SUPPLEMENTARY MATERIALS AND METHODS

Colonic mucosal-luminal interface (MLI) aspirate sample collection

Colonic mucosal luminal interface (MLI) aspirates were obtained at time of diagnostic colonoscopy following a standard 1 day clean-out preparation[1]. Sampling occurred at the mid-ascending colon (AC) and/or at the site of the lower descending colon and upper sigmoid colon region (DC), and annotated to be from a normal, non-inflamed (CoN) or an affected, inflamed region (CoA) based on macroscopic evaluation. Briefly, upon insertion of the colonoscope, initial fluid and debris in the fluid were aspirated away. Thereafter sterile water was flushed onto the mucosa of the selected region and the fluid was aspirated into sterile collection vials as the MLI aspirate sample. These samples were used for analysis and the order of collection was distal to proximal sites. The samples were put on ice in the endoscopy room and immediately delivered to laboratory for further processing. A complete protease inhibitor cocktail (Roche Diagnostic GmbH, Mannheim, Germany) was added to the intestinal aspirates upon receipt in the lab. Following debris depletion by centrifugation at 700g for 5min at 4°C, the supernatant was subsequently subjected to 14,000g centrifugation for 20 minutes at 4°C to remove bacteria. The resultant supernatant was then filtered through a 0.2μm syringe driven filter for removal of any residual bacterial cells and stored at -80°C.

Heavy isotopic-labeled reference proteome

Heavy reference proteins for quantification were prepared from 5 isotopically-labeled commercially available human cell lines, namely lymphocytic Jurkat (ATCC), HEK-293 (ATCC), colorectal carcinoma HCT 116 (ATCC), monocytic THP-1 (ATCC) and hepatic HuH-7 (JCRB Cell Bank). HCT116, HuH-7 and HEK293 cells were grown in custom prepared media as previously described [2]. THP-1 and Jurkat cells were grown in RPMI media (#0422 AthenaES

Baltimore, MD, USA) supplemented with 15 mg/L methionine, 40 mg/L [13C6,15N2]-L-lysine, 200 mg/L [13C6,15N4]-L-arginine (Sigma Aldrich, Oakville, ON, Can), 10% dialyzed FBS (GIBCO-Invitrogen; Burlington, ON,CAN), 1 mM sodium pyruvate (Gibco-Invitrogen), 0.0059g/L Phenol Red (Sigma Aldrich, Oakville, ON, Can) and 28 µg/mL gentamicin (Gibco-Invitrogen). Heavy amino acid incorporation (>95%) was confirmed by MS analysis as previously described [3]. For protein isolation, cells were lysed in lysis buffer (4% SDS, 50 mM Tris, pH 8.0, protease inhibitor cocktail (Roche Diagnostic GmbH, Mannheim, Germany)) and sonicated three times in 10 second pulses with a 30 second incubation on ice between pulse intervals. Lysates were centrifuged at 10,000g for 10 minutes, supernatants collected and protein concentrations were quantified by DC (detergent compatible) protein assay (BIORAD, California, USA). The heavy internal reference was prepared once, aliquoted and stored at -80°C and thawed just prior to combining with colonic aspirate proteins.

Processing of colonic MLI aspirate proteins for mass spectrometry analysis

Filtered colonic MLI aspirates were thawed on ice, and proteins precipitated overnight with trichloroacetic acid (20% v/v). Protein pellets were washed three times with acetone (100%) and dried before resuspension and sonication in lysis buffer, as described above. Protein concentration was quantified by DC protein assay (BIORAD, California, USA). Colonic aspirate proteins (45µg) were combined with an equal amount of heavy isotopic-labeled cell lysate (9µg of each cell type), used as a representative internal standard to permit quantitative proteomic data, an approach known as Super-SILAC (stable-isotope labeling by amino acids in cell culture) [4]. The mixture containing proteins from colonic aspirates and internal reference cells was digested with trypsin by filter aided sample preparation method (FASP) [5], fractionated

into 5 fractions (pH 4, 6, 8, 10 and 12) using SCX resin (Agilent Technologies, CA, USA), and desalted with a 10 µm AQUA-C_{18} resin (Dr Maisch, GmbH, Ammerbuch, Germany).

Stool collection and protein extraction from stool

Stool samples were collected from treatment-naive participants within 8 weeks of diagnostic colonoscopy with 82.2% collected within 4 days of colonoscopy (supplemental table 3). This cohort includes both independent participants (n= 26) and participants for which their MLI samples were utilized for biomarker discovery (n=15). Stool samples were frozen after collection and brought into the clinic on ice. Extraction buffer (50mM Tris, pH 7.2, 150mM NaCl, protease inhibitor cocktail (Complete, Mini (Roche Diagnostic GmbH, Mannheim, Germany)) was added to stool in a 5:1 ratio (extraction buffer volume: stool weight). The resultant slurry was mixed by agitation for 30 seconds, followed by rotation at 4°C for 20 minutes. Following centrifugation for 20 minutes at 10,000g at 4°C to remove large sediment, the supernatant was filtered (0.2µm) and the protein-containing filtrate was collected and stored at -80°C. Protein concentration was assessed by BCA protein assay kit (ThermoFisher Scientific, San Jose, CA).

LC-MS/MS and Bioinformatic analysis

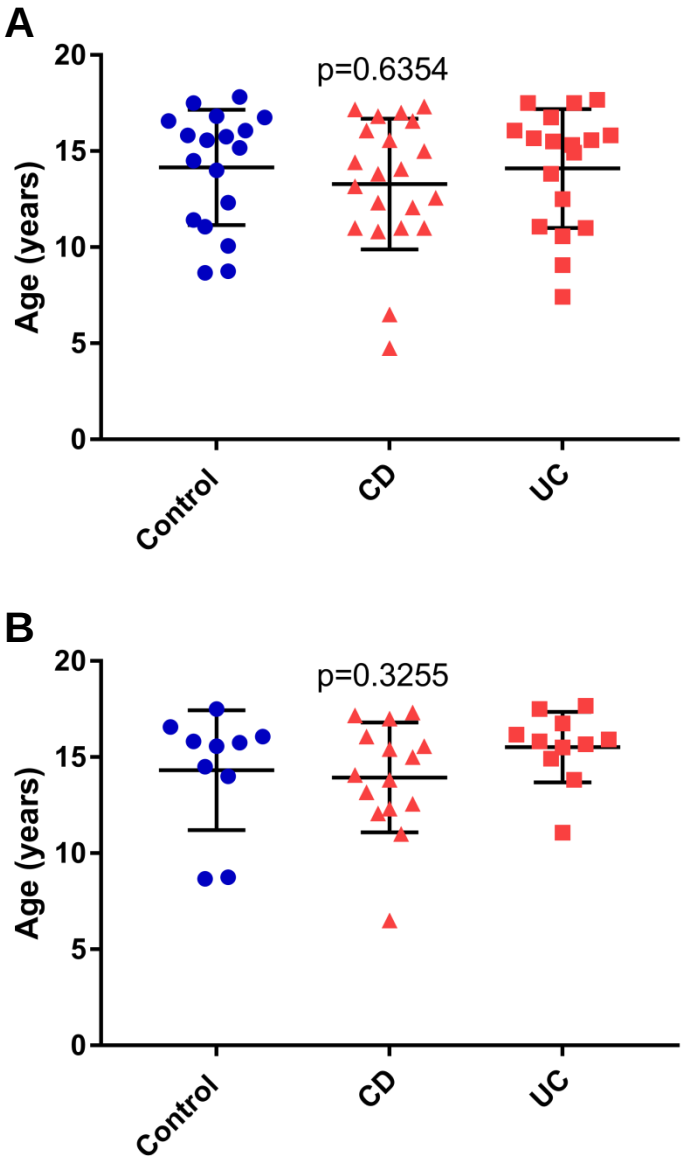
High-performance liquid chromatography/electrospray ionization tandem mass spectrometry (HPLC-ESI-MS/MS) was performed as previously described[6] using an Ekspert nanoLC 400 (Eksigent, Dublin, CA, USA) coupled to an Orbitrap ELITE MS (ThermoFisher Scientific, San Jose, CA).

Peptides were assigned and quantified using MaxQuant version 1.5.3.30 [7] in a single run against the human Uniprot database (downloaded 2012/07/11). Patients with inconclusive IBD diagnosis at the time of sample collection were excluded from downstream bioinformatics analysis. The following parameters were used: a multiplicity of two with Arg10 and Lys8 selected as the heavy

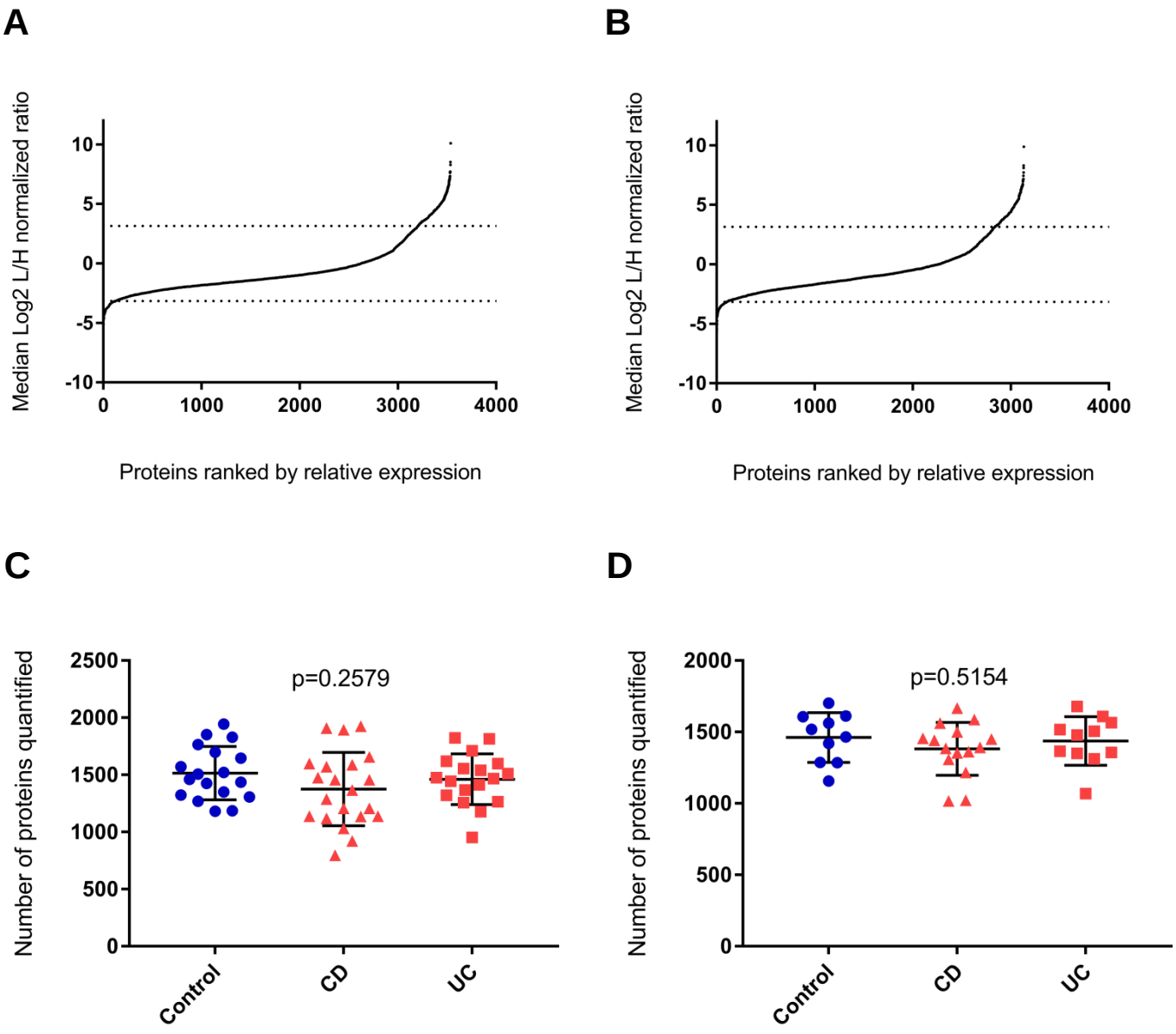
labels; a specific digestion mode was implemented with trypsin selected as the enzyme with a maximum of two missed cleavages; cysteine carbamidomethylation as a fixed modification; methionine oxidation and acetylation (protein N-termini) as variable modifications; the re-quantify and match between runs parameters were enabled; minimum peptide length of seven amino acids; ion mass tolerance of 0.5 Da; protein and peptide false discovery rate (FDR) of 1%. The AC and DC proteomes were analyzed separately post database search.

To minimize the effects of overfitting, MetaboAnalyst uses the Monte-Carole cross validation approach (MCCV) sub-sampling wherein at random 2/3 of the samples are used to evaluate the importance of the features, and these are then validated on the remaining 1/3 of samples; this process is repeated 50 times.

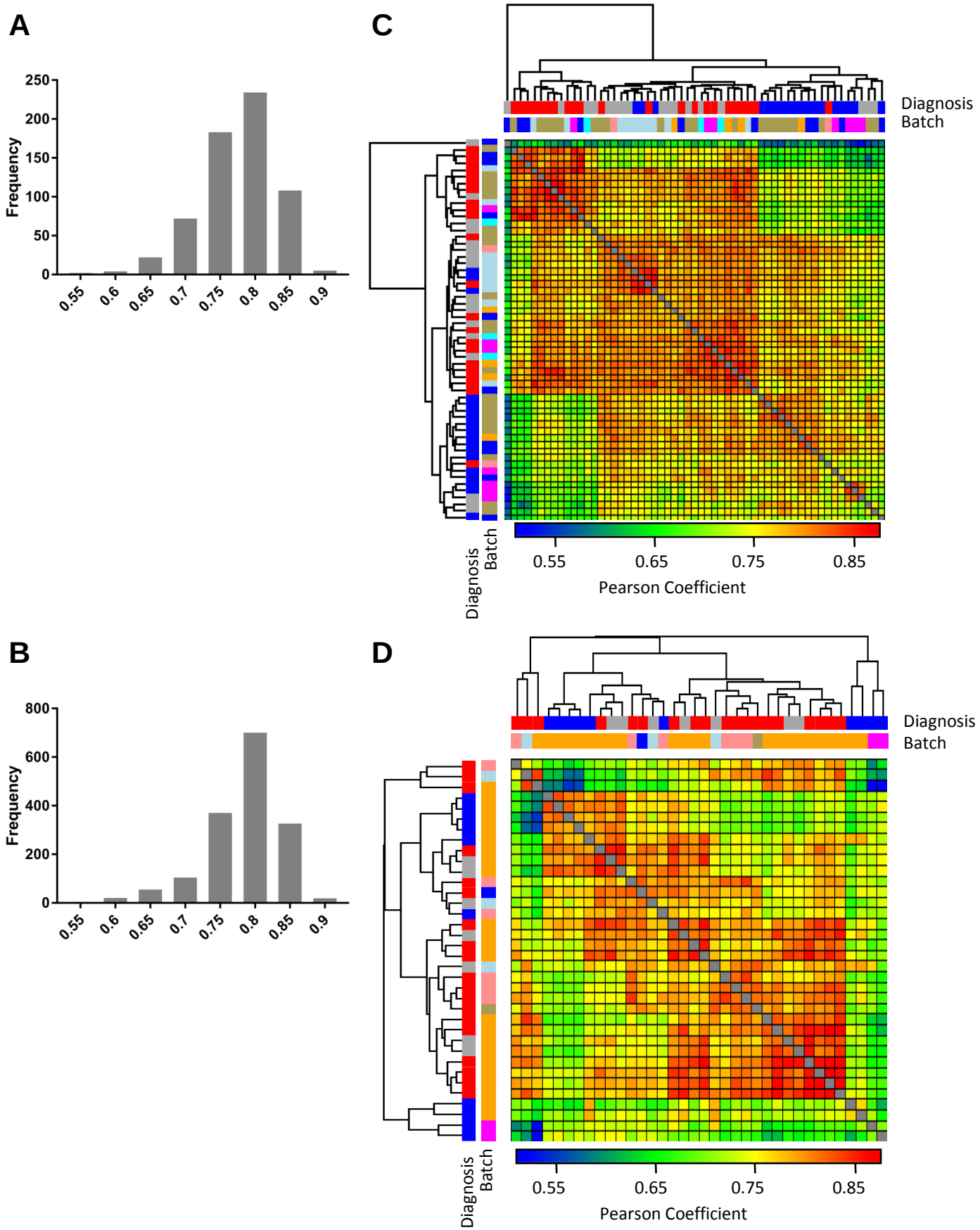
1. Jimenez-Rivera C, Haas D, Boland M, et al. Comparison of two common outpatient preparations for colonoscopy in children and youth. Gastroenterol Res Pract 2009;2009:518932.
2. Mottawea W, Chiang C-K, Mühlbauer M, et al. Altered intestinal microbiota–host mitochondria crosstalk in new onset Crohn's disease. Nature Communications 2016;7:13419.
3. Ong SE, Mann M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc 2006;1:2650-60.
4. Geiger T, Cox J, Ostasiewicz P, et al. Super-SILAC mix for quantitative proteomics of human tumor tissue. Nat Methods 2010;7:383-5.
5. Wisniewski JR, Zougman A, Mann M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J Proteome Res 2009;8:5674-8.
6. Starr AE, Deeke SA, Ning ZB, 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.
7. Cox J, Matic I, Hilger M, et al. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc 2009;4:698-705.



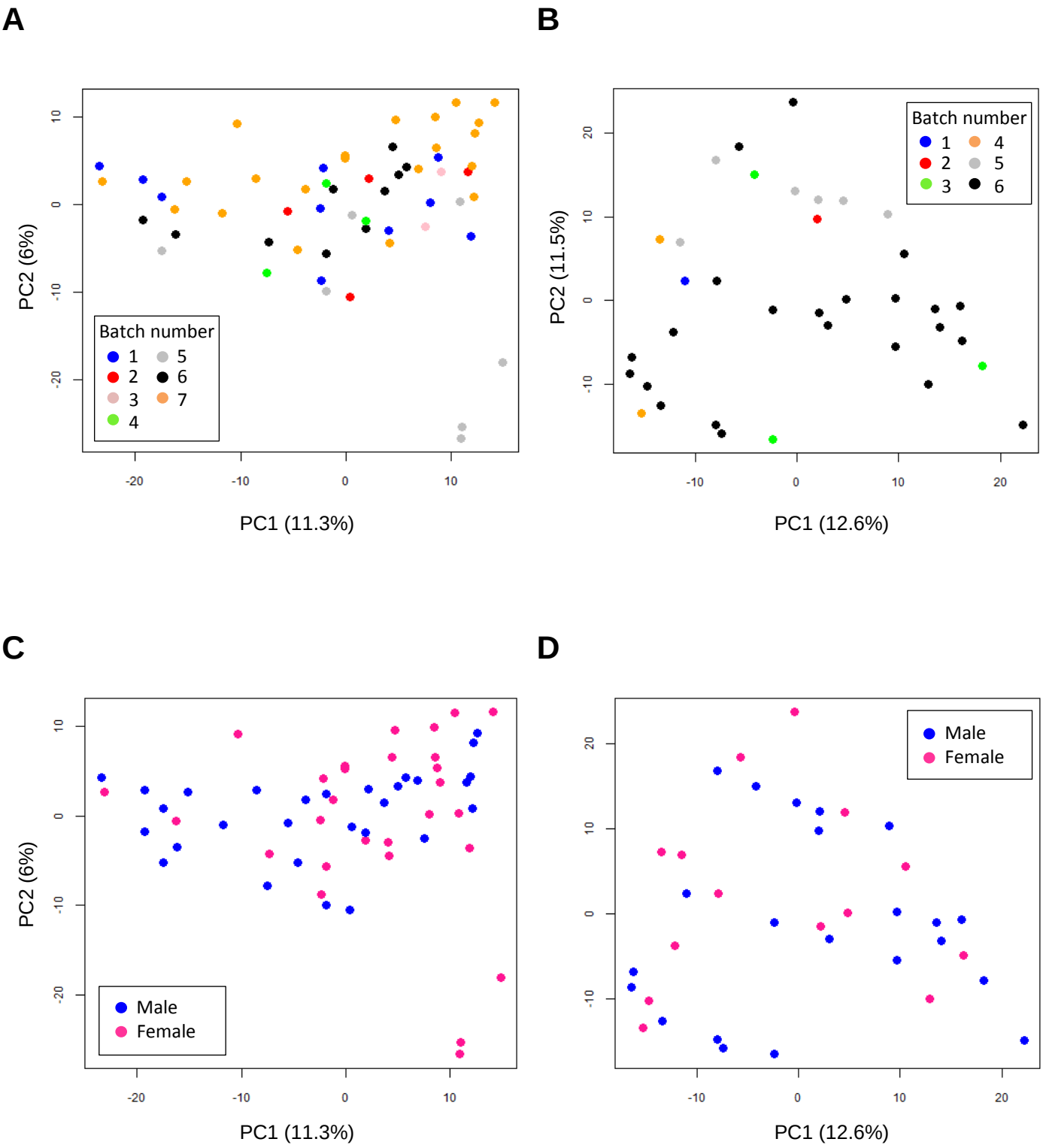
Supplementary Figure 1: Patient cohort characteristics. Age of patients included in cohort for the (A) AC and (B) DC . No significant differences were observed between patient subgroups by one-way ANOVA. CD = Crohn's disease, UC = ulcerative colitis



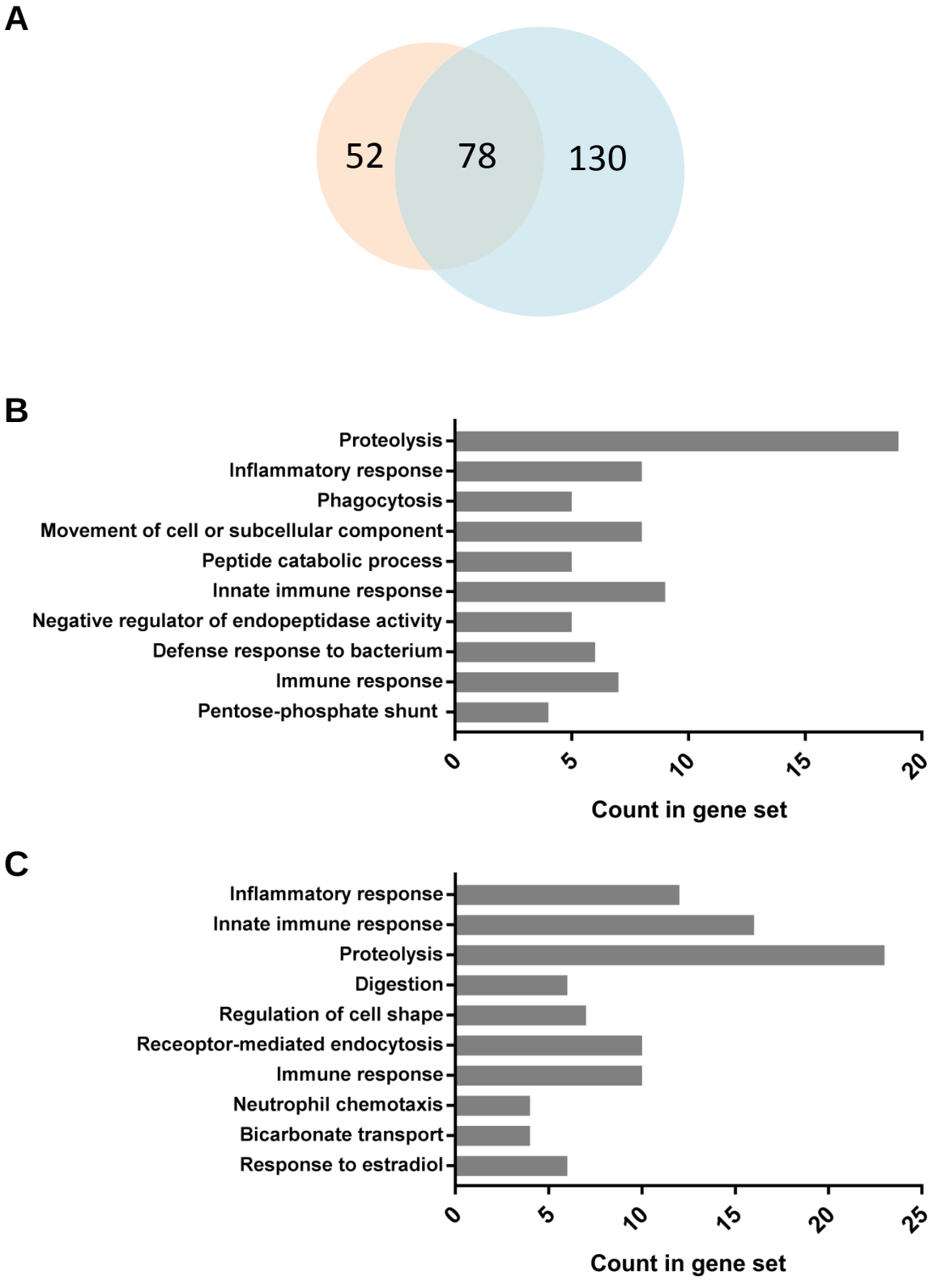
Supplementary Figure 2: MS Data Evaluation. Heavy reference proteome evaluation; median Log2 L/H normalized ratio (MLI proteins/super-SILAC reference proteome) of proteins quantified in the (A) AC and (B) DC . Dotted lines indicate 10-fold ratio threshold. Number of proteins quantified per patient in the (C) AC and (D) DC. No significant differences in the number of proteins were observed between patient subgroups by one-way ANOVA.



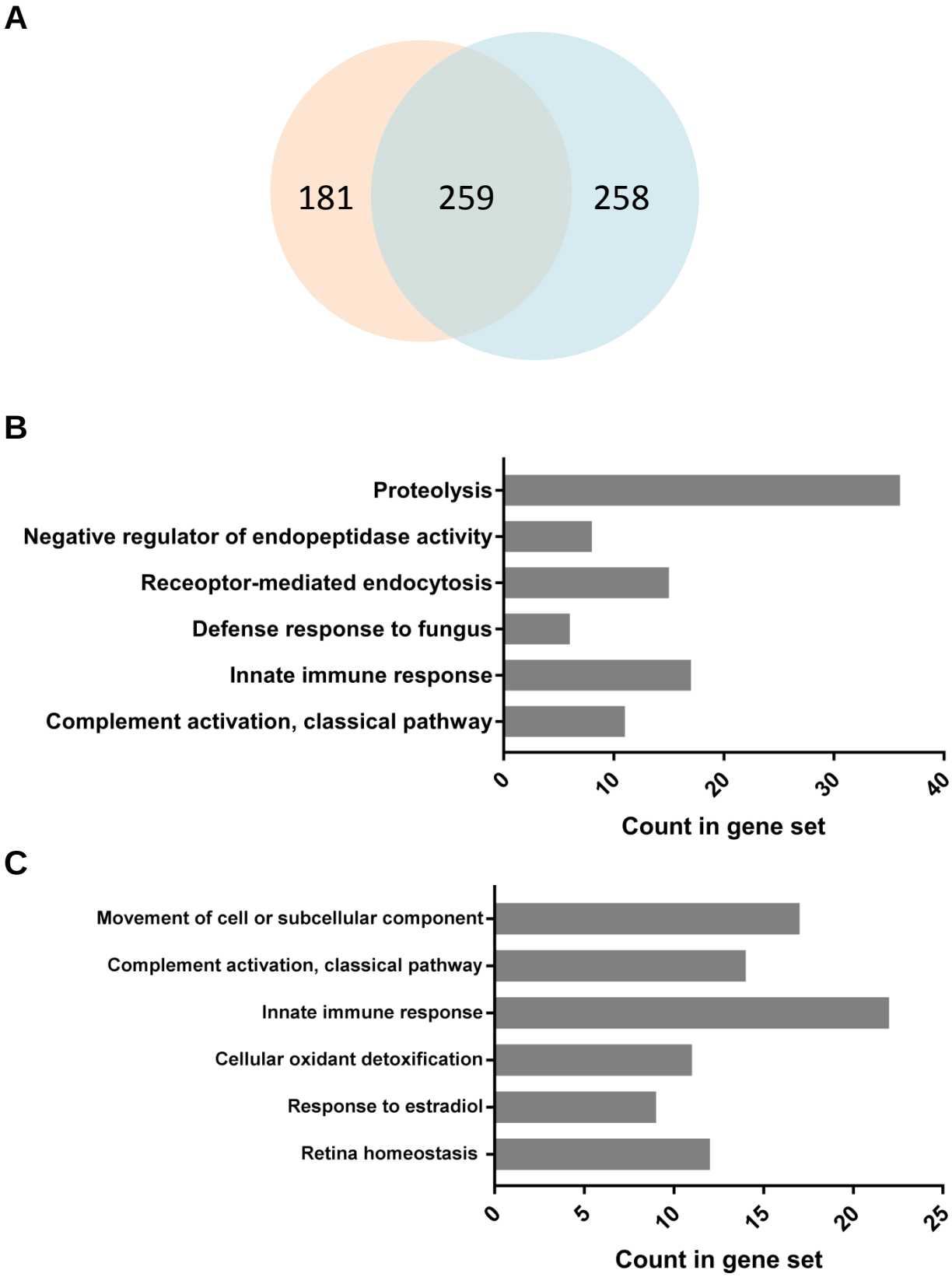
Supplementary Figure 3 : MS Data Evaluation. Pearson correlations of Q75 proteome log2 (light/heavy) from (A) AC and (B) DC MLI samples. Hierarchical clustering of Pearson correlation (C) AC (D) DC. Diagnosis: blue = control, grey = IBD without macroscopic inflammation , red = IBD with macroscopic inflammation. MS batch over time is indicated by blocks of colour .



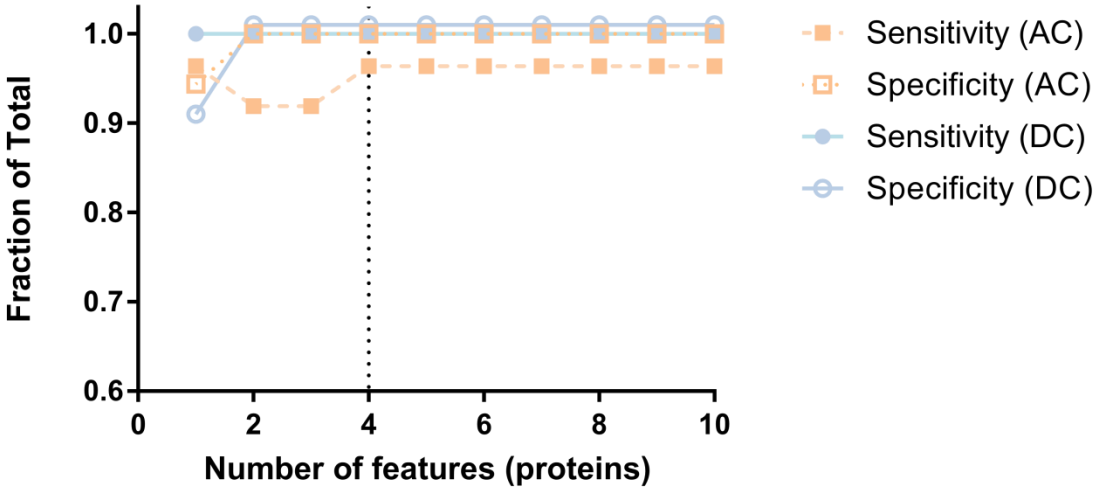
Supplementary Figure 4: MS Data Evaluation. PCA of Q75 (A) and (C) from the AC and (B) and (D) from the DC . Intestinal MLI aspirate samples segregate according to diagnosis status rather than by batch (A) and (B) or sex (C) and (D) in either colon sub-region.



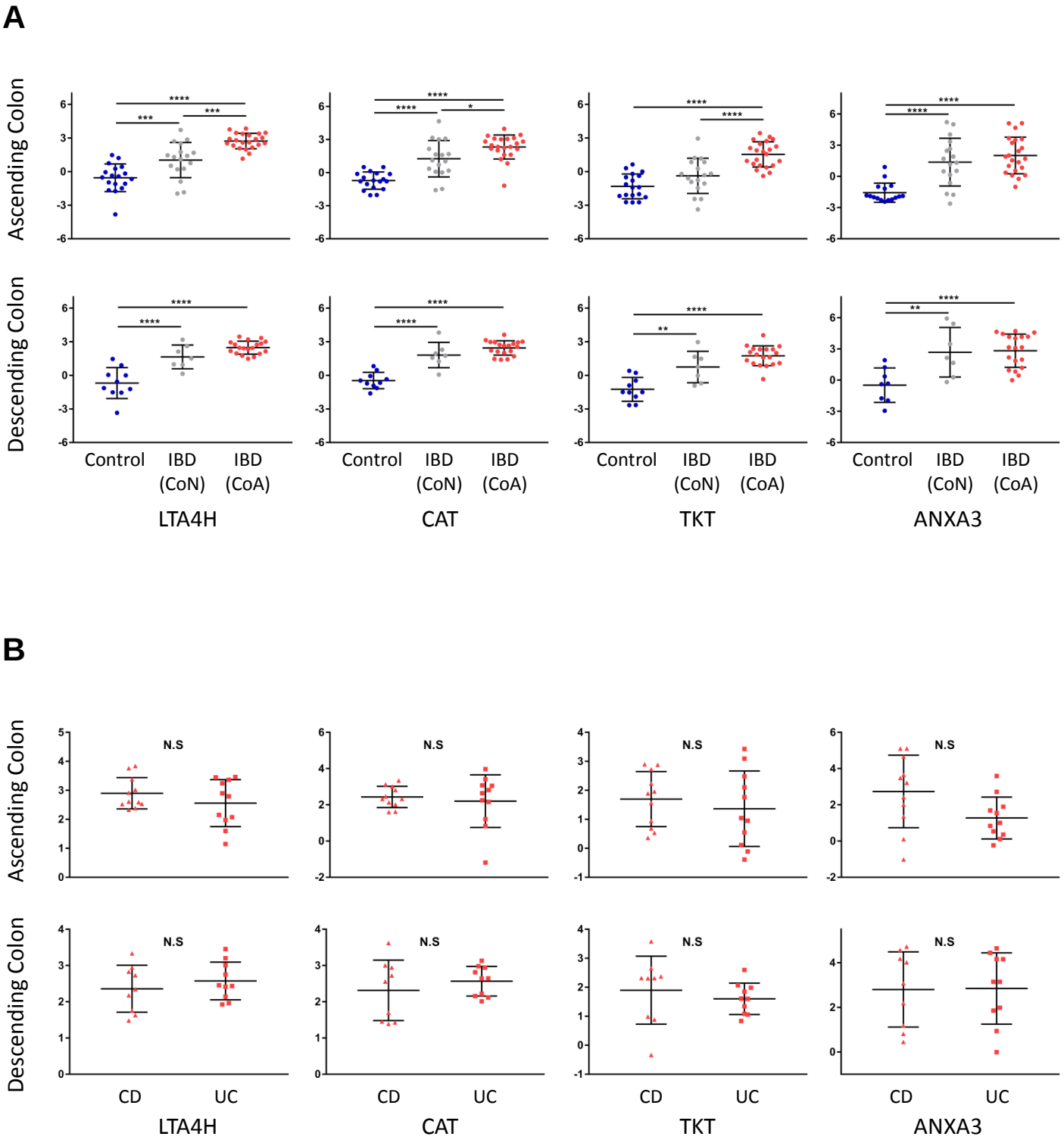
Supplementary Figure 5: Proteomic landscape evaluation at the colonic MLI. (A) Number of features identified in AC (orange) and DC (blue) between control patients and IBD patients with macroscopic evidence of inflammation by PLS-DA. Top ten biological processes of discriminant features identified by comparison of control and IBD CoA patient samples by PLSDA , enrichment is relative to Q75 from the (B) AC and (C) DC .



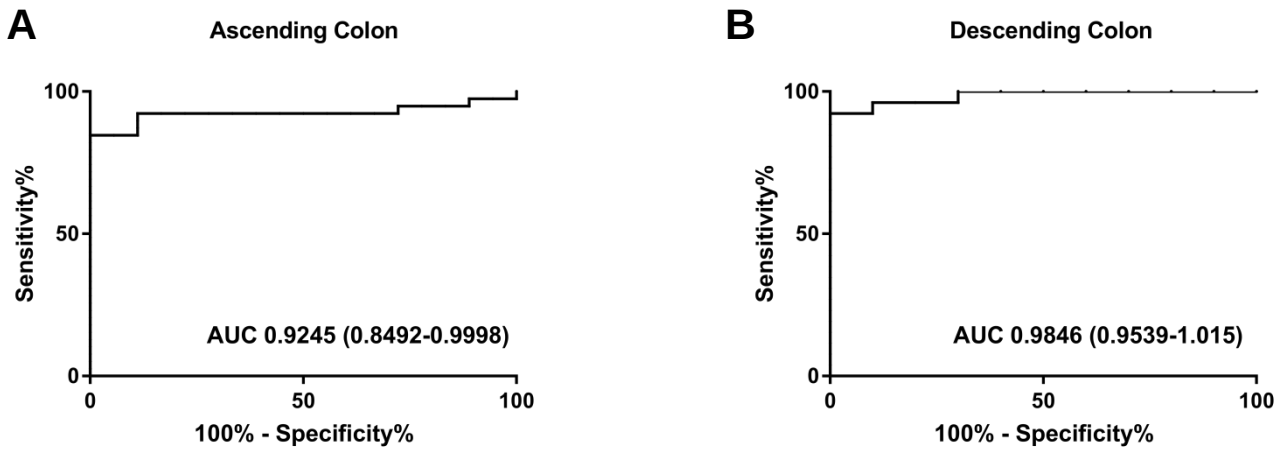
Supplementary Figure 6 : Proteomic landscape alteration in IBD CoN compared to control (A) Number of features identified in AC (orange) and DC (blue) between control and IBD patient samples without macroscopic evidence of inflammation (CoN) by PLS-DA. (B) AC (C) DC of top biological processes of discriminant features identified by comparison of control and IBD CoN patient samples by PLSDA; enrichment is relative to the Q75.



Supplementary Figure 7 : IBD Biomarker panel generation. Minimum number of proteins required to achieve maximum sensitivity and specificity in both colon sub-regions (indicated by the doted line). Protein biomarker candidates were added based on highest combined (AC and DC)AUC values , using the PLSDA model to classify controls from IBD CoA.

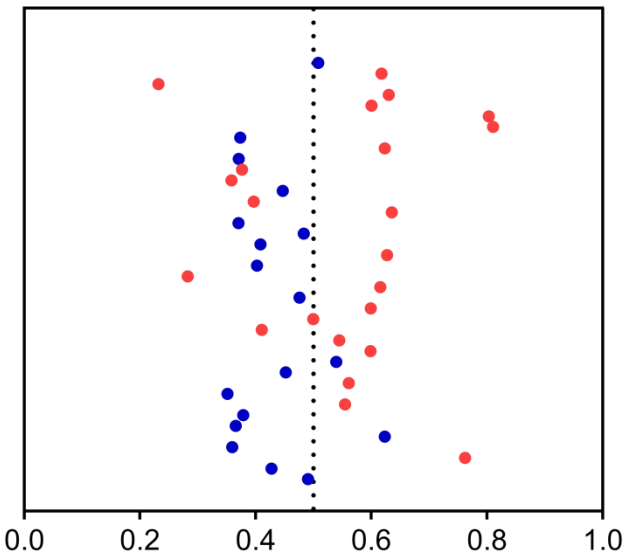


Supplementary Figure 8 : Relative expression of proteins featuring in the IBD diagnosis biomarker panel ; (A) comparison with MLI samples lacking macroscopic evidence of inflammation (CoN) and (B) between CD (CoA) and UC (CoA). One-way ANOVA with Tukey's multiple comparison test * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$



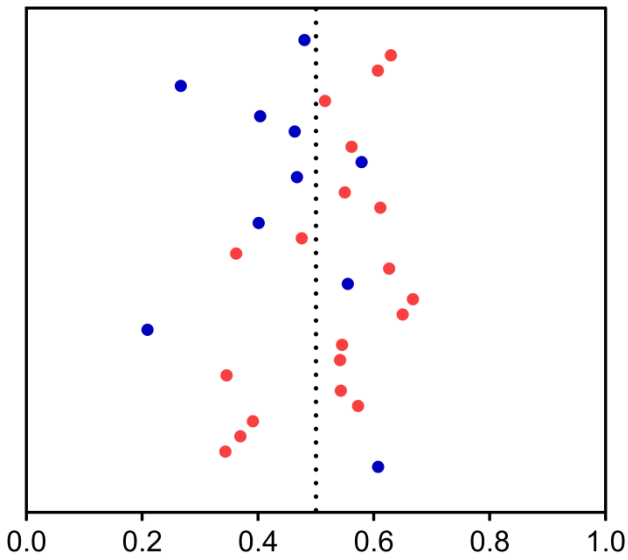
Supplementary Figure 9 : Application of the active IBD biomarker panel to IBD samples with and without evidence of macroscopic inflammation: ROC curve for the (A) AC and (B) DC.

A



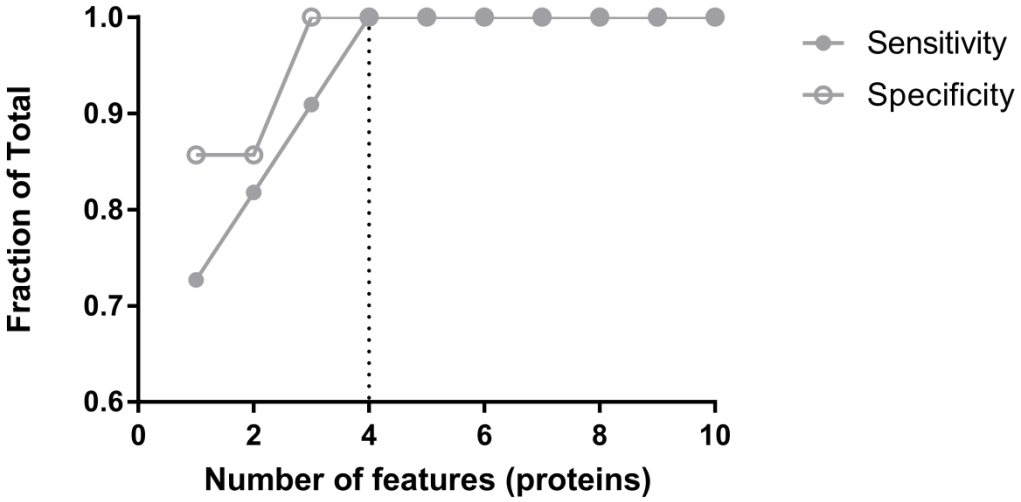
	True Control	True IBD
Test Control	15	7
Test IBD	3	15

B

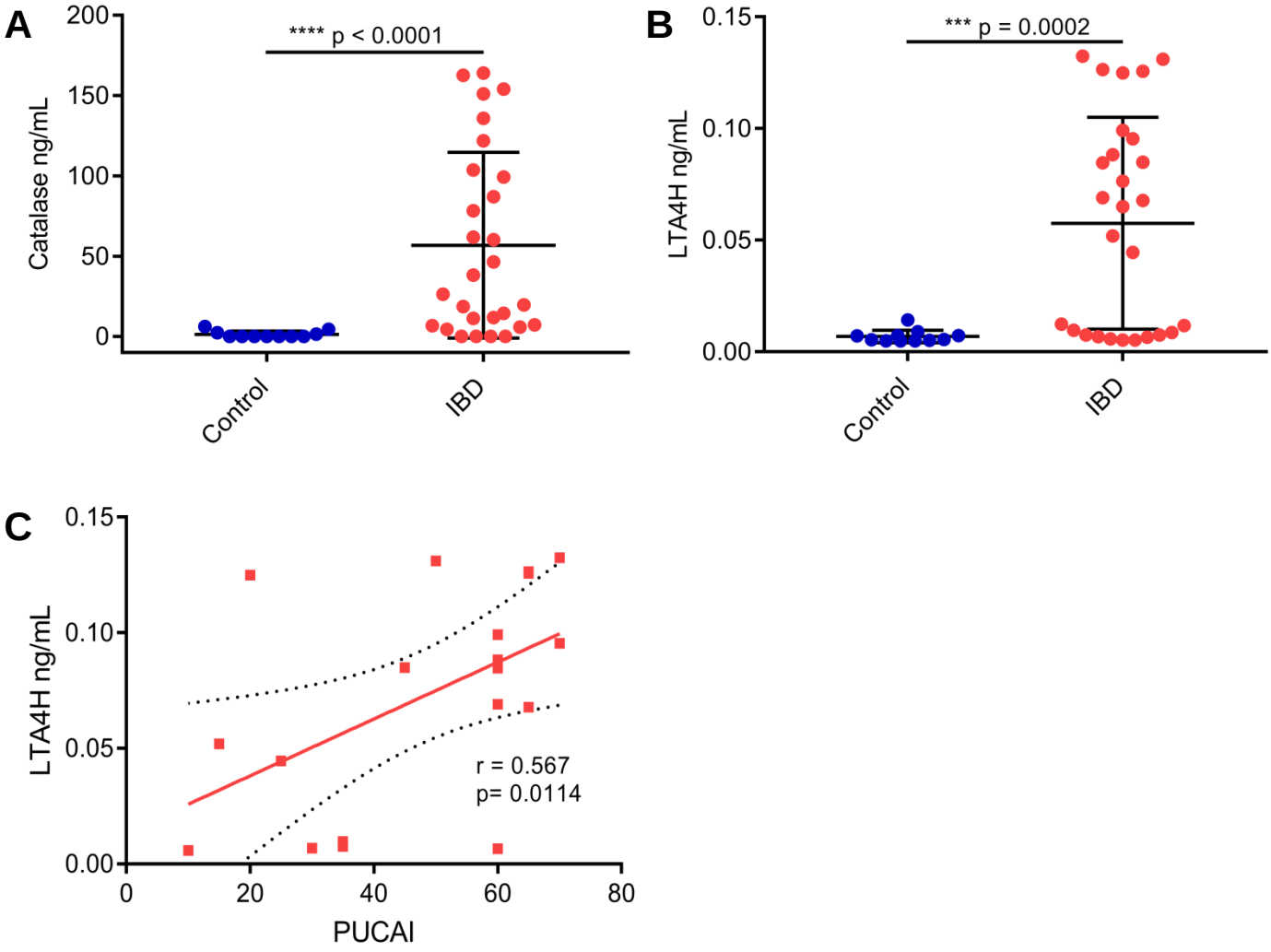


	True Control	True IBD
Test Control	7	6
Test IBD	3	13

Supplementary Figure 10 : Predictive class probabilities of Calprotectin (S100-A8 and S100-A9) at the MLI using PLS-DA in the (A) AC and (B) DC .



Supplementary Figure 11 : Minimum number of proteins required to achieve maximum sensitivity and specificity (indicated by the doted line) for extent of disease biomarker panel (pancolitis vs non-pancolitis). Protein biomarker candidates added based on highest AUC values , using the PLSDA model to classify UC CoN from UC CoA in the ascending colon.



Supplementary Figure 12 : Validating differential protein expression in stool: (A) Catalase and (B) Leukotriene A-4 hydrolase, biomarker candidates proposed for the diagnosis of pediatric IBD, were assessed by ELISA from a cohort consisting of independent patients and patients for which their MLI samples were utilized to develop the biomarker panel. P values were calculated using the Mann Whitney test. (C) The expression level of LTA4H in stool correlates with the PUCDAI. Analysis performed using Spearman two-tailed.