Published in final edited form as:

J Proteome Res. 2014 February 7; 13(2): 395–407. doi:10.1021/pr400422g.

Discrimination between adenocarcinoma and normal pancreatic ductal fluid by proteomic and glycomic analysis

Mindy Porterfield[†], Peng Zhao[†], Haiyong Han, John Cunningham, Kazuhiro Aoki, Daniel D. Von Hoff, Michael J. Demeure, J. Michael Pierce, Michael Tiemeyer^{*}, and Lance Wells^{*}

¹Complex Carbohydrate Research Center, University of Georgia, Athens, GA

²Chemistry Department, University of Georgia, Athens, GA

³Translational Genomics Research Institute, Phoenix, AZ

⁴Section of Gastroenterology, College of Medicine, University of Arizona, Tucson, AZ

⁵Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA

Abstract

Sensitive and specific biomarkers for pancreatic cancer are currently unavailable. The high mortality associated with adenocarcinoma of the pancreatic epithelium justifies the broadest possible search for new biomarkers that can facilitate early detection or monitor treatment efficacy. Protein glycosylation is altered in many cancers, leading many to propose that glycoproteomic changes may provide suitable biomarkers. In order to assess this possibility for pancreatic cancer, we have performed an in-depth LC-MS/MS analysis of the proteome and MSⁿ-based characterization of the N-linked glycome of a small set of pancreatic ductal fluid obtained from normal, pancreatitis, intraductal papillary mucinous neoplasm (IPMN), and pancreatic adenocarcinoma patients. Our results identify a set of seven proteins that were consistently increased in cancer ductal fluid compared to normal (AMYP, PRSS1, GP2-1, CCDC132, REG1A, REG1B, and REG3A) and one protein that was consistently decreased (LIPR2). These proteins are all directly or indirectly associated with the secretory pathway in normal pancreatic cells. Validation of these changes in abundance by Western blotting revealed increased REG protein glycoform diversity in cancer. Characterization of the total N-linked glycome of normal, IPMN, and adenocarcinoma ductal fluid clustered samples into three discrete groups based on the prevalence of 6 dominant glycans. Within each group, the profiles of less prevalent glycans were able to distinguish normal from cancer on this small set of samples. Our results emphasize that individual variation in protein glycosylation must be considered when assessing the value of a glycoproteomic marker, but also indicate that glycosylation diversity across human subjects can be reduced to simpler clusters of individuals whose N-linked glycans share structural features.

Keywords

Pancreatic cancer; Proteomics; Biomarker; N-linked glycan; Glycomics

^{*}Co-corresponding authors: Michael Tiemeyer, mtiemeyer@ccrc.uga.edu and Lance Wells, lwells@ccrc.uga.edu.

[†]The first two authors contributed equally to this work.

SUPPORTING INFORMATION

Supplementary material includes 6 tables and 4 figures. One table presents patient information, three tables present quantitative proteomic data, and two tables present quantitative glycomic profiles. One supplementary figure contains all of the compiled MS/MS spectra for single-peptide protein identifications, one supplementary figure presents the hierarchical clustering for total glycan profiles and for blood group antigens, neither of which distinguish disease diagnosis, and the other two supplementary figures present deconvoluted full MS spectra for each of the analyzed glycan profiles.

INTRODUCTION

In 2010, the American Cancer Society estimated 41,000 diagnoses of pancreatic cancer in the U.S.¹. With a very low percentage of five-year survival, early stage biomarkers for this disease are urgently needed, although there are markers that are used to monitor the course of disease; e.g., the glycan-specific serum marker, CA19-9^{2, 3}. Proteomic analyses of serum samples from patients with pancreatic ductal adenocarcinoma have yielded important information for developing potential biomarkers⁴. Recent data has demonstrated that pancreatic cancer cells are not always of ductal epithelial origin, but may in fact more frequently arise from acinar cells⁵, the primary secretory cell responsible for producing the proteins of pancreatic ductal fluid. Therefore, pancreatic ductal fluid, which is likely to contain proteins released from pancreatic adenocarcinoma, has also been subjected to proteomic analysis in search of markers that could also be present in sera⁶⁻¹². These previous studies have produced proteomes that overlap with each other and with the results reported here. However, full validation of a single proteomic marker or set of proteomic markers has not been achieved for pancreatic cancer.

In addition to altered protein expression during oncogenesis, many studies have clearly documented that the glycans expressed on glycoproteins secreted or released from various types of cancer cells exhibit changes in structure that are cell-type specific. For example, glycoproteins that express N-glycans with a “core fucose” residue (alpha1,6 fucose) are secreted into serum from hepatocellular carcinoma (HCC) but not from cirrhotic hepatocytes. An assay for core-fucosylated alpha-fetoprotein is in use to test for HCC, and there is evidence that including other core-fucosylated glycoproteins, such as GP73, in the analysis yields an HCC diagnostic test with higher specificity and sensitivity¹³⁻¹⁶. Aberrant glycosylation in pancreatic carcinoma is apparent by increased serum levels of CA19-9 and by the detection of circulating antibodies directed against the mucin MUC1 that expresses truncated O-linked glycans (Tn antigens)¹⁷.

In order to identify additional serum markers for pancreatic carcinoma, particularly those for early detection, our approach has first focused on applying proteomic and glycomic analytical technologies for in-depth analysis of pancreatic ductal fluid. Markers identified in ductal fluid are then candidates for validation as serum markers using antibodies that recognize the glycan and protein differences that are identified between ductal fluid samples from patients with pancreatic cancer and from controls, including pancreatitis and intraductal papillary mucinous neoplasms (IPMN). Here, we report in-depth analysis of the proteome and the N-linked glycome of a training set of ductal fluid samples. The results provide potential targets for full validation and highlight important considerations for analyzing human glycoproteomes.

EXPERIMENTAL METHODS

Pancreatic Ductal Fluid Samples

Pancreatic ductal fluid samples along with matching serum and plasma samples were collected from patients who underwent endoscopic retrograde cholangiopancreatogram (ERCP) or endoscopic ultrasound (EUS) procedures. The ductal fluid samples were snap frozen in liquid nitrogen following aspiration from the patients. Sample collection protocols were reviewed and approved by the Institutional Review Boards (IRBs) at the University of Arizona (Tucson, AZ) and the Translational Genomics Research Institute (Phoenix, AZ), and written informed consent was obtained from all patients. Pancreatic ductal fluid samples from patients with the following four diagnoses were used in this study: pancreatic cancer, intraductal papillary mucinous neoplasm (IPMN), pancreatitis, and normal pancreas (Sphincter of Oddi Dysfunction, SOD Type II or III). The age ranges and gender

distributions for each diagnostic class were as follows: pancreatic cancer, 54-79 years, 5 male, 4 female; IPMN, 72-77 years, 2 male, 1 female; pancreatitis, 41-56 years, 1 male, 2 female; normal pancreas, 31-72 years, 2 male, 7 female (Supplementary Table S1). For protein identification and quantification, 12 samples in total were analyzed (three samples for each diagnosis). Clear (no visible blood or bile contamination) pancreatic ductal fluid samples were thawed on ice and filtered by 0.2 μ m spin columns (Nanosep). Protein concentration of all the samples was determined using the micro BCA protein assay kit (Pierce) following the manufacturer's instructions. Equal amounts of protein (1 mg) were used for analysis. For glycomic analysis, 3 normal, 4 pancreatic cancer, and 2 IPMN samples were chosen. As described for proteomic analysis, ductal fluid samples that were clear (no visible blood or bile contamination) were selected for glycomic analysis. For each sample, 300-500 μ l of ductal fluid were extracted with organic solvent as previously described¹⁸. Briefly, for 300 μ l of ductal fluid, total volume was adjusted to 4:8:3 (chloroform:methanol:water) by the addition of 0.4 ml water, 1.87 ml methanol, and 0.93 ml chloroform. The adjusted sample was extracted overnight by nutation. The next morning, proteins were harvested by centrifugation. The protein pellet was washed three times with cold 80% acetone in water and the final pellet was dried under a stream of nitrogen. The resulting solids were harvested as a uniform, white protein powder, which was stored desiccated at -20 °C. The amount of material harvested from most samples was insufficient to allow both glycomic and proteomic analysis to be performed on all samples (Supplementary Table S1).

Trypsin Digestion for Proteomic Analysis

The fluid samples were reduced with 10 mM DTT for 1 h at 56 °C, alkylated (carboxyamidomethylated) with 55 mM iodoacetamide (Sigma) in dark for 45 min, and digested with trypsin (Promega) in 40 mM NH_4HCO_3 overnight at 37 °C. The digestion was quenched with 1% trifluoroacetic acid (TFA), and the resulting peptides were desalted with C18 spin columns (Vydac Silica C18, The Nest Group, Inc.) and dried in a Speed Vac.

Protein Fractionation

Protein fractionation was performed by reverse phase liquid chromatography (RP-LC) using the Agilent 1100 series HPLC system (Agilent Technologies). Solvent A (0.1% TFA) and solvent B (0.085% TFA/80% acetonitrile) were used to develop a linear gradient starting with 5 minutes at 5% solvent B (95% solvent A), followed by a 60-minute gradient at variable slope from 5% to 95% solvent B and staying for 3 minutes at 95% solvent B, then returning to 5% solvent B (95% solvent A) in 1.5 minutes and staying for 4.5 minutes at 5% solvent B (95% solvent A). Dried peptides were dissolved in solvent A and separated on a 2.1 $\times$ 250 mm silica-based C18 column (VYDAC) at a flow rate of 100 μ l/min over the linear gradient. Eluted peptides were collected every 4 min, and subsequently combined into 5 fractions (F1, 15-32%; F2, 32-40%; F3, 40-45%; F4, 45-55%; and F5, 55-85%), desalted and dried as described above.

Reverse Phase nanoLC-MS/MS Analysis

Dried peptides from each fraction generated by RP-LC (12 $\times$ 5 in total) were resuspended in 0.5 μ l of solvent B (0.1% formic acid/80% acetonitrile) and 19.5 μ l of solvent A (0.1% formic acid) and loaded onto a 75 μ m $\times$ 105 mm C18 reverse phase column (packed in house, YMC GEL ODS-AQ120ÅS-5, Waters) by nitrogen bomb. Peptides were eluted directly into the nanospray source of an LTQ Orbitrap XLTM (Thermo Fisher Scientific) with a 140-min linear gradient consisting of 5-100% solvent B over 90-95 min at a flow rate of ~250 nl/min. In order to optimize the separation of peptides eluted into the mass spectrometer, gradients were expanded over a 70-min period in the appropriate region corresponding to each fraction collected from the previous offline RP-LC separation (F1,

4-30%; F2, 9-35%; F3, 15-42%; F4, 20-55%; and F5, 28-85%). The spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to 200 °C. Full scan MS spectra were acquired from m/z 300 to 2000 with a resolution of 60000 at m/z 400 after accumulation of 1000000 ions (mass accuracy < 5 ppm). MS/MS events were triggered by the 6 most intense ions from the preview of full scan and a dynamic exclusion window was applied which prevents the same m/z value from being selected for 6 seconds after its acquisition. All 5 sub-fractions were analyzed in technical triplicates and data were acquired using Xcalibur® (ver. 2.0.7, Thermo Fisher Scientific). Spectra will be made available upon request.

Proteomic Data Analysis

The acquired MS/MS spectra were searched against the UniProt human proteome database (58831 entries, updated at May 10, 2009) using SEQUEST (Bioworks 3.3, Thermo Fisher Scientific) with the following settings: 50-ppm and 0.5-Da deviation were set for monoisotopic precursor and fragment masses, respectively; trypsin was specified as the enzyme; only fully tryptic peptide identifications were retained; a maximum of 3 missed cleavage sites, 3 differential amino acids per modification and 3 differential modifications per peptide were allowed; oxidized methionine (+15.9949 Da) and carbamidomethylated cysteine (+57.0215 Da) were set as differential modifications. All of the raw spectra were searched against both normal and reversed database under the same parameters, and all of the output files from SEQUEST search were filtered and grouped by different biological samples and replicates in ProteoIQ™. The cutoff value of peptides was set to an Xcorr of 0.5 and the minimum peptide length was set to 4 amino acids. For protein identification, false discovery rate was set to 1% at protein level and peptides matched to multiple proteins were excluded; for protein quantification, the 1% protein-level false discovery rate data was further filtered to achieve a 10% peptide-level false discovery rate, and only proteins that are identified by more than one peptide and in more than one biological sample were considered. The validated result was submitted to Gene Ontology (www.geneontology.org)¹⁹ for protein subcellular localization and biological function annotation.

In order to compare the protein expression levels across samples with different diagnoses, normalized spectral abundance factors (NSAF) were calculated for each protein that was commonly observed in all four diagnoses. In this approach, the spectral counts (SpC) of each protein in a given dataset were divided by its length (L) and normalized to the sum of SpC/L values in the given dataset^{20, 21}:

$$NSAF_x = \frac{\left(\frac{SpC}{L}\right)_x}{\sum_{i=1}^N \left(\frac{SpC}{L}\right)_x}$$

To further resolve shared peptides between protein isoforms, a distribution factor was introduced into the calculation of NSAF²²:

$$dNSAF = \frac{uSpC + [(d)(sSpC)]}{uL + sL}$$

$$d = \frac{uSpC}{\sum uSpC}$$

According to the equations above, dNSAF is calculated where spectral counts from shared peptides are distributed among protein isoforms based on a distribution factor, *d*. Spectral counts from peptides uniquely mapping to a protein are denoted as “uSpC”, while spectral counts from peptides shared between isoforms are labeled “sSpC”. Protein amino acid lengths mapped to unique and shared peptides are denoted as “uL” and “sL”, respectively.

Immunoblotting

Protein concentrations of normal and cancer pancreatic juice samples were determined by micro BCA protein assay. Equal amounts of protein from normal and cancer samples (ranging from 2-8 µg for different antibodies) were separated by 4-20% Tris-HCl precast minigels (Bio-Rad), and semi-dry transferred to Immobilon-P transfer membrane (Millipore). The membranes were blocked with 5% BSA in TBST (TBS with 0.1% Tween 20), and probed with each antibody at 4 °C overnight as follows: 1:1000 dilution for REG1A (Abcam), REG1B (Abcam), and REG3A (Abnova) blots, and 1:2000 dilution for phospholipase A2 (Abcam) and pancreatic lipase-related protein 2 (Abnova) blots. After the addition of secondary antibodies conjugated to horseradish peroxidase (HRP) at room temperature for 1 h, the final detection of HRP activity was performed using SuperSignal West Pico chemiluminescent substrates (Thermo Fisher Scientific). The films were exposed to CL-XPosure film (Thermo Fisher Scientific). The amount of material harvested from any individual sample was insufficient to allow glycomic, proteomic, and western blotting (with all antibodies) to be performed on each sample. In some cases, proteomic and western blot analysis with a subset of antibodies were performed on the same sample (Supplementary Table S1).

N-linked Glycan Analysis

N-linked glycans were prepared from tryptic/chymotryptic digests of total ductal fluid proteins as described previously¹⁸. Briefly, protein powder produced by organic extraction of ductal fluid (described above) was resuspended in 200 µl of trypsin buffer (0.1 M Tris-HCl, pH 8.2 containing 1 mM CaCl₂) by sonication and boiling for 5 minutes. After cooling to room temperature, 25 µl of trypsin solution (2 mg/ml in trypsin buffer) and 25 µl of chymotrypsin solution (2mg/ml in trypsin buffer) were added. Digestion was allowed to proceed for 18 hours at 37 °C before the mixture was boiled for 5 minutes. Insoluble material was removed by centrifugation and the supernatant was removed and dried by vacuum centrifugation. The dried peptide mixture was resuspended in 250 µl of 5% (v/v) acetic acid and loaded onto a Sep-Pak C18 cartridge column. The cartridge was washed with 10 column volumes of 5% acetic acid. Glycopeptides were then eluted step-wise, first with 3 volumes of 20% isopropanol in 5% acetic acid and then with 3 volumes of 40% isopropanol in 5% acetic acid. The 20% and 40% isopropanol steps were pooled and evaporated to dryness. Dried glycopeptides were resuspended in 50 µl of 20 mM sodium phosphate buffer, pH 7.5, for digestion with PNGaseF (Prozyme, San Leandro, CA). Following PNGaseF digestion for 18 hr at 37 °C, released oligosaccharides were separated from peptide and enzyme by passage through a Sep-Pak C18 cartridge. The digestion mixture was adjusted to 5% acetic acid and loaded onto the Sep-Pak. The column run-through and an additional wash with 3 column volumes of 5% acetic acid, containing released oligosaccharides, were collected together and evaporated to dryness.

Following enzymatic release and clean-up, liberated N-linked glycans were permethylated²³ and analyzed by direct infusion, nanospray ionization, ion trap mass spectrometry (NSI-LTQ/Orbitrap, Thermo Fisher). An automated MS workflow was employed to sequentially capture MS/MS spectra for all detectable ions. In this workflow, full MS spectra were obtained in the Orbitrap and the highest intensity peak was then selected for fragmentation in the linear trap (collision energy was 35-55% based on instrument calibration). Following acquisition of each MS/MS spectra, the next most intense parent ion was selected for fragmentation. In order to limit the fragmentation of redundant isotopes, an m/z window extending from -1.2 mass units below to +2.1 mass units above the parent ion was excluded. The cycle was repeated until fragmentation profiles revealed only background noise, generally 200 rounds. The resulting MS and MS/MS files were processed using SimGlycan (Premier Biosoft, Palo Alto, CA) to provide initial structural assignments for all m/z values associated with glycans²⁴. SimGlycan assignments were subsequently validated by manually inspecting MS/MS spectra for the presence of signature fragments consistent with the proposed structure for all glycans that demonstrated signal intensity differences in cancer or IPMN greater than 2-fold above or below normal samples (61 assignments). SimGlycan assignments were also validated manually if the assigned structure was not considered to be a likely component of the human ductal fluid glycoproteome (Xylose-containing glycans, inappropriately degraded structures, biosynthetic impossibilities). Such artifactual assignments arise because these structures are contained within the database used by SimGlycan. When MS/MS spectra for such candidate glycans were manually inspected, they uniformly revealed a lack of glycan-based fragmentation and their intensities were excluded from the total profile. Signal intensities for valid glycan assignments were retrieved from full MS spectra as peak areas obtained using the Orbitrap FT. Signals associated with different charge states of the same glycans were combined. The prevalence of each glycan was calculated by normalizing its signal intensity to the total signal intensity for all detected glycans and is expressed as “% Total Profile” for each glycan. The associations of glycans with clinical status were queried by hierarchical clustering methods using Euclidean distance calculations as previously described^{25, 26}.

Structural assignments for the glycans detected at the reported m/z values were based on the compositions determined by accurate mass of the intact molecule (detected by Orbitrap FT), the presence of diagnostic MS/MS fragments that report specific N-glycan features, and the limitations imposed on structural diversity by known glycan biosynthetic pathways. Key structural features that were used to assign glycan topologies included the detection of B-ion fragments and their Y-ion neutral loss counterparts corresponding to terminal LacNAc (Hex-HexNAc, assumed to be Gal-GlcNAc; fragment at m/z = 486.2, [m+Na]⁺), internal LacNAc (Hex-HexNAc, fragment at m/z = 472.2, [m+Na]⁺) sialic acid (fragment at m/z = 398.2, [m+Na]⁺), outerarm Fuc (as fucosylated LacNAc; deoxyHex-Hex-HexNAc and/or Hex-(deoxyHex)-HexNAc; fragment at m/z = 660.3, [m+Na]⁺), terminal Fuc (deoxyHex-Hex; fragment at m/z = 415.2, [m+Na]⁺); core Fuc (as Fuc-HexNAc at the reducing terminal; fragment at m/z = 474.2, [m+Na]⁺). It is frequently not possible to unambiguously assign non-reducing terminal modifications to a specific arm of a complex N-linked glycan solely using MS/MS spectra. For consistency of presentation and ease of comparison, outer arm modifications are presented as elaborations on the increasingly complex products of the known branching N-acetylglucosamine transferases (GlcNAcT) in the following succession: GlcNAcT1, T2, T4, T5. For example, a monosialylated, fully galactosylated triantennary glycan is depicted with a single sialic acid on the arm initiated by GlcNAcT1 (the 3-arm) and the three antennae would be represented as products of GlcNAcT1, 2, and 4 (two GlcNAc residues on the 3-arm and 1 on the 6-arm). The disialylated form of the same triantennary glycan would be depicted with the second sialic acid added to the arm initiated by GlcNAcT2 (the 6-arm). Structural ambiguity is also annotated by brackets, which are meant to indicate equally likely sites for elongation.

RESULTS

Protein Identification

For each of the 4 diagnoses, 3 patient samples were analyzed in technical triplicates. Each sample was trypsin digested and separated by off-line RP-HPLC separation. Five fractions were collected for each sample and analyzed by LC-MS/MS to yield a total of 45 LC-MS/MS experiments for each diagnosis (180 total LC-MS/MS analyses). After filtering and removing duplicates, the combined data set consists of 368 unique proteins identified by 1995 peptides corresponding to 58930 spectra, 74% (273/368) of which were identified by more than one peptide. Specifically, 112 proteins were identified by 750 peptides with 11598 spectra in normal samples; 138 proteins were identified by 743 peptides with 6590 spectra in pancreatitis samples; 124 proteins were identified by 808 peptides with 22581 spectra in IPMN samples; and 188 proteins were identified by 1068 peptides with 18161 spectra in pancreatic cancer samples (Table 1, Supplementary Table S2 and S3, Supplementary Figure 1). All the identified proteins were submitted to Gene Ontology (www.geneontology.org) for subcellular localization and biological function annotation. Based on the spectral counts assigned to each identified protein, the majority of the proteins are secreted proteins (81%) involved in proteolysis (52%) and metabolic process (29%) (Fig. 1).

Protein Quantification

To evaluate the variation in protein expression across pancreatic ductal fluid samples with different diagnoses, the identified protein dataset was further filtered to achieve a 10% peptide-level false discovery rate at 1% protein-level false discovery rate. After filtering, the resulting dataset was examined manually to eliminate proteins that were only identified by one peptide or in only one patient. In the final quantified dataset, a total of 47 proteins were quantified with 590 peptides and 46172 spectra across three diagnoses and normal controls (Fig. 2). Specifically, 22 proteins were quantified with 300 peptides and 8674 spectra in normal samples; 19 proteins were quantified with 215 and 3774 spectra in pancreatitis samples; 35 proteins were quantified with 414 peptides and 18632 spectra in IPMN samples; and 36 proteins were quantified with 422 peptides and 15092 spectra in pancreatic cancer samples.

By comparing the dNSAF values of proteins that were commonly observed in the samples from normal control and the three diagnoses, we were able to discover the differential expression of 22 proteins in our dataset (Fig. 3, Table 2). In comparison to normal controls, several proteins, such as REG1A, alpha-amylase, trypsin-1, chymotrypsinogen B, and glycoprotein GP2-1, showed significant elevation in IPMN and cancer samples. Several other proteins, such as pancreatic amylase, elastase 2A, 3B and 3A, carboxypeptidase A1, and pancreatic lipase-related protein 2, were downregulated in IPMN and cancer samples compared to normal controls. We also found several proteins that were uniquely expressed in IPMN and/or cancer samples on the quantifiable level (Table 3), such as REG1B, REG3A, CCDC132, phospholipase A2, and elastase 2B. As we re-examined the uniquely expressed proteins on the identifiable level, we discovered that even though some of those proteins were unique in IPMN and/or cancer samples on quantifiable level, they may be observed universally in the other biological samples on identifiable level (Table 4). For example, REG1B was only seen in two cancer patients on the quantifiable level, however, it was identified in patients with all three diagnoses and normal controls, suggesting the protein is likely present in all samples but upregulated in cancer samples.

Biological variation was also investigated by calculating the standard deviation across the biological triplicates based on the normalized spectral counts of each quantified protein

(Figure 3, Supplementary Table S4). The pronounced biological variances represented by the data are likely contributed by the differences of individual patients, such as gender, age, blood type and other medical conditions. The statistical data also indicates the need to increase the number of biological samples, and possibly to further stratify the samples based on multiple biological and medical factors instead of solely on diagnosis.

Orthogonal Validation of Protein Identifications

Antibodies were obtained for a subset of candidate biomarkers in order to validate the proteomic results by Western blotting (Fig. 4). While normal samples demonstrated 2 major bands for REG1A, additional bands were observed in the cancer sample (Fig. 4A). A similar pattern was observed in REG1B and REG3A blots with more prominent increases in abundance and multiple bands present in cancer samples (Fig. 4B,C). The molecular weight heterogeneity of REG proteins is believed to result from glycoform heterogeneity and proteolytic processing²⁷. Distinctive bands of immunoreactive phospholipase A2 were observed at 32 kDa (full length) and 16 kDa (mature) in the cancer samples and were absent in the normal controls (Fig. 4D). Therefore, phospholipase A2 (PLA2) can be considered as a positive marker for pancreatic malignancy. In contrast to REG proteins and PLA2, immunoreactive bands of pancreatic lipase-related protein 2 (LIPR2) at 37 kDa (mature) and at 52 kDa (full length) were decreased in cancer (Fig. 4E). Therefore, REG proteins and PLA2 may be positive indicators for pancreatic cancer while LIPR2 may be considered a negative indicator.

Total N-linked Glycan Profile

A total of 80 glycans were analyzed by NSI-MS/MS (nanospray ionization-MS/MS) and MSⁿ as needed to elucidate the structural features of N-linked glycans harvested from 3 normal, 2 IPMN, and 4 pancreatic cancer ductal fluid samples. Comparisons of the prevalence of all N-linked glycans did not detect glycan markers or even glycan patterns that could distinguish cancer from normal (Supplementary Figure 1A). However, the total glycan profiles for the samples analyzed were dominated by a small set of glycans whose prevalences ranged from 6 – 38% of the total profile. These driver glycans overwhelmed the contribution of less prevalent glycans and did not sort with normal, cancer, or IPMN, nor were they correlated with patient gender or age (Supplementary Table S1). After removing the driver glycans from the total profile and recalculating the prevalence of the remaining glycans, differences in the profile of minor glycans became apparent. By Wilcoxon rank-sum test, 9 of the remaining glycans showed increased or decreased prevalence ($p < 0.05$) in cancer or IPMN compared to normal (Fig. 5). Several of the discriminating glycan structures carry blood group epitopes of the H, Lewis X/A or Lewis Y/B type. However, in their entirety, blood group epitopes or secretor status were not able to sort the samples by diagnosis, indicating that blood group by itself does not account for the observed segregation of cancer, normal, and IPMN (Supplementary Figure 1B).

A striking division of the 9 samples was detected by comparing the prevalences of the driver glycans. All of the analyzed samples could be assigned to 1 of 3 groups based on driver glycan prevalence: S-Group, dominated by sialylated glycans; F-Group, dominated by fucosylated glycans; M-Group, characterized by a mixture of the dominant S and F glycans (Fig. 6 and Supplementary Figures 2,3). Within each group, N-linked glycan profiles segregated normal from cancer (Fig. 7,8). For the S-group, cancer samples showed increases in the major driver glycans (structures 34, 50, 55) as well as increases in branching and additional sialylation of less prevalent glycans. For the M-group, an obvious trend was not discernable among the major driver glycans. However, among the less prevalent glycans in the M-group, increases in high mannose (structures 4, 6, 13, 22, 21) and less complex glycans (structures 9, 17) mirrored decreases in highly branched, fully galactosylated and

fucosylated glycans in the cancer and IPMN samples. For the F-group, the cancer sample showed decreases in the major driver glycans (structures 26, 33, 40) as well as increases in glycan branching, outer arm fucosylation, and poly-LacNAc extension (structures 75, 79, 81, 82, 85). The glycan structures that define the S-, M-, and F-groups (Fig. 6) are not biosynthetic precursors for the glycan structures that differentiate cancer from control samples within each group (Fig. 7,8), indicating that the generation of putative marker glycans does not simply reflect the up-regulation of a dominant glycan processing pathway.

Interestingly, the glycan profiles of the IPMN samples tested here exhibited characteristics of normal and cancer samples. Consequently, the IPMN sample assigned to the F-group clustered with normal and the IPMN sample in the M-group clustered with cancer. The segregation of the IPMN samples likely reflects the transitional nature of this diagnosis, with IPMN patients exhibiting a continuum of clinical presentation, including the possible progression to adenocarcinoma.

DISCUSSION

Glycomic and Proteomic Biomarker Strategy

The identification of proteomic markers for human disease holds promise for improving early diagnosis and for enhancing clinicians' ability to monitor treatment efficacy. Glycomic markers offer similar opportunities and several are currently in service as cancer diagnostics (CA125, CA19-9, core fucosylation of α FP). We have proposed that a marker or set of markers that reports changes in both protein and glycan composition could potentially yield higher sensitivity and specificity than a single protein or glycan marker. This is especially true in a disease that affects a small percentage of the population such as pancreatic cancer in the United States. With an incidence rate of approximately 1 in 10,000 Americans, a screen with a 1% false-positive rate and 1% false-negative rate of one million Americans would be expected to come back with 10,099 positive identifications. 99 of these would be expected to be true-positives (from an expected 100 in the population group) while 10,000 of them would be false-positives. This simple example illustrates the need for multiplexing and orthogonal analyses.

Predictive protein glycosylation changes might be found on proteins that are themselves biomarkers or might be detected as a change to the whole glycan profile of a biological sample. Here, we have concentrated on defining the scope of the proteomic and glycomic changes associated with pancreatic cancer, as detected in ductal fluid. Our quantitative proteomic results identified several proteins that are distinctively upregulated or downregulated in pancreatic cancer and/or IPMN samples, and some proteins that are only detected in cancer. Our analysis of N-linked glycans released from ductal fluid glycoproteins revealed unexpected diversity and interesting commonalities between subjects and also detected changes in glycan expression that correlate with pancreatic cancer. Future work will investigate whether the glycosylation changes that we have detected can be mapped to the putative glycoprotein markers that we have identified. The current work (using a training set of samples) establishes important parameters for validating candidate glycomic and proteomic biomarkers (using a confirmatory set of samples) and for interpreting and expanding this biomarker discovery effort.

Secreted Pancreatic Enzyme Proteomics

The ductal fluid proteins most predictive of cancer in our training set are primarily secreted proteins that possess degradative enzyme activities consistent with the exocrine function of the pancreas. A subset of lipases, glycosidases, and proteases exhibit changes ranging from a 6-fold increase to a 22-fold decrease in cancer. We detected decreased pancreatic amylase

(AMYP) in cancer ductal fluid, consistent with previous studies in a rat model that demonstrated significant loss of amylase from pancreatic tumor cells by immunocytochemistry²⁸. Likewise, decreased levels of elastase activity in duodenal aspirates have been reported following secretin-induced secretion in chronic pancreatitis, pancreatic cancer, and liver cirrhosis patients compared to normal controls²⁹. Consistent with this result, and with another recent proteomic analysis of pancreatic ductal fluid by Gao, et al., we detected decreased elastase proteins (ELA2A, 3A, and 3B) in cancer¹⁰. However, increased levels of elastase 3b (formerly designated elastase 1) in pancreatic cancer tissue samples have been observed by several groups³⁰⁻³⁵. Considering the full range of proteomic analyses completed to date, the utility of elastase 3b as a biomarker for pancreatic cancer is in doubt, especially in light of its reported changes in pancreatitis as well as cancer. Pancreatic lipase-related protein 2 (LIPR2), which is the major colipase-dependent lipase in the pancreas³⁶, has been implicated in tumor cell killing through apoptotic and necrotic death induced by high levels of unsaturated fatty acids³⁷⁻⁴⁴. Consistent with the proteomic results presented here, pancreatic lipase immunoreactivity in serum was shown to decrease in pancreatic cancer⁴⁵. However, the recent proteomic characterization by Gao, et al. reported increased LIPR1 in cancer and failed to detect LIPR2, while we detect no change in LIPR1 and decreased LIPR2¹⁰.

The proteomic analysis by Gao, et al. also reported increased serine proteinase-2 (PRSS2, trypsin 2) in cancer, a protein that we failed to detect, but our analysis did identify PRSS1 (trypsin 1) as a significantly increased candidate biomarker for cancer¹⁰. Disagreements between the current study and the results of Gao, et al. likely reflect the very different techniques used (2D-gel electrophoresis followed by MALDI-TOF/MS versus LC-iontrap MS/MS). Another protein that we measured to be increased in pancreatic cancer is phospholipase A2 (PLA2), which has been previously associated with breast, lung and prostate cancers⁴⁶⁻⁵⁶. We were unable to detect PLA2 in normal ductal fluid, although it was present in ductal fluid of cancer, IPMN, and pancreatitis patients, thereby failing to discriminate based on presence between pancreatic disease types. Therefore, among the major secreted enzymes detected in pancreatic ductal fluid, we have identified increased AMYP and PRSS1, as well as decreased LIPR2 as candidate biomarkers worthy of further validation.

Non-enzyme Pancreatic Proteomics

Non-enzyme proteins, such as GP2-1, CCDC132, and REG family members also show significant changes in pancreatic ductal fluid. GP2-1 is a major glycoprotein of pancreatic acinar cell secretory granules and our analysis demonstrates that it is significantly increased in ductal fluid of IPMN and cancer patients compared to normal⁵⁷. GP2-1 exists as a GPI-anchored form and as a truncated form that is secreted into the ductal fluid. GP2-1 function is incompletely characterized but the protein has significant similarity to uromodulin (Tamm-Horsfall protein), a kidney protein secreted into the urine and associated with renal innate immunity and ionic homeostasis⁵⁸. Similarly, the coiled-coil domain-containing protein 132 (CCDC132) was only quantifiable or detectable in cancer and in one IPMN sample in our dataset; it was not detected in normal ductal fluid. Like GP2-1, the function of CCDC132 is currently unknown, but this cytoplasmic phosphoprotein possesses an N-terminal domain with homology to vacuolar sorting factors, suggesting a role in protein trafficking and association with secretory or transport vesicles⁵⁹. The REG proteins are a group of structurally related proteins that stimulate proliferation and differentiation of liver, pancreatic, gastric, and intestinal cell populations²⁷. Members of the REG protein family have been linked to gastric, liver, and pancreatic cancer and we detected increased REG1a, REG1b, and REG3a proteins in ductal fluid from cancer patients⁶⁰⁻⁶³. Interestingly, we not only detected changes in REG protein amounts, but also in REG protein glycoform

distribution by Western blotting. In general, REG proteins exhibited greater heterogeneity in cancer than in normal ductal fluid. The differences that generate this heterogeneity are currently uncharacterized and may reflect distinct glycosylation profiles or differential proteolytic processing of the apoproteins. Therefore, among the non-enzyme proteins that we detected in pancreatic ductal fluid, we have identified increased GP2-1, CCDC132, REG1a, REG1b, and REG3a, as well as increased heterogeneity of REG proteins as candidate biomarkers requiring further validation.

N-linked Glycomics of Pancreatic Cancer

Changes in N-linked protein glycosylation have been described in many cancers including pancreatic cancer, in which altered glycosylation of serum proteins has been demonstrated⁶⁴⁻⁶⁷. The serum proteins previously reported to exhibit altered N-linked glycosylation in pancreatic cancer are acute phase proteins normally produced by hepatocytes, not pancreatic cells, in response to systemic inflammation. By monitoring the glycan profile of pancreatic ductal fluid, we have accessed the secretory products of normal and cancerous pancreatic cells in the compartment of closest proximity to their biosynthetic origin. Our purpose was to assess glycan profiles in a biological sample that would provide the greatest opportunity to detect relevant changes. Surprisingly, our N-linked glycomic analysis revealed 3 distinct glycosylation signatures within the 9 samples analyzed. Each signature was defined by a set of highly prevalent driver glycans.

These signatures were independent of cancer diagnosis, blood group status, and any other characteristic captured by our sample collection protocol, including age and sex. However, the structural features of the driver glycans that define these groups suggest that expression of the Secretor α 1-2 Fucosyltransferase (Se FucT), which generates blood group H epitopes in epithelial cells, might account for some of the differences. If true, the S-group could be assigned as $Se-/Se-$, the M-group as $Se-/Se+$, and the F-group as $Se+/Se+$. These assignments make sense when considering the reciprocal gradation of fucosylation/sialylation across the groups; decreasing Se FucT activity increases the prevalence of unmodified terminal Gal residues that are substrates for sialyltransferases. But the changes we observed in the total glycan profiles of cancer samples within each group indicate a more complex scenario. In particular, it is difficult to propose how altered terminal fucosylation might affect branching, which we detect as increased in the S- and F-groups but decreased in the M-group. Likewise, increased poly-LacNAc in the F-group was countered by decreased poly-LacNAc in the M-group. Such divergent changes in cancer glycan expression cannot be linked simply to the level of active Se FucT. Rather, the divergent glycan profiles must reflect underlying changes in protein glycosylation that accompany cancer progression.

It remains to be determined whether all human samples can be clustered into these three groups or whether additional glycan signatures will be defined by other principal components as we increase the statistical power of our analysis. No analogous balkanization of glycan profiles has been described for serum glycomics, perhaps reflecting the more restricted cellular origin of pancreatic ductal fluid in comparison to systemically circulating blood. Regardless of its origin, the sorting of human subjects into discrete glycomic bins provides unique opportunities to pursue personalized glycan-based diagnostics. Our data indicates that proteomic and glycomic analysis of pancreatic ductal fluid must first assign samples to a driver glycan class before attempting to decipher the relevance of candidate glycoproteomic markers. On one hand, this classification requires that more samples must be characterized to achieve statistical power within each population. But, once sorted, glycomic and proteomic differences may provide markers capable of discriminating clinical diagnoses with greater specificity and sensitivity than is currently available.

CONCLUSIONS

Increases (AMYP, PRSS1) and decreases (LIPR2) of secreted pancreatic enzymes and increases of non-enzyme pancreatic proteins (GP2-1, CCDC132, REG1a, 1b, 3a) were detected in cancerous pancreatic ductal fluid in comparison to normal pancreatic ductal fluid in a small training set of samples. In addition, heterogeneity of the REG proteins was also found to increase in cancerous ductal fluid. A comprehensive analysis of the N-linked glycome of pancreatic ductal fluid identified unexpected clustering of patient samples into discrete subgroups that are enriched in sialylation or fucosylation, or are mixed with respect to both types of glycans independent of diagnosis. Within each group, changes in glycan prevalences are detected comparing normal to cancer albeit on a small number of samples. But, across groups, the glycan expression changes are different, even opposite in some cases. Therefore, interpretation of glycomic and glycoproteomic profiles must consider the heterogeneity of glycosylation across human populations before assessing the meaningfulness of changes in candidate biomarkers. The proteomic and glycomic features extracted from the training set of samples reported here establish important parameters for expanded validation and emphasize the need for large sample sets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Will York (CCRC) for building and sharing an intuitive interface for hierarchical clustering and René Renzinger for developing software tools that facilitated the extraction of MS and MS/MS glycomic features. This work was supported by U01 CA128454 from NIH/NCI and P41 GM103490 from NIH/NIGMS.

REFERENCES

1. American Cancer Society Cancer Facts & Figures 2010. American Cancer Society; Atlanta: 2010. p. 1-66.2010
2. Bernhard J, Dietrich D, Glimelius B, Hess V, Bodoky G, Scheithauer W, Herrmann R. Estimating prognosis and palliation based on tumour marker CA 19-9 and quality of life indicators in patients with advanced pancreatic cancer receiving chemotherapy. *Br J Cancer*. 2010; 103(9):1318–24. [PubMed: 20877359]
3. Magnani JL, Steplewski Z, Koprowski H, Ginsburg V. Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer Res*. 1983; 43(11):5489–92. [PubMed: 6193872]
4. Xue A, Scarlett CJ, Chung L, Butturini G, Scarpa A, Gandy R, Wilson SR, Baxter RC, Smith RC. Discovery of serum biomarkers for pancreatic adenocarcinoma using proteomic analysis. *Br J Cancer*. 2010; 103(3):391–400. [PubMed: 20588270]
5. Kopp JL, von Figura G, Mayes E, Liu FF, Dubois CL, Morris J. P. t. Pan FC, Akiyama H, Wright CV, Jensen K, Hebrok M, Sander M. Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. *Cancer Cell*. 2012; 22(6):737–50. [PubMed: 23201164]
6. Chen R, Brentnall TA, Pan S, Cooke K, Moyes KW, Lane Z, Crispin DA, Goodlett DR, Aebersold R, Bronner MP. Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer. *Mol Cell Proteomics*. 2007; 6(8):1331–42. [PubMed: 17496331]
7. Chen R, Pan S, Cooke K, Moyes KW, Bronner MP, Goodlett DR, Aebersold R, Brentnall TA. Comparison of pancreas juice proteins from cancer versus pancreatitis using quantitative proteomic analysis. *Pancreas*. 2007; 34(1):70–9. [PubMed: 17198186]

8. Chen R, Pan S, Yi EC, Donohoe S, Bronner MP, Potter JD, Goodlett DR, Aebersold R, Brentnall TA. Quantitative proteomic profiling of pancreatic cancer juice. *Proteomics*. 2006; 6(13):3871–9. [PubMed: 16739137]
9. Doyle CJ, Yancey K, Pitt HA, Wang M, Bemis K, Yip-Schneider MT, Sherman ST, Lillemoe KD, Goggins MD, Schmidt CM. The Proteome of Normal Pancreatic Juice. *Pancreas*. 2011
10. Gao J, Zhu F, Lv S, Li Z, Ling Z, Gong Y, Jie C, Ma L. Identification of pancreatic juice proteins as biomarkers of pancreatic cancer. *Oncol Rep*. 2010; 23(6):1683–92. [PubMed: 20428826]
11. Gronborg M, Bunkenborg J, Kristiansen TZ, Jensen ON, Yeo CJ, Hruban RH, Maitra A, Goggins MG, Pandey A. Comprehensive proteomic analysis of human pancreatic juice. *J Proteome Res*. 2004; 3(5):1042–55. [PubMed: 15473694]
12. Rosty C, Goggins M. Identification of differentially expressed proteins in pancreatic cancer using a global proteomic approach. *Methods Mol Med*. 2005; 103:189–97. [PubMed: 15542907]
13. Comunale MA, Rodemich-Betesh L, Hafner J, Wang M, Norton P, Di Bisceglie AM, Block T, Mehta A. Linkage specific fucosylation of alpha-1-antitrypsin in liver cirrhosis and cancer patients: implications for a biomarker of hepatocellular carcinoma. *PLoS One*. 2010; 5(8):e12419. [PubMed: 20811639]
14. Marrero JA, Romano PR, Nikolaeva O, Steel L, Mehta A, Fimmel CJ, Comunale MA, D'Amelio A, Lok AS, Block TM. GP73, a resident Golgi glycoprotein, is a novel serum marker for hepatocellular carcinoma. *J Hepatol*. 2005; 43(6):1007–12. [PubMed: 16137783]
15. Nakagawa T, Takeishi S, Kameyama A, Yagi H, Yoshioka T, Moriwaki K, Masuda T, Matsumoto H, Kato K, Narimatsu H, Taniguchi N, Miyoshi E. Glycomic analyses of glycoproteins in bile and serum during rat hepatocarcinogenesis. *J Proteome Res*. 2010; 9(10):4888–96. [PubMed: 20731380]
16. Sato Y, Nakata K, Kato Y, Shima M, Ishii N, Koji T, Taketa K, Endo Y, Nagataki S. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. *N Engl J Med*. 1993; 328(25):1802–6. [PubMed: 7684823]
17. Wandall HH, Blixt O, Tarp MA, Pedersen JW, Bennett EP, Mandel U, Ragupathi G, Livingston PO, Hollingsworth MA, Taylor-Papadimitriou J, Burchell J, Clausen H. Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes. *Cancer Res*. 2010; 70(4):1306–13. [PubMed: 20124478]
18. Aoki K, Perlman M, Lim J, Cantu R, Wells L, Tiemeyer M. Dynamic developmental elaboration of N-linked glycan complexity in the *Drosophila melanogaster* embryo. *J Biol Chem*. 2007; 282(12):9127–9142. [PubMed: 17264077]
19. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dummer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004; 32(Database issue):D258–61. [PubMed: 14681407]
20. Zybilov B, Mosley AL, Sardu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res*. 2006; 5(9):2339–47. [PubMed: 16944946]
21. Zybilov BL, Florens L, Washburn MP. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol Biosyst*. 2007; 3(5):354–60. [PubMed: 17460794]
22. Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem*. 2010; 82(6):2272–81. [PubMed: 20166708]
23. Ciucanu I, Kerek F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research*. 1984; 131:209–217.

24. Apte A, Meitei NS. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods Mol Biol.* 2010; 600:269–81. [PubMed: 19882135]
25. Ferguson MW, Wycoff KL, Ayers AR. Use of cluster analysis with monoclonal antibodies for taxonomic differentiation of phytopathogenic fungi and for screening and clustering antibodies. *Curr Microbiol.* 1988; 17:127–132.
26. Pattathil S, Avci U, Baldwin D, Swennes AG, McGill JA, Popper Z, Bootten T, Albert A, Davis RH, Chennareddy C, Dong R, O'Shea B, Rossi R, Loeff C, Freshour G, Narra R, O'Neil M, York WS, Hahn MG. A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol.* 2010; 153(2):514–25. [PubMed: 20363856]
27. De Reggi M, Capon C, Gharib B, Wieruszkeski JM, Michel R, Fournet B. The glycan moiety of human pancreatic lithostathine. Structure characterization and possible pathophysiological implications. *Eur J Biochem.* 1995; 230(2):503–10. [PubMed: 7607222]
28. Hilmy AM, Kandeel KM, Selim NM. Pancreatic amylase as a tumour marker for pancreatic cancer. *Arch Geschwulstforsch.* 1984; 54(6):475–82. [PubMed: 6084495]
29. Mizuno R, Hayakawa T, Noda A. Elastase secretion in pancreatic disease. *Am J Gastroenterol.* 1985; 80(2):113–7. [PubMed: 3844284]
30. Achilles K, Bednarski PJ. Quantification of elastase-like activity in 13 human cancer cell lines and in an immortalized human epithelial cell line by RP-HPLC. *Biol Chem.* 2003; 384(5):817–24. [PubMed: 12817479]
31. Iwase K, Miyata M, Yamaguchi T, Kawaguchi T, Tanaka Y, Matsuda H. Pancreatic ductal cell carcinoma producing pancreatic elastase 1. *J Surg Oncol.* 1993; 54(3):199–202. [PubMed: 8412180]
32. Ventrucci M, Pezzilli R, Gullo L, Plate L, Sprovieri G, Barbara L. Role of serum pancreatic enzyme assays in diagnosis of pancreatic disease. *Dig Dis Sci.* 1989; 34(1):39–45. [PubMed: 2463138]
33. Wendorf P, Geyer R, Sziegoleit A, Linder D. Localization and characterization of the glycosylation site of human pancreatic elastase 1. *FEBS Lett.* 1989; 249(2):275–8. [PubMed: 2737288]
34. Wendorf P, Linder D, Sziegoleit A, Geyer R. Carbohydrate structure of human pancreatic elastase 1. *Biochem J.* 1991; 278(Pt 2):505–14. [PubMed: 1898343]
35. Wu D, Qian JM, Deng RX, Jiang WJ, Chen YJ, Liu XH, Lu XH. Evaluating the role of serum elastase 1 in the diagnosis of pancreatic cancer. *Chin J Dig Dis.* 2006; 7(2):117–20. [PubMed: 16643340]
36. D'Agostino D, Lowe ME. Pancreatic lipase-related protein 2 is the major colipase-dependent pancreatic lipase in suckling mice. *J Nutr.* 2004; 134(1):132–4. [PubMed: 14704305]
37. Cury-Boaventura MF, Curi R. Regulation of reactive oxygen species (ROS) production by C18 fatty acids in Jurkat and Raji cells. *Clin Sci (Lond).* 2005; 108(3):245–53. [PubMed: 15563273]
38. Cury-Boaventura MF, Gorjao R, de Lima TM, Newsholme P, Curi R. Comparative toxicity of oleic and linoleic acid on human lymphocytes. *Life Sci.* 2006; 78(13):1448–56. [PubMed: 16236329]
39. Cury-Boaventura MF, Pompeia C, Curi R. Comparative toxicity of oleic acid and linoleic acid on Jurkat cells. *Clin Nutr.* 2004; 23(4):721–32. [PubMed: 15297111]
40. Cury-Boaventura MF, Pompeia C, Curi R. Comparative toxicity of oleic acid and linoleic acid on Raji cells. *Nutrition.* 2005; 21(3):395–405. [PubMed: 15797684]
41. Finstad HS, Heimli H, Kolset SO, Drevon CA. Proliferation and types of killing of leukemia cell lines by very long chain polyunsaturated fatty acids. *Lipids.* 1999; (34 Suppl):S107. [PubMed: 10419110]
42. Finstad HS, Myhrstad MC, Heimli H, Lomo J, Blomhoff HK, Kolset SO, Drevon CA. Multiplication and death-type of leukemia cell lines exposed to very long-chain polyunsaturated fatty acids. *Leukemia.* 1998; 12(6):921–9. [PubMed: 9639421]
43. Heimli H, Finstad HS, Drevon CA. Necrosis and apoptosis in lymphoma cell lines exposed to eicosapentaenoic acid and antioxidants. *Lipids.* 2001; 36(6):613–21. [PubMed: 11485166]

44. Lima TM, Kanunfre CC, Pompeia C, Verlengia R, Curi R. Ranking the toxicity of fatty acids on Jurkat and Raji cells by flow cytometric analysis. *Toxicol In Vitro*. 2002; 16(6):741–7. [PubMed: 12423658]
45. Hayakawa T, Kondo T, Shibata T, Kitagawa M, Ono H, Sakai Y, Kiriya S. Enzyme immunoassay for serum pancreatic lipase in the diagnosis of pancreatic diseases. *Gastroenterol Jpn*. 1989; 24(5):556–60. [PubMed: 2478412]
46. Denizot Y, Chianea T, Labrousse F, Truffinet V, Delage M, Mathonnet M. Platelet-activating factor and human thyroid cancer. *Eur J Endocrinol*. 2005; 153(1):31–40. [PubMed: 15994743]
47. Dong Q, Patel M, Scott KF, Graham GG, Russell PJ, Sved P. Oncogenic action of phospholipase A2 in prostate cancer. *Cancer Lett*. 2006; 240(1):9–16. [PubMed: 16182442]
48. Graff JR, Konicek BW, Deddens JA, Chedid M, Hurst BM, Colligan B, Neubauer BL, Carter HW, Carter JH. Expression of group IIA secretory phospholipase A2 increases with prostate tumor grade. *Clin Cancer Res*. 2001; 7(12):3857–61. [PubMed: 11751475]
49. Jiang J, Neubauer BL, Graff JR, Chedid M, Thomas JE, Roehm NW, Zhang S, Eckert GJ, Koch MO, Eble JN, Cheng L. Expression of group IIA secretory phospholipase A2 is elevated in prostatic intraepithelial neoplasia and adenocarcinoma. *Am J Pathol*. 2002; 160(2):667–71. [PubMed: 11839587]
50. Laye JP, Gill JH. Phospholipase A2 expression in tumours: a target for therapeutic intervention? *Drug Discov Today*. 2003; 8(15):710–6. [PubMed: 12927514]
51. Sved P, Scott KF, McLeod D, King NJ, Singh J, Tsatralis T, Nikolov B, Boulas J, Nallan L, Gelb MH, Sajinovic M, Graham GG, Russell PJ, Dong Q. Oncogenic action of secreted phospholipase A2 in prostate cancer. *Cancer Res*. 2004; 64(19):6934–40. [PubMed: 15466184]
52. Tribler L, Jensen LT, Jorgensen K, Brunner N, Gelb MH, Nielsen HJ, Jensen SS. Increased expression and activity of group IIA and X secretory phospholipase A2 in peritumoral versus central colon carcinoma tissue. *Anticancer Res*. 2007; 27(5A):3179–85. [PubMed: 17970059]
53. Yamashita J, Ogawa M, Sakai K. Prognostic significance of three novel biologic factors in a clinical trial of adjuvant therapy for node-negative breast cancer. *Surgery*. 1995; 117(6):601–8. [PubMed: 7778023]
54. Yamashita S, Ogawa M, Sakamoto K, Abe T, Arakawa H, Yamashita J. Elevation of serum group II phospholipase A2 levels in patients with advanced cancer. *Clin Chim Acta*. 1994; 228(2):91–9. [PubMed: 7988039]
55. Yamashita S, Yamashita J, Ogawa M. Overexpression of group II phospholipase A2 in human breast cancer tissues is closely associated with their malignant potency. *Br J Cancer*. 1994; 69(6):1166–70. [PubMed: 8198986]
56. Yamashita S, Yamashita J, Sakamoto K, Inada K, Nakashima Y, Murata K, Saishoji T, Nomura K, Ogawa M. Increased expression of membrane-associated phospholipase A2 shows malignant potential of human breast cancer cells. *Cancer*. 1993; 71(10):3058–64. [PubMed: 8490834]
57. Scheele GA, Fukuoka S, Freedman SD. Role of the GP2/THP family of GPI-anchored proteins in membrane trafficking during regulated exocrine secretion. *Pancreas*. 1994; 9(2):139–49. [PubMed: 8190715]
58. Rampoldi L, Scolari F, Amoroso A, Ghiggeri G, Devuyst O. The rediscovery of uromodulin (Tamm-Horsfall protein): from tubulointerstitial nephropathy to chronic kidney disease. *Kidney Int*. 2011; 80(4):338–47. [PubMed: 21654721]
59. Matsumoto Y, Imai Y, Sugita Y, Tanaka T, Tsujimoto G, Saito H, Oshida T. CCDC132 is highly expressed in atopic dermatitis T cells. *Mol Med Report*. 2010; 3(1):83–7.
60. Sekikawa A, Fukui H, Fujii S, Ichikawa K, Tomita S, Imura J, Chiba T, Fujimori T. REG Ialpha protein mediates an anti-apoptotic effect of STAT3 signaling in gastric cancer cells. *Carcinogenesis*. 2008; 29(1):76–83. [PubMed: 18024479]
61. Usami S, Motoyama S, Koyota S, Wang J, Hayashi-Shibuya K, Maruyama K, Takahashi N, Saito H, Minamiya Y, Takasawa S, Ogawa J, Sugiyama T. Regenerating gene I regulates interleukin-6 production in squamous esophageal cancer cells. *Biochem Biophys Res Commun*. 2010; 392(1):4–8. [PubMed: 20056108]
62. Yamagishi H, Fukui H, Sekikawa A, Kono T, Fujii S, Ichikawa K, Tomita S, Imura J, Hiraishi H, Chiba T, Fujimori T. Expression profile of REG family proteins REG Ialpha and REG IV in

- advanced gastric cancer: comparison with mucin phenotype and prognostic markers. *Mod Pathol*. 2009; 22(7):906–13. [PubMed: 19329938]
63. Zhou L, Zhang R, Wang L, Shen S, Okamoto H, Sugawara A, Xia L, Wang X, Noguchi N, Yoshikawa T, Uruno A, Yao W, Yuan Y. Up-regulation of REG Ialpha accelerates tumor progression in pancreatic cancer with diabetes. *Int J Cancer*. 2010
64. Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. *Biochim Biophys Acta*. 2011
65. Cazet A, Julien S, Bobowski M, Burchell J, Delannoy P. Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res*. 2010; 12(3):204. [PubMed: 20550729]
66. Reis CA, Osorio H, Silva L, Gomes C, David L. Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol*. 2010; 63(4):322–9. [PubMed: 20354203]
67. Sarrats A, Saldoval R, Pla E, Fort E, Harvey DJ, Struwe WB, de Llorens R, Rudd PM, Peracaula R. Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis. *Proteomics Clin Appl*. 2010; 4(4):432–48. [PubMed: 21137062]

Figure 1. Subcellular localization and biological function of proteins identified in 12 pancreatic ductal fluid samples

Distributions were calculated based on spectral counts of identified proteins.

Figure 2. Data filter process flow chart

368 proteins were identified by Sequest after filtering at 1% protein-level false-discovery rate (FDR). The dataset was then further filtered at 10% peptide-level FDR and 1-hit proteins were eliminated. In the resulting dataset, only proteins that were observed in at least 2 out of 3 patients were considered for quantification, and finally 47 proteins were quantified.

Figure 3. Variations in protein expression for pancreatitis, IPMN, and cancer samples
Protein expression variation in pancreatitis, IPMN, and cancer samples in reference to normal controls. The ratios are calculated based on dNASF values of quantified proteins and are plotted on a Log2 scale.

Figure 4. Validation of proteomic data by immunoblotting

Pancreatic ductal fluid samples with diagnosis of pancreatic cancer (C5, C6, C7, C8, and C9) were compared to normal controls (N4, N5, N6, N7, and N8) by probing with respective antibodies: (A) REG1A, (B) REG1B, (C) REG3A, (D) Phospholipase A2 (PLA2), and (E) Pancreatic lipase-related protein 2 (LIPR2). Numbers on the left side of the blots indicate molecular weights in kDa. The split panels in A and D were originally part of the same blot, one for A and one for D. The lanes of interest were originally separated by irrelevant samples and have been brought together to facilitate direct comparison.

Figure 5. N-linked glycans that differentiate between normal and cancer/IPMN identified from whole glycan profiles

A set of 6 dominant glycans (see Fig. 6) was removed from the whole profile and the prevalence of the remaining 79 glycans was recalculated for each sample. After recalculation, 9 glycans exhibited statistically significant changes comparing cancer (C1 - C4) and IPMN (IP1 - IP2) patients to normal (N1 - N3, Wilcoxon rank-sum $p = 0.05$). Hierarchical clustering of the prevalences of these 9 glycans demonstrates that their prevalences segregate normal ductal fluid glycan profiles from the glycan profiles of cancer or IPMN. Graphic representation of glycan structures are in accordance with the guidelines proposed by the Consortium for Functional Glycomics (CFG): blue square, N-acetylglucosamine (GlcNAc); green circle, mannose (Man); yellow circle, galactose (Gal); red triangle, fucose (Fuc); pink diamond, sialic acid as N-acetylneuraminic acid (NeuAc); light blue diamond, sialic acid as N-glycolylneuraminic acid (NeuGc). Glycan numbers are provided as arbitrary identifiers and refer consistently to the same structure throughout the manuscript and in the supplementary information (figures and tables). Brackets across the top of the cluster diagram provide a graphic presentation of the relatedness of the profile defined by each column. Thus, the total path length separating any two samples is directly proportional to the similarity of the glycan profile presented by those samples. For instance, N1 and N3 are more similar to each other than either is to N2 and all of the N samples are more similar to each other than any of the IP or C samples.

Figure 6. Dominant glycans define three distinct sample groups

Analysis of 9 pancreatic ductal fluid samples identified 6 glycans that dominate the total glycan profiles of discrete sample subsets. These driver glycans defined three groups: S-group, dominated by sialylated glycans; F-group, dominated by fucosylated glycans; and M-group, presented a balance of the S- and F-group drivers. Hierarchical clustering robustly segregates S-, M-, and F-group samples. N1, N2, and N3: normal samples. C1, C2, C3, and C4: cancer samples. IP1, and IP2: IPMN samples. Glycan notation, numeric assignments, and clustering representations are as described in the legend to Figure 5.

NIH-PA Author Manuscript

Figure 8. Summary of group-specific changes in glycan structural features that discriminate normal from cancer

For S-group samples, cancer glycan profiles are characterized by increased driver glycan expression and decreased high-mannose glycans. For F-group samples, driver glycans are decreased in cancer samples and glycan complexity is increased. For M-group samples, driver glycan expression is not significantly altered, but high-mannose structures are increased and the most complex glycan structures are decreased in prevalence. Thus, each group exhibits unique changes in ductal fluid glycan expression when comparing normal to cancer. N1, N2, and N3: normal samples. C1, C2, C3, and C4: cancer samples. IP1, and IP2: IPMN samples.

Table 1

Number of identifiable and quantifiable proteins in pancreatic ductal fluid.

Identified Proteins				
Diagnosis	Proteins	Peptides	Spectral counts	Single-hit proteins
Cancer	213	1094	18206	77
IPMN ^a	149	831	22641	57
Pancreatitis	163	769	6645	61
Normal	136	775	11635	49
Combined	451	2082	59127	184
Quantified Proteins				
Diagnosis	Proteins	Peptides	Spectral counts	Unique proteins
Cancer	36	422	15092	7
IPMN	35	414	18632	8
Pancreatitis	19	215	3774	1
Normal	22	300	8674	0

^a IPMN, Intraductal papillary mucinous neoplasm

Table 2

Quantified pancreatic ductal fluid proteins differentially expressed in IPMN, and cancer relative to normal.

Uniprot Accession	Abbreviation	Protein Name	Gene Name	Protein Length (AA)	Protein Weight (kDa)	PT/N ^a (log2 of ratio)	IP/N (log2 of ratio)	C/N (log2 of ratio)
P02787	TRFE	Serotransferrin	IF	698	76.982	3.49	NQ ^b in IP	2.54
P04118	COL	Colipase	CLPS	112	11.928		0.44	0.04
P04745	AMY1	Alpha-amylase 1	AMY1 A	511	57.713	-0.60	2.30	2.95
P04746	AMYP	Pancreatic alpha-amylase	AMY2A	511	57.652	1.05	-0.24	-1.96
P05451	REGIA	Lithostathine-1-alpha	REGIA	166	18.701	0.23	0.52	1.13
P07477	TRY1	Trypsin-1	PRSS 1	247	26.523	0.37	1.11	1.48
P08217	ELA2A	Elastase-2A	ELA2A	269	28.851	-0.33	-0.29	-1.65
P08861	ELA3B	Elastase-3B	ELA3B	270	29.256	-2.39	-1.17	-1.93
P09093	ELA3A	Elastase-3A	ELA3A	270	29.438	-0.23	-0.75	-0.67
P15085	CBPA1	Carboxypeptidase A1	CPA1	419	47.093	-0.39	-0.73	-0.70
PI 5086	CBPBI	Carboxy peptidase B	CPBI	417	47.320	-0.32	-0.80	0.08
P16233	LIPP	Pancreatic triacylglycerol lipase	PNLIP	465	51.106	0.57	0.88	0.18
P17538	CTRB1	Chymotrypsinogen B	CTRB1	263	27.834	0.54	0.86	0.78
P19835-1	CEL	Isoform Long of Bile salt-activated lipase	CEL	742	78.278	-0.01	-0.19	-0.72
P19961	AMY2B	Alpha-amylase 2B	AMY2B	511	57.655	NQ in PT	1.30	NQ in C
P48052	CBPA2	Carboxy peptidase A2	CPA2	417	46.781	NQ in PT	0.31	-0.54
P54315-1	LIPR1	Isoform 1 of Pancreatic lipase-related protein 1	PNLIPRP1	467	51.797	NQ in PT	-0.20	NQ in C
P543 17	LIPR2	Pancreatic lipase-related protein 2	PNLIPRP2	469	51.895	NQ in PT	NQ in IP	-4.56
P55259-1	GP2-1	Isoform 1 of Pancreatic secretory granule membrane major glycoprotein GP2	GP2	537	59.424	NQ in PT	2.87	3.73
P55259-3	GP2-3	Isoform Alpha of Pancreatic secretory granule membrane major glycoprotein GP2	GP2	534	59.071	NQ in PT	1.37	-0.65
Q3SY19	PRSSI	PRSSI protein	PRSSI	247	26.521	-0.20	0.50	1.98

^a N: Normal; PT: Pancreatitis; IP: Intraductal papillary mucinous neoplasm; C: Cancer.^b NQ: Not quantifiable in the indicated diagnosis.

Table 3

Quantified pancreatic ductal fluid proteins differentially expressed in IPMN and cancer relative to pancreatitis.

Uniprot Accession	Abbreviation	Protein Name	Gene Name	Protein Length (AA)	Protein Weight (kDa)	IP/PT ^a (log2 of ratio)	C/PT (log2 of ratio)
P68871	HBB	Hemoglobin subunit beta	HBB	147	15.970	-2.89	-1.00
A8K008	A8K008	cDNA FLJ78387	NA ^b	472	51.546	NQ ^c m IP	-2.84
Q5EFE6	Q5EFE6	Anti-RhD monoclonal T125 kappa light chain	NA	234	25.664	NQ in IP	-0.57

^aPT: Pancreatitis; IP: Intraductal papillary mucinous neoplasm; C: Cancer.^bNA: None assigned.^cTSTQ: Not quantifiable in the indicated diagnosis.

Table 4

Pancreatic ductal fluid proteins unique to each diagnosis.

Abbreviation	Protein Name	Gene Name	Quantifiable In		Identifiable In			
			C ^a	IP	C	IP	PT	N
			(# Patients)					
COS	Complement C3	C3	2	0	3	0	1	0
IGHG3	Ig gamma-3 chain C region	IGHG3	2	0	2	1	0	0
REGIB	Lithostathine-1 -beta	REGIB	2	0	2	1	1	1
CC132	Isoform 1 of Coiled-coil domain-containing protein 132	CCDC132	2	0	2	1	0	0
A0A5E4	Putative uncharacterized protein	NA ^b	2	0	2	1	1	1
Q569I7	Putative uncharacterized protein	NA	2	0	2	1	1	0
Q6ZP64	CDNA FLJ26451 fis, clone KDN03041	NA	2	0	2	0	0	0
PA21B	Phospholipase A2	PLA2G1B	2	3	2	1	2	0
ELA2B	Elastase-2B	ELA2B	2	3	3	1	1	0
TRY3	Isoform A of Trypsin-3	PRSS3	2	3	3	1	1	0
HBA	Hemoglobin subunit alpha	HBA1	2	3	3	1	0	0
REG3A	Regenerating islet-derived protein 3 alpha	REG3A	2	2	2	0	0	0
CTRC	Chymotrypsin-C	CTRC	2	3	3	1	1	0

^a N: Normal; PT: Pancreatitis; IP: Intraductal papillary mucinous neoplasm; C: Cancer.

^b NA: None assigned.