/v) DTT, and 1:50 ratio (v/v) TPCK-treated trypsin (Promega, Madison, WI). The solution was incubated at 37°C overnight and the tryptic digestion was terminated by addition of 2.5% (v/v) of TFA. The digested peptide mixture was analyzed by nano-flow reverse-phase LC/MS/MS using the LTQ mass spectrometer with a nano-spray ESI ion source (Thermo, San Jose, CA). The samples were separated using a (0.1 \times 150 mm) capillary reverse phase column (MichromBioresources, Auburn, CA) with a flow rate of 5 $\mu\text{L}/\text{min}$. An acetonitrile:water gradient method was used, starting with 5% acetonitrile which was ramped to 60% in 25 min and to 90% in another 5 min. Both solvent A (water) and B (ACN) contained 0.3% formic acid. The electrospray voltage was 2.6 kV, with a capillary temperature of 200°C and a capillary voltage of 4 kV. The normalized collision energy was set at 35% for MS/MS. The MS/MS spectra obtained were analyzed using the Sequest feature of Bioworks 3.1 SR1, allowing only one missed cleavage during SwissProt human protein database searching. To further validate data obtained from Sequest, Protein prophet/peptide prophet software modified in house was used to provide a confidence level in identification of 95%. Since there might be more than one protein in a protein spot on the microarray slide, we compared proteins identified in adjacent fractions. If the spot that responded to the humoral response was unique and did not have an adjacent spot that lit up then the highest scoring protein based on LC-MS/MS analysis and protein prophet/peptide prophet was considered as the likely identification. If more than one protein was identified in the spot, then we also performed mass spec analysis on the adjacent spots. If the proteins were identified in the adjacent spots that did not respond then they were likely not to be the protein with the humoral response in our unique spot. However, if adjacent spots also showed a humoral response then the protein present in all spots was considered as the most likely candidate.

4. Statistical analysis

GenePix 6.0 software was used to grid all spots, to determine the fluorescent intensities at wavelength 635 nm and median local background intensities for each spot. Background subtracted data of the spots was taken into analysis if the foreground measure was at least 2X the background intensity measure. The signal intensities from all the slides are normalized to minimize experimental slide-to-slide variation. The data for each individual sample in the columns is centered by the median and scaled by the interquartile range (IQR). Two types of statistical analysis were applied to the normalized data in search of biomarkers with up-regulated response in the cancer samples compared to the normal and pancreatitis samples. The non-parametric Wilcoxon rank-sum test was employed to identify fractions showing a universally increased reactivity in the cancer samples. The Outlier-sum test was performed to select the fractions that react with only a subset of the samples in the cancer group.

5. Wilcoxon Rank-Sum test

Two pair-wise Wilcoxon Rank-Sum Tests were performed between cancer versus normal and cancer versus pancreatitis. The fractions with the lowest p-value and minimal correlation were combined in Receiver Operation Characteristic (ROC) analyses to determine their sensitivity and specificity in differentiating the sample groups. The Wilcoxon Rank-Sum Tests and the ROC analyses are programmed in R.

6. Heatmap

The fractions with a p-value less than 0.02 in Wilcoxon Rank-Sum tests are clustered and shown in heatmaps. The $p < 0.02$ threshold was determined to have proper numbers of fractions to show in the heatmaps. The heatmap and dendrogram are drawn in R.

7. Outlier Sum Statistics (OS)

The dataset is first standardized for each fraction by subsequently subtracting the median and dividing the median absolute deviation (MAD). The 75% quartile ($q(75)$) plus the interquartile range ($q(75) + IQR$) is used as a threshold. The data points beyond this threshold are defined as the outliers. The outlier-sum statistic is the sum of the values of these data points in the disease groups. Fractions with outlier sum statistics ranked top 5% and no outliers in the normal groups were considered to be differential. The overlapping fractions found in the comparisons between cancer/normal and cancer/pancreatitis are presented in bar graph form (Fig. 4) (made in R with COPA package).

8. Confirmation using recombinant proteins

Recombinant proteins were purchased from Abnova Corporation (Taiwan), and Genway Biotech Inc., (SanDiego, CA). The concentration of each recombinant protein was 10 ug/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). Each spotting event that resulted in 500 pL of solution being deposited was programmed to occur 5 times/spot to ensure that 2.5 nL was deposited on each spot. 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 (PBS-T 0.1) and then blocked with 1% bovine serum albumin (Roche) in PBS-T 0.1 for one hour. 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 PBS-T 0.1 containing 0.1% Brij. One hundred microliters of each diluted sample was applied to the recombinant protein array and the hybridization was performed in a humidified chamber for one hour. The 165 samples from different groups were perfectly balanced on each slide to eliminate bias from block-to-block variation and slide-to-slide variation. Two blocks on each of the slides were hybridized with two specific samples and used as control blocks for data normalization. The slides were then rinsed three times to remove unbound proteins. 1 µg/mL goat anti-human IgG conjugated with Alexafluor647 (Invitrogen, Carlsbad, CA) solution was used for detection. After a second one-hour hybridization with anti-human 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. The signals from different slides were normalized with the averaged signal of the control blocks on each slide.

9. Results and discussion

The proteins from Panc-1 human pancreatic ductal adenocarcinoma (PDAC) cell line were used as bait to study the humoral response in pancreatic cancer since the Panc-1 cell line has been used as a good representative sample of human pancreatic cancer [27]. The analytical work flow is illustrated in Fig. 1. The solubilized protein solution extracted from Panc-1 cell line was fractionated using 2-D liquid separation methods as described consisting of chromatofocusing in the first dimension followed by nonporous reversed phase HPLC where intact proteins were collected as the final product. Fraction collection was performed where liquid eluent from each chromatographic peak was collected into 96 well plates. Each collected protein fraction was separated into two parts for further work. One portion was used for spotting the microarray plates and a second portion was used for protein identification based on LC-MS/MS. There were 1052 protein peaks obtained over a pH range of 8.0–4.0 spotted using the microarray device onto each nitrocellulose coated glass slide. Each slide was hybridized against a patient serum sample where the humoral response was run in this work against 37 cancer serum samples, 24 pancreatitis serum samples and 23 normal controls. Statistical analysis including non-parametric Wilcoxon Rank-Sum Tests and Outlier-Sum Statistics were then performed over this sample set to determine which proteins provided a significant response to patient sera. For the selection of identified proteins, a confirmation study using a second, independent set of 168 serum samples was performed where five recombinant proteins were arrayed on nitrocellulose slides and probed with serum from a separate cohort of normal, pancreatitis, diabetes and pancreatic cancer patients.

10. Microarray result of humoral response

The heterogeneity of humoral response has been displayed in a substantial percentage of patients with increased antibody expression to disease-related antigens, where only a subset of patients has an autoimmune response to a particular antigen. We herein assume that autoimmune markers show either an increased level of reactivity against most of the patient sera or an outlier pattern that exclusively appear in the cancer group. Two statistical methods, Wilcoxon Rank-Sum and Outlier Sum Test, were applied to the dataset to search fractions for auto-antibody response.

11. Statistical analysis

Compared to traditional T-Test, Wilcoxon Rank-Sum Test was preferred in several previous humoral response studies because the dataset do not always fit a Gaussian distribution. The test generates a list of fractions with significantly greater intensities in the cancer group (p-value set at < 0.02) in the pairwise comparisons in cancer versus normal and cancer versus

pancreatitis. Twenty-nine fractions were selected in the cancer/normal pair and only seventeen passed the threshold in the cancer/pancreatitis pair. Figure 2 shows two heatmaps of these fractions after they are clustered using a hierarchical clustering algorithm. The clustering tree is added on top of the heatmaps. In the first heatmap/dendrogram, 65% (24 out of 37) of the cancer samples and only 17% (4 out of 23) normal samples are clustered on the left side with more blue bands which indicate increased reactivity with serum. Similarly the left side of the second heatmap/dendrogram includes 70% (26 out of 37) of the cancer samples and 29% (7 out of 24) of pancreatitis samples. Most of the samples are clustered with their own groups while a portion of the samples are not. Several reasons can be taken into account for this result. The incorrectly clustered cancer samples may not contain the antibodies to some particular antigens in the Panc-1 cell line. Additionally, the non-cancer samples that incorrectly clustered with the majority of the cancer samples may be reactive to the co-eluted proteins in some fractions containing cancer-related antigens.

In a 2-dimensional separation, a protein often appears in two or more subsequent fractions rather than one because of the limited chromatographic resolution and also the post-column diffusion. In Fig. 2, it is worth noting that the fractions in the heatmap are often accompanied by their adjacent fractions (red circled) ex. 7B1-7B3 and 4E4-4E11. The consecutive bands with a smooth reactivity profile are better candidates for further investigation and confer important information for protein identification.

The result of the Wilcoxon Rank-Sum Tests can be transformed to a ROC curve which estimates the ability of the selected biomarkers to distinguish case from non-case. For each ROC curve, an area-under-the-curve (AUC) is reported, where 1.0 represents perfect separation of one group from the other, 0.5 represents a completely random result or no separation. In both cancer vs. normal and cancer vs. pancreatitis categories, the fractions top-ranked in Wilcoxon Rank-Sum Tests, exhibit AUC values of 0.70–0.72. To improve the AUC value, we combined the top three fractions with the lowest correlations in ROC analyses so as to avoid combining neighboring fractions. The AUC values of the ROC curves (Fig. 3) for the three combined fractions are 0.813 and 0.792 for cancer vs. normal and cancer vs. pancreatitis respectively.

12. Outlier sum statistics

Recently, statistical outlier sum methods, such as COPA [28] and OS [26] have been proposed as methods for searching cancer related genes with microarray techniques. The outlier sum analysis is able to detect a small number of significantly up-regulated signals from microarray data in the disease group while the signal from the majority may not necessarily change. Since the majority of cancers have heterogeneous activation for different individuals, it appears that the application of this method using the “subset” idea where some cancers respond to the humoral response and others do not respond may result in an improved performance for microarray data. Of the two outlier sum methods, COPA identifies pairs of biomarkers with mutually exclusive up-regulated samples because it was designed to search for gene activation with a mutually exclusive mechanism, while the protein biomarkers in this work may not have the same feature. OS identifies outliers in a similar manner as COPA, but it calculates an outlier score for each individual. Therefore, OS is the preferred method in this study.

After OS analysis was applied to the dataset, we found 9 fractions (listed in Table 1) that ranked in the top 5% in both of the comparisons of cancer/normal and cancer/pancreatitis. The reactivity profiles for the top 3 fractions are shown in bar graphs in Fig. 4. It appears that only subsets of the cancer group show increased reactivity against these fractions, while the signals from the other samples remain the same. The signals of subsets with increased

reactivity in the cancer group are much higher than the range of the non-cancer groups, which remain close to the baseline. Such an outlier pattern for these fractions indicates that they contain proteins that are only immunogenic for a subgroup of the cancer samples and not immunogenic for all the non-cancer samples. In clinical application, these fractions can provide information for accurate diagnosis as the immunogenic cancer samples distinguish themselves with a high signal.

The results from Wilcoxon Rank-Sum test and Outlier Sum analysis are compared, where the lists of marker fractions do not overlap. We are more interested in the candidates given by OS, as those fractions exhibiting an outlier pattern exclusively in the cancer group would be more useful in diagnosis. The list of candidate fractions from the OS were thus identified with mass spectrometry and their IDs and performance were confirmed with the recombinant protein array.

13. Mass spectrometry identification

LC-MS/MS is used to identify the proteins in these fractions and their adjacent fractions. As expected, multiple proteins were identified in each of the fractions. The identified proteins were screened based on an assumption that their reactivity profile should be consistent with their appearance in the neighboring fractions. The resulting protein IDs are listed in Table 1 for each of the fractions.

13.1. Biomarker confirmation

Due to the large number of fractions from the 2-dimension separation, a set of only 84 serum samples was used to search for the fractions that could be potential biomarkers, where a larger set is usually required for confident biomarker discovery. It is also necessary to confirm the protein IDs identified for the candidate fractions. To confirm these potential markers, recombinant proteins were tested with a different sample set. Five commercially available recombinant proteins were selected for the confirmation experiment with 48 samples from the cancer group, 40 samples from pancreatitis group, 40 samples from normal group, and 37 samples from diabetes group. Type 2 diabetes samples are included since some pancreatic patients also develop this condition which might be responsible for the autoimmune reactivity.

In order to measure the auto-antibody response that is elicited against the recombinant proteins correctly, care must be taken to avoid saturating the signal. Hence, the serum must be diluted sufficiently so that the amount of available auto-antibody in the serum is lower than the binding capacity of the specific recombinant protein. 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 ten-fold dilution, the recombinant proteins were not saturated and yielded a signal/background ratio of > 5 . Higher or lower dilution resulted in partial saturation or decreased signal intensity. A tenfold dilution factor was therefore used in the current pre-confirmation experiment using recombinant proteins. The microarray data (background subtracted) was also adjusted by the average signal of control blocks on each slide and standardized for each recombinant protein.

In Fig. 5, we show the plot of the distribution of the reactivity for each of the recombinant proteins against the sera. The recombinant protein that produces the best result is Phosphoglycerate kinase 1 (PGK1), where 10 outliers out of 48 total samples are observed in the cancer group, while no outlier is present in the other three non-cancer groups. PGK1 protein is a kinase in the glycolytic pathway and can be up-regulated by HIF-1 α in the cellular response to hypoxia to provide energy for tumor cell proliferation [29]. Genomics-

based studies have found that it acts as a suppressor of proangiogenic factor such as VEGF and triggers metastasis due to its effect on the increased expression level of β -catenin, chemokine CXCR4 and CXCL12 [30–32]. At the protein level, PGK-1 has been found over-expressed in pancreatic cancer tissue versus adjacent controls and also elevated significantly in the sera of pancreatic cancer patients (19% strongly up-regulated, 50% weakly-moderately up-regulated) [33]. The performance of PGK-1 in the confirmation experiment using recombinant protein indicates that a lower percentage of patients elicit auto-response. In a future study, it would be interesting to see whether there is a correlation between serum level of PGK-1 and auto-antibody level and how the production of antibody affects the development of the cancer.

For both Malate dehydrogenase (MDH1) and ADP-ribosylation factor interacting protein 2 (ARFIP2), there are 4 such outliers in the cancer group. The absence of outliers in the non-cancer group indicates that these 3 recombinant proteins are exclusively antigenic in cancer sera and could be tumor-associated. In Table 3, the performance of these 3 biomarkers used together to distinguish cancer is estimated. A cutoff equal to the highest signal in a certain non-cancer group is applied to define the reactive samples in the cancer group. The 3 recombinant proteins together distinguish more than 40% of the cancer samples from the normal and diabetes group, while only 29.2% from the pancreatitis group.

For Annexin A2 (ANXA2), the cancer group only has one outlier that is above all the other groups. This is not consistent with the OS analysis which showed differential humoral response in the fraction where ANXA2 was identified in pancreatic cancer sera. It could be due to the use of the recombinant proteins, which may lack the required PTMs to induce a humoral response or the protein may not be in a form to induce a humoral response [17]. Also, since multiple proteins are identified in the fraction, the protein that was responsible for the observed humoral response may not be ANXA2. Heterogeneous nuclear ribonucleoprotein A2 (HNRPA2) produced a more unexpected result in the confirmation experiment where it showed a universal increase in the reactivity against diabetes samples, while the signals of the other three groups remained at the same level.

14. Conclusion

We have presented a study of the cancer-related humoral response on pancreatic adenocarcinoma using 2-dimensional separation and protein microarray techniques. After analyzing the data with two statistical tools, the fractions showing outlier patterns in Outlier Sum Test were chosen for identification of the proteins and confirmation using recombinant proteins. In the confirmation experiment, 20.8% of the cancer samples demonstrated strongly elevated reactivity for PGK-1, while no proteins in the non-cancer groups were found to react. This result suggests that the auto-antibody level of PGK-1 in the serum is useful as a diagnostic biomarker indicating the presence of cancer. Future study of the correlation between the protein level and auto-antibody level of PGK-1 in cancer patients may provide a better understanding of the role of PGK-1 in cancer development.

Acknowledgments

This work was supported in part by the National Cancer Institute under grant R01CA106402 (D.M.L.), and the National Institutes of Health under grant R01GM49500 (D.M.L.). Support was also generously provided by Eprogen, Inc.

References

1. Jemal A, Siegel R, Ward E, Murray T, Xu JQ, Smigal C, Thun MJ. Cancer Statistics, 2006. *CA Cancer J Clin.* 2006; 56:106–130. [PubMed: 16514137]

2. Rosty C, Goggins M. Early detection of pancreatic carcinoma. *Hematol Oncol Clin North Am.* 2002; 16:37–52. [PubMed: 12063828]
3. Garcea G, Neal CP, Pattenden CJ, Steward WP, Berry DP. Molecular prognostic markers in pancreatic cancer: A systematic review. *Eur J Cancer.* 2005; 41:2213–2236. [PubMed: 16146690]
4. Tan EM. Autoantibodies as reporters identifying aberrant cellular mechanisms in tumorigenesis. *J Clin Invest.* 2001; 108:1411–1415. [PubMed: 11714730]
5. Casiano CA, Mediavilla-Varela M, Tan EM. Tumor-associated arrays for the serological diagnosis of cancer, *Mol & Cel. Proteomics.* 2006; 5:1745–1759.
6. Desmetz C, Maudelonde T, Mange A, Solassol J. Identifying autoantibody signatures in cancer: a promising challenge. *Expert Rev Proteomics.* 2009; 6:377–386. [PubMed: 19681673]
7. Hall JC, Casciola-Rosen L, Rosen A. Altered structure of autoantigens during apoptosis. *Rheum Dis Clin North Am.* 2004; 30:455–471. [PubMed: 15261336]
8. Pollard KM, Lee DK, Casiano CA, Bluthner M, Johnston MM, Tan EM. The autoimmunity-inducing xeno-biotic mercury interacts with the autoantigen fibrillarin and modifies its molecular and antigenic properties. *J Immunol.* 1997; 158:3521–3528. [PubMed: 9120314]
9. Utz PJ, Anderson P. Posttranslational protein modifications, apoptosis, and the bypass of tolerance to autoantigens. *Arthritis Rheum.* 1998; 41:1152–1160. [PubMed: 9663470]
10. Rosen A, Casciola-Rosen C, Wigley F. Role of metal-catalyzed oxidation reactions in the early pathogenesis of scleroderma. *Curr Opin Rheumatol.* 1997; 9:538–543. [PubMed: 9375283]
11. Ben-Mahrez K, Thierry D, Sorokine I, Danna-Muller A, Kohiyama M. Detection of circulating antibodies against c-myc protein in cancer patient sera. *Br J Cancer.* 1998; 57:529–534. [PubMed: 3044429]
12. Disis ML, Calenoff E, McLaughlin G, Murphy AE, Chen W, Groner B, Jeschke ML, Lydon N, McGlynn E, Livingston RB, Moe R, Cheever MA. Existent T-cell and antibody immunity to HER-2/neu protein in patients with breast cancer. *Cancer Res.* 1994; 54:16–20. [PubMed: 7505195]
13. Disis ML, Cheever MA. Oncogenic proteins as tumor antigens. *Curr Opin Immunol.* 1996; 8:637–642. [PubMed: 8902388]
14. Soussi T. p53 Antibodies in the sera of patients with various types of cancer: a review. *Cancer Res.* 2000; 60:1777–1788. [PubMed: 10766157]
15. Brichory FM, Misek DE, Yim AM, Krause MC, Giordano TJ, Beer DG, Hanash SM. An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc Natl Acad Sci USA.* 2001; 98:9834–9829.
16. Gnjatich S, Wheeler C, Ebner M, Ritter E, Murray A, Altorki NK, Ferrara CA, Hepburne-Scott H, Joyce S, Koopman J, McAndrew MB, Workman N, Ritter G, Fallon R, Old LJ. Seromic analysis of antibody responses in non-small cell lung cancer patients and healthy donors using conformational protein arrays. *Clin Canc Res.* 2009; 15:4733–4741.
17. Desmetz C, Bascoul-Mollevis C, Rochaix P, Lamy PJ, Kramar A, Rouanet P, Maudelonde T, Mange A, Solassol J. Identification of a New Panel of Serum Autoantibodies Associated with the Presence of In situ Carcinoma of the Breast in Younger Women. *J Immunol Methods.* 2009; 341:50–58. [PubMed: 19041653]
18. Looi KS, Nakayasu ES, de Diaz R, Tan EM, Almeida IC, Zhang JY. Using proteomic approach to identify tumor-associated antigens as markers in hepatocellular carcinoma. *J Proteome Res.* 2008; 7:4004–4012. [PubMed: 18672925]
19. Xia Q, Kong XT, Zhang GA, Hou XJ, Qiang H, Zhong RQ. Proteomics-based identification of DEAD-box protein 48 as a novel autoantigen, a prospective serum marker for pancreatic cancer. *Biochem Biophys Res Commun.* 2005; 330:526–532. [PubMed: 15796914]
20. Hamanaka Y, Suehiro Y, Fukui M, Shikichi K, Imai K, Hinoda Y. Circulating anti-MUC1 IgG antibodies as a favorable prognostic factor for pancreatic cancer. *Int J Cancer.* 2003; 103:97–100. [PubMed: 12455059]
21. Kotera Y, Fontenot JD, Pecher G, Metzgar RS, Finn OJ. Humoral immunity against a tandem repeat epitope of human mucin MUC-1 in sera from breast, pancreatic, and colon-cancer patients. *Cancer Res.* 1994; 54:2856–2860. [PubMed: 7514493]

22. Raedle J, Oremek G, Welker M, Roth WK, Caspary WF, Zeuze Fm S. p53 autoantibodies in patients with pancreatitis and pancreatic carcinoma. *Pancreas*. 1996; 13:241–246. [PubMed: 8884844]
23. Yan F, Sreekumar A, Laxman B, Chinnaiyan AM, Lubman DM, Barder TJ. Protein microarrays using liquid phase fractionation of cell lysates. *Proteomics*. 2003; 3:1228–1235. [PubMed: 12872223]
24. Patwa TH, Li C, Poisson LM, Kim HY, Pal M, Ghosh D, Simeone DM, Lubman DM. The identification of phosphoglycerate kinase-1 and histone H4 autoantibodies in pancreatic cancer patient serum using a natural protein microarray. *Electrophoresis*. 2009; 12:2215–2226.
25. Canelle L, Bousquet J, Pionneau C, Deneux L, Imam-Sghiouar N, Caron M, Joubert-Caron R. An efficient proteomics-based approach for the screening of autoantibodies. *J Immunol Methods*. 2005; 229:77–89. [PubMed: 15914192]
26. Tibshirani R, Hastie T. Outlier sums for differential gene expression analysis. *Biostatistics*. 2007; 8:2–8. [PubMed: 16702229]
27. Lieber M, Mazzetta J, Nelson-Rees W, Kaplan M, Todaro G. Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. *Int J Cancer*. 1975; 15:741–747. [PubMed: 1140870]
28. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao CH, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005; 310:644–648. [PubMed: 16254181]
29. Dayan F, Roux D, Brahimi-Horn MC, Pouyssegur J, Mazure NM. The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1alpha. *Cancer Res*. 2006; 66:3688–3698. [PubMed: 16585195]
30. Wang J, Dai J, Jung Y, Wei CL, Wang Y, Havens AM, Hogg PJ, Keller ET, Pienta KJ, Nor JE, Wang CY, Taichman RS. A glycolytic mechanism regulating an angiogenic switch in prostate cancer. *Cancer Res*. 2007; 67:149–159. [PubMed: 17210694]
31. Kurayoshi M, Oue N, Yamamoto H, Kishida M, Inoue A, Asahara T, Yasui W, Kikuchi A. Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion. *Cancer Res*. 2006; 66:10439–10448. [PubMed: 17079465]
32. Zieker D, Konigsrainer I, Traub F, Nieselt K, Knapp B, Schillinger C, Stirnkorb C, Fend F, Northoff H, Kupka S, Brucher BL, Konigsrainer A. PGK1 a potential marker for peritoneal dissemination in gastric cancer. *Cell Physiol. Biochem*. 2008; 21:429–436.
33. Shichijo S, Azuma K, Komatsu N, Ito M, Maeda Y, Ishihara Y, Itoh K. Two proliferation-related proteins, TYMS and PGK1, could be new cytotoxic T lymphocyte-directed tumor-associated antigens of HLA-A2 (+) colon cancer. *Clin Cancer Res*. 2004; 10:5828–5836. [PubMed: 15355913]

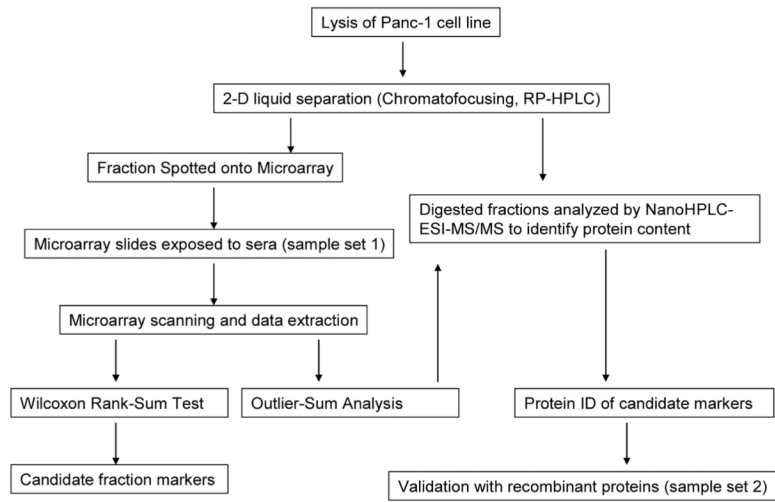


Fig. 1.
Flowchart of the experiment.

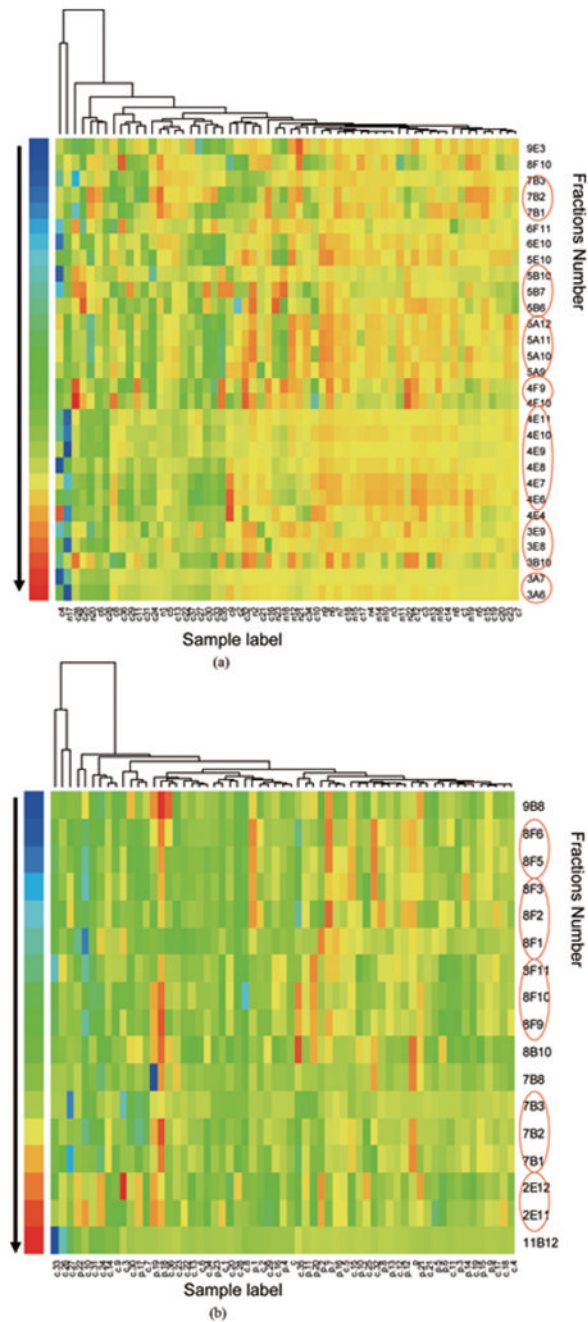


Fig. 2. Heatmap with dendrogram of the microarray data. The colors of the bands indicate the normalized intensities of the microarray signal of the fractions. Only the fractions with a p-value less than 0.02 in the pairwise comparisons using Wilcoxon Rank-Sum Tests are shown in the heatmap. The fractions appear with their adjacent ones are red-circled. a. cancer vs. normal; b. cancer vs. pancreatitis.

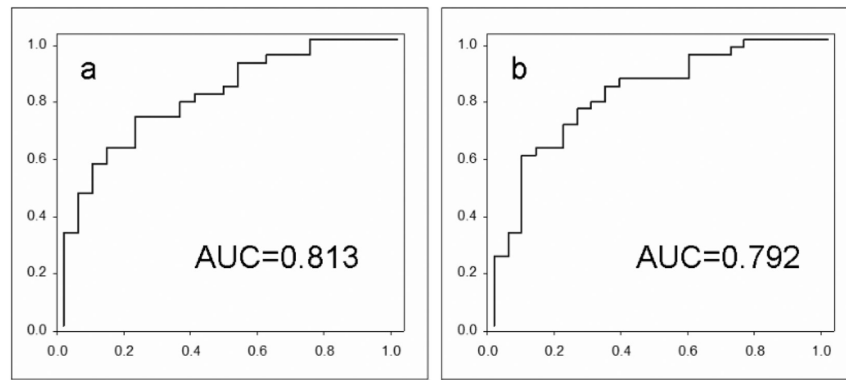


Fig. 3. Combined ROC curves for the top-ranked fractions with lowest correlations in the Wilcoxon Rank-Sum Tests. a) combined, cancer vs. normal; b) cancer vs. pancreatitis.

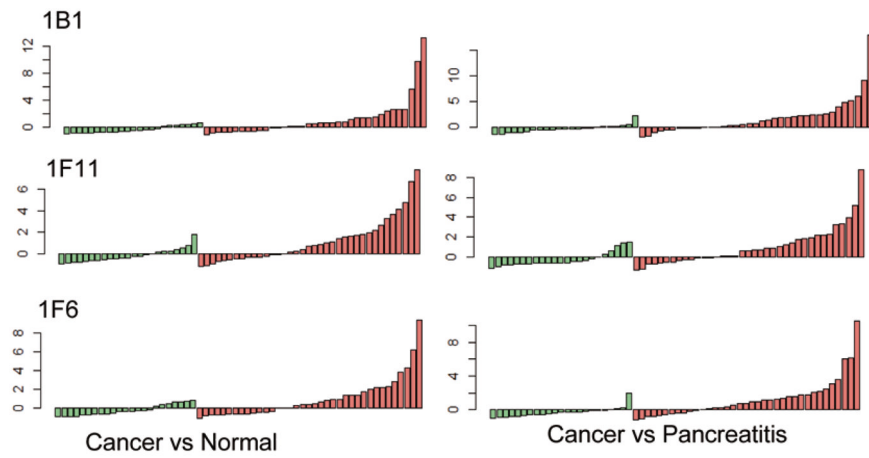


Fig. 4. Colored bar graphs of three fractions found responded exclusively to some cancer sera in both pairwise comparisons between cancer vs. normal and cancer vs. pancreatitis. The y-axis is the normalized microarray signal for each sample.

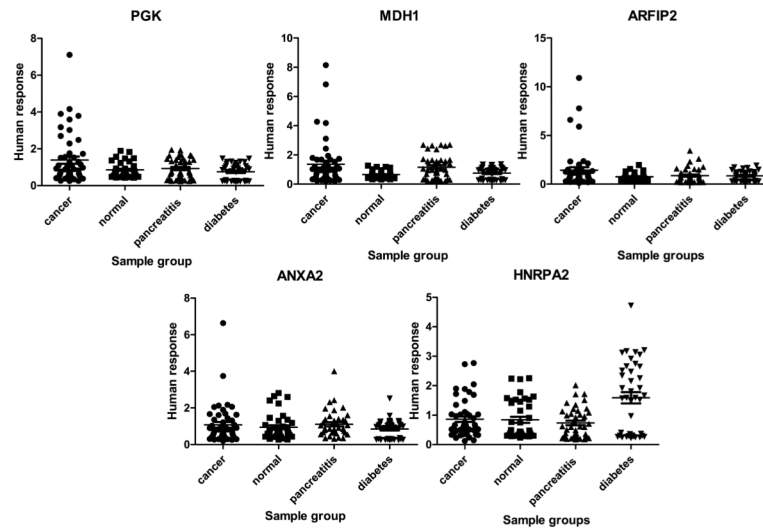


Fig. 5. Distribution of the level of reactivity of five biomarker candidates examined in the confirmation experiment. The plots present the normalized fluorescent intensities of the auto-antibody capture by the five recombinant proteins.

Table 1

Lists of differentiated fractions picked by OS analysis in both pairwise comparison between cancer vs. normal and cancer vs. pancreatitis. The information about the identified proteins in these fractions is also included

Fraction	Access number	Protein name	Fraction pH	MW	Seq Cov%	Theoretical pI	Unique peptides
1B1	P62847	40S ribosomal protein S24	7.9-7.6	15414	28.59	10.79	3
1F11	P00558	Phosphoglycerate kinase 1	7.9-7.6	44587	18.21	8.30	5
1F6	Q15369	Transcription elongation factor B polypeptide 1	7.9-7.6	12466	17.74	4.74	2
3E5	P04406	Glyceraldehyde-3-phosphate dehydrogenase	7.0-6.7	35900	18.97	8.58	4
4D6	Q9Y6N5	Sulfide:quinone oxidoreductase, mitochondrial precursor	6.7-6.4	49929	11.62	9.18	4
5G2	Q06830	Peroxioredoxin 1 (Thioredoxin peroxidase 2)	6.1-5.8	22097	29.35	8.27	6
8C3	O95881	Thioredoxin domain containing protein 12 precursor	4.9-4.6	19194	37.98	5.25	5
9A9	Q99729	Heterogeneous nuclear ribonucleoprotein A/B	4.6-4.3	36590	9.15	9.04	3
11D5	Q8NC51	Plasminogen activator inhibitor 1	IPA	44539	18.43	8.66	5
		RNA-binding protein	wash				

Table 2

Demographic and clinical characteristics of the samples used in the experiment

Characteristics	Cancer	Disease Groups		
		Pancreatitis	Normal	Diabetes
Age	66.6	55.8	52.8	65.2
Gender (Male)	56.2%	62.5%	65.0%	35.1%
Clinical	Pancreatic Adenocarcinoma Stage I/II 20.8% Stage III/IV 79.2%	Chronic Pancreatitis (No acute symptoms)	Healthy	Type II Diabetes

Table 3

Numbers of samples with reactivity above the cutoff (cutoff = highest signal in the non-cancer group) against recombinant proteins in the cancer group compared to 3 non-cancer groups. The numbers in the parentheses are the percentage of the positive reactors in the cancer category i.e. sensitivity at 100% specificity

Recombinant proteins	Cancer vs Normal	Pair of sample groups	Cancer vs. Pancreatitis	Cancer vs. Diabetes
PGK1	10 (20.8)	10 (20.8)		12 (25)
PGK1 or MPH1	18 (37.5)	12 (25)		19 (39.6)
PGK1 or MPH1 or ARFIP2	22 (45.8)	14 (29.2)		21 (43.8)