


ARTICLE

DOI: 10.1038/s41467-018-05357-4

OPEN

Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease

Xu Zhang ¹, Shelley A. Deeke¹, Zhibin Ning¹, Amanda E. Starr¹, James Butcher¹, Jennifer Li¹, Janice Mayne¹, Kai Cheng¹, Bo Liao¹, Leyuan Li¹, Ruth Singleton², David Mack², Alain Stintzi¹ & Daniel Figeys^{1,3,4}

Alterations in gut microbiota have been implicated in the pathogenesis of inflammatory bowel disease (IBD), however factors that mediate the host-microbiota interactions remain largely unknown. Here we collected mucosal-luminal interface samples from a pediatric IBD inception cohort and characterized both the human and microbiota proteins using meta-proteomics. We show that microbial proteins related to oxidative stress responses are upregulated in IBD cases compared to controls. In particular, we demonstrate that the expression of human proteins related to oxidative antimicrobial activities is increased in IBD cases and correlates with the alteration of microbial functions. Additionally, we reveal that many of these human proteins are present and show altered abundance in isolated free extracellular vesicles (EVs). Therefore, our study suggests that the alteration of intestinal EV proteomes is associated with the aberrant host-microbiota interactions in IBD.

¹Ottawa Institute of Systems Biology and Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, Canada. ²Department of Paediatrics, Faculty of Medicine, University of Ottawa and Children's Hospital of Eastern Ontario Inflammatory Bowel Disease Centre and Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada. ³Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie, Ottawa, ON K1N 6N5, Canada. ⁴Canadian Institute for Advanced Research, 661 University Ave, Toronto, ON M5G 1M1, Canada. These authors contributed equally: Xu Zhang, Shelley A. Deeke. Correspondence and requests for materials should be addressed to D.M. (email: dmack@cheo.on.ca) or to A.S. (email: astintzi@uottawa.ca) or to D.F. (email: dfigeys@uottawa.ca)

Inflammatory bowel disease (IBD) is a condition characterized by inflammation of the mucosal lining of the intestinal tract. Crohn's disease (CD) and ulcerative colitis (UC) are the two principal subtypes of IBD. Both subtypes can display similar clinical manifestations and follow similar treatment strategies at times, but they also can vary due to fundamental differences in the diseases¹. It has been estimated that >0.3% of the population in many developed countries suffer from IBD, with the incidence in developing countries on the rise². In particular, the worldwide incidence of pediatric-onset IBD is increasing³. Childhood-onset IBD is distinct from adult-onset IBD with differences observed in progression, anatomical location of disease, and treatment outcomes⁴. Childhood-onset IBD often leads to growth failure and has serious long-term health consequences that result in higher health-care costs⁵. Hence, there is an urgent need to fully understand the pathogenesis of childhood-onset IBD to allow for precision therapeutics and long-term disease management.

Although the exact mechanism underlying the development of IBD remains unclear, its pathogenesis is linked to multiple factors including genetic susceptibility, immune activity, environmental factors (such as diet and life style), and the intestinal microbiome⁴. Accumulating evidence has revealed alterations of the gut microbiota in both adults and children with IBD^{6–8}. However, most studies that characterize the microbiota in IBD patients use stool as proxy for sampling the gut microbiota. It has become increasingly clear that the mucosal-associated microbial dysbiosis in pediatric CD patients is only weakly reflected in fecal samples⁶ and unfortunately limited studies have been able to sample at the site of disease within the gastrointestinal tract^{9,10}. We previously demonstrated an important role of mitochondrial dysfunction and an aberrant host-microbiota crosstalk in the pathogenesis of CD using mucosal-luminal interface (MLI) aspirates⁹. This study highlighted the value of sampling directly at the site of disease in order to elucidate the complex microbiota-host interactions and their implications in IBD. In addition, the study of a pediatric IBD inception cohort also mitigates confounding factors such as medication, lifestyle and disease complications which are associated with the study of adult patients with long-standing IBD.

Deep profiling of the microbiota composition and functions is typically achieved through DNA and RNA sequencing¹¹. However, since gene/transcript presence does not necessarily indicate protein expression, metaproteomics, which directly measures the

expressed proteins, can provide more precise functional information¹². Moreover, mass spectrometry (MS)-based proteome analysis allows simultaneous measurement of proteins of both human and microbiota origin¹³, and hence is a promising approach to aide in deciphering host-microbe interactions in complex intestinal eco-systems. Limited metaproteomic studies of stool and MLI samples from adult IBD patients have been previously reported^{10,14} and there are no metaproteomic studies focusing on pediatric IBD patients. Therefore, in this study, we took advantage of the MLI aspirate sampling approach, high-resolution MS and our previously established MetaPro-IQ bioinformatic workflow¹⁵ to study the proteomes of both microbiota, human and the isolated extracellular vesicles (EVs) present at the MLI of children with new-onset IBD. We characterized the functional activity of microbiome and identified microbial functions and metabolic pathways that were upregulated in IBD compared to control and not previously reported, including those in response to oxidative stress, such as DNA damage repair/defense, mobilome, and L-cysteine degradation. We also showed that the intestinal commensal bacterium *Faecalibacterium prausnitzii* was associated with the CD at strain level. Importantly, we further demonstrated for the first time that the host defense proteins, including those for producing reactive oxidants, were present in the intestinal EVs and correlated with the functional alterations of microbiota in pediatric IBD.

Results

Pediatric IBD cohort and metaproteomic characterization. In this study, we enrolled 71 treatment-naïve pediatric patients, including 25 CD, 22 UC, and 24 non-IBD control subjects that showed no macroscopic-intestinal inflammation (Table 1; details in Supplementary Data 1). No significant difference in age was observed between control, CD or UC ($p = 0.7588$; Kruskal–Wallis test), nor did we observed a significant difference in the number of females and males in each group (control, $p = 0.5413$; CD, $p = 0.2295$ and UC, $p > 0.9999$; Binomial two-tailed test). The majority (81.2%) of pediatric UC participants had extensive disease, defined as a Paris classification of E3 or E4, which is consistent with previous studies¹⁶. Furthermore the majority (80%) of CD patients presented with colonic or ileocolonic disease, also in accordance with previous reports¹⁷.

Table 1 Patient characteristics of treatment-naïve pediatric IBD cohort

	Controls (n = 24)	CD (n = 25)	UC (n = 22)
Gender			
Male/female	10/14	16/9	11/11
Age			
Median (Q1–Q3)	15.0 (13.5–16.3)	13.8 (12.3–16.1)	14.9 (11.2–16.2)
Disease activity			
Severe/moderate/mild	NA	(13/6/6)	(5/14/3)
Paris classification			
A1a/A1b/A2	NA	(4/19/2)	NA
L1/L2/L3/L4a/L4b	NA	(3/4/16/12/4)	NA
B1/B2/B3/p	NA	(23/1/1/4)	NA
G0/G1	NA	(17/8)	NA
E1/E2/E3/E4	NA	NA	(1/3/4/14)
S0/S1	NA	NA	(14/8)
MLI aspirate samples (inflammatory status)			
DeC (w/o inflam/w inflam)	20/0	11/9	17/1
AsC (w/o inflam/w inflam)	22/0	11/11	12/6
TI (w/o inflam/w inflam)	21/0	16/3	16/0

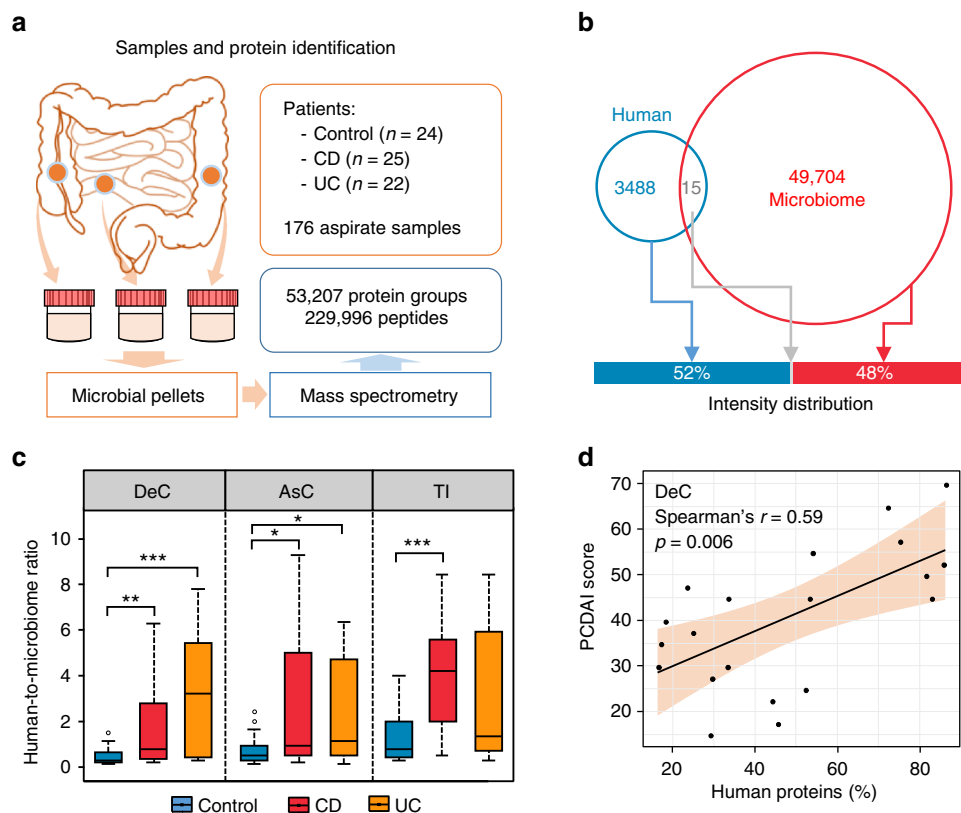


Fig. 1 Metaproteomics approach for assessing both human and microbiota proteins in intestinal aspirate samples. **a** Experimental workflow; **b** Distribution of quantified human- and microbiome-derived protein groups. Venn diagram indicates the numbers of quantified protein groups, while the bar graph shows the intensity of human or microbiome proteins. **c** Human-to-microbiome ratios of total protein intensity. For the box plot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. Whiskers are drawn from the ends of the interquartile range (IQR) to the furthest observations within 1.5 times the IQR range. Outliers >1.5 times the IQR are indicated with circle. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **d** Representative scatter plot between the percentage of human proteins and CD disease severity score

We used metaproteomics to characterize the functionality of intestinal microbiome from 176 MLI aspirate samples collected from the descending colon (DeC), or ascending colon (AsC), or terminal ileum (TI), either with or without inflammation, at the time of diagnostic colonoscopy (Table 1 and Fig. 1a). We quantified 229,996 unique peptides and 53,207 protein groups in these 176 samples, with an average identification rate of $39 \pm 8\%$ (mean \pm SD). On average, $32,882 \pm 8836$ unique peptides and $14,603 \pm 3328$ protein groups were identified per sample. There were 226 that showed biogeographic site-specific expressions (Supplementary Note 1, Supplementary Fig. 1 and Supplementary Data 2). To ensure consistent MS performance, a quality control (QC) sample was generated by combining 10 randomly selected MLI samples; this sample was then repeatedly run in every MS batch analysis. Sample-wise Pearson correlations of the quantified proteins showed that there was no obvious sample clustering based on MS date, and the eight QC samples closely clustered together (with Pearson's r of 0.86 ± 0.02 ; Supplementary Fig. 2). These results indicate reproducible MS measurements and metaproteomic workflow in this study.

Intestinal MLI metaproteome landscape alterations in IBD.

One advantage of metaproteomics is that both microbiota and host proteins can be measured in a single analysis. Among the 53,207 protein groups quantified, 93% were from intestinal microorganisms, with only 3488 (7%) protein groups derived from the host. However, human proteins constituted 52% of the total protein intensities measured in the MLI samples (Fig. 1b),

indicating that the MLI microbiome was associated with abundant host proteins. Our sample preparation workflow is also similar to procedures for isolating shed microvesicles (100–1000 nm), which often employ a 10,000–20,000 g centrifugation step¹⁸, and this could partially explain the high levels of human proteins in our samples. In evaluating the ratios between human and microbial proteins in MLI aspirate samples, we found that the samples from both CD and UC patients had an elevated proportion of human proteins (Fig. 1c). Furthermore, the percentage of human proteins significantly correlated with the clinical disease severity in CD ($p = 0.006, 0.026$, and 0.125 for DeC, AsC, and TI, respectively; Fig. 1d) but not in UC ($p = 0.152, 0.167$, and 0.052 for DeC, AsC, and TI, respectively).

Principal component analysis (PCA) was performed using the MLI human-derived proteins or non-human proteins (i.e. microbiota-derived) quantified in $\geq 50\%$ of the samples. The PCA score plot using the MLI human proteins showed an obvious separation between control and IBD along the first component (PC1), while no obvious difference was observed between CD and UC (Fig. 2a). In addition, the IBD samples emanating from intestinal areas without macroscopic inflammation tend to locate between those of controls and inflamed areas (Fig. 2a and Supplementary Fig. 3). By contrast, the PCA analysis of microbiota-derived proteins showed that the control subjects were clustered together with most of the IBD subjects, and only a subset of IBD samples were shifted from the control. No obvious segregation corresponding to intestinal locations, or the presence of inflammation, was observed (Fig. 2b and Supplementary Fig. 3). Within-group metaproteome variations were evaluated

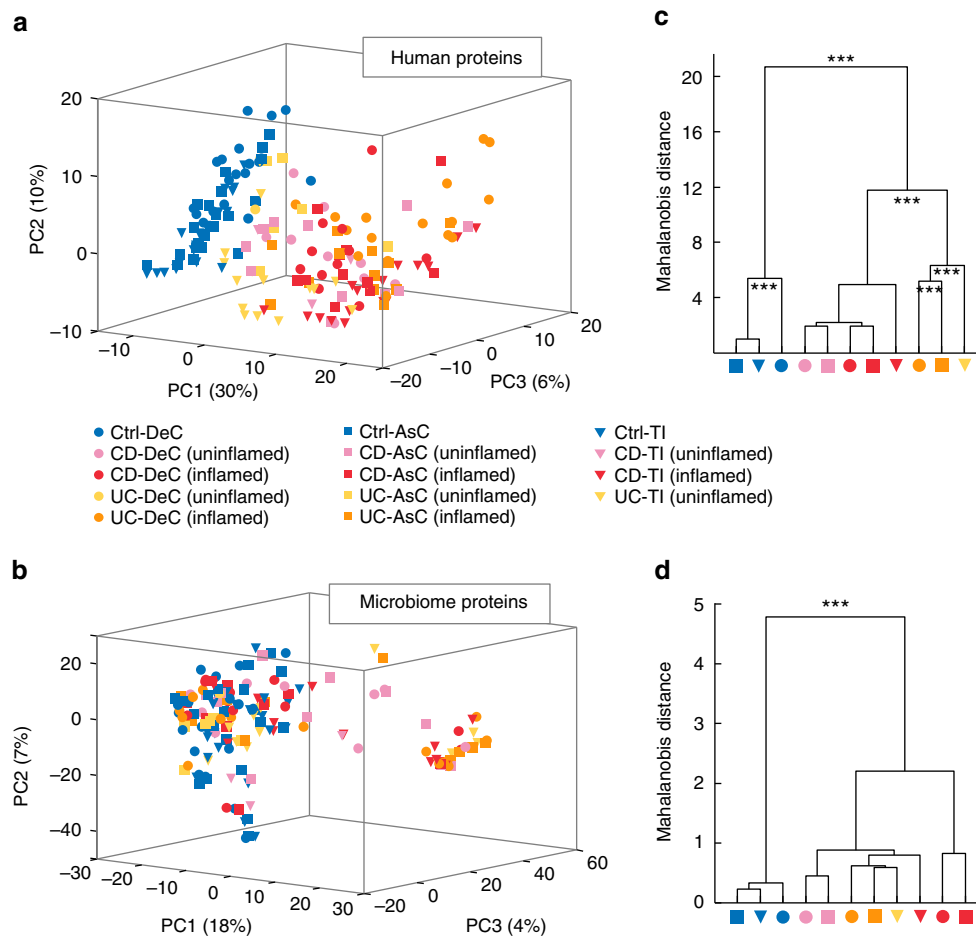


Fig. 2 MLI host and microbiota proteome landscape alterations in pediatric IBD. **a** PCA score plot of MLI human-derived proteins quantified in $\geq 50\%$ of the samples; **b** PCA score plot of MLI microbiome-derived proteins quantified in $\geq 50\%$ of the samples; **c, d** Group mean clustering based on Mahalanobis distances calculated with MANOVA test using the first 10 PCs of PCA on MLI human-derived proteins (**c**) and MLI microbiome-derived proteins (**d**). The clusters are computed by applying the single linkage method to the matrix of Mahalanobis distances between group means. *** $p < 0.001$

using the pairwise Bray-Curtis dissimilarity index, which demonstrated lower within-group heterogeneity in controls as compared to both CD and UC (Supplementary Fig. 4). Power calculations using pairwise-distance matrix and PERMANOVA¹⁹ suggested that we were sufficiently powered ($>99\%$) to detect relevant between-group differences (effect sizes/ P -values were 0.21/0.001, 0.19/0.001, 0.27/0.001 for human-derived proteins, and 0.09/0.001, 0.07/0.001, 0.10/0.001 for microbial proteins, at DeC, AsC, and TI, respectively).

To further study whether significant differences corresponded to disease, presence of inflammation, or intestinal location, we performed multivariate analysis of variance (MANOVA) to compare the multivariate means of subgroups (CD-TI-uninflamed, UC-DeC-uninflamed and UC-AsC-uninflamed were not included due to insufficient sample size). MANOVA analysis of human-derived proteins yielded the greatest differences, which revealed significant differences ($p < 0.001$) for the comparisons of Control vs. IBD, CD, vs. UC, as well as the different intestinal locations in UC and Control groups (Fig. 2c). By contrast, MANOVA analysis of microbiota-derived proteins only yielded a significant difference between IBD and control, and no significant difference was observed between samples emanating from inflamed and uninflamed areas within IBD (Fig. 2d). Therefore, to identify IBD-related host and microbiota alterations, we combined samples from inflamed and uninflamed areas in IBD for further analysis.

Elevated microbial response to oxidative stress in IBD. In order to assess the microbial functions, we annotated all the quantified microbial proteins using the COG database, obtaining 2251 COGs from 24 COG categories. PLS-DA was performed to identify differentially abundant functions in IBD, analyzing the three intestinal locations individually (the goodness of prediction of PLS-DA model, $Q^2 = 0.58$ for DeC, $Q^2 = 0.48$ for AsC, $Q^2 = 0.55$ for TI), which identified 71, 61, and 63 differentially abundant COGs for the DeC, AsC, and TI, respectively (Supplementary Data 3). There were 54 COGs identified with a VIP ≥ 2 in at least two intestinal locations and 42 COGs that were identified in all three locations; all were elevated in IBD compared to control. Functions related to the DNA replication, recombination, and repair (8 COGs in category L), bacterial outer membrane component biogenesis (5 COGs in category M), defense mechanisms (5 COGs in category V) and microbial mobilome (4 COGs in category X) were among the most significantly increased functions identified in IBD (Fig. 3a). We found that DNA damage repair proteins including those involved in base excise repair (BER) and mismatch repair (MMR) pathways, such as uracil-DNA glycosylase (UDG) and DNA MMR ATPase MutS, were increased in microbiome of pediatric IBD patients. In addition, we also revealed the increase of glutathione peroxidase, a microbial defense response protein induced under oxidative stress conditions²⁰, in IBD compared to control (Fig. 3a).

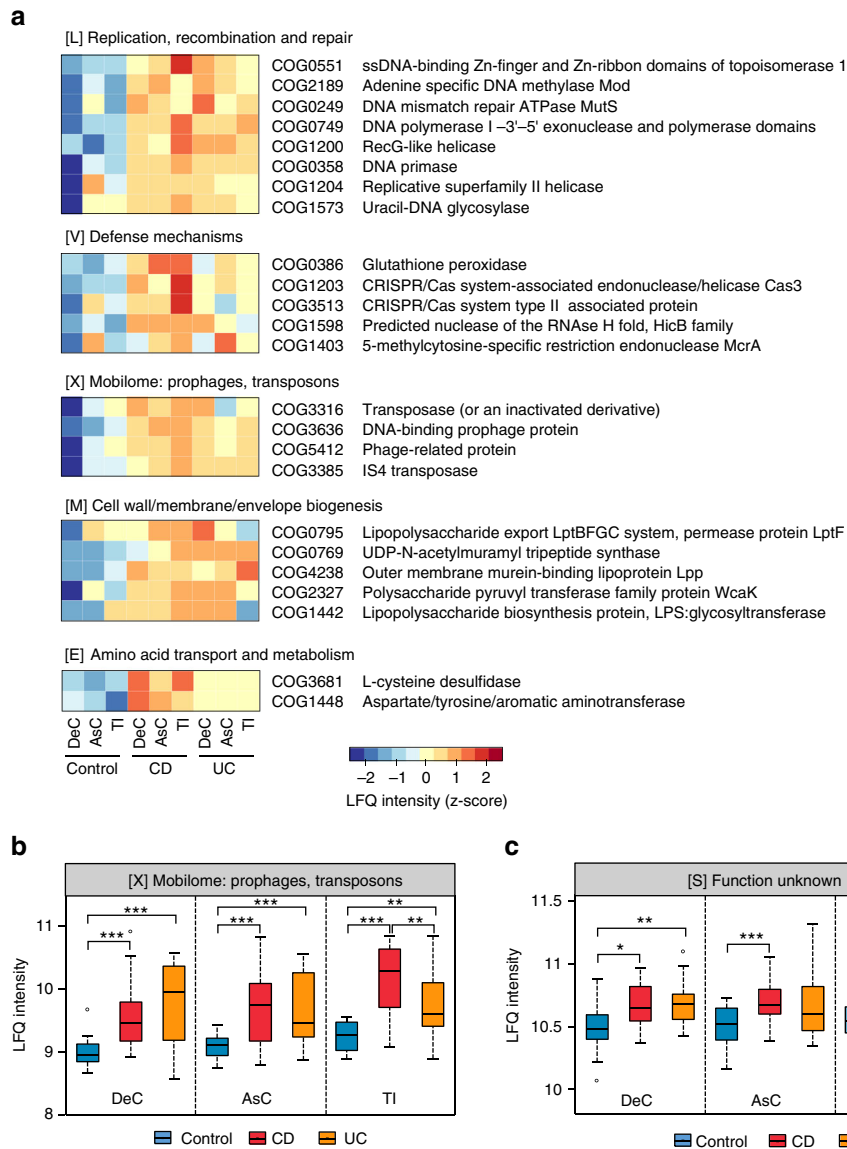


Fig. 3 Functional compositions of MLI microbiome in pediatric IBD patients. **a** Heatmap of differentially abundant COGs in IBD. Representative COG categories are shown and the colors indicate the average LFQ intensity for each subgroup of samples. Each row corresponds to a COG with the COG id and name indicated. **b, c** LFQ intensity of COG category X (mobilome) and category S (unknown function) in MLI microbiome of IBD patients. For the box plot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. Whiskers are drawn from the ends of the IQR to the furthest observations within the 1.5 times the IQR range. Outliers >1.5 times the IQR are indicated with circle. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

At the COG category level, we found that mobilome and category S (function unknown) were significantly increased in IBD compared to controls (Fig. 3b, c), in particular the mobilome displayed significant increase in both CD and UC at all three intestinal regions. Four COGs belonging to mobilome were identified as differentially abundant in IBD microbiome compared to that of controls (Fig. 3a), among which 2 were transposase and the other 2 were phage-related proteins. Furthermore, we found that 2 COGs with functions relating to CRISPR/Cas system, an important prokaryotic immune system against foreign nucleic acids, such as phage genome sequences^{21,22}, were significantly increased in the microbiome of pediatric IBD patients (Fig. 3a).

Altered microbial sulfur and cysteine metabolism in IBD. Sulfate reducing bacteria and the production of hydrogen sulfide

(H₂S) in the gut are known to be increased in IBD patients^{9,23}, however, the pathways underlying the production of H₂S in the microbiome of pediatric IBD patients remain largely unknown. Therefore, we investigated proteins involved in H₂S production, including those for L-cysteine metabolism enzymes as L-cysteine can act a substrate for H₂S production. We found that L-cysteine desulfidase, an enzyme that catalyzes the production of H₂S and ammonia from L-cysteine²⁴, was significantly increased in IBD at both the DeC and AsC (Fig. 3a). KEGG annotation identified 60 KOs corresponding to 912 proteins that belong to sulfur metabolism or cysteine/methionine metabolism pathways. The wilcoxon rank-sum test identified 41 proteins that were significantly different between either CD or UC and controls in at least two intestinal locations, among which there were 18 proteins that decreased and 23 that increased in IBD (Fig. 4a). Within these 41 proteins, 13 are directly involved in the synthesis of L-cysteine, including 8 O-acetylhomoserine (thiol)-lyase proteins (K01740)

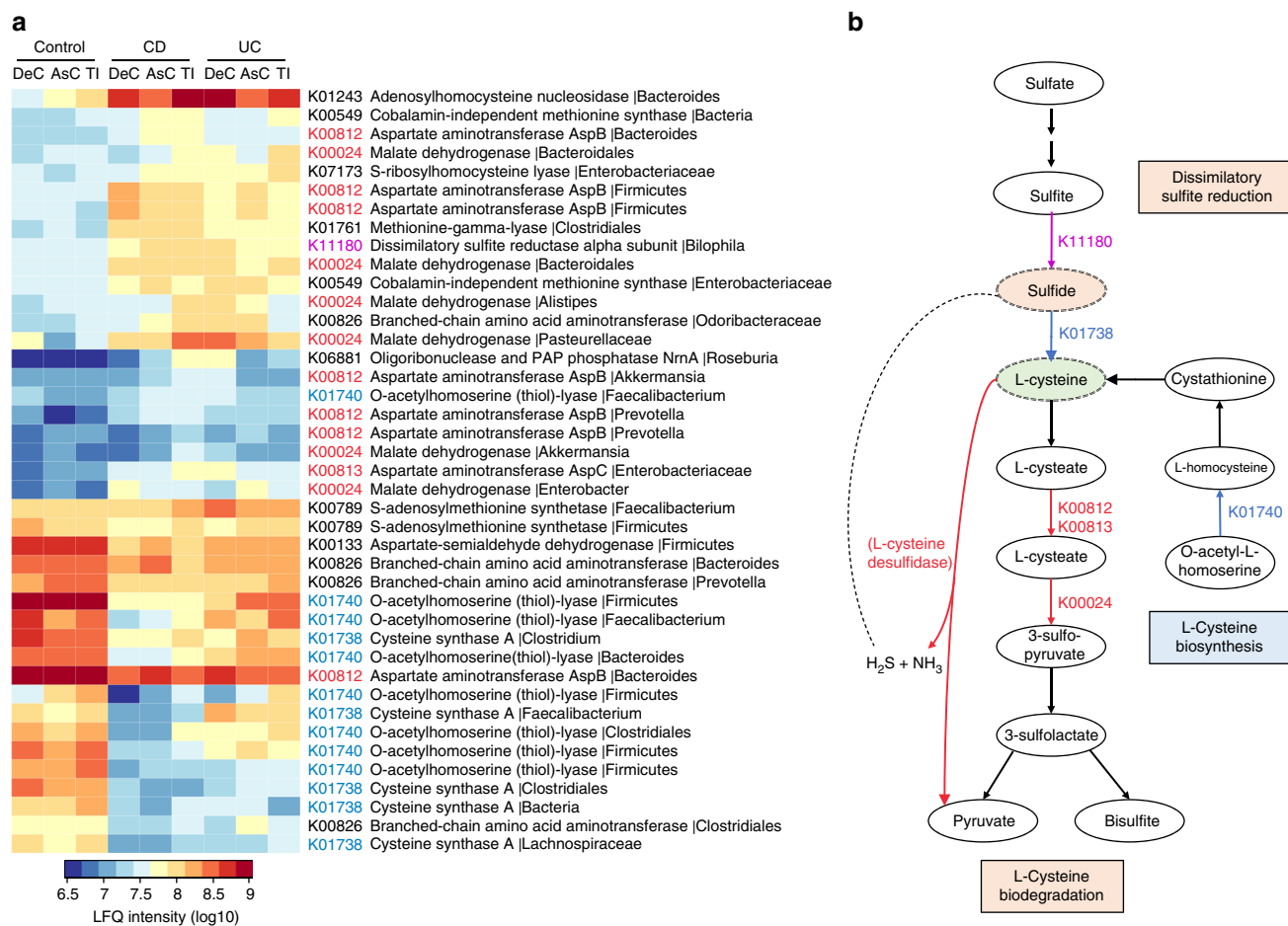


Fig. 4 Alterations of sulfur and cysteine/methionine metabolism pathways in the MLI microbiome of pediatric IBD patients. **a** Taxon-specific sulfur or cysteine/methionine metabolism-related functions that were significantly changed in IBD. Heatmap shows the average LFQ intensity for each subgroup of samples. Each row represents one protein group, and the corresponding KO id, KO name and taxonomic assignment were indicated in the right panel. **b** Pathways relating to H₂S production, L-cysteine biosynthesis, and degradation as represented by the identified differentially abundant proteins. Pathway was adapted from KEGG pathways (map00270 and map00920, version 5/8/2017)

and 5 cysteine synthase A proteins (K01738) (Fig. 4b). We found that 7 out of the 8 O-acetylhomoserine (thiol)-lyase proteins (including one from *Faecalibacterium*) and all 5 cysteine synthase A proteins were decreased in IBD. We also identified 14 microbial proteins involved in the degradation pathways of L-cysteine to generate hydrogen sulfite and pyruvate, including 8 aspartate aminotransferase (AST) proteins (K00812 or K00813) and 6 malate dehydrogenase (MDH) proteins (K00024) (Fig. 4b). All 6 MDH proteins and 7 of 8 AST proteins were increased in IBD. In addition, we found that a major component of dissimilatory sulfate reduction pathway, namely dissimilatory sulfite reductase alpha subunit (DsrA) from *Bilophila*, was significantly increased in the AsC/TI of CD and the DeC of UC patients (Fig. 4a). Moreover, significant correlations of DsrA with disease severity scores were also observed in the AsC/TI of CD and the DeC of UC patients (Supplementary Fig. 5). Given the regulating roles of L-cysteine and H₂S in the mucosal inflammation^{9,25}, our findings suggest possible microbial metabolic pathways that influence the host intestinal physiology and disease progression.

Strain-specific alterations of *F. prausnitzii* in CD. We identified 289 microbial species (189 ± 27 species per sample) with a minimum of 3 distinctive peptides from 4 different kingdoms (Bacteria, Fungi, Archaea, and Virus) (Supplementary Data 4). At

the kingdom level, we found that Fungi was significantly increased in both CD and UC compared to control subjects ($p < 0.01$, Fig. 5a), which is in accordance with previous studies²⁶. In addition, significant correlations were observed between fungal protein abundance and CD disease severity at both the AsC (Spearman's $r = 0.46$, $p = 0.03$; not shown) and TI (Spearman's $r = 0.70$, $p = 0.0008$; Fig. 5b). Linear discriminant analysis effect size (LEfSe)²⁷ analysis identified 4 phyla, 7 classes, 11 orders, 11 families, 16 genera, and 16 species as differentially abundant in IBD compared to control in at least two intestinal regions (Supplementary Data 5 and Supplementary Fig. 6). All four differentially abundant phyla, namely Proteobacteria, Verrucomicrobia, Ascomycota, and Spirochetes, were increased in IBD compared to control. We found that the bacteriophage *Caudovirales* was significantly increased in IBD compared to controls (Fig. 5c), which is also in agreement with a previous virome metagenomics study⁸. The most significantly changed genera in IBD were *Bacteroides* and *Faecalibacterium*, which were decreased and increased, respectively, as compared to controls (Supplementary Fig. 6).

Faecalibacterium prausnitzii, the major species of *Faecalibacterium*, showed higher relative abundance in both UC and CD patients compared to control subjects; in particular, significant differences were observed in UC at the AsC and TI (Fig. 5d). This finding is in agreement with a previous 16s rRNA gene

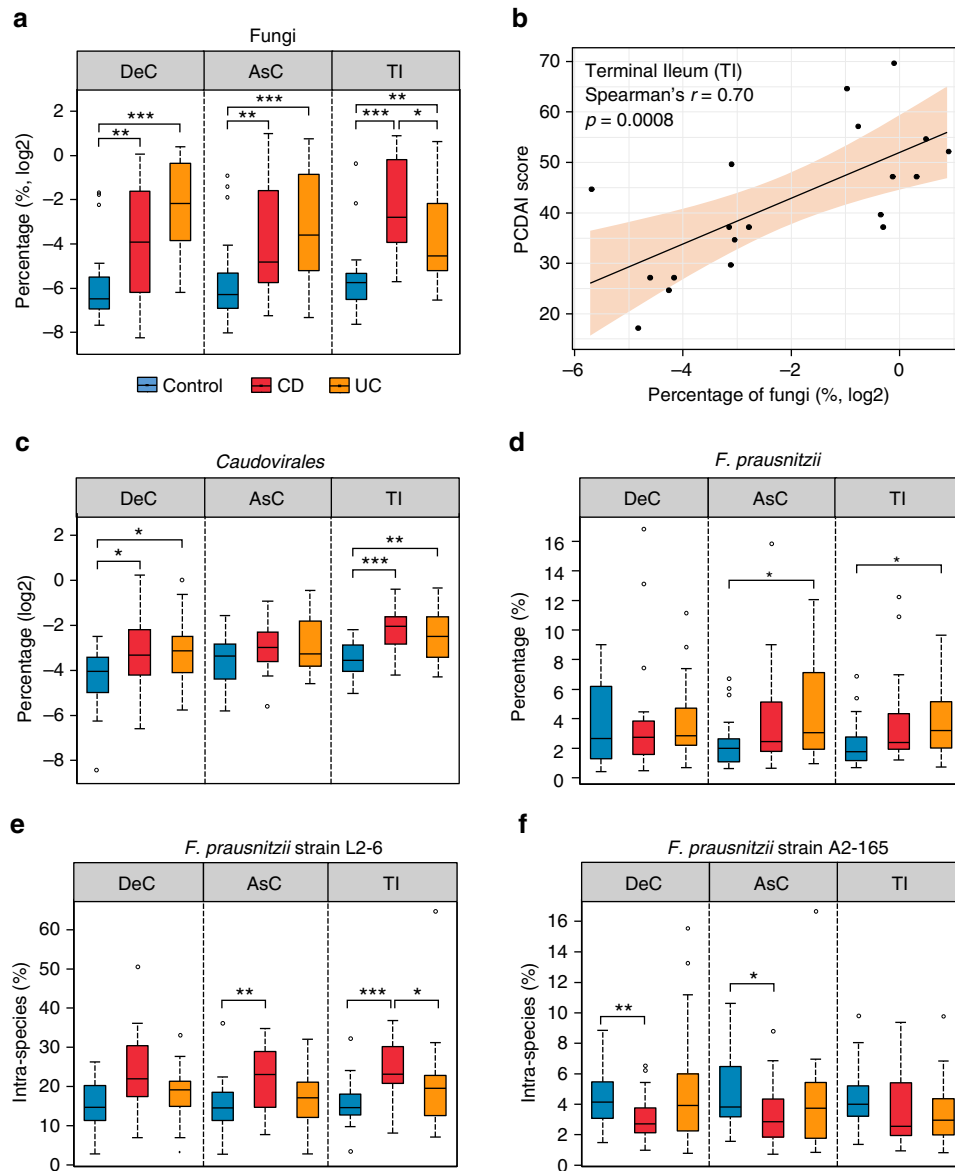


Fig. 5 Alterations of MLI microbiome composition in pediatric IBD patients. **a** Relative abundance of Fungi in MLI microbiome. **b** Representative scatter plot between the fungal abundance and CD disease severity scores. **c** Relative abundance of *Caudovirales* in MLI microbiome. **d** Relative abundance of *F. prausnitzii* in MLI microbiome. **e, f** Intra-species relative abundance of *F. prausnitzii* strain L2-6 and A2-165. For the box plot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. Whiskers are drawn from the ends of the IQR to the furthest observations within the 1.5 times the IQR range. Outliers >1.5 times the IQR are indicated with circle. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

sequencing study using the same sample type (i.e., AsC MLI aspirate) from pediatric IBD patients⁹. Similar findings have also been reported by Hansen et al.²⁸ and Assa et al.²⁹, who studied the microbiota compositions in biopsies collected from pediatric IBD patients using DNA sequencing and real-time PCR. However, controversial associations of *F. prausnitzii* with IBD, regarding their protective roles on gut barrier function, have been commonly reported³⁰. To obtain a deeper understanding of *F. prausnitzii*'s role in IBD, we performed strain-resolved analyses and for the first time demonstrated that L2-6 was the dominant strain in the pediatric MLI samples and only this strain showed higher intra-species abundance in CD (Fig. 5e). Strain A2-165, which is more adept in producing butyrate compared to L2-6³¹, showed significantly lower intra-species abundance in CD compared to control subjects at the DeC and AsC (Fig. 5f). These results are in agreement with a previous study by Song

et al.³¹ that demonstrated a subspecies-level dysbiosis of *F. prausnitzii* in the gut of atopic dermatitis patients with an increase of strain L2-6 and decrease of strain A2-165. In addition, pangenomic analysis of *F. prausnitzii* also reported that strain L2-6 uniquely encoded metabolic functions responsive to oxidative stress, such as *sufBC*³¹, enabling its competitive ability in stressful environment such as the gut of IBD patients. These findings provide evidence that the relationship between IBD and *F. prausnitzii* may be at the strain level instead of species level and its sub-species level dysbiosis may be related to their different competitive growth abilities in response to oxidative stress.

Elevated MLI host defense response proteins in IBD. To identify the MLI human proteins that were differentially abundant in IBD compared to control or CD compared to UC, PLS-

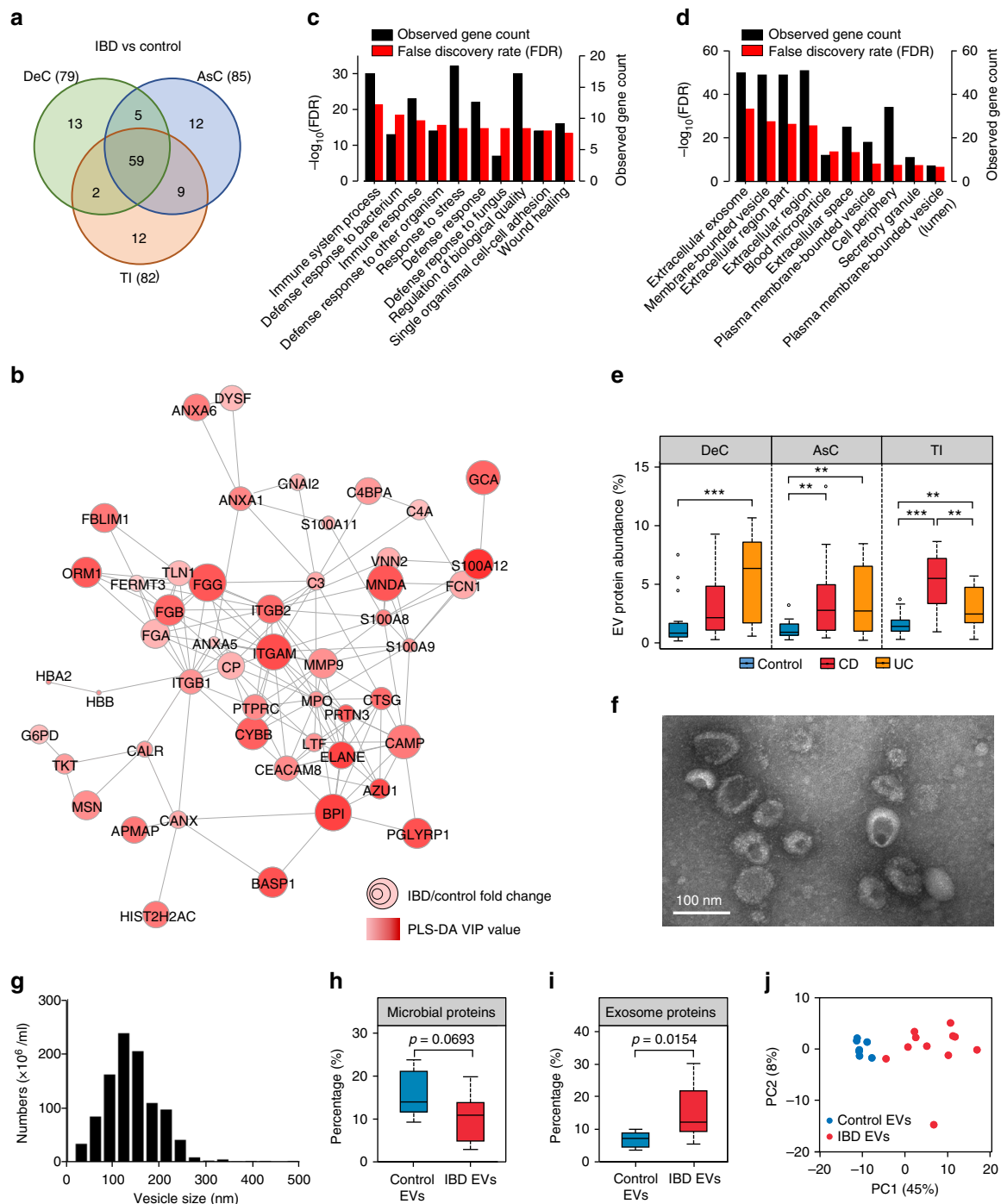


Fig. 6 Alteration of MLI human proteins in pediatric IBD patients. **a** Numbers of differentially abundant proteins identified by PLS-DA with a threshold of VIP ≥ 2 . The numbers in brackets indicate the total number of proteins identified for each intestinal region. **b** Protein-protein interaction network of the 59 differentially abundant MLI human proteins at all three intestinal regions in IBD. **c** Top biological processes and **d** top cellular components enriched by the differentially abundant MLI human proteins in IBD. **e** Percentage of EV related proteins in the MLI human proteomes. For the box plot, the bottom and top of the box are the first and third quartiles, respectively. The middle line represents the sample median. Whiskers are drawn from the ends of the IQR to the furthest observations within the 1.5 times the IQR range. Outliers > 1.5 times the IQR are indicated with circle. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **f** Electron microscopy of isolated vesicles. **g** Nanoparticle tracking analysis of the isolated vesicle size distribution from DeC aspirates of patient HM537. **h** Percentage of microbial proteins in the isolated EV proteomes. **i** Percentage of exosome proteins matching to the top 100 proteins in ExoCarta database. **j** PCA score plot of the free EV human proteomes from patients with or without IBD

DA was conducted for each intestinal region. We identified 79, 85, and 82 IBD-related differentially abundant human proteins for the DeC ($Q^2 = 0.82$), AsC ($Q^2 = 0.83$), and TI ($Q^2 = 0.73$), respectively, with a threshold of VIP ≥ 2 (Fig. 6a and Supplementary Data 6); 70, 71, and 96 disease subtype-related

differentially abundant human proteins were identified for the DeC ($Q^2 = 0.57$), AsC ($Q^2 = 0.48$), and TI ($Q^2 = 0.39$), respectively (Supplementary Data 7). Among the IBD-related differentially abundant proteins, 59 were common for all three intestinal regions, while there were only three disease subtype-related

differentially abundant proteins (C-reactive protein, dual oxidase 2 and complement component 4 binding protein alpha chain) that were common for all regions (Supplementary Fig. 7). Among the 59 IBD-related differentially abundant human proteins, 58 were increased in IBD, with only the chloride anion exchanger (S26A3) decreased. Protein interaction networks of these 59 proteins yielded significant protein interactions (PPI enrichment $p < 0.0001$ as assessed using STRING³²) (Fig. 6b). In the protein network, integrin alpha M (ITGAM; also known as CD11b), myeloperoxidase (MPO) and integrin beta-2 (ITGB2; also known as CD18) displayed the highest degree of interaction, connecting with 17, 17, and 15 proteins in the network, respectively (Fig. 6b), suggesting a “hub” role for these proteins.

Gene ontology enrichment analysis of the 59 differentially abundant human proteins demonstrated that the mostly enriched biological functions include immune and defense responses against bacterium or fungi (Fig. 6c and Supplementary Data 8). In addition, the majority of proteins were related to extracellular region (51/59) and/or extracellular exosome (50/59) (Fig. 6d). Interestingly, among the 3493 human proteins quantified in this study, 2340 proteins matched to the EVpedia database of EVs and 88 matched the top 100 EV proteins (defined by the frequency of identification in previous EV datasets)^{33,34}. The relative abundance of EV proteins that correspond to the top 100 list demonstrated a significant increase in both CD and UC as compared to control, in particular in the TI of CD patients and the DeC of UC patients ($p < 0.001$; Fig. 6e), regions of the intestine which are most often inflamed for their respective subtypes. In addition, we performed transmission electron microscopy (TEM) examination of the pellet fraction of aspirate samples, which confirmed the presence of vesicles consistent in size with that of shed intestinal microvesicles or exosomes (Supplementary Fig. 8a, b).

To gain insights on the host-microbiome interaction, we performed co-occurrence analysis with all identified differentially abundant human proteins and microbial functions, and found that both human proteins and microbial functions showed substantial associations with each other (Supplementary Fig. 9). Microbial uracil-DNA glycosylase (UDG, COG1573) displayed the highest connections ($n = 48$) to human proteins acting as important microbial hub in connecting the host and microbiome nodes. UDG is an enzyme which participates in mutagenesis prevention, as part of the base-excision repair pathway and cleaves uracil, as well as oxidized cytosine derivatives generated under conditions with high oxidative stress³⁵. As mentioned above, 13 out of the 54 significantly increased microbial COGs belonged to category L and V that were related to DNA damage/mismatch responses, indicating that in IBD, microbes are deploying a considerable number of defensive mechanisms against stressors derived from the host.

Pediatric IBD MLI extracellular vesicle proteome changes. To further expand the EV protein findings, we isolated free EVs that were present in MLI supernatants from a cohort of 11 new-onset IBD participants and 7 controls (Supplementary Data 9) and performed the first proteomic characterization of intestinal EVs from pediatric IBD patients. Electron microscopy and nanoparticle tracking analysis (NTA) analysis showed that the isolated vesicles were membrane-enclosed and consistent in size with that of exosomes, shed microvesicles and bacterial outer membrane vesicles (OMV, Fig. 6f, g). No significant difference was observed between the mean sizes of EVs isolated from children with IBD and controls (Supplementary Fig. 8c). Proteomic analysis identified a total of 2744 proteins (1519 human proteins; 1225 microbial proteins). The intensity-based microbial abundance

across all samples was 12.6% and was lower in the EVs isolated from IBD than those isolated from controls (Fig. 6h). Of those microbial peptide sequences which were taxonomically assigned at the phylum level, the majority were derived from Bacteroidetes (70.7%; genus *Bacteroides* and *Prevotella*) followed by Firmicutes (25.6%; *Faecalibacterium*) (Supplementary Fig. 10a). Phylum level comparative analyses indicated that Proteobacteria and Ascomycota were significantly increased in IBD EVs, coinciding with the abovementioned metaproteomic findings (Supplementary Fig. 10b–d). Of the top 100 proteins in ExoCarta, 81 were identified in this dataset and yielded a mean exosomal protein abundance of 12% across all samples with a significant increase observed in IBD (Fig. 6i). This observation may partly be due to the elevated MLI human proteins in IBD (Fig. 1c). We also identified exosomal markers implicated in exosome biogenesis including CD9, CD63, TSG101, ALIX/PDCD61P, FLOT1, FLOT2, Rab27A and the bacterial OMV marker *OmpA* (Supplementary Table 1)³⁶. Moreover, proteins previously deemed as under-represented in EVs³⁶ were either not identified or present at low abundances (Supplementary Table 1).

The PCA analysis of the free EV human proteins displays an obvious segregation between IBD and control subjects (Fig. 6j). A total of 64 proteins were identified as discriminant features that were either identified with PLS-DA ($Q^2 = 0.81$; 37 proteins) or overrepresented in IBD or control (identified in >70% of IBD and <30% control or vice versa; 18 proteins). Comparing the 64 free EV differentially expressed human proteins with the aforementioned 59 differentially abundant MLI microbiome-associated human proteins yielded considerable overlap with 20 proteins identified in both datasets, such as MPO and bactericidal/permeability-increasing protein (BPI) (Supplementary Fig. 8d, e). Interestingly, these 64 EV proteins yielded significant enrichment of biological processes relating to innate immunity and host defense, including phagocytosis, immune response, defense response to fungus, and defense response to bacterium (Supplementary Data 10).

Discussion

We, and others, have shown that aberrant microbiota–host crosstalk plays a central role in the pathogenesis of pediatric CD^{9,37}, however, the mechanisms underlying this complex interplay in disease remains unclear. Through a metaproteomic approach, the current study identified both functional alterations of microbiome and the associated human protein signatures at the intestinal MLI in new-onset pediatric IBD patients, including those that have not been readily reported with metagenomics or metatranscriptomics^{6,26,38}. Moreover, we demonstrated that the host defense proteins present in the intestinal EVs, including oxidative antimicrobial substrates, which would contribute to the excessive oxidative stress in the gut and thus abnormal host-microbiota interactions. A graphic summary of the study design and main findings is shown in Fig. 7. These findings provide new insights into the mechanisms of host–microbiome interaction and its role in the pathogenesis of IBD.

In evaluating the metaproteome of the intestinal MLI aspirate samples, we identified a substantial amount of human proteins. Interestingly a large proportion (67%) of these identified human proteins corresponded to proteins found in the EV protein databases³³. EVs are membrane-bounded, protecting their cargo from proteases and nucleases which enables the delivery of bioactive molecules to neighboring cells, and thus EVs act as important mediators of long distance cell–cell communication and interaction³⁹. In the intestinal lumen, EVs can be derived from multiple sources including intestinal epithelial/immune cells and microbial cells. Timár et al. demonstrated that EVs derived

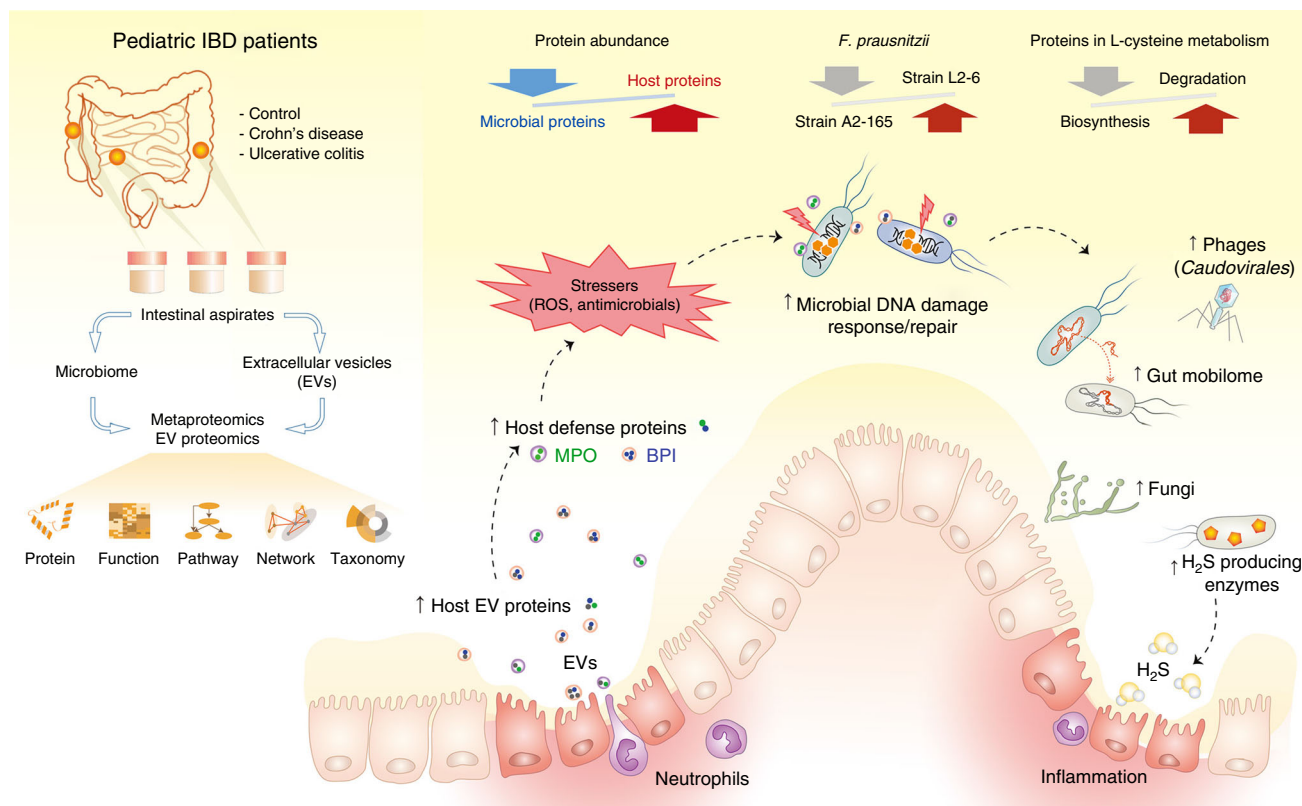


Fig. 7 Graphic summary of the study design and main findings. Intestinal MLI aspirate samples were collected from three different intestinal regions from pediatric IBD patients or controls; the microbiomes and EVs were isolated and characterized using techniques including metaproteomics/proteomics approach. The differentially abundant human/microbial proteins, functions, pathways and microbial taxa were identified using multivariate statistical approaches. The intestinal MLI of IBD patients features an increased relative abundance of human proteins, in particular those EV-associated defense proteins, such as MPO and BPI, and a concomitant microbial dysbiosis at both functional and taxonomic levels

from neutrophils, an important component of intestinal mucosal immunity, had anti-bacterial activities which were dependant on ITGB2 function⁴⁰. Accordingly, we found that ITGB2 (CD18) as well as ITGAM (CD11b, a marker that is consistently expressed on cell surface of neutrophils⁴¹), two subunits of $\alpha\text{M}\beta\text{2}$ integrin, were the most significantly increased MLI microbiome-associated human proteins in IBD and both also were hub proteins in the network of differentially abundant MLI human proteins identified in this study (Fig. 6b). The additional hub protein MPO, which is abundantly expressed in neutrophils and previously reported as a protein cargo of EVs released from neutrophils⁴⁰, was also significantly increased in IBD. This elevated expression of MPO is also in agreement with previous observations from fecal samples^{42–44}. Moreover, the proteomic characterization of isolated free EVs from intestinal aspirate samples also identified MPO as significantly increased in IBD compared to controls (Supplementary Fig. 8e). Given the ability of MPO to produce reactive oxidants such as hypohalous acids⁴⁵, the host cell-derived EVs that carry increased amount of oxidative antimicrobials, such as MPO, may contribute to the elevated level of oxidative stress against microbes in the intestine of IBD patients.

Neutrophil infiltration of the intestinal mucosa is a common phenomenon in IBD. Li et al. have performed a metaproteomic characterization of the MLI proteins in adult CD and UC patients and found that the neutrophil-derived products such as human neutrophil peptide (HNP) 1–3 were significantly increased in both the MLI and biopsy samples¹⁰. In addition, other anti-microbial proteins derived from epithelium cells such as α -defensin 5, β -defensin 1 and 2, as well as transferrin and hepcidin that are involved in iron homeostasis, were among the most

significantly increased proteins in the MLI adult patients¹⁰. In the current pediatric study, we also identified HNP3 and lactotransferrin (LTF) as increased in IBD compared to controls, however, we did not identify the other epithelium cell-derived ones. Instead, we identified neutrophil-derived proteins as the most significantly different MLI human proteins in IBD compared to control. This difference might be due to differences in the depth of metaproteomic measurement or sample processing procedures since we performed differential centrifugation to enrich the microbial cells and remove potential shedding intestinal cells or tissue debris. Therefore, human proteins directly binding to the microbial cells or carried by the human-derived vesicles were preferentially enriched and more readily identified in our study.

In addition to the MLI microbiome-associated human proteins, we also revealed marked functional alterations of microbiota in IBD patients. In particular, the microbial functions related to DNA damage repair and the microbial defense response were among the most significantly increased functions in IBD (Fig. 3a), which is in accordance with the observations of elevated host EV proteins, such as reactive oxidant-producing MPO, since reactive oxidants are known to mediate DNA damage⁴⁶. Interestingly, we showed that the abundance of the UDG protein, an enzyme which cleaves uracils and repairs oxidative stress related DNA damage such as oxidized cytosine derivatives³⁵, correlated to most of the host response proteins and aberrant microbial functions (Supplementary Fig. 9). These findings indicate that the elevated oxidative stress in the intestine of IBD patients might be a crucial driver for the aberrant host-microbiota interactions.

In addition, we found that the expression level of proteins related to transposase and prophage, namely mobilome, was increased in the microbiome of IBD patients compared to controls. The mobilome has been shown as a key factor in affecting the microbiome dynamics or adaptation in response to intestinal environmental changes⁴⁷ and Diard et al. reported that the presence of inflammation and oxidative stress in the intestine can promote the induction of prophage lysogenic conversion⁴⁸. Accordingly, we found a significant increase of proteins uniquely belonging to *Caudovirales* phages (Fig. 5c), in IBD compared to control, which is also in agreement with a previous virome metagenomics study⁸. In Prokaryotes, the CRISPR/Cas system acts as an important immune system against foreign nucleic acids, such as phage genome sequences^{21,22}. Interestingly, we identified two CRISPR/Cas system related functions that were significantly increased in both CD and UC (Fig. 3a), further providing evidence for the bloom of bacteriophage in the microbiome of IBD patients. Given that these mobile gene elements (MGEs) are usually associated with the horizontal gene transfer (HGT) of antibiotic resistance determinants^{49,50}, the findings in the current study may also suggest an increased susceptibility to antibiotic resistance of gut microbiota in patients with IBD.

We previously demonstrated by 16s rRNA gene sequencing an increase of H₂S producers and decrease of butyrate-producers in pediatric IBD, which was associated with the down-regulation of mitochondrial H₂S detoxifying proteins⁹. Here, we further revealed that microbial proteins with functions of L-cysteine desulfidase and dissimilatory sulfite reductase, two important pathways for H₂S production, were increased in the MLI microbiome from pediatric IBD patients. In addition to H₂S production, we found a significant increase of L-cysteine degradation (including the enzymes for H₂S production) and decrease of L-cysteine biosynthesis in the intestinal microbiome of IBD patients. L-cysteine plays a vital role in regulating cellular redox potential through the synthesis of anti-oxidant glutathione (GSH)⁵¹. L-cysteine has also been shown to play important roles in maintaining the gut barrier functions^{25,52}. Although it's still unknown whether the alteration of L-cysteine metabolism is a cause or consequence of the elevated oxidative stress, our findings suggest that the dysregulated gut microbial L-cysteine utilization together with the elevated intestinal oxidative stressors may influence the development of mucosal inflammation in IBD.

In summary, we performed a large-scale gut metaproteomic study of a pediatric IBD inception cohort to assess both the intestinal microbiome and its associated human proteins. We characterized the functional activity of the MLI microbiota at the early stage of IBD and identified new host-microbiome signatures of pediatric IBD at the MLI, including the alterations of microbial mobilome and L-cysteine metabolism, both of which are indicative of elevated oxidative stress. In addition, we performed the first proteomic characterization of intestinal EVs from children with new-onset IBD and demonstrated that the host defense proteins present in the isolated EV samples, in particular the reactive oxidant-producing enzymes, may contribute to the increased oxidative stress in the intestine. The latter may lead to microbial defense responses and functional adaptations, which may initiate the gut microbial dysbiosis and in turn contribute to the development of mucosal inflammation. This study thereby provides valuable protein-level information for aiding in understanding the host-microbiome interactions underlying the development of IBD.

Methods

Subjects and MLI aspirate sample collections. Seventy-one pediatric patients (<18 years old) were included in this study. All eligible patients were undergoing clinically indicated endoscopic examination following their consultation to General

Gastroenterology clinic for suspected IBD. None of the participants had ever been on any empiric IBD treatments. The most common reasons for colonoscopy were non-acute abdominal pain, reports of rectal bleeding, non-acute diarrhea and weight loss. Non-IBD controls had visually normal mucosa, histologically normal mucosal biopsies and normal imaging to definitively rule out IBD. The diagnosis of IBD was performed according to standard criteria following thorough clinical, endoscopic, histologic and radiological evaluation⁵³. In total, 25 patients were diagnosed with CD, 22 patients were diagnosed with UC, and 24 patients were deemed non-IBD controls. The study and clinical sample collection protocol was approved by the Research Ethics Board of the Children's Hospital of Eastern Ontario (CHEO), Ottawa, ON, Canada. All patients were newly diagnosed and naïve to treatment. Informed consent was obtained from all subjects. The following exclusion criteria were implemented: (1) presence of diabetes mellitus; (2) presence of infectious gastroenteritis within 2 months; (3) use of any antibiotics or probiotics within 4 weeks, or (4) irritable bowel syndrome. Clinical disease activity of CD was determined using the PCDAI (Pediatric Crohn's Disease Activity Index)⁵⁴ and of UC using the PUCAI (Pediatric Ulcerative Colitis Activity Index)⁵⁵.

Intestinal MLI aspirate samples were collected from the DeC, AsC and/or TI during diagnostic colonoscopy for each participant. During colonoscopy, any loose fluid and debris was first aspirated and discarded. Thereafter, sterile water was flushed onto the mucosa to dislodge the mucus layer from the mucosal epithelial cells and the resulting mixture was aspirated into a container, which was immediately put on ice and transferred to the laboratory for processing.

MLI aspirate sample processing for metaproteomic analysis. Aspirate samples were first centrifuged at 700 g, 4 °C for 5 min and the supernatant was transferred into a new tube for another centrifugation at 14,000 g, 4 °C for 20 min. This pellet fraction was then harvested for metaproteomics analysis. Protein extraction and trypsin digestion were performed according to our previously published procedure¹⁵. Briefly, proteins were extracted with protein lysis buffer containing 4 % (w/v) sodium dodecyl sulfate (SDS) and 6 M urea in 50 mM Tris-HCl buffer (pH 8.0). The protein lysates were then subjected to three ultrasonication (30 s each with 1 min interval on ice) using a Q125 Sonicator (Qsonica, LLC) with an amplitude of 25%. Remaining cell debris was removed through high-speed centrifugation at 16,000 g, 4 °C for 10 min. The resulting supernatant was then precipitated using 5-fold volume acidified acetone/ethanol buffer at -20 °C overnight. Proteins were pelleted by centrifugation at 16,000 g for 20 min at 4 °C, and washed with ice-cold acetone three times. The protein pellets were then re-suspended in 6 M urea, 50 mM ammonium bicarbonate buffer. For in-solution trypsin digestion, 50 µg of proteins for each sample were reduced and alkylated with 10 mM dithiothreitol and 20 mM iodoacetamide, respectively. One microgram of trypsin (Worthington Biochemical Corp., Lakewood, NJ) was then added for digestion at 37 °C overnight with agitation. The tryptic peptides were desalted with a 10-µm C18 column and the tryptic peptides were analyzed on a Q Exactive mass spectrometer (ThermoFisher Scientific Inc.). Briefly, peptides equivalent to 2 µg proteins were loaded and separated on an analytical column (75 µm × 50 cm) packed with reverse phase beads (1.9 µm; 120-Å pore size; Dr. Maisch GmbH, Ammerbuch, Germany) with 4 h gradient from 5 to 35% acetonitrile (v/v) at a flow rate of 200 nL/min. The instrument method consisted of one full MS scan from 300 to 1800 m/z followed by data-dependent MS/MS scan of the 12 most intense ions, a dynamic exclusion repeat count of 2, and repeat exclusion duration of 30 s. The MS data were exported as RAW format, which were then processed with MetaPro-IQ bioinformatics workflow for peptide/protein identification and quantification¹⁵.

To ensure better coverage of the reference protein database, we generated an integrated human gut microbiome protein database, containing the proteins from (1) integrated gene catalog database (IGC) of human gut microbiome (containing 9,879,896 proteins)⁵⁶; (2) NCBI reference sequence database of Virus (containing 275,423 proteins; downloaded at October 17, 2016 from <http://ftp.ncbi.nlm.nih.gov/refseq/release/viral/>); (3) 121,347 fungal protein sequences of 7 species that have been previously reported to present in pediatric IBD patients^{26,57}, namely the proteomes of *Candida albicans*, *Candida parapsilosis*, *Cladosporium cladosporioides*, *Clavispora lusitanae*, *Cyberlindnera jadinii*, *Kluyveromyces marxianus*, *Saccharomyces cerevisiae* (the proteome databases were downloaded at November 24, 2016 from UniprotKB: <http://www.uniprot.org/proteomes/>); (4) an in-house MLI gut microbial gene catalog database (containing 1,088,113 predicted gene sequences), generated from shotgun metagenomics sequencing for 28 MLI samples of a pediatric IBD sub-cohort (12 control, 7 CD and 9 UC) and with an approach described previously¹⁵. The two gene catalog databases were combined and the redundant sequences were removed using CD-HIT (-c 0.95 -M 0 -G 0 -a 5 0.9 -g 1 -d 0 -T 0)⁵⁸. Among the 1,088,113 predicted MLI genes, 75,425 genes (6.9%) were absent in the IGC database which therefore provides additional sequence information for microbial protein and peptide identification from MS data.

We used an iterative database search strategy, namely MetaPro-IQ, that has been implemented as an integrated analysis platform (MetaLab)^{15,59}, to increase the sensitivity of the large database search. Iterative database search strategies have previously been shown to improve the sensitivity of peptide identification using large databases and enable comparable performance to a matched metagenome database search strategy^{15,60}. Briefly, each of the raw files was searched against the above generated integrated microbiome protein database. All matched protein sequences were extracted from each run and combined to generate a refined and

size-reduced database. This refined database was then concatenated with the human protein database (downloaded from UniprotKB on 24 December 2016) for a target-decoy database search (at false discovery rate of 0.01) and label-free quantification using the MaxLFQ algorithm⁶¹. Both razor and unique peptides were used for protein quantification, and the minimum ratio count was set as 1. An alignment retention time window of 20 min and match time window of 5 min were applied to match the same accurate masses between different runs. Proteins identified by the same set or a subset of peptides were grouped together as one protein group. A protein group was considered as human or microbiome protein group only when all the majority proteins were from human or microbiome proteome database, respectively. Similarly, a peptide sequence was considered as being originated from human or microbiome only when all the matched proteins were from human or microbiome proteome database, respectively.

Free EV isolation, protein extraction and proteomics analysis. Debris and microbial cells in the MLI aspirate samples were depleted as described above. The resulting debris- and bacteria-depleted supernatant was filtered through a 0.45 µm syringe driven filter and the resulting supernatant was stored at -80 °C. EV isolation was performed according to the protocol used by Mitsuhashi et al. with modifications⁶², although it should be noted that contaminating soluble proteins may be co-isolated with the EVs. Briefly, the supernatant was subjected to ultracentrifugation at 100,000 g for 70 min at 4 °C. The EV pellet was washed with PBS buffer and subjected to another round of ultracentrifugation at 100,000 g for 70 min. A portion of the resuspended pellet was lysed in 4% (w/v) SDS, 50 mM Tris-HCl (pH 7.8). A total of 25 µg of proteins was digested by filter-aided sample preparation (FASP) procedure⁶³. The resulting tryptic peptides were desalted, analyzed, and data processed as described above.

Characterization of isolated free EVs. Size distribution of the isolated EVs was examined with NTA using the ZetaView nanoparticle tracking analyzer (Particle Metrix GmbH, Germany) using a shutter value of 40 and sensitivity of 85. Prior to each session, the size of standard silica beads was measured. The EV preparations were diluted 1:25,000–1: 250,000 in PBS to yield a concentration consistent with the accuracy range of the instrument. Measurement data obtained from the ZetaView output was analyzed in Excel and plotted in Prism 7.0; EV sizes were binned by 30 nm.

The morphology of EVs was characterized with TEM. Briefly, the differential centrifugation pellets (14,000 g and 100,000 g) were firstly fixed in 2.5% glutaraldehyde in a 0.1 M sodium cacodylate buffer. For negative staining, the pellets were suspended and a drop of suspension was added on to a formvar grid. When the suspension had partly dried, the grid was washed by adding a drop of distilled water and touching it three times to the surface of water droplet. Excess water was removed by touching the grid to a filter paper. After drying, a drop of 2% uranyl acetate stain was added to the grid and incubated for 10 s before removing the excess stain and drying at room temperature for imaging. For sectioning, the glutaraldehyde fixative was removed from the pellet and replaced with 2% OsO₄ in 0.1 M sodium cacodylate buffer. The OsO₄ fixative was replaced with 0.1 M sodium cacodylate before dehydrating in an increasing series of alcohol. The most concentrated alcohol was replaced with acetone. The material was penetrated by a growing series of Araldite diluted in acetone, and finally embedded in Araldite (Huntsman Advanced Materials LLC, United States). Ultrathin sections (80 nm) were prepared using an Ultracut Leica UC6 ultramicrotome (Leica Microsystems, Germany) and placed onto a copper grids coated with formvar film. Sections were stained with uranyl acetate and lead citrate solutions and examined with a transmission electron microscope (JEOL JEM 1230, Japan).

Microbial taxonomic analysis. All identified peptide sequences were subjected for taxonomic analysis using lowest common ancestor (LCA) algorithm implemented in MetaLab⁵⁹. The calculation of LCA was performed according to the principles implemented in Unipept with modifications⁶⁴. In the current study, NCBI taxonomic nodes containing words “uncultured”, “unspecified”, and “undetermined” were excluded, however species nodes with numbers in their name were kept for the LCA calculation. Two missed cleavages of peptides were allowed, which is equivalent to “Advanced missed cleavage handling” options in Unipept web application⁶⁵. To ensure confident taxonomic identification, taxa with ≥ 3 distinctive peptides were used further analysis. The sum intensity of all distinctive peptides for a taxon was used for calculating relative abundance of the taxon within a specific rank level. Strain-resolved analysis of *Faecalibacterium prausnitzii* was performed as previously described¹³. Briefly, strain-distinctive peptides of *Faecalibacterium prausnitzii* were identified by comparing *F. prausnitzii* species distinctive peptides with all the in silico tryptic peptides (with maximum of 2 missed cleavages) of *F. prausnitzii* strain databases available in UniprotKB proteome database (downloaded on October 25, 2016). A total of five strain proteomes of *F. prausnitzii* were obtained, including UP000006028 for strain KLE1255, UP000004619 for strain A2–165, UP000007059 for strain SL3/3, UP000008804 for strain L2–6, and UP000005945 for strain M21/2. Intra-species abundance of each strain was then calculated as the percentage of distinctive peptide intensity out of the total intensity of all strain-distinctive peptides.

Microbial function and pathway analysis. All quantified microbial protein sequences were annotated with the COG database (version 2014) as previously described¹⁵. KEGG ortholog (KO) annotation of protein sequences was conducted with the GhostKOALA web application⁶⁶. Taxonomy assignment of the proteins were performed using MEGAN 6⁶⁷. The taxonomy of a protein group was assigned with the lowest common ancestor (LCA) of all proteins within that protein group.

Protein–protein interaction and pathway enrichment analysis. Protein–protein interaction (PPI) and pathway enrichment analysis were performed with STRING (version 10.5)³². The protein interaction networks were exported from STRING and visualized using Cytoscape software (version 3.4.0).

Statistical and multivariate data analysis. The LFQ intensity of quantified protein group was log₁₀-transformed and used for statistical analysis. Univariate statistical difference between each pair of groups was assessed by Wilcoxon rank-sum test, except when indicated otherwise. For multivariate statistical analysis, only those protein groups with valid LFQ intensity values in ≥ 50% of the samples were used. Missing values were imputed with the K-nearest neighbors method. PCA was performed in MATLAB (The MathWorks Inc.). Hierarchical clustering analysis and heatmap plotting was performed with R (<https://www.r-project.org/>). Partial least squares discriminant analysis (PLS-DA) was performed in MetaboAnalyst 3.0⁶⁸ and important features were selected with a threshold of VIP (Variable Importance in Projection) ≥ 2.0.

Data availability

All MS proteomics data that support the findings of this study have been deposited to the ProteomeXchange Consortium (<http://www.proteomexchange.org>) with the dataset identifiers PXD007819 [<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX007819>] and PXD007959 [<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX007959>].

Received: 12 January 2018 Accepted: 29 June 2018

Published online: 20 July 2018

References

1. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307–317 (2011).
2. Ng, S. C. et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* **390**, 2769–2778 (2017).
3. Benchimol, E. I. et al. Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflamm. Bowel Dis.* **17**, 423–439 (2011).
4. Ruel, J., Ruane, D., Mehandru, S., Gower-Rousseau, C. & Colombel, J. F. IBD across the age spectrum: is it the same disease? *Nat. Rev. Gastroenterol. Hepatol.* **11**, 88–98 (2014).
5. Kappelman, M. D. et al. Direct health care costs of Crohn’s disease and ulcerative colitis in US children and adults. *Gastroenterology* **135**, 1907–1913 (2008).
6. Gevers, D. et al. The treatment-naive microbiome in new-onset Crohn’s disease. *Cell Host Microbe* **15**, 382–392 (2014).
7. Sokol, H. et al. Fungal microbiota dysbiosis in IBD. *Gut* **66**, 1039–1048 (2017).
8. Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
9. Mottawea, W. et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn’s disease. *Nat. Commun.* **7**, 13419 (2016).
10. Li, X. et al. Microgeographic proteomic networks of the human colonic mucosa and their association with inflammatory bowel disease. *Cell. Mol. Gastroenterol. Hepatol.* **2**, 567–583 (2016).
11. Fraher, M. H., O’Toole, P. W. & Quigley, E. M. Techniques used to characterize the gut microbiota: a guide for the clinician. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 312–322 (2012).
12. Verberkmoes, N. C. et al. Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* **3**, 179–189 (2009).
13. Zhang, X. et al. Deep metaproteomics approach for the study of human microbiomes. *Anal. Chem.* **89**, 9407–9415 (2017).
14. Erickson, A. R. et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn’s disease. *PLoS ONE* **7**, e49138 (2012).
15. Zhang, X. et al. MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota. *Microbiome* **4**, 31 (2016).
16. Sauer, C. G. & Kugathasan, S. Pediatric inflammatory bowel disease: highlighting pediatric differences in IBD. *Gastroenterol. Clin. North Am.* **38**, 611–628 (2009).

17. Vernier-Massouille, G. et al. Natural history of pediatric Crohn's disease: a population-based cohort study. *Gastroenterology* **135**, 1106–1113 (2008).
18. Szatanek, R., Baran, J., Siedlar, M. & Baj-Krzyworzeka, M. Isolation of extracellular vesicles: determining the correct approach (Review). *Int. J. Mol. Med.* **36**, 11–17 (2015).
19. Kelly, B. J. et al. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* **31**, 2461–2468 (2015).
20. Arenas, F. A. et al. The *Escherichia coli* *btuE* gene, encodes a glutathione peroxidase that is induced under oxidative stress conditions. *Biochem. Biophys. Res. Commun.* **398**, 690–694 (2010).
21. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
22. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
23. Loubinoux, J., Bronowicki, J. P., Pereira, I. A., Mouguel, J. L. & Faou, A. E. Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *FEMS Microbiol. Ecol.* **40**, 107–112 (2002).
24. Yoshida, Y. et al. Production of hydrogen sulfide by two enzymes associated with biosynthesis of homocysteine and lanthionine in *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586. *Microbiology* **156**, 2260–2269 (2010).
25. Song, Z. et al. L-cysteine protects intestinal integrity, attenuates intestinal inflammation and oxidant stress, and modulates NF- κ B and Nrf2 pathways in weaned piglets after LPS challenge. *Innate Immun.* **22**, 152–161 (2016).
26. Lewis, J. D. et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* **18**, 489–500 (2015).
27. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
28. Hansen, R. et al. Microbiota of de-novo pediatric IBD: increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am. J. Gastroenterol.* **107**, 1913–1922 (2012).
29. Assa, A. et al. Mucosa-associated ileal microbiota in new-onset pediatric Crohn's Disease. *Inflamm. Bowel Dis.* **22**, 1533–1539 (2016).
30. Sokol, H. et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).
31. Song, H., Yoo, Y., Hwang, J., Na, Y. C. & Kim, H. S. *Faecalibacterium prausnitzii* subspecies-level dysbiosis in the human gut microbiome underlying atopic dermatitis. *J. Allergy Clin. Immunol.* **137**, 852–860 (2016).
32. Szklarczyk, D. et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
33. Kim, D. K. et al. EVpedia: a community web portal for extracellular vesicles research. *Bioinformatics* **31**, 933–939 (2015).
34. Simpson, R. J., Kalra, H. & Mathivanan, S. ExoCarta as a resource for exosomal research. *J. Extracell. Vesicles* **1**, 18374 (2012).
35. David, S. S. & Williams, S. D. Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.* **98**, 1221–1261 (1998).
36. Lotvall, J. et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J. Extracell. Vesicles* **3**, 26913 (2014).
37. Kugathasan, S. et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* **389**, 1710–1718 (2017).
38. Schirmer, M. et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* **3**, 337–346 (2018).
39. Raposo, G. & Stoorvogel, W. Extracellular vesicles: exosomes, microvesicles, and friends. *J. Cell Biol.* **200**, 373–383 (2013).
40. Timar, C. I. et al. Antibacterial effect of microvesicles released from human neutrophilic granulocytes. *Blood* **121**, 510–518 (2013).
41. Lakschevitz, F. S. et al. Identification of neutrophil surface marker changes in health and inflammation using high-throughput screening flow cytometry. *Exp. Cell Res.* **342**, 200–209 (2016).
42. Saiki, T. Myeloperoxidase concentrations in the stool as a new parameter of inflammatory bowel disease. *Kurume Med. J.* **45**, 69–73 (1998).
43. Masoodi, I. et al. Evaluation of fecal myeloperoxidase as a biomarker of disease activity and severity in ulcerative colitis. *Dig. Dis. Sci.* **57**, 1336–1340 (2012).
44. Wagner, M., Peterson, C. G., Ridefelt, P., Sangfelt, P. & Carlson, M. Fecal markers of inflammation used as surrogate markers for treatment outcome in relapsing inflammatory bowel disease. *World J. Gastroenterol.* **14**, 5584–5589 (2008). discussion 5588.
45. Davies, M. J. Myeloperoxidase-derived oxidation: mechanisms of biological damage and its prevention. *J. Clin. Biochem. Nutr.* **48**, 8–19 (2011).
46. Friedberg, E. C., Walker, G. C., Siede, W. & Wood, R. D. *DNA repair and mutagenesis*, (American Society for Microbiology Press, Washington, 2005).
47. Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
48. Diard, M. et al. Inflammation boosts bacteriophage transfer between *Salmonella* spp. *Science* **355**, 1211–1215 (2017).
49. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
50. Raymond, F. et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. *ISME J.* **10**, 707–720 (2016).
51. Yin, J. et al. L-Cysteine metabolism and its nutritional implications. *Mol. Nutr. Food Res.* **60**, 134–146 (2016).
52. Bauchart-Thévret, C., Stoll, B., Chacko, S. & Burrin, D. G. Sulfur amino acid deficiency upregulates intestinal methionine cycle activity and suppresses epithelial growth in neonatal pigs. *Am. J. Physiol. Endocrinol. Metab.* **296**, E1239–E1250 (2009).
53. North American Society for Pediatric Gastroenterology, H. et al. 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.* **44**, 653–674 (2007).
54. Hyams, J. et al. Evaluation of the pediatric Crohn disease activity index: a prospective multicenter experience. *J. Pediatr. Gastroenterol. Nutr.* **41**, 416–421 (2005).
55. Turner, D. et al. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. *Gastroenterology* **133**, 423–432 (2007).
56. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
57. Chehoud, C. et al. Fungal signature in the gut microbiota of pediatric patients with inflammatory bowel disease. *Inflamm. Bowel Dis.* **21**, 1948–1956 (2015).
58. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
59. Cheng, K. et al. MetaLab: an automated pipeline for metaproteomic data analysis. *Microbiome* **5**, 157 (2017).
60. Jagtap, P. et al. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* **13**, 1352–1357 (2013).
61. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
62. Mitsuhashi, S. et al. Luminal extracellular vesicles (EVs) in inflammatory bowel disease (IBD) exhibit proinflammatory effects on epithelial cells and macrophages. *Inflamm. Bowel Dis.* **22**, 1587–1595 (2016).
63. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
64. Mesuere, B. et al. Unipept: tryptic peptide-based biodiversity analysis of metaproteome samples. *J. Proteome Res.* **11**, 5773–5780 (2012).
65. Mesuere, B. et al. Unipept web services for metaproteomics analysis. *Bioinformatics* **32**, 1746–1748 (2016).
66. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
67. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
68. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–W257 (2015).

Acknowledgements

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada, Genome Canada (OGI-114 and OGI-067), Canadian Institutes of Health Research (GPH-129340, ECD-144627, and MOP-114872), Ontario Ministry of Economic Development and Innovation (REG1-4450), IBD Foundation of Crohn's and Colitis Canada, and the Children's Hospital of Eastern Ontario (CHEO) Foundation. D.F. acknowledges a Canada Research Chair in Proteomics and Systems Biology. S.A.D. acknowledges support from the Ontario Graduate Scholarship. We thank Dr. Derrick Gibbins and Ryan Reshke for help in nanoparticle tracking analysis.

Author contributions

D.F., A.S., D.M., X.Z., and S.A.D. designed the study. D.M., J.B., J.L., and R.S. collected samples. X.Z. and S.A.D. performed the experiments. All authors participated in the data

analysis and interpretation. X.Z., S.A.D., D.F., and A.S. wrote the manuscript. All authors read and approved the final manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-05357-4>.

Competing interests: A.S., D.M., and D.F. have co-founded Biotagenics, a clinical microbiomics company. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018