Clinical and Experimental Gastroenterology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Microbial Patterns in Newly Diagnosed Inflammatory Bowel Disease Revealed by Presence and Transcriptional Activity - Relationship to Diagnosis and Outcome

Simen Svendsen Vatn[1–3,*], Simen Hyll Hansen[1,4,*], Tone Møller Tannæs[5], Stephan Brackmann[1,2], Christine Olbjørn[1,6], Daniel Bergemalm[7], Åsa V Keita[8], Fernando Gomollon[9], Trond Espen Detlie[1,2], Rahul Kalla[10], Jack Satsangi[10,11], Jørgen Jahnsen[1,2], Morten Harald Vatn[1], Jonas Halfvarson[7], Johannes Roksund Hov[1,4,12], Petr Ricanek[2,13], Aina EF Moen[1,5,14] On behalf of the IBD-Character Consortium

[1]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [2]Department of Gastroenterology, Division of Medicine, Akershus University Hospital, Lørenskog, Norway; [3]Innlandet Hospital Trust, Gjøvik, Norway; [4]Norwegian PSC Research Center and Research Institute of Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital, Oslo, Norway; [5]Section for Clinical Molecular Biology (Epigen), Akershus University Hospital, Lørenskog, Norway; [6]Department of Pediatric and Adolescent Medicine, Akershus University Hospital, Lørenskog, Norway; [7]Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; [8]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden; [9]Digestive Diseases Unit, IIS Aragón, Zaragoza, Spain; [10]Gastrointestinal Unit, Centre for Genomics and Molecular Medicine, Division of Medical and Radiological Sciences, University of Edinburgh, Edinburgh, UK; [11]Translational Gastroenterology Unit, Medical Sciences/ Experimental Medicine Division, University of Oxford, Oxford, UK; [12]Section of Gastroenterology, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway; [13]Department of Gastroenterology, Iovisenberg Diaconal Hospital, Oslo, Norway; [14]Department of Virology, Norwegian Institute of Public Health, Oslo, Norway

*These authors contributed equally to this work

Correspondence: Simen Svendsen Vatn, Innlandet Hospital Trust, Gjøvik, Norway, Tel +47 94277594, Email bikkjas@hotmail.com

Background: As part of the IBD Character initiative, we examined an inception cohort and investigated mucosal microbiota composition and transcriptional activity in relation to clinical outcomes.

Methods: A cohort of 237 individuals were included from five countries: Crohn's disease (CD, n = 72), ulcerative colitis (UC, n = 57), symptomatic non-IBD controls (SC, n = 78) and healthy controls (HC, n = 30). Rectal/colonic biopsies were obtained at inclusion, and DNA and RNA were extracted from the same biopsy and examined by sequencing the 16S rRNA V4 region.

Results: Beta diversity measurements separated IBD from both HC and SC. IBD and SC exhibited reduced intra-individual diversity compared with HC. When comparing taxonomy at DNA and RNA level, six bacteria were found to differ in abundance and/or transcriptional activity between IBD and symptomatic control, while there were 14 and three between symptomatic control and CD and UC, respectively. A limited number of bacterial taxa were responsible for the largest difference between presence and activity, separating patients and controls. Multiple bacterial taxa were associated with treatment escalation in both UC and CD. Machine-learning models separated IBD from symptomatic controls and treatment escalators from non-escalators (AUC >0.8). However, the differential effects were mainly driven by clinical biomarkers, such as f-calprotectin, s-albumin, and b-hemoglobin.

Conclusion: Differences between presence and transcriptional activity were found among multiple taxa when assessing 16S rRNA at DNA and RNA level. Symptomatic controls were more similar to the IBD patients compared to HC. The analyses suggest that the mucosal microbiota carries a moderate diagnostic and predictive potential, outcompeted by f-calprotectin.

Keywords: microbiota, RNA, DNA, IBD, biomarkers

Introduction

Disturbances of the gut microbiota composition, as well as the mutualistic relationship between host and gut microbiota, have been demonstrated to play ubiquitous roles in inflammatory bowel disease (IBD).[1–3] Studies of both fecal and mucosa-associated

Received: 6 November 2024
Accepted: 29 April 2025
Published: 21 May 2025

microbiota have revealed a skewed pattern, including a reduction in biodiversity and abnormal microbial composition with a decrease in several taxa within the phylum *Firmicutes* and an increase in the phyla *Bacteroidetes* and *Proteobacteria*, in both Crohn's disease (CD)[4] and ulcerative colitis (UC).[1,5] Furthermore, compositional differences between CD and UC have been demonstrated, and microbial signatures on species and genus levels have also been found.[6–8] The mucosa-associated microbiota composition differs from the fecal microbiota and is more host-specific than the latter,[9,10] suggesting possible strengths of metagenomics based on mucosal biopsies. Swidsinski et al showed that IBD patients have a greater overall density of bacteria attached to the mucus layer compared to healthy controls (HC),[5] and this mucosa-associated microbiota is proposed to have a major impact on the host-microbiota crosstalk and maintenance of gut barrier functions.[11,12] Assessing the transcriptionally active mucosa-associated microbiota has the potential of discovering novel aspects of IBD pathobiology at the barrier interface separating host and the luminal content. By utilizing the transcribed 16S rRNA count, which correlates with the protein synthesis potential of a microbe, the activity rather than the presence of the microbe, can be assessed.[13–15] In a subcohort of UC patients in the IBD Character study, we have already demonstrated differential abundances among taxa between the total and transcriptionally active microbiota by simultaneous purification and sequencing of microbial DNA and RNA in single colonic mucosal biopsies.[16]

CD and UC patients represent a highly heterogenous group with wide variability in clinical outcomes.[17–19] Some patients have a mild disease course, whilst others experience severe morbidity because of sustained treatment-refractory inflammation of the intestinal wall. Nevertheless, tools for predicting the disease course in IBD are insufficient. The identification of microbiota signatures in early disease might be a tool for personalized medicine and more targeted therapy.[20–23] In this study, we assessed the mucosal microbiota composition and the transcriptional activity in inflamed and non-inflamed biopsies from newly diagnosed patients with CD and UC, HC and symptomatic non-IBD controls (SC). We further related mucosal microbiota in patients with IBD to need of treatment escalation and anti-TNF response in search of novel prognostic biomarkers.

Methods

Patient Recruitment, Sample Collection and Follow-Up

Patients referred to colonoscopy with suspected IBD were recruited prospectively in the EU FP7 IBD-Character study,[24] involving six university clinics across Europe (Akershus, Norway; Örebro, Sweden; Linköping, Sweden; Edinburgh, United Kingdom; Zaragoza, Spain; Maastricht, Netherlands). The diagnosis of IBD was based on internationally accepted diagnostic criteria.[25,26] Biopsies from inflamed and non-inflamed tissue were collected from ileum and different colorectal segments during colonoscopy and immediately stored in Allprotect Tissue Reagent (Qiagen, Hilden, Germany) at −80°C, preserving both DNA and RNA. Patients with gastrointestinal symptoms, but with no endoscopic or histologic signs of IBD at inclusion and no evidence of IBD during follow-up were classified as SC. Patients with a possible, but no definite IBD diagnosis, were excluded from the study. Subjects without gastrointestinal symptoms or disease were recruited as HC at the Swedish centers.

The majority of the IBD patients were treatment naïve (92%), while others had been started on 5-ASA and/or glucocorticoids in primary care, before inclusion and biobanking in a tertiary center. The patients were followed in clinical routine according to national and European guidelines[25–27] up to five years after inclusion in this study. Treatment escalation was defined as introduction of anti-TNF agents, cyclosporine or surgery in the case of non-response to treatment with 5-ASA, thiopurines and glucocorticoids. In individuals who experienced a severe disease course, which led to the introduction of anti-TNF agents in the time after enrollment to the study, response was evaluated 14 weeks after induction of anti-TNF. For CD, response was defined as no use of concomitant steroids, a decrease in the Harvey-Bradshaw Index (HBI) of at least 3 points, or to ≤4, and at least one of the following two criteria: (a) a reduction in CRP by at least 50% or to ≤3 mg/l from baseline, or (b) a reduction in fecal calprotectin by at least 30% or to <250mg/kg. For UC, response was defined as no use of concomitant steroids, a reduction from the baseline partial Mayo score of at least 3 points, and at least 30%, and a reduction in the rectal bleeding subscore of at least 1 point or an absolute rectal bleeding score of 0 or 1. Additionally, at least one of the following three criteria was achieved: (a) a reduction in CRP fall by at least 50% or to ≤3 mg/l from baseline, or (b) a reduction in fecal calprotectin by at least 30% or to <250mg/kg, or (c) a reduction of at least 1 in the Endoscopic Mayo score.

Mucosa Associated Microbiota Profiling from DNA and RNA

Mucosal microbiota composition and potential activity were assessed using both a 16S rRNA transcript and gene amplicon sequencing. The methods used for nucleic acid purification, cDNA synthesis, 16S rRNA sequencing and taxonomic identification have been described in our previous publication.[28] Briefly, RNA and DNA were purified using the AllPrep DNA/RNA Mini kit (Qiagen) following a modified protocol.[28] The concentrations of the DNA and RNA samples were assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA integrity number (RIN) was assessed with an Agilent 2100 Bioanalyzer, Agilent 2100 Expert software and Agilent RNA 6000 Nano Kit (Agilent Technologies Inc., Santa Clara, CA, USA). The cDNA was synthesized from 1 µg RNA from each sample using the AccuScript High Fidelity 1st Strand cDNA Synthesis Kit (Agilent Technologies Inc.) and random hexamers, according to the manufacturer's instructions. Two random RNA samples were run in the absence of reverse transcriptase to assess the degree of contaminating genomic DNA. Blank samples were also run through the wet lab procedure, from the nucleic acid purification step through the sequencing process, to detect possible reagent contamination.

16S rRNA Amplicon Sequencing

Two 16S amplicon libraries targeting the V4 segment were made according to a protocol published by Kozich and colleagues[29] using a 500 ng DNA and 250 ng cDNA template, respectively. The sequencing was performed using the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA) and the MiSeq reagent kit v2 (500 cycles) according to the manufacturer's instructions, with the addition of custom sequencing primers and index and 8% PhiX, as described in the MiSeq Wet Lab SOP.[30]

Sequence Processing and Quality Control

The minimum numbers of reads from each sample was set from the number of reads on the DNA and RNA blanks. MiSeq Reporter Software (Illumina Inc.) was used for demultiplexing the reads and for fastQ file generation.

The total sequence data were processed with QIIME2 v/2021.4. The raw fastQ reads were quality filtered, trimmed, de-noised, and paired end sequences merged using Deblur[31] and the q2-deblur plugin implemented in QIIME2. Default settings in the q2-deblur plugin were used.

Taxonomy was assigned to the amplicon sequence variants (ASVs) using the Silva V.138 reference sequence database, VSEARCH,[32] and the q2-feature-classifier plugin.

After library prep, 15 DNA samples and 21 RNA samples did not pass quality control (QC), and after sequencing, 12 DNA and nine RNA samples had lower number of raw reads than what was produced from the DNA and RNA blanks, respectively, and hence were removed.

Statistics/Bioinformatics

Analysis Setup

We examined each combination of sequencing data type (either DNA only, RNA only or a combination of DNA + RNA samples) and biopsy type (biopsy from either inflamed or non-inflamed segments of the mucosa, or all biopsies included), resulting in a total of nine different datasets. Due to biases relating raw read counts to patient diagnosis and prognosis, all samples were rarefied to the lowest number of reads for any sample within each dataset, ranging from 217 reads in the RNA dataset of non-inflamed biopsies to 887 reads in the DNA dataset of non-inflamed biopsies.

Main Software Versions

R version 4.2.3[33] was used for all analyses and modelling except for those relating to machine learning. The alpha diversity metrics were calculated using the R package *vegan* 2.6–4,[34] while beta diversity distances and coordinates were produced with *phyloseq*.[35] Machine learning was performed with the *gpboost 0.7.3* package[36] in Python 3.6.7,[37] using the *reticulate* R package[38] to allow Python access of the datasets in R. The *gpboost* models were trained iteratively, repeating n = 100 times for each combination of input data and outcome variable of interest. As input data, either only microbial sequencing data or only "traditional clinical markers" (Hb, leukocytes, platelets, albumin, ALP, CRP and f-calprotectin) were used, or a combination of the two. After training and assessing model performances with area under the ROC curve

(AUC values), the impacts of individual variables were assessed using SHapley Additive exPlanations-values (SHAP-values), calculated with the *shap* package in Python.[39]

Mixed Models and Random Effects

To account for multiple samples per patient in the study, mixed models were applied where possible using patient ID, country of sample origin and biopsy location as random effects. For testing differences in microbial diversity, we designed such mixed models using the *glmer* function of the *lme4* package.[40] Likewise, the *MaAsLin 2* package[41] was used with random effects to test for differential abundance of bacterial taxa. When looking at beta diversity, restrictions in the software made it necessary to instead calculate centroids as average values for all samples per patient within each dataset. In the *gpboost* machine learning models,[36] the same random effects were accounted for using mixed models' design. For simple statistical comparisons, nominal p-values are reported, while for differential abundance analysis using MaAsLin 2 we report q-values calculated as false discovery rate (FDR) corrected p-values. Additionally, nominal p-values are included in differential abundance figures as colored areas lacking an "x".

Visualization

All visualizations were produced using *ggplot2*, for some plots in a combination with *ggpubr*,[42] *gghalves*,[43] *ggdist*[44] or *gridExtra*.[45]

Results

In the present study a total of 237 subjects were included and grouped into four groups; CD (N = 72), UC (N = 57), SC (N=78), and HC (N = 30). Table 1 demonstrates demographics and clinical characteristics of the groups. In IBD, mean time from diagnosis to inclusion was 14 days, while the median was 0 days, as the majority of IBD was diagnosed at the date of inclusion and biobanking. SC in this study had gastrointestinal symptoms with no signs of inflammation or infection, and 70% were retrospectively diagnosed with irritable bowel syndrome (IBS) according to the ROME IV criteria.[46]

From the 72 CD, 57 UC, 78 SC and 30 hC included in the present study, 107, 119, 103 and 30 biopsies were respectively collected, resulting in a total of 359 samples from 237 persons undergoing combined DNA/RNA extraction (Table 1). Following library preparation and sequencing, and subsequent quality filtering, denoising and assembling of overlapping paired-end reads, 359 DNA samples yielding3428086merged sequences and 356 RNA samples yielding2000855merged sequences were included in the final datasets. The sequences were clustered into 2470 and 2281 ASVs from the DNA and RNA datasets, respectively. After rarefying and filtering of taxa, the data comprised 40 and 37 genera from nine and eight phyla, when using DNA and RNA, respectively.

Reduced Inter-Individual (Beta) Diversity and Intra-Individual (Alpha) Diversity in IBD and Symptomatic Controls

In an analysis of global microbiota composition, as measured by beta diversity, there were statistically significant differences between IBD, HC and SC, mainly driven by increased dispersion (variability) in the IBD group, irrespective of data being RNA or DNA-based. The differences were larger in the RNA dataset than in the DNA dataset (Figure 1A). Regarding intra-individual (alpha) diversity, a higher diversity was seen among HC compared to CD and UC in both datasets when all samples were pooled (both inflamed and non-inflamed samples), while SC were indistinguishable from CD and UC with similar reduced diversity (Figure 1B).

Altered Microbial Taxa in IBD Defined by DNA and RNA

In the differential abundance analyses, we wanted to define bacterial taxa that separated IBD from HC and SC, respectively. This was performed by including all IBD biopsies (both inflamed and non-inflamed), to account for the entire microbiota of IBD, but we also made a separate comparison including only non-inflamed biopsies, and both DNA- and RNA-based microbiota profiles were used (Figure 2A and B).

Comparing IBD to HC revealed a total of 19 taxa with different abundances ($q < 0.05$) in either of these analyses (Figure 2A). Consistent findings in common for both the DNA- and RNA-based microbial profiles, were a significantly

Table 1 Demographics, Clinical Characteristics and Biobanking for Patients and Controls

	CD (N = 72)	UC (N = 57)	SC (N = 78)	HC (N = 30)
Age at inclusion				
Mean (SD)	33.3 (13.5)	35.9 (12.6)	37.1 (12.6)	26.4 (6.52)
Median [Min, Max]	29 [18, 66]	35 [18, 66]	33 [20, 71]	24.5 [19, 48]
CRP				
Mean (SD)	31.5 (52.8)	19.5 (36.8)	5.66 (11.8)	NA (NA)
Median [Min, Max]	10.5 [0, 270]	5.00 [0, 210]	3.00 [0, 85.0]	NA [NA, NA]
Missing	8 (11.1%)	5 (8.8%)	16 (20.5%)	30 (100%)
Fecal calprotectin				
Mean (SD)	822 (1410)	1390 (1420)	155 (260)	NA (NA)
Median [Min, Max]	504 [0, 10,000]	993 [0, 6000]	32.0 [0, 1060]	NA [NA, NA]
Missing	7 (9.7%)	10 (17.5%)	20 (25.6%)	30 (100%)
Montreal (UC)[a]				
Proctitis (E1)		13 (22.8%)		
Left-sided (E2)		18 (31.6%)		
Extensive (E3)		26 (45.6%)		
Missing		0 (0%)		
Montreal (CD)[b]**- Age (years)**				
<17 (A1)	0 (0%)			
17–40 (A2)	55 (76.4%)			
>40 (A3)	17 (23.6%)			
Missing	0 (0%)			
Montreal (CD)[b] **- Location**				
Ileal (L1)	26 (36.1%)			
Colonic (L2)	21 (29.2%)			
Ileocolonic (L3)	25 (34.7%)			
Missing	0 (0%)			
Montreal (CD)[b]**- Behavior**				
Non-stricturing non-penetrating (B1)	57 (79.2%)			
Stricturing (B2)	8 (11.1%)			
Penetrating (B3)	2 (2.8%)			
Perianal non-stricturing non-penetrating (B1p)	3 (4.2%)			
Perianal penetrating (B3p)	2 (2.8%)			
Missing	0 (0%)			
Biopsies				
Terminal Ileum				
0[c]	45 (62.5%)	21 (36.8%)	50 (64.1%)	30 (100%)
1[d]	27 (37.5%)	36 (63.2%)	28 (35.9%)	0 (0%)
Proximal Colon				
0	50 (69.4%)	25 (43.9%)	57 (73.1%)	30 (100%)
1	22 (30.6%)	32 (56.1%)	21 (26.9%)	0 (0%)
Distal Colon				
0	52 (72.2%)	28 (49.1%)	77 (98.7%)	30 (100%)
1	20 (27.8%)	29 (50.9%)	1 (1.3%)	0 (0%)
Rectum				
0	34 (47.2%)	35 (61.4%)	25 (32.1%)	0 (0%)
1	38 (52.8%)	22 (38.6%)	53 (67.9%)	30 (100%)

Notes: [a]Montreal classification for disease extension in ulcerative colitis. There are three categories: E1, E2 and E3. [b]Montreal classification for CD, defined by Age, Location and behavior. [c]number of patients without biopsy from this location. [d]number of patients with biopsy from this location.

Abbreviations: CD, Crohn's disease; CRP, C-reactive protein; HC, healthy controls; NA, not available; SD, standard deviation; SC, symptomatic controls; UC, ulcerative colitis.

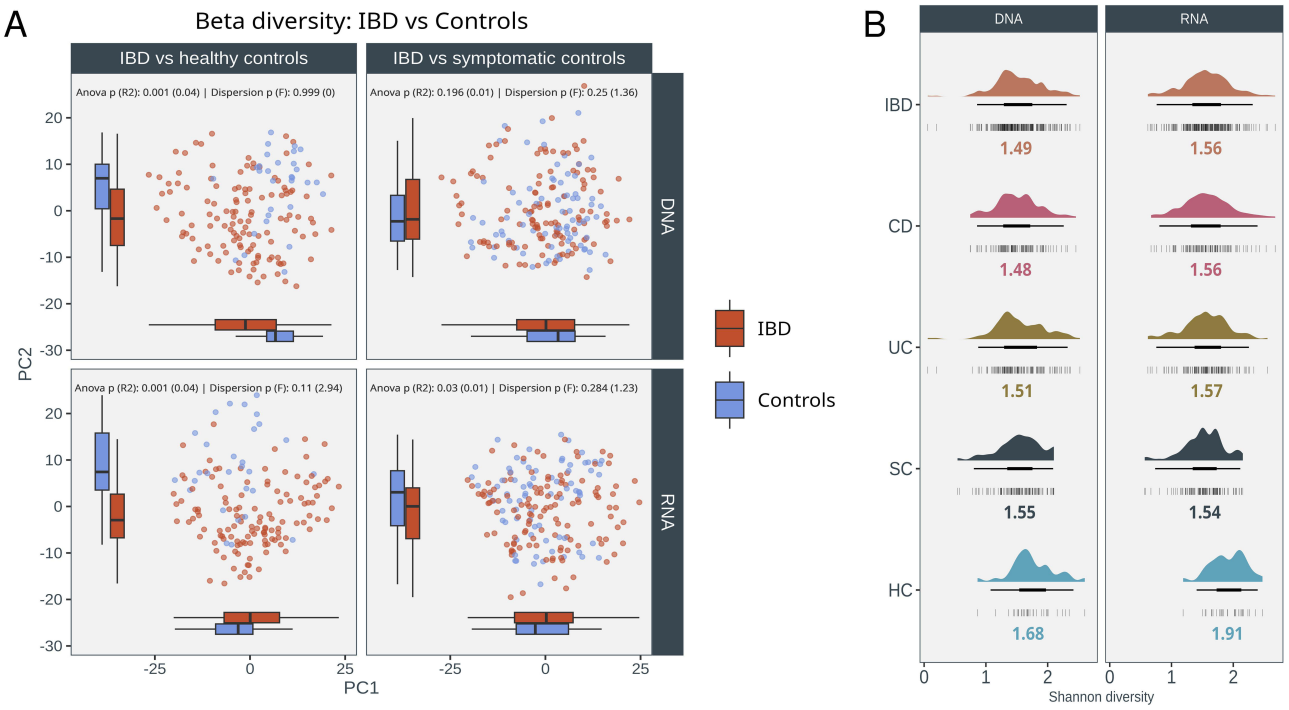

Figure 1 (A) Beta diversity measurements demonstrated by principal component analyses (pCoA) plots and boxplots. IBD in red. Controls in blue. The upper two plots are based on the DNA dataset and the lower two plots are based on the RNA dataset. Comparison of IBD vs HC to the left and comparison of IBD vs SC to the right. Horizontal Boxplots for PC1 (principal component 1). Vertical Boxplots for PC2. Black line in the box demonstrates the median value. Standard inter-quartile range (IQR) is applied, with 50% of samples within the box. Extended lines from the boxes illustrates dispersion/variation within the group. **(B)** Alpha-diversity as estimated by the Shannon diversity index in the DNA (left) and RNA dataset (right). The median index values are given per group.
Abbreviations: IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; SC, symptomatic controls; HC, healthy controls; PC1/2, principal component 1/2.

lower load of *Clostridium senso strictu 1, Ruminococcaceae* and *Bifidobacterium* and a higher load of *Bacteroides* in IBD. Notable differences between the analytical strategies included a higher abundance of eg *Pseudomonas and Enterobacterales*, which was only seen in the DNA dataset. Furthermore, a high abundance of *Burkholderiales indet*. and a low abundance of *Bacillaceae* in IBD compared to HC were only found in the RNA dataset.

Comparing IBD to SC revealed a total of six taxa with different abundances at $q < 0.05$ (Figure 2B) with limited overlap with the findings that separated IBD and HC. The differences between IBD and SC were mainly driven by the distinct microbiota composition in CD.

Notable observations included increased relative abundance of *Pseudomonas* in IBD, which was mainly driven by CD. Furthermore, multiple taxa within the phylum Proteobacteria demonstrated higher abundance in IBD compared to SC, while the differences to phylum Firmicutes was smaller than the comparison to HC. The relative abundance of *Akkermansia* was high in CD in non-inflamed samples.

There were multiple taxa with different relative abundance when comparing CD vs UC, even after correction for multiple comparisons (Figure 2C), and there were few overlapping findings when comparing the associations observed in the DNA and RNA datasets, respectively.

Integrated Analysis of Clinical Variables and Microbiota Profiles to Separate IBD and Symptomatic Controls

In the next step, we applied machine-learning methodologies and both microbial and clinical variables to assess their ability to separate IBD, CD and UC from controls. RNA and DNA showed differing signatures in the differential abundance analysis suggesting that they carry differing information, thus both datasets were included in an integrated predictive model. The combination of clinical and microbiota data had a strong ability to separate IBD from SC (AUC > 0.80, Figure 3A).

Figure 2 Heatmaps summarizing results from differential abundance analyses using MaAsLin 2 in **(A)** IBD vs healthy controls, **(B)** IBD vs symptomatic controls and **(C)** CD vs UC. Each taxon is grouped according to its phylum. In each heatmap, the two leftmost columns show combined DNA/RNA signatures, while those in the middle represents the DNA dataset, and the rightmost columns show the RNA dataset. "All biopsies" refers to biopsies from both inflamed and non-inflamed tissue, and separate columns for inflamed-only and non-inflamed only are also included. A log2 fold change scale ranges from deep red (more abundant in IBD, CD and UC (**A** and **B**), or in CD **(C)**) to deep blue (less abundant in IBD, CD and UC **(A, B)**, or in CD **(C)**). All results with nominal p-values < 0.05 are shown, while q-values < 0.05 (FDR-adjusted p-values) are marked with "x".
Abbreviations: HC, healthy controls; SC, symptomatic controls; IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis.

https://doi.org/10.2147/CEG.S504459

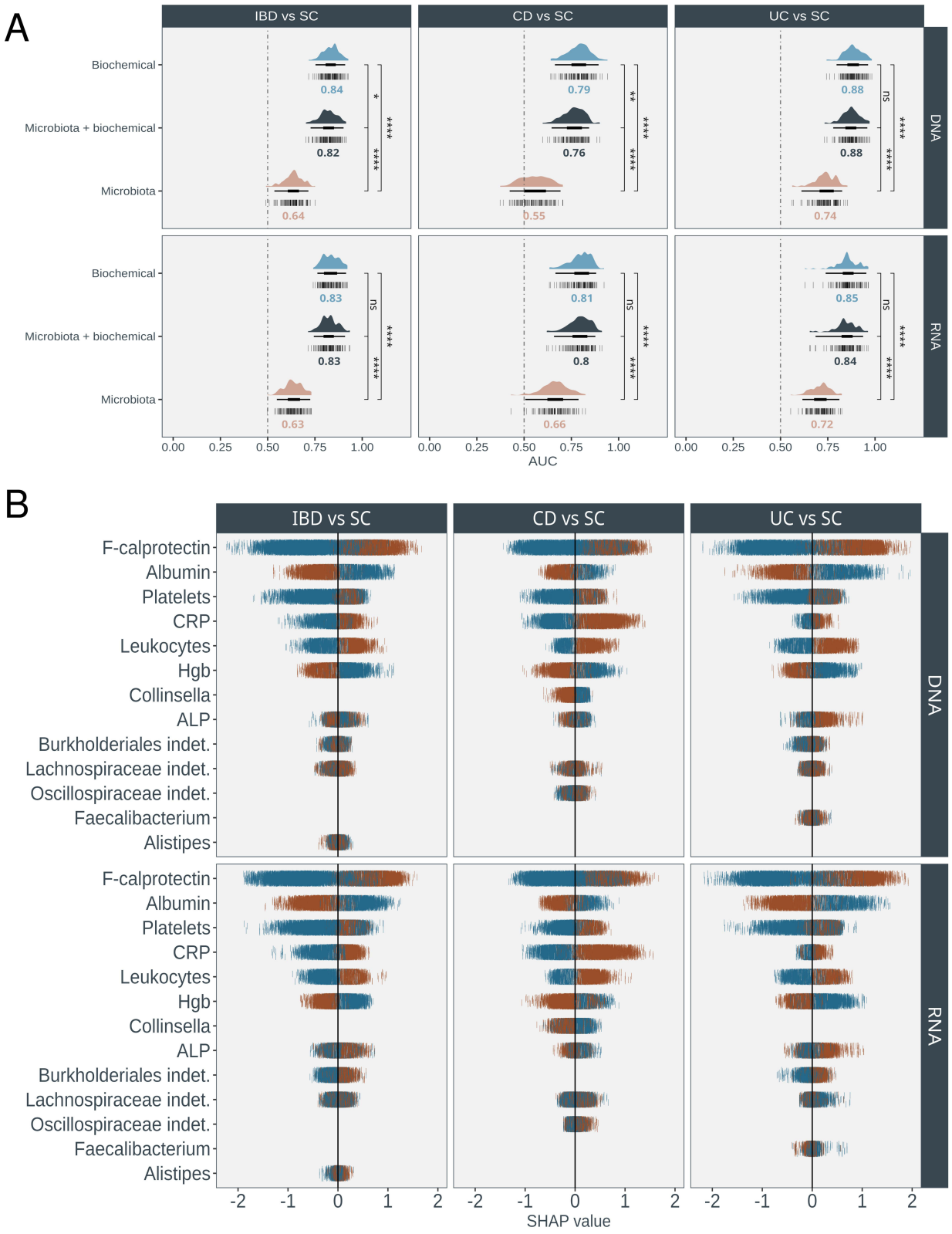

Figure 3 Machine learning results for separating IBD from symptomatic controls, performed with the *gpboost* in Python 3.6.7 (see Methods). Three different data sets were used as input in the models: microbial data, biochemical markers (Hbg, leukocytes, platelets, albumin, ALP, CRP and f-calprotectin), or a combination of the two. **(A)** Model performance was estimated with area under the receiver operating curve (AUC), and compared between model categories using t-tests. * = p<0.05. ** = p<0.01. **** = p<0.0001. **(B)** The impacts of individual variables were assessed using SHapley Additive exPlanations-values (SHAP-values), calculated with the *shap* package in Python, and the top 10 variables per classification task was visualized.

Abbreviations: AUC, area under curve; Hbg, Blood-hemoglobin; Leucocytes, Blood-leucocytes; Platelets, Blood-platelets; Albumin, serum-albumin; ALP, serum-alkaline phosphatase; CRP, Serum-C-Reactive Protein; F-Calprotectin, fecal calprotectin; HC, healthy controls; SC, symptomatic controls; IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis.

https://doi.org/10.2147/CEG.S504459

However, in all models, the contribution from microbiota was small compared to other parameters, and the difference was mainly driven by biomarkers already in clinical use, including f-calprotectin, albumin, Hb and CRP (Figure 3B).

Differential Bacterial Transcription Defined by RNA/DNA Ratio of Individual Taxa in IBD and Controls

So far, we identified multiple compositional differences between IBD and control groups and between UC and CD. There were some, but not complete overlap between the results defined by the taxonomic profiles based on DNA and RNA. In the next step, we decided to investigate the ratio of relative abundance on RNA level (expression/activity) to DNA level per taxon as a measure of the transcriptional activity of the individual taxon. The RNA/DNA ratio was separately measured in IBD and in controls, where HC and SC were merged to one group. The ratio differed extensively between taxa in both groups. The RNA/DNA ratio was similar in IBD and in controls for certain taxa: *Rhodanobacteriaceae* and *Pseudomonas* both exhibited high transcriptional activity compared to their presence, while members from the phylum Bacteroidetes demonstrated low activity in both groups (Figure 4). However, for several taxa, there was also evidence of reduced transcriptional activity (indeterminate genera of the *Pasteurellaceae* and *Actinobacteria* families) or more transcriptionally active (a genus of the *Burkholderiales* family) in IBD (Figure 4).

Clinical Impact of Microbial Signatures to Predict Treatment Escalation or Failure

Among 62 CD patients, 25 (40.3%) escalated treatment, while among 53 UC patients, 6 (11.3%) escalated treatment during a median follow-up of 24 months. There were no significant associations between alpha diversity and treatment escalation in neither CD nor UC (Supplementary Figure 1). In contrast, treatment escalation in CD was associated with differential relative abundances of multiple taxa in the differential abundance analysis ($q < 0.05$). Alterations of similar magnitude, but with limited overlap with CD were observed in UC (Figure 5A).

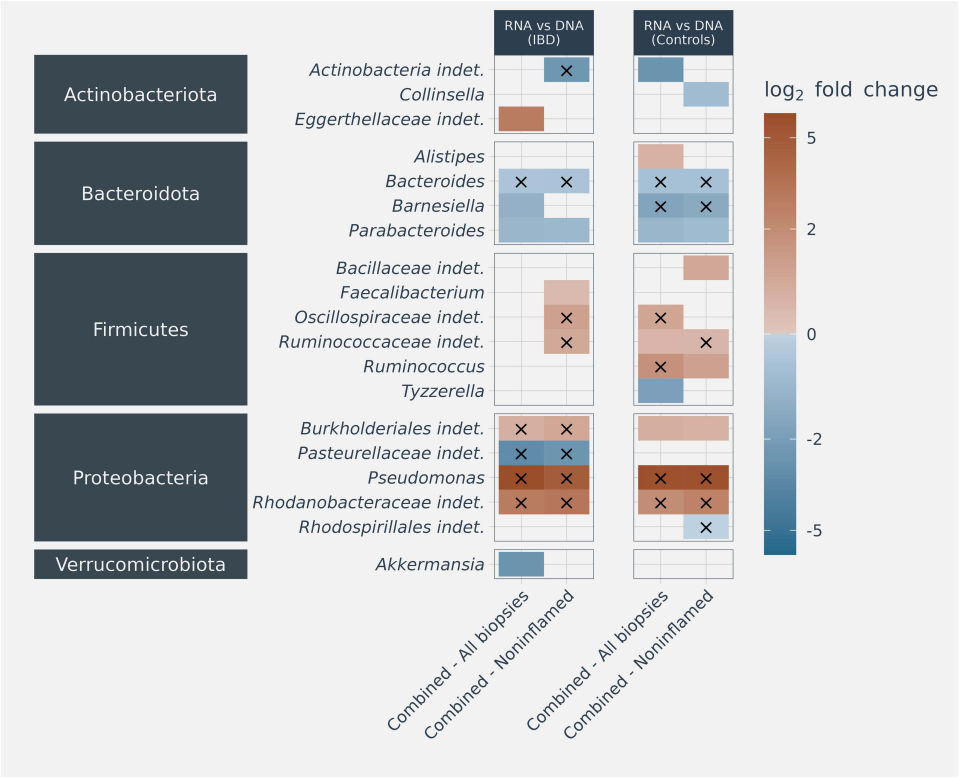

Figure 4 Heatmaps showing results from differential abundance analyses using MaAsLin 2 directly comparing RNA and DNA-derived abundances in IBD and controls (symptomatic and healthy combined). Each taxon is grouped according to its phylum. Only the categories "All biopsies" or noninflamed-only biopsies are shown. A log2 fold change scale ranges from deep red (more abundant RNA/active microbiota) to deep blue (more abundant in DNA/present microbiota). All results with nominal p-values < 0.05 are shown, while q-values < 0.05 (FDR-adjusted p-values) are marked with "x".
Abbreviations: IBD, inflammatory bowel disease; DNA, Deoxyribonucleic acid; RNA, Ribonucleic acid.

https://doi.org/10.2147/CEG.S504459

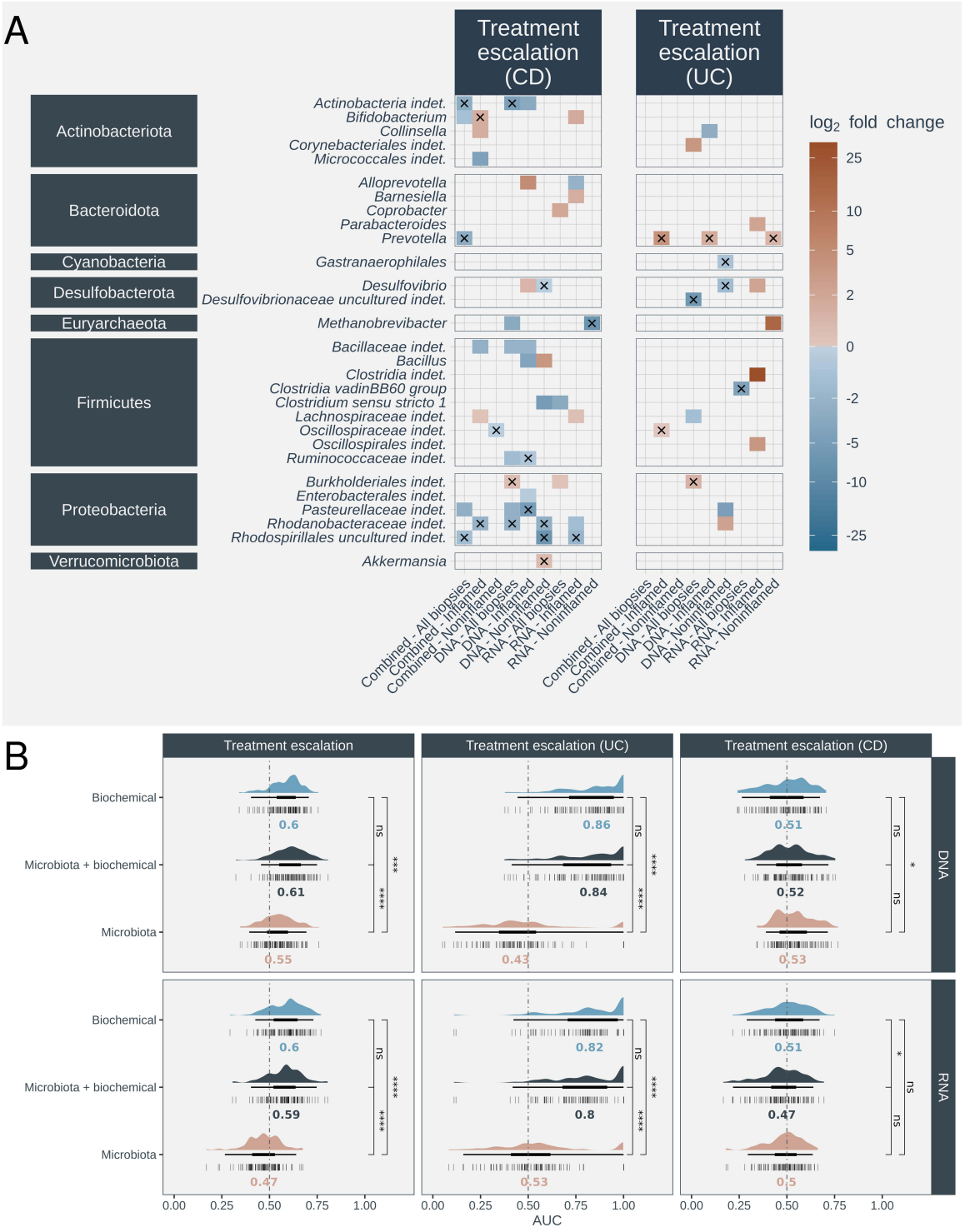

Figure 5 (A) Heatmaps summarizing results from differential abundance analyses of treatment escalation in CD and UC using MaAsLin 2. Each taxon is grouped according to its phylum. In each heatmap, the two leftmost columns show combined DNA/RNA signatures, while those in the middle represents the DNA dataset, and the rightmost columns show the RNA dataset. "All biopsies" refers to biopsies from both inflamed and non-inflamed tissue, and separate columns for inflamed-only and non-inflamed only are also included. A log2 fold change scale ranges from deep red (more abundant in treatment escalators) to deep blue (less abundant in treatment escalators). All results with nominal p-values < 0.05 are shown, while q-values < 0.05 (FDR-adjusted p-values) are marked with "x". **(B)** Machine learning results for separating treatment escalators from non-escalators, performed with the *gpboost* in Python 3.6.7 (see Methods). DNA dataset is applied in the upper comparison and the RNA dataset is applied in the lower comparison. IBD to the left, CD in the middle and UC to the right. Three different data sets were used as input in the models: microbial data, biochemical markers (Hbg, leukocytes, platelets, albumin, ALP, CRP and calprotectin), or a combination of the two. Model performance was estimated with area under the receiver operating curve (AUC), and compared between model categories using t-tests. * = p<0.05*** = p<0.001. **** = p<0.0001.

Abbreviations: HC, healthy controls; SC, symptomatic controls; IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis. Hbg, Blood-hemoglobin; Leucocytes, Blood-leucocytes; Platelets, Blood-platelets; Albumin, serum-albumin; ALP, serum-alkaline phosphatase; CRP, Serum-C-Reactive Protein; F-Calprotectin, fecal calprotectin.

As previously, we also performed an integrated analysis of RNA- and DNA-based microbiota profiles and clinical data to separate escalators from non-escalators using machine-learning methods, showing an excellent performance in UC (AUC > 0.80), but not in CD. However, in all models, the microbiota did not contribute to the predictive capacity (Figure 5B).

Finally, we also evaluated anti-TNF response for 29 IBD patients, 22 CD and seven UC, respectively. With a low sample size, the mucosal microbiota was assessed for IBD in total. In the DNA dataset, a high abundance of *Alloprevotella* was found among anti-TNF responders, while in the RNA dataset, a signature of high abundance of *Pseudomonas* and *Rhodanobacteraceae* combined with a low abundance of *Barnesiella* was associated with anti-TNF response ($q < 0.05$, Supplementary Figure 2).

Discussion

The present study identified differences between IBD patients and controls when assessing 16S rRNA at both the DNA and RNA levels, representing the total and transcriptionally active microbiota. The RNA dataset complemented the DNA dataset in separating CD from UC, supporting our hypothesis that the transcriptionally active microbiota can provide novel insights beyond the exploration of the total microbiota.

Microbial Differences Between IBD and Controls

In line with previous studies,[16,47,48] multiple bacterial genera differed between IBD and controls, with notable variations depending on whether 16S rRNA was assessed at the DNA or RNA level. These differences partly reflect variations in gut microbiota transcriptional activity between patients and controls.

We identified differences in gut microbiota composition between IBD and HC, including a reduced amount of bacteria producing short-chain fatty acids (SCFA) in IBD. Additionally, we observed a reduction of Actinobacteria, which is known to promote a healthy gut.[4,49–51] Furthermore, pathobionts within the phylum Proteobacteria were increased in non-inflamed biopsies of IBD patients, in both the active and total microbiota. Consistent with the findings by Rehman et al, UC exhibited increased abundance and activity of *Bacteroides*, while CD showed a reduction of *Ruminococcus* and *Lactobacillus*, both in terms of activity and presence.[47] Schirmer et al pointed to strains of *Alistipes* and *Bacteroides* being associated with IBD, particularly at the transcript level.[48] In our cohort, high levels of the genera *Alistipe*s and *Bacteroides* were similarly associated with IBD compared with HC.

There was no difference in alpha diversity of the active nor the total microbiota when comparing IBD to SC, indicating that decreased diversity is a feature associated with symptoms, regardless of diagnosis. Notably, IBD and SC differed compositionally only in the RNA dataset. This could indicate a benefit of assessing the active microbiota, as this difference was not detectable when sequencing DNA and hence the total microbiota. SC participants in this were mainly diagnosed retrospectively with IBS, and IBS is known to be associated with dysbiosis.[52] The similarity in the gut microbiota composition between IBD and SC supports the concept of IBS as a disease more than a "functional disorder" of the gut. In previous publications, we have demonstrated the fecal microbiota[53] and human intestinal wall gene expression[54] in SC to share similarities with IBD in the IBD-character cohort. Despite similarities in diversity, differential abundance analyses identified several taxa that differed significantly between IBD and SC ($q < 0.05$). The difference was more pronounced in CD, and mainly in the total microbiota. In particular, the signature was dominated by an overrepresentation of Proteobacteria, an underrepresentation of Actinobacteria, and shifts within the phylum Bacteroidetes. *Akkermansia* was overrepresented in CD in the total microbiota, but in the active microbiota this difference could not be found, indicating an overall inactivity of *Akkermansia* in CD. Interestingly, the overrepresentation of *Akkermansia* was only found in non-inflamed samples. Since *Akkermansia* uses colonic mucin as its substrate, the microbe may be lost with the mucin layer being reduced in inflamed areas.[55] *Pseudomonas* was found to be highly increased in CD compared to SC, in both the DNA and RNA datasets. *Pseudomonas* has previously been found to be overrepresented in ileal biopsies in pediatric CD compared to non-IBD controls.[56] Furthermore, studies of serological markers have demonstrated high levels of the *Pseudomonas* associated[57] T-cell superantigen I2 in CD.[58,59] Altogether, this points to a possible pathogenic role of *Pseudomonas* in IBD.

The microbiota signature of UC deviated from SC to a lesser extent than CD did. In UC, only three taxa were differentially abundant compared to SC, even when exploring the DNA and RNA dataset both independently and combined. Nevertheless, an increase in the facultative anaerobes of the family *Enterobacteriaceae* was found in UC, consistent with earlier findings in a subcohort of UC patients in the IBD-Character cohort.[16]

Microbial Changes Between CD and UC

We found differences in the mucosal microbiota of CD compared with UC patients, confirming previous studies.[6,8] *Rhodanobacteraceae, Alloprevotella* and *Pseudomonas* were all overrepresented in the active microbiota of CD compared to UC, suggesting a role in disease pathobiology. *Alloprevotella*, an anaerobe gram negative, is found in the oral cavity and has been associated with oral infectious disease,[60] suggesting a possible link between the oral microbiota and manifestations of CD more distally in the GI tract. Meanwhile, *Pseudomonas* has previously been suggested as a pathobiont in CD.[58,59] In contrast, the generally health-associated genera *Alistipes* and *Oscillospiraceae*[4,48] were transcriptionally more active in UC compared to CD.

Differential Bacterial Transcription Defined by RNA/DNA Ratio of Individual Taxa in IBD and Controls

We hypothesize that taxa with the largest discrepancies between their presence and transcriptional activity may contribute to our understanding of the pathobiology of IBD. In CD, one of the most striking discrepancies between the total and active microbiota was found for *Akkermansia*. A high proportion of *Akkermansia* appeared dormant, as it was underrepresented in the RNA dataset. In UC, there were fewer taxa separating the total and active microbiota. Nevertheless, several common findings emerged, such as Actinobacteria, *Akkermansia, Barnesiella* and *Pasteurellaceae* being underrepresented in RNA compared to DNA, resembling the findings from CD. These four taxa have common features promoting a healthy gut, and we can see that they are suppressed on the transcript level in CD and UC. The analysis of RNA/DNA ratio furthermore confirmed that the transcriptionally active gut microbiota differed from the total microbiota, in IBD as well as in controls. These results are supported by the literature, where some differences between presence and activity among microbes have been described.[14,48]

This study supports previous results from our group in a subcohort of UC patients from the IBD Character cohort[16] as *Ruminococcaceae* showed to be overrepresented in RNA vs DNA and Bacteroidetes was underrepresented in RNA vs DNA in both analyses. On phylum level, Firmicutes showed high activity while Bacteroidetes was rather inactive in IBD. Taking the Firmicutes/Bacteroidetes ratio into account, which have repeatedly been found to be reduced in IBD on the DNA level, our results suggest a compensatory mechanism where the changes in presence could be partially neutralized by effects on the transcription level. On the other hand, within the phylum Proteobacteria many taxa showed high transcriptional activity. The high presence and transcriptional activity of Proteobacterial in IBD suggests a significant role for this phylum in IBD pathobiology.

The Diagnostic and Prognostic Ability of the Microbial Data is Weaker Than Established Biomarkers

Although microbial signatures indicative of IBD were identified, their diagnostic utility remains limited. In a machine-learning analysis including the available microbial and clinical data, the ability to separate IBD from SC was high (AUC > 0.80). However, diagnostic ability was mainly driven by biomarkers already in clinical use, as f-calprotectin, albumin, Hb and CRP, while the contribution from microbiota was small, both in RNA and DNA datasets, compared to these parameters.

Beyond diagnostics, there is a clinical need for biomarkers and algorithms for disease development and treatment response. Multiple bacterial genera were associated with a need for treatment escalation during follow-up, in both CD and UC, but established clinical biomarkers outperformed the mucosal microbiota profile as predictive biomarker in the machine-learning models. Nonetheless, some microbial associations could be of interest as markers of key pathogenetic mechanisms. In the differential abundance analyses signature, there was a relative underrepresentation of taxa both in the total and active microbiota among escalators, pointing to a generally deteriorated microbiota in persons headed for

a severe disease course. Moreover, the CD escalation signature was dominated by reduced abundances of several taxa like *Prevotella, Rhodanobacteriaceae, Methanobrevibacter, Desulfovibrio* and *Pasteurellaceae. Methanobrevibacter*[21] and *Micrococcales*[4,8,49] have previously been shown to be reduced in IBD patients, and here we see the most reduced abundances among patients with the poorest prognosis. A high level of *Akkermansia* was associated with treatment escalation in CD. *Akkermansia* has been affiliated with IBD as a mucin-degrader and pathobiont,[55] and the observation underscores the possibility of a central role for this bacterial strain in IBD pathogenesis. In UC, high abundances of *Prevotella* could predict treatment escalation. Members of the *Prevotella* genus have been shown to exacerbate gut inflammation,[61–63] in line with our report of an overabundance of this genus in those UC patients with the poorest prognosis. Finally, microbiota at diagnosis was also associated with anti-TNF response later in the disease course. Previously, Ananthakrishnan et al found a stronger ability to predict anti-TNF response when combining microbiota signatures with clinical data.[20] In the present study, high abundances of *Alloprevotella, Pseudomonas* and *Rhodanobacteraceae* and a low abundance of *Barnesiella* were found at baseline among anti-TNF responders. The observations demonstrate how taxa involved in initiation of IBD, might play other roles in disease progression. Furthermore, 3 out of 4 hits in the anti-TNF responder signature originates from the RNA dataset, empowering the motivation of further exploration of the transcriptionally active microbiota in future studies.

Strengths and Limitations

A strength of this study is the access to inception cohorts in different European centers, where the participants are newly diagnosed and mainly treatment naïve, resulting in the microbial compositional changes still being in early stages of disease. Studying a treatment naïve population is of great potential value in giving insights into pathobiology at the stage of disease initiation. The analysis of microbial DNA and RNA from the same biopsy gives a more complex picture of the gut microbiota composition, where RNA represents the active members of the microbial community. Another strength of the study is the detailed microbial signatures based on the application of both inflamed and non-inflamed samples, and a determination of the mucosa-associated microbiota both on the DNA and RNA level in the same biopsy, which may contribute to a better understanding of the gut microbiota in the context of IBD.

There are several limitations in this study. The study size was too small for achieving satisfactory statistical power in several analyses, such as in creating models for predicting anti-TNF response. Only 29 IBD patients had simultaneous 16S data and anti-TNF response data, thus hindering separate analysis of CD and UC. Furthermore, participants were given laxatives as part of a routine protocol prior to the diagnostic colonoscopy where the biopsies were sampled. Laxatives may have both short- and long-term effects on the microbiome composition, and these effects are thought to be independent of stool consistency.[64,65] This introduces a potential source of bias, possibly masking results reflective of the natural microbiota state. A third limitation is that the study relies on 16S rRNA sequencing, which provides limited taxonomic resolution compared to whole-genome metagenomics.[66,67] The current study does not capture strain-level differences or functional potential, which could be relevant for understanding important features of the microbiota in IBD.

Conclusion

The mucosal IBD microbiota differs significantly from both healthy and symptomatic controls. Differences between abundance as well as transcriptional activity were found among multiple taxa when assessing 16S rRNA on DNA and RNA level, and the datasets complemented each other, building a comprehensive microbiota signature. A limited number of bacterial taxa were responsible for the largest difference between activity and presence (RNA/DNA ratio), among patients as well as among controls. The diagnostic and prognostic capacity of the microbiota were outcompeted by established clinical biomarkers in machine learning models.

Abbreviations

ASVs, (amplicon sequence variants); AUC, (area under curve); CD, (Crohn's disease); FDR, (false discovery rate); HC, (healthy controls); IBD, (inflammatory bowel disease); IBD-U, (inflammatory bowel disease unclassified); IBS, (irritable bowel syndrome); UC, (ulcerative colitis); SC, (symptomatic controls).

Data Transparency Statement

Clinical and descriptive data of the cohort, analytic methods and microarray raw data can be shared on request to the corresponding author s.s.vatn@studmed.uio.no with permission from the IBD Character data access committee.

Ethics

All centers were granted local ethics approval for this study and all patients gave written and informed consent prior to participating in this study.

In Norway, the study was approved by the Regional Committee for Medical Research Ethics, South-Eastern Norway, with reference REK sør-øst 2009/2015. In Sweden, the study was approved by the Uppsala Regional Ethics Committee, with reference 2010/313, and in the Netherlands by the Medical Ethics Review Committee. In Scotland, the study was approved by the Research Ethics Committee, with reference 15/ES/0094. The NHS Lothian Bioresource granted local approval for this study, with reference SR171. In Spain, the study was approved by Comité de Ética e Investigación Clínica de Aragón, with reference 12/00151.

All experiments were performed in accordance with and following the Declaration of Helsinki Principles. All methods were performed in accordance with the relevant guideline and regulations. All patients gave their written informed consent prior to colonoscopy and their inclusion in the IBD-Character study.

Acknowledgments

The authors thank the IBD-Character Consortium for contributions in study conception and design: Elaine R Nimmo, Hazel E Drummond, Ray K Boyapati, Nicholas T Ventham, Nicholas A Kennedy, Alex Adams, David C Wilson, Charles W Lees, Colin L Noble, Ian D Arnott, Gwo-Tzer Ho, Alan G Shand, Kate R O'Leary, Anna Frengen, Panpan You, Janne Sølvernes, Fredrik A Dahl, Jonas Christoffer Lindstrøm, Gunn S Ekeland, Haldor Husby, Johan D Söderholm, Henrik Hjortswang, Mauro D'Amato, Leif Törkvist, Christina Casén, Magdalena K Karlsson, Fredrik Hjelm, Mats Gullberg, Niklas Nordberg, Anette Ocklind, Erik Pettersson, Daniel Ekman, Mikael Sundell, Eddie Modig, Anne-Clémance Veillard, Renaud Schoemans, Dominique Poncelet, Céline Sabatel, Ivo G Gut, Marta Gut, Simon Heath, Monica Bayes, Angelika Merkel, and Ferdinando Bonfiglio.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work. All authors had access to the study data and reviewed and approved the final manuscript.

Funding

The study was funded by the following EU FP7 grant: IBD-CHARACTER (contract # 2858546). PR, SV and SHH were funded by South-Eastern Norway Regional Health Authority (project numbers 2014011, 2018001 and 2020066).

Disclosure

Dr Daniel Bergemalm reports personal fees from BMS, personal fees from pharmacosmos, personal fees from Janssen, personal fees from Takeda, personal fees from Sandoz, personal fees from Pfizer, personal fees from Sandoz, outside the submitted work. Dr Trond Espen Detlie reports personal fees from AbbVie, personal fees from Ferring, personal fees from Pfizer, personal fees from Pharmacosmos, personal fees from Takeda, personal fees from Tillotts, personal fees from CLS Vifor Pharma, outside the submitted work. Dr Rahul Kalla reports grants from MRC, grants from Crohn's and Colitis UK, outside the submitted work. Dr Jonas Halfvarson reports personal fees from AbbVie, personal fees from BMS, personal fees from Eli Lilly, personal fees from Alfasigma, personal fees from Aqilion, personal fees from Celltrion, personal fees from Ferring, personal fees from Galapagos, personal fees from Gilead, personal fees from Index Pharma, grants, personal fees from Janssen, personal fees from Medtronic, personal fees from Merck, grants, personal

fees from MSD, personal fees from Novartis, personal fees from Pfizer, personal fees from Prometheus Laboratories Inc., personal fees from Sandoz, personal fees from Shire, personal fees from STADA, grants, personal fees from Takeda, personal fees from Thermo Fisher Scientific, personal fees from Tillotts Pharma, outside the submitted work. Prof Dr Johannes Hov reports grants from South-Eastern Norway Regional Health Authority, during the conduct of the study. Dr Petr Ricanek reports grants from EU FP7 grant: IBD-CHARACTER (contract # 2858546), grants from South-Eastern Norway Regional Health Authority (project numbers 2014011, 2018001 and 2020066), during the conduct of the study. The authors declare no conflicts of interest related to this work.

References

1. Stange EF, Schroeder BO. Microbiota and mucosal defense in IBD: an update. *Expert Rev Gastroenterol Hepatol*. 2019;13(10):963–976. doi:10.1080/17474124.2019.1671822
2. Imhann F, Vich Vila A, Bonder MJ, et al. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*. 2018;67(1):108. doi:10.1136/gutjnl-2016-312135
3. Hold GL, Smith M, Grange C, et al. Role of the gut microbiota in inflammatory bowel disease pathogenesis: what have we learnt in the past 10 years? *World J Gastroenterol*. 2014;20(5):1192–1210. doi:10.3748/wjg.v20.i5.1192
4. Gevers D, Kugathasan S, Denson L, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15(3):382–392. doi:10.1016/j.chom.2014.02.005
5. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*. 2012;13(9):R79. doi:10.1186/gb-2012-13-9-r79
6. Bibiloni R, Mangold M, Madsen KL, et al. The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. *Journal of Medical Microbiology*. 2006;55(8):1141–1149. doi:10.1099/jmm.0.46498-0
7. Pascal V, Pozuelo M, Borruel N, et al. A microbial signature for Crohn's disease. *Gut*. 2017;66(5):813–822. doi:10.1136/gutjnl-2016-313235
8. Forbes JD, Van Domselaar G, Bernstein CN. Microbiome survey of the inflamed and noninflamed gut at different compartments within the gastrointestinal tract of inflammatory bowel disease patients. *Inflamm Bowel Dis*. 2016;22(4):817–825. doi:10.1097/MIB.0000000000000684
9. Zoetendal EG, von Wright A, Vilpponen-Salmela T, et al. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol*. 2002;68(7):3401–3407. doi:10.1128/AEM.68.7.3401-3407.2002
10. Ouwehand AC, Salminen S, Arvola T, et al. Microbiota composition of the intestinal mucosa: association with fecal microbiota? *Microbiol Immunol*. 2004;48(7):497–500. doi:10.1111/j.1348-0421.2004.tb03544.x
11. Swidsinski A, Ladhoff A, Pernthaler A, et al. Mucosal flora in inflammatory bowel disease. *Gastroenterology*. 2002;122(1):44–54. doi:10.1053/gast.2002.30294
12. Johansson ME. Mucus layers in inflammatory bowel disease. *Inflamm Bowel Dis*. 2014;20(11):2124–2131. doi:10.1097/MIB.0000000000000117
13. Blazewicz SJ, Barnard RL, Daly RA, et al. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME J*. 2013;7(11):2061–2068. doi:10.1038/ismej.2013.102
14. Heinsen F-A, Knecht H, Neulinger SC, et al. Dynamic changes of the luminal and mucosa-associated gut microbiota during and after antibiotic therapy with paromomycin. *Gut Microbes*. 2015;6(4):243–254. doi:10.1080/19490976.2015.1062959
15. Gosalbes MJ, Durbán A, Pignatelli M, et al. Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One*. 2011;6(3):e17447. doi:10.1371/journal.pone.0017447
16. Moen AEF, Lindstrøm JC, Tannæs TM, et al. The prevalence and transcriptional activity of the mucosal microbiota of ulcerative colitis patients. *Sci Rep*. 2018;8(1):17278. doi:10.1038/s41598-018-35243-4
17. Monstad I, Hovde O, Solberg IC, et al. Clinical course and prognosis in ulcerative colitis: results from population-based and observational studies. *Ann Gastroenterol*. 2014;27(2):95–104.
18. Solberg IC, Vatn MH, Høie O, et al. Clinical course in Crohn's disease: results of a Norwegian population-based ten-year follow-up study. *Clin Gastroenterol Hepatol*. 2007;5(12):1430–1438. doi:10.1016/j.cgh.2007.09.002
19. Hovde Ø, Småstuen MC, Høivik ML, et al. Mortality and causes of death in ulcerative colitis: results from 20 years of follow-up in the IBSEN study. *Inflamm Bowel Dis*. 2016;22(1):141–145. doi:10.1097/MIB.0000000000000582
20. Ananthakrishnan AN, Luo C, Yajnik V, et al. Gut microbiome function predicts response to anti-integrin biologic therapy in inflammatory bowel diseases. *Cell Host Microbe*. 2017;21(5):603–610.e3. doi:10.1016/j.chom.2017.04.010
21. Ghavami SB, Rostami E, Sephay AA, et al. Alterations of the human gut Methanobrevibacter smithii as a biomarker for inflammatory bowel diseases. *Microb Pathog*. 2018;117:285–289. doi:10.1016/j.micpath.2018.01.029
22. Franzin M, Stefančič K, Lucafò M, et al. Microbiota and drug response in inflammatory bowel disease. *Pathogens*. 2021;10(2):211. doi:10.3390/pathogens10020211
23. Borg-Bartolo SP, Boyapati RK, Satsangi J, Kalla R, et al. Precision medicine in inflammatory bowel disease: concept, progress and challenges. F1000Research 2020;9.
24. *IBD-Character consortium*. 2020; Available from: https://cordis.europa.eu/project/rcn/106191_en).htm. Accessed May 13, 2025.
25. Gomollón F, Dignass A, Annese V, et al. 3rd European evidence-based consensus on the diagnosis and management of crohn's disease 2016: part 1: diagnosis and medical management. *J Crohn's Colitis*. 2016;11(1):3–25. doi:10.1093/ecco-jcc/jjw168
26. Gionchetti P, Dignass A, Danese S, et al. 3rd European evidence-based consensus on the diagnosis and management of crohn's disease 2016: part 2: surgical management and special situations. *J Crohn's Colitis*. 2016;11(2):135–149. doi:10.1093/ecco-jcc/jjw169
27. Harbord M, et al. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. part 2. *CurrManag J Crohns Colitis*. 2017;11(7):769–784.

28. Moen AEF, Tannæs TM, Vatn S, et al. Simultaneous purification of DNA and RNA from microbiota in a single colonic mucosal biopsy. *BMC Res Notes*. 2016;9. doi:10.1186/s13104-016-2110-7
29. Kozich JJ, Westcott SL, Baxter NT, et al. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013;79(17):5112–5120. doi:10.1128/AEM.01043-13
30. James Kozich PS, Baxter N, Jenior M, Koumpouras C, Bishop L. *16S rRNA Sequencing with the Illumina MiSeq: Library Generation, QC, & Sequencing*. 2013; Available from: https://github.com/SchlossLab/MiSeq_WetLab_SOP/blob/master/MiSeq_WetLab_SOP.md. Accessed May 13, 2025.
31. Amir A, McDonald D, Navas-Molina JA, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems*. 2017;2(2). doi:10.1128/mSystems.00191-16.
32. Rognes T, Flouri T, Nichols B, et al. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584. doi:10.7717/peerj.2584
33. R Core Team. R: a language and environment for statistical computing. *R Foundation Statistical Computing*. 2023. Available from. https://www.R-project.org/.: Accessed 2025.
34. Oksanen J, Chen K, Vähänissi V, et al. vegan community ecology package version 2.6-4 April 2022. *The Journal of Physical Chemistry Letters*. 2022;13:5648–5653. doi:10.1021/acs.jpclett.2c01302
35. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217. doi:10.1371/journal.pone.0061217
36. Gpboost. 2024. doi:10.48550/arXiv.2004.02653
37. Rossum GV, Drake FL. *Python 3 Reference Manual*. CreateSpace; 2009.
38. Ushey K. Tang Y *_reticulate: Interface to* 'Python'*_. R package version 1.27*. 2023. Available from: https://CRAN.R-project.org/package=reticulate. Accessed May 13, 2025.
39. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nature Mach Intell*. 2020;2(1):56–67. doi:10.1038/s42256-019-0138-9
40. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1–48. doi:10.18637/jss.v067.i01
41. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol*. 2021;17(11):e1009442. doi:10.1371/journal.pcbi.1009442
42. Kassambara A. *ggpubr*: 'ggplot2' *Based Publication Ready Plots_. R package version 0.5.0*. 2023. Available from: https://CRAN.R-project.org/package=ggpubr. Accessed May 13, 2025.
43. Tiedemann F. *Gghalves: Compose Half-Half Plots Using Your Favourite Geoms_. R package version 0.1.4*. 2022. Available from: https://CRAN.R-project.org/package=gghalves. Accessed May 13, 2025.
44. Kay M. *ggdist: Visualizations of Distributions and Uncertainty*. 2022. Available from: https://mjskay.github.io/ggdist/. Accessed May 13, 2025.
45. Auguie B. Gridextra_gridExtra: miscellaneous Functions for "Grid" Graphics 2017.
46. Drossman DA, Hasler WL. Rome IV-functional gi disorders: disorders of gut-brain interaction. *Gastroenterology*. 2016;150(6):1257–1261. doi:10.1053/j.gastro.2016.03.035
47. Rehman A, Rausch P, Wang J, et al. Geographical patterns of the standing and active human gut microbiome in health and IBD. *Gut*. 2016;65(2):238–248. doi:10.1136/gutjnl-2014-308341
48. Schirmer M, Franzosa EA, Lloyd-Price J, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol*. 2018;3(3):337–346. doi:10.1038/s41564-017-0089-z
49. Chen L, Wang W, Zhou R, et al. Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease. *Medicine*. 2014;93(8):e51–e51. doi:10.1097/MD.0000000000000051
50. Alipour M, Zaidi D, Valcheva R, et al. Mucosal barrier depletion and loss of bacterial diversity are primary abnormalities in paediatric ulcerative colitis. *J Crohn's Colitis*. 2016;10(4):462–471. doi:10.1093/ecco-jcc/jjv223
51. Campieri M, Gionchetti P. Bacteria as the cause of ulcerative colitis. *Gut*. 2001;48(1):132–135. doi:10.1136/gut.48.1.132
52. Wang L, Alammar N, Singh R, et al. Gut microbial dysbiosis in the irritable bowel syndrome: a systematic review and meta-analysis of case-control studies. *J Acad Nutr Diet*. 2020;120(4):565–586. doi:10.1016/j.jand.2019.05.015
53. Vatn S, Carstens A, Kristoffersen A, et al. Faecal microbiota signatures of IBD and their relation to diagnosis, disease phenotype, inflammation, treatment escalation and anti-TNF response in a European Multicentre Study (IBD-Character). *Scand J Gastroenterol*. 2020;1–11.
54. Vatn SS, Lindstrøm JC, Moen AE, et al. Mucosal gene transcript signatures in treatment naïve inflammatory bowel disease: a comparative analysis of disease to symptomatic and healthy controls in the European IBD-Character Cohort. *Clin Exp Gastroenterol*. 2022;15:5–25. doi:10.2147/CEG.S343468
55. Earley H, Lennon G, Balfe Á, et al. The abundance of Akkermansia muciniphila and its relationship with sulphated colonic mucins in health and ulcerative colitis. *Sci Rep*. 2019;9(1):15683. doi:10.1038/s41598-019-51878-3
56. Wagner J, Short K, Catto-Smith AG, et al. Identification and characterisation of Pseudomonas 16S ribosomal DNA from ileal biopsies of children with Crohn's disease. *PLoS One*. 2008;3(10):e3578. doi:10.1371/journal.pone.0003578
57. Wei B, Huang T, Dalwadi H, et al. Pseudomonas fluorescens encodes the crohn's disease-associated i2 sequence and T-cell superantigen. *Infect Immun*. 2002;70(12):6567–6575. doi:10.1128/IAI.70.12.6567-6575.2002
58. Landers CJ, Cohavy O, Misra R, et al. Selected loss of tolerance evidenced by Crohn's disease-associated immune responses to auto- and microbial antigens. *Gastroenterology*. 2002;123(3):689–699. doi:10.1053/gast.2002.35379
59. Arnott ID, Landers CJ, Nimmo EJ, et al. Sero-reactivity to microbial components in Crohn's disease is associated with disease severity and progression, but not NOD2/CARD15 genotype. *Am J Gastroenterol*. 2004;99(12):2376–2384. doi:10.1111/j.1572-0241.2004.40417.x
60. Ulger Toprak N, Duman N, Sacak B, et al. Alloprevotella rava isolated from a mixed infection of an elderly patient with chronic mandibular osteomyelitis mimicking oral squamous cell carcinoma. *New Microbes and New Infect*. 2021;42:100880. doi:10.1016/j.nmni.2021.100880
61. Iljazovic A, Roy U, Gálvez EJC, et al. Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation. *Mucosal Immunol*. 2021;14(1):113–124. doi:10.1038/s41385-020-0296-4
62. Larsen JM. The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology*. 2017;151(4):363–374. doi:10.1111/imm.12760
63. Prasoodanan PKV, Sharma AK, Mahajan S, et al. Western and non-western gut microbiomes reveal new roles of Prevotella in carbohydrate metabolism and mouth-gut axis. *NPJ Biofilms Microbi*. 2021;7(1):77. doi:10.1038/s41522-021-00248-x

64. Weersma RK, Zhernakova A, Fu J. Interaction between drugs and the gut microbiome. *Gut*. 2020;69(8):1510–1519. doi:10.1136/gutjnl-2019-320204
65. Gorkiewicz G, Thallinger GG, Trajanoski S, et al. Alterations in the colonic microbiota in response to osmotic diarrhea. *PLoS One*. 2013;8(2): e55817–e55817. doi:10.1371/journal.pone.0055817
66. Tremblay J, Singh K, Fern A, et al. Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol*. 2015;6.
67. Rintala A, Pietilä S, Munukka E, et al. Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. *J Biomol Tech*. 2017;28(1):19–30. doi:10.7171/jbt.17-2801-003