ORIGINAL ARTICLE

Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate Crohn's disease from UC

Amanda E Starr,[1] Shelley A Deeke,[1] Zhibin Ning,[1] Cheng-Kang Chiang,[1] Xu Zhang,[1] Walid Mottawea,[1,2] Ruth Singleton,[3] Eric I Benchimol,[3,4,5] Ming Wen,[1] David R Mack,[3,4] Alain Stintzi,[1] Daniel Figeys[1,6]

ABSTRACT

Objective Accurate differentiation between Crohn's disease (CD) and UC is important to ensure early and appropriate therapeutic intervention. We sought to identify proteins that enable differentiation between CD and UC in children with new onset IBD.

Design Mucosal biopsies were obtained from children undergoing baseline diagnostic endoscopy prior to therapeutic interventions. Using a super-stable isotope labeling with amino acids in cell culture (SILAC)-based approach, the proteomes of 99 paediatric control and biopsies of patients with CD and UC were compared. Multivariate analysis of a subset of these (n=50) was applied to identify novel biomarkers, which were validated in a second subset (n=49).

Results In the discovery cohort, a panel of five proteins was sufficient to distinguish control from IBD-affected tissue biopsies with an AUC of 1.0 (95% CI 0.99 to 1.0); a second panel of 12 proteins segregated inflamed CD from UC within an AUC of 0.95 (95% CI 0.86 to 1.0). Application of the two panels to the validation cohort resulted in accurate classification of 95.9% (IBD from control) and 80% (CD from UC) of patients. 116 proteins were identified to have correlation with the severity of disease, four of which were components of the two panels, including visfatin and metallothionein-2.

Conclusions This study has identified two panels of candidate biomarkers for the diagnosis of IBD and the differentiation of IBD subtypes to guide appropriate therapeutic interventions in paediatric patients.

Significance of this study

What is already known on this subject?
► Accurate diagnosis of Crohn's disease (CD) and UC is important, so appropriate therapy can be initiated to reduce disease progression, complications of disease related to permanent bowel damage and avoid unnecessary adverse drug events.
► Differentiation through clinical symptoms, site of disease, presence of granulomas, genetics and current serological tests each have limitations.
► Proteomic analyses are powerful tools to assess quantitative and qualitative data about biological systems.

What are the new findings?
► An unbiased proteomics screen of biopsies from an inception cohort of patients with IBD led to the identification of a panel of proteins that can be used to differentiate patients with CD from those with UC.
► Candidate biomarkers suggest elevated fatty acid metabolism in paediatric CD and elevated protein metabolism in paediatric UC.

How might it impact on clinical practice in the foreseeable future?
► The utilisation of these protein biomarkers into the clinical setting could improve accuracy of the current methods used to differentiate subtypes of IBD and do so in a rapid time frame. Further understanding of the pathogenesis of paediatric IBD may also be realised.

INTRODUCTION

The IBDs are chronic, relapsing and remitting, GI inflammatory conditions that affect over 4 million people worldwide.[1] In Canada, over 230 000 individuals suffer from IBD, with an additional 10 000 cases diagnosed each year.[2] IBD includes Crohn's disease (CD) and UC, which differ in clinical presentation and complications and may have different responses to treatment.[3–5] Paediatric patients, 10%–25% of the IBD population,[6 7] often present with different features from adult patients with IBD, and have more aggressive and extensive disease and more long-term severe outcomes.[4–8] Fewer children have the so-called classical symptoms, and instead have a range of presenting features including atypical symptoms such as short stature or weight loss leading to delayed recognition and diagnosis.[9] Generally, CD is characterised by discreet mucosal and submucosal lesions, which can occur anywhere throughout the GI tract, and are transmural in nature. In contrast, UC extends proximally from the rectum, with contiguous, but generally superficial inflammation. However, non-

specific gastric involvement, relative rectal sparing and periappendiceal patch involvement in UC[10] can add to diagnostic confusion and misclassification of disease. Differential diagnosis is important for the induction and maintenance of appropriate therapy, which can differ between CD and UC. In the paediatric population, exclusive enteral nutrition is an effective induction therapy in CD but not in UC.[11] For maintenance therapy, methotrexate is established for CD,[12] but not for UC,[13] whereas mesalamine is not.[14] Furthermore, complications related to the diseases are very different. For those unfortunate patients who do not respond to therapy, surgical options are very different with surgery for UC offering the end of the UC disease process and the opportunity for reservoir formation and reconstitution of the bowel continuity. For CD, with the very real possibility of recurrence elsewhere in the GI tract, surgery to remove inflamed segments of inflamed mucosa is generally not offered unless disease complications necessitate such action. Thus, for both consideration of therapeutic options and discussion of outcomes, correct diagnosis is an important factor in effective treatment in IBD.

Genome-wide association studies have identified nearly 170 susceptibility loci associated with IBD, including many loci that overlap in CD and UC,[15–17] yet disease manifestation due to these variants is less than 15%.[17] Biomarkers have been sought to complement current IBD diagnostic tools such as endoscopy, imaging and histology to reduce ambiguous diagnosis of IBD, assist in subtype differentiation and provide objective measures of disease. Previous studies have identified proteins that are elevated and measurable in serum or stool;[3] however, these proteins have been found to perform best in the more obvious cases of CD or UC in the paediatric population.[18] [19] Serum-detected antibodies directed against neutrophil or bacterial components tend to have low sensitivities (true positive rate <50%; reviewed[20]). Other biomarkers, namely faecal calprotectin, are clinically useful to identify patients with IBD from populations without mucosal inflammation (eg, IBS, healthy controls) but cannot differentiate IBD subtypes.[21] Furthermore, faecal calprotectin is not sufficient to distinguish between mild, moderate or severe disease,[19] which is important in deciding appropriate therapeutic intervention. Considering the limitations of current genetic and protein markers, atypical presentations and progression of IBD in the paediatric population, there is a clear need for new biomarkers and approaches that can rapidly and accurately provide diagnosis of CD and UC.

To identify novel protein biomarkers, proteomic approaches have been applied to serum, cell or tissue isolates from adult patients with IBD.[22] Findings of these early studies indicate the usefulness of proteomics in biomarker selection, but have been limited by the lack of an inception cohort and differences in patient therapies[23–28] and the inability to identify the differentiating proteins (use of MALDI/surface-enhanced laser desorption/ionization (SELDI),[25] [26] [29] or are low throughput (two-dimensional gels).[24] [27] [30] In this study, we overcame these limitations by employing a quantitative, high-throughput proteomic approach to evaluate and compare ascending colonic biopsies from 99 (39 control, 30 CD, 30 UC) therapy-naive children at the time of diagnosis with IBD. We conducted the first proteomic analysis of paediatric biopsies, and aimed to identify candidate biomarkers to contribute to the accurate subdiagnosis of IBD.

MATERIAL AND METHODS
Study design
We performed a cross-sectional study, the design for which is outlined graphically in online supplementary figure S1. The

patient biopsies, for which the variance of the proteomic data were <2.0, were randomly divided into equal groups between the discovery and the validation phase using a balanced stratification approach for gender and diagnosis (Etcetera in WinPepi, BixtonHealth.ca). The study was approved by the Children's Hospital of Eastern Ontario (CHEO) Research Ethics Board. Study data were collected and managed using Research Electronic Data Capture (REDCap), hosted at the CHEO Research Institute. REDCap is a secure, web-based application designed to support data capture for research studies.[31]

Study population
All patients under 18 years of age and scheduled to undergo diagnostic colonoscopy between October 2011 and February 2015 were considered eligible for recruitment. Exclusion criteria, related to conditions known to affect the intestinal microbiome and mucosal gene expression, included (1) a body mass index >95th percentile for age; (2) diabetes mellitus (insulin and non-insulin-dependent); (3) infectious gastroenteritis within the preceding 2 months or (4) use of any antibiotics, probiotics or immunomodulatory agents within the preceding 4 weeks. All IBD cases met the standard diagnostic criteria for either UC or CD following thorough clinical, microbiological, endoscopic, histological and radiological evaluation.[32] Phenotyping of disease was based on endoscopy and radiological findings and based on the Paris modification of the Montreal Classification for IBD.[33] Clinical disease activity of CD or UC was determined using the Pediatric Crohn's Disease Activity Index (PCDAI)[34] or the Pediatric UC Activity Index (PUCAI),[35] respectively. Since ascending colon and terminal ileum are the most common sites of CD, and pancolitis dominates in children with UC,[36] the ascending colon was chosen as the site for mucosal biopsy to eliminate the region of the bowel biopsied as a confounder. As such, patients with only IBD from whom affected/inflamed ascending colon (CoA) biopsies were obtained were included in the proteomic study. When possible, biopsies in these patients were also obtained from non-inflamed ascending colon (CoN). In contrast, only non-inflamed biopsies were available from control patients, who had macroscopically and histologically normal mucosa, and did not carry a diagnosis for any known chronic intestinal disorder.

Sample collection and processing
Detailed methodologies are provided in the online supplementary materials and methods. Briefly, frozen biopsies were lysed by mechanical homogenisation and proteins isolated following centrifugation. Sample protein was combined with an equal amount of isotopically labelled reference protein lysate to permit for relative quantification of proteins. Tryptic digestion of proteins was performed with filter-aided sample preparation,[37] and resulting peptides analysed on an Orbitrap Elite mass spectrometer (MS).

Statistical analysis and validation
All MS raw files were analysed in a single run with MaxQuant V.1.5.1, against the human Uniprot database (downloaded 2014/07/11). Data filtering and statistical analysis were performed in Perseus, Excel (Microsoft) and Prism (Graphpad). Quality assessment of MS data was performed (see online supplementary methods), and outliers determined using Robust regression and Outlier[38] were removed from the study.

Proteins quantified by ≥2 unique peptides in ≥95% of the biopsies (Q95) were determined, as were proteins considered to be subgroup specific due to the over-representation in one

subgroup (>70% of subgroup biopsies) when compared with at least one other subgroup (<50% of subgroup biopsies). To limit the effects due to imputation of missing data, only the data from the Q95+subgroup-specific proteins were used in principal component analysis (PCA; Matlab) and candidate biomarker identification.

Mathematical models for disease classification (control vs IBD CoA; CD CoA vs UC CoA) were developed in Receiving Operating Characteristics (ROC) Curve Explorer and Tester (ROCCET)[39] using proteomic data from a subset of the patients (discovery cohort), and the models substantiated with data from the remaining patients (validation cohort). Specifically, after cohort allocation, the Q95 and subgroup-specific proteins of the discovery cohort were determined and the associated proteomic data evaluated by Partial Least Squares Discriminant Analyses (PLSDA), Support Vector Machine (SVM) and Random Forest (RF) in the Explorer module of ROCCET.[39] Ultimately, proteins commonly identified in all three models were considered to be candidate biomarkers, and then ranked by the respective Area Under the ROC Curve (AUC) value. Biomarker panels were developed in the Tester module of ROCCET by iterative analysis with a PLSDA model using a step-forward method, with candidate biomarkers added by protein-specific AUC values (starting with the highest). The minimal number of proteins selected for inclusion in the panel was based upon the point of plateau for the AUC, to balance specificity and sensitivity. Biomarker panels were independently validated by applying the validation cohort data to the discovery-trained PLSDA models.

PCDAI or PUCAI scores in the discovery cohort were compared with all proteins in the cohort Q95+subgroup-specific proteins to determine the Pearson correlation (Graphpad, Prism). Pathway analyses were performed using DAVID (david. ncifcrf.gov) and Panther (Pantherdb.org) and visualised with iPATH2 interactive pathways explorer (pathways.embl.de) using Uniprot accession numbers. Enzyme-linked immunosorbent assays (ELISAs) for visfatin (Ezno Life Sciences, New York, USA) and metallothionein-2 (MT2) (Cloud-Clone, Texas, USA) were performed as per the manufacturer's protocol on biopsy lysate diluted to a final sodium dodecyl sulfate (SDS) concentration of 0.08%.

RESULTS
Descriptive characteristics
Biopsies from the ascending colon of 100 children undergoing diagnostic colonoscopy were obtained (see online supplementary table S1). The mean age at endoscopy in patients with IBD (n=61) was 13.6±0.4 years (range 4.8–17.8), and in controls (n=39) was 14.3±0.5 years (range 6.1–17.6), and were comparable between groups. Following quality assessment, one subject was considered an outlier and so removed from the bioinformatic analysis. A summary of characteristics for remaining 99 included patients are shown in table 1 (individual details are provided in online supplementary table S1). There was no difference in gender distribution between controls or patients with UC. Subjects diagnosed with CD were more likely to be male than female, characteristic for CD in a paediatric population.[40] The majority of patients with CD (90%) had active inflammatory colonic or ileocolonic disease; 86.7% of patients with UC exhibited pancolitis. While CoA biopsies were obtained from all 60 patients with IBD, due to the extensive nature of disease, CoN samples were obtained from only 23/30 CD and 2/30 patients with UC. For statistical significance to be achieved, only the CD CoN biopsies were analysed.

Evaluation of full proteomic data set
One hundred and twenty-four biopsies were processed over a 15-month period and analysed by high performance liquid

Table 1 Patient characteristics

	Discovery cohort			Validation cohort		
	Control	CD	UC	Control	CD	UC
Patients, n (m/f)	20 (10/10)	15 (10/5)	15 (7/8)	19 (9/10)	15 (10/5)	15 (6/9)
Age, y*	15.5 (11.6–16.6)	13.9 (10.8–16.6)	14.8 (13.3–16.3)	15.3 (12.8–16.5)	14.1 (10.8–15.8)	13.7 (11.1–16.5)
PCDAI/PUCAI*	NA	37.5 (32.5–57.5)	45.0 (30.0–70.0)	NA	47.5 (32.5–57.5)	50.0 (40.0–70.0)
Classification						
A1a		3			2	
A1b		12			13	
L1		3			0	
L2		2			2	
L3		10			13	
L4a		4			9	
L4b		1			2	
B1		15			15	
B2		0			0	
P		4			2	
G0		12			13	
G1		3			2	
E1			0			0
E2			1			1
E3			1			1
E4			13			13
S0			10			10
S1			5			5

*Values shown as median (IQR).
CD, Crohn's disease; NA, not applicable; PCDAI, Pediatric Crohn's Disease Activity Index; PUCAI, Pediatric UC Activity Index.

chromatography electrospray ionization mass spectrometry/mass spectrometry (HPLC-ESI-MSMS) to identify and quantify proteins that are differentially expressed between disease conditions. Detailed evaluation of MS data quality is provided in the online supplementary results. One affected (CoA) CD biopsy was determined to be an outlier (see online supplementary figure S2A) and so both CoA and CoN biopsies from the patient were removed from further analyses. The remaining samples showed consistent MS profiles over time (see online supplementary figure S2B–E).

From the 122 biopsies analysed, 3644 proteins were identified by ≥2 unique peptides, 949 of which were in the Q95; 225 additional proteins were subgroup specific (see online supplementary table S2). Control and affected (CoA) IBD proteomes are distinguished by PCA using proteomic data from these 1174 proteins, whereas non-inflamed (CoN) CD proteomes are more similar to the control group (figure 1). A similar segregation was obtained even when proteins annotated as involved in immunological response (see online supplementary tables S2 and S3) were removed from the dataset (see online supplementary figure S3A). Consistent with previous studies, blood-based parameters were insufficient to segregate IBD from control patients by PCA (see online supplementary figure S3B).

Establishment of biomarker models
Control versus affected IBD
Evaluation of the Q95+subgroup-specific proteins in control versus CoA IBD (combined CoA CD and CoA UC; see online supplementary table S2) by SVM, PLSDA and RF resulted in 106 common candidate biomarkers (see online supplementary table S4). By step-forward analysis of a PLSDA model, peak and stabilisation of the AUC, specificity and sensitivity were observed with five proteins (see online supplementary figure S5A). This panel of five proteins (table 2), the relative expressions of which are shown in figure 2A, was sufficient to differentiate patients with IBD from controls with an AUC of 1.0 (95% CI 0.99 to 1.0), and a classification accuracy of 96% (figure 2B, C). Notably, the relative expression of all five proteins from CD CoN biopsies was significantly different than both control and CD CoA, with CoN showing intermediary expression (see online supplementary figure S6A).
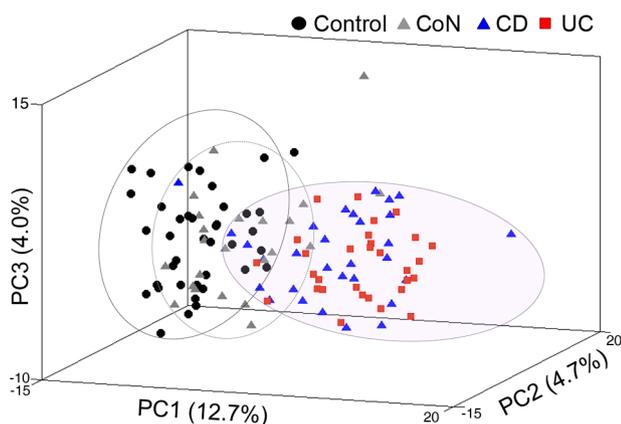

Figure 1 (A) Principal component analysis (PCA) of the Q95 +subgroup-specific proteins showing separation of IBD inflamed ascending colon patients proteomes (UC, red; Crohn's disease (CD), blue) from controls (black) and from non-inflamed CD ascending colon (CoN) (grey) based on the of Q95+subgroup-specific proteins.

Table 2 Panel 1 candidate protein biomarkers for the segregation of IBD from control patients

Protein	Uniprot	Gene	#Unique peptides	Ratio (IBD/Control)	AUC
Fatty acid-binding protein, epidermal	Q01469	FABP5	21	0.32	0.98875
Visfatin	P43490	NAMPT	24	2.60	0.98792
UDP-glucose 6-dehydrogenase	O60701	UGDH	31	0.22	0.98083
leucine-rich PPR motif-containing protein, mitochondrial	P42704	LRPPRC	73	0.45	0.97708
Inorganic pyrophosphatase	Q15181	PPA1	16	1.82	0.97542

Rows in dark grey indicate proteins elevated in patients with IBD.
LRPPRC, leucine rich pentatricopeptide repeat containing; NAMPT, nicotinamide phosphoribosyltranserase; PPR, pentatricopeptide repeat; UGDH, uracil-diphosphate-glucose 6-dehydrogenase.

CD versus UC
From the proteins evaluated in the 15 CD and 15 UC discovery cohort proteomes (see online supplementary table S2), 252 were common to SVM, PLSDA and RF (see online supplementary table S5). By step-forward analysis in a PLSDA model, a plateau in specificity and sensitivity was observed at 12 proteins (see online supplementary figure S5B), and thus determined to be the minimal number of proteins required for optimal classification. The relative expression of the 12 proteins is shown (figure 3A). Notably, the mitochondrial proteins trifunctional enzyme hydroxyacyl-CoA dehydrogenase/3-ketoacylCoA thiolase/enoyl-CoA hydratase Beta subunit (HADHB) and tricarboxylate transport protein (SLC25A1) were not significantly different between CD and UC by t test, though they contribute to the specificity and sensitivity of the panel (figure 3B). The panel of 12 proteins (table 3) resulted in an overall AUC of 0.95 (95% CI 0.86 to 1.0) (figure 3B), with a sensitivity and specificity of 1.0 (95% CI 0.78 to 1.0) and 0.933 (95% CI 0.68 to 1.0), respectively (table 4), with only one patient incorrectly classified (figure 3C). As observed for the control versus IBD panel proteins, the relative expression of the 12 proteins identified to separate CD from UC had relative expression in CD CoN biopsies that were between control and CD CoA (see online supplementary figure S7A).

Application and performance evaluation of the panels to an independent validation cohort
As outlined (see online supplementary figure S1), the biomarker panel PLSDA models were independently assessed with proteomic data from the validation cohort. Shown in the ROC curve (figure 2B), panel 1 proteins applied to the classification of the validation cohort result in an AUC of 0.99 (95% CI 0.99 to 1.0) with 47 of 49 (95.9%) patients accurately classified as either control or IBD; the associated prediction overview (figure 2C) is shown. PCA using the five panel 1 proteins shows good separation of the control and IBD CoA populations (figure 2D, see online supplementary figure S6B). Similarly, the 12 proteins in panel 2 differentiate CD CoA from UC CoA with an AUC of 0.86 (95% CI 0.72 to 1.0) (figure 3B, see online supplementary figure S7B), with 24 of 30 (80%) patients accurately classified (figure 3C). The four misidentified patients with CD all have ileocolonic disease (Paris L3), whereas the two patients with UC were Paris E2 and E4. Notably, all patients with limited
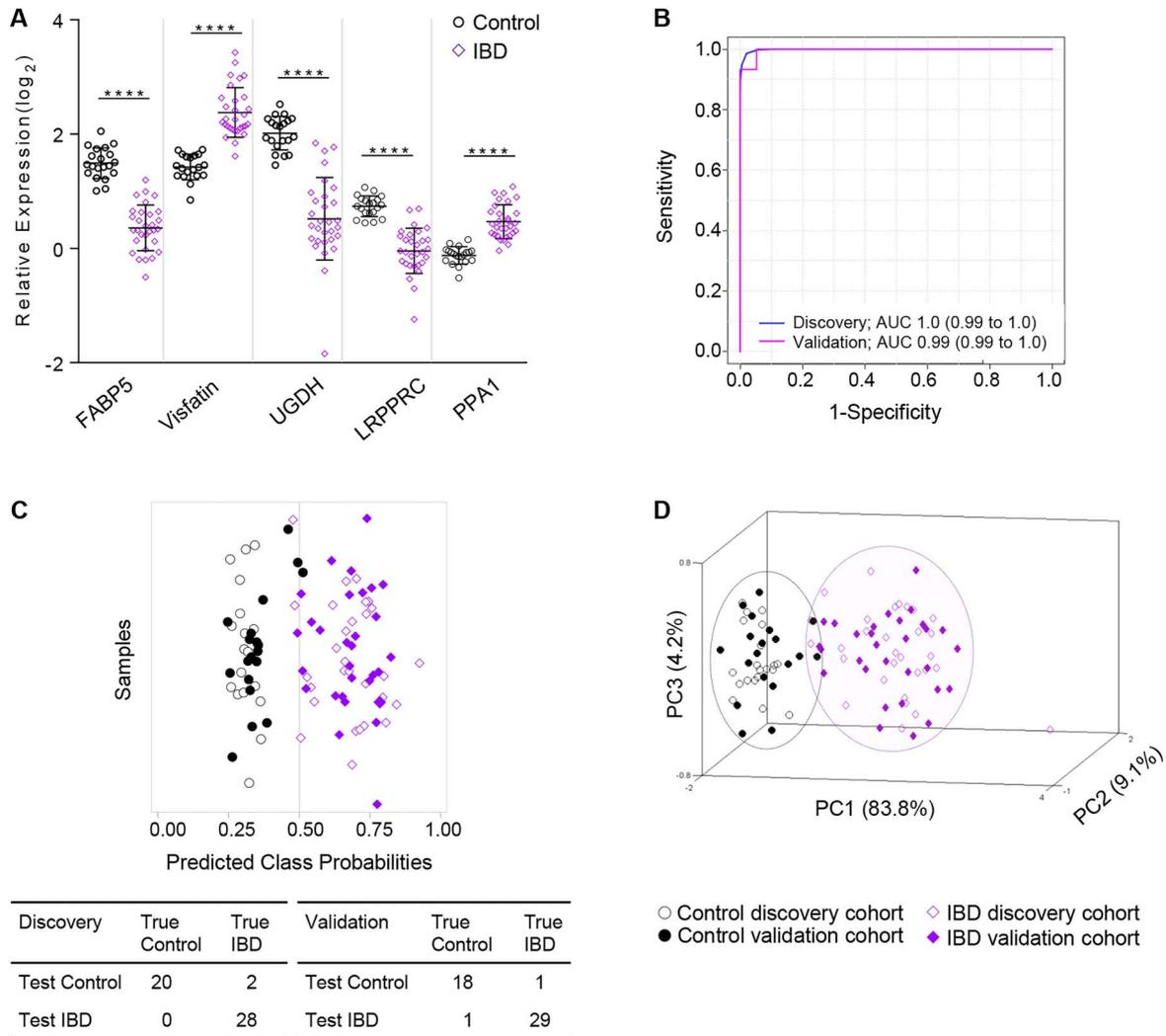
Figure 2 Partial Least Squares Discriminant Analyses (PLSDA) models were trained using data from the discovery cohort and then tested with the data from the validation cohort. They were used to classify control and patients with IBD (inflamed ascending colon biopsy). (A) Relative expression levels of the five candidate biomarkers (panel 1) to separate control from patients with IBD. (B) Panel 1 receiving operating characteristics (ROC) curve of the discovery cohort (blue) and validation cohort (pink) and (C) associated prediction overview for classification using the panel 1 PLSDA model wherein patients to the left of 0.5 would classify as controls and to the right of 0.5 would classify as IBD; true diagnoses of individual patient samples from the discovery and validation cohorts are shown in open or closed symbols, respectively. (D) Principal component analysis using five biomarkers to distinguish IBD (purple) from control (black) population in the discovery and validation cohorts. Statistical significance by Student's t test with ****p<0.0001.

colonic CD (Paris L2) were correctly classified, indicating that despite the common localisation of disease between these patients with CD and UC, the biomarker panel was able to accurately differentiate the patients. PCA performed using the 12 proteins from panel 2 shows good separation of the CD and UC populations (figure 3D). Despite reduced sensitivity and specificity in the validation cohort compared with the discovery group (table 4), these results indicate the utility of the biomarker panels in diagnosis and subdiagnosis of patients with IBD.

Candidate biomarkers are biologically relevant

Pathway analysis was performed to evaluate the functional roles of the 106 IBD and 252 differential diagnostic candidate biomarkers. Most proteins that segregate IBD from control are involved in metabolic processes, and function predominantly in catalysis, specifically oxidoreductase activity (figure 4A, see online supplementary table S6). Canonical pathways identified

to differ in IBD are related to metabolism (figure 4B; see online supplementary tables S6 and S7). In particular, proteins elevated in CD are related to fatty acid metabolism, whereas proteins elevated in UC function in energy and amino acid metabolism (figure 4B).

Correlation with severity

From the 944 Q95+subgroup-specific proteins in the discovery cohort, 118 proteins (12.5%) correlated significantly with PCDAI or PUCAI (figure 5A; see online supplementary table S8). PCDAI was significantly correlated with 83 proteins (see online supplementary table S8), 10% of which are components of the protein ubiquitination pathway. In contrast, 10% of the 43 proteins that correlated with PUCAI were components of the mammalian target of rapamycin signalling pathway. Fifteen of the CD-associated and nine of the UC-associated proteins are regulated by HNF4A, which
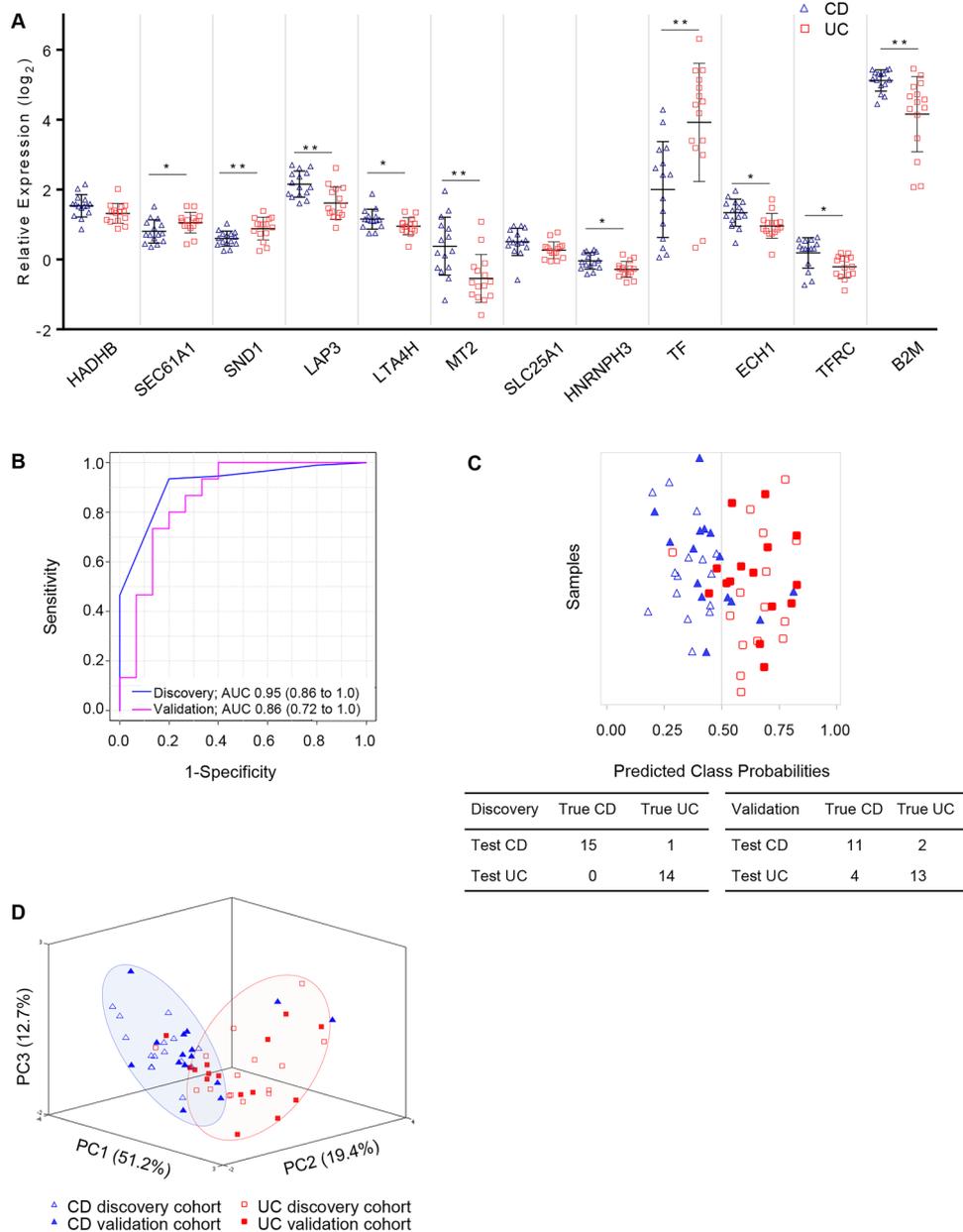
Figure 3 Partial Least Squares Discriminant Analyses (PLSDA) models were trained using data from the discovery cohort and then tested with the data from the validation cohort to classify patients with Crohn's disease (CD) and UC from inflamed ascending colon biopsies. (A) Relative expression levels of the 12 candidate biomarkers (panel 2) to separate patients with CD from those with UC. (B) Panel 2 receiving operating characteristics (ROC) curve of the discovery cohort (blue) and validation cohort (pink) and (C) associated prediction overview for classification using the panel 2 PLSDA model wherein patients to the left of 0.5 would classify as CD and to the right of 0.5 would classify as UC; true diagnoses of individual patient samples from the discovery and validation cohorts are shown in open or closed symbols, respectively. (D) Principal component analysis using 12 biomarkers to distinguish CD (blue) from UC (red) population in the discovery and validation cohorts. Statistical significance by Student's t test with *p<0.05, **p<0.005.

was identified in a paediatric population to be associated with CD[41] and is a UC susceptibility loci.[42] Of the 118 proteins showing correlation with severity, 39 proteins were identified as biomarker candidates, four of which were in the biomarker panels. In panel 1 the relative expression of both visfatin and inorganic phosphatase showed significant correlation with CD severity (figure 5B,C). Similarly, the relative expression of panel 2 protein MT2 correlated with CD severity (figure 5D), whereas heterogeneous nuclear ribonucleoprotein H3 (HNRNP H3) was inversely related to UC severity (figure 5E). A previous study found a correlation between MT2 and grade of inflammation in adult IBD biopsies;[43] the

correlation with disease severity of the other three proteins is a new finding.

ELISAs of visfatin and MT2 are consistent with proteomic data

With the ultimate intent of translating our findings into the clinical setting, the absolute amount of two candidate biomarkers (one from each of the panels) was measured by ELISA from validation cohort patient biopsy samples. The amount of visfatin was within the detection limits for 23 of 24 (95.8%) samples tested. The relative amount of visfatin determined by proteomics in the discovery cohort (figure 1C) is consistent in the

Table 3 Panel 2 candidate protein biomarkers for the segregation of patients with CD from those with UC

Protein	Uniprot	Gene Name	#Unique peptides	Ratio (CD/UC)	AUC
Trifunctional enzyme subunit β, mitochondrial;3-ketoacyl-CoA thiolase	P55084	HADHB	26	1.25	0.809
Protein transport protein Sec61 subunit α isoform 1	B4DR61	SEC61A1	10	0.78	0.803
Staphylococcal nuclease domain-containing protein 1	Q7KZF4	SND1	47	0.75	0.792
Cytosol aminopeptidase	P28838	LAP3	36	1.72	0.781
Leukotriene A-4 hydrolase	P09960	LTA4H	31	1.22	0.778
Metallothionein-2	P02795	MT2A	3	2.51	0.767
Tricarboxylate transport protein, mitochondrial	P53007	SLC25A1	13	1.27	0.766
HNRNP H3	P31942	HNRNP H3	7	1.28	0.750
Serotransferrin	P02787	TF	72	0.15	0.748
Delta(3,5)-Delta (2,4)-dienoyl-CoA isomerase, mitochondrial	Q13011	ECH1	14	1.46	0.746
Transferrin receptor protein 1	P02786	TFRC	30	1.49	0.742
β-2-microglobulin	P61769	B2M	6	2.62	0.740

Rows in dark grey indicate proteins elevated in patients with CD.
CD, Crohn's disease; CoA, inflamed ascending colon; HNRNP H3, heterogeneous nuclear ribonucleoprotein H3; TF, transferrin; TFRC, transferrin receptor.

validation cohort ELISA data (figure 6A), with a significantly higher amount in patients with IBD. Notably, there was no significant difference in the amount of visfatin between CD CoN and CoA biopsies by paired analysis (see online supplementary figure S6C). MT2 was quantified in all validation cohort samples tested by ELISA, and was significantly higher in patients with CD than in those with UC (figure 6B). Consistent with the discovery cohort proteomic data, the ELISA results of the validation cohort showed correlation between the absolute amount of MT2 and the PCDAI in patients with moderate or severe CD (PCDAI>30) (figure 6C). Due to the limited number of patients with mild CD, we could not determine any association with MT2 levels. There was no significant difference in absolute MT2 levels in CoN samples compared with paired CoA samples from patients with CD (see online supplementary figure S7C).

DISCUSSION
The accurate classification of IBD subtype remains a significant clinical challenge, particularly in the paediatric population where clinical features are less distinctive and may change with time.[44 45] Here, we evaluated the proteomes of biopsies taken from 99 paediatric patients at the time of diagnosis, and prior to therapeutic intervention, for the discovery and validation of potential protein biomarkers.

We quantified over 3500 proteins, and identified five proteins that are sufficient to segregate IBD from control patients, and 12 that distinguish CD from UC with an accuracy of >80% (table 4). All candidate biomarkers in our study were quantified in ≥95% of patient biopsies, which is in contrast to markers of disease measuring antibodies to inflammatory or microbial components (eg, perinuclear anti-neutrophil cytoplasmic antibodies (pANCA), anti-Saccharomyces cerevisiae antibody (ASCA), CBir1) that are identified in a limited number of patients (range 2%–85%).[46] The performance of our model (table 4) is greater than the serological panel that is often applied in difficult-to-diagnose cases, despite a sensitivity of 0.67 and specificity of 0.76 on a paediatric cohort.[18]

Current serum and faecal biomarkers for the diagnosis of IBD have limited clinical application due to low selectivity. Standard serological tests can provide information on general inflammation, yet 21% and 54% of paediatric patients with mild CD or UC, respectively, and up to 4% with moderate/severe disease test as normal for haemoglobin, albumin, platelets and erythrocyte sedimentation rate,[47] and show limited improvement even with the addition of C-reactive protein (CRP).[48] CRP and haemoglobin have utility in differentiating IBD from non-inflammatory conditions[21] but do not differentiate IBD from other inflammatory states nor differentiate IBD subtypes.[49 50] Faecal calprotectin performs well to differentiate conditions with mucosal inflammation from those with similar symptoms, yet are non-inflammatory in nature, such as IBS, but has modest selectivity.[51] In our study, biopsy-associated calprotectin levels were not significantly different between groups. This may be reflective of the difference in sample collection for measurement (tissue vs faeces).

The candidate biomarkers identified herein contribute to multiple biological processes, predominantly metabolism (figure 4). Candidate biomarkers in both panel 1 and panel 2 are components of fatty acid metabolism, and the contribution to the pathogenesis of IBD has previously been recognised.[52] In our study, we found decreased FABP5 protein levels in children with IBD compared with control patients (table 2). FABP was previously identified in pooled adult IBD colonic specimens to have higher RNA expression in patients with IBD when compared with controls,[53] whereas a second study identified a reduction in RNA expression of L-FABP in inflamed UC mucosa.[54] Notably, we found proteins involved in fatty acid metabolism to be elevated in the CD over UC paediatric patients (figure 4). Specifically, the proteins leukotriene A-4 hydrolase, tricarboxylate transport protein, trifunctional enzyme and delta(3,5)-delta (2,4)-dienoyl-CoA isomerase were all elevated in patients with CD compared with those with UC. Leukotriene A-4 hydrolase is

Table 4 Characteristics of ROC curves based on panel 1 and panel 2, for the discovery and validation cohorts

Model	AUC	Sensitivity	Specificity	PPV	NPV
Panel 1: Control vs IBD (5 proteins)					
Discovery	1.00 (0.99 to 1.0)	1.0 (0.83 to 1.0)	0.93 (0.78 to 0.99)	0.91	1.0
Validation	0.99 (0.99 to 1.0)	0.95 (0.75 to 1.0)	0.97 (0.83 to 1.0)	0.95	0.97
Panel 2: CD vs UC (12 proteins)					
Discovery	0.95 (0.86 to 1.0)	1.0 (0.78 to 1.0)	0.933 (0.68 to 1.0)	0.94	1.0
Validation	0.86 (0.72 to 1.0)	0.73 (0.45 to 0.92)	0.87 (0.59 to 0.98)	0.85	0.76

CD, Crohn's disease; NPV, negative predictive value; PPV, positive predictive value; ROC, receiving operating characteristics.
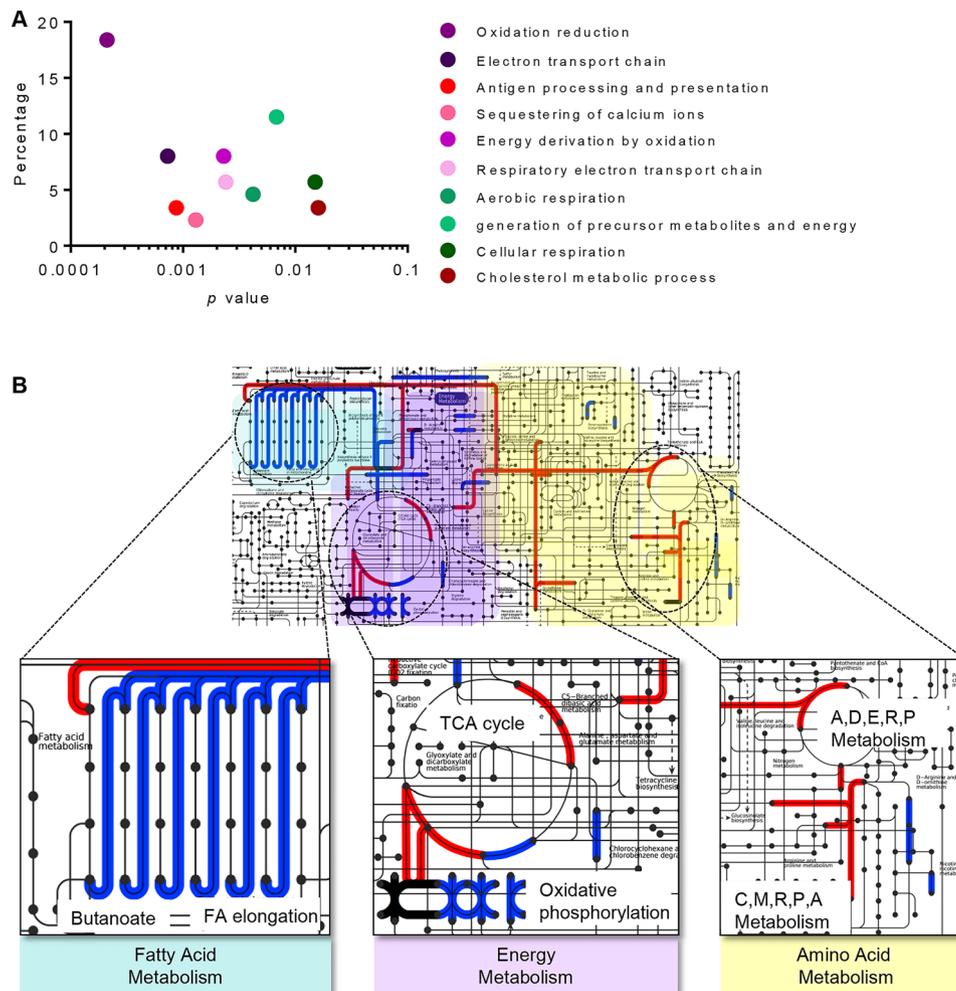
Figure 4 (A) Biological processes of 106 candidate biomarker proteins that contribute to segregation of IBD from control. (B) Metabolic pathways that differ within IBD subtypes; proteins that are upregulated in Crohn's disease (blue lines) are associated with fatty acid metabolism and oxidative phosphorylation, whereas amino acid and energy metabolism are elevated in UC candidate biomarkers (red lines). Pathways with some overlap are shown (dark purple lines). Amino acid metabolism shown with single letter code for alanine (A), aspartic acid (D), glutamic acid (E), arginine (R), proline (P), cysteine (C) and methionine (M).

involved in production of the proinflammatory mediator leukotriene B4, the latter of which was previously found to be elevated in the colon from active patients with IBD.[55]

Energy metabolism was also identified herein to be altered in IBD, which includes the candidate biomarkers inorganic pyrophosphatase, visfatin and UDP-glucose 6-dehydrogenase. Poulsen et al[27] showed an elevation of inorganic pyrophosphatase in inflamed colonic biopsies of patients with UC, concurring with our findings of elevated levels in paediatric IBD biopsies. Waluga et al[56] found elevated levels of visfatin in serum of patients with CD and UC, and reduced in patients with CD following treatment. In our study, we found an elevation of visfatin in biopsies of patients with IBD, and a correlation between visfatin levels in the CD paediatric biopsies and the PCDAI score (figure 5B). While the study by Waluga did not find a correlation between visfatin levels and clinical indices of IBD activity,[56] extracellular levels were found to correlate with inflammatory indices in colorectal tumours.[57] Unlike inorganic pyrophosphatase and visfatin, UDP-glucose 6-dehydrogenase has not previously been associated with IBD. However, the enzyme, which acts on NAD+ downstream of visfatin activity, is a predictive biomarker of colitis-associated cancer risk in patients with both CD and UC.[58]

Nine biomarker panel proteins are not components of metabolic pathways, but have roles in binding and transport of metals, protein and RNA. Cytosol aminopeptidase, found in our study to be elevated in CD when compared with UC (table 3), was previously identified by label-free proteomics to have higher expression in patients with CD compared with those with UC and control patients.[23] Interestingly, serum levels of the protein were not elevated in a cohort of adult patients with UC, but the authors indicated that further testing of colon samples should be performed.[59] The relative expression of metallothioneins in IBD, observed in multiple studies, appears to be dependent on study-specific factors including the location of isolation and therapeutic intervention.[60] In agreement with our observed correlation between relative MT2 levels and PCDAI (figures 5D and 6C), correlation between grade of inflammation with metallothionein in adult IBD biopsies was shown.[43] Further, extracellular antibody blockage of metallothionein resulted in a reduction of colitis in a mouse model,[43] suggesting a contribution of this candidate biomarker to disease pathogenesis. Levels of serotransferrin and the circulating form of transferrin receptor are indicators of anaemia in paediatric patients with IBD.[61] Herein, transferrin receptor protein-1 was elevated in patients with CD compared with those with UC,

Figure 5 (A) Venn diagram of the number of proteins that correlated with severity of disease in patients with Crohn's disease (CD) (blue) or UC (red) as determined by Pearson correlation analysis of the Pediatric Crohn's Disease Activity Index (PCDAI) or Pediatric UC Activity Index (PUCAI) score with the Q95 proteins +subgroup-specific proteins. (B–E) Scatter plots of the relative expression of proteins with significant correlation to severity score that are part of the (B and C) IBD versus control panel, and (D and E) CD versus UC panel are shown.
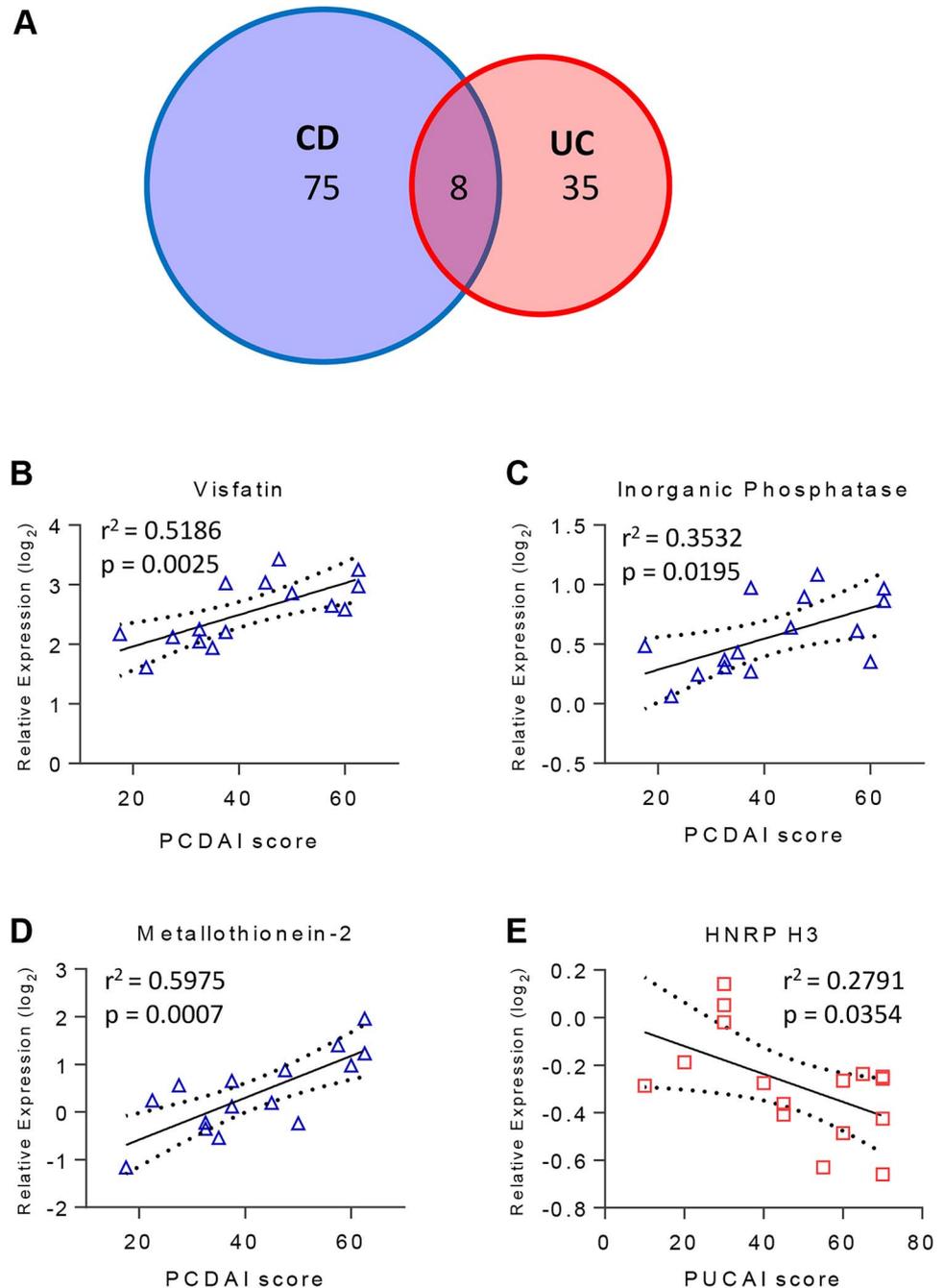
whereas serotransferrin was found to be higher in patients with UC than those with CD (figure 3A). Interestingly, serotransferrin was found to be elevated in adult patients with CD that did not respond to infliximab treatment.[62] Leucine PPR motif-containing protein, Sec61a1, Staphylococcal nuclease domain-containing protein, HNRNP H3 and β-2 microglobulin have not previously been linked with IBD. In ongoing proteomic analysis, further characterisation of the contribution of these and other non-biomarker proteins to the pathogenesis of IBD will be evaluated. While the biomarker study presented herein has been limited to the Q95 (to ensure the identification of broadly applicable biomarkers), in future studies we will assess all proteins that are different from controls or between IBD subtypes at inception, as well as those that change in the same patients following therapeutic intervention.

Overall, we have identified candidate biomarkers from inflamed tissue in samples of initial disease in therapy-naïve patients. The proteins identified in panels 1 and 2 are quantified in ≥95% of the paediatric cohort, indicating the applicability in diagnosis at initial endoscopy. Though several of the panel 1 and 2 proteins have previously been associated with IBD, in this study we show the utility of these proteins together as biomarkers of IBD and, importantly, for the differentiation of CD and UC. The panels were able to classify patient populations with >80% accuracy. At this stage, the identified proteins are reflective of relative changes; further work is required to translate these findings into the clinical setting, specifically to develop methods for absolute quantification of the proteins and development of an appropriate multivariable prediction model for disease classification. The evaluation of visfatin and MT2 by ELISA and validation of the MS data strengthen the possibility of translating this knowledge to the clinic with relative ease. Our IBD and control subjects were those with subacute and chronic complaints. We have not applied these analyses to those
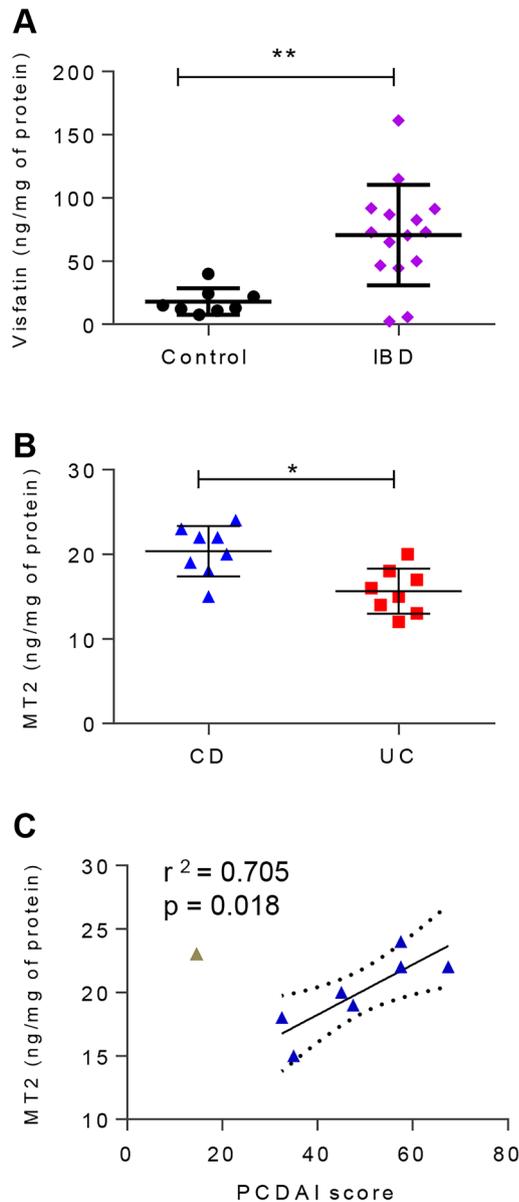
A

B

C

Figure 6 (A) Amount of visfatin per milligram of biopsy protein, determined by enzyme-linked immunosorbent assay (ELISA) in a subset of the validation cohort. (B) Amount of metallothionein-2 (MT2) per milligram of biopsy total protein determined by ELISA in a subset of the validation cohort. (C) Correlation between the ELISA-measured amount of MT2 and the corresponding patient Pediatric Crohn's Disease Activity Index (PCDAI) score in a subset of the Crohn's disease (CD) validation cohort. n=8 each for control, patients with CD and patients with UC.

with acute self-limited infectious colitis or other inflammatory conditions that may be found elsewhere in the intestinal tract such as coeliac disease. However, herein we have identified biomarkers that are able to subdiagnose paediatric patients with subacute to chronic presentations of disease at inception. The use of an inception cohort is an important distinction of this study, and increases the applicability of these findings from studies that have searched for disease biomarkers from patients undergoing treatment. The application of these panels of proteins for use as biomarkers could enable for differential diagnosis of IBD subtypes to permit for appropriate therapeutics.

REFERENCES

1 Ng SC, Bernstein CN, Vatn MH, et al. Geographical variability and environmental risk factors in inflammatory bowel disease. *Gut* 2013;62:630–49.
2 Rocchi A, Benchimol EI, Bernstein CN, et al. Inflammatory bowel disease: a Canadian burden of illness review. *Can J Gastroenterol* 2012;26:811–17.
3 Fengming Y, Jianbing W. Biomarkers of inflammatory bowel disease. *Dis Markers* 2014;2014:710915.
4 Gower-Rousseau C, Savoye G, Colombel JF, et al. Are we improving disease outcomes in IBD? A view from the epidemiology side. *Gut* 2014;63:1529–30.
5 Vernier-Massouille G, Balde M, Salleron J, et al. Natural history of pediatric Crohn's disease: a population-based cohort study. *Gastroenterology* 2008;135:1106–13.
6 Benchimol EI, Manuel DG, Guttmann A, et al. Changing age demographics of inflammatory bowel disease in Ontario, Canada: a population-based cohort study of epidemiology trends. *Inflamm Bowel Dis* 2014;20:1761–9.
7 Griffiths AM. Specificities of inflammatory bowel disease in childhood. *Best Pract Res Clin Gastroenterol* 2004;18:509–23.
8 Sauer CG, Kugathasan S. Pediatric inflammatory bowel disease: highlighting pediatric differences in IBD. *Gastroenterol Clin North Am* 2009;38:611–28.
9 Day AS, Ledder O, Leach ST, et al. Crohn's and colitis in children and adolescents. *World J Gastroenterol* 2012;18:5862–9.
10 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;58:795–806.
11 Day AS, Whitten KE, Lemberg DA, et al. Exclusive enteral feeding as primary therapy for Crohn's disease in Australian children and adolescents: a feasible and effective approach. *J Gastroenterol Hepatol* 2006;21:1609–14.
12 McDonald JW, Tsoulis DJ, Macdonald JK, et al. Methotrexate for induction of remission in refractory Crohn's disease. *Cochrane Database Syst Rev* 2012;12: CD003459.
13 Wang Y, MacDonald JK, Vandermeer B, et al. Methotrexate for maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev* 2015;8:CD007560.
14 Moja L, Danese S, Fiorino G, et al. Systematic review with network meta-analysis: comparative efficacy and safety of budesonide and mesalazine (mesalamine) for Crohn's disease. *Aliment Pharmacol Ther* 2015;41:1055–65.
15 Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011;43:246–52.
16 Franke A, Balschun T, Sina C, et al. Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat Genet* 2010;42:292–4.

17   Jostins L, Ripke S, Weersma RK, *et al*. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–24.

18   Benor S, Russell GH, Silver M, *et al*. Shortcomings of the inflammatory bowel disease Serology 7 panel. *Pediatrics* 2010;125:1230–6.

19   Shaoul R, Sladek M, Turner D, *et al*. Limitations of fecal calprotectin at diagnosis in untreated pediatric Crohn's disease. *Inflamm Bowel Dis* 2012;18:1493–7.

20   Cioffi M, Rosa AD, Serao R, *et al*. Laboratory markers in ulcerative colitis: Current insights and future advances. *World J Gastrointest Pathophysiol* 2015;6:13–22.

21   Alibrahim B, Aljasser MI, Salh B. Fecal calprotectin use in inflammatory bowel disease and beyond: a mini-review. *Can J Gastroenterol Hepatol* 2015;29:157–63.

22   Bennike T, Birkelund S, Stensballe A, *et al*. Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. *World J Gastroenterol* 2014;20:3231–44.

23   Han NY, Choi W, Park JM, *et al*. Label-free quantification for discovering novel biomarkers in the diagnosis and assessment of disease activity in inflammatory bowel disease. *J Dig Dis* 2013;14:166–74.

24   Hsieh SY, Shih TC, Yeh CY, *et al*. Comparative proteomic studies on the pathogenesis of human ulcerative colitis. *Proteomics* 2006;6:5322–31.

25   M'Koma AE, Seeley EH, Washington MK, *et al*. Proteomic profiling of mucosal and submucosal colonic tissues yields protein signatures that differentiate the inflammatory colitides. *Inflamm Bowel Dis* 2011;17:875–83.

26   Nanni P, Parisi D, Roda G, *et al*. Serum protein profiling in patients with inflammatory bowel diseases using selective solid-phase bulk extraction, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and chemometric data analysis. *Rapid Commun Mass Spectrom* 2007;21:4142–8.

27   Poulsen NA, Andersen V, Møller JC, *et al*. Comparative analysis of inflamed and non-inflamed colon biopsies reveals strong proteomic inflammation profile in patients with ulcerative colitis. *BMC Gastroenterol* 2012;12:76.

28   Seeley EH, Washington MK, Caprioli RM, *et al*. Proteomic patterns of colonic mucosal tissues delineate Crohn's colitis and ulcerative colitis. *Proteomics Clin Appl* 2013;7:541–9.

29   Meuwis MA, Fillet M, Geurts P, *et al*. Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. *Biochem Pharmacol* 2007;73:1422–33.

30   Shkoda A, Werner T, Daniel H, *et al*. Differential protein expression profile in the intestinal epithelium from patients with inflammatory bowel disease. *J Proteome Res* 2007;6:1114–25.

31   Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.

32   Bousvaros A, Antonioli DA, Colletti RB, *et al*., North American Society for Pediatric Gastroenterology H, Nutrition, Colitis Foundation of A. Differentiating ulcerative colitis from Crohn disease in children and young adults: report of a working group of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the Crohn's and Colitis Foundation of America. *J Pediatr Gastroenterol Nutr* 2007;44:653–74.

33   Levine A, Griffiths A, Markowitz J, *et al*. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm Bowel Dis* 2011;17:1314–21.

34   Hyams J, Markowitz J, Otley A, *et al*. Evaluation of the pediatric crohn disease activity index: a prospective multicenter experience. *J Pediatr Gastroenterol Nutr* 2005;41:416–21.

35   Turner D, Otley AR, Mack D, *et al*. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. *Gastroenterology* 2007;133:423–32.

36   Gower-Rousseau C, Dauchet L, Vernier-Massouille, *et al*. The natural history of pediatric ulcerative colitis: a population-based cohort study. *Am J Gastroenterol* 2009;104:2080–8.

37   Wiśniewski JR, Zougman A, Nagaraj N, *et al*. Universal sample preparation method for proteome analysis. *Nat Methods* 2009;6:359–62.

38   Motulsky HJ, Brown RE. Detecting outliers when fitting data with nonlinear regression—a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* 2006;7:123.

39   Xia J, Broadhurst DI, Wilson M, *et al*. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* 2013;9:280–99.

40   Ruel J, Ruane D, Mehandru S, *et al*. IBD across the age spectrum: is it the same disease? *Nat Rev Gastroenterol Hepatol* 2014;11:88–98.

41   Marcil V, Sinnett D, Seidman E, *et al*. Association between genetic variants in the HNF4A gene and childhood-onset Crohn's disease. *Genes Immun* 2012;13:556–65.

42   Consortium UIG, Barrett JC, Lee JC, *et al*., UK IBD Genetics Consortium. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* 2009;41:1330–4.

43   Devisscher L, Hindryckx P, Lynes MA, *et al*. Role of metallothioneins as danger signals in the pathogenesis of colitis. *J Pathol* 2014;233:89–100.

44   Alexander F, Sarigol S, DiFiore J, *et al*. Fate of the pouch in 151 pediatric patients after ileal pouch anal anastomosis. *J Pediatr Surg* 2003;38:78–82.

45   Carvalho RS, Abadom V, Dilworth HP, *et al*. Indeterminate colitis: a significant subgroup of pediatric IBD. *Inflamm Bowel Dis* 2006;12:258–62.

46   Iskandar HN, Ciorba MA. Biomarkers in inflammatory bowel disease: current practices and recent advances. *Transl Res* 2012;159:313–25.

47   Mack DR, Langton C, Markowitz J, *et al*. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics* 2007;119:1113–19.

48   Tsampalieros A, Griffiths AM, Barrowman N, *et al*. Use of C-reactive protein in children with newly diagnosed inflammatory bowel disease. *J Pediatr* 2011;159:340–2.

49   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;110:802–19.

50   Murdoch T, O'Donnell S, Silverberg MS, *et al*. Biomarkers as potential treatment targets in inflammatory bowel disease: a systematic review. *Can J Gastroenterol Hepatol* 2015;29:203–8.

51   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;109:637–45.

52   Shores DR, Binion DG, Freeman BA, *et al*. New insights into the role of fatty acids in the pathogenesis and resolution of inflammatory bowel disease. *Inflamm Bowel Dis* 2011;17:2192–204.

53   Heimerl S, Moehle C, Zahn A, *et al*. Alterations in intestinal fatty acid metabolism in inflammatory bowel disease. *Biochim Biophys Acta* 2006;1762:341–50.

54   Okahara S, Arimura Y, Yabana T, *et al*. Inflammatory gene signature in ulcerative colitis with cDNA macroarray analysis. *Aliment Pharmacol Ther* 2005;21:1091–7.

55   Jupp J, Hillier K, Elliott DH, *et al*. Colonic expression of leukotriene-pathway enzymes in inflammatory bowel diseases. *Inflamm Bowel Dis* 2007;13:537–46.

56   Waluga M, Hartleb M, Boryczka G, *et al*. Serum adipokines in inflammatory bowel disease. *World J Gastroenterol* 2014;20:6912–7.

57   Neubauer K, Misa IB, Diakowska D, *et al*. Nampt/PBEF/visfatin upregulation in colorectal tumors, mirrored in normal tissue and whole blood of colorectal cancer patients, is associated with metastasis, hypoxia, il1beta, and anemia. *Biomed Res Int* 2015;2015:523930.

58   Park JM, Han NY, Han YM, *et al*. Predictive proteomic biomarkers for inflammatory bowel disease-associated cancer: where are we now in the era of the next generation proteomics? *World J Gastroenterol* 2014;20:13466–2476.

59   Witts LJ, Perry SW. Leucine aminopeptidase with special reference to ulcerative colitis. *Gut* 1960;1:140–1.

60   Waeytens A, De Vos M, Laukens D. Evidence for a potential role of metallothioneins in inflammatory bowel diseases. *Mediators Inflamm* 2009;2009:729172.

61   Revel-Vilk S, Tamary H, Broide E, *et al*. Serum transferrin receptor in children and adolescents with inflammatory bowel disease. *Eur J Pediatr* 2000;159:585–9.

62   Gazouli M, Anagnostopoulos AK, Papadopoulou A, *et al*. Serum protein profile of Crohn's disease treated with infliximab. *J Crohns Colitis* 2013;7:e461–70.