Comprehensive Profiling of the Human Fecal Proteome from IBD Patients with DIA-MS Enables Evaluation of Disease-Relevant Proteins

Brandon J. Harder[#], Annemarie N. Lekkerkerker[#], Ellen P. Casavant, Jason A. Hackney, Allen Nguyen, Jacqueline M. McBride, W. Rodney Mathews, and Veronica G. Anania[*]

[#]These authors contributed equally to this work

*Corresponding author – Veronica G. Anania, Dept. of Translational Medicine, 1 DNA Way, South San Francisco, CA 94080. Email: anania.veronica@gene.com

Department of Translational Medicine, Genentech, South San Francisco, California, 94080, United States

Running Title: DIA-MS analysis of the IBD human fecal proteome

Abbreviations: IBD, Inflammatory bowel disease; HC, healthy control; UC, ulcerative colitis; CD, Crohn's disease; GI, gastrointestinal; LTF, lactoferrin; DDA, data-dependent acquisition; DIA, data-independent acquisition; fCAL: fecal calprotectin; FEB, fecal extraction buffer; MS, mass spectrometry; LC, liquid chromatography; ABS, absolute

ABSTRACT

Inflammatory bowel disease (IBD), which includes ulcerative colitis (UC) and Crohn's disease (CD), is characterized by chronic gastrointestinal inflammation. A high unmet need exists for non-invasive biomarkers in IBD to detect mucosal inflammation to monitor changes in disease severity and guide treatment decisions. Fecal proteomics has the potential to allow frequent, non-invasive monitoring of biomarkers in stool of IBD patients, however the fecal proteome remains under explored. Here a data-independent acquisition LC-MS/MS approach was used to profile the human fecal proteome in two independent cohorts of IBD patients and healthy controls (HC) to identify non-invasive biomarkers reflective of disease activity. 688 human proteins were quantified, with 523 measured in both cohorts. In UC stool 96 proteins were differentially abundant and in CD stool 126 proteins were differentially abundant compared to HC stool (absolute log2 fold change >1, p-value <0.05). Many of these fecal proteins are associated with infiltrating immune cells and ulceration/rectal bleeding, which are hallmarks of IBD pathobiology. Mapping of the identified fecal proteins to a whole blood single-cell RNA sequencing data set revealed the involvement of various immune cell subsets to the IBD fecal proteome. Findings from this study not only confirmed the presence of established fecal biomarkers for IBD, such as calprotectin and lactoferrin, but also revealed new fecal proteins from multiple pathways known to be dysregulated in IBD. These novel proteins could serve as potential non-invasive biomarkers to monitor specific aspects of IBD disease activity which could expedite clinical development of novel therapeutic targets.

Keywords: DIA-MS, fecal proteomics, biomarkers, IBD, Crohn's disease, ulcerative colitis, stool

INTRODUCTION

IBD is a chronic disorder of the gut characterized by an abnormal inflammatory response by the immune system [1]. Ulcerative colitis (UC) and Crohn's disease (CD) are the two most common forms of IBD, affecting over 3 million adult patients in the United States as of 2016 and over 6.8 million worldwide as of 2017 [2, 3]. Symptoms include abdominal pain, rectal bleeding, persistent diarrhea, and intestinal damage which often necessitates surgery leading to gut resection [4]. Diagnosis of disease, differentiation of disease subtype, and assessment of disease activity in IBD requires endoscopy, a procedure to examine the patient's digestive tract [5-9]. The invasive nature of this procedure is a burden to patients and therefore clinicians are unable to frequently monitor changes in ongoing disease activity or therapeutic response. Non-invasive biomarkers reflective of intestinal disease activity could allow more frequent, longitudinal, monitoring of mucosal disease and could support early futility analyses in clinical trials and earlier discontinuation of ineffective therapies. Furthermore, non-invasive biomarkers can increase our understanding of underlying disease pathways in patients with an inadequate or loss of response to an interventional treatment. These types of biomarkers are desirable in the treatment paradigm in IBD that aims to achieve disease remission based on normalization of intestinal inflammation which is currently evaluated by endoscopic and histological healing [10, 11].

Stool is a disease-proximal matrix that can be collected non-invasively and could enable more frequent, longitudinal assessment of mucosal disease activity in IBD. In addition to food waste and bacterial products, stool is known to contain secreted human proteins and cellular debris, shed into the lumen as a result of immune cell activity and a damaged mucosal epithelial barrier of the gastrointestinal (GI) tract. Therefore, stool is a rich source of proteins that reflect overall gut health status and biologically relevant pathway activation information [12-14].

Fecal calprotectin (fCAL), a heterodimer of S100A8 and S100A9, is a protein released by activated neutrophils recruited to the gut during inflammation and found in the stool of IBD patients [15]. Currently, fCAL is used as a proxy for intestinal inflammation in clinical practice and in clinical studies to monitor both IBD disease activity and treatment response [15-18]. Recent studies indicate that fCAL measured at baseline and up to 14 weeks after treatment with biologics had long-term prognostic value; however, fCAL levels were affected by duration of disease, making it less reliable as an inflammatory signal indicative of flare or disease activity [19, 20]. Furthermore, fCAL is also elevated in patients with NAFLD and obesity, and therefore, fCAL is not necessarily reflective of gut specific inflammation or IBD flare [21-24].

IBD is a heterogeneous disease and different biological mechanisms are likely responsible for residual disease burden across patient subgroups and fCAL does not provide information beyond neutrophil-associated inflammation. Other human fecal proteins have been reported as potential biomarkers in IBD, however, these have not been validated in multiple cohorts and none have been employed in the context of a

clinical trial (reviewed in [25, 26],[27, 28]). A more comprehensive view of the fecal proteome could enable discovery of additional disease mechanisms that may lead to development of novel therapeutic strategies. Novel, non-invasive biomarkers that inform on specific IBD disease pathways are needed to improve monitoring of patients and inform drug development.

Recent advances in mass spectrometry instrumentation and analytical tools have increased sensitivity in complex clinical matrices and enabled more robust label-free quantitation [29-34]. To deeply profile the human fecal proteome, proteins were extracted from stool, proteolytically digested, and resulting peptides were quantified using a data-independent acquisition LC-MS/MS (DIA-MS) approach. These techniques were employed in a proteomic approach to characterize differences in the human fecal proteome from two independent cohorts both with stool from healthy controls and patients with UC or CD. The findings presented herein provide peptide level results on 688 proteins in the human IBD fecal proteome, many of which are differentially abundant in IBD stool across two independent cohorts including valuable information spanning multiple pathways in IBD. These results support further development of fecal proteomics as a tool for non-invasive detection of gut mucosal inflammation in IBD patients.

EXPERIMENTAL PROCEDURES

Patients and Sample Collection

Two independent sample cohorts were compiled: Cohort 1, n=15 stool samples and Cohort 2, n= 40 stool samples. Informed consent for use of stool in this study was covered by informed consent forms (ICFs) under different IRBs. Samples sourced from Genentech were collected for this use and covered under study number 1096262, wcg IRB tracking number 20080040. Samples sourced from BioIVT were prospectively collected for this use and covered under SeraTrials sponsor protocol number 05035, Western IRB tracking number 20170439.

Procured samples from BioIVT were collected from patients within 24 hours of being produced then frozen at -70 °C before being shipped on dry ice. Upon receipt, samples were thawed and homogenized by stirring. Genentech volunteer samples were collected the same morning they were produced and then homogenized by stirring. These samples were then processed into single-use aliquots and stored at -70 °C. Prior studies have demonstrated that proteins in stool remain stable when frozen [35, 36]. Cohort 1 consisted of n=5 UC and n=5 CD samples purchased from BioIVT and n=5 HC samples from Genentech volunteers. Cohort 2 consisted of a total of 40 samples: 20 HC samples (n=15 from Genentech volunteers and n= 5 from BioIVT), 10 UC samples and 10 CD samples purchased from BioIVT. Inclusion criteria for all prospectively collected IBD patient samples required a clinical diagnosis of UC or CD as determined by endoscopic and histologic examination as well as a minimum Mayo Clinical Score (MCS) ≥6 as reported on day of collection for UC donors and a minimum Simple Endoscopic Score for Crohn's disease (SES-CD) score of ≥4 for CD donors. A

summary of available patient characteristics can be viewed in Table 1 and Supplemental Table 1.

Fecal Sample Preparation

10 different solutions were assessed for their ability to extract fecal calprotectin as measured by immunoassay as a surrogate for total protein extraction. Detergent free buffers tested included water, phosphate buffered saline (PBS), fecal extraction buffer ((FEB), 0.1 M Tris, 0.15 M NaCl, 1.0 M urea, 10 mM $CaCl_2$, 0.1 M citric acid monohydrate, pH 8, plus protease inhibitors (Roche, cat# 5892970001)), 8M urea, and normal saline (0.9% NaCl in water). Buffers containing detergent or proprietary formulations included: PBST (PBS + 0.1% Tween-20), B-CAL-EX (Buhlmann, proprietary), RIPA (50mM Tris-HCl pH 7.4, 1mM EDTA, 1% Triton X-100, 0.25% sodium deoxycholate, 150mM NaCl, in water), CTAB (1% cetyltrimethyl ammonium bromide (CTAB), 1.4M NaCl, 100mM Tris pH 8.0, 20mM EDTA, 2% polyvinyl pyrrolidone MW 40kDA, 0.4% beta-mercatpoethanol, in water) and ScheBo (ScheBo Biotech AG, proprietary). FEB performed well for extracting fCAL and this buffer was used throughout the remainder of the study (Supplemental Figure 1). All study samples were randomized prior to sample extraction. Raw stool samples from Cohort 1 were thawed, weighed, placed in a SmartPrep extraction device (Buhlmann Diagnostics, cat# B-CAL-RAD), and a 1:50 sample weight to volume of FEB was added. During protein extraction, samples were normalized by wet weight. The device was vortexed until sample and buffer homogeneity was visually confirmed. This homogenate was centrifuged at 15K g for 10 minutes and the resulting supernatant (extract) was

analyzed for total protein content by BCA assay (Thermo Fisher Scientific, cat# 23227). Extracts were aliquoted and frozen at -70 °C for further analyses. Cohort 2 samples were prepared similarly to Cohort 1, except with a 1:20 weight to buffer volume. The sample to buffer ratio for Cohort 2 was changed due to the detection of polymer in the injected samples which caused interference. No determination could be made as to the polymer source. The change in sample to buffer ratio reduced this polymer signal, allowing analysis to proceed.

Sample extracts (60 μg total protein per sample) were denatured using Biognosys' Denature Buffer and reduced/alkylated using Biognosys' Reduction/Alkylation solution for 1 h at 37 °C. Digestion was carried out using trypsin (Promega) at a 1:50 ratio (trypsin:protein) overnight at 37 °C. Peptides were desalted using a C18 microspin plate (Nest Group) according to the manufacturer's instructions and dried using a SpeedVac system. Peptides were resuspended in 30 μL solvent A (1% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA), in water) containing iRT kit calibration peptides (Biognosys). Peptide concentrations were determined using a Micro BCA assay and Nanodrop spectrophotometer (Thermo Fisher Scientific, cat# 23235).

Spectral Library Sample Fractionation

Equal amounts of peptides from each Cohort 1 sample were pooled by group (pool 1: HC samples, pool 2: UC samples, pool 3: CD samples) then fractionated by high pH reversed phase chromatography. Ammonium hydroxide was added to achieve a pH >10. For each set of pooled peptides (UC, CD, HC), fractionation was performed using a Dionex UltiMate 3000 RS pump (Thermo Fisher Scientific) on a NanoAcuity

UPLC CSH C18 1.7μm, 2.1 x 150mm column (Waters). The gradient was 1% to 40% solvent B in 20 min, solvents were A: 20 mM ammonium formate in water, B: ACN. Fractions for each set of pooled peptides (UC, CD, HC) were taken every 40 sec and sequentially combined to 6 fraction pools for a total of 18 fractions. These were dried by SpeedVac and resolved in solvent A containing iRT kit calibration peptides.

For shotgun DDA LC-MS/MS measurements, 2 μg of peptides per fraction was injected to an in-house packed C18 column (Dr. Maisch ReproSil Pur, 1.9 μm particle size, 120 Å pore size; 75 μm inner diameter, 50 cm length, New Objective) on a Thermo Fisher Scientific Easy nLC 1200 nano-liquid chromatography system connected to a Thermo Fisher Scientific Q Exactive HF mass spectrometer equipped with a standard nano-electrospray source. LC solvents were A: 1% ACN in water with 0.1% formic acid (FA); B: 15% water in ACN with 0.1% FA. The nonlinear LC gradient was 1-52% solvent B in 60 min followed by 52–90% B in 10 sec, 90% B for 10 min, 90% -1% B in 10 sec and 1% B for 5 min. A modified TOP 15 method was used; full MS1 covered the m/z range of 350-1650 with a resolution of 60,000 (AGC target value was 3e6) and 15 data-dependent MS2 scans with a resolution of 15,000 (AGC target value was 2e5). MS2 acquisition precursor isolation width was 1.6 m/z while normalized collision energy was centered at 27 (10% stepped collision energy) and the default charge state was 2+.

Raw data were analyzed using the SpectroMine 1.0 search engine (Biognosys), and the false discovery rate on peptide and protein level was set to 1%. A human UniProt FASTA database (downloaded 2018-07-01) was used to search against, and the search parameters allowed for 2 missed cleavages and variable modifications N-

terminal acetylation and methionine oxidation. The resultant identified peptides from the 18 separate fractions (6 fractions from each set of pooled peptides from UC, CD, and HC samples) were compiled into a spectral library using Spectronaut 14.10 (Biognosys) which contained 6914 peptides and 1102 human proteins.

DIA-MS Analysis of Stool Samples

Individual samples were prepared as indicated above. For Cohort 1 DIA-MS measurements, 2 μg of peptide per sample were injected in a randomized order on the same column and MS system described above. LC solvents were A: 1% ACN in water with 0.1% FA; B: 15% water in ACN with 0.1% FA. The nonlinear LC gradient was 1-55% solvent B in 120 minutes followed by 55–90% B in 10 seconds, 90% B for 10 minutes, 90% -1% B in 10 seconds and 1% B for 5 minutes. A DIA method with one full range MS1 survey scan and 22 DIA variable width windows was used as previously described (37).

For Cohort 2 DIA-MS measurements, 1 μg of peptides per sample were injected in a randomized order to an in-house packed reverse phase column (PicoFrit emitter with 75μm inner diameter, 60 cm length and 10 μm tip from New Objective packed with 1.7 μm Charged Surface Hybrid C18 particles from Waters) on a Thermo Fisher Scientific Easy nLC 1200 nano-liquid chromatography system connected to a Thermo Fisher Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with a Nanospray Flex Ion Source. LC solvents were A: 1% ACN in water with 0.1% FA; B: 20% water in ACN with 0.1% FA. The nonlinear LC gradient was 1-59% solvent B in 95 minutes followed by 59–90% B in 10 seconds, 90% B for 8 minutes, 90% -1% B in 10

seconds and 1% B for 5 minutes at 60 °C and a flow rate of 250 μL/min. A DIA-MS method with one full range MS1 survey scan and 29 DIA windows was used as previously described [37].

Mass Spectrometry Data Analysis

DIA-MS data were analyzed using Spectronaut 14.10 software (Biognosys) for peptide identification and MSstats "highQuality" feature subset for peptide to protein roll up [38, 39]. The false discovery rate on peptide and protein level roll-up was set to 1%. The assay library (protein inventory) generated in this project was used for the analysis. DIA-MS measurements analyzed with Spectronaut were not normalized in order to capture sample and disease state variation but were normalized at extraction level by wet weight and analysis level by equivalent peptide injection mass. Q-value filtered data (identifications included if q-value < 0.01) and sparse filtered data (no missing values, identifications included if target was identified in a minimum of one sample with q-value < 0.01) were reported. Distance in heatmaps was calculated using the “manhattan” method, and the clustering was performed using “ward.D” for both axes. General plotting was done in R using the ggplot2 package. Fold change was computed using the mean log2 transformed abundance values by group. P-values were calculated using the Mann-Whitney U test. Values for undetected proteins (no abundance data) were imputed by identifying the lowest detected log2 abundance value within the data set and dividing by the total sample number. For the analysis of the proteomic abundance data set p-values were calculated using the Mann-Whitney U test in Excel after testing for significance with the Shapiro-Wilk test as outlined by Aguilan et al. [40]. For box and

violin plots, GraphPad Prism was used and reported p-values were computed using two-way ANOVA analysis with Tukey's or Šídák's multiple comparison correction, as appropriate. Principal component analysis (PCA) was performed using the FactoMineR package in R.

Fecal Calprotectin Immunoassay

Samples were analyzed for fecal calprotectin using the EK-CAL IA kit (Buhlmann Diagnostics) per manufacturer's instructions using the recommended final dilution of 1:7500 except for HC samples, which were diluted to a final ratio of 1:2500 for analysis due to low fecal calprotectin levels. An aliquot of the same sample extract, prepared and described above, was used to determine fCAL levels as was used for tryptic digestion.

Preparation of Whole Blood for Single-Cell RNA Sequencing

Whole blood was collected from 2 healthy volunteers in K2EDTA tubes. Red blood cell (RBC) lysis buffer (BioLegend, cat# 420301) was added to 200 μL of the whole blood and rotated end to end at room temperature for 15 minutes. At the end of incubation, tubes were then centrifuged at 1250 RPM for 5 minutes. The supernatant was decanted and discarded, and the resulting cell pellet was resuspended in 1% bovine serum albumin (BSA) (Gibco, cat# 15260-037), then repeated the spin and decant steps. The pellet was then resuspended in 1% BSA with an additional RNase inhibitor (Roche, cat# 3335399001) and proceeded to library preparation.

Single-cell RNA sequencing

Sample processing for single-cell RNA sequencing (scRNAseq) was completed using Chromium Single Cell 3' Library and Gel bead kit v3 following the manufacturer's

protocol (10x Genomics, cat# PN-1000121). The cell density and viability of single-cell suspension were determined by Vi-CELL XR cell counter (Beckman Coulter). The cell density was used to infer the volume of single cell suspension needed in the reverse transcription (RT) master mix, aiming to achieve ~6,000 cells per sample. To bring the neutrophil-specific transcripts above background, PCR cycles were increased by 2 extra cycles during cDNA amplification in a non-enriched sample. cDNAs and libraries were prepared following the manufacturer's user guide (10x Genomics). Libraries were profiled by Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, cat# 5067-4626) and quantified using Kapa Library Quantification Kit (Kapa Biosystems, cat# KK4824). Each library was sequenced in one lane of HiSeq4000 (Illumina) following the manufacturer's sequencing specification (10x Genomics).

Single-cell data analysis and clustering

This data will be made publicly available in the European Genome-Phenome Archive (EGA, ID pending). Reads from scRNAseq libraries were mapped to the reference human genome (GRCh38), with gene models based on GENCODE (v27), using cellranger and default parameters. The emptyDrops function from the DropletUtils package was used to identify cells with > 200 UMIs at an FDR of 1%. Droplets with fewer than 100 detected genes were removed from the analysis, as were droplets with >35% of UMIs mapping to mitochondrial genes. Libraries were normalized using the logNormCounts function from the scater package. The top 5000 most variable genes were identified and principal components analysis was performed, computing the top 10 principal components. Mutual nearest neighbors (MNN) was used to merge the two

libraries and cells were clustered by calculating a shared nearest neighbor graph using the scran package, and identified clusters using the Louvain algorithm, as implemented by the igraph R package, setting k=50. Marker genes for each cluster were identified using the findMarkers function from the scater package, and manually annotated clusters based on prior knowledge and literature searches to identify the major cell types present in the dataset.

Proteins that were differentially regulated in stool from UC or CD patients versus healthy controls were chosen and mapped to gene names. The expression of these selected genes in each cell type were obtained from the annotated single-cell dataset and subsequently aggregated and summed by cell type. Gene expression for each cell type was plotted using pheatmap in R. Data was scaled by gene, with hierarchical clustering applied by cell type and gene.

RESULTS

An overview of the mass spectrometry workflow used in this study is outlined in Figure 1A. Briefly, stool samples were collected from HC subjects or IBD patients, homogenized, and then aliquoted for LC-MS/MS and fCAL analysis. For LC-MS/MS, stool was extracted, digested, and analyzed by DIA-MS on orbitrap mass spectrometer systems with Spectronaut software used to analyze the resulting data for peptide identification and MSstats used for peptide-to-protein roll up. To identify and confirm differentially abundant human proteins in stool, the fecal proteomes of patients with IBD

from two independent cohorts were investigated (Figure 1B). A summary of patient characteristics and sample numbers are listed in Table 1.

Assessment of the IBD Fecal Proteome

Proteins from stool samples from two independent cohorts that contained both HC subjects and patients with UC or CD were extracted, proteolytically digested, and analyzed by DIA-MS. Resultant spectra were analyzed and interrogated with the spectral library built from pooled Cohort 1 samples used in this study. In total, 6076 peptides from 672 proteins were observed in Cohort 1 and 5,127 peptides from 539 proteins were observed in Cohort 2 (Supplemental Table 2). In Cohort 1, 149 proteins were identified that were not detected in Cohort 2 while 16 proteins were identified in Cohort 2 that were not detected in Cohort 1. In total, 688 human proteins were quantified in at least one cohort, with 523 proteins quantified in both cohorts.

Identification of IBD Linked Fecal Proteins – Cohort 1 Analysis

Cohort 1 consisted of samples from n=5 HC donors, n=5 patients with UC, and n=5 patients with CD. In total, 495 proteins were detected in HC samples, 670 proteins detected in UC, and 581 proteins detected in CD (Supplemental Table 2). Assessment of protein abundance using unsupervised hierarchical clustering demonstrated that HC samples have fewer human proteins and they are present at lower levels compared to IBD patient samples (Figure 2A). The same amount of total protein was analyzed for every sample; however, HC samples have higher levels of microbial proteins and proteins from food, and therefore, it is expected that human proteins would be lower abundance in stool in this group compared to IBD patients. Principal component

analysis indicates that differences in protein abundance drives separation of HC samples from samples from IBD patients by PC1 (Figure 2B). In addition, CD samples are more similar to each other, clustering closely together by PC2, while in this small cohort, UC samples tended to be less similar to one another, demonstrating a large distribution on both PC1 and PC2.

Using a cutoff value of p-value < 0.05 and log2 fold change (FC) > 1 to filter results, 288 proteins showed significant differential abundance in the UC cohort compared to HC (Figure 2C). The top 10 most changed proteins in UC stool, all of which showed an increased abundance, included numerous neutrophil derived proteins such as fecal calprotectin (S100A8 and S100A9), lactotransferrin (LTF), myeloperoxidase (MPO), proteinase 3 (PRTN3), neutrophil elastase (ELANE), and lipocalin 2 (LCN2). In addition, plasma proteins hemoglobin subunit B (HBB) and granulin precursor (GRN) were detected. Interestingly, olfactomedin 4 (OLFM4), a glycoprotein produced by specialized intestinal cells called Paneth cells and secreted into the mucus of the gut [41], was also significantly elevated in the UC fecal samples (Figure 2C).

For the CD cohort, 276 proteins showed significant differential abundance in the CD cohort compared to HC, and 113 proteins showed significant differential abundance in the UC cohort compared to the CD cohort (Figure 2D, Supplemental Table 2). Similar to UC stool, the proteins most elevated in CD stool included neutrophil derived proteins: S100A8, S100A9, LTF, MPO, and PRTN3 as well as plasma proteins: HBB, complement factor C3 (C3), immunoglobulin heavy constant gamma 1 (IGHG1), and

alanyl aminopeptidase (ANPEP). Notably, enterocyte derived tropomyosin 3 (TPM3) showed a reduction in abundance in CD stool compared to HC (Figure 2D). The distribution of shared proteins between HC, UC, and CD samples is summarized in Figure 2E. Overall, 177 proteins out of 672 proteins were only found in stool from IBD patients and 58 proteins were unique in stool from UC patients.

To determine how well fCAL data generated with an established immunoassay aligned with DIA-MS data, fCAL was measured by immunoassay and compared to the DIA-MS results for the two heterodimeric protein subunits of calprotectin (S100A8 and S100A9). This analysis demonstrated that DIA-MS measurements for both proteins were very strongly correlated with fCAL results (Pearson r =0.945 and r =0.962, for S100A8 and S100A9, respectively) and in both assays, fCAL was significantly lower in HC samples than in samples from patients with IBD (Figure 2F, 2G, p-value < 0.0001).

Identification of IBD Linked Fecal Proteins – Cohort 2 Analysis

A second, independent cohort of stool samples consisting of samples from HC (n=20), UC (n=10), and CD (n=10), was analyzed. In Cohort 2 a total of 5,127 peptides from 539 proteins were identified. Again, fewer overall proteins were identified in HC donor samples compared to IBD patients (HC: n=481, UC: n=535, CD: n=537) suggesting the composition of HC donor stool contained fewer human proteins (Supplemental Table 2, Supplemental Figure 2). 125 proteins were determined to have significant differential abundance in stool from UC patients compared to HC stool, and there were 186 proteins with significant differential abundance in stool from CD patients compared to HC stool (Supplemental Figure 2). In total, 96 proteins were found to have

significant differences in abundance in both Cohort 1 and Cohort 2 UC samples compared to HC (33% confirmation rate) and 126 proteins were found to be significantly different in both Cohort 1 and Cohort 2 CD samples compared to HC (68% confirmation rate) (Figure 3A, 3B, Supplemental Table 2 and 3). In Cohort 2, 28 proteins were found to have significant differential abundance between UC and CD samples, and only 3 out of 28, namely lactate dehydrogenase A (LDHA), vimentin (VIM), and alpha cardiac actin 1 (ACTC1), showed significant differential abundance in Cohort 1 as well (Figure 3C, Supplemental Table 2 and 3).

There was a robust correlation between the fold change in protein abundance between healthy and diseased samples for proteins confirmed in Cohort 1 and Cohort 2 ($r = 0.56$, $r = 0.5$ for fold change of UC versus HC and for fold change of CD versus HC, respectively) (Figure 3D, 3E). Additionally, similar to Cohort 1, levels of fecal calprotectin measured by immunoassay were very strongly correlated with DIA-MS measurements (Supplemental Figure 2). UC samples from both cohorts and CD samples from Cohort 2 showed an increased number of protein identification compared to healthy control samples (Figure 3F). IBD patient samples from Cohort 2 showed an overall lower level of fCAL compared to Cohort 1, suggestive of a lower level of neutrophil infiltration compared to Cohort 1.

Four-way plots were utilized to assess the directional changes in all proteins measured in both cohorts, including 200 proteins in UC stool and 205 proteins in CD stool that only demonstrated significant differences in disease compared to healthy control in one cohort and not the other. In UC samples there were 170 proteins that only

showed significant differences in Cohort 1 and 30 proteins that only showed significant differences in Cohort 2 while in CD samples there were 49 proteins that only showed significant differences in Cohort 1 and 156 proteins that only showed significant differences in Cohort 2. For these proteins that only reached significance in one cohort, the overwhelming majority (81% in UC and 80% in CD) demonstrate similar trends across both cohorts (Figure 3G, 3H).

Single Cell RNAseq of Whole Blood Mapped to Confirmed Proteins

In order to characterize the cellular origin of the fecal proteins enriched in IBD stool samples, a whole blood single cell RNA sequencing (scRNAseq) dataset was generated. Confirmed proteins from the proteomic data set (96 from UC stool or 126 from CD stool) were mapped to the single cell expression dataset and the resulting gene set was used to create a heatmap which was populated with expression data from the whole blood scRNAseq data (Figure 4A, 4B). This analysis showed that many of the proteins enriched in IBD stool samples are highly expressed in, and likely derived from, neutrophils and T cells. Other immune cell subsets such as B cells and monocytes potentially contribute to the IBD fecal proteome as well. This analysis revealed the diverse contributions of innate and adaptive immune cells to the human fecal proteome.

Utilizing the scRNAseq analysis, IBD stool proteins enriched in specific immune cell subsets were identified as biomarker candidates (Figure 5A, 5B). T cell produced proteins alpha-2-macroglobulin (A2M) and annexin A6 (ANXA6) and B cell produced proteins immunoglobulin heavy constant gamma 4 (IGHG4) and immunoglobulin heavy constant gamma 1 (IGHG1) were elevated in stool collected from UC and/or CD

patients compared to HC. Additionally, monocyte derived proteins (complement 3 (C3) and phospholipase B domain containing 1 (PLBD1)) and neutrophil derived proteins (S100 calcium binding protein P (S100P), leucine rich alpha-2-glycoprotein 1 (LRG1), and matrix metallopeptidase 9 (MMP9)) were also increased in IBD stool.

DISCUSSION

Here, a comprehensive, unbiased, analysis of the human fecal proteome from IBD patients, focusing on secreted human proteins was performed. In this study, a robust DIA-MS method was used to interrogate the human fecal proteome of IBD patients in two independent clinical cohorts of IBD patient stool where the abundance of 688 proteins was characterized. In total, 96 proteins were confirmed to be differentially abundant in in UC patient stool and 126 proteins were consistently differentially abundant in CD patient stool compared to HC. Interestingly, 40 proteins were differentially abundant in both UC and CD compared to HC suggesting overlapping disease mechanisms between IBD subtypes. Many of these fecal proteins were associated with disease relevant biology such as adaptive immune cell subsets and neutrophilic infiltration into the gut lumen. In addition, the study identified fecal proteins indicative of disruption of the epithelial barrier leading to cell damage and ulceration and/or rectal bleeding.

fCAL is the most widely used fecal biomarker of intestinal mucosal inflammation and used in clinical practice to monitor disease activity and guide treatment. Results from this study demonstrate very high correlations between fCAL measured using

ELISA and abundance of the two fCAL protein subunits S100A8 and S100A9 measured by DIA-MS, corroborating the fecal proteomic approach to interrogate gut mucosal inflammation (Figure 2F, Supplemental Figure 2).

Infiltration of immune cells into the gut is a hallmark of IBD pathology and understanding which populations of cells are present and potentially driving disease activity is imperative to develop better targeted therapies. Recruitment and activation of immune cells leads to release of proteases and other effectors that disrupt the epithelial barrier and cause mucosal injury [42]. Some of the neutrophil-derived proteins identified in this study have been reported previously as fecal biomarkers in stool (reviewed in [43, 44]). Cohort 1 data demonstrated higher levels of neutrophil derived proteins known to cause tissue injury and loss of epithelial barrier function of the gut such as lactotransferrin and myeloperoxidase which were detected in both UC and CD patient stool [45-50].

Cross referencing the fecal proteome results with scRNAseq data derived from whole blood suggested the involvement of neutrophilic and T cell derived proteins and indicated that other relevant immune cell subsets such as B cells and NK cells may also be represented in this proteomic dataset. The limitation to this analysis is that this dataset lacked gut-specific cell types such as epithelial cells, which may result in a more restricted view of the cellular origin of these fecal proteins.

Some of the top-ranked differentially abundant proteins detected in this study were linked to T and B cell biology. Examination of proteins mapped to T cells, revealed increased levels of the phospholipid membrane binding protein ANXA6, that has been

implicated in T cell activation and signaling [51]. Interestingly, A2M, a broad protease inhibitor, was also found to be elevated in IBD patients and has been suggested to protect against structural damage by inhibition of proteases released by activated neutrophils. A2M was previously found in stool to be a potentially novel marker of IBD disease activity [52-54].

Historical studies have identified an increase in mucosal IgG responses in IBD patients [55, 56] directed to intestinal bacteria and this has been linked to mucosal inflammation [57, 58]. Many immunoglobulins detected in this study were found to be more abundant in IBD stool compared to healthy control stool including the IgG isotypes IGHG1 and IGHG4. The role of B cells in IBD biology has gained attention and characterization of IgG subclasses in stool could further our understanding of UC vs CD disease biology [59, 60].

Innate immune cell related proteins were also found to be increased in IBD stool compared to healthy control. C3, an important player in the classical and non-classical complement pathways, was found to be upregulated in IBD stool. Similarly, PLBD1, a putative phospholipase produced by monocytes as well as other cell types, was also found to be elevated in IBD stool. PLBD1 is thought to be important for antimicrobial responses and has been shown to contribute to both intestinal epithelial injury and repair [61-63].

This study also identified proteins related to gut ulceration and/or rectal bleeding, and disruption of the epithelial barrier leading to cell damage. Ulceration and rectal bleeding are signs of UC and CD, though less frequently observed in CD, and

contribute to both the partial and complete Mayo score (pMCS/MCS) for UC and Crohn's disease activity index (CDAI), each representing clinical indices of disease, respectively [64, 65]. Here, increased levels of plasma proteins were detected in IBD stool, including albumin (ALB), hemoglobin (HBA1, HBD) and haptoglobin (HP). These plasma proteins were absent or only present at low abundance levels in HCs. Blood in the stool is generally thought to be the result of hemorrhaging from the intestinal mucosa as a result of ulcers formed during inflammation of the large intestine and rectum and from constipation and straining during bowel movements [66]. Non-invasive monitoring of blood proteins in stool by fecal proteomics could be useful as an early indicator of ulceration and/or rectal bleeding, while the reduction of these proteins in stool could indicate mucosal healing.

Intracellular and membrane bound proteins such as mitochondrial glutathione-disulfide reductase (GSR), and major histocompatibility complex, class I, A (HLA-A) were also present in IBD stool and absent or present at low abundance levels in healthy stool. Injury to the gut mucosa due to inflammation, causing cellular damage, ulceration, and release of membrane bound proteins and cytoplasmic contents into the gut lumen, is a likely source of these proteins in IBD stool. These proteins, which may be a proxy for damage to the gut epithelial barrier, could be potential biomarkers of mucosal inflammation and injury.

Whereas the search for fecal biomarkers to monitor disease activity and response to therapy has focused on known fecal proteins [43], the search for novel non-invasive fecal biomarkers in stool through proteomic approaches has been sparse.

Casavant et al. used a DDA approach to profile 10 pediatric IBD stool samples resulting in the detection of over 500 human proteins, 30 of which were found to have an increased differential abundance in flare vs. remission patients [11]. Interestingly, this study identified 28 out of these 30 significantly changed proteins in the high inflammatory IBD patient stool. As was found in this study, Casavant et al. identified proteins involved in neutrophil degranulation (S100A8, S100A9, MPO, ELANE) and plasma proteins (HBA1, HBB, and HP). Similar neutrophilic and plasma proteins were previously reported by others as well [12, 13, 67]. Interestingly, Basso et al. were able to identify 193 human proteins in their study, of which only 161 showed any differential abundance in IBD stool, whereas in our study we identified 688 proteins, 422 of which showed statistically significant differential abundance in IBD stool. The lower number of identified proteins could be due to their use of water as an extraction buffer, which performs poorly at extracting calprotectin (and presumably other soluble stool proteins) compared to other tested buffers (Supplemental Figure 1).

Several studies have utilized 2-D gels combined with MALDI-TOF to profile IBD stool, however, this approach is less sensitive in terms of proteome coverage and is often is only capable of identifying features rather than proteins making this approach less informative as a tool to assess IBD biological pathway activation [13, 68]. Other studies have investigated other sample types from IBD patients to discover novel disease activity biomarkers in IBD including intestinal mucosal luminal interface aspirates and mucosal biopsies [27, 28]. While this previous work provides important tissue level context for proteomic changes in the IBD gut, obtaining these sample types

requires invasive procedures which prevents frequent patient sampling. Therefore, proteins measured from these sample types are not suitable as non-invasive disease monitoring tools which was the focus of this study.

This work began as a feasibility study to determine if differences in the human fecal proteome could be detected in IBD stool compared to normal stool as a way to identify possible biomarkers of disease in IBD from non-invasively collected samples using DIA-MS. One differentiating factor of this study was the use of two independent cohorts, adding strength to the reproducibility of our findings. A possible limitation to the generalization of these results is that these cohorts are small and not demographically balanced; therefore, these findings should be further investigated in larger clinical cohorts. Additionally, this study was not designed to identify biomarkers related to prediction of flare, or specific disease location, and future larger cohorts will be required to specifically address those questions. This study focused on identifying human secreted proteins as well as proteins introduced to the gut lumen due to cell damage and the spectral libraries used were entirely constructed of human proteins in order to monitor immune pathway activation. The DIA-MS datasets presented herein can be further interrogated to identify secreted microbial proteins to provide additional insights into the biology of IBD.

In summary, DIA-MS is a valid approach to non-invasively identify novel human fecal proteins in IBD and further analysis of these proteomic data sets may reveal insights into IBD disease pathways. A qualified method was developed to perform proteomics on stool samples that enabled identification of known and novel proteins

relevant to mucosal UC and CD disease biology. Analysis of these proteomic datasets has identified fecal proteins that may reflect activation of specific immune subsets. Follow-up studies to confirm these findings will provide further insight into relevant IBD pathways. If confirmed, quantification of specific stool proteins could allow for frequent and non-invasive monitoring of mucosal disease activity in patients, and potentially guide therapeutic development of novel interventional therapies in IBD.

DATA AVAILABILITY

All mass spectrometry (MS) raw data sets and spectral libraries have been deposited to the MASSIVE database (http://massive.ucsd.edu/) and can be downloaded by the identifier MSV000089237.

SUPPLEMENTAL DATA

This article contains supplemental data.

ACKNOWLEDGEMENTS

We thank Claudia Escher and Magdalena Bober from Biognosys for help with collecting the DIA-MS data and Sumedh Sankhe for assistance with data analysis.

REFERENCES

[1] Xavier, R. J., Podolsky, D. K., Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 2007, *448*, 427-434.

[2] Dahlhamer, J. M., Zammitti, E. P., Ward, B. W., Wheaton, A. G., Croft, J. B., Prevalence of Inflammatory Bowel Disease Among Adults Aged >/=18 Years - United States, 2015. *MMWR Morb Mortal Wkly Rep* 2016, *65*, 1166-1169.
[3] Collaborators, G. B. D. I. B. D., The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* 2020, *5*, 17-30.
[4] Colombel, J. F., Shin, A., Gibson, P. R., AGA Clinical Practice Update on Functional Gastrointestinal Symptoms in Patients With Inflammatory Bowel Disease: Expert Review. *Clin Gastroenterol Hepatol* 2019, *17*, 380-390 e381.
[5] Waye, J. D., Endoscopy in inflammatory bowel disease: indications and differential diagnosis. *Med Clin North Am* 1990, *74*, 51-65.
[6] Pera, A., Bellando, P., Caldera, D., Ponti, V.*, et al.*, Colonoscopy in inflammatory bowel disease. Diagnostic accuracy and proposal of an endoscopic score. *Gastroenterology* 1987, *92*, 181-185.
[7] Abreu, M. T., Harpaz, N., Diagnosis of colitis: making the initial diagnosis. *Clin Gastroenterol Hepatol* 2007, *5*, 295-301.
[8] Spiceland, C. M., Lodhia, N., Endoscopy in inflammatory bowel disease: Role in diagnosis, management, and treatment. *World J Gastroenterol* 2018, *24*, 4014-4020.
[9] Narula, N., Wong, E. C. L., Colombel, J. F., Sandborn, W. J.*, et al.*, Predicting endoscopic remission in Crohn's disease by the modified multiplier SES-CD (MM-SES-CD). *Gut* 2021.
[10] Peyrin-Biroulet, L., Sandborn, W., Sands, B. E., Reinisch, W.*, et al.*, Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): Determining Therapeutic Goals for Treat-to-Target. *Am J Gastroenterol* 2015, *110*, 1324-1338.
[11] Ungaro, R., Colombel, J. F., Lissoos, T., Peyrin-Biroulet, L., A Treat-to-Target Update in Ulcerative Colitis: A Systematic Review. *Am J Gastroenterol* 2019, *114*, 874-883.
[12] Casavant, E., Park, K. T., Elias, J. E., Proteomic Discovery of Stool Protein Biomarkers for Distinguishing Pediatric Inflammatory Bowel Disease Flares. *Clin Gastroenterol Hepatol* 2020, *18*, 2618-2619 e2611.
[13] Basso, D., Padoan, A., D'Inca, R., Arrigoni, G.*, et al.*, Peptidomic and proteomic analysis of stool for diagnosing IBD and deciphering disease pathogenesis. *Clin Chem Lab Med* 2020, *58*, 968-979.
[14] Lehmann, T., Schallert, K., Vilchez-Vargas, R., Benndorf, D.*, et al.*, Metaproteomics of fecal samples of Crohn's disease and Ulcerative Colitis. *J Proteomics* 2019, *201*, 93-103.
[15] Heida, A., Park, K. T., van Rheenen, P. F., Clinical Utility of Fecal Calprotectin Monitoring in Asymptomatic Patients with Inflammatory Bowel Disease: A Systematic Review and Practical Guide. *Inflamm Bowel Dis* 2017, *23*, 894-902.
[16] Reinisch, W., Bressler, B., Curtis, R., Parikh, A.*, et al.*, Fecal Calprotectin Responses Following Induction Therapy With Vedolizumab in Moderate to Severe Ulcerative Colitis: A Post Hoc Analysis of GEMINI 1. *Inflamm Bowel Dis* 2019, *25*, 803-810.

[17] Mumolo, M. G., Bertani, L., Ceccarelli, L., Laino, G.*, et al.*, From bench to bedside: Fecal calprotectin in inflammatory bowel diseases clinical setting. *World J Gastroenterol* 2018, *24*, 3681-3694.
[18] Chen, F., Hu, Y., Fan, Y.-H., Lv, B., Clinical Value of Fecal Calprotectin in Predicting Mucosal Healing in Patients With Ulcerative Colitis. *Frontiers in Medicine* 2021, *8*.
[19] Facciorusso, A., Ramai, D., Ricciardelli, C., Paolillo, R.*, et al.*, Prognostic Role of Post-Induction Fecal Calprotectin Levels in Patients with Inflammatory Bowel Disease Treated with Biological Therapies. *Biomedicines* 2022, *10*, 2305.
[20] Ishida, N., Kaneko, M., Asai, Y., Miyazu, T.*, et al.*, Effect of disease duration on fecal biomarkers in ulcerative colitis: a prospective cohort study. *BMC Gastroenterology* 2022, *22*, 420.
[21] Demirbaş, F., Çaltepe, G., Comba, A., Abbasguliyev, H.*, et al.*, Association of obesity and non-alcoholic fatty liver disease with the fecal calprotectin level in children. *Arab J Gastroenterol* 2020, *21*, 211-215.
[22] Liu, X., Wang, Y., Ming, Y., Song, Y.*, et al.*, S100A9: A Potential Biomarker for the Progression of Non-Alcoholic Fatty Liver Disease and the Diagnosis of Non-Alcoholic Steatohepatitis. *PLoS One* 2015, *10*, e0127352.
[23] Biemann, R., Buß, E., Benndorf, D., Lehmann, T.*, et al.*, Fecal Metaproteomics Reveals Reduced Gut Inflammation and Changed Microbial Metabolism Following Lifestyle-Induced Weight Loss. *Biomolecules* 2021, *11*.
[24] Sydor, S., Dandyk, C., Schwerdt, J., Manka, P.*, et al.*, Discovering Biomarkers for Non-Alcoholic Steatohepatitis Patients with and without Hepatocellular Carcinoma Using Fecal Metaproteomics. *Int J Mol Sci* 2022, *23*.
[25] Sutherland, A. D., Gearry, R. B., Frizelle, F. A., Review of fecal biomarkers in inflammatory bowel disease. *Dis Colon Rectum* 2008, *51*, 1283-1291.
[26] Kopylov, U., Rosenfeld, G., Bressler, B., Seidman, E., Clinical utility of fecal biomarkers for the diagnosis and management of inflammatory bowel disease. *Inflamm Bowel Dis* 2014, *20*, 742-756.
[27] Deeke, S. A., Starr, A. E., Ning, Z., Ahmadi, S.*, et al.*, Mucosal-luminal interface proteomics reveals biomarkers of pediatric inflammatory bowel disease-associated colitis. *Am J Gastroenterol* 2018, *113*, 713-724.
[28] Starr, A. E., Deeke, S. A., Ning, Z., Chiang, C. K.*, et al.*, Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate Crohn's disease from UC. *Gut* 2017, *66*, 1573-1583.
[29] Coscia, F., Doll, S., Bech, J. M., Schweizer, L.*, et al.*, A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* 2020, *251*, 100-112.
[30] Kimura, Y., Nakai, Y., Shin, J., Hara, M.*, et al.*, Identification of serum prognostic biomarkers of severe COVID-19 using a quantitative proteomic approach. *Sci Rep* 2021, *11*, 20638.
[31] Messner, C. B., Demichev, V., Bloomfield, N., Yu, J. S. L.*, et al.*, Ultra-fast proteomics with Scanning SWATH. *Nat Biotechnol* 2021, *39*, 846-854.

[32] Shu, Y., Gao, M., Zhou, Y., Liu, H., Sun, X., DIA Comparative Proteomic Analysis of Retro-oil Fluid and Vitreous Fluid From Retinal Detachment Patients. *Front Mol Biosci* 2021, *8*, 763002.
[33] Cho, K. C., Oh, S., Wang, Y., Rosenthal, L. S.*, et al.*, Evaluation of the Sensitivity and Reproducibility of Targeted Proteomic Analysis Using Data-Independent Acquisition for Serum and Cerebrospinal Fluid Proteins. *J Proteome Res* 2021, *20*, 4284-4291.
[34] Nakajima, D., Ohara, O., Kawashima, Y., Data-Independent Acquisition Mass Spectrometry-Based Deep Proteome Analysis for Hydrophobic Proteins from Dried Blood Spots Enriched by Sodium Carbonate Precipitation. *Methods Mol Biol* 2022, *2420*, 39-52.
[35] Lasson, A., Stotzer, P. O., Öhman, L., Isaksson, S.*, et al.*, The intra-individual variability of faecal calprotectin: a prospective study in patients with active ulcerative colitis. *J Crohns Colitis* 2015, *9*, 26-32.
[36] Oyaert, M., Van den Bremt, S., Boel, A., Bossuyt, X., Van Hoovels, L., Do not forget about pre-analytics in faecal calprotectin measurement! *Clin Chim Acta* 2017, *473*, 124-126.
[37] Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y.*, et al.*, Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol Cell Proteomics* 2017, *16*, 2296-2309.
[38] Choi, M., Chang, C. Y., Clough, T., Broudy, D.*, et al.*, MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 2014, *30*, 2524-2526.
[39] Tsai, T. H., Choi, M., Banfai, B., Liu, Y.*, et al.*, Selection of Features with Consistent Profiles Improves Relative Protein Quantification in Mass Spectrometry Experiments. *Mol Cell Proteomics* 2020, *19*, 944-959.
[40] Aguilan, J. T., Kulej, K., Sidoli, S., Guide for protein fold change and p-value calculation for non-experts in proteomics. *Mol Omics* 2020, *16*, 573-582.
[41] Gersemann, M., Becker, S., Nuding, S., Antoni, L.*, et al.*, Olfactomedin-4 is a glycoprotein secreted into mucus in active IBD. *J Crohns Colitis* 2012, *6*, 425-434.
[42] Fournier, B. M., Parkos, C. A., The role of neutrophils during intestinal inflammation. *Mucosal Immunol* 2012, *5*, 354-366.
[43] Di Ruscio, M., Vernia, F., Ciccone, A., Frieri, G., Latella, G., Surrogate Fecal Biomarkers in Inflammatory Bowel Disease: Rivals or Complementary Tools of Fecal Calprotectin? *Inflamm Bowel Dis* 2017, *24*, 78-92.
[44] Lehmann, F. S., Burri, E., Beglinger, C., The role and utility of faecal markers in inflammatory bowel disease. *Therap Adv Gastroenterol* 2015, *8*, 23-36.
[45] Ginzberg, H. H., Cherapanov, V., Dong, Q., Cantin, A.*, et al.*, Neutrophil-mediated epithelial injury during transmigration: role of elastase. *Am J Physiol Gastrointest Liver Physiol* 2001, *281*, G705-717.
[46] Barry, R., Ruano-Gallego, D., Radhakrishnan, S. T., Lovell, S.*, et al.*, Faecal neutrophil elastase-antiprotease balance reflects colitis severity. *Mucosal Immunol* 2020, *13*, 322-333.

[47] Hansberry, D. R., Shah, K., Agarwal, P., Agarwal, N., Fecal Myeloperoxidase as a Biomarker for Inflammatory Bowel Disease. *Cureus* 2017, *9*, e1004.
[48] Peterson, C. G., Eklund, E., Taha, Y., Raab, Y., Carlson, M., A new method for the quantification of neutrophil and eosinophil cationic proteins in feces: establishment of normal levels and clinical application in patients with inflammatory bowel disease. *Am J Gastroenterol* 2002, *97*, 1755-1762.
[49] Buisson, A., Vazeille, E., Minet-Quinard, R., Goutte, M.*, et al.*, Fecal Matrix Metalloprotease-9 and Lipocalin-2 as Biomarkers in Detecting Endoscopic Activity in Patients With Inflammatory Bowel Diseases. *J Clin Gastroenterol* 2018, *52*, e53-e62.
[50] Kolho, K. L., Sipponen, T., Valtonen, E., Savilahti, E., Fecal calprotectin, MMP-9, and human beta-defensin-2 levels in pediatric inflammatory bowel disease. *Int J Colorectal Dis* 2014, *29*, 43-50.
[51] Cornely, R., Pollock, A. H., Rentero, C., Norris, S. E.*, et al.*, Annexin A6 regulates interleukin-2-mediated T-cell proliferation. *Immunol Cell Biol* 2016, *94*, 543-553.
[52] Becker, K., Niederau, C., Frieling, T., Fecal excretion of alpha 2-macroglobulin: a novel marker for disease activity in patients with inflammatory bowel disease. *Z Gastroenterol* 1999, *37*, 597-605.
[53] Wu, S. M., Patel, D. D., Pizzo, S. V., Oxidized alpha2-macroglobulin (alpha2M) differentially regulates receptor binding by cytokines/growth factors: implications for tissue injury and repair mechanisms in inflammation. *J Immunol* 1998, *161*, 4356-4365.
[54] Vandooren, J., Itoh, Y., Alpha-2-Macroglobulin in Inflammation, Immunity and Infections. *Front Immunol* 2021, *12*, 803244.
[55] Brandtzaeg, P., Baklien, K., Fausa, O., Hoel, P. S., Immunohistochemical characterization of local immunoglobulin formation in ulcerative colitis. *Gastroenterology* 1974, *66*, 1123-1136.
[56] Rüthlein, J., Ibe, M., Burghardt, W., Mössner, J., Auer, I. O., Immunoglobulin G (IgG), IgG1, and IgG2 determinations from endoscopic biopsy specimens in control, Crohn's disease, and ulcerative colitis subjects. *Gut* 1992, *33*, 507-512.
[57] Castro-Dopico, T., Dennison, T. W., Ferdinand, J. R., Mathews, R. J.*, et al.*, Anti-commensal IgG Drives Intestinal Inflammation and Type 17 Immunity in Ulcerative Colitis. *Immunity* 2019, *50*, 1099-1114.e1010.
[58] Masu, Y., Kanazawa, Y., Kakuta, Y., Shimoyama, Y.*, et al.*, Immunoglobulin subtype-coated bacteria are correlated with the disease activity of inflammatory bowel disease. *Sci Rep* 2021, *11*, 16672.
[59] Uzzan, M., Martin, J. C., Mesin, L., Livanos, A. E.*, et al.*, Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat Med* 2022, *28*, 766-779.
[60] Spencer, J., Bemark, M., Human intestinal B cells in inflammatory diseases. *Nat Rev Gastroenterol Hepatol* 2023, *20*, 254-265.
[61] Okada, K., Itoh, H., Ikemoto, M., Serum complement C3 and α(2)-macroglobulin are potentially useful biomarkers for inflammatory bowel disease patients. *Heliyon* 2021, *7*, e06554.

[62] Sampath, P., Moideen, K., Ranganathan, U. D., Bethunaickan, R., Monocyte Subsets: Phenotypes and Function in Tuberculosis Infection. *Front Immunol* 2018, *9*, 1726.
[63] Zhang, J., Ye, J., Ren, Y., Zuo, J.*, et al.*, Intracellular activation of complement C3 in Paneth cells improves repair of intestinal epithelia during acute injury. *Immunotherapy* 2018, *10*, 1325-1336.
[64] Henao, M. P., Bewtra, M., Osterman, M. T., Aberra, F. N.*, et al.*, Measurement of Inflammatory Bowel Disease Symptoms: Reliability of an Abbreviated Approach to Data Collection. *Inflamm Bowel Dis* 2015, *21*, 2262-2271.
[65] Differentiating Ulcerative Colitis from Crohn Disease in Children and Young Adults: Report of a Working Group of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the Crohn's and Colitis Foundation of America. *Journal of Pediatric Gastroenterology and Nutrition* 2007, *44*, 653-674.
[66] Vancamelbeke, M., Vermeire, S., The intestinal barrier: a fundamental role in health and disease. *Expert Rev Gastroenterol Hepatol* 2017, *11*, 821-834.
[67] Soomro, S., Venkateswaran, S., Vanarsa, K., Kharboutli, M.*, et al.*, Predicting disease course in ulcerative colitis using stool proteins identified through an aptamer-based screen. *Nat Commun* 2021, *12*, 3989.
[68] Vitali, R., Palone, F., Armuzzi, A., Fulci, V.*, et al.*, Proteomic analysis identifies three reliable biomarkers of intestinal inflammation in the stools of patients with Inflammatory Bowel Disease. *J Crohns Colitis* 2022.

Table 1. Subject Demographics by Cohort and Group[a]

	Cohort 1			Cohort 2		
	HC	**UC**	**CD**	**HC**	**UC**	**CD**
Number of donors	5	5	5	20	10	10
Age (mean ± SD, years)	39.6 ±10.2	36.4 ±15.0	46.2 ±15.6	40.4 ±8.2	34.8 ±12.9	39.6 ±8.9
Female	0/5 (0%)	3/5 (60%)	3/5 (60%)	14/20 (70%)	3/13 (23.1%)	3/7 (42.9%)
Asian	3	-	-	8	-	-
Black	-	-	1	3	-	1
White	1	4	3	6	11	6
Hispanic	1	1	1	2	2	-
Middle Eastern	-	-	-	1	-	-
Mayo Score (UC)	N/A	7.8 ±2.0	N/A	N/A	9.2 ±2.0	N/A
SES-CD Score (CD)	N/A	N/A	5.8 ±1.3	N/A	N/A	7.1 ±0.7
fCAL *μ*g/g	12.0 ±12.1	3074 ±4395	627 ±354	38.9 ±57.4	246 ±414	282 ±339

[a]Summary of baseline demographics for each patient used in the study analysis by cohort and group.

FIGURES

Figure 1. Study design and overview. (A) Sample processing workflow: stool was collected from subjects, and protein was extracted and quantified. Proteins were digested into peptides, desalted, and analyzed by DIA-MS. Data were analyzed using Spectronaut to identify peptides and protein level quantitation was performed in MSstats. (B) Study design. Cohort 1 consisted of stool collected from n=5 UC patients, n=5 CD patients, and n=5 HC. Cohort 2 consisted of stool collected from n=10 UC patients, n=10 CD patients, and n=20 HC. Differentially abundant proteins from Cohort 1 and Cohort 2 were further investigated as potential biomarkers.

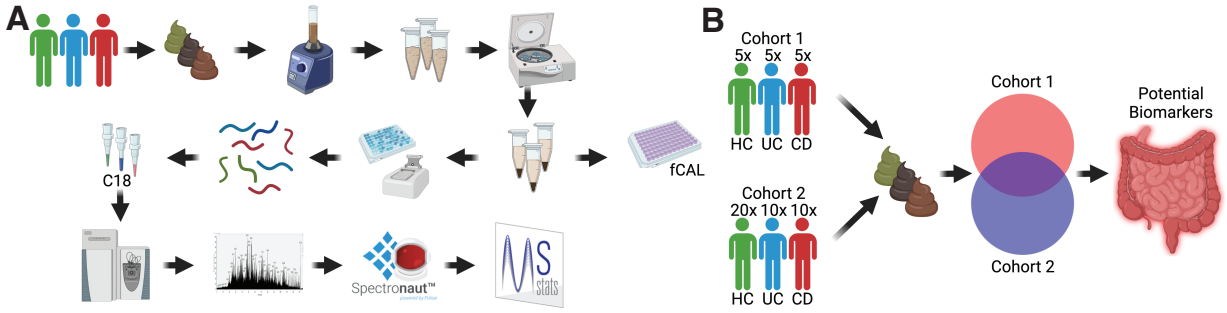

Figure 2. **Cohort 1 analysis.** (A) Heatmap analysis comparing protein abundances in UC, CD and HC samples. (B) PCA plot comparing sample similarity in UC, CD, and HC samples by principal component. Volcano plots comparing differentially abundant proteins in UC vs. HC (C) and CD vs. HC (D) are shown. The top 10 most changed proteins are labelled. (E) Upset plot showing distribution of shared proteins across UC, CD, and HC groups. (F) Pearson correlation of fCAL measured by immunoassay vs S100A8 and S100A9 detected by DIA-MS. The solid line is the linear regression line for S100A8 (blue) and S100A9 (red), dotted black lines are the 95% confidence intervals. (G) Box plots of S100A8 and S100A9 levels in UC, CD, and HC samples as measured by DIA-MS; statistical significance was determined by 2-way ANOVA analysis with Tukey's multiple comparison correction. All significant (p-value < 0.05) comparisons shown, ****, p-value <0.0001.

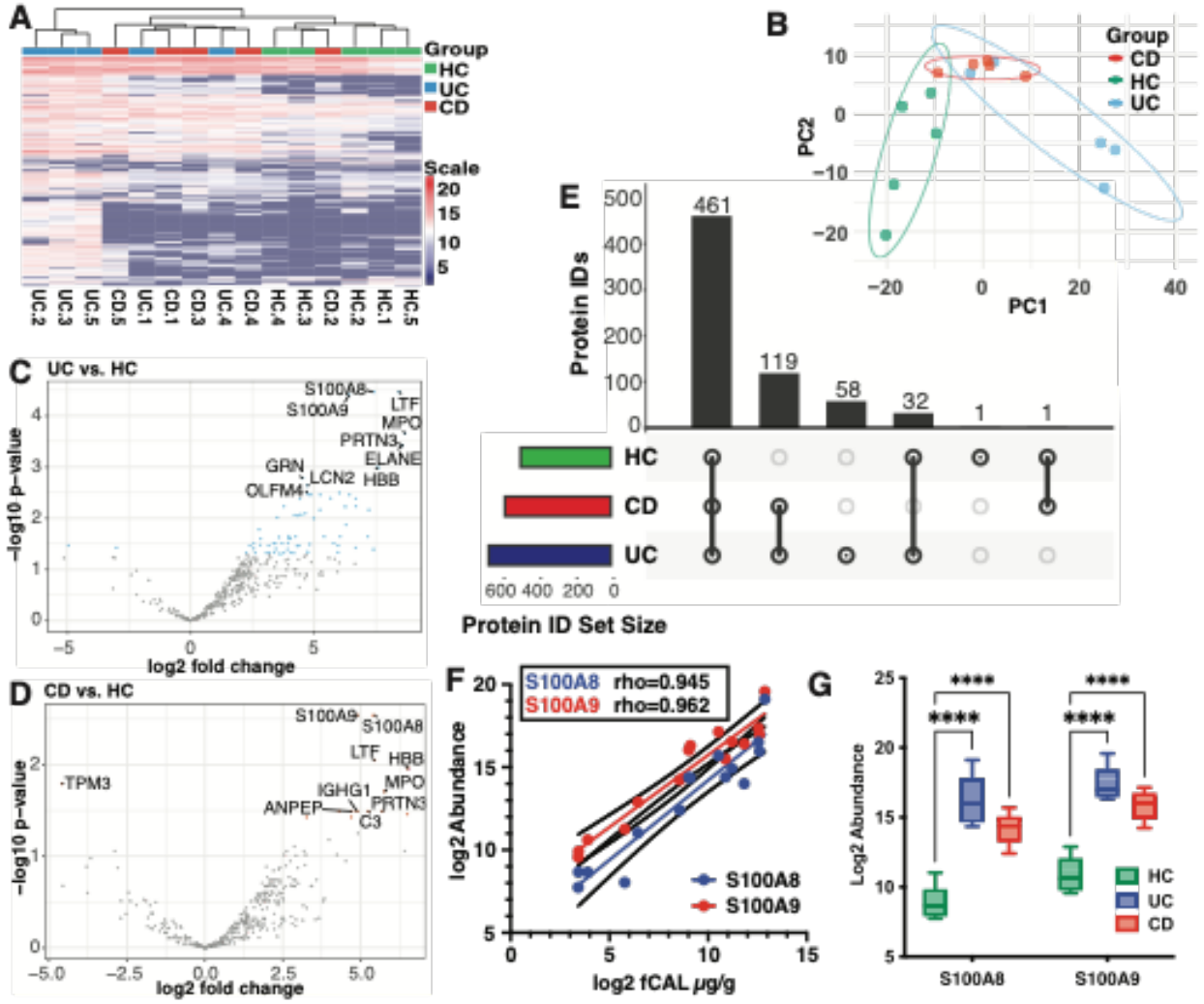

Figure 3. Comparison of DIA-MS findings across Cohort 1 and Cohort 2. Venn diagrams showing the overlap of significant, differentially abundant proteins from Cohort 1 and Cohort 2 for (A) UC versus HC, (B) CD versus HC, and (C) UC versus CD. Pearson correlations between log2 FC values of significant, differentially abundant, proteins from Cohort 1 versus Cohort 2 from 96 confirmed proteins from UC stool (D) and 126 confirmed proteins from CD stool (E). Dotted lines show the linear regression fit. Total protein identifications were compared between groups from both cohorts (F). Four-way plots show distribution and concordance of all proteins identified in Cohort 1 and Cohort 2 in (G) UC versus HC and (H) CD versus HC where red dots indicate proteins significantly changed in both cohorts, green dots indicate proteins significantly changed in Cohort 1, blue dots indicate proteins significantly changed in Cohort 2, and black dots indicate proteins present in both cohorts but not significantly different from HC.

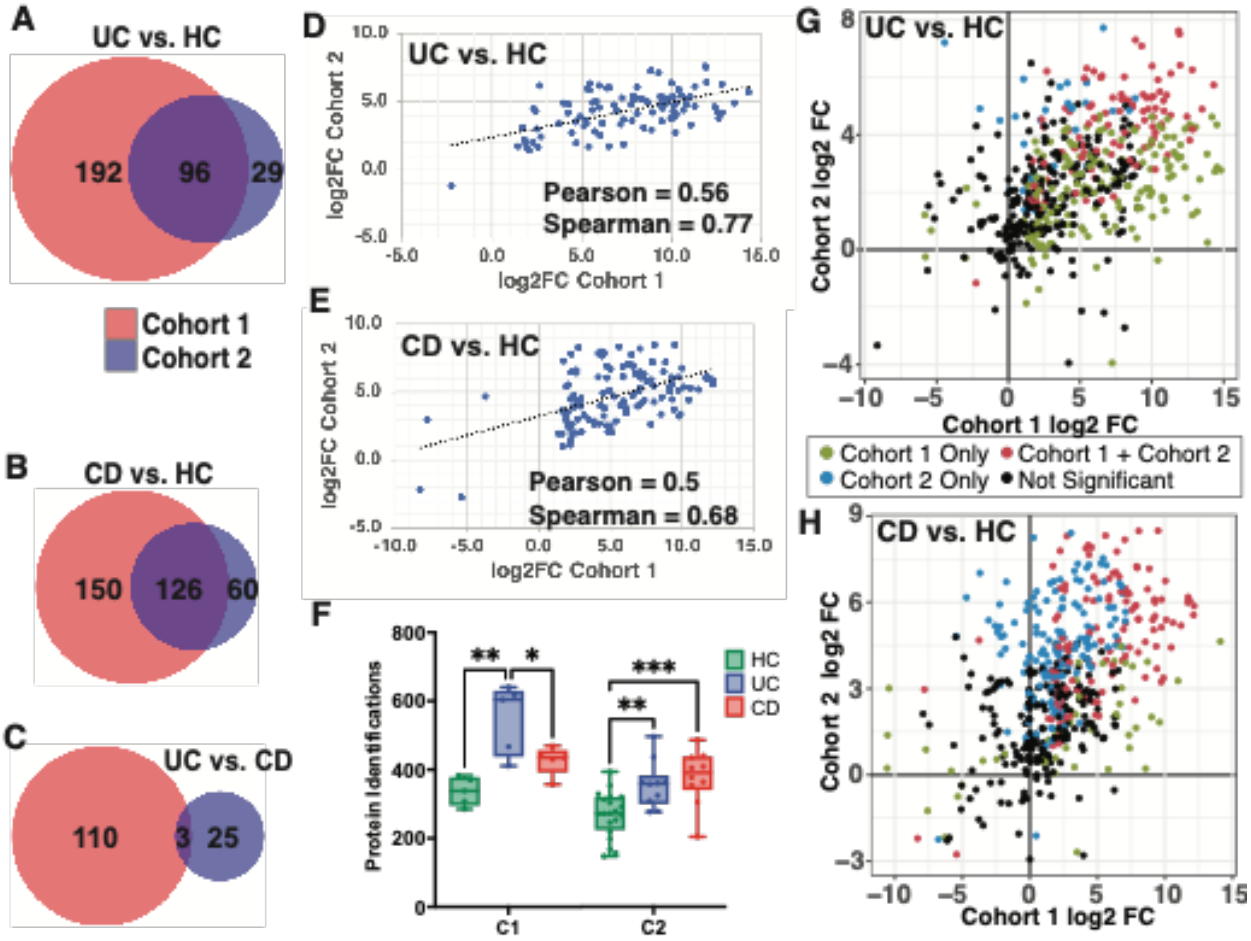

Figure 4. Gene expression profiles from proteins found to be upregulated in stool samples from (A) UC patients or (B) CD patients were assessed in using scRNAseq of whole blood from HC subjects. A heatmap containing expression levels of these proteins in T cells, neutrophils, monocytes, B cells, platelets, and NK cells are shown.

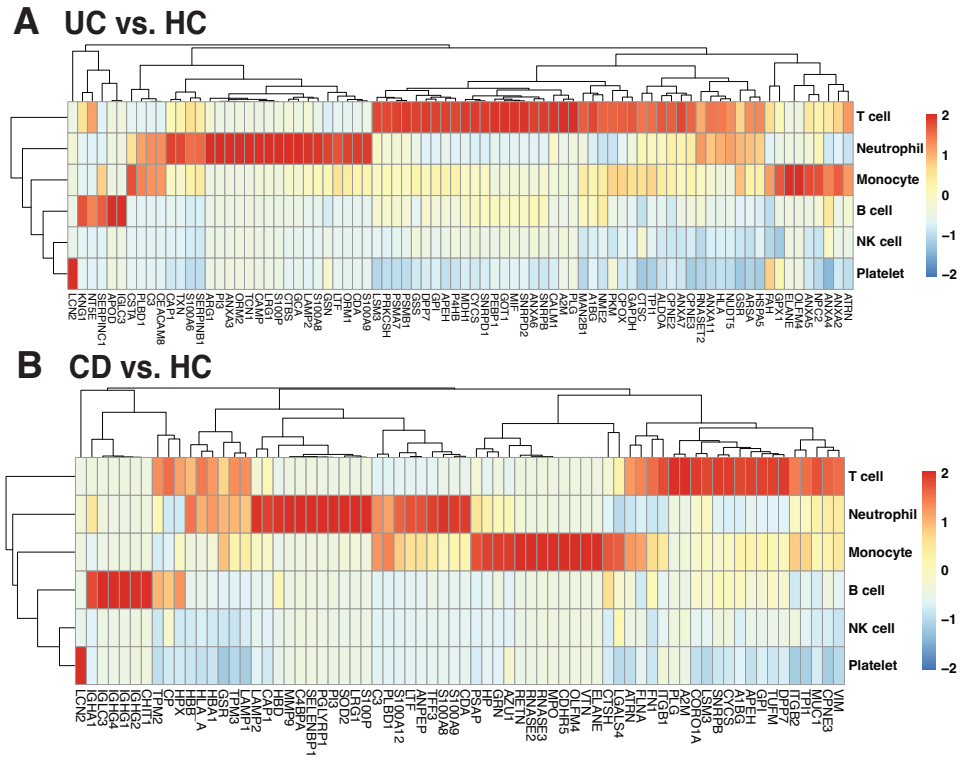

Figure 5. **Significantly changed proteins in IBD stool may reflect immune cell activity**. Cohort 1 and cohort 2 samples were combined in this analysis. Violin plots showing levels of immune cell enriched proteins in stool from UC patients (n= 15) and HC (n= 25) (A) and in stool from CD patients (n=15) and HC (n=25) (B). Proteins included are alpha-2 macroglobulin (A2M), Annexin A6 (ANXA6), Immunoglobulin G4 (IGHG4), complement 3 (C3), phospholipase B domain containing 1 (PLBD1), S100 calcium binding protein P (S100P), leucine rich alpha-2-glycoprotein 1 (LRG1), and matrix metallopeptidase 9 (MMP). Two-way ANOVA analysis with Siadak's multiple comparison correction was used to test statistical difference, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value <0.0001.