

Short Gel, Long Gradient Liquid Chromatography Tandem Mass Spectrometry to Investigate the Urine Proteome of Chronic Pancreatitis

Joao A. Paulo^{1,2,3,4}, Vivek Kadiyala^{3,4}, Scott Brizard^{3,4}, Peter A. Banks^{3,4}, Darwin L. Conwell^{3,4,#,*} and Hanno Steen^{1,2,4,#,*}

¹Departments of Pathology Children's Hospital Boston, Boston, MA, USA

²Proteomics Center at Children's Hospital Boston, Boston, MA, USA

³Center for Pancreatic Disease, Division of Gastroenterology, Hepatology and Endoscopy,

⁴Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston, MA, USA

Abstract: Chronic pancreatitis (CP) is currently diagnosed using invasive endoscopic and imaging techniques. However, urine can be collected safely and noninvasively and as such may offer a superior alternative to current techniques of CP diagnosis. We use mass spectrometry-based methods to discover proteins which are exclusive to or differentially abundant in urine of chronic pancreatitis patients.

We have performed a comparative quantitative proteomic analysis of urine collected from 5 healthy controls, 5 severe CP patients, and 5 patients of a mixed cohort with clinical representation typical of patients referred for CP, but not diagnosed with the disease. Proteins from urine were fractionated *via* SDS-PAGE and digested in-gel with trypsin prior to reversed-phase liquid chromatography in-line with a mass spectrometer. ProteinPilot software and the QSPEC algorithm identified proteins and determined statistically significant differences between cohorts.

We identified over 600 proteins from urine, of which several hundred were either exclusive to or differ quantitatively in severe CP patients. Members of the cathepsin protein family were of significantly higher abundance in the severe CP cohort. In addition, we have identified a core set of 50 proteins in all 15 samples, 25 of which showed no significant difference among the cohorts.

The differentially abundant proteins in severe CP patients represent an initial set of targets for directed proteomics experiments for further validation studies. However, larger matched cohorts will be required to determine if these differences have statistically significant diagnostic potential.

Key words: pancreas, cathepsin, urine, chronic pancreatitis, biomarkers, clinical proteomics.

BACKGROUND

Chronic pancreatitis (CP) is characterized by chronic inflammation and progressive fibrosis, clinically manifested as intense pain, and pancreatic exocrine and endocrine insufficiency. Annually in the United States, exocrine pancreatic disorders affect over one million persons and cost approximately \$3 billion. During the past decade, diseases of the exocrine pancreas have resulted in 277,000 hospitalizations and 475,000 ambulatory care visits per year [1]. Clinical diagnosis of chronic pancreatitis is primarily based on morphological and functional findings. Complications, such as

bleeding and fistulae formation, preclude pancreatic biopsy for histologic diagnosis. The non-histological "surrogate" gold standard is pancreas function testing [2], which can diagnose only moderate to late stage chronic pancreatitis with currently irreversible tissue damage and fibrosis [3]. Radiologic imaging is also limited to diagnosing later-stage disease as objective morphologic changes are only associated with moderate to advanced disease. Evaluation of chronic pancreatitis using proteomics would revolutionize diagnosis and potentially lead to novel therapies designed to retard, or modify disease progression, before irreversible organ damage and dysfunction become apparent.

We [4] and others [5] have shown that pathologic changes in the pancreas may be reflected in the pancreatic fluid proteome. Pancreatic fluid is the proximal body fluid of the pancreas and is a reservoir of locally secreted biomolecules from surrounding cells that are likely to include specific markers of disease. However, pancreatic fluid collection is a relatively invasive procedure requiring endoscopy, sedation and/or anesthesia, and is performed at limited specialized centers.

*Address correspondence to these authors at the Department of Pathology Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115, USA; Tel: +1-617-919-2629; Fax: +1-617-730-0168; E-mail: Hanno.steen@childrens.harvard.edu and Harvard Medical School, Department of Cell Biology, 240 Longwood Ave, Boston, MA 02115, USA; Tel: (401) 368-2925; Fax: (617) 264-5277; E-mail: joao_paulo@post.harvard.edu

#these authors contributed equally to this work

A diagnosis of chronic pancreatitis via proteomic methodologies, using a non-invasively collected fluid, such as urine, would thus be preferred.

The analyses of proximal and systemic fluids have inherent advantages in medical testing. In contrast to proximal fluids, systemic fluids, such as urine, represent the entire body and thus provide a snapshot of the whole organism under a given set of systemic conditions. As such, analysis of systemic fluids may be confounded by proteins representing the normal or abnormal physiology of other organs and/or their pathophysiologic response to the disease of interest [6], limiting the utility of systemic fluids in investigating pathogenic and pathophysiologic mechanisms. However, systemic fluid analysis would be appropriate, clinically, for assays targeted toward diagnosis and staging. Compared to other body fluids, urine is readily available in large volumes, non-invasively collected, and the same individual can be sampled repeatedly and safely. Compared to pancreatic fluid and blood, urine has been underutilized in mass spectrometry-based proteomic analyses related to chronic pancreatitis.

Mass spectrometry-based proteomic techniques can identify disease-specific proteins in urine to complement established diagnostic methods and improve diagnosis of chronic pancreatitis [4d]. Traditionally, pre-fractionation of proteins and/or peptides is performed prior to LC-MS/MS analysis. Greater sample complexity requires increased fractionation to achieve maximum proteome coverage [7]. However, the proteome of urine is less complex compared to larger proteomes, such as that of whole cell lysates [8]. As such, we have used a modified strategy to minimize sample processing by altering two parameters: 1) the time of SDS-PAGE fractionation and 2) the length of the column and gradient during liquid chromatography. First, during a typical minigel SDS-PAGE fractionation, proteins usually are allowed to migrate 7-8 cm and the gel lane is divided into 6-24 fractions, each to be processed separately (termed GeLC-MS/MS). However, we performed SDS-PAGE fractionation on the TCA-extracted protein sample so that the proteins

migrated only 1 cm into the gel, and subsequently subjected to in-gel tryptic digestion. Second, following peptide extraction from gel slices, traditional LC-MS/MS methods then analyze the eluent from a 12-15 cm long reversed-phase capillary column over a 1 hr period, however, we used a 45 cm-long column and eluted over a 3 hr period. This strategy simplifies sample processing while allowing efficient peptide fractionation upstream of mass spectrometric analysis.

We present the first comparative mass spectrometry-based proteomic analysis of urine in the study of chronic pancreatitis. The aim of this study is to determine the feasibility of identifying chronic pancreatitis-specific proteins in urine using a variation of well-established GeLC-MS/MS strategy employing minimal pre-fractionation.

MATERIALS AND METHODS

Study Population

This protocol was approved by the Institutional Review Board at Brigham and Women's Hospital (BWH) (IRB # 2007-P-002480/1 for severe chronic pancreatitis (sCP) and mixed disease (MD) cohorts, and 2011P-001391 for healthy control volunteers). The experimental cohorts were assembled from adult patients referred to the Center for Pancreatic Disease (BWH) for evaluation of abdominal pain, and the healthy control (HC) adult volunteers were recruited from the general population. Patient characteristics are listed in Table 1. The diagnosis of chronic pancreatitis was determined using the M-ANNHEIM (Multiple risk factors, Alcohol, Nicotine, Nutrition, Hereditary factors, Efferent duct factors, Immunological factors, and Miscellaneous and metabolic factors) classification [9]. The M-ANNHEIM classification is a standardized system designed to classify chronic pancreatitis according to etiology, clinical staging, and severity of the disease [9]. In addition, for chronic pancreatitis patients, the mean peak bicarbonate is listed. A threshold of 80 meq/L was two standard deviations below the mean and thus, ≤ 80 meq/l was considered abnormal [10].

Table 1. Cohort Characteristics

	Healthy Controls (HC)					Severe Chronic Pancreatitis (sCP)					Mixed Diagnoses (MD)				
	HC1	HC2	HC3	HC4	HC5	sCP1	sCP2	sCP3	sCP4	sCP5	MD1	MD2	MD3	MD4	MD5
Gender	M	M	M	M	M	F	M	M	M	M	M	F	M	F	M
Race	W	B	W	W	W	W	W	W	W	W	W	H	W	W	W
Age (years)	n/a	n/a	n/a	n/a	n/a	53	44	78	68	33	74	42	69	39	54
MANNHEIM diagnosis	n/a	n/a	n/a	n/a	n/a	Def CP	Def CP	Def CP	Def CP	Def CP	AP	AP	Prob CP	CAP	CAP
Atlanta AP severity	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Severe	Mild	n/a	n/a	n/a
[HCO₃], (meq/L)	n/a	n/a	n/a	n/a	n/a	38	50	36	39	60	n/a	n/a	n/a	n/a	100
Smoking	n/a	n/a	n/a	n/a	n/a	2	0	1	2	1	0	0	0	0	2
Alcohol	n/a	n/a	n/a	n/a	n/a	1	2	1	2	1	2	0	0	2	2

AP, acute pancreatitis; n/a, not available; M, male; F, female; W, white; H, Hispanic; B, black; CP, chronic pancreatitis; Def CP, definite chronic pancreatitis; Prob CP, probable chronic pancreatitis; CAP, chronic abdominal pain. Smoking and Alcohol: 0, never smoking; 1, past; 2, current.

Materials

SeeBluePlus2 Pre-Stained standard (LC5925), LDS (lithium dodecyl sulfate) sample buffer (NP0008), NuPAGE 4-12% Bis-Tris polyacrylamide gels (NP0335), Simply Blue Coomassie stain (LC0665), and MES-SDS (2-(N-morpholino) ethanesulfonic acid-sodium dodecyl sulfate) electrophoresis buffer (NP002) were from Invitrogen (Carlsbad, CA). Sequencing-grade modified trypsin (V5111) was obtained from Promega (Madison, WI). Other reagents and solvents were from Sigma-Aldrich and Burdick & Jackson, respectively.

Experimental Workflow

The general experimental workflow (Fig. 1) was as follows: 1) Urine samples were collected (clean-catch); 2) proteins were extracted from the sample by trichloroacetic acid (TCA) precipitation; 3) proteins were reduced with dithiothreitol (DTT) and alkylated with acrylamide; 4) proteins were electrophoresed into an SDS-PAGE gel for a length of 1 cm; 5) proteins from each sample were in-gel tryptically digested in a single gel slice; 6) peptides were fractionated using a 45 cm capillary reversed-phase (C18) column with a 3 hr linear gradient, in-line with an LTQ-FT Ultra mass spectrometer; 7) bioinformatics processing was performed using ProteinPilot for database searching, QSPEC [11] to determine statistically significant differences in spectral counts between cohorts, and DAVID [12] to perform gene ontology analysis.

Sample Preparation

Urine collection and storage. Urine from the individuals' second void of the day following an overnight fast was collected using the clean-catch method. We did not control for hydration/drinking, however, we were able to equate protein amounts loaded onto the gel and subsequently analyzed via mass spectrometry. The urinary protein concentration was determined using the bicinchoninic acid (BCA) colorimetric assay. The urine was promptly aliquoted and stored at -80°C until analysis.

TCA precipitation of urine. A total of 125 μL of ice-cold 100% TCA was added to 1 mL of each urine sample. The samples were vortexed for 5 seconds and incubated at 4°C for 2 hours. The samples were then centrifuged at $20,000\times g$ at 4°C for 30 minutes and the supernatant carefully aspirated and discarded. One milliliter of 100% ice-cold acetone was added to the pellets, which were then briefly vortexed and incubated at -20°C for 1 hour. The samples were centrifuged again at $20,000\times g$ at 4°C for 30 minutes, and the pellets were washed twice with 1 mL of 100% ice-cold acetone. The final pellets were allowed to air dry at room temperature.

SDS-PAGE Analysis

LDS sample buffer, with 50mM DTT, was added to each sample to achieve a 1X concentration. The samples were incubated at 56°C for 30 min and allowed to cool. The samples were then alkylated with 1% acrylamide and incubated

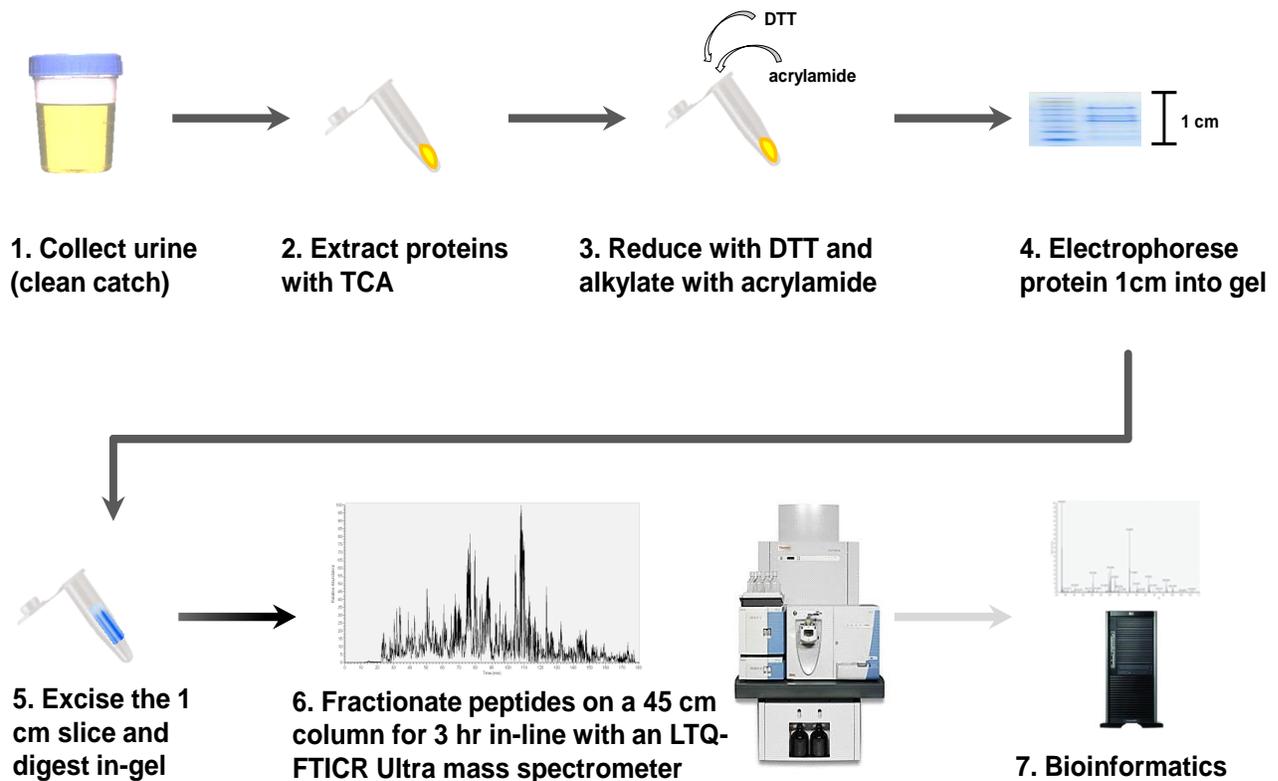


Fig. (1). Experimental workflow. The method was as follows: 1) Urine was collected via clean-catch. 2) Proteins were extracted with trichloroacetic acid (TCA) precipitation. 3) Proteins were reduced with dithiothreitol (DTT) and alkylated with acrylamide. 4) Proteins were separated via SDS-PAGE for a length of 1 cm. 5) The single band was excised and in-gel tryptically digested. 6) Peptides were fractionated on a 45 cm column with a 3 hr gradient in-line with an LTQ-FTICR Ultra mass spectrometer. 7) Bioinformatics tools, such as ProteinPilot, QSPEC, and DAVID, were used to identify, quantify, and characterize the collected data.

for 30 min at 23°C. Proteins (approximately 100 µg) were fractionated by SDS-PAGE at 150 volts in MES buffer for a total migration length of 1 cm. Gels were rinsed in deionized water for 10 min, fixed in 45% methanol/45% water/10% acetic acid for 30 min, stained with SimplyBlue Coomassie for 1 hour, and destained overnight in deionized water.

GeLC-MS/MS Analysis

Proteins in the entire 1 cm gel section underwent standard in-gel tryptic digestion using established techniques [13]. Peptides were extracted from each gel section and fractionated by a nanoflow reversed-phase ultra-high pressure liquid chromatography system (nanoLC, Eksigent) in-line with a linear trap quadrupole-Fourier transform ion cyclotron mass spectrometer (LTQ-FT Ultra, Thermo Scientific). The reversed-phase liquid chromatography columns (45 cm long x 150 µm ID, New Objective, Woburn, MA) were packed in-house (Magic C18, 5 µm, 100 Å, Michrom Bio Resources). Samples were analyzed with a 3-hr linear gradient (5-29% acetonitrile with 0.2% formic acid), and data were acquired in a data-dependent manner using 6 MS/MS scans for every full scan spectrum. RAW files are available for download upon request.

Bioinformatics and Data Analysis

Mascot generic files (“mgf”) were generated using MSconvert software [14]. All MS data generated from the gel sections were searched against the UniProt database (downloaded November 11, 2011) using the Paragon Algorithm [15] integrated into the ProteinPilot search engine (v. 4; ABSciex). Search parameters were set as follows: sample type, identification; Cys alkylation, acrylamide; Instrument, Orbitrap/FT (1-3 ppm), LTQ MS/MS; special factors, gel-based ID; ID focus, Biological Modifications; database, UniProt database; detection protein threshold, 95.0%; and search effort, thorough ID. Species specificity was *Homo sapiens*. We used Proteomics System Performance Evaluation Pipeline (PSPEP) to determine the cutoff that would result in a 1% global false discover rate at the protein level. We allowed onto our lists proteins identified by a single peptide, as has been recommended previously [16].

Venn diagrams

The VENNY on-line Venn diagram plotter was used to obtain lists of unique and common proteins among the cohorts investigated [17].

QSPEC Spectral Counting Analysis

We performed relative protein quantification using a label-free technique, spectral counting, which compared the number of identified peptide-spectra matches for the same protein across multiple data sets. To search for differences in the protein profile among data sets, spectral counts were normalized based on the total spectral counts, as outlined previously [18]. We performed significance analysis of our normalized spectral count data using QSPEC, a published algorithm for determining the statistical significance of differences in spectral count data [11]. This algorithm used Bayes Factors, in lieu of p-values, as measures of evidential strength, more specifically as indicators of statistical significance of two competing statistical models where the likelihood of each competing model is averaged over all possible parameter values by numerical integration [19]. This Bayesian model calculates posterior probabilities for the differential abundance for thousands of proteins in a given experiment. By convention, a Bayes factor greater than 10 suggested strong evidence that a particular protein was differentially expressed between the two cohorts; thus a value of 10 was used as our significance threshold [20].

DAVID Functional Annotation Bioinformatics Microarray Analysis

We used the DAVID (Database for Annotation, Visualization and Integrated Discovery) Bioinformatics Database (<http://david.abcc.ncifcrf.gov/>) interface [12] to classify proteins according to the Panther Protein Class [21].

RESULTS

Several Hundred Proteins were Identified in Each Specimen

SDS-PAGE analysis (Fig. 2) showed, as expected, that simple visualization of gels is insufficient in determining proteomic differences between cohorts. Our subsequent mass spectrometry-based analysis of the gel lanes identified a total of 609 unique proteins in 15 samples (Supplemental Table 1). Of interest is the redundancy of each protein over all samples, as determined by grouping proteins in bins according to the number of samples in which they were identified. In Supplemental Fig. (1A) we illustrated the percentage of proteins identified in all 15 samples down to those only identified in a single sample. For example, the diamond pattern between 67% and 72% indicates that 5% of proteins were identified in 8 samples, and so it follows that 67% were iden-

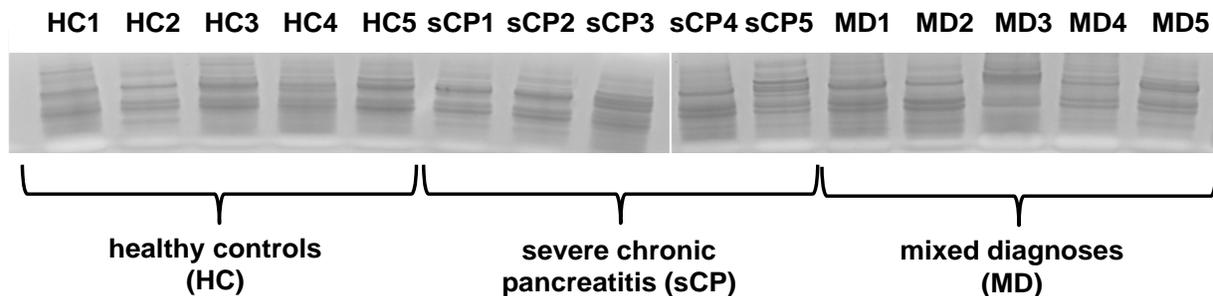


Fig. (2). SDS-PAGE 1 cm separation gel image. Fifteen samples - 5 each for healthy controls (HC), severe chronic pancreatitis (sCP), and mixed diagnoses (MD) - were analyzed via SDS-PAGE.

tified in less than 8 samples and 28% were identified in more than 8 samples. Approximately 25% of the proteins were identified in 10 or more samples and greater than 45% of the proteins were identified in 5 or more samples.

Similarly, when examining each cohort individually (Supplemental Fig. 1B), the protein redundancy distribution within each cohort was similar to that of the whole sample set; approximately 30 - 45% of the identified proteins were identified in more than half the samples (i.e., ≥ 3) and of the 5 bins, the largest proportion of proteins (20- 40%) were identified in only one sample.

Differentially-expressed Proteins were Identified in Severe Chronic Pancreatitis (sCP) Versus Healthy Controls (HC)

We identified a total of 430 unique proteins in healthy control individuals and 475 unique proteins in individuals with sCP. Comparing proteins identified in both cohorts, 127 proteins (Supplemental Table 2) were exclusive to the sCP cohort and 82 (Supplemental Table 3) were exclusive to the HC cohort (Fig. 3A). The majority of the proteins exclusive to a specific cohort were identified in only 1 or 2 individual samples, making it difficult to assess the potential of these proteins as markers of pancreatic disease. However, we did identify several proteins in 3 or more of the 5 samples. In the sCP cohort, Ig lambda-7 chain C region, carbonic anhydrase 1 and neutrophil gelatinase-associated lipocalin were identified in 3 of the 5 samples. In the HC cohort, 6-phosphogluconolactonase appeared in all 5 HC samples; ephrin type-B receptor, phosphatidylcholine-sterol acyltransferase, and N(G),N(G)-dimethylarginine dimethylaminohydrolase appeared in 4 samples; and Poliovirus receptor-related protein 4 and CD276 appeared in 3 of the 5 samples. These exclu-

sive proteins, particularly those appearing in 4 or 5 samples, represent promising targets for further investigation.

Additionally, a total of 348 proteins were found to be common to both cohorts. Using the QSPEC algorithm, we compared the spectral counts of these common proteins to identify those present in statistically different abundances between the two cohorts. We identified 49 proteins (Supplemental Table 4) significantly more abundant in the sCP cohort and 52 proteins (Supplemental Table 5) significantly more abundant in the HC cohort (Fig. 3B). Combining these two sets of proteins (those exclusive to and those statistically more abundant in a particular cohort) we determined that 176 (i.e., 127+49) and 134 (i.e., 82+52) proteins were differentially abundant in the sCP and HC cohorts, respectively, which we refer to as sCP' and HC'.

Current limitation of mass spectrometry technology precludes absolute proteome identification. As such, it is possible that those proteins which were not identified in a cohort were not sampled by the mass spectrometer, but may be present at lower abundance and so statistical significance for such proteins cannot be ascertained. It follows that proteins identified readily across all samples, and for which statistical significance can be determined (Supplemental Tables 4 and 5), are far more informative than those exclusive proteins (Supplemental Tables 2 and 3).

A third (mixed diagnosis, MD) cohort was analyzed to control for pathologic changes in the urine proteome that may not be directly due to chronic pancreatitis. A cohort composed of patients with chronic abdominal pain, acute pancreatitis, and a participant with "borderline" chronic pancreatitis was included in our analysis as these individuals represent the intended use population, i.e., clinical patients commonly presenting to a tertiary care pancreas center for

Table 2. Proteins Exclusive to sCP' Cohort (i.e., Exclusive to or of Higher Abundance in sCP Cohort Versus HC Cohort and Not Identified in the MD Cohort) and Appearing in ≥ 3 of the 5 Samples

Protein Names	Entry	Σ Spectral Counts	# of Samples	Spectral Counts				
				sCP1	sCP2	sCP3	sCP4	sCP5
Ig lambda-7 chain C region	A0M8Q6	1052	3	0	243	0	255	554
Carbonic anhydrase 1	P00915	86	3	0	30	0	11	45
Neutrophil gelatinase-associated lipocalin	P80188	85	3	29	0	18	38	0

Table 3. Proteins Exclusive to HC' Cohort (i.e., Exclusive to or of Higher Abundance in HC Cohort Versus CP Cohort) and Not Identified in the MD Cohort) and Appearing in ≥ 3 of the 5 Samples

Protein Names	Entry	Σ Spectral Counts	# of Samples	Spectral Counts				
				HC1	HC2	HC3	HC4	HC5
6-phosphogluconolactonase	O95336	171	5	63	24	30	15	39
Ephrin type-B receptor 4	P54760	189	4	0	53	53	38	45
Phosphatidylcholine-sterol acyltransferase	P04180	97	4	0	29	12	24	32
N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	O95865	89	4	42	21	14	12	0
Poliovirus receptor-related protein 4	Q96NY8	109	3	0	0	26	28	55
CD276 antigen	Q5ZPR3	19	3	0	0	5	7	7

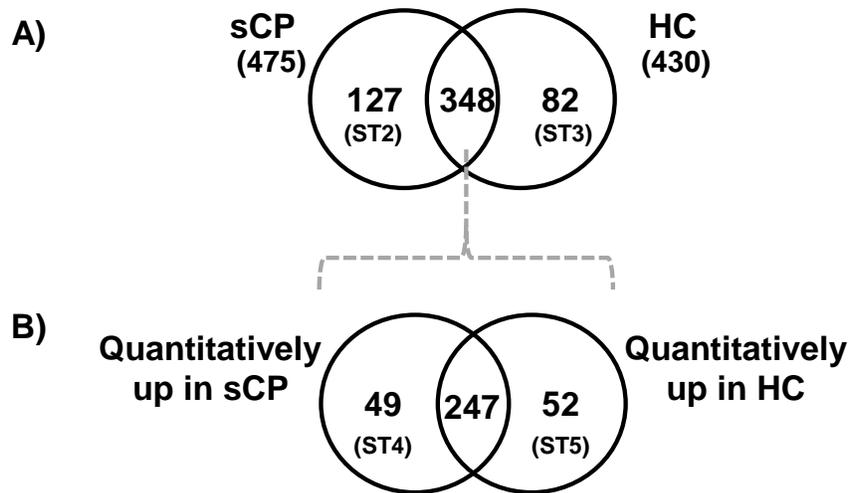


Fig. (3). Venn diagrams comparing identified proteins for the sCP and HC cohorts. **A)** Proteins identified in sCP and HC cohorts were compared qualitatively to identify proteins exclusive to a particular cohort. **B)** The 348 common proteins were analyzed further using the QSPEC algorithm to identify statistically significant differences between the two cohorts. ST#, Supplementary Table #.

evaluation of abdominal pain and potential CP.

We identified a total of 386 unique proteins in this mixed diagnosis (MD) cohort. We compared the sCP' protein subset to proteins identified in the MD cohort, as well as the HC' protein subset to proteins identified in the MD cohort. In each of these comparisons, overlapping proteins (not exclusive to either the sCP' or the HC' cohort) will be eliminated from our list of chronic pancreatitis-specific proteins. Compared to the MD cohort, 104 of the 176 proteins in the sCP' subset remained exclusive to the sCP cohort, while 67 of the 134 proteins in the HC' subset remained exclusive to the HC cohort (Fig. 4A). As in the comparison between sCP and HC, many of these "exclusive" proteins were identified only in 1 or 2 individuals. Such would not be ideal candidates for initial follow-up studies, as those identified in ≥ 3 samples would fit this role better. We have listed these "exclusive" proteins appearing in ≥ 3 samples for sCP' in Table 2 and HC' in Table 3.

We also investigated if differences existed quantitatively in the proteins determined to be in common between the sCP' and MD cohorts, as well as the HC' and MD cohorts. Using spectral counting and QSPEC, we compared the MD cohort to the HC' and sCP' for significant differences in the abundance of common proteins (Fig. 4B). Those proteins of similar abundances likely represent confounding pathophysiological processes, and not necessarily chronic pancreatitis and will be eliminated as potential diagnostic chronic pancreatitis-specific urinary proteins. First, we examined the comparison of the HC' and MD cohorts (Fig. 4B left). This analysis showed that of the 72 proteins common to both sCP' and MD, 1 was of greater abundance in the MD cohort, 42 were of similar abundance. More importantly, 29 proteins (Table 4) were of significantly greater abundance in the sCP' subset, representing proteins that are potentially specific to chronic pancreatitis. Second, we examined the comparison of the HC' and MD cohorts (Fig. 4B right). Of the 67 proteins common to both HC' and MD, 18 proteins (Table 5) were of significantly higher abundance in HC' subset and thus remained potentially useful targets for diagnostic testing. Additionally, 1 protein was more abundant in the MD cohort and 48 did not differ in abundance. Thus, a total

of 49 proteins were eliminated from those specific to urine of individuals with a normal pancreas. Therefore, our filtering strategy reduced potential confounders by eliminating many proteomic changes representing a general disease state, but not necessarily chronic pancreatitis.

We Identified a Core Set of Proteins Common to All 15 Samples

Of the 609 proteins identified, 50 were present in all 15 samples (Supplemental Table 6). These 50 proteins are identified using our minimal sample processing GeLC-MS/MS workflow. Differentially expressed proteins readily identified with mass spectrometry are advantageous over those requiring large-scale preparation, so long as sensitivity and specificity are not compromised. Comparisons of the cohorts using QSPEC analysis identified 25 proteins common to all cohorts with no significant difference in abundance. These proteins include beta-2-glycoprotein, fibronectin, insulin-like growth factor-binding protein 7, osteopontin, prothrombin, and serotransferrin. If validated in larger cohorts, several of these proteins may serve as candidates for normalization controls, which may be integrated into future, directed mass spectrometry workflows.

Gene ontology analysis revealed different classes of proteins which were identified exclusively or in higher abundance in the sCP and HC cohorts. Using the DAVID interface, we categorized the potential chronic pancreatitis-specific proteins according to their functional class (Fig. 5). We compared proteins in the sCP' subset (which included the 104 proteins that were exclusive to the sCP cohort, and the 29 proteins that were of higher abundance after filtering out those which overlap with the mixed diagnosis (MD) cohort) and those in the HC' subset (which included the 67 proteins that were exclusive to the HC cohort and the 18 proteins that were of higher abundance, after filtering out those which overlap with the mixed diagnosis (MD) cohort). We considered only categories represented by $\geq 5\%$ of at least one cohort. We compared functional classes of protein identified in the sCP and HC samples to each other, rather than comparing sCP to the entire human proteome, as urine is an overrep-

resentation of secreted proteins with unique functions, and such a comparison to the global proteome would represent a biased measure of significance. Various protein classes were identified; of note, proteases and other enzyme classes (hydrolase, oxidoreductase, and transferase) were more commonly identified in the sCP cohort. Proteins in this class included both lysosomal and secreted proteins: lysosomal protective protein, dipeptidyl peptidase 1 and 4, plasma glutamate carboxypeptidase, pepsin, and members of the cathepsin family (B, D, and Z). The HC cohort urine was rich in cell adhesion molecules, defense/immunity-related proteins, extracellular matrix (ECM) proteins, and receptors. Such proteins include Ig gamma-1 chain C region, endothelial

protein C receptor, apolipoprotein E, and poliovirus receptor-related protein 4. The investigation of these protein classes will allow us to target, not only single proteins, but classes, with protein microarrays, ELISA, western blotting and/or targeted mass spectrometry experiments for downstream validation.

DISCUSSION

We used a modified GeLC-MS/MS strategy to identify potential chronic pancreatitis-specific protein targets in urine for subsequent validation. Using this strategy, we identified a total of 609 proteins in the urine of 15 individuals. We qualitatively and quantitatively compared the proteins identified

Table 4. Proteins (n=29) of Higher Abundance in sCP' Cohort (i.e., Exclusive to or of Higher Abundance in sCP Cohort Versus HC Cohort) Compared to MD Cohort

Protein names	Entry	Σ sCP Spectral Count	# sCP Samples	Compare with HC		Compare with MD	
				Bayes Factor	Fold Change	Bayes Factor	Fold Change
Aggrecan core protein	P16112	110	3	16.83	5.24	33.50	4.69
Agrin	O00468	518	5	6.7E+06	6.97	7.9E+09	16.07
Beta-galactosidase	P16278	465	5	4.7E+04	6.47	3.4E+04	8.90
Cadherin-13	P55290	274	5	49.51	1.89	2.1E+04	5.45
Cathepsin B	P07858	198	5	291.77	4.79	3.6E+04	10.15
Cathepsin D	P07339	760	5	4497.61	2.14	2.9E+05	2.91
Cathepsin Z	Q9UBR2	88	4	36.48	9.32	152.97	11.79
Collagen alpha-3	P12111	1237	5	1.0E+04	4.61	1750.90	4.38
Complement decay-accelerating factor	P08174	328	5	17.75	2.43	13.38	2.12
Dipeptidyl peptidase 1	P53634	221	5	1083.51	6.00	2.8E+06	13.38
Dipeptidyl peptidase 4	P27487	275	5	1109.62	5.86	6.4E+05	11.01
Ganglioside GM2 activator	P17900	182	4	33.65	7.47	66.35	5.01
Ig gamma-4 chain C region	P01861	984	4	93.74	7.42	183.78	8.11
Ig kappa chain C region	P01834	8024	5	559.32	1.94	1253.38	2.06
Ig kappa chain V-I region HK102	P01602	212	5	27.35	2.24	3730.57	10.08
Ig kappa chain V-IV region JI	P06313	418	5	3469.47	6.90	20.25	2.45
Ig lambda chain V-I region WAH	P04208	254	5	1241.14	6.15	10.55	2.87
Ig lambda chain V-III region LOI	P80748	211	5	138.80	2.98	2693.09	6.49
Ig lambda-1 chain C regions	P0CG04	1758	5	2.1E+04	6.27	2.8E+04	5.75
Ig lambda-2 chain C regions	P0CG05	2285	5	7.4E+04	5.28	34.07	2.30
Lymphatic vessel endothelial hyaluronic acid receptor 1	Q9Y5Y7	191	4	114.41	12.67	28.45	6.81
Lysosomal protective protein	P10619	325	5	1242.00	3.80	29.77	2.82
Lysozyme C	P61626	108	5	6.2E+04	15.67	2.5E+05	13.78
Pepsin A	P00790	408	5	688.91	6.04	261.65	2.10
Plasma glutamate carboxypeptidase	Q9Y646	180	5	3.2E+04	7.92	1.8E+04	10.94
Prostaglandin-H2 D-isomerase	P41222	3798	5	2.2E+04	2.51	4747.91	2.09
Protein AMBP	P02760	6600	5	8.6E+06	2.96	3.1E+04	1.86
SH3 domain-binding glutamic acid-rich-like protein 3	Q9H299	129	4	91.15	5.69	87.76	15.04
Triosephosphateisomerase	P60174	147	5	21.03	3.72	40.70	3.42

Table 5. Proteins (n=18) of Higher Abundance in HC' Cohort (i.e., Exclusive to or of Higher Abundance in HC Cohort Versus CP Cohort) Compared to MD Cohort

Protein names	Entry	Σ HC Spectral Counts	# HC Samples	Compare with sCP		Compare with MD	
				Bayes Factor	Fold Change	Bayes Factor	Fold Change
Alpha-amylase 2B	P19961	2042	5	6.9E+16	21.39	7.55E+16	21.18
Annexin A11	P50995	218	3	n/a	n/a	1.5E+06	18.12
Apolipoprotein E	P02649	289	4	284.92	12.07	3.4E+07	14.38
Collagen alpha-1	P39059	718	5	1.1E+04	8.49	155.32	12.45
Dipeptidase 1	P16444	326	5	1.2E+07	24.85	102.23	6.90
Endothelial protein C receptor	Q9UNN8	329	5	299.78	2.94	56.36	6.22
Extracellular superoxide dismutase	P08294	215	5	7.5E+04	8.36	23.88	6.04
Ig gamma-1 chain C region	P01857	1949	5	58.56	1.46	21.85	4.71
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	P33908	324	4	569.87	12.10	1363.04	4.45
Multimerin-2	Q9H8L6	219	4	n/a	n/a	160.77	3.57
Pro-epidermal growth factor	P01133	1730	5	4.0E+04	2.87	40.30	2.81
Roundabout homolog 4	Q8WZ75	740	5	7.4E+05	10.67	21.01	2.23
Tetranectin	P05452	158	5	9.4E+02	6.83	18.37	1.98
Twisted gastrulation protein homolog 1	Q9GZX9	56	4	n/a	n/a	17.43	1.79
Urokinase-type plasminogen activator	P00749	138	3	12.26	4.56	98.88	1.77
Uromodulin	P07911	8539	5	1.1E+06	2.55	27.47	1.27
Vesicular integral-membrane protein VIP36	Q12907	847	5	1395.01	2.11	25.50	1.04
Vitamin K-dependent protein Z	P22891	345	5	8.8E+06	14.22	101.02	1.01

in severe chronic pancreatitis (sCP) and healthy control (HC) urine. We established subsets of differentially abundant proteins (i.e., those exclusive or of statistically greater abundance): 176 proteins in the sCP cohort and 134 proteins in the HC cohort. We eliminated proteins which were also present in a mixed diagnosis (MD) cohort to isolate chronic pancreatitis as the cause for the enrichment of certain proteins. The MD cohort represents the typical clinical presentation of referred patients: individuals with chronic abdominal pain, acute pancreatic, and probable, but not severe, chronic pancreatitis. Alternatively, a pancreatic cancer cohort could be used, as differentially diagnosing pancreatic cancer and chronic pancreatitis is difficult. Our strategy removed approximately one third of the proteins in each subset. Furthermore, gene ontology analysis identified certain classifications of proteins, such as proteases (particularly serine proteases) and defense/immune proteins as potential protein class-based targets for directed analysis. In addition, of the 609 proteins identified, 50 were identified in all 15 samples, and as such may be useful for normalization in downstream assays.

Of particular interest for further investigation may be the cathepsin family of proteins. Three cathepsins (cathepsin B, D, and Z) were more abundant in sCP compared to the HC

and MD cohorts. This family of proteins is involved in mammalian cellular turnover, including bone absorption [22]. In a prior study, low-trauma bone fracture appeared to have a higher prevalence in chronic pancreatitis [23]. However, cathepsins have a wide range of cellular functions, including breakdown of cellular matrix proteins and cancer [24], but have not been thoroughly studied in pancreatic disease. The role of these particular members of the cathepsin family in bone turnover and its relationship to chronic pancreatitis merits further investigation.

Relative to recent studies, in which thousands of proteins have been identified in urine, 609 proteins is a modest number to be identified via mass spectrometry-based proteomics [8]. In two of the most commonly cited and comprehensive datasets by Adachi, *et al*, [8c] and Kentsis *et al*, [8b], over 1500 and 2000 proteins were identified in urine, respectively. Unlike our single band, short gel strategy, these studies analyzed up to 15 gel slices per lane which significantly increases the resources needed for analysis. We agree that fractionation in its various forms is important for greater analytical depth [25]. Here the additional processing does limit the throughput of the analysis. Here, our goal was not to obtain a comprehensive proteomic profile of urine, but

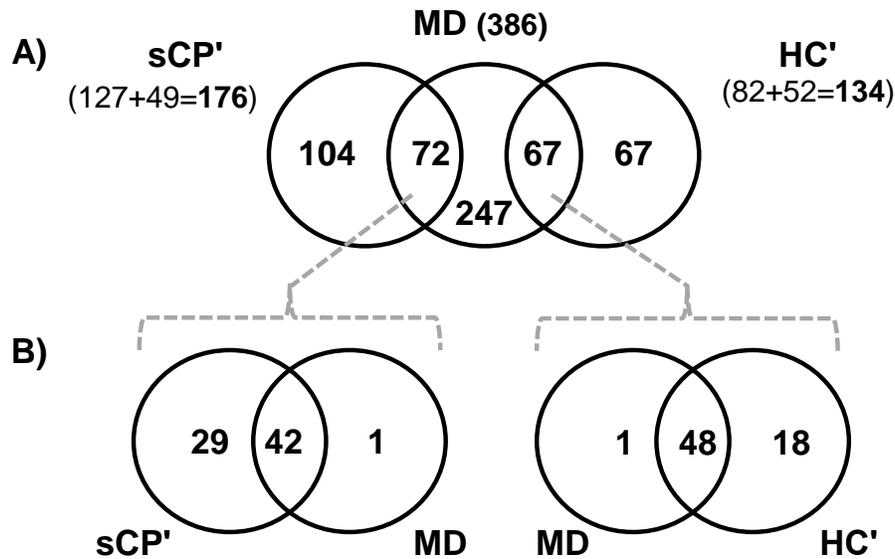


Fig. (4). Venn diagrams comparing differentially expressed proteins for the sCP and HC cohorts with those identified in the MD cohort. **A)** Exclusive and statistically different proteins were combined appropriately into 2 subsets, sCP' and HC', which were compared qualitatively with the MD cohort to eliminate proteins that are not specific to the chronic pancreatitis disease state. **B)** Proteins that were shared between MD and the other cohorts were subjected to QSPEC analysis and proteins with no significant difference between sCP' and MD, as well as those with no significant difference between HC' and MD, were removed from the set of potential chronic pancreatitis targets.

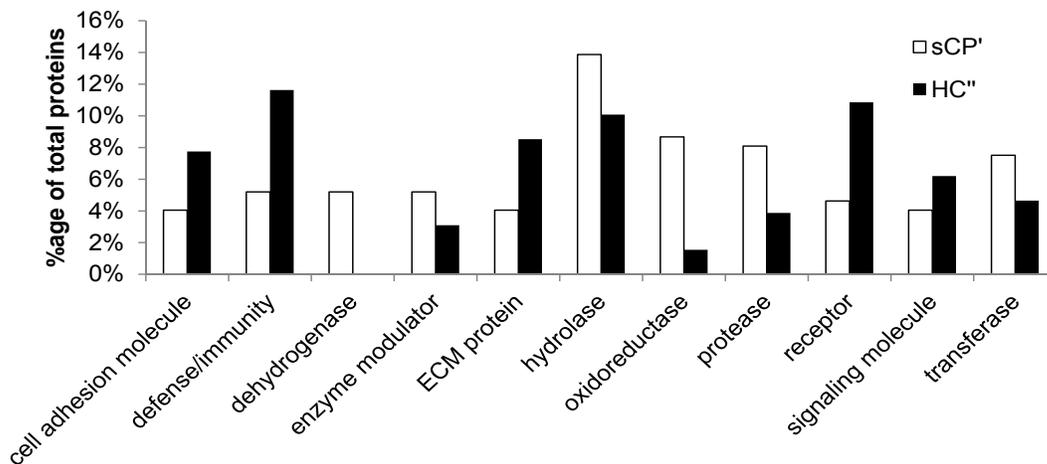


Fig. (5). Gene ontology classification of proteins differentially abundant in severe chronic pancreatitis (sCP) and healthy controls (HC). Functional class categorization of the aforementioned protein subsets according to Panther classification using the DAVID interface.

rather to use minimal sample processing to discover potential protein targets which could be utilized in efficient directed mass spectrometry experiments on larger cohorts. By limiting fractionation, we will likely identify the most abundant proteins. In fact, 18 of the 20 most abundant proteins in urine [26] were among the 50 proteins that were identified in all 15 samples (Supplemental Table 6). However, our intention is to pursue a non-invasively-collected fluid (i.e., urine) as an alternative proteomic strategy due to the relative ease of collection and analysis, a favored characteristic of a diagnostic clinical test. As such, for translational studies seeking an ELISA or dip-stick test as a diagnostic, enrichment and fractionation is a deterrent to general acceptance, making the characterization of abundant proteins a more practical strategy.

A major advantage of the methodology described herein is the minimal sample processing necessary for proteomic analysis. Apart from TCA precipitation to concentrate and extract proteins from the non-protein constituents of urine, no sample enrichment is performed. Although SDS-PAGE was used to visualize proteins prior to tryptic digestion, gel-free in-solution digestion may be an equivalent alternative. We do caution that for protease-rich fluids, such as pancreatic fluid, in-solution digestion may promote undesired proteolysis by endogenous enzymes producing non-tryptic peptides, which will complicate database searching and protein identification. Regarding pancreatic fluid, we postulate that the denaturing conditions of SDS-PAGE and subsequent fixing of the protein in the polyacrylamide matrix prevents refolding of protease active sites, limiting non-tryptic cleav-

age of proteins. However, in urine such spurious protease activity may not be as prevalent, but caution may be warranted. The staining of these gels is also optional as this omission, along with reduced time of trypsinization from overnight to 30 minutes at 55°C [27], would allow the entire sample preparation procedure to be performed in one day.

Using lengthy reversed-phase columns and extended liquid chromatography gradients provide fractionation that is directly in-line with the mass spectrometer. The typical column length for a nanospray mass spectrometry experiment is 12-18 cm, however, we used 45 cm for the present study, along with an extension of the gradient from 1 to 3 hr. Coinciding with the advent of ultra-high pressure liquid chromatography (UHPLC) systems, longer columns, typically packed with reversed-phase C18 particles of smaller diameter ($\leq 2\mu\text{m}$), show promise for the in-line fractionation of complex digests [28].

The advantages of our strategy include its simplicity, speed, and its amenity to multiplexing. Traditional GeLC-MS/MS methods slice gel lanes into typically 6-24 segments, each of which must be processed and analyzed individually thus requiring more time and effort, even with the use of robotics. For complex cell lysates which contain thousands of proteins, such a strategy is warranted, as to acquire the highest degree of proteomic depth. For urine, a less complex proteome, such fractionation maybe unnecessary. Our strategy also allows for multiplexing of 6 or greater samples [29]. However, as with all strategies, caveats also persist. As alluded to above, the lack of fractionation may result in an overly complex sample, as ions may be suppressed, resulting in fewer identified peptides. With next generation mass spectrometry instrumentation, such barriers are being lowered and great proteomic depth may be achieved. We are confident that limited fractionation will be the norm as mass spectrometry instruments continue to achieve greater sequencing speeds and analytical depths.

We identified 50 proteins which appeared across all 15 samples (Supplemental Table 6), thereby establishing a core proteome under our experimental conditions. We realize that this is only a partial inventory of the core proteome of urine. Greater analytical depth resulting from advances in sensitivity and dynamic range, whether it be due upstream fractionation or technological advances, will enhance proteome coverage thereby allowing for further increases in the number of proteins identified across samples. Additional pre-fractionation, at the subcellular level and/or via full 24 fractions on SDS-PAGE, strong cation exchange chromatography, isoelectric focusing, or OFFGEL fractionation would increase the proteome coverage. However, such procedures would concomitantly increase sample processing times and may result potentially in sample loss due to the additional processing procedures. Future studies may further lengthen the column to provide greater protein loading capacity and elongate the gradient for greater fractionation.

Proteins specific to chronic pancreatitis in urine could potentially offer an innovative and non-invasive modality for diagnosis, staging, and prognostication of the disease. In

addition to urine, blood can also be collected non-invasively and used in protein-based chronic pancreatitis analysis. The human blood proteome is a reflection of the specific physiological state at a given point in time from a variety of organ systems. Consistently detected alterations in the proteome can potentially be used for disease diagnosis, staging and prognosis [30]. However, blood (both serum and plasma) has one of the highest dynamic ranges of any body fluid, with specific protein concentrations spanning over 10 orders of magnitude [31]. Although the dynamic range of the urine proteome still spans several orders of magnitude, its simplicity compared to serum makes it an attractive alternative for proteomic research in chronic pancreatitis [32]. Specifically, during urine production, renal blood flow undergoes glomerular ultrafiltration, greatly reducing urine protein complexity, thereby simplifying analysis and preparation methods [33]. In addition, up to 30% of proteins in urine are not of urinary tract origin, but are from distant sites [34]. Although urine was chosen over blood derivatives for analysis because of this decreased complexity and narrower dynamic range, the methods described herein are also relevant to proteomic investigation of blood, as well as other body fluids [35].

The successful elucidation of chronic pancreatitis-specific proteins will rely on not only the methods we describe herein, but also specimen selection and cohort design. Proper statistical analysis must be performed to determine the correct sample size needed to observe a true difference among controls and diseased patients. In addition, proper study participant matching must also be implemented. Proteins with high spectral counts, and which appear in all samples are those which should be targeted for specific analyses. These proteins are abundant and more readily identified via mass spectrometry. As such, the chronic pancreatitis-specific proteins that we describe herein will require orthogonal strategies and well-designed studies using cohorts matched in age, gender, comorbid diseases, and ethnicity among other baseline patient characteristics. Analysis of the urine from a greater number of patients is also key to improve the statistical significance of our results. We present here a first step upon which we and others can build and move forward.

In conclusion, the diagnosis of early chronic pancreatitis remains elusive and is a major pillar of pancreatic research. Early detection is vital to initiate treatment within the therapeutic window before irreversible pancreatic scarring and dysfunction, and before symptoms become debilitating. Here we have presented a mass spectrometry-based strategy for proteomic analysis of urine for chronic pancreatitis research. Our data show both qualitative and quantitative differences between the urine proteomes of healthy individuals and that of individuals with severe chronic pancreatitis. To control for potential confounding diagnoses, we also compared a mixed disease cohort representing typical patients who are commonly referred to a pancreas center, whose protein profile may reflect a diseased state, but not necessarily that of chronic pancreatitis. Based on these investigations, we have assembled a subset of proteins which may be specific to chronic pancreatitis upon further validation. Several of these proteins can be categorized as having roles in defense and

immunity as well as proteases. As such, future studies may develop assays to target these particular classes of proteins. Certain proteins may also serve as targets for further studies using directed mass spectrometry approaches or classical biochemistry techniques, such as ELISA or western blotting.

ABBREVIATIONS

CP	=	chronic pancreatitis
GeLC-MS/MS	=	in-gel tryptic digestion followed by liquid chromatography-tandem mass spectrometry
HC	=	healthy controls
LTQ-FTICR	=	linear trap quadrupole-in middle Fourier transform ion cyclotron resonance mass spectrometry
MD	=	mixed diagnosis
sCP	=	severe chronic pancreatitis
TCA	=	trichloroacetic acid

CONFLICTS OF INTERESTS

The authors confirm that this article content has no conflicts of interest.

AUTHOR CONTRIBUTIONS

JP carried out the experiments. SB and VK provided technical assistance. JP drafted the original manuscript. JP, HS, PB and DC conceived of the study, and participated in its design and coordination. All authors helped to draft the manuscript and approved the final manuscript.

ACKNOWLEDGEMENTS

Funds were provided by the following NIH grants: 1 F32 DK085835-01A1 (JP), 1 R21 DK081703-01A2 (DC) and 5 P30 DK034854-24 (Harvard Digestive Diseases Center; DC). We would like to thank the Burrill family for their generous support through the Burrill Research Grant. We would also like to thank members of the Steen Laboratory at Boston Children's Hospital, particularly John FK Sauld and Ali Ghoulidifor their technical assistance and critical reading of the manuscript. In addition, we thank members of the Center for Pancreatic Disease at Brigham and Women's Hospital, particularly Shadeah Suleiman for her technical assistance.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

- [1] James, S. Opportunities and challenges at NIDDK in digestive diseases research. *Gastroenterology*, **2007**, *132*, 1219-1220.
- [2] DiMagno, E. P.; Go, V. L.; Summerskill, W. H. Relations between pancreatic enzyme outputs and malabsorption in severe pancreatic insufficiency. *N. Engl. J. Med.*, **1973**, *288* (16), 813-815.
- [3] Chowdhury, R.; Bhutani, M. S.; Mishra, G.; Toskes, P. P.; Forsmark, C. E. Comparative analysis of direct pancreatic function testing versus morphological assessment by endoscopic ultrasonography for the evaluation of chronic unexplained abdominal pain of presumed pancreatic origin. *Pancreas*, **2005**, *31* (1), 63-68.
- [4] (a) Paulo, J. A.; Kadiyala, V.; Lee, L. S.; Banks, P. A.; Conwell, D. L.; Steen, H. Proteomic Analysis (GeLC-MS/MS) of ePFT- Collected Pancreatic Fluid in Chronic Pancreatitis. *J. Proteome Res.*, **11**(3), 1897-1912. **2012**; (b) Paulo, J. A.; Lee, L. S.; Banks, P. A.; Steen, H.; Conwell, D. L. Difference gel electrophoresis identifies differentially expressed proteins in endoscopically collected pancreatic fluid. *Electrophoresis*, **2011**, *32* (15), 1939-1951; (c) Paulo, J. A.; Lee, L. S.; Wu, B.; Banks, P. A.; Steen, H.; Conwell, D. L. Cytokine profiling of pancreatic fluid using the ePFT collection method in tandem with a multiplexed microarray assay. *J. Immunol. Methods*, **2011**, *369* (1-2), 98-107; (d) Paulo, J. A.; Lee, L. S.; Wu, B.; Banks, P. A.; Steen, H.; Conwell, D. L. Mass spectrometry-based proteomics of endoscopically collected pancreatic fluid in chronic pancreatitis research. *Proteomics Clin. Appl.*, **2011**, *5* (3-4), 109-120; (e) Paulo, J. A.; Lee, L. S.; Wu, B.; Repas, K.; Banks, P. A.; Conwell, D. L.; Steen, H. Optimized sample preparation of endoscopic collected pancreatic fluid for SDS-PAGE analysis. *Electrophoresis*, **2010**, *31* (14), 2377-2387; (f) Paulo, J. A.; Lee, L. S.; Wu, B.; Repas, K.; Mortelet, K. J.; Banks, P. A.; Steen, H.; Conwell, D. L. Identification of pancreas-specific proteins in endoscopically (endoscopic pancreatic function test) collected pancreatic fluid with liquid chromatography-tandem mass spectrometry. *Pancreas*, **2010**, *39* (6), 889-896.
- [5] (a) Chen, R.; Brentnall, T. A.; Pan, S.; Cooke, K.; Moyes, K. W.; Lane, Z.; Crispin, D. A.; Goodlett, D. R.; Aebersold, R.; Bronner, M. P. Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer. *Mol. Cell Proteomics*, **2007**, *6* (8), 1331-1342; (b) Chen, R.; Pan, S.; Cooke, K.; Moyes, K. W.; Bronner, M. P.; Goodlett, D. R.; Aebersold, R.; Brentnall, T. A. Comparison of pancreas juice proteins from cancer versus pancreatitis using quantitative proteomic analysis. *Pancreas*, **2007**, *34* (1), 70-79; (c) Chen, R.; Yi, E. C.; Donohoe, S.; Pan, S.; Eng, J.; Cooke, K.; Crispin, D. A.; Lane, Z.; Goodlett, D. R.; Bronner, M. P.; Aebersold, R.; Brentnall, T. A. Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. *Gastroenterology*, **2005**, *129* (4), 1187-1197; (d) Cui, Y.; Tian, M.; Zong, M.; Teng, M.; Chen, Y.; Lu, J.; Jiang, J.; Liu, X.; Han, J. Proteomic analysis of pancreatic ductal adenocarcinoma compared with normal adjacent pancreatic tissue and pancreatic benign cystadenoma. *Pancreatol.*, **2008**, *9* (1-2), 89-98; (e) Gronborg, M.; Bunkenborg, J.; Kristiansen, T. Z.; Jensen, O. N.; Yeo, C. J.; Hruban, R. H.; Maitra, A.; Goggins, M. G.; Pandey, A. Comprehensive proteomic analysis of human pancreatic juice. *J. Proteome Res.*, **2004**, *3* (5), 1042-1055; (f) Gronborg, M.; Kristiansen, T. Z.; Iwahori, A.; Chang, R.; Reddy, R.; Sato, N.; Molina, H.; Jensen, O. N.; Hruban, R. H.; Goggins, M. G.; Maitra, A.; Pandey, A. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol Cell Proteomics*, **2006**, *5* (1), 157-171; (g) Ke, E.; Patel, B. B.; Liu, T.; Li, X. M.; Haluszka, O.; Hoffman, J. P.; Ehya, H.; Young, N. A.; Watson, J. C.; Weinberg, D. S.; Nguyen, M. T.; Cohen, S. J.; Meropol, N. J.; Litwin, S.; Tokar, J. L.; Yeung, A. T. Proteomic Analyses of Pancreatic Cyst Fluids. *Pancreas*, **2009**, *38* (2), 33-42; (h) Li, C.; Simeone, D. M.; Brenner, D. E.; Anderson, M. A.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M. Pancreatic Cancer Serum Detection Using a Lectin/Glyco-Antibody Array Method. *J. Proteome Res.*, **2008**, *8* (2), 483-492; (i) Zhao, Y.; Lee, W. N.; Lim, S.; Go, V. L.; Xiao, J.; Cao, R.; Zhang, H.; Recker, R. R.; Xiao, G. G. Quantitative proteomics: measuring protein synthesis using 15N amino acid labeling in pancreatic cancer cells. *Anal. Chem.*, **2009**, *81* (2), 764-771.
- [6] Issaq, H. J.; Xiao, Z.; Veenstra, T. D. Serum and plasma proteomics. *Chem. Rev.*, **2007**, *107* (8), 3601-3620.
- [7] Paulo, J. A.; Gaun, A.; Kadiyala, V.; Ghoulidi, A.; Banks, P. A.; Conwell, D. L.; Steen, H. Subcellular fractionation enhances

- proteome coverage of pancreatic duct cells. *Biochim. Biophys. Acta*, **2013** 1834(4), 791-797
- [8] (a) Nagaraj, N.; Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J. Proteome Res.*, **2011**, *10* (2), 637-645; (b) Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steen, H. Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clin. Appl.*, **2009**, *3* (9), 1052-1061; (c) Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J. V.; Mann, M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol.*, **2006**, *7* (9), R80; (d) Pieper, R.; Gatlin, C. L.; McGrath, A. M.; Makusky, A. J.; Mondal, M.; Seonarin, M.; Field, E.; Schatz, C. R.; Estock, M. A.; Ahmed, N.; Anderson, N. G.; Steiner, S. Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics*, **2004**, *4* (4), 1159-1174.
- [9] Schneider, A.; Lohr, J. M.; Singer, M. V. The M-ANNHEIM classification of chronic pancreatitis: introduction of a unifying classification system based on a review of previous classifications of the disease. *J. Gastroenterol.*, **2007**, *42* (2), 101-119.
- [10] Stevens, T.; Conwell, D.; Zuccaro, G.; Van Lente, F.; Khandwala, F.; Hanaway, P.; Vargo, J. J.; Dumot, J. A. Analysis of pancreatic elastase-1 concentrations in duodenal aspirates from healthy subjects and patients with chronic pancreatitis. *Dig. Dis. Sci.*, **2004**, *49* (9), 1405-1411.
- [11] Choi, H.; Nesvizhskii, A. I. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J. Proteome Res.*, **2008**, *7* (1), 47-50.
- [12] (a) Dennis, G., Jr.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.*, **2003**, *4* (5), P3; (b) Huang da, W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **2009**, *4* (1), 44-57.
- [13] (a) Neubauer, G.; Mann, M. Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal. Chem.*, **1999**, *71* (1), 235-242; (b) Steen, H.; Kuster, B.; Fernandez, M.; Pandey, A.; Mann, M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.*, **2001**, *73* (7), 1440-1448.
- [14] Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*, **2008**, *24* (21), 2534-2536.
- [15] Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell Proteomics*, **2007**, *6* (9), 1638-1655.
- [16] (a) Gupta, N.; Pevzner, P. False discovery rates of protein identifications: a strike against the two-peptide rule. *J. Proteome Res.*, **2009**, *8* (9), 4173-4181; (b) Veenstra, T.; Conrads, T.; Issaq, H., What to do with "one-hit wonders"? *Electrophoresis*, **2004**, *25* (9), 1278-1279.
- [17] Oliveros, J. VENNY: An interactive tool for comparing lists with Venn diagrams. Available from: <http://bioinfopg.cnb.csic.es/tools/venny/index.html>
- [18] Dong, M. Q.; Venable, J. D.; Au, N.; Xu, T.; Park, S. K.; Cociorva, D.; Johnson, J. R.; Dillin, A.; Yates, J. R. 3rd Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science*, **2007**, *317* (5838), 660-663.
- [19] (a) Goodman, S. N. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann. Intern. Med.*, **1999**, *130* (12), 995-1004; (b) Goodman, S. N. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann. Intern. Med.*, **1999**, *130* (12), 1005-1013.
- [20] Jeffreys H. *Theory of probability*. 3rd ed.; Clarendon Press: Oxford, **1961**; p 447.
- [21] Thomas, P. D.; Kejariwal, A.; Campbell, M. J.; Mi, H.; Diemer, K.; Guo, N.; Ladunga, I.; Ulitsky-Lazareva, B.; Muruganujan, A.; Rabkin, S.; Vandergriff, J. A.; Doremieux, O. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.*, **2003**, *31* (1), 334-341.
- [22] Authier, F.; Posner, B. I.; Bergeron, J. J. Endosomal proteolysis of internalized proteins. *FEBS Lett.*, **1996**, *389* (1), 55-60.
- [23] Tignor, A. S.; Wu, B. U.; Whitlock, T. L.; Lopez, R.; Repas, K.; Banks, P. A.; Conwell, D. High prevalence of low-trauma fracture in chronic pancreatitis. *Am. J. Gastroenterol.*, **2010**, *105* (12), 2680-2686.
- [24] (a) Guzińska-Ustymowicz, K.; Zalewski, B.; Kasacka, I.; Piotrowski, Z.; Skrzydlewska, E. Activity of cathepsin B and D in colorectal cancer: relationships with tumour budding. *Anticancer Res.*, **2004**, *24* (5A), 2847-2851; (b) Leto, G.; Gebbia, N.; Rausa, L.; Tumminello, F. Cathepsin D in the malignant progression of neoplastic diseases (review). *Anticancer Res.*, **1992**, *12* (1), 235-240; (c) Sevenich, L.; Schurig, U.; Sachse, K.; Gajda, M.; Werner, F.; Müller, S.; Vasiljeva, O.; Schwinde, A.; Klemm, N.; Deussing, J.; Peters, C.; Reinheckel, T. Synergistic antitumor effects of combined cathepsin B and cathepsin Z deficiencies on breast cancer progression and metastasis in mice. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (6), 2497-2502; (d) Sloane, B.; Yan, S.; Podgorski, I.; Linebaugh, B.; Cher, M.; Mai, J.; Cavallo-Medved, D.; Sameni, M.; Dosesescu, J.; Moin, K. Cathepsin B and tumor proteolysis: contribution of the tumor microenvironment. *Semin. Cancer Biol.*, **2005**, *15* (2), 149-157.
- [25] Paulo, J. A.; Gaun, A.; Kadiyala, V.; Ghouli, A.; Banks, P. A.; Conwell, D. L.; Steen, H. Subcellular fractionation enhances proteome coverage of pancreatic duct cells. *Biochim. Biophys. Acta.*, **2013**, *1834* (4), 791-797.
- [26] Nagaraj, N.; Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J. Proteome Res.*, **2011**, *10* (2), 637-645.
- [27] Havlis, J.; Thomas, H.; Sebela, M.; Shevchenko, A. Fast-response proteomics by accelerated in-gel digestion of proteins. *Anal. Chem.*, **2003**, *75* (6), 1300-1306.
- [28] (a) Nagaraj, N.; Alexander Kulak, N.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. System-wide Perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra hplc runs on a bench top orbitrap. *Mol. Cell Proteomics*, **2012**, *11* (3), M111 013722; (b) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Frohlich, F.; Cox, J.; Mann, M., Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell Proteomics*, **2011**, *10* (8), M110 003699.
- [29] Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.*, **2003**, *75* (8), 1895-1904.
- [30] Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell Proteomics*, **2004**, *3* (4), 311-326.
- [31] Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics*, **2002**, *1* (11), 845-867.
- [32] Fliser, D.; Novak, J.; Thongboonkerd, V.; Argiles, A.; Jankowski, V.; Girolami, M. A.; Jankowski, J.; Mischak, H. Advances in urinary proteome analysis and biomarker discovery. *J. Am. Soc. Nephrol.*, **2007**, *18* (4), 1057-1071.
- [33] (a) Rosty, C.; Christa, L.; Kuzdzal, S.; Baldwin, W. M.; Zahurak, M. L.; Carnot, F.; Chan, D. W.; Canto, M.; Lillemoe, K. D.; Cameron, J. L.; Yeo, C. J.; Hruban, R. H.; Goggins, M. Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein 1 as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res.*, **2002**, *62* (6), 1868-1875 (b) Kishi, T.; Grass, L.; Soosaipillai, A.; Scorilas, A.; Harbeck, N.; Schmalfeldt, B.; Dorn, J.; Mysliwiec, M.; Schmitt, M.; Diamandis, E. P. Human kallikrein 8, a novel biomarker for ovarian carcinoma. *Cancer Res.*, **2003**, *63* (11), 2771-2774; (c) Alatas, F.; Alatas, O.;

- Metintas, M.; Colak, O.; Harmanci, E.; Demir, S. Diagnostic value of CEA, CA 15-3, CA 19-9, CYFRA 21-1, NSE and TSA assay in pleural effusions. *Lung Cancer*, **2001**, *31* (1), 9-16.
- [34] Decramer, S.; Gonzalez de Peredo, A.; Breuil, B.; Mischak, H.; Monsarrat, B.; Bascands, J. L.; Schanstra, J. P. Urine in clinical proteomics. *Mol. Cell. Proteomics*, **2008**, *7* (10), 1850-1862.
- [35] Paulo, J.; Vaezzadeh, A.; Conwell, D.; Lee, R.; Steen, H. Sample Handling of Body Fluids for Proteomics. In *Sample Preparation in Biological Mass Spectrometry*, Ivanov, A.; Lazarev, A., Eds. Springer: New York, NY, **2011**.

Received: May 29, 2012

Revised: July 15, 2012

Accepted: July 28, 2012

© Paulo *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.