Serum proteomics in paediatric inflammatory bowel disease from a case–control study: biomarker discovery and ulcerative colitis–Crohn's disease differentiation

Mmeyeneabasi Omede,[a,f,h] Hasan H. Otu,[b,h] Laurie B. Grossberg,[c,f] Jui-Yen Huang,[a,f] Xuesong Gu,[d,e,f] Simon T. Dillon,[d,e,f] Handan Can,[b] Tina Lung Morhardt,[g] Catherine Pursley,[a] Harland S. Winter,[a,f,i] and Towia A. Libermann[d,e,f,*,i]

[a]Center for Pediatric Inflammatory Bowel Disease, Mass General Hospital for Children, Pediatric Gastroenterology, Hepatology and Nutrition, Harvard Medical School, Boston, MA, USA
[b]Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA
[c]Division of Gastroenterology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA
[d]Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Boston, MA, USA
[e]Beth Israel Deaconess Medical Center Genomics, Proteomics, Bioinformatics and Systems Biology Center, Boston, MA, USA
[f]Harvard Medical School, Boston, MA, USA
[g]Dartmouth Hitchcock Clinics, Manchester, Geisel School of Medicine, Dartmouth, NH, USA

eBioMedicine
2026;128: 106311
Published Online xxx
https://doi.org/10.1016/j.ebiom.2026.106311

Summary

Background The diagnosis of inflammatory bowel disease (IBD), ulcerative colitis (UC), and Crohn's disease (CD), relies on clinical and pathological criteria. Non-invasive precision medicine tools to diagnose IBD and discriminate between UC and CD are needed to personalise management. Serum proteomics identified protein biomarkers capable of diagnosing IBD and differentiating CD from UC subtypes.

Methods We obtained serum samples from a retrospective study of 47 patients with IBD and non-IBD patients seen in a tertiary care paediatric gastroenterology clinic and applied SomaScan proteomics to measure 1305 proteins to discriminate between IBD and non-IBD and UC and CD. Four proteins were further validated by immunoassays in two retrospective cohorts of 295 and 105 individuals and multi-protein predictors were developed using Support Vector Machines (SVM).

Findings The SomaScan discovery phase identified 95 serum protein biomarkers (BH $p < 0.01$) that differentiated IBD from non-IBD and 70 proteins ($p < 0.01$) that distinguished UC from CD. Pathway analysis linked specific inflammatory processes and vascular functions to IBD and UC versus CD. An 8-protein classifier achieved an AUC of 0.95 for identifying IBD. Significant elevation of four key predictor proteins (MMP1, MMP3, Resistin, Haptoglobin) in IBD was validated by ELISA in the expanded cohort (N = 295). The 4-protein SVM predictor achieved an AUC of 0.86 and 0.90 for IBD discrimination in two independent cohorts. A separate 4-protein SVM predictor for differentiating UC from CD achieved an AUC of 0.93 in independent validation.

Interpretation Patients with paediatric-onset IBD have a unique serum protein signature associated with pro-inflammatory and vascular pathways. Additional studies are needed to determine whether these dysregulated proteins can be used in conjunction with traditional risk factors to support non-invasive biomarkers that identify IBD and discriminate between its subtypes.

Funding Martin Schlaff, The Diane and Dorothy Brooks Foundation, The Manessis Family, Thomas and Lynn Kuzma, and The Hasso Serrano Foundation.

Copyright © 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Corresponding author. Beth Israel Deaconess Medical Center, 99 Brookline Avenue, Boston, MA 02115, USA.
E-mail addresses: tliberma@bidmc.harvard.edu (T.A. Libermann), momede@mgh.harvard.edu (M. Omede), hotu2@unl.edu (H.H. Otu), lgrossbe@bidmc.harvard.edu (L.B. Grossberg), jhuang69@mgh.harvard.edu (J.-Y. Huang), xgu@bidmc.harvard.edu (X. Gu), sdillon1@bidmc.harvard.edu (S.T. Dillon), handancan03@gmail.com (H. Can), tinaklung@gmail.com (T.L. Morhardt), katiepursley13@gmail.com (C. Pursley), hwinter@mgh.harvard.edu (H.S. Winter).
[h]Equal authors.
[i]Co-senior authors.

Keywords: SomaScan; Inflammatory bowel disease; Crohn's disease; Ulcerative colitis; Proteomics; Paediatrics

Research in context

Evidence before this study

Inflammatory bowel disease (IBD) is a chronic, relapsing-remitting condition of the gastrointestinal tract with frequent onset in children and adolescents. Distinguishing functional abdominal pain from IBD can be challenging; diagnostic tests including MR enteroscopy and colonoscopy are often ordered. Most prior blood-based IBD diagnostics are based on individual protein biomarkers or a small set of molecules and have not yet achieved the diagnostic accuracy needed for standard-of-care IBD diagnostics. However, recent advances in technology have led to the emergence of high-plex proteomics as a next-generation tool for biomarker discovery and development of precise multiprotein diagnostic assays.

Added value of this study

Our study used comprehensive SomaScan proteomic profiling with immunoassay validation to investigate the underlying pathophysiology of paediatric IBD. New serum protein biomarkers were identified and used to develop predictive models that accurately distinguish patients with IBD from patients without IBD and differentiate ulcerative colitis from Crohn's disease. These findings strengthen the evidence supporting precision medicine approaches in IBD and suggest that serum-based biomarkers may enable non-invasive diagnostics and inform the development of new therapeutic targets.

Implications of all the available evidence

Our findings measuring proteomic profiles in blood samples demonstrate the utility of proteomics in identifying disease-specific biomarkers and new tests for differential diagnosis in paediatric IBD as well as defining the underlying pathophysiological mechanisms. Future large-scale prospective replication studies in independent paediatric and adult populations are warranted to validate these biomarkers and further explore their role in differentiating IBD phenotypes and subtypes. Such advances could inform individualised treatment strategies and improve the accuracy of non-invasive diagnostics.

Introduction

Inflammatory bowel disease (IBD), consisting of ulcerative colitis (UC), Crohn's disease (CD), and IBD unclassified (IBD-U), are chronic relapsing gastrointestinal autoimmune diseases characterised by mucosal ulceration, bleeding, and complications of inflammation including fibrosis.[1] The aetiology of IBD is multifactorial and thought to be related to genetic, immunologic, environmental, and microbial factors.[2,3] With the global incidence of IBD increasing,[4,5] a diagnosis based on clinical symptoms, imaging, histopathology, and endoscopic findings may be challenging, and distinguishing UC from CD, especially in paediatric patients, may be difficult. Therefore, there is an unmet need for more accurate and reliable biomarkers that can aid in the non-invasive differential diagnosis of IBD and its subtypes and provide insights into its pathogenesis, prognosis, and management.

Noninvasive biomarkers such as serum C-reactive protein (CRP), Erythrocyte sedimentation rate (ESR) and faecal calprotectin are important in the diagnostic evaluation of IBD,[6–8] but they are not specific for IBD and can be elevated in infection, ischaemia and malignancy.[9] In patients with an established diagnosis of IBD, these biomarkers help in monitoring disease activity and response to treatment, but not in distinguishing CD from UC or predicting outcome.[10] Proteomic analysis plays a potentially important role in precision medicine for identifying diagnostic biomarkers characterising IBD.[11]

SomaScan is a state-of-the-art proteomics platform that measures the levels of many proteins simultaneously across the broad dynamic range of proteins in blood for biomarker discovery that has been applied to various diseases.[12–14] Few studies have evaluated SomaScan in patients with IBD and have primarily focused on defining changes in protein profiles in response to treatment. Using SomaScan, one study showed certain proteins such as Insulin-like growth factor binding protein 1 and 2 (IGFBP1 and IGFBP2), α-1 antitrypsin (SERPINA1), resistin (RETN), and C–C motif chemokine 23 (CCL23) were significantly decreased after treatment with infliximab.[15] Another study demonstrated a significant decrease in circulating 17-hydroxyprogesterone, corticosterone, and 11-deoxycortisol in glucocorticoid-treated patients with IBD and noted a significant alteration in levels of the proinflammatory protein, C–C Motif Chemokine Ligand 22 (CCL22), and insulin in post-treatment patients with IBD.[16]

In this study, we aimed to assess the use of aptamer-based SomaScan in a discovery cohort to identify serum protein biomarkers that distinguish paediatric patients with IBD from non-IBD and differentiate between clinical subtypes. Two larger cohorts, using ELISA and the immunoassay-based Omni 1000 platform, validated predictor models to achieve excellent discrimination

between paediatric IBD and non-IBD as well as UC and CD.

Methods

Population cohort

Paediatric patients with and without IBD (non-IBD controls) seen in the Mass General Brigham for Children (MGBfC) paediatric GI program were enrolled in a cohort study. Patients who had a clear diagnosis of IBD based on standard clinical-pathological criteria were included. Patients were selected consecutively. The average years of follow-up for the ELISA cohort is 16.7 years, with only four patients with UC (4.9% or 4/82) later diagnosed with CD. Among the SomaScan cohort, the average follow-up is 13.7 years, with only one patient with UC (9.1%, 1/11) later being diagnosed with CD. No patients initially diagnosed with CD had their diagnosis changed to UC. Serum samples from MGBfC, Children's Hospital of Philadelphia, Riley Children's Hospital, Columbia University Irving Medical Center, University of California, San Francisco, Medical Center, and Upstate Golisano Children's Hospital were analysed in an independent dataset by the Nomic Omni 1000 and 11k SomaScan v5.0 proteomic platforms. Serum samples for the biorepository cohort at MGBfC were obtained from paediatric patients undergoing clinically indicated gastrointestinal endoscopic procedures. Patients with a confirmed diagnosis of IBD were classified as UC, CD, or IBD-U based on clinical, radiographic, endoscopic, and pathological criteria.[17] Sex was included as a biological variable in the study design and recorded as assigned at birth. Sex was obtained from the medical record, Epic, which is self-reported by patients $\geq$18 years of age or by a parent/guardian in patients <18 years of age. There were no exclusions based on sex. Blood samples were collected in both IBD and non-IBD subjects. Serum samples were aliquoted and stored at −80 °C until assayed.

Ethics

The Institutional Review Board at MGBfC approved the paediatric GI biorepository and patients (18 years and older), parents/guardians (<18 years of age) signed informed consent/assent (IRB Protocol No: 2009P001287 and IRB Protocol No: 2019P002322). For subjects in all of the cohorts and at all of the sites with samples in the Omni 1000, SomaScan and ELISA cohorts, informed consent for children <7 years of age was given by at least one parent. Subjects aged 7–13 years signed an assent form, and at least one parent signed an informed consent form. Subjects 14–17 years of age signed the informed consent form along with at least one parent. Subjects 18 years of age and older signed the informed consent form. Subjects in the Omni 1000 dataset were enrolled in a prospective biomarker study for which IRB approval and informed consent were obtained at The Children's Hospital of Philadelphia (IRB Protocol No: 19-016969), Riley Children's Hospital (IRB Protocol No: 1911802085), Columbia University Irving Medical Center (IRB Protocol No: AAAT3906), University of California, San Francisco, Medical Center (IRB Protocol No: 19-28156), or Upstate Golisano Children's Hospital (IRB Protocol No: 2020-E). The study conformed to the Declaration of Helsinki.

SomaScan and Omni 1000 Assays

1.3k SomaScan v3.2 Assay

Serum specimens obtained from 12 paediatric patients with CD, 11 paediatric patients with UC, and 24 non-IBD controls were analysed by SomaScan® (SomaLogic, Inc., Boulder, CO) at the BIDMC Genomics, Proteomics, Bioinformatics and Systems Biology Center using the SomaScan Assay Kit for human serum 1.3k (cat. #900-00012) that measures expression of 1305 biologically relevant proteins according to the manufacturer's standard protocol (SomaLogic; Boulder, CO) that has been described elsewhere.[18] Five pooled human serum controls and one no-protein buffer control were run in parallel with the serum test samples for each set of 24 test samples. Sample-to-sample variability was further controlled by several hybridisation spike-in controls. Hybridisation normalisation, plate scaling, median normalisation, and calibration of the SomaScan data, performed according to the standard quality control (QC) protocols at SomaLogic, demonstrated that all samples passed the established QC criteria and were fit for further analysis. Full SomaScan data are available at GEO GSE317366.

11k SomaScan v5.0 Assay

Serum specimens (55 µl) obtained from 56 paediatric patients with CD and 25 paediatric patients with UC were analysed by SomaScan proteomics analysis using 96-well plates on a Tecan Fluent 780 liquid handling robot at the BIDMC Genomics, Proteomics, Bioinformatics and Systems Biology Center using the 11k SomaScan v5.0 Assay Kit for human serum (SomaLogic, Inc., Boulder, CO) that measures expression of 10,775 clinically relevant proteins using highly selective single-stranded modified Slow Off-rate Modified DNA Aptamers (SOMAmer) according to the manufacturer's standard protocol. Five pooled serum Calibrator replicates, 3 pooled serum Quality Control replicates, and 3 buffer replicates were used to control for batch effects and to estimate accuracy, precision, and buffer background over time. Twelve hybridisation controls and 299 non-human SOMAmers were added alongside the 10,775 SOMAmers to control for readout variability for each set of 85 test samples. Sample to sample variability was further controlled by several hybridisation spike-in controls. Hybridisation normalisation, plate scaling,

median normalisation, calibration, and Adaptive Normalisation by Maximum Likelihood (ANML) of the SomaScan data, performed according to the standard quality control (QC) protocols at SomaLogic, demonstrated that all samples passed the established QC criteria and were fit for further analysis.

Nomic Omni 1000 Assay

Serum specimens (50 μl) obtained from 54 paediatric patients with CD, 25 paediatric patients with UC, and 26 non-IBD controls were analysed by the Omni 1000 (Nomic, Inc., Montreal, Canada) assay. The Omni 1000 nELISA multiplex immunoassay platform enables absolute quantification of 1000 protein simultaneously. Serum samples were frozen and shipped on dry ice to Nomic's Montreal facility and profiled using the 1000 nELISA assay targets on the automated Omni 1000 platform according to standard protocols established by Nomic. Flow cytometry data were spectrally decoded using emFRET to map each antibody barcode to its target protein. A dilution series of calibration standards of recombinant proteins was run in parallel to generate the standard curves for the 1000 proteins. Calibration standards were fit to a four-parameter logistic curve to derive pg/ml values from fluorescence units.

Statistics

Step 1: SomaScan proteomics discovery, systems biology, and development of IBD predictor models

Identification of proteomic signatures of IBD and UC vs CD. Differential expression analysis was performed on log-transformed relative fluorescence unit (RFU) values of the SomaScan data using t-tests and Benjamini-Hochberg (BH) method for multiple hypotheses testing.[19] We tested for normality using the Kolmogorov-Smirnov test and 1247, 1232, 1287, and 1302 proteins exhibited log-normal distribution in non-IBD, IBD, CD, and UC groups respectively with 1304 proteins being log-normal in either IBD or non-IBD groups. Fold change (FC) of protein expression was calculated as the ratio of average signal values (IBD/non-IBD, UC/CD) with decreases being indicated by a negative sign.[20] For example, a FC of –2 implies a 2-fold decrease in expression in IBD compared to non-IBD.

Samples and proteins were clustered using the Unweighted Pair Group Method with Arithmetic-mean (UPGMA) method.[21] Data were standardised along the individual protein expression, and average linkage clustering was used in the iterations to update the similarity matrix in the UPGMA implementation. Samples were represented using principal components analysis (PCA)[22] of protein expression with Support Vector Machines (SVM)[23] on the PCA results using Gaussian, polynomial (degree = 2, 3, 4), and linear kernels. Computational analyses were performed using MATLAB (v.2023a, The MathWorks Inc., MA). All prediction models developed followed log-normal distribution and remained significant based on the non-parametric Mann-Whitney U-test.

Model performance was evaluated using leave-one-out cross-validation (L1OXV). In each iteration, one sample was withheld as the test set, PCA and SVM model training were performed on the remaining samples, and the trained classifier was used to predict the class label of the held-out sample. Classification accuracy was calculated as the proportion of correctly classified samples across all L1OXV iterations. Receiver operating characteristic (ROC) curves were generated by aggregating prediction scores across L1OXV iterations, and the area under the ROC curve (AUC) was calculated as a threshold-independent measure of classifier performance. Confidence intervals for AUC were computed using bootstrapping.

Systems biology analysis. To acquire new insights into potential pathophysiological pathways and biological functions underlying IBD- and IBD subtype-specific serum protein signatures and to better understand the complex interactions between the differentially expressed proteins and candidate upstream regulators, we performed functional category, canonical pathway, interactive network, upstream regulator, and regulator effect analyses of dysregulated proteins (IBD vs non-IBD and UC vs CD) using the Ingenuity Pathway Analysis (IPA) software tool, version 01-22-01 (QIAGEN, Redwood City, CA).[24]

The dysregulated proteins were included in further network analysis using the STRING database version 12.0 for protein–protein functional and physical interactions, the results of which were displayed as a functional network.[25]

Interactions were considered with a STRING confidence score of 0.4 (UC vs CD) and 0.7 (IBD vs non-IBD) or higher, respectively, garnered from the "experimental" and "databases" categories. Proteins without associations to other proteins in the network were removed. A k-means clustering algorithm was performed to select connected proteins (k-means = 6 for IBD vs non-IBD and 5 for UC vs CD). Functional description of clusters was assigned based on a manually curated evaluation of enriched KEGG pathway, Gene Ontology (GO), Reactome, STRING local network clusters terms, and PubMed literature search.

Step 2: ELISA validation

ELISA validation of proteomic data. Based on their biological plausibility, we identified four proteins (Matrix metalloproteinase-1 [MMP1], Matrix metalloproteinase-3 [MMP3], resistin [RETN], haptoglobin [HP]) selected from the IBD predictor developed using SomaScan data analysis. Expression levels of MMP1 (RRID tag AB_355700, AB_356776) and RETN (RRID tag AB_3668726) were measured using the fully automated immunoassay platform, Ella (Cat # SPCKB-PS-000513 and SPCKB-PS-001017; Protein Simple/Biotechne; San

Jose, CA) and commercial immunoassays optimised for Ella, following the manufacturer's protocols. A standard curve was included on each Ella cartridge. Inter-assay coefficients of variation (CVs) of triplicate measures were generally <5%. If a CV was >10%, the assay was repeated. MMP3 (AB_355402, AB_356540) and HP (AB_3659178, AB_3644596) expression levels were measured using quantikine ELISAs (Cat # DMP300 and DHAPG0; R&D/Biotechne; San Jose, CA), following the manufacturer's protocols. Assay personnel were blinded to case and control status. Immunoassay concentrations of the biomarkers were reported as median with interquartile range and t-test was used to determine the statistical significance of the difference among cases and controls. P-values (t-test) <0.05 were considered significant. Box Whisker plots of median protein expression values were created using XLSTAT 2024.3. ELLA data was log-normal for all 4 proteins and remained significant based on the non-parametric Mann-Whitney U-test.

Step 3: Modelling of ELISA and Omni 1000 data

Similar to SomaScan data analysis, clustering and classification of ELISA and Omni 1000 data was performed using the UPGMA algorithm, SVM applied on PCA of samples and ROC curve analysis.

Stratification

Stratification was performed to evaluate the robustness of biomarker associations across clinically relevant subgroups.

Non-IBD controls were stratified into three subtypes based on the presence and anatomical location of inflammation (non-IBD without inflammation, non-IBD with upper gastrointestinal inflammation, and non-IBD with colonic inflammation). Biomarker levels were compared across control subtypes using one-way analysis of variance (ANOVA). Multinomial logistic regression was additionally used to model associations between biomarker levels and control subtype membership, with the non-inflammatory control group specified as the reference category.

Patients with IBD were stratified by disease activity using PUCAI and PCDAI scores and categorised as "in remission" or "active disease" groups. Biomarker levels were compared between active and remission groups using two-sample t-tests, and associations were evaluated using univariate binomial logistic regression.

In all stratified analyses, IBD subgroups were additionally compared separately with non-IBD controls using both descriptive (t-test) and inferential (logistic regression) approaches.

Permutation testing

Permutation testing was performed to assess the extent to which observed discovery-stage results could arise from random class assignment in the high-dimensional SomaScan dataset.

For IBD versus non-IBD and UC versus CD comparisons, sample labels were randomly permuted 1000 times while preserving group sizes. For each permuted dataset, differential protein analysis was repeated using the same criteria as in the original discovery analysis (Benjamini–Hochberg–adjusted p-value threshold and FC cutoff). The number of proteins meeting significance criteria was recorded. In addition, classification models were refit using the permuted labels, and model performance metrics (AUC and leave-one-out cross-validation accuracy) were computed.

Permutation results were compared with those obtained using the true class labels to evaluate whether observed discrimination and feature selection exceeded that expected by chance.

Power calculation

Sample size and power calculations were done using: Qiu W, Lee MT, Whitmore GA (2025). Sizepower: Sample Size and Power Calculation in Micorarray Studies. https://doi.org/10.18129/B9.bioc.sizepower, R package version 1.80.0. Using data from 56 CD and 25 UC samples from an independent 11k SomaScan v5.0 dataset that was not included in the SomaScan data presented in this manuscript as a pilot for power calculation, our analysis indicated that 11 samples in each group is sufficient to obtain 80% power controlling for a false discovery rate of 5%.

Role of funders

The funders had no role in the design, conduct, data collection, data analyses, interpretation, reporting or writing of this study.

Results

Characteristics of the SomaScan discovery and orthogonal Ella immunoassay validation cohort

Among a paediatric cohort at MGBfC of individuals undergoing colonoscopy for clinical evaluation of gastrointestinal symptoms, a total of 47 patients (11 UC (23%), 12 CD (26%), and 24 non-IBD controls (51%)) were included in the SomaScan discovery cohort, and 295 patients (82 UC (28%), 97 CD (33%), 7 IBD-U (2%), and 109 non-IBD controls (37%)) were included in the Ella immunoassay validation cohort. As depicted in Table 1, in the discovery SomaScan cohort, 100% (N = 11) of UC subjects and 83% (N = 10) of CD subjects were not on treatment, compared to 55% (N = 45) of UC subjects and 57% (N = 55) of CD subjects in the ELISA validation cohort. The majority of patients with UC had pancolitis (82% (N = 9) and 60% (N = 49) in the discovery and validation cohorts, respectively), and ileocolonic disease (Paris classification L3) was present in 58% (N = 7) and 39% (N = 38) of the CD discovery and validation cohorts respectively. Further details of the patient demographics and IBD characteristics for

	Control (SomaScan)	Control (ELISA)	UC (SomaScan)	UC (ELISA)	CD (SomaScan)	CD (ELISA)	IBD-U (ELISA)
Analysis set, no	24	109	11	82	12	97	7
Sex at birth, (Female, %)	12 (50)	49 (45)	5 (45)	41 (50)	7 (58)	39 (40)	3 (43)
Sex at birth, (Male, %)	12 (50)	60 (55)	6 (55)	41 (50)	5 (42)	58 (60)	4 (57)
Age, years (Mean ± SD)	11.93 ± 2.57	13.79 ± 5.70	15.25 ± 2.80	16.13 ± 4.88	13.58 ± 3.10	15.03 ± 4.64	11.35 ± 5.20
Race (n, %)							
Asian		2 (2)		3 (4)		4 (4)	
Black or African American	1 (4.2)	1 (1)		1 (1)		3 (3)	1 (14)
Hispanic or Latino	2 (8.3)	6 (5)	2 (18)	3 (4)	1 (9)	3 (3)	
White	20 (91.6)	95 (87)	9 (82)	70 (85)	11 (91)	83 (86)	6 (86)
Unknown	1 (4.2)	5 (5)		5 (6)		4 (4)	
Disease location – no (%)							
L1					1 (8)	24 (24.7)	
L2					2 (17)	17 (17.5)	
L3					7 (58)	38 (39.17)	
L4a					2 (17)	17 (17.5)	
L4b					–	1 (1)	
Disease phenotype – no (%)							
B1					11 (92)	76 (78.4)	
B2					–	9 (9.3)	
B3					–	4 (4.1)	
B2B3					1 (8)	8 (8.2)	
Disease location – no (%)							
E1				7 (8.5)			2 (28.6)
E2			1 (9)	13 (15.85)			
E3			1 (9)	13 (15.85)			
E4			9 (82)	49 (59.8)			5 (71.4)
Perianal disease – no (%)					3 (25)	29 (29.9)	
VEO-IBD[a] - no (%)			0	7 (8.5)	0	6 (6.2)	1 (14.3)
Treatment- no (%)							
On Treatment			0	36 (43.9)	2 (17)	42 (43.3)	3 (42.9)
5-ASA[b]				34 (41.5)	2 (17)	41 (42.3)	2 (28.6)
Corticosteroid				6 (7.3)	1 (8)	5 (5.2)	0
Topical agents[c]				5 (6.1)		4 (4.1)	1 (14.3)
Biologics[d]				0		0	0
Not on treatment			11 (100)	45 (54.9)	10 (83)	55 (56.7)	4 (57.1)
Activity Index- Median, IQR							
PUCAI/PCDAI			55.0 (37.5–72.5)	27.5 (0.0–45.0)	40.0 (20.0–50.0)	25.0 (10.0–45.0)	–
Inflammatory marker							
ESR (mean ± SD, mm/hr)			22.4 ± 20.3	20.9 ± 19.2	35.7 ± 18.7	29.4 ± 23.6	22.8 ± 16.4
CRP (mean ± SD, mg/L)			13.8 ± 16.2	10.0 ± 22.8	42.4 ± 45.8	29.6 ± 40.4	5.4 ± 5.0
Inflammation[e] - no (%)	14 (58.3)	66 (60.6)	11 (100)	50 (61.0)	12 (100)	67 (69.1)	–

E1-ulcerative proctitis. E2 Left-sided UC (distal UC). E3: Extensive (hepatic flexure distally). E4: pancolitis (proximal to hepatic flexure). B1 non-stricturing. B2 stricturing. B3 penetrating. B2B3: both penetrating and stricturing disease, either at the same or different times. L1 ileal. L2 colonic. L3 ileocolonic. L4a upper disease proximal to the ligament of Treitz. L4b upper disease distal to the ligament of Treitz and proximal to distal 1/3 ileum. [a]VEO-IBD: Very Early Onset IBD, defined as disease in children less than 6 years of age. [b]5-ASA: 5-aminosalicylic acid. [c]Topical agents include topical 5-ASA and topical corticosteroid. [d]Biologics include anti-tumour necrosis factor, Integrin blockers, and Interleukin blockers. [e]In control group, evidence of inflammation is defined by positive pathologic report of biopsy; in the IBD group, active inflammation is defined by those with activity index (PCDAI or PUCAI) >10.

Table 1: **Demographics and clinical characteristics of the study cohort.**

the discovery set and validation set including percent of patients with very early onset (VEO)-IBD (onset <6 years of age), prevalence of perianal disease, disease activity, inflammatory markers (CRP and ESR), and medications at the time of sample collection are described in Table 1. Non-IBD controls included paediatric patients undergoing endoscopy for gastrointestinal symptoms but without confirmed IBD. GI diagnoses with intestinal inflammation in this group were obtained via chart review (Supplemental Table S1) and included eosinophilic oesophagitis, gastritis, celiac disease, collagenous gastritis, infectious enterocolitis,

lymphoid hyperplasia, autism, and microscopic colitis. A subset of non-IBD controls (61% (N = 66) in the ELISA group and 58% (N = 14) in the SomaScan group) had evidence of inflammation.

Proteomics analysis of circulating serum protein biomarkers associated with paediatric IBD

Fig. 1 schematically outlines the overall study and results. To identify differentially expressed proteins that discriminate between IBD and non-IBD, t-test results used to analyse SomaScan data for 1305 proteins were visualised by a Volcano plot (Fig. 2A). Supplemental Table S2 lists raw data for the entire data set. 98% of all 1305 proteins measured on the SomaScan platform had inter-assay coefficients of variations less than 10% (Supplemental Table S3). Proteins with Benjamini-Hochberg (BH) adjusted p value < 0.01 (t-test) were considered significant, resulting in 95 serum proteins that differentiated IBD from non-IBD (Supplemental Table S2). Twenty-eight proteins were increased and 67 were decreased in IBD compared with non-IBD controls. Table 2 depicts the 32 up- and down-regulated proteins with a fold change of at least 1.5. Among proteins with the highest fold increases in IBD were multiple proteases, Matrix metalloproteinase 1 (MMP1), Matrix metalloproteinase 3 (MMP3), Matrix metalloproteinase 8 (MMP8), and Pappalysin-1 (PAPPA), all of which are involved in extracellular matrix degradation; whereas, the most downregulated protein was haemoglobin, consistent with the anaemia

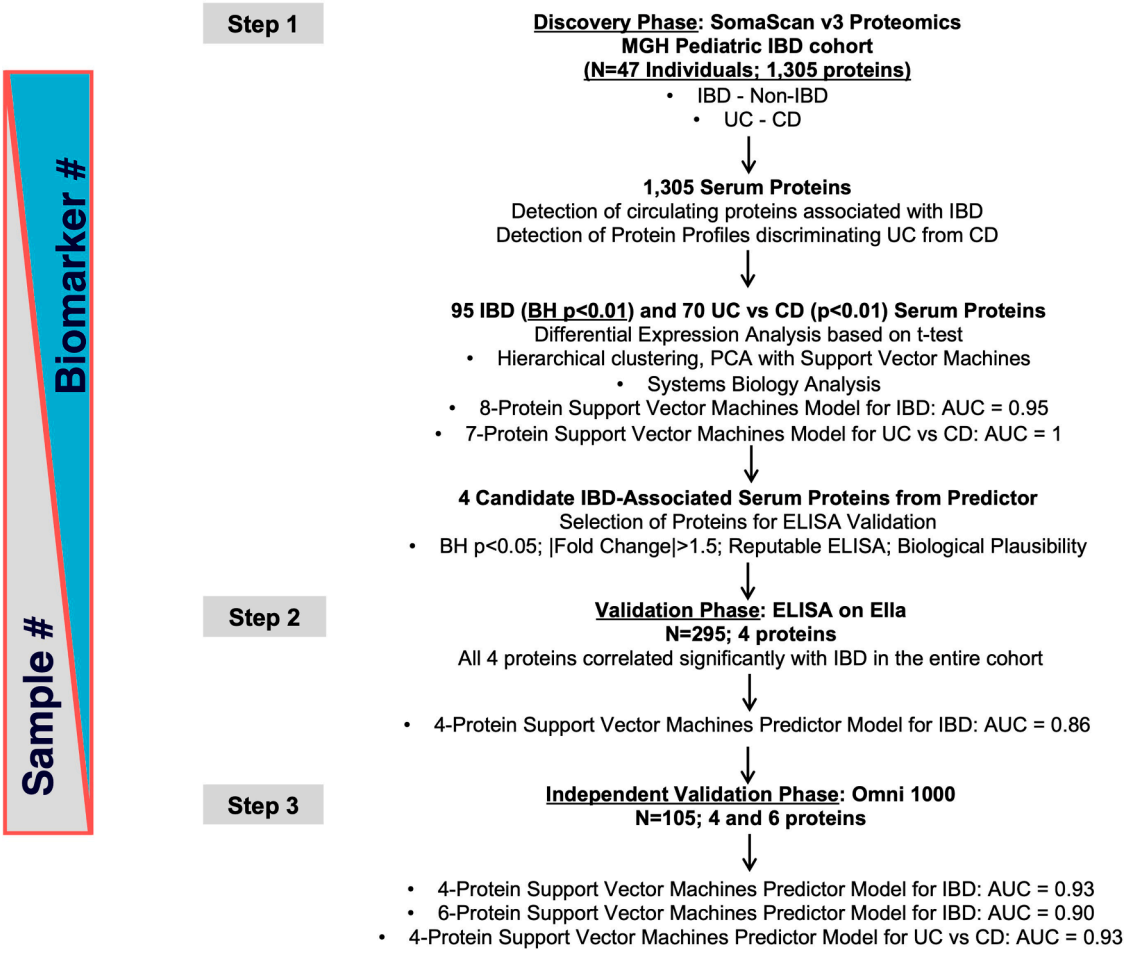

***Fig. 1:* Schematic overview of study design and methodologic steps for identification of IBD serum protein biomarkers and ELISA validation.** Step 1: Discovery Phase: SomaScan v3 proteomics to identify IBD- and UC-specific serum proteins. Statistical analysis to identify the biomarker proteins. Both IBD- and UC-specific protein sets were used for System Biology Analysis to identify their biological implications using IPA and STRING. IBD and UC predictor model was generated. Four candidate IBD biomarkers from predictor model were selected for ELISA validation. Step 2: Validation Phase: ELISA on Ella. Four candidate IBD biomarkers were analytically validated using multiplex immunoassays on the Ella platform in 295 samples. Step 3. Independent Validation Phase: Omni 1000. Immunoassay-based IBD and UC vs CD predictor models confirmed association in 105 serum samples.

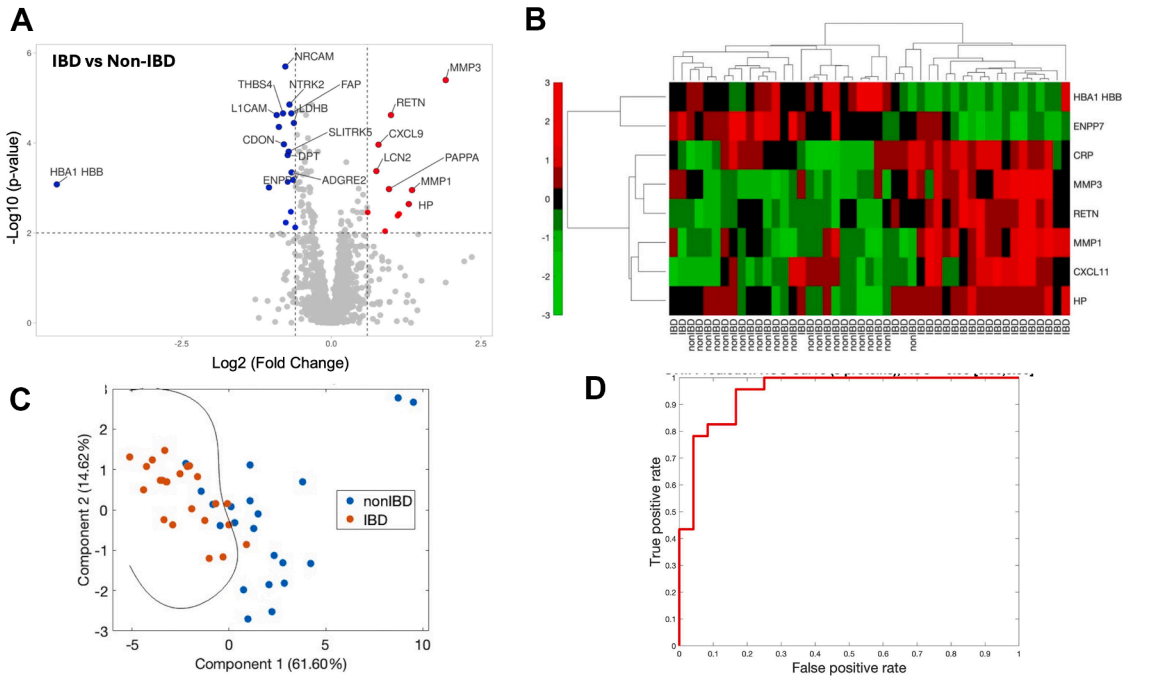

***Fig. 2:* Discovery proteomics in paediatric subjects identifies serum proteins associated with IBD and a multi-protein IBD predictor.** (A) Association of serum proteins with IBD for SomaScan Data. Volcano plot displaying log2 fold change (FC) (x axis) against -log10 p-value (t-test) differences (y axis) for all proteins measured on SomaScan. Red dots show significantly upregulated proteins and blue dots show significantly downregulated proteins in IBD vs non-IBD above a threshold of BH $p < 0.01$ and absolute FC > 1.5. Grey dots indicate proteins not significantly changed in IBD. Discriminatory analysis using top 8 IBD vs. non-IBD proteins with BH $p < 0.01$ (t-test) and |FC|>2. (B) Hierarchical clustering of proteins and samples (N = 47) where red and green represent up and down regulation in IBD compared to non-IBD, respectively. Colour bar represents range of row-normalised signal values and associated colour map. (C) SVM classification with Gaussian kernel applied on 2D PCA of the 47 samples based on the 8-protein expression profile. Axes labels indicate percent variance explained by principal components. L1OXV accuracy was 85.11%, L1OXV: Leave-one-out cross-validation. Blue circles, non-IBD; Red circles, IBD. (D) AUC of the ROC for the classifier model shown in (C). Values in brackets show 95% confidence interval for the AUC. The AUC was 0.95 (0.86, 0.99).

commonly noted in patients with IBD. Additional top increased proteins in IBD were resistin (RETN) and haptoglobin (HP), and proteins strongly associated with inflammation and other immune system functions, C-reactive protein (CRP), lipocalin 2 (LCN2), C-X-C motif chemokine 11 (CXCL11), and C-X-C motif chemokine 9 (CXCL9) (Table 2 and Supplemental Table S2). Other decreased IBD-associated proteins were primarily linked to adhesion such as L1 cell adhesion molecule (L1CAM), cell adhesion molecule-related/down-regulated by oncogenes (CDON), neuronal cell adhesion molecule (NRCAM), aggrecan (ACAN), brevican (BCAN), adhesion G protein-coupled receptor E2 (ADGRE2), neural cell adhesion molecule 1 (NCAM1), periostin (POSTN), and lumican (LUM) (Table 2 and Supplemental Table S2). Hierarchical clustering of the top 8 differentially expressed proteins with BH $p < 0.01$ (t-test) and FC > 2 (Haemoglobin subunit alpha 1/Beta (HBA1 HBB), Ectonucleotide pyrophosphatase/phosphodiesterase 7 (ENPP7), C-reactive protein (CRP), Matrix metallopeptidase 3 (MMP3), Resistin (RETN), Matrix metallopeptidase 1 (MMP1), C-X-C motif chemokine ligand 11 (CXCL11), Haptoglobin (HP)), discriminated well between IBD cases and non-IBD controls as shown in Fig. 2B, with three IBD samples clustering with non-IBD and one non-IBD sample clustering with IBD. Good discrimination between IBD and non-IBD was also observed when performing SVM classification with a Gaussian Kernel applied on the PCA of samples based on the 8-protein profile. A high predictive leave-one-out cross-validation (L1OXV) accuracy of 85% (Fig. 2C) was attained with an area under the curve (AUC) of the receiver operating characteristic (ROC) of 0.95 [95% CI: 0.86, 0.99] (Fig. 2D). An additional multivariate model with direct interpretability and explicit handling of correlated predictors was provided using multivariable logistic regression with elastic-net regularisation. Model tuning was performed using cross-validation, and performance is reported using AUC. Multivariable logistic regression using the minimum-deviance-coefficients (Supplemental Figure S1A and Supplemental Results) resulted in an AUC [95% CI] of 0.99 [0.9266, 1.0000] (Supplemental Figure S1B).

SomaId	Target full name	Target	UniProt	Entrez gene ID	Entrez gene symbol	FC	p-val	BH p-val
SL000524	**Stromelysin-1**	**MMP-3**	**P08254**	**4314**	**MMP3**	**3.76**	**0**	**0.000004**
SL000521	**Interstitial collagenase**	**MMP-1**	**P03956**	**4312**	**MMP1**	**2.55**	**0.000043**	**0.001114**
SL000437	**Haptoglobin**	**Haptoglobin, Mixed Type**	**P00738**	**3240**	**HP**	**2.45**	**0.000098**	**0.002277**
SL003326	C-X-C motif chemokine 11	I-TAC	O14625	6373	CXCL11	2.19	0.000214	0.00378
SL000051	C-reactive protein	CRP	P02741	1401	CRP	2.15	0.000241	0.004133
SL004260	**Resistin**	**resistin**	**Q9HD89**	**56729**	**RETN**	**2.00**	**0**	**0.000024**
SL002755	Pappalysin-1	PAPP-A	Q13219	5069	PAPPA	1.95	0.000038	0.001058
SL000526	Neutrophil collagenase	MMP-8	P22894	4317	MMP8	1.86	0.000649	0.009121
SL003188	C-X-C motif chemokine 9	MIG	Q07325	4283	CXCL9	1.73	0.000001	0.000109
SL000695	Neutrophil gelatinase-associated lipocalin	Lipocalin 2	P80188	3934	LCN2	1.68	0.00001	0.000422
SL000542	Integrin alpha-IIb: beta-3 complex	gpIIbIIIa	P08514 P05106	3674 3690	ITGA2B ITGB3	1.53	0.000187	0.003482
SL003341	Fibrinogen gamma chain	Fibrinogen g-chain dimer	P02679	2266	FGG	1.51	0.000125	0.002704
SL003764	Neural cell adhesion molecule 1, 120 kDa isoform	NCAM-120	P13591	4684	NCAM1	−1.51	0.000013	0.000485
SL005084	Periostin	Periostin	Q15063	10631	POSTN	−1.51	0.00003	0.000943
SL007471	Collectin-12	COLEC12	Q5KU26	81035	COLEC12	−1.52	0.00052	0.007519
SL000493	L-lactate dehydrogenase B chain	LDH-H 1	P07195	3945	LDHB	−1.54	0	0.000036
SL006230	Lumican	Lumican	P51884	4060	LUM	−1.56	0.000019	0.000665
SL008822	Adhesion G protein-coupled receptor E2	EMR2	Q9UHX3	30817	ADGRE2	−1.58	0.000011	0.000451
SL006528	Prolyl endopeptidase FAP	SEPR	Q12884	2191	FAP	−1.59	0	0.000022
SL009089	Brevican core protein	PGCB	Q96GW7	63827	BCAN	−1.60	0.000174	0.003377
SL004160	BDNF/NT-3 growth factors receptor	TrkB	Q16620	4915	NTRK2	−1.62	0	0.000014
SL014070	SLIT and NTRK-like protein 5	SLIK5	O94991	26050	SLITRK5	−1.64	0.000003	0.000156
SL008178	Dermatopontin	DERM	Q07507	1805	DPT	−1.66	0.000003	0.000186
SL001947	Melanoma-derived growth regulatory protein	MIA	Q16674	8190	MIA	−1.66	0.000021	0.000733
SL004661	Aggrecan core protein	Aggrecan	P16112	176	ACAN	−1.69	0.000387	0.005876
SL005210	Neuronal cell adhesion molecule	Nr-CAM	Q92823	4897	NRCAM	−1.70	0	0.000002
SL014092	Cell adhesion molecule-related/down-regulated by oncogenes	CDON	Q4KMG0	50937	CDON	−1.73	0.000001	0.000107
SL007207	Thrombospondin-4	TSP4	P35443	7060	THBS4	−1.75	0	0.000022
SL008416	C-type mannose receptor 2	MRC2	Q9UBG0	9902	MRC2	−1.84	0	0.000044
SL004154	Neural cell adhesion molecule L1	NCAM-L1	P32004	3897	L1CAM	−1.88	0	0.000024
SL009045	Ectonucleotide pyrophosphatase/phosphodiesterase family member 7	ENPP7	Q6UWV6	339221	ENPP7	−2.06	0.000032	0.000975
SL000836	Haemoglobin	Haemoglobin	P69905, P68871	3039 3043	HBA1 HBB	−24.04	0.000026	0.000834

The 4 proteins seleted for ELISA validation are bolded.

Table 2: **Top up- and down-regulated proteins between IBD cases and non-IBD controls.**

Functional annotation and pathway analysis of IBD serum biomarkers

To better understand pathophysiological mechanisms and complex protein interactions that may be altered in individuals with paediatric-onset IBD and involved in IBD pathogenesis, Ingenuity Pathway Analysis (IPA) and STRING analysis were performed using the 95 IBD-associated proteins as input. IPA biological function analysis revealed that the highest enriched pathways were inflammatory response (inflammatory response, inflammation of joint, inflammation of absolute anatomical region, chronic inflammatory disorder), as well as migration, activation, and quantity of immune cells, especially of the myeloid lineage (leucocyte migration, cell movement of phagocytes, cell movement of myeloid cells, systemic autoimmune syndrome, cell movement of granulocytes, cell movement of neutrophils, immune response of leukocytes, accumulation of myeloid cells, T cell response, activation of myeloid cells), and vascular function (development of vasculature, angiogenesis, vasculogenesis, migration of vascular cells) (Fig. 3A). Inflammatory response was linked to 27 out of 95 proteins, which was supported by top associated proteins CRP, MMP1, TNFRSF1B, LCN2, and IL17A with known roles as pro-inflammatory factors in IBD (Fig. 3B). Predicted enhanced angiogenesis was associated with 35 out of 95 proteins, as supported by proteins such as metalloproteinases MMP1, MMP2, MMP3, and MMP8, as well as LCN2, TIMP2, and NOTCH3 (Fig. 3C). These results not only highlight the involvement of immune pathways in IBD, but also indicate altered vascular

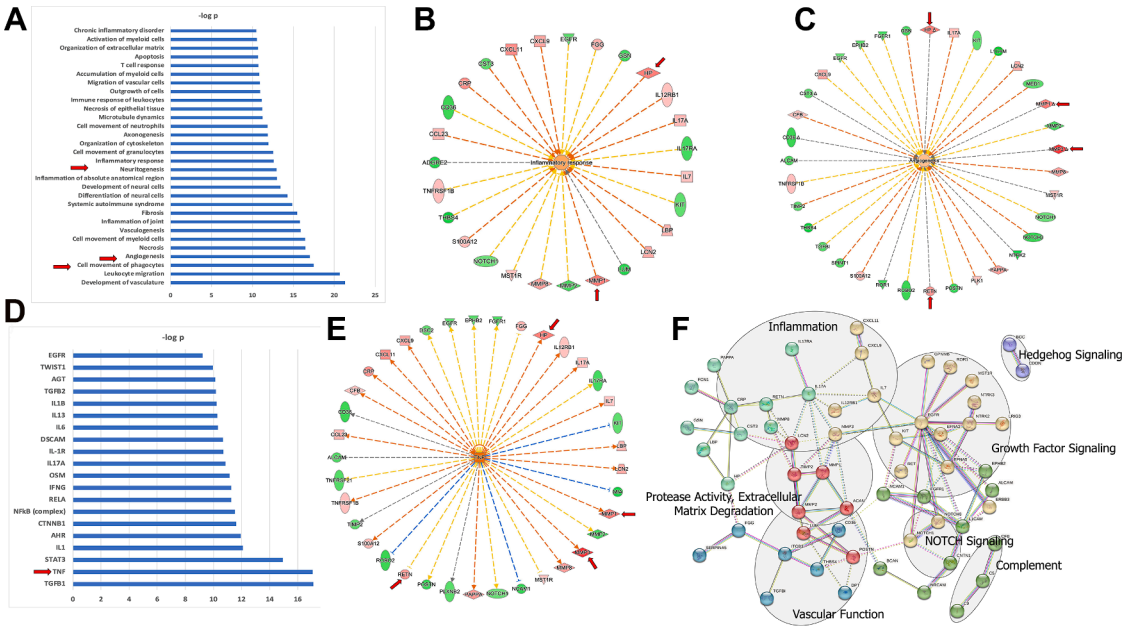

***Fig. 3:* Systems biology analysis of 95 SomaScan IBD proteins (BH p < 0.01[t-test]) identifies plausible pathophysiological pathways linked to IBD pathophysiology.** (A) IBD Biological functions that are significantly enriched (i.e., statistically relative high number of proteins dysregulated in IBD cases versus non-IBD) by 95-input protein list. X-axis indicates -log p-values (Fisher's exact test). Level of Significance: $p < 10^{-5}$ (Fisher's exact test). (B and C) Two significant classes of biological functions, Inflammatory Response (B) and Angiogenesis (C), enriched among the IBD-associated 95 proteins and proteins linked to these biological functions. IPA colour and symbol guide: blue = inhibition; orange = activation; red = increased; green = decreased; yellow = contrary to published evidence; grey = unknown; dashed line = indirect; arrowhead (pointed) = activating; arrowhead (blunt) = inhibitory. (D) IBD Upstream regulators with highest statistical significance that best explain observed expression changes in input 95 protein list as their targets. X-axis indicates -log p-values (Fisher's exact test). Level of Significance: $p < 10^{-5}$ (Fisher's exact test). (E) IBD Downstream targets of TNF from among the 95-protein list. Proteins that are validated by immunoassays are indicated by red arrow. The colours are as in part C. Red colour of upstream regulator and biological function shown in center implies "activation," and colour shade implies confidence in activation with darker shades implying more confidence in prediction. (F) STRING analysis and visualisation of interaction clusters (k-means = 6 clusters indicated by node colour) formed by the 95 significantly altered IBD proteins and labelled on related functional categories. Solid line represents within-cluster, dashed grey line represents between-cluster interactions. Line thickness indicates strength of data support.

functions and angiogenesis may play an important role in IBD pathogenesis.

To contextualise our findings, we used IPA's Upstream Regulator analysis to identify factors likely influencing expression of IBD-associated proteins. This helps uncover potential regulatory mechanisms driving these protein changes and sheds light on pathways relevant to IBD development. Among the most statistically significant upstream regulators identified to be associated with the IBD serum proteins, pro-inflammatory cytokines (TNF, IL1, IL6, IL17A) and their downstream signalling transcription factors (NFKB, STAT3) were predicted to have enhanced activity with positive activation z-scores, while TGFB1 and CTNNB1 activity may be reduced in IBD (Fig. 3D and E and Supplemental Figure S2). Key transcription factors (NFKB, STAT3) were predicted to regulate expression of many differentially expressed proteins, linking them to the inflammatory and hypoxic responses characteristic of IBD. The IBD-associated proteins downstream of TNF are highlighted in Fig. 3E.

Using STRING database for network analysis of functional and physical protein associations curated across major data repositories is a powerful method to explore interactions among the 95 dysregulated IBD-associated proteins by uncovering interconnected relationships and potential pathways underlying the disease. Fifty-seven of the 95 proteins assembled into distinct interacting protein networks, forming 6 major clusters that pointed toward enrichment of pathways associated with complement, inflammation, vascular function, NOTCH signalling, Hedgehog signalling, growth factor signalling, protease activity and extracellular matrix degradation with IBD (Fig. 3F). EGFR and MMP3, occupying key focus hubs, connected several biological functions (Fig. 3F).

Profiling the serum proteome to identify biomarkers discriminating between UC and CD

SomaScan identified 70 serum proteins that significantly (p < 0.01 [t-test]) differentiated between patients with UC and CD (Supplemental Table S4), with 7

proteins having a BH p < 0.05 (t-test), as depicted in the Volcano plot (Fig. 4A). Twenty proteins were increased and 50 decreased in UC as compared to CD (Supplemental Table S4). Table 3 shows the top 25 of these 70 proteins increased or decreased in UC vs CD. Metalloproteinase MMP12 had the lowest p-value (t-test) and the highest fold increase in UC compared to CD. Indeed, metalloproteinases were prominent among the differentiating proteins as MMP10 was also increased in UC while MMP8 was decreased. The complement factor C3b as well as several additional members of the complement pathway (C1S, C1QBP, C9) were increased in CD vs UC, supporting an enhanced complement activation in CD compared to UC. Additional biomarkers with enhanced expression in UC were various proteins linked to immune system functions, including myeloid and neutrophil proteins, CCL22 and CXCL13 as well as the cytokine IL19, whereas some pro-inflammatory proteins such as CRP and S100A12 were elevated in CD (Table 3 and Supplemental Table S4).

Hierarchical clustering of the top 7 proteins with a BH p < 0.05 (t-test) (legumain (LGMN), cadherin 15 (CDH15), protein disulfide isomerase (PDIA3), kallikrein-related peptidase 14 (KLK14), protein phosphatase 3 regulatory subunit B (PPP3R1), CCL22, matrix metallopeptidase 12 (MMP12)) that were differentially expressed between UC and CD cases (Fig. 4B) accurately separated UC from CD cases with only 1 CD case clustering with UC. A 7-protein SVM model using PCA demonstrated high accuracy in predicting UC vs CD, achieving an L1OXV accuracy of 95.65%, indicating high diagnostic performance (Fig. 4C). Receiver operating characteristic (ROC) curve

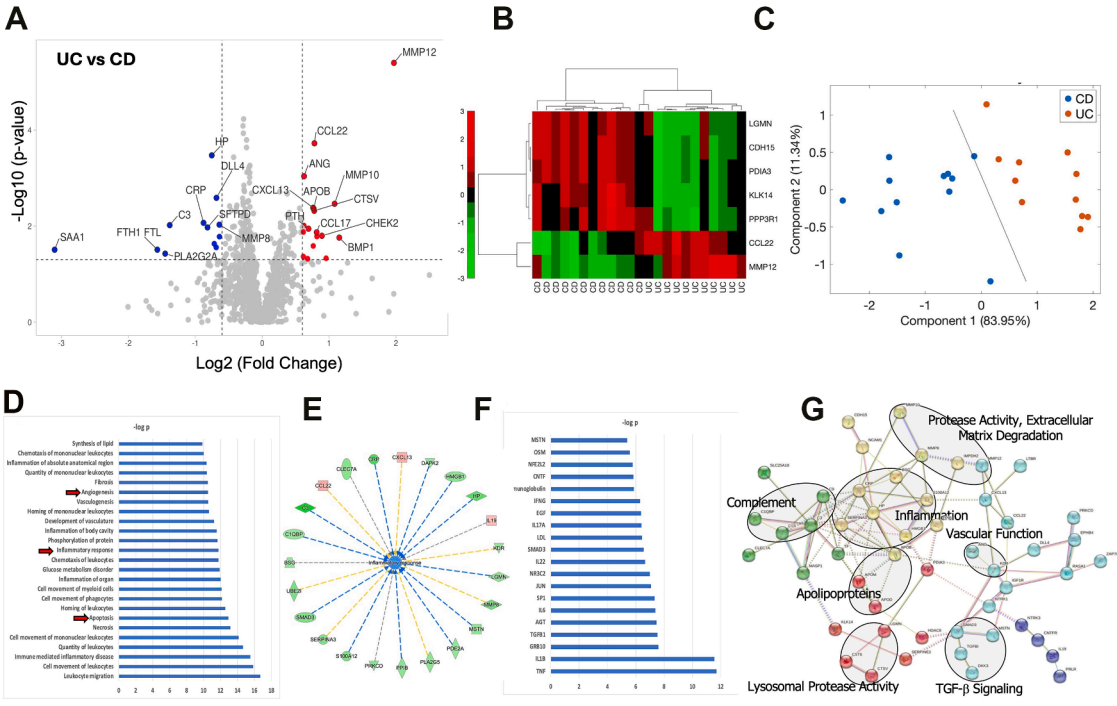

***Fig. 4:* Discovery proteomics in paediatric subjects identifies serum proteins discriminating between UC vs CD and a multi-protein UC vs CD predictor.** (A) Volcano plot displaying log2 fold change (FC) (x axis) against -log10 p-value (t-test) differences (y axis) for all proteins measured on SomaScan. Red dots show significantly upregulated proteins and blue dots show significantly downregulated proteins in UC vs CD above threshold of $p < 0.05$ and absolute FC > 1.5. Grey dots indicate proteins not significantly changed in UC vs CD. Discriminatory analysis using top 7 UC vs. CD proteins with BH $p < 0.05$. (B) Hierarchical clustering of proteins and 23 samples where red and green represent up and down regulation in UC compared to CD, respectively. (C) SVM classification of 23 samples with linear kernel applied on PCA based on the 7-protein expression profile. L1OXV accuracy was 95.65%. (D) Systems biology analysis of 70 SomaScan UC vs CD proteins ($p < 0.01$ [t-test]). UC vs CD biological functions that are significantly enriched (i.e., statistically relative high number of proteins dysregulated in UC cases versus CD controls) by 70-input protein list. X-axis indicates -log p-values (Fisher's exact test). Level of Significance: $p < 10^{-5}$ (Fisher's exact test). (E) Two significant classes of biological functions enriched among UC vs CD-associated 70 proteins and proteins linked to these biological functions. (F) UC vs CD Upstream regulators with highest statistical significance that best explain observed expression changes in input 70 protein list as their targets. X-axis indicates -log p-values (Fisher's exact test). Level of Significance: $p < 10^{-5}$ (Fisher's exact test). (G) STRING analysis of 70 significantly altered UC vs CD proteins and visualisation of interaction clusters (k-means = 5 clusters indicated by node colour) formed by 70 UC vs CD plasma proteins and labelled on related functional categories.

SomaId	Target full name	Target	UniProt	Entrez gene ID	Entrez gene symbol	FC	p-val	BH p-val
SL000522	Macrophage metalloelastase	MMP-12	P39900	4321	MMP12	3.92	0.000004	0.005325
SL000645	Stromelysin-2	MMP-10	P09238	4319	MMP10	2.12	0.003431	0.143062
SL003187	C–C motif chemokine 22	MDC	O00626	6367	CCL22	1.72	0.000189	0.041599
SL006910	Cathepsin L2	Cathepsin V	O60911	1515	CTSV	1.72	0.004801	0.145102
SL000020	Apolipoprotein B	Apo B	P04114	338	APOB	1.71	0.004346	0.143062
SL003167	C-X-C motif chemokine 13	BLC	O43927	10563	CXCL13	1.69	0.004149	0.143062
SL000003	Angiogenin	Angiogenin	P03950	283	ANG	1.54	0.000915	0.074995
SL004438	Cystatin-M	Cystatin M	Q15828	1474	CST6	1.54	0.009758	0.176164
SL003764	Neural cell adhesion molecule 1, 120 kDa isoform	NCAM-120	P13591	4684	NCAM1	1.51	0.007247	0.171078
SL010375	Neutral ceramidase	ASAH2	Q9NR71	56624	ASAH2	1.47	0.004259	0.143062
SL005153	Ciliary neurotrophic factor receptor subunit alpha	CNTFR alpha	P26992	1271	CNTFR	1.36	0.009271	0.176164
SL004354	Interleukin-19	IL-19	Q9UHD0	29949	IL19	1.35	0.006852	0.164704
SL004747	Apolipoprotein M	ApoM	O95445	55937	APOM	1.35	0.009441	0.176164
SL003304	Insulin-like growth factor 1 receptor	IGF-I sR	P08069	3480	IGF1R	1.33	0.006848	0.164704
SL012561	Regenerating islet-derived protein 4	REG4	Q9BYZ8	83998	REG4	1.32	0.00523	0.150242
SL018548	Alpha-1-antichymotrypsin complex	alpha-1-antichymotrypsin complex	P01011	NaN	SERPINA3	−1.31	0.00412	0.143062
SL000325	Complement component C9	C9	P02748	735	C9	−1.31	0.004939	0.145102
SL004097	Mothers against decapentaplegic homologue 3	SMAD3	P84022	4088	SMAD3	−1.34	0.000368	0.053991
SL004306	SUMO-conjugating enzyme UBC9	UBC9	P63279	7329	UBE2I	−1.38	0.004715	0.145102
SL004783	Protein S100-A12	S100A12	P80511	6283	S100A12	−1.40	0.003844	0.143062
SL000526	Neutrophil collagenase	MMP-8	P22894	4317	MMP8	−1.56	0.009362	0.176164
SL010457	Delta-like protein 4	DLL4	Q9NR61	54567	DLL4	−1.61	0.002581	0.128894
SL000437	Haptoglobin	Haptoglobin, Mixed Type	P00738	3240	HP	−1.69	0.000336	0.053991
SL000051	C-reactive protein	CRP	P02741	1401	CRP	−1.84	0.008559	0.176164
SL000314	Complement C3b	C3b	P01024	718	C3	−2.61	0.009571	0.176164

Table 3: **Top up- and down-regulated proteins between UC and CD.**

of the 7-protein model showed excellent diagnostic performance with an area under of the curve (AUC) of 1.00 [95% CI: 1.00, 1.00] (Supplemental Figure S3). Multivariable logistic regression using the minimum-deviance-coefficients (Supplemental Figure S1A and Supplemental Results) resulted in an AUC [95% CI] of 1.00 [1.00, 1.00] (Supplemental Figure S1C). Adjusted for age each of 7 UC vs. CD proteins and each of 8 IBD vs. non-IBD proteins remained significant. Permutation testing demonstrated that the IBD vs non-IBD proteomic signature is highly robust and unlikely to reflect overfitting, while UC vs CD discrimination shows evidence of biological signal but with reduced stability consistent with smaller effect sizes and sample numbers. The clear separation observed in the clustering and PCA plots and the ROC performance highlight the potential for these protein markers in differential diagnostics and personalised treatment strategies for subtypes of paediatric patients with IBD.

Functional annotation and pathway analysis of biomarkers discriminating between UC and CD

To provide more in-depth understanding into the disease pathogenesis and pathways that differentiate between patients with UC and CD, IPA and STRING analysis were employed. For IPA analysis, we identified significant enrichment in various biological pathways differentiating UC from CD (Fig. 4D and Supplemental Figure S4) using the top 70 proteins ($p < 0.01$), of which immune cell migration (leucocyte migration, cell movement of leucocytes, cell movement of mononuclear leucocytes, homing of leukocytes, cell movement of phagocytes, cell movement of myeloid cells, chemotaxis of leukocytes, homing of mononuclear leucocytes, chemotaxis of mononuclear leucocytes), inflammation (immune-mediated inflammatory disease, inflammatory response, inflammation of organ, inflammation of body cavity, inflammation of absolute anatomical region), vascular function (angiogenesis, vasculogenesis), cell death (apoptosis, necrosis), and fibrosis were most prominent and predicted to exhibit reduced activation in UC relative to CD (Fig. 4D). These results suggest that paediatric patients with CD exhibit higher levels of inflammation than patients with UC. Inflammatory processes were linked to 23 out of 70 proteins, most of them with decreased expression in UC (Fig. 4E). These differences in immune pathway regulation may signify distinct immune mechanisms driving UC compared to CD. Apoptosis, necrosis as well as angiogenesis and vasculogenesis were also less active in UC vs CD (Fig. 4D and Supplemental Figure S4), suggesting differences

in tissue damage and repair mechanisms in patients with UC and CD.

Upstream Regulator analysis confirmed that pro-inflammatory cytokines such as TNF, IL1B, IL6, IL17A, and IFNG were predicted to exhibit reduced activity in UC vs CD (Fig. 4F and Supplemental Figure S4). Supplemental Figure S4 shows that 27 of 70 proteins are downstream of TNF. Interaction analysis of these 70 dysregulated proteins using STRING demonstrated that 51 proteins have interactions with at least one other protein. These interacting proteins formed 5 major modules that were enriched in pathways related to vascular function, inflammation, TGF-β signalling, protease activity/extracellular matrix degradation, lysosomal protease activity, apoliproteins, and complement (Fig. 4G).

Orthogonal immunoassay verification of four IBD-associated predictor proteins across the paediatric MGH cohort

To confirm robustness of SomaScan measures and to further validate IBD-associated proteins, we performed validation of a selected set of four IBD-associated proteins from the 8-protein SomaScan-derived predictor model across the larger paediatric MGH cohort (N = 295) (Table 1) utilising orthogonal immunoassays on the Ella platform. The four proteins from the 8-protein IBD predictor established from SomaScan analysis (MMP1, MMP3, HP, RETN) were selected based on criteria detailed in Fig. 1 which included p-value <0.05 (t-test), absolute FC > 1.5, availability of reputable ELISA on the Ella platform, and biological plausibility.

Serum samples from the MGH paediatric IBD validation cohort of 295 subjects, including 186 IBD (82 UC, 97 CD, 7 IBD-U) and 109 non-IBD controls, underwent immunoassay analysis. The demographic and clinical characteristics of the cohort are described in Table 1. We performed Box Whisker plot analysis of serum concentrations for these 4 candidate IBD biomarker proteins, demonstrating that median levels of all proteins were increased with high significance ($p < 0.0001$ [t-test]) for all four selected proteins in IBD as compared to non-IBD (Fig. 5A), supporting the diagnostic potential of these biomarkers and suggesting more extensive tissue remodelling and inflammation in IBD as compared to non-IBD. All of the proteins on the ELLA platform had inter-assay coefficients of variations far less than the acceptable 15% CV for ELLA data (Supplemental Table S5). Adjusted for age and sex, each of the 4 proteins remained significant in the IBD vs. non-IBD comparison. When comparing UC and CD separately to non-IBD, each of these four proteins remained statistically significant (Supplemental Figure S5). MMP1 and haptoglobin were also statistically significantly increased in CD compared to UC (Supplemental Figure S5). We next performed stratified analyses to assess the robustness of biomarker associations within clinically relevant subgroups. Biomarker levels and associations with IBD status were examined separately within treatment-naïve and treated patients, medication classes (Systemic 5-ASA, Steroid, Topical 5-ASA/Steroid), disease activity states (active vs remission by PUCAI/PCDAI), and non-IBD control subtypes (non-IBD with no inflammation, non-IBD with inflammation in upper GI, non-IBD with inflammation in colon). Comparing ELISA data between the treated cohort and non-treated cohort showed no significant differences between both groups. MMP1, Haptoglobin, Resistin and MMP3 remained statistically significant in both treated and non-treated cohorts (Supplemental Table S6). To test whether any particular drug class significantly altered protein expression levels, we performed a stratified analysis for 3 classes of medication: 1. Systemic 5-ASA (N = 77); 2. Steroid (N = 11); 3. Topical 5-ASA/Steroid (N = 10). Due to the age of the serum samples none of the patients at the time had received anti-TNF agents. For patients with IBD treated with systemic 5-ASA, protein levels do not change significantly and remain significant when compared to non-IBD (Supplemental Table S7). For steroid treated patients, RETN shows significant difference due to medication in IBD but remains significant for discrimination of non-IBD (Supplemental Table S7). For topical 5-ASA/Steroid, MMP3 and HP show borderline significant changes which result in borderline significance for differentiating non-IBD (Supplemental Table S7).

When patients were stratified according to disease activity measured by PUCAI and PCDAI (InRemission: PUCAI/PCDAI<10; Active: PUCAI/PCDAI>10), RETN and HP were increased significantly in active disease compared to remission (behaving like activity-sensitive markers) but MMP1 and MMP3 did not differ between remission and active IBD (behaving like state markers) (Supplemental Table S8). Importantly, active and in remission IBD subgroups exhibited significantly higher levels for all four proteins compared with non-IBD controls (Supplemental Table S8). Active IBD showed stronger separation from non-IBD controls, as expected. These results suggest that IBD is associated with persistent biomarker alterations even in remission, while a subset of markers (RETN, HP) additionally track disease activity, supporting both disease specificity and potential utility for assessment of disease activity.

Non-IBD controls are comprised of three different subtypes (non-IBD without inflammation, non-IBD with inflammation in the upper GI (for example oesophagitis, gastritis or duodenitis), non-IBD with colonic inflammation). Differences in biomarker levels across non-IBD control subtypes were assessed using one-way ANOVA. To model the association between biomarker levels and control subtype membership, multinomial

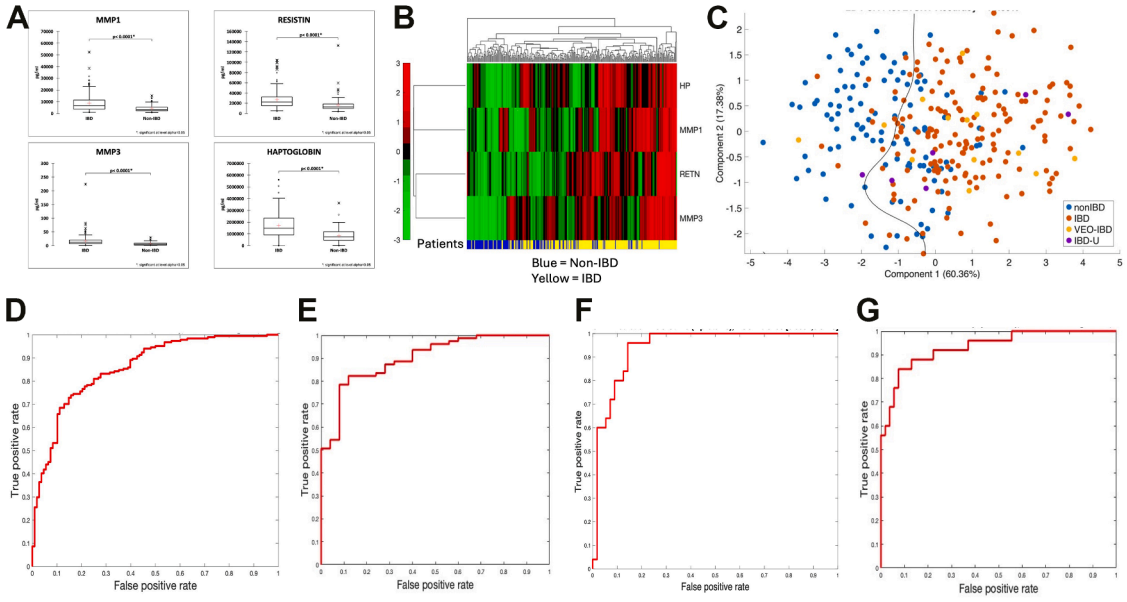

***Fig. 5*: Discriminatory analysis of IBD and non-IBD samples using ELISA.** (A) Box and Whisker Plots for IBD and non-IBD represent median expression indicated by a horizontal line and mean expression depicted as + and interquartile range of serum protein concentration (N = 295). All four proteins had significant p-value <0.001 (t-test). (B) Hierarchical clustering of proteins and 295 samples where red and green represent up and down regulation in IBD compared to non-IBD, respectively. Blue and yellow colours at the bottom of the heatmap indicate non-IBD and IBD samples, respectively. (C) SVM classification with Gaussian kernel applied on PCA of the 295 samples based on 4-protein ELISA expression profile. L1OXV accuracy was 75.69%. IBD = inflammatory bowel disease; VEO-IBD = very early onset IBD (<6yrs of age); IBD-U = IBD Unclassified. (D) AUC of ROC for 4-protein IBD classifier model shown in (C) for ELISA data (N = 295). The AUC was 0.86 (0.81, 0.90). (E) AUC of ROC for 4-protein IBD classifier model for Omni 1000 data on independent validation set (N = 105). The AUC was 0.90 (0.82, 0.95). (F) AUC of ROC for 7-protein UC vs CD classifier model for 11k SomaScan v5.0 data on independent validation set (N = 81). The AUC was 0.94 (0.87, 0.98). (G) AUC of ROC for 4-protein UC vs CD classifier model for Omni 1000 data on independent validation set (N = 79). The AUC was 0.93 (0.84, 0.97).

logistic regression was performed with the non-inflammatory control group as the reference category. No significant differences in biomarker levels were observed across non-IBD control subtypes by ANOVA (all $p > 0.05$) or multinomial logistic regression with the non-inflammatory control group as reference (all $p > 0.05$), and all four biomarkers were significantly different in IBD compared with each Non-IBD control subtype by both t-tests and binomial logistic regression (all $p < 0.05$) (Supplemental Table S9). These results suggest that the inflammatory status in the non-IBD controls did not substantially influence biomarker levels and differentiation from IBD. The biomarkers differentiate IBD from non-IBD regardless of the presence or location of inflammation in controls, while remaining stable across control subtypes themselves.

Hierarchical clustering using these immunoassay data showed significant differentiation between IBD and non-IBD controls but not perfect separation (Fig. 5B). Similarly, PCA representation of the 4-protein SVM model achieved a reasonable separation with an L1OXV accuracy of 75.69% (Fig. 5C). Among the seven patients with IBD-U five were classified correctly as IBD (Fig. 5C). Similarly, when evaluating the classification of patients with very early onset IBD (VEO-IBD) 11 of 13 classified correctly as IBD (Fig. 5C). Thus, both IBD-U and VEO-IBD show similar protein expression levels in serum for the selected proteins as the rest of the patients with IBD. The ROC curve of the 4-protein SVM model performed very well, showing excellent diagnostic performance with an AUC of 0.86 [0.81, 0.90] (Fig. 5D). Multivariable logistic regression using the minimum-deviance-coefficients (Supplemental Figure S1A and Supplemental Results) resulted in an AUC [95% CI] of 0.86 [0.8085, 0.8969] (Supplemental Figure S1D). To compare this model to established, routinely used biomarkers we calculated the AUC for CRP and ESR. The AUCs [95% CI] for CRP was 0.6987 [0.6154, 0.7730] (Supplemental Figure S6A) and for ESR 0.7425 [0.6568, 0.8083] (Supplemental Figure S6B), indicating that our 4-protein model outperforms established markers. Faecal calprotectin levels were not available for these samples, since these patient samples were collected before faecal calprotectin became a standard test.

Validation of prediction models on an independent test cohort

To independently clinically validate the prediction models from the SomaScan and Ella measures, we applied the Omni 1000 immunoassay platform, measuring 1000 proteins, to serum samples from a second multi-center paediatric IBD cohort of 105 subjects from the MGH, Columbia University Irving Medical Center, Upstate Golisano Children's Hospital, Riley Children's Hospital, Children's Hospital of Philadelphia, and University of California San Francisco Medical Center, including 79 IBD (25 UC, 54 CD) and 26 non-IBD controls (Supplemental Table S10).

PCA representation of the 4-protein IBD SVM model achieved a similar separation as the verification cohort with an L1OXV accuracy of 74.04% (Supplemental Figure S7A). The ROC curve of the 4-protein IBD SVM model performed with an AUC of 0.90 [0.82, 0.95] (Fig. 5E). While not all 8 proteins from the SomaScan-derived IBD model were covered by Omni 1000, we were able to test a model with 6 (HP, MMP1, MMP3, RETN, CRP, CXCL11) out of the 8 proteins. Not all 8 proteins are currently available with well-validated immunoassays. PCA representation of the 6-protein IBD SVM model achieved a separation with an L1OXV accuracy of 81.73% (Supplemental Figure S7B). The ROC curve of the 6-protein IBD SVM model performed with an AUC of 0.89 [0.77, 0.93] on the Omni 1000 Data (Supplemental Figure S7C). Reanalysis of the SomaScan data using the 6-protein IBD model resulted in an AUC of 0.98 [0.90, 1.00]. Supplemental Table S11 provides all the Omni 1000 data for the selected proteins.

For the UC vs CD SVM SomaScan-derived 7-protein model 56 CD and 25 UC samples were analysed on the 11k SomaScan v5.0 assay measuring 10,775 proteins. Supplemental Table S12 provides all 11k SomaScan v5.0 data for the selected proteins. PCA representation of the 7-protein UC vs CD SVM model achieved a similar separation as the discovery SomaScan cohort with an L1OXV accuracy of 77.78% (Supplemental Figure S8A). The ROC curve of the 7-protein UC vs CD SVM model performed with an AUC of 0.94 [0.87, 0.98] (Fig. 5F), close to the AUC of 1 for the discovery cohort. A subset of 4 proteins (MMP12, CCL22, KLK14, LGMN) was measured on Omni 1000 (same samples as 11k SomaScan except 2 CD samples less) and a 4-protein UC vs CD SVM model achieved an L1OXV accuracy of 81.01% (Supplemental Figure S8B) and AUC of 0.93 [0.84, 0.97] (Fig. 5G).

Discussion

Currently, no single test is used in clinical practice to diagnose IBD. Diagnosis is based on clinical, endoscopic, pathological and radiologic findings along with a combination of laboratory tests.[26] For children who are more often in the acute phase of IBD with intermittent symptoms, diagnostic evaluation may result in treatment delays, ultimately impacting patient outcomes.[27] Biomarkers such as CRP and faecal calprotectin may increase suspicion for IBD and lead to scheduling diagnostic tests. Non-invasive diagnostic tools such as ELISA which are cost-effective and reliable,[28] could contribute to the diagnostic process.

This study used SomaScan proteomics for biomarker discovery in a paediatric population to identify 95 serum proteins that distinguish IBD from non-IBD controls and 70 proteins differentiating UC from CD, creating the potential for non-invasive diagnostic biomarkers that could enhance the evaluation of children suspected of having IBD. To validate the biomarker discovery from SomaScan, serum samples from 295 paediatric subjects were analysed using the orthogonal Ella immunoassay platform. Validation of a 4-protein subset (MMP1, MMP3, HP, RETN) from the SomaScan-derived 8-protein IBD prediction model by ELISA across the 295 paediatric patients still achieved a remarkable AUC of 0.86 which was significantly higher than the AUCs obtained with CRP (0.70) or ESR (0.74) for the same samples. The robustness of the 4-protein IBD prediction model was further confirmed in an independent multi-center validation cohort, achieving an AUC of 0.90, highlighting the potential of this set of serum proteins to accurately diagnose IBD. Patients with IBD-U and VEO-IBD were included, in order to generate a predictor for IBD that is generally applicable to diagnose IBD. IBD-U and VEO-IBD cases were separated from non-IBD as efficiently as the rest of the IBD cases. Notably, the paediatric cohort evaluated in this study included individuals with inflammatory and non-inflammatory non-IBD gastrointestinal disorders, many of whom were being evaluated for IBD. Importantly, expression levels of these four proteins in IBD were not significantly impacted whether patients were on or off therapy. Among the treated IBD patients, stratification for specific medication classes showed no changes in protein levels for systemic 5-ASA, but a significant reduction in RETN for patients on steroids, even though RETN remained significant for differentiation from non-IBD. This indicates that RETN may be modulated by steroid use, but not sufficiently to go back to normal levels. Our cohort did not include patients treated with anti-TNF or other biologics, so we cannot draw any conclusions how this treatment impacts any of the 4 proteins. Disease activity stratification of the IBD cases into active disease and in remission revealed that all 4 proteins were significantly elevated in both active and in remission IBD compared to non-IBD; however, levels of RETN and HP were significantly higher in active disease when compared to values in remission. Thus, RETN and HP may be not only IBD markers but also disease activity markers. Since RETN levels are impacted by at least some therapies such as corticosteroids and anti-TNF

agents in other studies, RETN may be a relevant biomarker for both IBD status, disease activity, and treatment response.

While a subset of individuals in the non-IBD group did not have any identified inflammatory gastrointestinal disease, most were evaluated to determine the cause of abdominal pain. Nevertheless, many non-IBD subjects had inflammatory gastrointestinal conditions unrelated to IBD (eosinophilic oesophagitis, gastritis, celiac disease, collagenous gastritis, infectious enterocolitis, microscopic colitis), demonstrating that the IBD-associated proteins and predictor model not only discriminate between IBD and non-IBD but also between different causes of gastrointestinal inflammation. Differences in inflammatory components were also observed when comparing UC to CD, suggesting that UC and CD share some inflammatory pathways as shown by common discrimination of both from non-IBD but also have additional distinguishing pathways that differentiate between CD and UC. To determine whether inflammatory gastrointestinal conditions unrelated to IBD are similar to inflammation in IBD we also modelled the association between the biomarker levels and inflammation status in the non-IBD group, focussing on non-IBD without inflammation, non-IBD with upper GI inflammation, and non-IBD with inflammation in the colon. All 4 proteins remained significantly elevated for each of the 3 non-IBD classes, further supporting that these biomarkers discriminate IBD from non-IBD independent of inflammation status.

Our model outputs are consistent with prior studies from Prometheus[29] demonstrating that serologic, genetic, and inflammatory biomarkers can diagnose IBD. While currently no blood test definitively diagnoses IBD, Prometheus is offering a test for guiding IBD diagnosis, IBD sgi Diagnostic®, that combines 17 serologic, genetic and inflammatory markers[29] with an AUC of 0.871 which is very similar to the AUCs of 0.86 and 0.90 achieved with our 4-protein IBD predictor on two different cohorts. None of our 4 proteins overlap with markers in IBD sgi Diagnostic®. Compared with previously published models, our approach shows comparable performance while using fewer input variables. Since our SomaScan-derived 8-protein predictor reached an AUC of 0.95, we anticipate that the AUC for the full 8-protein IBD predictor, once validated by ELISA, will reach an AUC that is at least as accurate as or superior to the IBD sgi Diagnostic®. Therefore, our prediction model could be a valuable addition to the evaluation of gastrointestinal complaints in children or in differentiating between UC and CD.

The four ELISA-validated proteins were elevated in patients with IBD compared to controls and MMP1 as well as HP may also be higher in CD compared to UC. These four proteins may be involved in the pathogenesis of IBD.[30,31] Further experimental validation is needed to evaluate their relevance for IBD pathogenesis. MMPs that are upregulated in IBD play a significant role in inflammatory processes, particularly in mucosal ulceration, tissue remodelling, and repair.[32,33] Von Lampe et al. noted an elevation in MMP1 and MMP3 in the inflamed colon of patients with UC and CD.[34] Makitalo et al. have studied MMP expression in intestinal epithelium and noted a greater MMP7 expression in CD compared to UC and increased MMP10 in CD and UC samples compared to controls.[30] Interestingly, in our paediatric patient with IBD cohort, MMP12 and MMP10 were the most highly elevated proteins in UC when compared to CD; whereas MMP8 was elevated in CD vs UC, indicating that these metalloproteinases discriminate well between UC and CD.

Resistin, a pro-inflammatory adipokine, is significantly increased in chronic inflammation and higher resistin levels have been observed in IBD.[31,35] Resistin along with elastase and lactoferrin have been shown to be significantly elevated in both UC and CD.[35] A decrease in resistin following infliximab therapy has been reported.[36] Haptoglobin is reported to be elevated in the stool of paediatric patients with IBD when compared to non-IBD controls.[37]

IPA and STRING analysis provided deeper insights into biological systems involved in paediatric-onset IBD. Key upstream regulators and biological functions associated with the differentially expressed proteins included vasculature development, leucocyte migration, phagocytosis, and angiogenesis. IPA of IBD-associated proteins confirmed inflammatory response, as expected, as a significantly enriched key biological function. That observation was further confirmed by enrichment of proinflammatory cytokines, including TNF, a well-established therapeutic target for biologics,[38] as predicted upstream regulators of most IBD-associated proteins. Notably, these differentially expressed proteins were also associated with NF-kappa B (NF-κB), and Interleukin-6, which play a role in inflammatory processes related to IBD.[39] NF-κB's role in IBD has been associated with nucleotide-binding oligomerization domain-containing protein 2 (NOD2) activation, which is crucial in intestinal barrier function and activation of the Interleukin-23 (IL-23) pathway. This activation through Toll like receptor (TLR) signalling leads to elevated production of pro- and anti-inflammatory cytokines which are essential for the immune response. Additionally, TLR signalling stimulates secretion of immunoregulatory proteins and antimicrobial peptides, which play pivotal roles in modulating inflammation and defending against pathogens.[39,40]

In our study, 35 significantly altered proteins including MMP1, MMP2, MMP3, MMP8, LCN2, TIMP2, PAPPA, and NOTCH3 were linked to angiogenesis and endothelial integrity, indicating a role for vascular dysfunction in IBD pathology.[32] Increased

mesenteric vasculature is a hallmark especially in CD where a "comb" sign, prominence of mesenteric vessels, can be observed radiographically in areas of active inflammation. Abnormal angiogenesis and endothelial dysfunction play important roles in the pathogenesis and progression of IBD, presumably because of chronic inflammation impacting vascular permeability, allowing immune cells and inflammatory mediators to migrate into the gut tissue. The inflammatory process may also trigger release of pro-angiogenic factors and result in excessive and disorganised blood vessel growth, resulting in increased vascularity, and tissue damage.[41]

To differentiate between UC and CD our original 7-protein SomaScan-derived model generated a perfect AUC of 1 on a relatively small set of samples, potentially resulting in overfitting. To address this issue, we analysed an independent, larger test set (N = 56) on the latest 11k version of SomaScan v5.0[42] as well as on a new immunoassay-based proteomics platform, Omni 1000 from Nomic.[43] The 7-protein model predicted UC vs CD with an AUC of 0.94 on the 11k SomaScan, close to the original dataset. A 4-protein subset model on Omni 1000 also achieved an AUC of 0.93. The biomarkers for discrimination between UC and CD that we identified appear to be determined by the immunopathological mechanisms in the host that result in UC or CD. Although the distribution of inflammation differs across clinical phenotypes, what determines which areas are inflamed remains unknown. Nevertheless, the immunologic mechanisms known in UC and CD appear to be determined by the disease itself rather than by its location.[44] This study is not powered to evaluate biomarkers across the various subtypes of CD or UC. In future studies, we will seek to identify biomarkers that will personalise medical decisions based on the patient, not the disease.

Our comparison of UC to CD demonstrated that not only were MMP12 and MMP10 significantly increased in UC compared to CD but also lysosomal proteases such as CST6 and CTSV were differentially expressed in UC and CD. MMP12 is known to be elevated in colonic mucosa in UC compared to healthy individuals, possibly associated with intestinal fibrosis and epithelial cell degradation.[45] Functional relevance of MMP12 is supported by mouse model studies where MMP12 knockout protected against colitis.[46] Thus, differential expression profiles of MMPs in UC vs CD may contribute to differences in UC and CD phenotypes.

Components of the complement cascade (C3, C9, C1S, C1QBP) were up-regulated in CD compared to both UC and non-IBD, suggesting that dysregulation in the complement system enhanced innate immune response in patients with CD compared to UC or non-IBD. While complement system activation is crucial for eliminating pathogens, overactivity and lack of control enhance inflammatory processes. This is also reflected in the elevated CRP levels seen in CD vs UC in our study. Elevated serum levels of C3 in CD compared to UC are previously described.[47] The complement system appears to play an important role in intestinal immune response during chronic inflammation. C3 knockout mice show divergent results in different models.[48] C3 knockout protects mice from acute dextran sulphate sodium (DSS)-induced colitis but in a different knockout model C3 deficiency promotes inflammatory responses in the mid colon.[48,49]

Strengths of this study include use of SomaScan proteomics to discover serum biomarkers to discriminate IBD from non-IBD and UC from CD, further validation of 4 proteins by an orthogonal, "gold standard" ELISA method and a new proteomics assay, Omni 1000, across two larger cohorts of 295 and 105 individuals, and development of a multi-protein models that diagnose IBD and differentiate UC from CD with high accuracy. The identified biomarkers provide new insights into etiopathological mechanisms underlying IBD development and differentiating UC from CD.

We acknowledge potential limitations in our study. First, while immunoassay validation in two cohorts included significantly more patients than SomaScan discovery, we did not evaluate validity of these biomarkers in adult IBD. Therefore, it is not known whether these data are generalisable from paediatric to adult IBD. Second, the total sample size is relatively modest for diagnostic biomarker discovery and validation. Although the sample size was limited, we identified proteins with high predictive accuracies for differentiating IBD from non-IBD and UC from CD. Larger cohort studies are necessary to further validate these results. The retrospective nature of the study might introduce biases related to patient selection and data collection. Prospective studies may provide an unbiased way to confirm these findings. The promising performance of the multi-protein prediction models underscores the need for prospective validation in an incident cohort—such as children presenting with suspected IBD—to more accurately assess diagnostic performance and determine their potential to reduce reliance on invasive procedures. Third, the patient cohorts included primarily Caucasian patients and, consequently, generalisation to individuals in other ethnic populations is not possible. Fourth, as our analysis was limited to serum proteins, important biomarkers present in other biological samples such as faecal samples and tissue biopsies were not considered. Fifth, this is an observational study and while serum proteins fit into biological pathways relevant to IBD, they could be surrogate markers for IBD manifestation without being causally involved. Ultimately, clinical trials targeting such IBD-associated proteins may confirm their mechanistic relevance. Sixth, since the SomaScan version used here only measured 1305 proteins, we most likely missed many additional proteins

associated with IBD. Nevertheless, we identified and validated 4 proteins with high biological plausibility and involvement in key pathophysiological pathways postulated to be linked to IBD. Seventh, while we have validated 4 proteins of the 8-protein IBD predictor as IBD disease biomarkers by immunoassays and achieved an excellent AUC, not testing the additional other four proteins could affect diagnostic performance. Seventh, while we compared the diagnostic performance to CRP and ESR performance, we were unable to do this for faecal calprotectin, because most blood samples had been collected prior to faecal calprotectin testing and the non-IBD controls lacked any faecal calprotectin testing.

This study highlights the potential of serum proteomics in diagnosing paediatric IBD pathogenesis and distinguishing UC and CD. Evaluating pathophysiology and biomarkers in paediatric patients could be more relevant than previous studies in adults who are more likely to have well established chronic disease; whereas, children with IBD are more likely to be in the initial active inflammatory phase of these chronic illnesses. We identified 95 serum proteins that distinguished IBD from non-IBD healthy controls and 70 serum proteins that differentiated UC from CD and linked them to highly relevant pathways. Importantly, we validated four IBD predictor proteins by ELISA and demonstrated their high predictive performance when incorporated into a multi-protein IBD predictor model. These findings offer new insights into pathophysiological mechanisms contributing to paediatric IBD development and pathways differentiating UC from CD. Performance of the IBD predictor model and additional proteins that may further enhance its diagnostic performance provide new opportunities for non-invasive clinical differential diagnosis in paediatric patients with GI complaints, inform patient pathways, as well as monitoring patients with IBD for remission and flares. Furthermore, our multi-protein proteomic IBD prediction model requires further prospective studies of clinical validity in large and diverse multicenter cohorts as well as comparison of identified biomarkers with available biomarkers to confirm clinical use. Future studies will also add other protein biomarkers based on broader proteomics analysis. By utilising two distinct proteomic platforms (SomaScan, Ella), we have substantially increased the list of promising IBD biomarkers, including several proteins associated with inflammation and vascular dysfunction. Some of these proteins may emerge as therapeutic targets, potentially modifiable by existing or new medications to treat IBD.

Contributors

HSW and TAL conceived the study, planned, designed, and interpreted experiments, and contributed to funding acquisition, supervision, writing, review and editing. MO wrote the original draft and contributed to literature search, review and editing of the manuscript. CP provided support to the collection of biosamples, clinical data curation, selected and aliquoted the serum samples. XG, STD, and TLM performed the proteomic studies, and contributed to data curation, formal analysis and methodology. HHO and HC performed statistical, bioinformatics, and modelling analysis of the protein data, and contributed to methodology and data curation. TAL performed the pathway analysis. HSW, TAL, HC, JYH, and HO made the figures and tables. HSW, TLM, MO, and JYH provided support to the collection and curation of clinical data. LBG contributed to writing, interpretation, reviewing and editing the manuscript. All authors read, critically reviewed and approved the final manuscript. HSW, HHO, and TAL have accessed and verified the underlying data.

Data sharing statement

We declare that all data supporting the findings of this study are available including code identifier so that it can be accessed and available prior to acceptance. Additional data can be obtained from the corresponding author upon reasonable request. Raw SomaScan data have been deposited at NCBI GEO and are available at GEO GSE317366. Other raw data are presented in Supplemental Tables S11 and S12. All analyses were conducted using established, previously published methods implemented in standard, publicly available and appropriately cited software packages. There is no separately deposed study-specific analysis code.

Declaration of interests

LBG—Consultant: Pfizer, Research Support: Sanofi. HSW—Consultant: Abbvie, Janssen, Pfizer, Phathom Pharmaceuticals; Research Support: Janssen, Pfizer, Nestle, Women's Wellness Foundation, Autism Research Institute, Abbvie, Paediatric IBD Foundation. TAL—Research Support: Sanofi. All other authors have no interests to declare.

Acknowledgements

The authors gratefully acknowledge the contributions of the patients, family members, nurses, physicians and staff members.

This study was supported by philanthropic support from Martin Schlaff, The Diane and Dorothy Brooks Foundation, The Manessis Family, Thomas and Lynn Kuzma, and The Hasso Serrano Foundation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2026.106311.

References

1 Day AS, Lemberg DA. Identification and diagnosis of Crohn disease and ulcerative colitis in children. *J Paediatr Child Health*. 2020;56(11):1731–1734.
2 Khan R, Kuenzig ME, Benchimol EI. Epidemiology of pediatric inflammatory bowel disease. *Gastroenterol Clin N Am*. 2023;52(3):483–496.
3 Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet (London, England)*. 2016;387(10014):156–167. https://doi.org/10.1016/S0140-6736(15)00465-1.
4 Kuenzig ME, Fung SG, Marderfeld L, et al. Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. *Gastroenterology*. 2022;162(4):1147–1159.e4. https://doi.org/10.1053/j.gastro.2021.12.282.
5 Ye Y, Manne S, Treem WR, Bennett D. Prevalence of inflammatory bowel disease in pediatric and adult populations: recent estimates from large national databases in the United States, 2007-2016. *Inflamm Bowel Dis*. 2020;26(4):619–625. https://doi.org/10.1093/ibd/izz182.
6 Mosli MH, Zou G, Garg SK, et al. C-Reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. *Am J Gastroenterol*. 2015;110(6):802–820. https://doi.org/10.1038/ajg.2015.120.
7 Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: a systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(5):637–645.

8 Solem CA, Loftus EV Jr, Tremaine WJ, Harmsen WS, Zinsmeister AR, Sandborn WJ. Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease. *Inflamm Bowel Dis*. 2005;11(8):707–712. https://doi.org/10.1097/01.mib.0000173271.18319.53.

9 Bray C, Bell LN, Liang H, et al. Erythrocyte sedimentation rate and C-reactive protein measurements and their relevance in clinical medicine. *WMJ*. 2016;115(6):317–321.

10 Noor NM, Lee JC, Bond S, et al. A biomarker-stratified comparison of top-down versus accelerated step-up treatment strategies for patients with newly diagnosed Crohn's disease (PROFILE): a multicentre, open-label randomised controlled trial. *Lancet Gastroenterol Hepatol*. 2024;9(5):415–427. https://doi.org/10.1016/S2468-1253(24)00034-7.

11 Cannarozzi AL, Latiano A, Massimino L, et al. Inflammatory bowel disease genomics, transcriptomics, proteomics and metagenomics meet artificial intelligence. *United Eur Gastroenterol J*. 2024;12(10):1461–1480. https://doi.org/10.1002/ueg2.12655.

12 Webber J, Stone TC, Katilius E, et al. Proteomics analysis of cancer exosomes using a novel modified aptamer-based array (SOMAscan™) platform. *Mol Cell Proteomics*. 2014;13(4):1050–1064. https://doi.org/10.1074/mcp.M113.032136.

13 Ngo D, Sinha S, Shen D, et al. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation*. 2016;134(4):270–285. https://doi.org/10.1161/CIRCULATIONAHA.116.021803.

14 Posavi M, Diaz-Ortiz M, Liu B, et al. Characterization of Parkinson's disease using blood-based biomarkers: a multicohort proteomic analysis. *PLoS Med*. 2019;16(10):e1002931. https://doi.org/10.1371/journal.pmed.1002931.

15 Heier CR, Fiorillo AA, Chaisson E, et al. Identification of pathway-specific serum biomarkers of response to glucocorticoid and infliximab treatment in children with inflammatory bowel disease. *Clin Transl Gastroenterol*. 2016;7(9):e192. https://doi.org/10.1038/ctg.2016.49.

16 Hathout Y, Conklin LS, Seol H, et al. Serum pharmacodynamic biomarkers for chronic corticosteroid treatment of children. *Sci Rep*. 2016;6:31727. https://doi.org/10.1038/srep31727.

17 Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795–806. https://doi.org/10.1097/MPG.0000000000000239.

18 Shubin AV, Kollar B, Dillon ST, Pomahac B, Libermann TA, Riella LV. Blood proteome profiling using aptamer-based technology for rejection biomarker discovery in transplantation. *Sci Data*. 2019;6(1):314. https://doi.org/10.1038/s41597-019-0324-y.

19 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc*. 1995;57:289–300.

20 Hoaglin D, Mosteller F, Tukey JW. Introduction to more refined estimators. In: Hoaglin D, Mosteller F, Tukey JW, eds. *Understanding Robust and Exploratory Data Analysis*. New York: John Wiley; 2000. https://www.wiley.com/en-us/Understanding+Robust+and+Exploratory+Data+Analysis-p-9780471384915.

21 Sneath PHA, Sokal R. Numerical taxonomy. The principles and practice of numerical classification. *Quart Rev Biol*. 1975;50:525–526.

22 Pearson KL III. On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philosoph Mag J Sci*. 1901;2:559–572.

23 Brereton RG, Lloyd GR. Support vector machines for classification and regression. *The Analyst*. 2010;135(2):230–267. https://doi.org/10.1039/b918972f.

24 Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523–530. https://doi.org/10.1093/bioinformatics/btt703.

25 Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–D613. https://doi.org/10.1093/nar/gky1131.

26 Nikolaus S, Schreiber S. Diagnostics of inflammatory bowel disease. *Gastroenterology*. 2007;133(5):1670–1689.

27 Nguyen VQ, Jiang D, Hoffman SN, et al. Impact of diagnostic delay and associated factors on clinical outcomes in a U.S. inflammatory bowel disease cohort. *Inflamm Bowel Dis*. 2017;23(10):1825–1831. https://doi.org/10.1097/MIB.0000000000001257.

28 Vakil N, Rhew D, Soll A, Ofman JJ. The cost-effectiveness of diagnostic testing strategies for Helicobacter pylori. *Am J Gastroenterol*. 2000;95(7):1691–1698.

29 Plevy S, Silverberg MS, Lockton S, et al. Combined serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients. *Inflamm Bowel Dis*. 2013;19(6):1139–1148. https://doi.org/10.1097/MIB.0b013e318280b19e.

30 Mäkitalo L, Kolho KL, Karikoski R, Anthoni H, Saarialho-Kere U. Expression profiles of matrix metalloproteinases and their inhibitors in colonic inflammation related to pediatric inflammatory bowel disease. *Scand J Gastroenterol*. 2010;45(7–8):862–871. https://doi.org/10.3109/00365520903583863.

31 Konrad A, Lehrke M, Schachinger V, et al. Resistin is an inflammatory marker of inflammatory bowel disease in humans. *Eur J Gastroenterol Hepatol*. 2007;19(12):1070–1074. https://doi:10.1097/MEG.0b013e3282f16251.

32 Arihiro S, Ohtani H, Hiwatashi N, Torii A, Sorsa T, Nagura H. Vascular smooth muscle cells and pericytes express MMP-1, MMP-9, TIMP-1 and type I procollagen in inflammatory bowel disease. *Histopathology*. 2001;39(1):50–59. https://doi.org/10.1046/j.1365-2559.2001.01142.x.

33 Rath T, Roderfeld M, Halwe JM, Tschuschner A, Roeb E, Graf J. Cellular sources of MMP-7, MMP-13 and MMP-28 in ulcerative colitis. *Scand J Gastroenterol*. 2010;45(10):1186–1196. https://doi.org/10.3109/00365521.2010.499961.

34 von Lampe B, Barthel B, Coupland SE, Riecken EO, Rosewicz S. Differential expression of matrix metalloproteinases and their tissue inhibitors in colon mucosa of patients with inflammatory bowel disease. *Gut*. 2000;47(1):63–73. https://doi.org/10.1136/gut.47.1.63.

35 Louis Sam Titus ASC, Vanarsa K, Soomro S, et al. Resistin, elastase, and lactoferrin as potential plasma biomarkers of pediatric inflammatory bowel disease based on comprehensive proteomic screens. *Mol Cell Proteomics*. 2023;22(2):100487. https://doi.org/10.1016/j.mcpro.2022.100487.

36 Karmiris K, Koutroubakis IE, Xidakis C, Polychronaki M, Kouroumalis EA. The effect of infliximab on circulating levels of leptin, adiponectin and resistin in patients with inflammatory bowel disease. *Eur J Gastroenterol Hepatol*. 2007;19(9):789–794. https://doi.org/10.1097/MEG.0b013e3282202bca.

37 Soomro S, Venkateswaran S, Vanarsa K, et al. Predicting disease course in ulcerative colitis using stool proteins identified through an aptamer-based screen. *Nature communications*. 2021;12(1):3989. https://doi.org/10.1038/s41467-021-24235-0.

38 Hyams J, Crandall W, Kugathasan S, et al. Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. *Gastroenterology*. 2007;132(3):863–1166. https://doi.org/10.1053/j.gastro.2006.12.003.

39 Mukherjee T, Kumar N, Chawla M, Philpott DJ, Basak S. The NF-κB signaling system in the immunopathogenesis of inflammatory bowel disease. *Sci Signal*. 2024;17(818):eadh1641. https://doi.org/10.1126/scisignal.adh1641.

40 Dejban P, Nikravangolsefid N, Chamanara M, Dehpour A, Rashidian A. The role of medicinal products in the treatment of inflammatory bowel diseases (IBD) through inhibition of TLR4/NF-kappaB pathway. *Phytother Res*. 2021;35(2):835–845. https://doi.org/10.1002/ptr.6866.

41 Deban L, Correale C, Vetrano S, Malesci A, Danese S. Multiple pathogenic roles of microvasculature in inflammatory bowel disease: a Jack of all trades. *The American journal of pathology*. 2008;172(6):1457–1466. https://doi.org/10.2353/ajpath.2008.070593.

42 Candia J. SomaScan Bioinformatics: normalization, quality control, and assessment of pre-analytical variation. *Methods Mol Biol*. 2025;2929:107–127.

43 Dagher M, Ongo G, Robichaud N, et al. nELISA: a high-throughput, high-plex platform enables quantitative profiling of the inflammatory secretome. *Nat Methods*. 2025;22:2375–2385.

44 Bai X, Liu W, Chen H, Zuo T, Wu X. Immune cell landscaping reveals distinct immune signatures of inflammatory bowel disease. *Front Immunol*. 2022;13:861790.

45 Arosa L, Camba-Gómez M, Lorenzo-Martín LF, et al. RNA expression of MMP12 is strongly associated with inflammatory bowel disease and is regulated by metabolic pathways in RAW 264.7 macrophages. *Int J Mol Sci*. 2024;25:3167.

46 Nighot M, Ganapathy AS, Saha K, et al. Matrix metalloproteinase MMP-12 promotes macrophage transmigration across intestinal epithelial tight junctions and increases severity of experimental colitis. *J Crohns Colitis*. 2021;15:1751–1765.
47 Bene L, Füst G, Kovács A, et al. High normal serum levels of C3 and C1 inhibitor, two acute-phase proteins belonging to the complement system, occur more frequently in patients with Crohn's disease than ulcerative colitis. *Dig Dis Sci*. 2003;48:1186–1192.
48 Choi YJ, Kim JE, Lee SJ, et al. Promotion of the inflammatory response in mid colon of complement component 3 knockout mice. *Sci Rep*. 2022;12:1700.
49 Schepp-Berglind J, Atkinson C, Elvington M, Qiao F, Mannon P, Tomlinson S. Complement-dependent injury and protection in a murine model of acute dextran sulfate sodium-induced colitis. *J Immunol*. 2012;188(12):6309–6318. https://doi.org/10.4049/jimmunol.1200553.