Mass Spectrometry-Based (GeLC-MS/MS) Comparative Proteomic Analysis of Endoscopically (ePFT) Collected Pancreatic and Gastroduodenal Fluids

Joao A. Paulo, PhD1,2,3, Vivek Kadiyala, MD3, Peter A. Banks, MD3, Hanno Steen, PhD1,2,4 and Darwin L. Conwell, MD, MS3,4

OBJECTIVES: The secretin-stimulated endoscopic pancreatic function test (ePFT) allows for the safe collection of gastroduodenal and pancreatic fluid from the duodenum. We test the hypothesis that these endoscopically collected fluids have different proteomes. As such, we aim to show that the ePFT method can be used to collect fluid enriched in pancreatic proteins to test for pancreatic function.

METHODS: Gastroduodenal and pancreatic fluid were collected sequentially from chronic pancreatitis patients undergoing an ePFT. Proteins from each fluid type were extracted using previously published optimized methods and subjected to GeLC-MS/MS analysis for protein identification and bioinformatics analysis.

RESULTS: Mass spectrometry analysis identified proteins that were exclusive in either gastroduodenal (46) or pancreatic fluid (234). Subsequent quantitative analysis revealed proteins that were differentially abundant with statistical significance. As expected, proteolytic enzymes and protease inhibitors were among the differentially detected proteins. The proteases pepsinogens and gastrin were enriched in gastroduodenal fluid, while common pancreatic enzymes (e.g., aminopeptidase N, chymotrypsin C, elastase-3A, trypsin, and carboxypeptidase A1, and elastase 2B) were found in greater abundance in pancreatic fluid. Similarly for protease inhibitors, members of the cystatin family were exclusive to gastroduodenal fluid, while serpins A11, B4, and D1 were exclusive to pancreatic fluid.

CONCLUSIONS: We have shown that ePFT collection coupled with mass spectrometry can be used to identify differentially detected proteins in gastroduodenal and pancreatic fluids. The data obtained using GeLC-MS/MS techniques provide further evidence supporting the feasibility of using ePFT-collected fluid to study specific diseases of the upper gastrointestinal tract, such as chronic pancreatitis.

Clinical and Translational Gastroenterology (2012) 3, e14; doi:10.1038/ctg.2012.7; published online 3 May 2012

INTRODUCTION

Diseases of the upper gastrointestinal tract are major burdens on the healthcare system. In the United States alone, over 22,000 new cases of gastric cancer were diagnosed and over 11,000 gastric cancer-related deaths were reported in 2006.1 Nationally, the financial cost of gastric cancer in 2010 is estimated to be more than $1.8 billion.2 Similarly, disorders of the pancreas affect more than 1 million persons in the United States and cost nearly $3 billion annually. Over the past decade, pancreatic diseases have resulted in 277,000 hospitalizations and 475,000 annual ambulatory care visits, of which nearly 25% are because of chronic pancreatitis.3 Novel methods are needed to enable the diagnosis of early disease and to understand better the pathogenesis of the disease. Proteomic experiments directed toward the study of gastroduodenal and pancreatic disease present a unique opportunity to accelerate the pace of disease-specific biomarker discovery.

Although tissue biopsies are often utilized for screening and diagnosis of gastrointestinal diseases, such methods are invasive, may result in infections and complications, and only sample a small region of the tissue being investigated, potentially missing the diseased region. Proximal body fluids, however, bathe the diseased organ and represent the proteins in the nearby tissue.4 In contrast, the proteins in systemic body fluids, such as urine, blood, and its derivatives (i.e., plasma or serum), correlate to the entire body and likely include those not related to the disease of interest. Furthermore, disease-specific markers that are secreted or shed directly into proximal fluids are likely to be present in a higher concentration than in blood or urine, as no dilution or filtration has occurred.

The endoscopic pancreatic function test (ePFT) allows for the safe collection of gastroduodenal and pancreatic fluid from the duodenum. During this test, pancreatic secretion is typically stimulated by secretin, which acts on pancreatic duct
cell receptors.5,6 Stimulation of these duct cells by secretin results in secretion of bicarbonate-enriched fluid, which facilitates excretion of proteins. This fluid is readily and safely collected endoscopically and can be used for mass spectrometry-based proteomic analysis.7,8

Pancreatic and gastroduodenal fluid are excellent clinical specimens for the identification of disease-specific biomarkers by proteomic analysis, as each is a proximal fluid of relatively low complexity, thereby facilitating the identification of low-abundant proteins.9–11 The application of body fluid proteomics in the study of pancreatic and gastroduodenal disease may reveal physiologically and clinically relevant markers of disease.12,13 Gastroduodenal and pancreatic fluid are proximal body fluids of the digestive system having roles in protein digestion, and therefore are expected to be rich in digestive proteins. Moreover, many gastric and pancreatic diseases are exacerbated by adverse proteolytic events. As such, we expect to discover differences in secreted proteolytic enzymes and protease inhibitors between the two fluids. Further knowledge of the underlying molecular events affecting proteolytic dysregulation will have a positive impact on the understanding of the natural history of upper gastrointestinal tract diseases.

Using our previously optimized methods of sample preparation for gastroduodenal14 and pancreatic fluid,15,16 and mass spectrometry-based protein identification, we present a comparative proteomic analysis of these ePFT-collected fluid samples. We aim to show that ePFT-collected gastroduodenal and secretin-stimulated pancreatic fluids have different proteomes. As such, it would follow that the ePFT method can be used to collect secretin-stimulated fluid enriched in proteins of pancreatic origin to test for pancreatic function.

To achieve our aims, we will (1) collect gastroduodenal and pancreatic fluids from each subject, (2) extract proteins from gastroduodenal and pancreatic fluids using our previously optimized methods, (3) perform GeLC-MS/MS analysis, (4) identify proteins using Mascot (Matrix Science, Boston, MA, USA) and ProteomeDiscoverer software (Thermo Fisher Scientific, Waltham, MA, USA), and (5) compare identified proteins using qualitative (Scaffold3) and quantitative (QSPEC) bioinformatic techniques.

**METHODS**

**Materials.** ChiRhoStim synthetic human secretin was from ChiRhoClin (Burtons ville, MD, USA). SeeBluePlus2 Pre-Stained standard (LC5925), 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) running buffer (NP002) were from Invitrogen (Carlsbad, CA, USA). Other reagents and solvents were from Sigma-Aldrich (St Louis, MO, USA) and Burdick and Jackson (Muskegon, MI, USA), respectively.

**Study cohort.** This protocol was approved by the Institutional Review Board at Brigham and Women’s Hospital (IRB 2007-P-002480/1). The study cohort included adult patients seen in the Center for Pancreatic Diseases at Brigham and Women’s Hospital. Subjects were referred for evaluation of pancreatic etiologies for their gastrointestinal symptoms. All subjects underwent the following: (1) comprehensive history and physical examination, (2) review of radiological and endoscopic data, and (3) upper endoscopy with ePFT followed by mucosal biopsy. The diagnosis of chronic pancreatitis was determined using the M-ANNHEIM (Multiple risk factors, Alcohol, Nicotine, Nutrition, Hereditary factors, Eff e rent duct factors, Immunological factors, and Miscellaneous and metabolic factors) classification.17 The M-ANNHEIM classification is a standardized system designed to classify chronic pancreatitis according to etiology, clinical staging, and severity of the disease.17 This system considered clinical data from a wide array of laboratory tests, and radiological imaging techniques including ultrasound, endoscopic ultrasound, magnetic resonance imaging, computed tomography, as well as risk factors of chronic pancreatitis.17,18 The study cohort (Table 1) was comprised of three subjects classified as “Definite Chronic Pancreatitis,” according to the M-ANNHEIM classification system.

**Experimental workflow.** Figure 1 illustrates the general workflow for the overall analysis as follows: (1) collect gastroduodenal and pancreatic fluids from each subject, (2) extract proteins from gastroduodenal and pancreatic fluids using our previously optimized methods,12,14,16 (3) perform GeLC-MS/MS analysis, (4) identify proteins using Mascot and ProteomeDiscoverer software, and (5) compare identified proteins using qualitative (Scaffold3) and quantitative (QSPEC) bioinformatic techniques.

**Gastroduodenal and pancreatic fluid collection (ePFT method).** The ePFT procedure was performed as previously described.7 Gastroduodenal fluid was collected immediately before secretin stimulation.14 Pancreatic fluid that was used for the ensuing analysis was collected at the 30-min time point following secretin stimulation, as was previously published.15 Protein concentration was determined using the bicinchoninic acid protein assay.19

**Protein precipitation.** We have previously compared a series of protein precipitation procedures, establishing that proteins from gastroduodenal and pancreatic fluid were most efficiently extracted by acetone and trichloroacetic acid (TCA), respectively.14,16 Although we used different methods of precipitation, it was our intention to maximize the proteins identified via our previously optimized protocols. These precipitation processes limit protein degradation by instantaneously deactivating enzymes, concentrating the protein sample, and removing salts that will interfere with the subsequent electrophoretic mobility-based fractionation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), as described below.

**Acetone precipitation of gastroduodenal fluid.** Four sample volumes (800 μl) of ice-cold 100% acetone were added to 200 μl of gastroduodenal fluid, vortexed briefly, and incubated at −20 °C for 3 h. Subsequently, the samples were centrifuged at 20,000 × g at 4 °C for 30 min. The super-
natants were carefully aspirated and the pellets were allowed to air dry at 23°C. TCA precipitation of pancreatic fluid.

Aliquots of pancreatic fluid samples for proteomic analysis were collected on ice (as described above), centrifuged at 4°C at 14,000 r.p.m. for 15 min to remove cellular debris, and aliquoted (500 ml) before storage at -80°C. Ice-cold 100% TCA (25 ml) was added to 200 ml of pancreatic fluid, vortexed, and incubated at 4°C for 2 h. The sample was centrifuged at 20,000 g at 4°C for 30 min and the supernatant was carefully aspirated. One milliliter of 100% ice-cold acetone was added to the pellets, which were briefly vortexed and incubated at 0°C for 1 h. The sample was centrifuged at 20,000 g at 4°C for 30 min and the final pellets were allowed to air dry at room temperature.

SDS-PAGE prefractionation and liquid chromatography-tandem mass spectrometry (GeLC-MS/MS) of gastroduodenal pancreatic fluid specimens.

Protein concentration was determined using the BioRAD protein assay according to the manufacturer's instructions, so that 100 mg of protein was added per gel lane. Each precipitated protein pellet was re-dissolved in 50 ml of reducing Laemmli buffer (with 10 mM DTT) for 1 h at 56°C and alkylated with 1% acrylamide at room temperature for 30 min or subsequent SDS-PAGE prefractionation.

Table 1: Demographic and clinical data for study cohort

| Patient | Reason for referral | Age (year) | Gender | CT Scan | MRI | PE-1 (units/mg) | EUS Score (0-9) | ePFT: peak HCO3- (meq/l) | GDF protein conc. (mg/ml) | GDF volume (ml) | PF protein conc. (mg/ml) | PF volume (ml) | Endoscopic biopsy | M-ANNHEIM category |
|---------|---------------------|------------|--------|---------|-----|----------------|----------------|-----------------|----------------|----------------|----------------|----------------|-----------------|------------------|
| 1       | Chronic pancreatitis | 64         | Female | Atrophy, calcifications, ductal dilation and stricture | Atrophy, dudal dilation and stricture, filling defects, low T1 signal, dilated side branches | 221 | N/A | 37 | 1.4 | 3.3 | 1.1 | 5.3 | Duodenitis, gastropathy | Definite chronic pancreatitis |
| 2       | Chronic pancreatitis | 60         | Female | Atrophy, calcifications, ductal dilation and stricture | N/A | N/A | 8 | 54 | 2.1 | 1.5 | 0.6 | 4.5 | Normal | Definite chronic pancreatitis |
| 3       | Chronic pancreatitis | 52         | Female | Atrophy, calcifications, ductal dilation and stricture | N/A | < 15 | 5 | 38 | 1.6 | 2.0 | 0.7 | 6.5 | Normal | Definite chronic pancreatitis |

conc., concentration; CP, chronic pancreatitis; CT, computed tomography; EUS, endoscopic ultrasound; ePFT, endoscopic pancreas function test; GDF, gastroduodenal fluid; HCO3-, bicarbonate; M-ANNHEIM, Multiple risk factors, Alcohol, nicotine, nutrition, heredity, different duct factors, immunological factors, and miscellaneous and metabolic factors; MRI, magnetic resonance imaging; N/A, not available; PE-1, pancreatic elastase 1; PF, pancreatic fluid.

Figure 1: General workflow. (1) As part of the standard endoscopic pancreas function test (ePFT) procedure, gastroduodenal fluid is collected from the duodenum before secretin injection. In addition, we collected a pancreatic fluid sample 30 min after secretin stimulation. (2) Protein was extracted using acetone precipitation for gastroduodenal fluid and trichloroacetic acid (TCA) precipitation for pancreatic fluid. (3) Following precipitation, proteins from each sample were fractionated via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and gel lanes were subsequently processed via standard gel LC-MS/MS methods. (4) Database searching was performed via the Proteome Discoverer graphical user interface with Mascot database searching. (5) Data were analyzed further using Scaffold3 for qualitative comparison and QSPEC to identify statistically significant quantitative differences. GDP, gastroduodenal fluid; PF, pancreatic fluid.
GeLC-MS/MS analysis. The proteins were fractionated on 4–12% NuPAGE pre-cast SDS-PAGE gels at 175 V for 45 min using MES-SDS running buffer. Subsequently, each gel lane was divided into 10 sections. Proteins in each gel section were digested in-gel with trypsin. The extracted peptides from each gel section were subjected to peptide fractionation using reversed-phase high performance liquid chromatography (Thermo Scientific, Waltham, MA, USA) and the gradient-eluted peptides were analyzed by a hyphenated LTQ-FTICR (linear trap quadrupole-Fourier Transform ion cyclotron resonance) mass spectrometer (Thermo Scientific). The liquid chromatography columns (15 cm × 100 μm ID) were packed in-house (Magic C18, 5 μm, 100 Å beads, Michrom BioResources, into PicoTips, New Objective, Woburn, MA, USA). Samples were analyzed with a 60-min linear gradient (5–35% acetonitrile with 0.2% formic acid) and data were acquired in a data-dependent manner, in which MS/MS fragmentation was performed using the six most intense peaks of every full MS scan.

**Bioinformatics and data analysis**

**Database search.** All data generated from the gel sections were searched against the international protein index human database (v3.69) using the Mascot search engine (v.2.204; Matrix Science, Boston, MA, USA) through the ProteomeDiscoverer graphical user interface (v 1.2; Thermo Fisher Scientific). One miscleavage per peptide was allowed of the peptides of all the entries in the international protein index database (v3.69).23,24

**Scaffold.** Scaffold (version Scaffold 3.00.07, Proteome Software, Portland, OR, USA) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at >95% probability as specified by the Peptide Prophet algorithm. Protein identifications were accepted if they could be established at greater than 99% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

**Spectral counting.** Relative protein quantitation was accomplished using a label-free technique, spectral counting, which compared the number of identified tandem mass spectra 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 previously suggested. Specifically, spectral counts of each protein were divided first by the total spectral counts of all proteins from the same sample, and then multiplied by the total spectral counts of the sample with the maximum total number of spectral counts. Significance analysis of our spectral count data was performed using OSPEC, a recently published algorithm for determining the statistical significance of differences in spectral counting data from two sample sets. This algorithm uses the Bayes factor, in lieu of the $P$ value, as a measure of statistical significance. According to convention, a Bayes factor > 10 suggests strong evidence (analogous to a type-I error of $\alpha < 0.05$) that a particular protein was differentially detected with statistical significance between the two cell states, thus a value of 10 was used as our significance threshold.

**Results**

**Gastroduodenal and pancreatic fluid were collected successfully for proteomic analysis.** Gastroduodenal and pancreatic fluid from three patients were collected safely with a sterile trap in-line with the vacuum. Table 1 lists the volumes and protein concentrations (as estimated using the bicinchoninic acid assay) of the collected fluids. For pancreatic fluid, the mean volume was 5.4 ± 1.0 ml, while the mean protein concentration was 0.8 ± 0.3 mg/ml. Similarly, for gastroduodenal fluid, the mean volume was 2.3 ± 0.9 ml, while the mean protein concentration was 1.7 ± 0.4 mg/ml. Samples were stored on ice and centrifuged (3 000 × g for 15 min at 4°C) within 30 min of collection. The collected supernatants were frozen at –80°C until proteins were extracted for SDS-PAGE analysis.

**SDS-PAGE analysis revealed characteristic protein patterns.** In agreement with what we have published previously, the protein patterns as illustrated by SDS-PAGE imaging differed substantially between gastroduodenal fluid (Figure 2a) and pancreatic fluid (Figure 2b). In addition, when comparing inter-fluid variation, fewer protein bands were present in gastroduodenal fluid samples compared with samples from pancreatic fluid. Analogous differences in the number of proteins were determined in subsequent mass spectrometry-based protein identifications. Also, when comparing samples within a fluid type, the protein patterns were similar, but inter-patient variations were visible when comparing the protein banding patterns within both the gastroduodenal and pancreatic fluid sample sets.

**Qualitative comparative mass spectrometry data analysis identified proteins that were exclusive to gastroduodenal or pancreatic fluid.** Using our mass spectrometry-based strategy in which a non-redundant list of proteins were determined using the Scaffold3 software, we identified a total of 285 proteins in gastroduodenal fluid and 473 proteins in pancreatic fluid (Figure 3). Of these proteins, 46 (Supplementary Table 1) were determined to be exclusive to gastroduodenal fluid and 234 (Supplementary Table 2) were determined to be exclusive to pancreatic fluid. Proteins common to both the gastroduodenal and pancreatic fluid samples (239) were analyzed further by QSPEC to determine proteins that were not exclusive to either fluid, but were enriched in one fluid or the other. In Table 2 we summarized the protein identification data. In addition, Supplementary Figure 1 illustrates the differences in the
number of proteins identified in pancreatic and gastroduodenal fluid for each subject.

Quantitative comparative mass spectrometry data analysis using QSPEC identified proteins with statistically significant enrichment in either gastroduodenal or pancreatic fluid. We performed QSPEC analysis of the spectral count values for the 239 proteins that were common to both gastroduodenal and pancreatic fluid to determine statistically significant differences between the two fluids. QSPEC was recently published as a Bayesian statistics-based algorithm for determining the statistical significance of differences in spectral counting data from two sample sets—in our case, gastroduodenal and pancreatic fluid. This analysis revealed 41 proteins enriched in gastroduodenal fluid (Supplementary Table 3) and 67 proteins enriched in pancreatic fluid (Supplementary Table 4) according to Bayesian statistical methods.

We defined “differentially detected proteins” as those either exclusive to or of statistically significant enrichment in one fluid relative to the other. As such, we have identified a total of 87 differentially detected proteins (46 exclusive plus 41 enriched) in gastroduodenal fluid (Table 2). Likewise, we have identified a total of 301 differentially detected proteins (234 exclusive plus 67 enriched) in pancreatic fluid (Table 2). The differentially detected proteins were used for subsequent gene ontology analysis.

Gene ontology analysis of differentially detected proteins using Scaffold3 detected differences in localization and protein function. Using Scaffold3, we investigated the molecular function of the differentially detected proteins (Figure 4). We noted that a large percentage of the proteins identified from both fluids were classified as binding, catalytic, and enzyme regulators. The binding category is very broad and includes ion, protein, lipid, nucleotide, carbohydrate, and oxygen binding. As such, it is expected that this category would be relatively large. The catalytic and enzyme regulator categories were of particular interest in regard to proteases and protease inhibitors related to digestion. Table 3 lists proteases that were exclusive to a particular fluid, while Table 4 lists those proteases that were enriched, to a statistically significant degree, in a particular fluid. Similarly, for the identified protease inhibitors, Table 5 lists those that were exclusive to a particular fluid, whereas Table 6 lists those that were enriched in a particular fluid.

Discussion

We have shown that the ePFT collection method coupled with mass spectrometry can characterize proteins in gastroduodenal and pancreatic fluids. Our analysis revealed the presence of proteases and protease inhibitors that are enriched in either gastroduodenal or pancreatic fluid. We determined that (1) gastroduodenal and pancreatic fluid can be sequentially collected from the same patient, (2) proteins can be readily extracted by fluid-specific chemical precipitation methods, and (3) proteases and enzyme regulators comprise the major protein functions, as determined by gene ontology annotation, for both gastroduodenal and pancreatic fluid.

Proteomic investigations of human body fluids for clinical applications necessitate the establishment of clear and consistent sample collection and processing methodologies as one of the initial stages in assay development. The effects...
Table 2  Summary of identified proteins

| Sample | Patient # | Total | Non-redundant in sample | Exclusive to particular fluid | Statistically significant enrichment<sup>a</sup> | Total differentially detected |
|--------|-----------|-------|-------------------------|-------------------------------|---------------------------------------------|------------------------------|
| GDF    | 1         | 244   |                        |                               |                                             | 87                           |
|        | 2         | 225   |                        |                               |                                             |                              |
|        | 3         | 206   |                        |                               |                                             |                              |
| PF     | 1         | 350   |                        |                               |                                             | 301                          |
|        | 2         | 351   |                        |                               |                                             |                              |
|        | 3         | 437   |                        |                               |                                             |                              |

GDF, gastroduodenal fluid; IPI #, international protein index number; PF, pancreatic fluid.

<sup>a</sup>Statistically significant Bayes factor > 10. Total differentially detected = exclusive to particular fluid + statistically significant enrichment.

Figure 4  Gene ontology (GO) analysis revealed the molecular function of proteins that were exclusive to or enriched in gastroduodenal or pancreatic fluid. Using Scaffold<sup>3</sup>, the two sets of differentially detected proteins were used for GO classification of molecular function. GDF, gastroduodenal fluid; PF, pancreatic fluid.

of any such variations are especially pronounced in gastroduodenal and pancreatic fluids as a result of the inherent high concentration of active proteolytic enzymes. Previously, we have made efforts to standardize sample handling of these two ePFT-collected fluids.<sup>14–16</sup> Significant changes in the proteomic profile may also be introduced during sample preparation if no consistent methodology is used. In the development of diagnostic and prognostic biomarkers of disease, these procedural artifacts may obscure a potentially significant result or be more likely to identify false positive biomarkers.

To ensure high sample quality, we are particularly careful that our ePFT-collected samples are immediately placed on ice in chilled tubes following collection, promptly centrifuged at 4 °C to remove particulates, and immediately frozen at −80 °C for future analysis. Frozen samples are thawed on ice before desalting and protease inactivation via the precipitation. We have shown previously that protein degradation of ePFT-collected pancreatic fluid is prominent after as little as 30 min at room temperature.<sup>16</sup>

Several proteases and protease inhibitors are found exclusively in gastroduodenal fluid. One such protein, pepsinogen is the precursor of pepsin, an abundant protease secreted by the chief cells in the stomach. The cystatin class of protease inhibitors, specifically cystatins A, D, SA, and SN, is also identified exclusively in gastroduodenal fluid. Cystatins mainly inhibit peptidases belonging to peptide families C1 (papain family) and C13 (legumain family).<sup>32,33</sup> Significantly, cystatin SN has recently been implicated in gastric cancer.<sup>34</sup> Gastricsin, a precursor of gastrin (aspartic protease), has been determined to be enriched in gastroduodenal fluid. Gastricsin is produced in the stomach and is a major component of the gastric mucosa. Elevated levels of gastrin have been associated previously with susceptibility to gastric cancer.<sup>35,36</sup>

Similarly, several proteins were identified exclusively in our pancreatic fluid analysis. Among them are elastase-2B and neprilysin. Elastase-2B has been identified previously in pancreatic fluid from chronic pancreatitis patients, but not from pancreatic cancer and thus is a potential biomarker for differentiating the two diseases.<sup>37</sup> In addition, neprilysin, also known as CD10 or MME, is a metalloprotease that has been associated previously with a variety of signaling peptide cascades and cancer.<sup>38,39</sup> We also identified several common pancreatic protease enzymes enriched in pancreatic fluid including: aminopeptidase N, chymotrypsin C, elastase-3A, trypsin 1, and carboxypeptidase A1.<sup>40</sup> Serpins A11, B4, and D1 are among the protease inhibitor proteins that are identified exclusively in pancreatic fluid, while serpin B6 is enriched in pancreatic fluid. This finding is expected as serpins, a group of similarly structured proteins, are the predominant class of protease inhibitors in pancreatic fluid. The members of the extracellularly secreted serpin family have various functions, including involvement in the proteolytic cascades central to inflammatory responses, blood clotting, and tissue remodeling.<sup>41</sup> Although serpin A11 is uncharacterized, serpin B4 is an inhibitor of cathpsin G and chymase, serpin B4 is an inhibitor of cathpsin G, while serpin D1 is a thrombin inhibitor.<sup>42–44</sup> We suspect this family of protease inhibitors may have a large role in pancreatic disease, and as such, merits further investigation.

It is worth noting, however, that we deem certain proteins as "exclusive" to a particular cohort, as they were not detected with the current methodology and technology. In fact such "exclusive" proteins may be present in the other fluid, albeit at a substantially (several orders of magnitude) lower concentration, due to being below the threshold of detection. Using
the current technology that is limited by attomolar concentration and instrumental sampling, it is not possible to identify every protein present in a sample. In the future, with improvements in depth of proteome coverage, it may be possible to detect particular proteins—which have been currently deemed “exclusive” to one cohort—in both fluids. However, as we are using quantitative techniques, it is expected that such proteins would remain statistically more abundant in the “exclusive” fluid.

Although both gastroduodenal and pancreatic fluids are secreted from the upper gastrointestinal tract and can be collected using the ePFT method, our data emphasize these fluids require different sample preparation techniques for optimal GeLC-MS/MS analysis. The argument could be made that the differences in protein detection patterns that we observed are a result of the different precipitation techniques used. However, our aim is to compare the fluid samples prepared using the technique best suited for maximum protein yield in each fluid type. In essence, the presence of the proteolytic enzymes in the fluids being analyzed precludes the use of identical precipitation conditions in this study.

We have shown that acidification of gastroduodenal fluid (e.g., by TCA) does not prevent protein degradation.14 In fact, concordant with known gastric physiology, the conversion of inactive pepsinogen to the protease pepsin is typically activated by the acidic pH.45 Thus, acidic TCA-based protein precipitation is counterproductive in the case of gastroduodenal fluid, as it may activate gastric protease precursors. Acetone is chosen as our precipitation reagent for gastroduodenal fluid as little protein degradation is evident and protein yield was maximized when compared with other precipitation methods, as shown previously.14 Similarly, using SDS-PAGE analysis we have determined in a prior study that the highest amount of protein could be extracted from pancreatic fluid using TCA.16 TCA precipitation has the

### Table 3
Proteases exclusive to gastroduodenal fluid and pancreatic fluid

| Proteases                        | IPI #           | GDF1 | GDF2 | GDF3 |
|---------------------------------|-----------------|------|------|------|
| **Exclusive to gastroduodenal fluid** |                 |      |      |      |
| Cathepsin E                     | IPI 00025062    | 9    | 5    | 8    |
| Leukocyte elastase              | IPI 00027769    | 3    | 4    | 4    |
| Pepsinogen 3                    | IPI 00736755    | 175  | 196  | 185  |
| Transmembrane protease, serine 11D | IPI 0003542    | 3    | 2    | 7    |
| **Exclusive to pancreatic fluid** |                 |      |      |      |
| Chymotrypsin-like protease CTRL-1 | IPI 00643847   | 11   | 11   | 6    |
| Dipeptidyl peptidase 4          | IPI 00018953    | 2    | 11   | 4    |
| Elastase-2B                     | IPI 0027723     | 64   | 61   | 53   |
| Glutamate carboxypeptidase 2    | IPI 00028514    | 1    | 1    | 1    |
| Glutamyl aminopeptidase         | IPI 0014375     | 3    | 4    | 4    |
| MEP1A protein                   | IPI 0004372     | 10   | 7    | 5    |
| Meprin A beta                   | IPI 00178015    | 6    | 3    | 4    |
| Neprilysin                      | IPI 00247063    | 15   | 19   | 11   |
| Plasminogen                     | IPI 0019580     | 4    | 4    | 3    |
| Prostasin                      | IPI 00329538    | 2    | 3    | 3    |
| Protein disulfide-isomerase A3  | IPI 00025252    | 0    | 1    | 3    |
| Similar to Complement factor B  | IPI 00019591    | 0    | 3    | 6    |
| Xaa-Pro aminopeptidase 2        | IPI 00485344    | 4    | 2    | 5    |

GDF, gastroduodenal fluid; IPI #, international protein index number; PF, pancreatic fluid.

### Table 4
Proteases of statistically significant enrichment in gastroduodenal fluid and pancreatic fluid

| Proteases                        | IPI #           | GDF1 | GDF2 | GDF3 | PF1 | PF2 | PF3 | Bayes factor | Fold change |
|----------------------------------|-----------------|------|------|------|-----|-----|-----|--------------|-------------|
| **Higher abundance in gastroduodenal fluid** |                 |      |      |      |     |     |     |              |             |
| Gastricin                        | IPI 00022213    | 86   | 103  | 94   | 4   | 6   | 1   | 3.9E+08      | 26.58       |
| Kallikrein-1                     | IPI 00304808    | 11   | 17   | 5    | 3   | 2   | 3   | 3.4E+01      | 3.87        |
| **Higher abundance in pancreatic fluid** |                 |      |      |      |     |     |     |              |             |
| Aminopeptidase N                 | IPI 00221224    | 16   | 12   | 7    | 91  | 108 | 90  | 3.0E+07      | 7.56        |
| Protease serine 4 isofrom B      | IPI 00385250    | 8    | 7    | 13   | 48  | 76  | 50  | 7.4E+04      | 6.18        |
| Chymotrypsin B2                  | IPI 00515087    | 48   | 65   | 26   | 149 | 153 | 136 | 1.6E+05      | 3.24        |
| Chymotrypsin-C                   | IPI 00018553    | 16   | 24   | 8    | 60  | 38  | 33  | 1.8E+02      | 2.76        |
| Elastase-2A                      | IPI 00027722    | 75   | 107  | 43   | 186 | 222 | 170 | 1.6E+04      | 2.50        |
| Carboxypeptidase A1              | IPI 00009823    | 142  | 162  | 93   | 255 | 320 | 340 | 7.9E+03      | 2.25        |
| Trypsin-1                        | IPI 00011694    | 82   | 99   | 76   | 143 | 182 | 164 | 1.3E+04      | 1.91        |
| Elastase-3A                      | IPI 00295663    | 49   | 80   | 38   | 128 | 88  | 88  | 3.1E+01      | 1.83        |
| PRSS1 protein                    | IPI 00815665    | 71   | 73   | 76   | 91  | 149 | 138 | 1.0E+02      | 1.69        |

GDF, gastroduodenal fluid; IPI #, international protein index number; PF, pancreatic fluid.
advantages of concentrating and desalting the solution, while simultaneously acidifying it, thereby inactivating pancreatic proteases via denaturation. In addition, lower molecular weight band smears, indicating proteolysis, are minimized when TCA is used compared with the other protein extraction strategies.\(^6\) In the case of pancreatic fluid (generally pH 8-8.5), the decrease in pH resulting from TCA addition successfully precipitates proteins and inhibits the activity of pancreatic proteases. As mentioned above, this effect is in contrast to gastroduodenal fluid, in which gastric enzymes, such as gastrin and pepsin, are active at very acidic pH in which the addition of TCA to gastroduodenal fluid promotes proteolysis.

We aim to overcome several potential limitations to our methodology in future studies. We acknowledge that the secretin-stimulated ePFT-collected gastroduodenal and pancreatic fluid samples are admixtures of several upper gastrointestinal fluids. Regarding the presence of pancreatic proteins in gastroduodenal fluid, such an admixture is an expected result of basal pancreatic secretions before secretin stimulation during ePFT. However, as we have shown, the majority of the proteins identified in gastroduodenal fluid are from the stomach and the duodenum. Likewise, the presence of a nominal amount of gastroduodenal proteins in pancreatic fluid is to be expected, but the contribution of this fluid is minimized by fluid aspiration in the duodenum before ePFT. Moreover, any residual gastroduodenal fluid proteins are subsequently diluted by the protein-rich secretin-stimulated pancreatic secretions, particularly as the sample for proteomic analysis is collected 30 min post-secretin stimulation. Furthermore, duodenal protein secretion is minimal, and potential gastric fluid efflux is decreased by placing the patient in the left lateral decubitus position. As we show herein, such differences are apparent when comparing gastroduodenal and pancreatic fluid both visually by the SDS-PAGE protein banding pattern profile and by qualitative and quantitative mass spectrometry analysis.

In conclusion, using ePFT collection coupled with mass spectrometry, we have identified proteins which are differentially detected in either gastroduodenal or pancreatic fluid. These data obtained using GeLC-MS/MS techniques provide further evidence supporting the feasibility of using ePFT-collected fluid to study specific diseases of the upper gastrointestinal tract. As such, our research team is currently performing a study searching for biomarkers of chronic pancreatitis via a comparative proteomic analysis of secretin-stimulated, ePFT-collected pancreatic fluid. We aim to identify biomarkers that can be traced back to physiological events of the pancreas. Such a study is supported by the premise that the majority of identified proteins are indeed of pancreatic origin, as we illustrate herein. In addition, other studies may be designed to target the role of these proteases and protease inhibitors—such as cystatins and serpins—in gastroduodenal and pancreatic diseases. Further elucidation of differences in the proteomes of each ePFT-collected fluid in diseased and non-diseased patients, may provide a better understanding of the molecular mechanisms leading to the onset and progression of upper gastrointestinal disease. In summary, we have shown that there is indeed a unique proteome in the ePFT-collected secretin-stimulated pancreatic fluid when compared with gastroduodenal fluid. Future studies may exploit such ePFT-based fluid collections to study particular diseases of the upper gastrointestinal tract.

### Table 5 Protease inhibitors exclusive to gastroduodenal fluid and pancreatic fluid

| Protease inhibitor proteins | IPI # | GDF1 | GDF2 | GDF3 | PF1 | PF2 | PF3 |
|----------------------------|-------|------|------|------|-----|-----|-----|
| **Exclusive to gastroduodenal fluid** | | | | | | | |
| Alpha-2-macroglubulin-like protein 1 | IPI 00419215 | 10 | 10 | 5 |
| Cystatin-A | IPI 00032325 | 13 | 7 | 15 |
| Cystatin-D | IPI 00002851 | 0 | 5 | 2 |
| Cystatin-SA | IPI 00132982 | 9 | 14 | 7 |
| Cystatin-SN | IPI 00305477 | 16 | 23 | 3 |
| Lipocalin-1 | IPI 00009650 | 86 | 122 | 55 |
| Uteroglobin | IPI 00006705 | 4 | 3 | 0 |
| **Exclusive to pancreatic fluid** | | | | | | | |
| Alpha-2-antiplasmin | IPI 00879231 | 1 | 1 | 1 |
| Annexin A5 | IPI 00239801 | 3 | 1 | 3 |
| Heat shock protein HSP 90-alpha | IPI 00784295 | 0 | 1 | 4 |
| Histidine-rich glycoprotein | IPI 00023371 | 1 | 2 | 15 |
| Inter-alpha-trypsin inhibitor H1 | IPI 00292530 | 1 | 1 | 3 |
| Isoform HMW of Kininogen-1 | IPI 00303238 | 0 | 1 | 4 |
| Kallistatin | IPI 00328609 | 1 | 0 | 1 |
| Serpin A11 | IPI 00333828 | 0 | 1 | 0 |
| Serpin B4 | IPI 00100303 | 0 | 1 | 37 |
| Serpin D1 | IPI 00292950 | 1 | 0 | 3 |
| Thyroxine-binding globulin | IPI 00292946 | 1 | 1 | 1 |

GDF, gastroduodenal fluid; HMW, high molecular weight; IPI #, international protein index number; PF, pancreatic fluid.

### Table 6 Protease inhibitors of statistically significant enrichment in gastroduodenal fluid and pancreatic fluid

| Protease inhibitor proteins | IPI # | Spectral counts | Bayes factor | Fold change |
|----------------------------|-------|-----------------|-------------|------------|
| **Higher abundance in gastroduodenal fluid** | | | | |
| Annexin A1 | IPI 002168918 | 25 | 46 | 3 | 0 | 0 | 4 | 1.8E+01 | 15.26 |
| Antileukoproteinase | IPI 00008580 | 22 | 19 | 8 | 1 | 2 | 0 | 2.9E+03 | 13.28 |
| Leukocyte elastase inhibitor | IPI 00027444 | 10 | 27 | 17 | 8 | 8 | 8 | 8.3E+00 | 2.13 |
| **Higher abundance in pancreatic fluid** | | | | |
| Annexin A4 | IPI 00872780 | 0 | 1 | 0 | 6 | 10 | 15 | 4.4E+02 | 19.39 |
| Alpha-2-macroglobulin | IPI 00478003 | 9 | 9 | 5 | 161 | 145 | 169 | 1.0E+10 | 16.83 |
| Antithrombin-III | IPI 00032179 | 0 | 1 | 8 | 9 | 13 | 23 | 3.1E+01 | 5.38 |
| Serpin B6 | IPI 00749398 | 4 | 2 | 2 | 11 | 6 | 14 | 3.2E+01 | 3.77 |
| Annexin A2 isoform 1 | IPI 001418169 | 9 | 10 | 5 | 18 | 11 | 32 | 1.1E+01 | 2.34 |

GDF, gastroduodenal fluid; IPI #, international protein index number; PF, pancreatic fluid.
CONFLICT OF INTEREST
Guarantor of the article: Joao A. Paulo, PhD.
Specific author contributions: J.P. and V.K. carried out the experiments and drafted the original manuscript. J.P., J.B., H.S., and D.C. conceived of the study, and participated in its design and coordination. All authors helped to draft the manuscript and approved the final manuscript.

Financial support: Funds were provided by the following NIH grants: 1 F32 DK085835-01A2 (J.P.), 1 R21 DK081703-01A2 (D.C.) and 5 P30 DK034854-24 (Harvard Digestive Diseases Center; D.C.).

Potential competing interests: None.

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). In addition, we would like to thank the Burrill family for their generous support through the Burrill Research Grant.

We also would like to thank members of the Steen Laboratory at Children’s Hospital Boston, in particular John FK Sauld, Ali Ghoulidi, Aleksander Gaun, and Dominic Winter for 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 Jessica Rosenblum for her technical assistance.

Study Highlights

WHAT IS CURRENT KNOWLEDGE
- ePFT allows for the safe collection of gastroduodenal and pancreatic fluid from the duodenum.
- ePFT-collected secretions are rich in proteins.
- Mass spectrometry-based proteomics can be used to identify proteins from body fluids.

WHAT IS NEW HERE
- Gastroduodenal and pancreatic fluid can be sequentially collected from the same patient.
- Proteins can be readily extracted and identified from gastroduodenal and pancreatic fluid by fluid-specific chemical precipitation methods.
- Proteases and enzyme regulators comprise the major molecular functions of the identified proteins for both gastroduodenal and pancreatic fluid.

1. Lam KWK, Lo SCL. Discovery of diagnostic serum biomarkers of gastric cancer using proteomics. Proteomics Clin Appl 2008; 2: 219–226.
2. Mariotto AB, Yabroff KR, Shao Y et al. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst 2011; 103: 117–128.
3. James S. Opportunities and challenges at NIDDK in digestive diseases research. Gastroenterology 2007; 132: 1219–1220.
4. Issaq HJ, Xiao Z, Veenstra TD. Serum and plasma proteomics. Chem Rev 2007; 107: 3601–3620.
5. McGill JNM, Basavappa S, Gethye TW et al. Secretin activates Cl- channels in bile duct epithelial cells through a cAMP-dependent mechanism. Am J Physiol 1994; 266 (4 Pt 1): G731–G736.
6. Gray MA, Greenwell JR, Argent BE. Secretin-regulated chloride channel on the apical plasma membrane of pancreatic duct cells. J Membr Biol 1988; 105: 131–142.
7. Wu B, Conwell DL. The endoscopic pancreatic function test. Am J Gastroenterol 2009; 104: 2381–2383.
8. Conwell DL, Zuccaro Jr G, Vargo JJ et al. An endoscopic pancreatic function test with synthetic porcine secretin for the evaluation of chronic abdominal pain and suspected chronic pancreatitis. Gastroenterol Endosc 2003; 37: 37–40.
39. Ehruma M, Köbel M, Mustafa T et al. Expression of neutral endopeptidase (NEP/CD10) on pancreatic tumor cell lines, pancreatitis and pancreatic tumor tissues. Int J Cancer 2007; 120: 2393–2400.
40. Whitcomb DC, Lowe ME. Human pancreatic digestive enzymes. Dig Dis Sci 2007; 52: 1–17.
41. Gettins PG. Serpin structure, mechanism, and function. Chem Rev 2002; 102: 4751–4804.
42. Schick C, Kamachi Y, Bartuski AJ et al. Squamous cell carcinoma antigen 2 is a novel serpin that inhibits the chymotrypsin-like proteinases cathepsin G and mast cell chymase. J Biol Chem 1997; 272: 1849–1855.
43. Scott FL, Hirst CE, Sun J et al. The intracellular serpin proteinase inhibitor 6 is expressed in monocytes and granulocytes and is a potent inhibitor of the azurophilic granule proteinase, cathepsin G. Blood 1999; 93: 2089–2097.
44. Vicente CP, He L, Pavão MS et al. Antithrombotic activity of dermatan sulfate in heparin cofactor II-deficient mice. Blood 2004; 104: 3965–3970.
45. Norris SH, Hersey SJ. pH dependence of pepsinogen and acid secretion in isolated gastric glands. Am J Physiol 1983; 245: G730–G738.

Supplementary Information accompanies this paper on the Clinical and Translational Gastroenterology website (http://www.nature.com/ctg)