Published in final edited form as:

J Proteome Res. 2014 January 3; 13(1): 289–299. doi:10.1021/pr400813u.

Characterization of glycoproteins in pancreatic cyst fluid using a high performance multiple lectin affinity chromatography platform

Francisca Owusu Gbormittah¹, Brian B. Haab², Katie Partyka², Carolina Garcia-Ott¹, Marina Hancapie³, and William S. Hancock^{1,*}

¹Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

²Centers for Cancer Genomics and Quantitative Biology, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA

³Genzyme, a Sanofi Company, 45 New York Avenue, Framingham, MA. 01701

Abstract

Currently, pancreatic cancer is the fourth cause of cancer death. In 2013, it is estimated that approximately 38,460 people will die of pancreatic cancer. Early detection of malignant cyst (pancreatic cancer precursor) is necessary to help prevent late diagnosis of the tumor. In this study, we characterized glycoproteins and non-glycoproteins on pooled mucinous (n=10) and non-mucinous (n=10) pancreatic cyst fluid to identify ‘proteins of interest’ to differentiate between mucinous cyst from non-mucinous cyst and investigate these proteins as potential biomarker targets. An automated multi-lectin affinity chromatography (M-LAC) platform was utilized for glycoprotein enrichment followed by nano-LC-MS/MS analysis. Spectral count quantitation allowed for the identification of proteins with significant differential levels in mucinous cysts from non-mucinous cysts of which one protein (periostin) was confirmed via immunoblotting. To exhaustively evaluate differentially expressed proteins, we used a number of proteomic tools including; gene ontology classification, pathway and network analysis, Novoseek data mining and chromosome gene mapping. Utilization of complementary proteomic tools, revealed that several of the proteins such as mucin 6 (MUC6), bile salt-activated lipase (CEL) and pyruvate kinase

*Corresponding Author: wi.hancock@neu.edu. Telephone: 617-373-4881. Fax: 617-373-8795..

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

Figures S1-S3 and Tables S1-S3 are contained in one file. A summary of proteins identified in the two samples set and used for quantitation is presented in Tables S4-S7, as an additional file. Figure S1: Workflow diagram showing experimental process used in glycoproteomic studies of two analysis sample set. Pancreatic cyst fluid samples were purified from immunoglobulins and albumin using an immobilized antibody HPLC packed PEEK column. Glycoprotein enrichment followed by one dimensional gel electrophoresis were used as further fractionation steps leading to nano-LC-MS/MS analysis of M-LAC bound and unbound fractions. Figure S2: 1D-SDS PAGE of two sample sets before (2A) and after (2B) M-LAC fractionation. Variations in albumin levels and individual differences in each sample set accounts for variability in proteins identified in sample set one compared to sample set two. Figure S3: A representative MS/MS fragmentation of diagnostic peptide TYAYLFPSR of CEL-long isoform. Sequence coverage is 13.95, Charge: +2, Monoisotopic m/z: 671.333 Da (accuracy: +0.05 mmu/+0.08 ppm), ionscore: 44. This material is available free of charge via the Internet at <http://pubs.acs.org>. Nano-LC-MS/MS proteomic data can be found at Global Proteome Machine Database (GPMDB) using the following link <http://gpmdb.thegpm.org/data/keyword/pancreatic%20cyst>.

lysozyme M1/M2 with significant differential expression have strong association with pancreatic cancer. Further, chromosome gene mapping demonstrated co-expressions and co-localization of some proteins of interest including 14-3-3 protein epsilon (YWHAE), pigment epithelium derived factor (SERPINF1) and oncogene p53.

Keywords

Pancreatic cancer; pancreatic cyst fluid; mucinous cyst; non-mucinous cyst; glycoproteins; biomarker discovery

1. INTRODUCTION

Pancreatic cancer is one of the deadliest cancers with a 95% mortality rate within 5 years after diagnosis.¹ The main cause for an almost 100% death rate of pancreatic cancer is attributed to late detection of the tumor and subsequent late diagnosis of the disease.²⁻⁴ It is a difficult task to accurately make a prognosis of pancreatic cancer because pancreatic tumors are pathologically diverse with similar clinical and radiological characteristics.⁵⁻⁶ The most effective means to reduce mortality from pancreatic cancer may be to identify and remove precursor lesions before they progress to invasive cancer. Pancreatic cysts are potential precursors of pancreatic cancer that can be identified through non-invasive imaging and therefore detectable prior to progression.⁷ Some pancreatic cyst lesions do not have malignant potential, including; pseudocysts and serous cystadenomas (referred to as non-mucinous cysts), and others are established cancer precursors, including mucinous cystic neoplasms and intraductal papillary mucinous neoplasms (referred to as mucinous cysts). Unfortunately, it is sometimes difficult to distinguish the mucinous from the non-mucinous cysts by imaging or by clinical symptoms.⁶ Although, there are a number of parameters and techniques currently available for classifying malignant lesions and non-malignant lesions, more needs to be done since none of these methods provides definitive results.⁸

Glycoproteomics play an essential role in biomarker discovery studies of biological samples since an alteration in glycan structures and cellular glycosylation profile are closely related to cellular regulation and malignancy.⁹⁻¹¹ Investigating and analyzing glycoproteins of pancreatic cyst fluid represents a potentially valuable source for information and can benefit in differentiating mucinous from non-mucinous cysts. In glycoproteomics, specific glycoproteins and glyco-isoforms are enriched followed by matrix assisted laser desorption or liquid chromatography mass spectrometry analysis. Previous studies have demonstrated that by using antibody-lectin sandwich microarray some protein families and their glycan variants can be glyco-biomarker targets for accurate differentiation of mucinous cyst from non-mucinous cyst.⁹ Also, lectin based glycoproteomics of pancreatic cyst fluid have identified specific glycans and glycoforms as possible biomarker targets to differentiate mucinous cyst from non-mucinous cyst¹⁰⁻¹¹.

We have focused on the application of glycoproteins enrichment via a high performance multi-lectin affinity chromatography (HP-MLAC) method¹² to differentiate mucinous pancreatic cyst fluid subtypes from non-mucinous pancreatic cyst fluid subtypes. HP-MLAC is a robust and high throughput high performance multi-lectin affinity chromatography

platform that combines the depletion of two highly abundant proteins (human immunoglobulins and albumin), enrichment of glycoproteins and glyco-isoforms by multiple lectins (ConA, WGA and Jac) followed by a reversed phase sample clean up on an HPLC system. This platform is shown to result in the identification of potential glyco-biomarker targets, in plasma of breast cancer patients.¹²⁻¹³ We present in this report the glycoproteome as well as the non-glycoproteome landscape of pancreatic cyst fluid, which allows us to study the differences between mucinous cyst fluid vs. non-mucinous cyst fluid. First, we present data analysis based on our glycoprotein fractionation platform which indicated that a combination of high abundance protein depletion and enrichment by M-LAC followed by 1D SDS-PAGE fractionation allows us to characterize not only glycoproteins but also low abundant proteins which may be potential glyco-biomarker targets of interest.¹⁴ By using different complementary proteomics tools such as; gene ontology, Novoseek, pathway analysis, network interactions and chromosome gene mapping analysis, we show that ‘proteins of interest’ selected via spectral count have significant cancer associations and provides a good list for selection of target proteins for pancreatic cyst biomarker discovery.

2. MATERIALS AND METHODS

2.1. Reagents

Pancreatic cyst fluid samples used in this study were obtained from Dr. Brian Haab's laboratory at the Van Andel Research Institute (Grand Rapids, MI). Lectins for M-LAC column were purchased from Vector laboratories (Burlingame, CA). Capture select^(R) ligand with specificity for albumin proteins and protein G with affinity for immunoglobulin's (IgA, IgM, and IgG) were obtained from BAC. B. V., Netherlands and Life Technologies, Inc. (Carlsbad, CA) respectively. POROS 20 R1 and POROS beads for conjugation were also purchased from Applied Biosystems (Framingham, MA). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). HPLC-MS grade water, formic acid, acetonitrile and other buffer reagents were all purchased from Thermo Fisher Scientific (Waltham, MA).

2.2. Pancreatic cyst fluid samples

All pancreatic cyst samples involved in this study were collected in compliance with the guidelines of the local Institutional Review Boards at the University of Michigan Medical Center, Memorial Sloan-Kettering Cancer Center, the University of Arizona Medical Center, the University of Pittsburgh Medical Center and Ospedale Sacro Cuore-Don Calabria Negrar, Italy. Due to limited amount of materials, pancreatic cyst fluid samples for glycoproteomics were pooled into mucinous subtypes (potential malignant lesions) which includes; intraductal papillary mucinous neoplasms (IPMN) and mucinous cystic neoplasms (MCN) and non-mucinous subtypes (benign lesions) comprise of serous cystadenomas (SC) and pseudocysts (PC) for glycoproteomic studies. Mucin-like proteins, fats and other particulate matter were depleted from cyst fluid samples as follows; 200 μ L volume of each pool was mixed with equal volumes of 20 mM phosphate buffer pH 7.4 in 1.5 mL centrifuge tubes. Samples were vortexed briefly and centrifuged for 15 minutes at 6,000 $\times$ g. Supernatants were carefully separated from precipitated materials into a conditioned (equal ratio water/

ethanol mixture) MWCO 3kDa micron centrifugal filter (Millipore, Billerica, MA) to buffer exchange with HP MLAC binding buffer (25 mM Tris, 0.5 M NaCl, 1 mM MnCl₂, 1 mM CaCl₂ and 0.05% sodium azide, pH 7.4). This was followed by BCA protein assay and immediate storage at -80°C until glycoproteomic analysis. Samples were thawed not more than twice for each experiment.

2.3. Sample fractionation and glycoproteins enrichment

200 µg each of mucinous subtypes and non-mucinous subtypes were fractionated and enriched for glycoproteins on an HP-MLAC platform as previously described¹². Briefly, high abundance proteins (IgG and albumin) were depleted followed by the HP-MLAC glycoprotein enrichment and sample desalting all on-line; preventing sample degradation and sample losses. The on-line sample preparation was performed on a 2D HPLC System (Shimadzu, Columbia, MD) equipped with three valves to allow for switching among the three columns connected serially. 0.1 M glycine, pH 2.5 was the elution buffer for POROS protein G and Capture select^(R) ligand albumin-IgG columns. For the HP-MLAC glycoprotein affinity column, 0.1 M acetic acid pH 2.5 was used to elute M-LAC bound proteins. POROS 50 R1 reversed-phase PEEK column was used to desalt unbound and bound fractions collected. For desalting, solvent A was made of 0.1% trifluoroacetic acid in milli-q water and that of B was composed of 0.1% trifluoroacetic acid in acetonitrile. 70% solvent B step gradient was used to elute bound proteins on the POROS R1 column. Unbound protein fractions (no specificity for M-LAC column) were collected separately from M-LAC bound fractions. All fractions collected were speed vacuum to dryness.

2.4. 1D SDS-PAGE analysis followed by in-gel Trypsin digestion

20 µg of each fraction was loaded on 10% Mini-PROTEAN® TGX™ Precast Gels and run on a Mini-PROTEAN Tetra cell system (Bio-Rad Laboratories, Hercules, CA) for 40 minutes at a constant voltage of 200V. Gels were stained after electrophoresis following manufacturer's instructions with SimplyBlue™ SafeStain (Invitrogen, Carlsbad, CA). Each fraction was excised into 10 separate bands and processed in a low binding centrifuge tubes as earlier described¹⁵, with some modifications. Briefly, each band was cut into 1mm × 1mm × 1mm pieces and destained as follows; 500 µL of 50 mM ammonium bicarbonate pH 8.0, 500 µL acetonitrile in that order for up to 5 cycles. In each instance, addition of the appropriate solvent, 5 minutes vortexing, short spin on a bench-top centrifuge and aspiration of solvent to waste were performed in a sequence. After the aspiration of acetonitrile in the last cycle, dehydrated gel pieces were dried in a speed vacuum, reduced at 56°C for 30 minutes with 0.5 M dithiothreitol (DTT) in 50 mM ammonium bicarbonate pH 8.0 to a final concentration of 25 mM, and alkylated with 0.5 M iodoacetamide in 50 mM ammonium bicarbonate pH 8.0 to a final concentration of 50 mM at room temperature for 30 minutes in the dark. Gel pieces were washed with 500 µL 50 mM ammonium bicarbonate 3 times and dehydrated with 500 µL acetonitrile. 50 µL of freshly prepared 0.04 µg/µL of sequencing grade trypsin (Promega, Madison, WI) prepared in 25mM ammonium bicarbonate containing 3% acetonitrile (v/v) pH 8.0 was added to the dehydrated gel pieces. Low binding centrifuge tubes containing gel pieces were incubated on ice for 45 minutes to allow for trypsin enzyme absorption. Excess trypsin was aspirated after 45 minutes and enough digestion buffer (25 mM ammonium bicarbonate containing 3% acetonitrile (v/v) pH 8.0)

was added to cover gel pieces and incubated at 37°C for approximately 20 hr. Peptides were extracted into labeled low binding centrifuge tubes in the following fashion; 200 µL of 5% formic acid/ 50% acetonitrile twice and 200 µL 100% acetonitrile once. Aspirated supernatants were pooled for each individual excised band and dried completely in a speed vacuum. Dried peptides were constituted in 20 µL 0.1% formic acid in HPLC grade water prior to nano-LC-MS/MS analysis.

2.5. Liquid chromatography mass spectrometry analysis

In-gel tryptic digested peptides were subjected to nano-LC-MS/MS analysis on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an Ultimate 3000 HPLC (LC Packings-Dionex, Marlton, NJ, USA). A 75µm reversed phase C18 column was packed in-house using a slurry of 5-µ particle, 200-Å pore size Magic C18 stationary phase (Michrom Bioresources, Auburn, CA) into a 150mm × 75mm i.d. capillary column (New Objective, Woburn, MA). The peptides were loaded onto the C18 capillary column using an auto sampler and desalted for 30 mins at a flow-rate of 300nl/min in isocratic mode. The following constituted the mobile phase buffers; mobile phase A consisted of 0.1% formic acid in HPLC grade water and mobile phase B consisted of 0.1% formic acid in HPLC grade acetonitrile. Flow rate was automatically adjusted to 200 nL/min for gradient separation of desalted peptides using the following gradient method; 5% B buffer to 40% B buffer for 80 mins; 40% B buffer to 90% B buffer for 15 mins; 90% B buffer to 2% B buffer for 5mins. The total time for nano-LC-MS/MS analysis was 140 mins. The mass spectrometer was operated in a data dependent fashion; eight most abundant precursor ions selected from the MS spectrum were MS/MS fragmented via collision induced dissociation (CID) using an isolation width of 3.0 mass unit. A resolution, R of 60,000 was used to acquire each full MS scan over a mass range of 400-2000 m/z. Dynamic exclusion was set with 1 repeating counts (repeat duration of 30s, exclusion list of 150, and exclusion duration of 30s, exclusion mass width 0.55 m/z low and 1.55 m/z high).

2.6. Data processing and Bioinformatics

Protein identification was obtained by searching the generated MS/MS spectra against Uniprot annotated human database (release 2012_1; 34,157 entries) using Thermo Fisher Proteome Discoverer 1.3 suite (Thermo Electron Corp, San Jose, CA). Both MASCOT (Matrix Science) version 2.3 and SEQUEST (Thermo Electron) algorithms present in the Thermo Fischer Proteome Discoverer suite were used simultaneously to perform the search. This approach was utilized because it is shown that combined algorithms increases the number of proteins identified and reduces false positive identification in shot gun proteomics.¹⁶⁻¹⁷ Confidence in identification was enhanced by applying the reverse database with a false discovery rate (FDR) targeted at 1% at the peptide level. The following are the other search parameters used; 2 maximum missed cleavages, enzyme was set at full trypsin, carbamidomethylation on cysteine as static modification, precursor ion mass tolerance and fragment ion mass tolerance were set at 5ppm and 0.8Da respectively. PANTHER (Protein ANALysis THrough Evolutionary Relationships) database (<http://pantherdb.org/>) was used to determine gene ontology (GO) Molecular Function. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, Gene A La Cart (provided by www.genecards.org) version 3.10 was used to assign gene symbols as well as protein disease relationships. Novoseek

(www.novoseek.com), a biomedical text mining tool was used to acquire disease relevance to gene scores.

2.7. Western Blot Analysis

10% Mini-PROTEAN® TGX™ Precast Gels (Bio-Rad Laboratories, Hercules, CA) and Tris/Glycine SDS running buffer were used to resolve 5ug of proteins for each sample. Blotting was performed on Bio-Rad's Transfer-Blot Turbo transfer system for 10mins. The primary antibody (rabbit polyclonal, 1:500) was purchased from Novus Biologicals, Littleton, CO. Goat anti Rabbit HRP (System Biosciences, Mountain View, CA) was used as the secondary antibody. Immuno-detected proteins were visualized using ECL Western Blotting reagents (GE Healthcare), and images were captured with a Fluorchem SP system (Alpha Innotech, Santa Clara, CA).

3. RESULTS AND DISCUSSION

3.1. Analytical Strategy

To achieve good depth of glycoproteome analysis of the cyst fluid samples we used a 4 step chromatographic strategy, namely 1) abundant protein depletion, 2) lectin based chromatography 3) 1D SDS-PAGE separation and 4) capillary LC reversed phase separation of the mixture of tryptic peptides generated by a trypsin digest of the resulting 10 gel bands. In previous publications we have described our lectin platform which is based on a physical admixture of 3 different lectins that are individually bound to different chromatographic beads (termed multi-lectin affinity chromatography or M-LAC). We showed that such an admixture gave approximately a 10-fold increase relative to total proteins in binding efficiency that was based on the glycoside clustering effect.¹⁸⁻¹⁹ Furthermore the ratio of an individual glycoprotein that was partitioned into the bound or unbound fraction was reproducible and reflected the types of glycan present in an individual protein. Also, changes in the ratios in a global protein analysis can yield information on changes in glycan motifs in glycoproteins resulting from diseases such as cancer and diabetes.²⁰⁻²¹ As mentioned in the introduction, pancreatic cancer is both difficult to diagnose and appropriate clinical samples are difficult to obtain. We initiated the study with a sample set of 20 individual cyst samples and performed protein concentration, 1D SDS-PAGE and mass spectrometric analysis. It was clear from this phase of the investigation that a study with meaningful depth of analysis of samples with such rare availability would not permit replicate analyses. Patients' samples were grouped (Table 1) into mucinous and non mucinous cyst subtypes based on clinical diagnosis and classification as previously discussed⁶ and then pooled (mucinous cyst fluid=10 and non mucinous cyst fluid=10 samples in the two groups respectively). While cyst fluid offers an ideal proximal fluid for the observation of 'proteins of interest' that change in either concentration or amount of glycosylation motifs in cancer, it is a sample that is highly variable and of limited amount. For this phase of the investigation we found that pools were necessary to permit an effective depletion of contaminating blood proteins and to achieve good depth of glycoproteome analysis of the bound and unbound M-LAC fractions. Another advantage of pooling is that individual patient variability would be minimized and that resulting 'proteins of interest' would likely to be of more general applicability.²² We also had access to sufficient patient numbers to permit the preparation of

a second pool of mucinous and non-mucinous cyst samples and thus to performed a second analysis (second sample set). Since the second pool contained a different set of individuals, the analysis would generate a second list of 'proteins of interest' that could be compared with the first analysis and be used to explore the differences in cyst fluid in mucinous and non-mucinous pancreatic cancer.

3.2. Glycoproteome and non-glycoproteome platform

As described earlier, our HP-MLAC based platform (see supplementary info fig 1) used two protein (2P, albumin and IgG) depletion followed by multi-lectin affinity chromatography (M-LAC) glycoproteins enrichment and then 1D SDS-PAGE fractionation of eluted M-LAC bound and unbound proteins. The decision to perform 2P depletion was based on a preliminary analysis of a set of pancreatic cyst fluid samples that showed highly variable and in many cases a significant contamination of albumin and IgG's (supplementary info figure 2) as well as demonstrating that the depth of resulting proteomic analysis was improved (data not shown). We selected lectins for the M-LAC column with broad specificity for glycans typically present in cancer and showed no non-specific binding in the M-LAC platform, namely Concanavalin A (ConA), Artocarpus integrifolia (Jacalin) and Wheat germ agglutinin (WGA) lectins (high mannose, glycans; galactose and O-linked glycans; and *N*-acetylglucosamine and sialic acid glycans respectively). As shown later (Table 2A), some proteins in the bound fraction are not known to be glycosylated (e.g. carboxypeptidases (CPA)) which are previously attributed to either non-specific binding or binding to a glycosylated carrier protein.¹² In a pilot study we showed that the 1D SDS-PAGE separation step followed by in-gel trypsin digestion of 10 gel bands improved depth of proteins identification by more than 2 fold compared to in-solution trypsin digestion of protein fractions with no prior proteins or peptides level separation (data not shown).

3.3. Summary of glycoproteome and non-glycoproteome data

To increase confidence and positive proteins identification, we used a 5ppm peptide mass tolerance and false discovery rate (FDR) targeted at 1% during MS/MS database search analysis. A total average of 520 unique proteins was identified between the two sample sets of which mucinous subtypes had 230 proteins and 290 proteins belonging to non-mucinous subtypes. A detailed table of the number of identified and quantified proteins can be found in supplementary table S1, S4-S7. A four way Venn diagram was used to understand the distribution of unique and shared proteins for individual M-LAC fractions in each sample set (Figure 1). Percent ratio for unbound vs. bound in mucinous and non mucinous fractions (79.7 vs. 72.2 and 65.5 vs. 66.0 respectively) is similar in both sample set 1 and sample set 2 (supplementary table 1). This observation is supported by molecular function characterization obtained from PANTHER, a web-based gene ontology classification system²³ showing protein abundance of each molecular function as their relative percentage. The corresponding unbound mucinous and non-mucinous fractions showed similar results (data not shown). The similarity in protein molecular function for the M-LAC fractions presents a similar picture between mucinous and non-mucinous (M-LAC bound and unbound fractions). As expected, we observed variability in identified proteins (427 unique proteins in analysis set 1 vs. 298 unique proteins in analysis set 2) due to two possibilities;

1) different levels of albumin contamination; and 2) individual differences between sample set 1 and sample set 2 (supplementary figure S2A & 2B).

Overall, we observed significant differences in protein levels between mucinous and non-mucinous fractions of lipid (fatty acid) metabolism proteins such as Glutathione S-transferase (Table 2B); energy associated proteins such as those involved in glucose metabolism and ATP synthesis e.g. Pyruvate kinase isozymes M1/M2 (Table 2A and 2B) as well as stress related proteins such as heat shock cognate 71kDa, heat shock 70kDa protein 1-like, and heat shock 70kDa protein 6 (Table 2B).

3.4. Quantitation of glycoproteins in different analysis set and selection of potential protein targets of interest

Based on BCA total protein assay, equal amounts of cyst fluid pools (200µg) were depleted and separated on the M-LAC column. Protein recoveries were determined using BCA total protein assay. Total recovery range of 84%-85% of yield of the starting material after M-LAC fractionation was recorded. This result is consistent with earlier published work.²⁴ Proteins were quantitated for differential expression by spectral counts (total MS/MS spectra collected for each protein), a label free semi-quantitative method developed for shot-gun proteomics and proteins with 2 unique peptides were used in quantitation. In selected cases, spectral counts were confirmed by peak area measurements of extracted ion chromatogram (EIC) as previously described¹⁴ as well as manual inspection of MS/MS spectra. Briefly, mass spectrometry data was first normalized by the reference ratio calculated from total spectral counts of mucinous and non-mucinous M-LAC fractions for each protein. Relative protein abundance changes were based on the ratios of spectral counts of mucinous and non mucinous M-LAC bound and unbound fractions after normalization. The algorithm used for reference ratio calculations and protein abundance calculations were previously published.¹³ Proteins with spectral count ratio ≥ 5.0 or ≤ 0.3 were assigned as differentially expressed. In cases where no peptides ("0") were observed for a protein, "1" was added for meaningful ratio calculations.

The development of a disease marker is preceded by a comprehensive discovery program such as in this study of pancreatic cyst fluid. In this type of study, it is premature to discuss biomarker candidates which will require a study of larger number of individual patient samples. Our goal is to determine 'proteins of interest' (see Table 2A & 2B) for the M-LAC bound and unbound subproteomes based on four criteria; 1) high protein abundance (spectral count), 2) presence of protein in both sample sets analyzed, 3) significance to pancreatic cancer and other related diseases, and 4) spectral count ratio changes (higher or lower protein levels) as grouped in table 2A and 2B. Of particular interest are proteins that are mostly expressed in high levels in mucinous fractions including pancreatic cancer related proteins (pancreatic lipases, amylases, mucin 2, calcium-activated chloride channel regulator 1, catalase, bile salt-activated lipase, carboxypeptidase etc.), and energy metabolism associated proteins (hexokinase-1, phosphoglycerate kinase 1, pyruvate kinase isozymes M1/M2, etc.). Also, chaperone function proteins (heat shock family proteins) were observed with higher protein levels (table 2B). It is interesting to note that periostin, an extracellular matrix protein and a low abundance protein implicated in pancreatic cancer and other

cancers²⁵⁻²⁹ was found in lower levels in mucinous cyst fluid. As part of a pilot pre-validation study, POSTN was investigated by western blots (see below). Some proteins in the M-LAC bound fraction potentially are not glycosylated (such as Alpha-amylase 2B, Carboxypeptidase A2, and Pyruvate kinase isozymes M1/M2), but may be associated with a glycoprotein due to their function (protein binding and metal ion binding) or bind to the column as a result of non-specific binding of lectins.

Unlike the analysis of molecular function classification for total proteins identified in the bound and unbound M-LAC fractions, the list of 'proteins of interest' showed significant differences between the fractions in specific proteins (Figure 2) when molecular function classification analysis was performed. Proteins with antioxidant activity, for example, catalase (CAT) and Glutathione S-transferase A1 were observed to be highly enriched in the unbound sub-proteome (Figure 2 (A)). On the other hand, proteins with receptor binding activity and proteins involved in structural functions were exclusively detected in the M-LAC bound glycoproteome (Figure 2 (B)). Bile salt-activated lipase (CEL) long and short isoforms were observed to be differentiated by our M-LAC column (see supplementary table 2A & 2B and supplementary figure 3 (MS/MS for long isoform diagnostic peptide)), with higher levels observed in M-LAC bound fraction of mucinous cyst fluid. CEL is a heavily O-linked glycosylated digestive enzyme³⁰ implicated in diabetes and pancreatic exocrine dysfunction.³² Previous studies of CEL^{31,32} shows O-linked glycosylation sites found at the C-tail fragment which binds to Jac lectin, a constituent of our M-LAC column. It is possible that due to glycosylation changes, CEL long isoform selectively binds to the M-LAC column while the short isoform flows through the M-LAC column. Recent studies by Mann and colleagues reported an observation of significant high levels of CEL in fucose enriched samples, suggesting CEL as a potential glyco-biomarker target in pancreatic cyst fluid.¹⁰ Our findings of M-LAC's ability to enrich CEL long isoform in mucinous cyst fluid subtypes contributes to recent observation,¹⁰ therefore M-LAC may contribute significantly to future structural and glycosylation alteration studies of CEL in pancreatic cyst fluid.

Novoseek, a data mining tool from Genecards human gene database (Version 3.10), was used to investigate the relationship to pancreatic disease of selected potential target proteins. The analysis revealed that, the majority of our selected targets are involved in a variety of pancreatic diseases such as pancreatic cancer and pancreatitis (Table 2A & 2B), consistent with earlier reports.³³ In supplementary document table 3, details of specific pancreatic disease association with each target protein is listed. Since we do not know which of these targets will be secreted into the blood, the next phase of study will determine which of these disease related targets are measurable in blood as such observations will depend on the biology of the disease and the huge dynamic range of plasma.

3.5. Pathway and network interaction analysis of potential targets of interest

Recent studies^{34,35} have suggested that a perturbed module (pathway or biological process) is a better disease marker than one or more biomarkers and thus discovery studies should have such a focus. Furthermore, it has been shown that prioritizing cancer associated protein observations together with protein-protein interaction network information can add further discrimination.³⁶ To determine the pathway of interest, we used the pathway listed in

Genecards with the greatest concentration of our 'proteins of interest' which resulted in the selection of the pancreatic secretion pathway (KEGG, Figure 3). The relevance of this pathway was shown by the clustering of proteins involved in the pancreatic secretion process that were present in either M-LAC bound fraction and/or unbound fractions (highlighted in red, blue and yellow in figure 4). Examples include: alpha-amylase 2B (AMY2B), bile salt activated lipase (CEL), pancreatic triacylglycerol lipase (PNLIP), pancreatic lipase-related protein 2 (PNLIPR2), and carboxypeptidases (CPA, CPB), and in the unbound fraction observed proteins include; chymotrypsinogen B1 (CTRB1) and pancreatic lipase-related (PNLIPR1) (denoted (^a) and (^b) in Figure 4 respectively). The presence of higher levels of proteins in mucinous cyst fluid (blue and red circles in figure 4), their clustering and the fact that some target proteins physically associate (CEL, PNLIP, and PNLIPR2), as shown in the interactome data (figure 4), is an indication of the significance of these identified proteins in pancreatic cyst biology.

3.6. Chromosome gene mapping analysis of potential targets of interest

It has been observed that protein coding genes which express proteins that have related functions, such as tissue location, cellular compartment, common pathways or interactants are more likely to be co-located in the same chromosomal region.^{37,38} In such situations co-expression can be facilitated by mechanisms such as cis-activation or suppression (gene slicing) of specific chromosomal regions. In this context, we submitted selected proteins of interest to the Gene A La Cart (provided by www.genecards.com, uploaded to Gene A La Cart for analysis in August, 2011) to obtain genomic location and Ensemble cytobands. It is of interest that some of the proteins identified in the M-LAC fractions with protein level change in mucinous vs. non-mucinous are located in specific chromosomal regions (table 3), e.g. chromosome 1, band p21 and 36 (amylase and elastase); chromosome 10, band 25&26 (lipases). Also catalase (CAT) is located in the same genomic region as the important cancer associated genes MUC2 and 6 (chromosome 11, bands p13 and p15). Further, several of the enzyme groups such as PNLIPR1, and PNLIP, are co-located in the same chromosome region and potentially co-expressed in cancer, which is consistent with other studies such as the ERBB2 amplicon.^{39,40} It is interesting to note that SERPINF1, YWHAE and NPEPPS genes (proteins involved in; proteolytic events of cell growth, various signaling pathways, and inhibitor of angiogenesis respectively) are located in the same band (p13 and p21) on chromosome 17 as the important apoptotic gene p53. Future studies will explore the potential role of co-expressions in the development of pancreatic cancer and the potential role of these genes since this is the first report on such observations.

3.7. Validation of Periostin

As part of a pilot pre-validation of differentially expressed target proteins identified, we analyzed periostin (POSTN) by western blot using six samples; three mucinous and three non-mucinous pancreatic cyst fluid subtypes. POSTN was chosen for investigation from the protein target list because of its potential significance in pancreatic cancer progression and other related diseases⁴¹ as well as its over expression in breast cancer⁴² which is in contrast to our observations. POSTN was immuno-precipitated using anti-periostin antibody and the blot detected by anti-periostin antibody measuring total POSTN protein levels at a molecular weight of 89kDa. Significant lower POSTN levels was found in mucinous cyst subtypes;

intraductal papillary mucinous neoplasm (IPMN) and mucinous cyst neoplasm (MCN), compared to non-mucinous cyst subtypes; serous cystadenoma (figure 5A). An 8 fold relative increase of POSTN in non-mucinous cyst fluid was observed (figure 5B) after densitometry quantification correlating spectral count observations of cyst pools. Periostin (POSTN) was previously reported to be a potential biomarker target in pancreatic cancer,⁴³ however its role in pancreatic cyst is not known. Although POSTN is known to be glycosylated with one *N*-linked site, its glycosylation patterns with pancreatic cancer has not been studied. Future studies are required to understand cancer related glycosylation changes of periostin and its overall role in pancreatic cyst fluid.

CONCLUSION

We have demonstrated that the high performance multi-lectin affinity chromatography (HP-MLAC) platform successfully allows the enrichment and characterization of glycoproteins which are present at different levels in mucinous and non-mucinous cyst fluid subtypes. Of particular interest is the observation of increased amounts in the mucinous vs. non mucinous cyst fluid of proteins with strong cancer associations such as; mucin 2 (MUC2), mucin 6 (MUC6), carboxypeptidase A2 (CPA), and hexokinase-1 (HK1), proteins with energy metabolism functions as well as various pancreatic enzymes. The significance of the identified proteins was further shown with pathway analysis, interaction and chromosomal location investigations. In addition, bile salt-activated lipase (CEL) long isoform was significantly enriched in M-LAC fractions and differentially bound to M-LAC column thus indicating possible glycosylation changes in mucinous cyst fluid. Since these observations are based on cyst fluid pools of 20 individuals (10 mucinous cyst subtypes and 10 non-mucinous cyst subtypes), we believe that the true picture will be confirmed by analyzing individual samples that constitute the pools. In this discovery study, glycoproteomics is used to explore differentially expressed proteins to differentiate mucinous cyst fluid from non-mucinous cyst fluid. In future studies we plan to; 1) investigate selected 'proteins of interest' using anti-body lectin sandwich microarray platform^{9,44} in a larger cohort; 2) potentially measure 'proteins of interest' in a readily available diagnostic fluid i.e. plasma; and 3) explore chromosome gene co-expressions and co-locations of 'proteins of interest' and their importance to pancreatic cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Early Detection Research Network grant # 1U01CA152653 01 from the National Cancer Institute (NCI). We are grateful to Somak Ray for his help in bioinformatics data analysis. This is contribution number 1043 from the Barnett Institute.

ABBREVIATIONS

M-LAC multiple lectin affinity chromatography

DTT	dithiothreitol
FDR	false discovery rate
ConA	concanavalin A
WGA	wheat germ agglutinin
Jac	Jacalin
1D-SDS-PAGE	one dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis

REFERENCES

1. Cleary SP, Gryfe R, Guindi M, Greig P, Smith L, Mackenzie R, Strasberg S, Hanna S, Taylor B, Langer B, Gallinger S. Prognostic factors in resected pancreatic adenocarcinoma: analysis of actual 5-year survivors. *J Am Coll Surg*. 2004; 198(5):722–31. [PubMed: 15110805]
2. Snow P. Pancreatic Cyst Fluid Analysis—A Review. *J Gastrointest Liver Dis*. 2011; 20(2):175–180. [PubMed: 21725515]
3. Cuoghi A, Farina A, Z'graggen K, Dumonceau J-M, Tomasi A, Hochstrasser DF, Genevay M, Lescuyer P, Frossard J-L. Role of Proteomics to Differentiate between Benign and Potentially Malignant Pancreatic Cysts. *Journal of Proteome Research*. 2011; 10(5):2664–2670. [PubMed: 21425880]
4. Kwon RS, Simeone DM. The Use of Protein-Based Biomarkers for the Diagnosis of Cystic Tumors of the Pancreas. *International Journal of Proteomics*. 2011; 2011:1–9.
5. Hutchins GF, Draganov PV. Cystic neoplasms of the pancreas: a diagnostic challenge. *World J Gastroenterol*. 2009; 15(1):48–54. [PubMed: 19115467]
6. Jeurnink SM, Vleggaar FP, Siersema PD. Overview of the clinical problem: facts and current issues of mucinous cystic neoplasms of the pancreas. *Dig Liver Dis*. 2008; 40(11):837–46. [PubMed: 18499541]
7. Matthaei H, Schulick RD, Hruban RH, Maitra A. Cystic precursors to invasive pancreatic cancer. *Nat Rev Gastroenterol Hepatol*. 2011; 8(3):141–50. [PubMed: 21383670]
8. Hara T, Kawashima H, Ishigooka M, Kashiya M, Takanashi S, Yamazaki S, Hosokawa Y. Mucinous cystic tumors of the pancreas. *Surg Today*. 2002; 32(11):965–9. [PubMed: 12444432]
9. Haab BB, Porter A, Yue T, Li L, Scheiman J, Anderson MA, Barnes D, Schmidt CM, Feng Z, Simeone DM. Glycosylation Variants of Mucins and CEACAMs As Candidate Biomarkers for the Diagnosis of Pancreatic Cystic Neoplasms. *Annals of Surgery*. 2010; 251(5):937–945. [PubMed: 20395854]
10. Mann BF, Goetz JA, House MG, Schmidt CM, Novotny MV. Glycomic and proteomic profiling of pancreatic cyst fluids identifies hyperfucosylated lactosamines on the N-linked glycans of overexpressed glycoproteins. *Mol Cell Proteomics*. 2012; 11(7) M111 015792.
11. Cao Z, Maupin K, Curnutte B, Fallon B, Feasley CL, Brouhard E, Kwon R, West CM, Cunningham J, Brand R, Castelli P, Crippa S, Feng Z, Allen P, Simeone DM, Haab BB. Specific glycoforms of MUC5AC and endorepellin accurately distinguish mucinous from non-mucinous pancreatic cysts. *Mol Cell Proteomics*. 2013
12. Kullolli M, Hancock WS, Hincapie M. Automated platform for fractionation of human plasma glycoproteome in clinical proteomics. *Anal Chem*. 2010; 82(1):115–20. [PubMed: 19957969]
13. Zeng Z, Hincapie M, Haab BB, Hanash S, Pitteri SJ, Kluck S, Hogan JM, Kennedy J, Hancock WS. The development of an integrated platform to identify breast cancer glycoproteome changes in human serum. *J Chromatogr A*. 2010; 1217(19):3307–15. [PubMed: 19782370]
14. Plavina T, Wakshull E, Hancock WS, Hincapie M. Combination of abundant protein depletion and multi-lectin affinity chromatography (M-LAC) for plasma protein biomarker discovery. *J Proteome Res*. 2007; 6(2):662–71. [PubMed: 17269723]

15. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc.* 2006; 1(6):2856–60. [PubMed: 17406544]
16. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics.* 2010; 73(11): 2092–123. [PubMed: 20816881]
17. Kapp EA, Schutz F, Connolly LM, Chakel JA, Meza JE, Miller CA, Fenyo D, Eng JK, Adkins JN, Omenn GS, Simpson RJ. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics.* 2005; 5(13):3475–90. [PubMed: 16047398]
18. Yang Z, Hancock WS. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *Journal of chromatography. A.* 2004; 1053(1-2):79–88. [PubMed: 15543974]
19. Yang Z, Hancock WS. Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *J Chromatogr A.* 2005; 1070(1-2):57–64. [PubMed: 15861788]
20. Kyselova Z, Mechref Y, Kang P, Goetz JA, Dobrolecki LE, Sledge GW, Schnaper L, Hickey RJ, Malkas LH, Novotny MV. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem.* 2008; 54(7):1166–75. [PubMed: 18487288]
21. Abd Hamid UM, Royle L, Saldova R, Radcliffe CM, Harvey DJ, Storr SJ, Pardo M, Antrobus R, Chapman CJ, Zitzmann N, Robertson JF, Dwek RA, Rudd PM. A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. *Glycobiology.* 2008; 18(12):1105–18. [PubMed: 18818422]
22. Batruch I, Lecker I, Kagedan D, Smith CR, Mullen BJ, Grober E, Lo KC, Diamandis EP, Jarvi KA. Proteomic analysis of seminal plasma from normal volunteers and post-vasectomy patients identifies over 2000 proteins and candidate biomarkers of the urogenital system. *J Proteome Res.* 2011; 10(3):941–53. [PubMed: 21142078]
23. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013; 41:D377–86. Database issue. [PubMed: 23193289]
24. Kullolli M, Hancock WS, Hincapie M. Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins. *J Sep Sci.* 2008; 31(14):2733–9. [PubMed: 18693314]
25. Kanno A, Satoh K, Masamune A, Hirota M, Kimura K, Umino J, Hamada S, Satoh A, Egawa S, Motoi F, Unno M, Shimosegawa T. Periostin, secreted from stromal cells, has biphasic effect on cell migration and correlates with the epithelial to mesenchymal transition of human pancreatic cancer cells. *Int J Cancer.* 2008; 122(12):2707–18. [PubMed: 18381746]
26. Baril P, Gangeswaran R, Mahon PC, Caulee K, Kocher HM, Harada T, Zhu M, Kalthoff H, Crnogorac-Jurcevic T, Lemoine NR. Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the beta4 integrin and the PI3k pathway. *Oncogene.* 2007; 26(14):2082–94. [PubMed: 17043657]
27. Kudo Y, Siriwardena BS, Hatano H, Ogawa I, Takata T. Periostin: novel diagnostic and therapeutic target for cancer. *Histol Histopathol.* 2007; 22(10):1167–74. [PubMed: 17616943]
28. Zhang Y, Zhang G, Li J, Tao Q, Tang W. The expression analysis of periostin in human breast cancer. *J Surg Res.* 2010; 160(1):102–6. [PubMed: 19524268]
29. Ouyang G, Liu M, Ruan K, Song G, Mao Y, Bao S. Upregulated expression of periostin by hypoxia in non-small-cell lung cancer cells promotes cell survival via the Akt/PKB pathway. *Cancer Lett.* 2009; 281(2):213–9. [PubMed: 19328625]
30. Hui DY, Hayakawa K, Oizumi J. Lipoamidase activity in normal and mutagenized pancreatic cholesterol esterase (bile salt-stimulated lipase). *Biochem J.* 1993; 291(Pt 1):65–9. [PubMed: 8471055]
31. Raeder H, Johansson S, Holm PI, Haldorsen IS, Mas E, Sbarra V, Nermoen I, Eide SA, Grevle L, Bjorkhaug L, Sagen JV, Aksnes L, Sovik O, Lombardo D, Molven A, Njolstad PR. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat Genet.* 2006; 38(1):54–62. [PubMed: 16369531]

32. Mechref Y, Chen P, Novotny MV. Structural characterization of the N-linked oligosaccharides in bile salt-stimulated lipase originated from human breast milk. *Glycobiology*. 1999; 9(3):227–34. [PubMed: 10024660]
33. Ke E, Patel BB, Liu T, Li XM, Haluszka O, Hoffman JP, Ehya H, Young NA, Watson JC, Weinberg DS, Nguyen MT, Cohen SJ, Meropol NJ, Litwin S, Tokar JL, Yeung AT. Proteomic analyses of pancreatic cyst fluids. *Pancreas*. 2009; 38(2):e33–42. [PubMed: 19136908]
34. Chuang HY, Rassenti L, Salcedo M, Licon K, Kohlmann A, Haferlach T, Foa R, Ideker T, Kipps TJ. Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood*. 2012; 120(13):2639–49. [PubMed: 22837534]
35. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol*. 2008; 4(11):e1000217. [PubMed: 18989396]
36. Lessner DJ, et al. An unconventional pathway for reduction of CO₂ to methane in CO-grown *Methanosarcina acetivorans* revealed by proteomics. *Proc Natl Acad Sci U S A*. 2006; 103(47):17921–6. [PubMed: 17101988]
37. Zhang EY, Cristofanilli M, Robertson F, Reuben JM, Mu Z, Beavis RC, Im H, Snyder M, Hofree M, Ideker T, Omenn GS, Fanayan S, Jeong SK, Paik YK, Zhang AF, Wu SL, Hancock WS. Genome Wide Proteomics of ERBB2 and EGFR and Other Oncogenic Pathways in Inflammatory Breast Cancer. *J Proteome Res*. 2013; 12(6):2805–17. [PubMed: 23647160]
38. Wu C, Zhu J, Zhang X. Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. *BMC Bioinformatics*. 2012; 13:182. [PubMed: 22838965]
39. Kauraniemi P, Barlund M, Monni O, Kallioniemi A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res*. 2001; 61(22):8235–40. [PubMed: 11719455]
40. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol*. 2003; 4(9):117. [PubMed: 12952525]
41. Ke E, Patel BB, Liu T, Li X-M, Haluszka O, Hoffman JP, Ehya H, Young NA, Watson JC, Weinberg DS, Nguyen MT, Cohen SJ, Meropol NJ, Litwin S, Tokar JL, Yeung AT. Proteomic Analyses of Pancreatic Cyst Fluids. *Pancreas*. 2009; 38(2):e33–e42. [PubMed: 19136908]
42. Abbott KL, Aoki K, Lim JM, Porterfield M, Johnson R, O'Regan RM, Wells L, Tiemeyer M, Pierce M. Targeted glycoproteomic identification of biomarkers for human breast carcinoma. *J Proteome Res*. 2008; 7(4):1470–80. [PubMed: 18271524]
43. Kosanam H, Makawita S, Judd B, Newman A, Diamandis EP. Mining the malignant ascites proteome for pancreatic cancer biomarkers. *Proteomics*. 2011; 11(23):4551–8. [PubMed: 21932441]
44. Orzechowski R, Hamelinck D, Li L, Gliwa E, vanBrocklin M, Marrero JA, Vande Woude GF, Feng Z, Brand R, Haab BB. Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer Res*. 2005; 65(23):11193–202. [PubMed: 16322270]

Figure 1. Four way Venn diagram showing distribution of proteins identified in unbound and M-LAC bound fractions of mucinous and non-mucinous subtypes after glycoproteomic analysis in sample set 1 and sample set 2

MucFT = mucinous unbound fraction

MucBD = mucinoc M-LAC bound fraction

Non-mucFT = non-mucinous unbound fraction

Non-mucBD = non-mucinous M-LAC bound fraction

Figure 2. Molecular functional characterization of differentially expressed proteins in M-LAC fractionation. (A) Unbound fraction and (B) M-LAC bound fraction. PANTHER was used for classification

Percentage (%) = relative abundance of each molecular function

Red boxes = protein molecular function differentiating bound and unbound subproteomes

Figure 3. Annotation of pancreatic secretion KEGG signaling pathway. The pathway was generated from <http://www.genome.jp/kegg/pathway.html>

Genes in the pathway are circled as follows; **Blue**: up-regulated potential target proteins in mucinous subtypes; **Red**: up-regulated proteins in mucinous subtypes with less fold change; **Yellow**: proteins identified in glycoproteomics with no change in relative abundance based on spectral count.

The following annotations were used; ^a Proteins observed more in M-LAC bound fraction; ^b Proteins observed more in unbound fraction; ^c Proteins observed equally in both unbound and bound fractions.

Figure 4. A string network interaction of CEL, PNLIP, and PNLIPRP1 genes significantly enriched in glycoproteomics and observed in pancreatic secretion pathway
 Red circle = target genes interacting

A **B**

Figure 5. Pre-validation of Periostin as a potential biomarker target through SDS-PAGE western blot analysis. (A) Six pancreatic cyst fluid samples were subjected to western blotting to validate the relative abundance of Periostin in mucinous and non-mucinous cyst fluid subtypes. (B) Densitometry quantitation of Periostin levels

IPMN= intraductal papillary mucinous neoplasm

MCN= mucinous cystic neoplasm

SC= serous cystadenoma

Table 1
Pancreatic cyst fluid samples for glycoproteomics analysis

Cyst fluid sample type and specimen notes	Clinical classification	Number of samples	
		Set 1	Set 2
Low IPMN	Mucinous	2	2
Low grade MUCN	Mucinous	0	1
Intraductal papillary mucinous neoplasms	Mucinous	1	1
Pancreatic ductal adenocarcinoma arising in intraductal papillary neoplasms	Mucinous	2	1
carcinoma-in situ (CIS) IPMN	Mucinous	2	2
Mucinous cystic neoplasm with low grade dysplasia	Mucinous	1	1
High grade dysplasia	Mucinous	1	0
Adenoma	Mucinous	1	2
Tissue origin: serous cystadenoma	Non-mucinous	1	1
Pseudocyst	Non-mucinous	2	1
Serous cystadenoma: Two lymph nodes negative for malignancy	Non-mucinous	1	2
Serous cystadenoma	Non-mucinous	3	2
Serous cystadenoma, macrocystic variant. Two benign lymph nodes	Non-mucinous	1	2
Benign retention cyst	Non-mucinous	2	1
Cystic lesion of the pancreas with fibrous walls but without epithelium	Non-mucinous	0	1

IPMN = intraductal papillary mucinous neoplasm

MUCN = mucinous neoplasm

Table 2A

Mucinous vs. non-mucinous proteins identified in M-LAC bound fraction with relative abundance changes (spectral counts)

Protein name ^a	Gene name ^b	Spectral Counts (Total peptides hits)						Pancreatic disease association ^b	
		Analysis set 1 (fractions)		Analysis set 2 (fractions)					
		Mucinous ^c	Non-mucinous ^c	Mucinous ^c	Non-mucinous ^c	Pancreatic cancer	Pancreatitis		
Bile salt-activated lipase	CEL	41	3	55	5	yes(3.9) ^d	yes(18.2)		
Carboxypeptidase A2	CPA2	31	4	12	2	yes(24.7)	yes(44.1)		
Mucin-6	MUC6	23	2	21	3	yes(50.0)	yes(2.4)		
Pancreatic triacylglycerol lipase	PNLIP	21	3	13	2	yes(1.0)	yes(49.0)		
Pyruvate kinase isozymes M1/M2	PKM	14	2	12	1 ^f	yes(22.3)	yes(6.3)		
Periostin	POSTN	3	20	4	23	yes(35.2)	yes(11.6)		
Alpha-amylase 2B	AMY2B	43	4	46	2	no	yes(22.9)		
Calcium-activated chloride channel regulator 1	CLCA1	78	3	23	1 ^f	no	yes*		
Carbonic anhydrase 1	CA1	2	ND	16	2	no	yes(17.3)		
Pancreatic lipase-related protein 2	PNLIPRP2	9	1 ^f	17	1 ^f	no	yes(25.5)		
Pancreatic alpha-amylase	AMY2A	15	3	25	5	no	yes(44.4)		
Leucine-rich alpha-2-glycoprotein	LRG1	5	31	4	19	no	yes(1.3)		
Phosphoglycerate kinase 1	PGK1	10	2	11	2	no	no		
Metalloproteinase inhibitor 1	TIMP1	2	15	2	11	no	no		
Alpha-1-acid glycoprotein 2	ORM2	37	5	21	3	no	no		
Interstitial collagenase	MMP1	36	6	42	3	no	no		
Fibronectin	FN1	63	2	77	2	no	no		
Tetranectin	CLEC3B	2	19	2	14	no	no		
Vitamin D-binding protein	GC	25	5	23	2	no	no		
Protein S100-A12	S100A12	6	ND	2	11	no	no		

Table 2B

Mucinous vs. non-mucinous proteins identified in unbound fraction with relative abundance changes (spectral counts)

Spectral Counts (Total peptides hits)									
Protein name ^a	Gene name ^b	Analysis set 1 (fractions)		Analysis set 2 (fractions)		Pancreatic disease association ^b			
		Mucinous ^c	Non-mucinous ^c	Mucinous ^c	Non-mucinous ^c	Pancreatic cancer	Pancreatitis		
Adenylyl cyclase-associated protein 1	CAP1	14	2	11	2	yes(8.5) ^d	yes(8.5)		
Hexokinase-1	HK1	57	4	49	5	yes(7.3)	yes(7.3)		
Isoform Short of Bile salt-activated lipase	CEL	8	1 ^f	11	2	yes(3.9)	yes(18.2)		
Mucin-2	MUC2	35	5	55	7	yes(49.0)	yes(10.6)		
Pancreatic triacylglycerol lipase	PNLIP	33	3	44	6	yes(1.0)	yes(49.0)		
Pyruvate kinase isozymes M1/M2	PKM	17	2	13	2	yes(22.3)	yes(6.3)		
Annexin A5	ANXA5	4	43	1 ^f	13	yes(36.0)	yes(9.4)		
Calcium-activated chloride channel regulator 1	CLCA1	59	2	27	1 ^f	no	yes [*]		
Kininogen-1	KNG1	18	2	9	ND	no	yes(3.7)		
Glycine amidinotransferase, mitochondrial	GATM	4	37	5	41	no	yes(8.2)		
Cadherin-17	CDH17	36	6	ND	2	no	yes(3.96)		
Pancreatic lipase-related protein 1	PNLIPRP1	17	1 ^f	15	ND	no	no		
Glutathione S-transferase A1	GSTA1	2	20	3	29	no	no		
Bifunctional purine biosynthesis protein PURH	ATIC	31	3	32	4	no	no		
Puromycin-sensitive aminopeptidase	NPEPPS	30	6	35	7	no	no		
14-3-3 protein zeta/delta	YWHAZ	5	31	7	37	no	no		
Vinculin	VCL	5	45	2	21	no	no		
Heat shock cognate 71 kDa protein	HSPA8	49	ND	41	ND	no	no		
Leukotriene A-4 hydrolase	LTA4H	41	ND	19	1 ^f	no	no		
14-3-3 protein epsilon	YWHAE	3	19	2	11	no	no		
Aldo-keto reductase family 1 member B10	AKR1B10	ND	25	1 ^f	11	no	no		
Aspartate aminotransferase, mitochondrial	GOT2	29	3	25	3	no	no		

Protein name ^a	Gene name ^b	Spectral Counts (Total peptides hits)				Pancreatic disease association ^b	
		Analysis set 1 (fractions)		Analysis set 2 (fractions)			
		Mucinous ^c	Non-mucinous ^c	Mucinous ^c	Non-mucinous ^c	Pancreatic cancer	Pancreatitis
Annexin A10	ANXA10	1 ^f	8	ND	23	no	no
Catalase	CAT	13	2	24	2	no	no

Red box highlights proteins expressed at lower levels in mucinous fractions

ND; not identified

^a protein names are from Swiss-Prot.

^b Gene name, and pancreatic disease association information are from Genecards

^c Relative expression levels based on spectral count.

^f In cases of no peptide identification, 1 replaced 0 for easier ratio calculations

^d Novoseek score (−log(P-Val)), based on literature mining information on the significance of disease to gene

* Genes without Novoseek score

Table 3

Chromosome gene analysis of some 'proteins of interest'

Protein Name	Gene Name	Chromosome #	Cytogenetic band
Pancreatic alpha-amylase	AMY2A↑	1	1p21.1
Alpha-amylase 2B	AMY2B↑	1	1p21.1
Calcium-activated chloride channel regulator 1	CLCA1↑	1	1p22.3
Heat shock 70 kDa protein 6	HSPA6	1	1q23.3
Isoform 2 of Adenylyl cyclase-associated protein 1	CAP1	1	1p34.2
Chymotrypsin-like elastase family member 3B	CELA3B	1	1p36.12
Chymotrypsin-like elastase family member 3A	CELA3A	1	1p36.12
Pancreatic lipase-related protein 1	PNLIPRP1↑	10	10q25.3
Pancreatic triacylglycerol lipase	PNLIP↑	10	10q26.1
Hexokinase-1	HK1↑	10	10q22.1
Catalase	CAT↑	11	11p13
Mucin-2	MUC2↑	11	11p15.5
Mucin-6	MUC6↑	11	11p15.5
Cellular tumor antigen p53	P53	17	17p 13.1
Pigment epithelium-derived factor	SERPINF1↓	17	17p 13.3
14-3-3 protein epsilon	YWHAE↓	17	17p 13.3
Puromycin-sensitive aminopeptidase	NPEPPS↓	17	17q21.32

Green highlights = co-expressed genes observed in the pancreatic secretion pathway

Red highlights = co-located genes with oncogene p53 (yellow highlight)

Arrows denotes relative proteins expression levels in mucinous vs. non-mucinous fractions; ↑(higher levels in mucinous), ↓ (lower levels in mucinous)

^a protein names are from Swiss-Prot^b Gene symbols, chromosome number, and cytogenic band are from Genecards