

Comparative Proteomic Analysis of Whole-Gut Lavage Fluid and Pancreatic Juice Reveals a Less Invasive Method of Sampling Pancreatic Secretions

Jana M. Rocker, PhD¹, Marcus C. Tan, MD^{1,2}, Lee W. Thompson, MD³, Carlo M. Contreras, MD⁴, Jack A. DiPalma, MD⁵ and Lewis K. Pannell, PhD¹

OBJECTIVES: There are currently no reliable, non-invasive screening tests for pancreatic ductal adenocarcinoma. The fluid secreted from the pancreatic ductal system (“pancreatic juice”) has been well-studied as a potential source of cancer biomarkers. However, it is invasive to collect. We recently observed that the proteomic profile of intestinal effluent from the bowel in response to administration of an oral bowel preparation solution (also known as whole-gut lavage fluid, WGLF) contains large amounts of pancreas-derived proteins. We therefore hypothesized that the proteomic profile is similar to that of pancreatic juice. In this study, we compared the proteomic profiles of 77 patients undergoing routine colonoscopy with the profiles of 19 samples of pure pancreatic juice collected during surgery. **METHODS:** WGLF was collected from patients undergoing routine colonoscopy, and pancreatic juice was collected from patients undergoing pancreatic surgery. Protein was isolated from both samples using an optimized method and analyzed by LC-MS/MS. Identified proteins were compared between samples and groups to determine similarity of the two fluids. We then compared our results with literature reports of pancreatic juice-based studies to determine similarity.

RESULTS: We found 104 proteins in our pancreatic juice samples, of which 90% were also found in our WGLF samples. The majority (67%) of the total proteins found in the WGLF were common to pancreatic juice, with intestine-specific proteins making up a smaller proportion.

CONCLUSIONS: WGLF and pancreatic juice appear to have similar proteomic profiles. This supports the notion that WGLF is a non-invasive, surrogate bio-fluid for pancreatic juice. Further studies are required to further elucidate its role in the diagnosis of pancreatic cancer.

Clinical and Translational Gastroenterology (2016) 7, e174; doi:10.1038/ctg.2016.27; published online 26 May 2016

Subject Category: Pancreas and Biliary Tract

INTRODUCTION

Pancreatic ductal adenocarcinoma is now the third leading cause of cancer-related death in United States, with 5-year survival rates of just 6%.¹ This poor prognosis is mostly due to advanced stage of disease at initial diagnosis. 85% of patients present with inoperable, usually metastatic disease, with median survival less than a year.² However, the time for pancreatic ductal adenocarcinoma to develop from the first dysplastic cell to frank malignancy is up to 20 years,³ suggesting that an early detection test may lead to more patients being diagnosed with earlier stage disease.

A biomarker is a molecule found in bodily fluids or tissues that reflects a physiologic response to a condition, and many in current clinical use are proteins. These can be used to diagnose a disease or to judge the effectiveness of a treatment.^{4,5} Mass spectrometry is one method used to identify the proteins that are differentially expressed between healthy and diseased patients and these have the potential to become biomarkers of that condition in a clinical test.⁴

Localized body fluids are ideal for biomarker studies.

Most biomarker studies use serum as the target fluid. The advantages of serum are that it is relatively non-invasive to collect and contains proteins derived from the entire body. This last advantage is also its principal disadvantage: serum is a very dilute and non-specific source of organ-specific proteins. Highly abundant proteins such as albumin can mask lower level proteins of interest, often requiring enrichment strategies to find biomarkers. Targeted, compartmentalized body fluids are a more attractive target for biomarker studies and represent a more concentrated source of proteins derived from local tissues. Since pancreatic adenocarcinoma arises from that organ's ductal epithelial lining, pancreatic ductal fluid (“pancreatic juice”) is the ideal fluid for studying this malignancy.

Pancreatic juice is difficult to collect.

Pancreatic juice is generally collected by suction directly from the pancreatic duct during surgery or endoscopic retrograde cholangio-pancreatography.^{6–11} Endoscopic retrograde cholangio-pancreatography-based suction from the pancreatic duct is

¹Proteomics and Mass Spectrometry Laboratory, University of South Alabama Mitchell Cancer Institute, Mobile, Alabama, USA; ²Department of Surgery, University of South Alabama, Mobile, Alabama, USA; ³Cancer Surgery of Mobile, PC, Mobile, Alabama, USA; ⁴Department of Surgery, University of Alabama at Birmingham, Birmingham, Alabama, USA and ⁵Division of Gastroenterology, University of South Alabama, College of Medicine, Mobile, Alabama, USA

Correspondence: Jana M. Rocker, PhD, University of South Alabama, Mitchell Cancer Institute, 1660 Springhill Ave, Mobile, Alabama 36604, USA.

E-mail: jrocker@health.southalabama.edu

Received 13 August 2015; accepted 5 April 2016

invasive and may cause severe pancreatitis.¹² Because of this, sample sizes are usually limited and it is difficult to obtain samples from healthy controls when the procedure is not medically indicated. These factors make pancreatic juice a less than ideal fluid for a screening test. Previous studies of pancreatic cancer using pancreatic juice have between 1 and 11 cases in each group, with controls being patients receiving endoscopic retrograde cholangiopancreatography for ultimately benign conditions.^{6,9–11,13} In an effort to address the lack of data of normal pancreatic juice, Doyle *et al.*⁸ published the proteome of normal pancreatic juice, taken from three female patients aged 29–32 years undergoing endoscopic retrograde cholangiopancreatography for abdominal symptoms but with no apparent pancreatic pathology.

Colonoscopy effluent as a source of pancreatic secretions. As the pancreas secretes ~1 l of pancreatic juice from the pancreatic duct into the bowel daily,¹⁴ it can be expected that the colon will also contain proteins that originate from the pancreas and we have found this to be true in colonoscopy effluent. The process of whole-gut lavage using a purgative solution prescribed for colonoscopy has been used in the past to examine various specific protein markers in many gastrointestinal conditions.¹⁵ The purgative causes an influx of water into the bowel and the resulting effluent is a lavage of the entire gut. This fluid, referred to as whole-gut lavage fluid (WGLF), can be collected either non-invasively by means of a toilet receptacle or during colonoscopy via endoscopic suction. Collection during colonoscopy is ideal for a potential screening test because it can be done during routine colonoscopy by attaching a trap to the endoscope during the initial suction of waste fluid from the rectum.

WGLF is more effective for the study of pancreatic secretions than fecal sampling for several reasons. Fecal detection of pancreatic proteins in the colon is compromised by degradation by other stool components and bacteria.¹⁶ In addition, fecal material has widely variable transit times through the colon between individuals, which leads to variability in both concentration and time exposed to these sources of degradation. Bacterial populations are significantly reduced in WGLF leading to reduced amounts of protein deterioration.^{17,18} There is also less interference from food particles in WGLF than in stool and it is easier to handle in the laboratory than solid fecal samples.^{19–21} Transit time through both the small bowel and the colon are also both significantly reduced after a bowel preparation solution is given, and the administration of a bowel preparation reduces age-related variability in gut transit.^{22,23} All of these factors make WGLF a more attractive target than solid fecal material.

To our knowledge there have been no prior mass spectrometry-based proteomic studies of WGLF. We developed a novel sample preparation method to analyze the protein from WGLF by liquid chromatography- tandem mass spectrometry (LC-MS/MS) and preliminary data showed that the majority of the proteins were pancreatic, not colonic, in origin. Our hypothesis is that the WGLF proteome is similar to the pancreatic juice proteome. To compare the proteomes of pancreatic juice with WGLF, pancreatic juice samples were collected directly from the pancreatic duct from patients during surgery and also analyzed by LC-MS/MS. Our findings were

then compared to select pancreatic juice proteomes in the literature.^{6,8,24}

METHODS

This study was exempted after Institutional Review Board review because WGLF is a waste product that is routinely discarded during colonoscopy and the pancreatic fluid was collected during surgery that would otherwise be discarded. In addition, no identifying patient information was accessible to the investigators.

WGLF sample collection. Colonoscopy effluent was collected from 149 patients undergoing colonoscopy at the University of South Alabama Gastrointestinal laboratory from November 2011 to November 2013. Patients were administered an oral sulfate solution (SuPrep, Braintree Laboratories, Braintree, MA) prior to colonoscopy as per manufacturer's instructions. Samples were collected from patients with normal colonoscopies ($n=77$) and patients found to have one or more colon polyps ($n=72$). Patients with colorectal cancer, inflammatory bowel disease, celiac disease, human immunodeficiency virus, hepatitis C, acute colitis, *Helicobacter pylori* infection, bowel resection, or gastric bypass surgery were excluded. Basic demographic and sample information (age, sex, race, and bowel preparation quality) for these samples are given in Supplementary Tables S1 and S2.

To obtain the WGLF samples, ~30 ml of effluent was suctioned from the rectum using the endoscope at the start of the colonoscopy procedure into a specimen trap and transferred to a 50 ml conical tube containing a pulverized protease inhibitor tablet (cOmplete Protease Inhibitor tablets, Roche Diagnostics, Indianapolis, IN). The tablet was crushed prior to sample collection to ensure rapid dissolution. The samples were then sealed and inverted to ensure the dispersal of the protease inhibitor into solution. Samples were transported to the laboratory on ice and immediately processed.

Pancreatic juice sample collection. Pancreatic juice samples were collected by Drs Lee Thompson and Russell Brown from Cancer Surgery of Mobile and Carlo Contreras from the University of South Alabama Mitchell Cancer Institute from pancreas tissue during pancreatic surgery. Approximately 0.2–5 ml of pancreatic juice was collected by syringe directly from the pancreatic duct of 19 patients undergoing pancreatic surgery for pancreatic lesions. Diagnoses of the patients who were available are given in Table 1. Liquid (tablet dissolved in water) protease inhibitor solution was immediately added to the samples at a concentration equivalent to WGLF samples described above. Pancreatic juice samples were rapidly transported to the laboratory on ice after collection where they were immediately processed.

Sample processing. Samples were centrifuged for 25 min at $120\times g$ to pellet large debris. Supernatants were then centrifuged for 25 min at $11,000\times g$ to pellet smaller particulates and bacteria. All centrifugation steps were performed at 4 °C. The supernatant was divided into aliquots of 1.8 ml and stored as aliquots at $-80\text{ }^{\circ}\text{C}$ until analysis.

Table 1 Pancreatic pathologies (where known) of 19 pancreatic juice samples taken

| Index | Sample ID | Operation | Size (cm) | Histology | T stage | N stage |
|-------|-----------|-----------------------|-------------------|----------------|---------|---------|
| 1 | 29 | Whipple | 4.2 | Adenocarcinoma | T3 | N1 |
| 2 | 30 | Whipple | Unable to measure | Adenocarcinoma | T3 | N0 |
| 3 | 37 | Whipple | 3.5 | Adenocarcinoma | T3 | N0 |
| 4 | 44 | Distal pancreatectomy | 5.2 | Adenocarcinoma | T3 | N0 |
| 5 | 47 | Unknown | | Adenocarcinoma | | |
| 6 | 48 | Unknown | | IPMN | | |
| 7 | 53 | Unknown | | Unknown | | |
| 8 | 55 | Unknown | | Unknown | | |
| 9 | 61 | Unknown | | Unknown | | |
| 10 | 69 | Unknown | | Unknown | | |
| 11 | 70 | Whipple | 3.5 | Adenocarcinoma | T3 | N1 |
| 12 | 74 | Unknown | | Unknown | | |
| 13 | 75 | Unknown | | Unknown | | |
| 14 | 76 | Whipple | 4.5 | Adenocarcinoma | T4 | N1 |
| 15 | 83 | Whipple | 2.5 | Adenocarcinoma | T3 | N0 |
| 16 | 84 | Whipple | 2.2 | Adenocarcinoma | T3 | N0 |
| 17 | 85 | Whipple | 2.7 | Adenocarcinoma | T3 | N1 |
| 18 | 90 | Whipple | 4.5 | Adenocarcinoma | T3 | N1 |
| 19 | 94 | Whipple | 2.5 | Adenocarcinoma | T3 | N1 |

To remove lipids and polyethylene glycol that interfere with LC-MS analysis, 200 μ l of sample was extracted three times with 1 ml of chloroform. After the final extraction, the sample was centrifuged at 16,100 \times g in a tabletop microcentrifuge (Eppendorf International, Hauppauge, NY) for 5 min and 100 μ l was removed from the top of the sample and placed in a new 1.5 ml microfuge tube. This step was incorporated into the procedure in order to allow the method to work with WGLF induced by both oral sulfate and polyethylene glycol based bowel preparations if needed. In addition, the protease inhibitor tablets used for sample stabilization contain small amounts of polyethylene glycol as a stabilizer, which is removed by this chloroform extraction step. The proteins were then precipitated from the solution using the methanol/chloroform/water precipitation method of Wessel and Flugge.²⁵ Briefly, 400 μ l of methanol was added to the 100 μ l of extracted sample, vortexed and spun for 30 s in a Galaxy mini microcentrifuge (VWR International, Atlanta, GA). Then 200 μ l of chloroform was added and the mixture was vortexed and again spun for 30 s. Next, 300 μ l of water was added and the mixture was vortexed and centrifuged at 16,100 \times g for 5 min. After centrifugation, the protein is located at the interface between the water and the chloroform layers. The water layer was then removed (750 μ l) without disturbing the protein layer and 300 μ l of methanol was added and the sample centrifuged at 16,100 \times g for 10 min in order to pellet the protein precipitate. The methanol/chloroform solution was decanted and the protein pellet was dried by centrifugal lyophilization in a speed vac (Savant, ThermoFisher Scientific, Waltham, MA) for 10 min.

At this point, the pancreatic juice samples were subjected to an extra step and the proteins were further purified on a C-2 solid phase extraction column (Sep-Pak, Waters, Milford, MA). Columns were run on a vacuum manifold and all washes were three column volumes. The column was first washed with acetonitrile (ACN; Fisher Scientific, Pittsburgh, PA), followed by equilibration with 0.1% trifluoroacetic acid (TFA; Fisher

Scientific, Pittsburgh, PA). The sample was loaded and desalted in 0.1% TFA. Proteins were eluted with 50% ACN and this fraction was collected and dried by centrifugal lyophilization in the speed vac. This extra step was added due to the fact that the pancreatic juice samples caused blockage of the MS spray tip unless purified further.

Dried pellets of both pancreatic juice and WGLF proteins were resuspended in 20 μ l of a solution containing 8 M urea (Sigma-Aldrich, St Louis, MO), 10 mM tris-carboxyethyl phosphine (Alfa-Aesar, Ward Hill, MA), 5 mM EDTA (Sigma-Aldrich, St Louis, MO) and 0.1 M ammonium bicarbonate (Sigma-Aldrich, St Louis, MO). Once dissolved, 60 μ l of 50 mM ammonium bicarbonate/10 mM tris-carboxyethyl phosphine was added to dilute the mixture to 2 M urea and the proteins were digested with 2 μ l of 10 μ M trypsin (Sequencing-grade modified porcine trypsin, Promega, Madison, WI) while shaking overnight at 37 $^{\circ}$ C and 600 r.p.m.

Post digestion, samples were centrifuged at 16,100 \times g for 15 min at 4 $^{\circ}$ C in the tabletop microcentrifuge (Eppendorf International) to pellet any insoluble debris. Then, 75 μ l of digested sample was removed from the supernatant of the digest and diluted with 20 μ l of water in a snap-top auto sampler vial. Seventy-five microliters were injected onto a C18 pre-column (5 μ m; 5 by 0.3 mm Zorbax, Agilent Technologies, Santa Clara, CA) connected to an Agilent 1200 HPLC with an auto sampler (a main column was not used). Solvent A was 3% ACN and 0.05% TFA in water, and solvent B consisted of 3% water and 0.05% TFA in ACN. A flow rate of 200 μ l/min was maintained throughout the run and the ultraviolet absorbance at 280 nm was monitored during the entire sample run to allow estimation of protein content in the samples. For the first 13 min, 2% solvent B was used to load the sample onto the C18 pre-column and wash it free from salts, buffers, and urea. From time 14 to 21 min, the peptides were eluted from the column with a step gradient and isocratic hold at 40% B. This was followed by a column wash with 90% B from time 22 to 30 min and re-equilibration to 2% B in the final 5 min. The entire run time was 35 min. The A280 (280 nm) peak area of

the eluted peptide peak from 14 to 21 min was used as an estimate of protein concentration. The eluted peptide fraction was dried in a speed vac and stored at -80°C until analysis by mass spectrometry.

Immediately prior to LC-MS/MS analysis, samples were resuspended in a volume of 0.1% TFA dependent upon the protein concentration in order to normalize total protein concentrations between samples. A single 0.5 μl test injection was performed prior to triplicate injections described below in order to further optimize the amount of protein injected into the LC-MS/MS in the final analysis run and ensure consistent intensities between samples.

Mass spectrometry (LC-MS/MS). Samples were injected in triplicate into an Agilent 1200 series nano-liquid HPLC coupled to a linear ion trap/Orbitrap hybrid MS (LTQ-Orbitrap XL, ThermoFisher Scientific, Waltham, MA). The HPLC mobile phases consisted of 3% ACN and 0.2% formic acid in water (solvent A), and 3% water and 0.2% formic acid in ACN (solvent B). A flow rate of 4 $\mu\text{l}/\text{min}$ of 5% solvent B was used to load the sample onto a C18 pre-column (5 μm ; 5 by 0.3 mm Zorbax, Agilent Technologies, Santa Clara, CA), and a flow rate of 1 $\mu\text{l}/\text{min}$ was used to elute the sample from the pre-column onto a separating Hypersil Gold C18 chromatography column (30 mm by 0.18 mm; ThermoFisher Scientific, Waltham, MA). The linear solvent gradient was slowly ramped to 40% B over 70 min in order to elute the peptides from the column and then ramped to 90% B over the final 20 min to wash the column. The total run time (pre-column and resolving chromatography) for each sample injection was 2 h. During the 70 min peptide elution, peptides were ionized by electrospray ionization and analyzed by the LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher Scientific) for MS analysis. Samples were analyzed in data dependent mode with full MS¹ scans from 400–2,000 m/z in the Orbitrap mass analyzer at a resolution of 60,000. The machine acquired data at a speed sufficient to ensure that at least 10 points were collected across each peak for accurate integration. The top five multiply charged ions per scan were selected for MS² scans with CID fragmentation in the LTQ mass analyzer. Selected ions were then added to a dynamic exclusion list to prevent repeated sampling of abundant peptides for 120 s following the initial fragment selection. XCalibur software (version 2.0.7, ThermoFisher Scientific) was used to generate RAW files of each MS experiment.

Protein identification and quantification. Raw files were processed with the Proteome Discoverer software package (Thermo Scientific, version 2.1.0.81) using the following processing workflow: Spectrum Files \rightarrow Spectrum Selector \rightarrow Sequest HT \rightarrow Target/Decoy PSM Validator \rightarrow Event Detector \rightarrow Precursor Ion area detector. Database searching was performed in Proteome Discoverer with the Sequest search engine against a custom database created by combining the non-redundant NCBI RefNCBI human database (version 120814) with the human immunoglobulin A, G, and M, lambda, and kappa constant region sequences from the SwissProt database (28 November 2012). The custom database contains 71485 sequences and 45902068

residues. Searches were performed using the following parameters: semispecific trypsin, dynamic modifications oxidation (M) and Deamidation (N,Q), precursor ion mass tolerance of 10 p.p.m., fragment ion mass tolerance of 0.6 Da, and allowance for up to two missed cleavages. Possible identifications were filtered to give a 1% false discovery rate (FDR). As a length cutoff, only peptides with at least six amino acids were used for identification. Six was used because most peptides below this length will not be retained on the pre-column and there are large numbers of false hits for peptides shorter than six amino acids in length even at higher scores.²⁶ Proteins were quantified using label free quantification of peak areas, averaging all peptides identified with high confidence to give a relative protein abundance value.

Data analysis. The WGLF ($n=149$) and pancreatic juice ($n=19$) samples were analyzed separately within Proteome Discoverer. Data from the three replicates of each individual sample were combined into 149 and 19 experiment files for WGLF and pancreatic juice, respectively. Presence of the protein was determined by the detection of at least three peptides.

Proteome Discoverer produces a spreadsheet which includes the names of the proteins identified and their individual intensities, sequence coverage and numbers of peptides identified in each sample. The percentage of individuals in which the protein was found was calculated by dividing the number of samples with the protein present by the total number of samples in each group. Total intensity in the WGLF and pancreatic juice samples was calculated by summing all intensities within sample group. The average intensity of each protein was calculated by summing its intensity across all samples in each group and dividing by the number of samples in that group. The average percent of the intensity of each protein in the WGLF and pancreatic juice groups was calculated by dividing its average intensity by the total intensity in each group.

Proteins found in common with WGLF and pancreatic juice were compared between the WGLF samples of patients with colon polyps ($n=72$) with those of the normal controls ($n=77$) in order to determine if the presence of polyps has any effect on the results. Data on bowel prep quality was also collected on these patients. Bowel prep quality was rated as "excellent", "good", "fair", "poor", or "none" by the physician at the time of colonoscopy and patients with "poor" or "none" ratings were excluded. Protein intensities were also statistically compared between bowel preparation qualities to determine whether an effect could be seen.

Intensities of each protein across individual samples were log transformed and statistical significance was determined without correction for multiple comparisons, with $\alpha=0.01$. Each row was analyzed individually, without assuming a consistent s.d. Missing values were replaced with zeroes. Analysis was performed using Graphpad Prism 6.05 (Graphpad Software, La Jolla, CA).

Ontology. Origins of proteins in WGLF and pancreatic juice samples were obtained using the DAVID Bioinformatics Tools (version 6.7)^{27,28} and the human protein atlas.²⁹

RESULTS

Comparison of WGLF and pancreatic juice proteomes.

There were 151 unique proteins identified: Of these, 94 were present in both WGLF and pancreatic juice, 10 were present in pancreatic juice only, and 47 were present in WGLF only (Figure 1). In proteomics, MS protein intensity is a measure of relative protein abundance. The total intensity of identified proteins in the WGLF group and pancreatic juice groups were similar, 6.52×10^9 and 4.19×10^9 ion counts, respectively. Ninety four percent of the proteins present in pancreatic juice (by total intensity) were present in WGLF (Figure 2). Furthermore, proteins present in pancreatic juice (mostly pancreatic enzymes and antibody fragments) comprise 87% of the protein mass in WGLF (Figure 3).

Most proteins were present in both WGLF and pancreatic juice.

Origins of proteins were determined using DAVID^{27,28} and the human protein atlas.²⁹ A breakdown of the total intensities of proteins originating from the pancreas, intestine, blood, immune system (antibodies), epithelial cells, and other sources in both pancreatic juice and WGLF are shown in

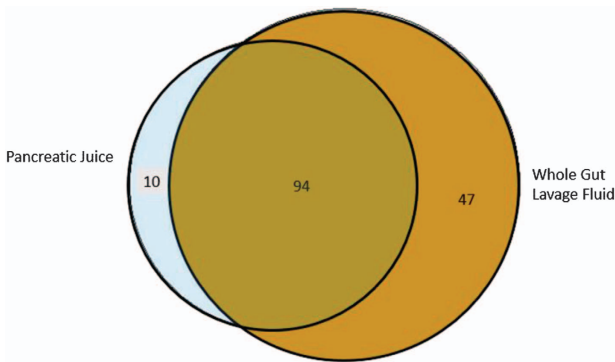


Figure 1 Venn diagram depicting the number of proteins positively identified with three or more peptides in whole-gut lavage fluid (WGLF), pancreatic juice, and the overlap between the two.

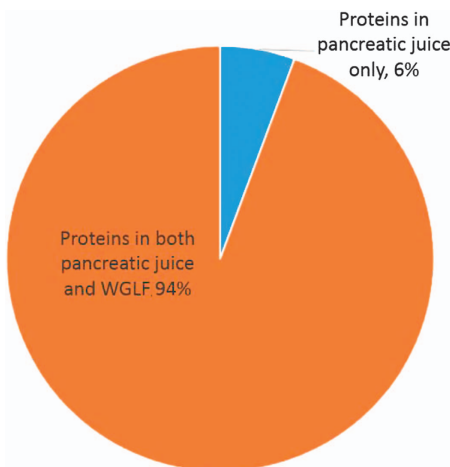


Figure 2 Percentage of protein by intensity in pancreatic juice that is present in pancreatic juice only, and in both whole-gut lavage fluid (WGLF) and pancreatic juice.

Figure 4. Comparing the mean abundance of the pancreatic derived proteins between WGLF and pancreatic juice reveals similar expression patterns between the two samples as shown in Figure 5.

The proteins present in pancreatic juice that were not seen in WGLF were mostly blood/plasma proteins. Proteins found in only WGLF were, as expected, mostly derived from the intestine with some from plasma and some from the liver and stomach.

Listings of all proteins by name including the accession number are given in Supplementary Tables S3–S6. The following additional information is also included: the origin of the protein, the percentage of samples in which the protein was present, the average number of total and unique peptides per sample, the average sequence coverage per sample, molecular weight, and sequence length for the proteins common between pancreatic juice and WGLF as well as those found in pancreatic juice and WGLF only groups are given in Supplementary Table S3–S6.

Proteins exclusive to WGLF.

The majority of the proteins exclusive to WGLF are intestinal in origin. Others include gastric (gi530396601 gastric intrinsic factor) proteins, also not surprising since WGLF is a whole-gut lavage. These proteins make up a small portion of the WGLF proteome by intensity.

Proteins present only in pancreatic juice.

The two pancreas-specific proteins seen in pancreatic juice and

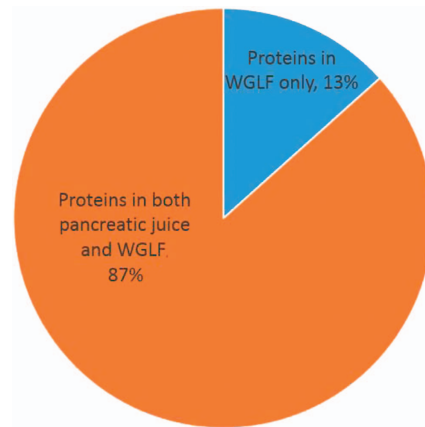


Figure 3 Percentage of protein by intensity in whole-gut lavage fluid (WGLF) that is present in WGLF only and in both WGLF and pancreatic juice.

Intensities of proteins by origin detected in both pancreatic juice and WGLF

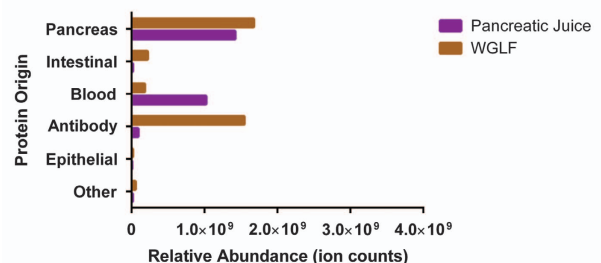


Figure 4 Intensities of proteins by origin of proteins found in both pancreatic juice and whole-gut lavage fluid (WGLF) compared.

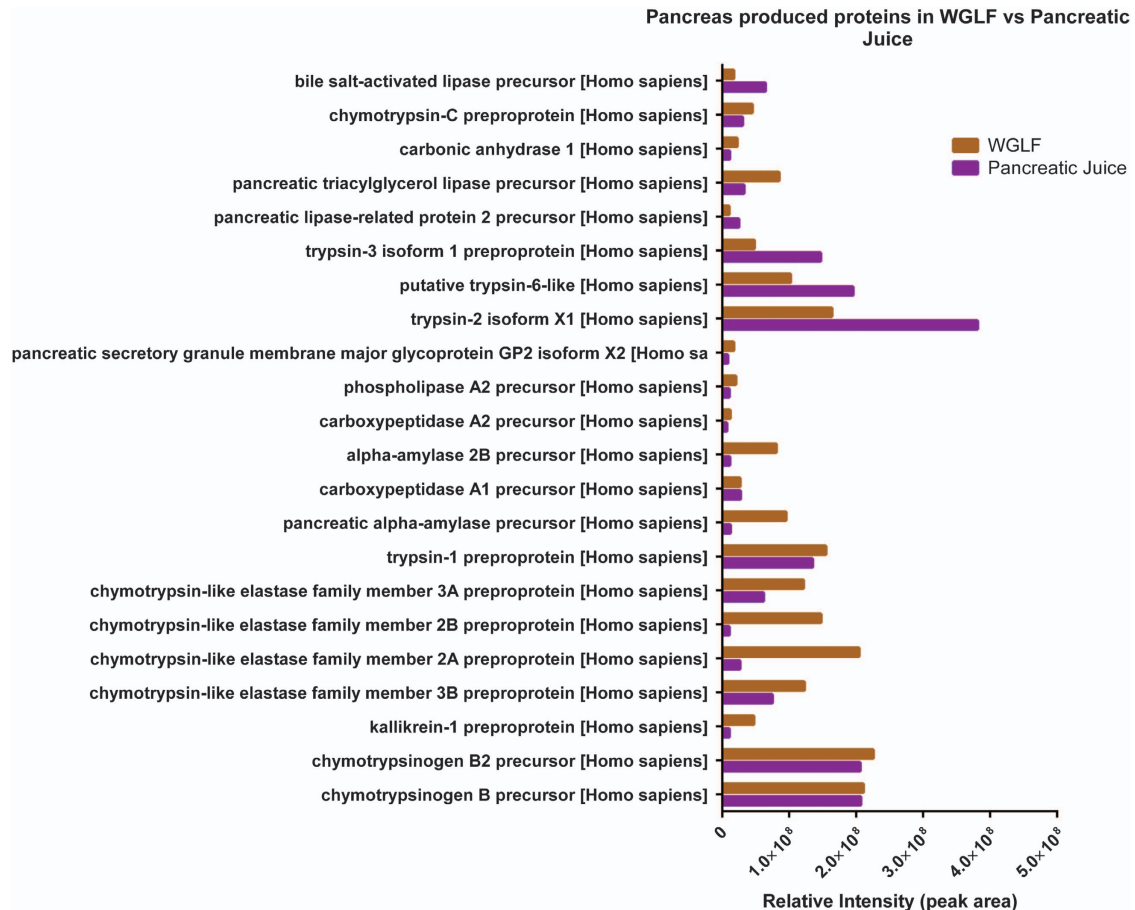


Figure 5 Average relative abundance of pancreatic exocrine enzymes compared between whole-gut lavage fluid (WGLF) and pancreatic juice groups.

not WGLF were gi122937329(syncollin) and gi38348213 (regenerating islet-derived protein 3-gamma). These were seen in 53 and 11% of pancreatic juice samples, respectively. When compared with the literature, syncollin was only seen in two published pancreatic juice proteomes^{7,24} and not in an ePFT-collected fluid proteome.³⁰ Regenerating islet-derived protein 3-gamma was not seen in any of the previously published proteomes. These are low intensity proteins, so it is possible that they are simply missed in the WGLF samples.

The other proteins exclusive to pancreatic juice were plasma and intracellular proteins. The presence of more plasma proteins is not surprising, as many of the pancreatic juice samples collected had visible hemolysis. The invasive nature of the collection causes bleeding, which can contaminate the sample and make study more difficult. WGLF contains less plasma proteins and is dominated by pancreas-specific proteins, making these samples easier to study.

Effect of polyps and suboptimal bowel preparation on pancreatic proteins in WGLF. The 94 proteins that WGLF had in common with pancreatic juice were compared statistically between WGLF obtained from healthy controls ($n=72$) and from patients found to have adenomatous polyps in the colon at colonoscopy ($n=77$). There were no statistically significant differences ($P<0.01$) between the groups when compared using multiple t -tests. Out of 149

patients, 118 (79%) had “excellent” quality bowel preparations, 18 (12%) had “good”, seven had “fair”, and six were not specified. There were two proteins that were statistically significantly different between the “excellent” and “good” groups, these were deleted in malignant brain tumors 1 ($P=0.002$) and leukocyte elastase inhibitor ($P=0.008$). Between the “excellent” and “fair”, and “good” and “fair” groups, there were no differences that reached statistical significance.

Literature comparison. We also compared our WGLF and pancreatic juice proteomic data with previous literature reports of pancreatic juice proteomes.^{6,8,24} On average, our pancreatic juice proteome matched 40% and our WGLF proteome matched 39% of the pancreatic juice proteins reported in previous studies. A summary of our comparisons between our WGLF and our pancreatic juice data with reports in pancreatic juice by Gronborg,²⁴ Chen,⁶ and Doyle⁸ is included in Table 2.

DISCUSSION

To our knowledge, this is the first published analysis of the proteome of WGLF. We developed a novel sample preparation method to facilitate the LC-MS analysis of colonoscopy effluent (WGLF) that causes no deterioration of the LC columns and is suitable for the analysis of individual patient

Table 2 Comparison of previous studies of pancreatic juice by number of proteins found, number of samples, method used, and percentage of proteins we found in our pancreatic juice and whole-gut lavage fluid (WGLF) samples that matched their findings

| Pancreatic Juice Study | # Of proteins in the study | # Of samples | Method | % Present in our WGLF Samples | % Present in our pancreatic juice samples |
|--|----------------------------|-------------------------------------|---|-------------------------------|---|
| Gronberg <i>et al.</i> , ²⁴ | 171 | 3 PDAC | ERCP Collect, fast spin, no inhibitor, 1 DE | 35% | 38% |
| Chen <i>et al.</i> , ⁶ | 136 | Pool:10 Benign Cysts, 1 CP, 1 PDAC | ERCP Collect, no spin, PMSF inhibitor, ICAT labeling | 49% | 46% |
| Doyle <i>et al.</i> , ⁸ | 116 | 3 Patients with normal ERCP results | ERCP collect, no spin, no inhibitor, in-solution digest | 35% | 33% |

samples. The method consists of several steps, all of which are necessary to remove potential contaminants within the samples. Although this study used oral sulfate-induced bowel preparation-induced samples, this method is valid for polyethylene glycol induced samples (data not shown).

We have clearly demonstrated that the vast majority (94%) of pancreatic juice proteins are also present in WGLF. Most pancreas-derived proteins are digestive enzymes, and we show that the relative intensities of these enzymes in WGLF are nearly identical to those in pancreatic juice.

These results support our hypothesis that WGLF and pancreatic juice have similar proteomic profiles, supporting the notion that the former may be a reasonable surrogate for the latter. Therefore, WGLF may be a suitable bio-fluid for the evaluation of pancreatic disease. In pancreatic ductal adenocarcinoma, pancreatic exocrine insufficiency is common;³¹ however, clinical symptoms of pancreatic exocrine insufficiency occur late, after about 90% of enzyme production has been lost.³² Therefore, the ability to detect and quantify enzyme loss may be useful in early detection of malignant or progressive pre-malignant disease before the onset of symptoms.

WGLF may be useful for the development of a non-invasive screening test for pancreatic cancer that could be coupled with a routine colonoscopy. This would be very useful as there are currently no screening tests for pancreatic cancer. The age of onset plus the long asymptomatic development time of pancreatic cancer mirrors that of colon cancer for which screening colonoscopy is suggested beginning at age 50 years. The simple preparation and collection methodology permits home sampling and safe, repeat sampling. These two features may permit longitudinal, time-course studies of pancreatic exocrine function and dysfunction (including screening of high-risk individuals for emergence of pancreatic neoplasia). Further studies to identify biomarkers of pancreatic disease (both neoplastic and inflammatory) from WGLF are ongoing.

It is very possible that the fact that the pancreatic juice samples were obtained from mostly pancreatic adenocarcinoma patients and the WGLF was obtained from patients with no obvious pancreatic pathology could have a role in the differences seen between the pancreatic juice and WGLF. As the focus of this study was to obtain an initial overall comparison of WGLF with pancreatic juice and no WGLF was collected from patients with pancreatic cancer, a direct statistical comparison of the WGLF and pancreatic juice samples for disease state was not performed. Although the two profiles are similar, the fact that they are two different bio fluids introduces enough variability that a direct comparison between the two would not be accurate.

Because we have seen pancreatic proteins consistently across WGLF samples, we believe the examination of WGLF in pancreatic cancer patients is justified. Because bowel preparations are often administered to patients prior to surgery, it is feasible to collect specimens at this time and compare them to samples obtained from other patients without pancreatic pathology. This has the potential to increase the number of overall patient samples (especially controls) and the subsequent power of future proteomic studies of pancreatic secretions.

Although there were no statistically significant differences between the patients with polyps and those without, care would need to be taken that any potential biomarkers be examined in patients with known colonic adenomas as well as other colonic pathologies as part of any validation studies of pancreatic disease using WGLF to rule out their effects on those proteins.

Our rate of successfully matching other studies was similar for both our pancreatic juice and WGLF samples and about the same as reported by other authors who compared their results with previous studies. Chen *et al.*⁷ identified 105 total proteins but only 45 of them were also found by Gronborg *et al.*²⁴ (43%). Doyle and coworkers' normal pancreatic juice proteome⁸ of 172 proteins had 42 proteins (25%) in common with Gronborg's study. The different protein isolation and analysis methods as well as differences in databases used for searching and stringency requirements for the successful identification of a protein make exact cross comparisons challenging.

Potential pitfalls. We do recognize that WGLF presents some potential pitfalls for analysis. However, many of them are the same issues present with the study of pancreatic juice directly, which include high human variability and potential degradation of protein. Studies on the amount of degradation in WGLF compared with pancreatic juice are underway.

CONCLUSIONS

WGLF has a similar proteomic profile to pancreatic juice, and can be collected non-invasively. Therefore, WGLF appears to be an excellent surrogate for the study of diseases of the pancreatic ductal system, especially pancreatic ductal adenocarcinoma. Further studies using this novel fluid are ongoing.

CONFLICT OF INTEREST

Guarantor of the article: Jana M. Rocker, PhD.

Specific author contributions: Study concept and design,

drafting of manuscript, acquisition, analysis, interpretation of data, and approved final draft submission: Jana M. Rocker; study concept and design, acquisition of data, critical revision of manuscript, and approved final draft submission: Marcus C. Tan, Lee W. Thompson, Carlo M. Contreras and Jack A. DiPalma; study concept and design, acquisition of data, critical revision of manuscript, study supervision, obtained funding, and approved final draft submission: Lewis K. Pannell.

Financial support: Creatics, sponsored this research.

Creatics, had no role in the design, collection, analysis and interpretation of the data, or the writing of the report.

Potential competing interests: Dr DiPalma is a consultant medical director for Braintree Laboratories, a maker of commercial bowel preparation solutions and has a financial interest in Creatics. Drs Rocker and Pannell also have a financial interest in Creatics, LLC.

Acknowledgments. We thank Carla Huber and Lynne Godwin and all of the physicians and nursing staff at the Gastroenterology clinic at the University of South Alabama and the staff at Cancer Surgery of Mobile and Mitchell Cancer Institute for their assistance with sample collection. We thank Joe J. Otto, Crystal L. Daniels, and Lindsay N. Schambeau in the Pannell laboratory at Mitchell Cancer Institute for their technical assistance and for critique of the manuscript. This work was funded by Creatics.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ There are no screening tests for pancreatic cancer.
- ✓ Pancreatic cancer has a long, asymptomatic development time and age of onset that mirrors colon cancer.
- ✓ A pancreatic ductal adenocarcinoma screening test which was performed at the time of screening colonoscopy would be of great benefit.
- ✓ Mass spectrometry is commonly used in biomarker studies.
- ✓ Pancreatic juice is difficult to collect.

WHAT IS NEW HERE

- ✓ The proteome of colonoscopy effluent contains a large amount of proteins produced in the pancreas and is similar to pancreatic juice in pancreatic protein content.
- ✓ Whole gut lavage fluid can be collected both during routine colonoscopy and in a toilet receptacle non-invasively.
- ✓ Study of whole-gut lavage fluid may enable development of a test for pancreatic cancer which can be coupled with routine colonoscopy.

6. Chen R, Pan S, Cooke K et al. Comparison of pancreas juice proteins from cancer versus pancreatitis using quantitative proteomic analysis. *Pancreas* 2007; **34**: 70–79.
7. Chen R, Pan S, Yi EC et al. Quantitative proteomic profiling of pancreatic cancer juice. *Proteomics* 2006; **6**: 3871–3879.
8. Doyle CJ, Yancey K, Pitt HA et al. The proteome of normal pancreatic juice. *Pancreas* 2012; **41**: 186–194.
9. Gao J, Zhu F, Lv S et al. Identification of pancreatic juice proteins as biomarkers of pancreatic cancer. *Oncol Rep* 2010; **23**: 1683–1692.
10. Park JY, Kim SA, Chung JW et al. Proteomic analysis of pancreatic juice for the identification of biomarkers of pancreatic cancer. *J Cancer Res Clin Oncol* 2011; **137**: 1229–1238.
11. Zhou L, Lu Z, Yang A et al. Comparative proteomic analysis of human pancreatic juice: methodological study. *Proteomics* 2007; **7**: 1345–1355.
12. Anderson MA, Fisher L, Jain R et al. Complications of ERCP. *Gastrointest Endosc* 2012; **75**: 467–473.
13. Makawita S, Smith C, Batruch I et al. Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. *Mol Cell Proteomics* 2011; **10**: M111 008599.
14. Yamada T, Alpers DH. *Textbook of Gastroenterology*, 4th edn. Vol 2. Philadelphia, PA, USA: Lippincott Williams & Wilkins. 2003.
15. Rocker JM, DiPalma JA, Pannell LK. Rectal effluent as a research tool. *Dig Dis Sci* 2015; **60**: 24–31.
16. Haug U, Wente MN, Seiler CM et al. Stool testing for the early detection of pancreatic cancer: rationale and current evidence. *Expert Rev Mol Diagn* 2008; **8**: 753–759.
17. Tooth D, Garsed K, Singh G et al. Characterisation of faecal protease activity in irritable bowel syndrome with diarrhoea: origin and effect of gut transit. *Gut* 2014; **63**: 753–760.
18. Harrell L, Wang Y, Antonopoulos D et al. Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. *PLoS One* 2012; **7**: e32545.
19. O'Mahony S, Barton JR, Crichton S et al. Appraisal of gut lavage in the study of intestinal humoral immunity. *Gut* 1990; **31**: 1341–1344.
20. Elson CO, Ealiding W, Lefkowitz J. A lavage technique allowing repeated measurement of IgA antibody in mouse intestinal secretions. *J Immunol Methods* 1984; **67**: 101–108.
21. Ferguson A, Humphreys KA, Croft NM. Technical report: results of immunological tests on faecal extracts are likely to be extremely misleading. *Clin Expe Immunol* 1995; **99**: 70–75.
22. Dai N, Gubler C, Hengstler P et al. Improved capsule endoscopy after bowel preparation. *Gastrointest Endosc* 2005; **61**: 28–31.
23. Singhal S, Nigar S, Paleti V et al. Bowel preparation regimens for colon capsule endoscopy: a review. *Therap Adv Gastroenterol* 2014; **7**: 115–122.
24. Gronborg M, Bunkenborg J, Kristiansen TZ et al. Comprehensive proteomic analysis of human pancreatic juice. *J Proteome Res* 2004; **3**: 1042–1055.
25. Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 1984; **138**: 141–143.
26. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008; **26**: 1367–1372.
27. Huang, da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1–13.
28. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
29. Uhlen M, Fagerberg L, Hallstrom BM et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015; **347**: 1260419.
30. Paulo JA, Kadiyala V, Banks PA et al. Mass Spectrometry-Based (GeLC-MS/MS) Comparative Proteomic Analysis of Endoscopically (ePFT) Collected Pancreatic and Gastrointestinal Fluids. *Clin Transl Gastroenterol* 2012; **3**: e14.
31. Hezel AF, Kimmelman AC, Stanger BZ et al. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev* 2006; **20**: 1218–1249.
32. Pezzilli R, Andriulli A, Bassi C et al. Exocrine pancreatic insufficiency in adults: a shared position statement of the Italian Association for the Study of the Pancreas. *World J Gastroenterol* 2013; **19**: 7930–7946.



Clinical and Translational Gastroenterology is an open-access journal published by Nature Publishing Group.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

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