

# UCLA

## UCLA Previously Published Works

### Title

Crohn's disease in endoscopic remission, obesity, and cases of high genetic risk demonstrate overlapping shifts in the colonic mucosal-luminal interface microbiome

### Permalink

<https://escholarship.org/uc/item/7f5258th>

### Journal

Genome Medicine, 14(1)

### ISSN

1756-994X

### Authors

Jacobs, Jonathan P  
Goudarzi, Maryam  
Lagishetty, Venu  
[et al.](#)

### Publication Date

2022

### DOI

10.1186/s13073-022-01099-7

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>


Peer reviewed

RESEARCH

Open Access



# Crohn's disease in endoscopic remission, obesity, and cases of high genetic risk demonstrates overlapping shifts in the colonic mucosal-luminal interface microbiome

Jonathan P. Jacobs<sup>1,2\*</sup>, Maryam Goudarzi<sup>3</sup>, Venu Lagishetty<sup>1</sup>, Dalin Li<sup>4</sup>, Tytus Mak<sup>5</sup>, Maomeng Tong<sup>6</sup>, Paul Ruegger<sup>7</sup>, Talin Haritunians<sup>4</sup>, Carol Landers<sup>4</sup>, Philip Fleshner<sup>4</sup>, Eric Vasiliauskas<sup>4</sup>, Andrew Ippoliti<sup>8</sup>, Gil Melmed<sup>4</sup>, David Shih<sup>4</sup>, Stephan Targan<sup>4</sup>, James Borneman<sup>7</sup>, Albert J. Fornace Jr<sup>9</sup>, Dermot P. B. McGovern<sup>4†</sup> and Jonathan Braun<sup>4†</sup> 

## Abstract

**Background:** Crohn's disease (CD) patients demonstrate distinct intestinal microbial compositions and metabolic characteristics compared to unaffected controls. However, the impact of inflammation and underlying genetic risk on these microbial profiles and their relationship to disease phenotype are unclear. We used lavage sampling to characterize the colonic mucosal-luminal interface (MLI) microbiome of CD patients in endoscopic remission and unaffected controls relative to obesity, disease genetics, and phenotype.

**Methods:** Cecum and sigmoid colon were sampled from 110 non-CD controls undergoing screening colonoscopy who were stratified by body mass index and 88 CD patients in endoscopic remission (396 total samples). CD polygenic risk score (GRS) was calculated using 186 known CD variants. MLI pellets were analyzed by 16S ribosomal RNA gene sequencing, and supernatants by untargeted liquid chromatography-mass spectrometry.

**Results:** CD and obesity were each associated with decreased cecal and sigmoid MLI bacterial diversity and distinct bacterial composition compared to controls, including expansion of *Escherichia/Shigella*. Cecal and sigmoid dysbiosis indices for CD were significantly greater in obese controls than non-overweight controls. CD, but not obesity, was characterized by altered biogeographic relationship between the sigmoid and cecum. GRS was associated with select taxonomic shifts that overlapped with changes seen in CD compared to controls including *Fusobacterium* enrichment. Stricturing or penetrating Crohn's disease behavior was characterized by lower MLI bacterial diversity and altered composition, including reduced *Faecalibacterium*, compared to uncomplicated CD. Taxonomic profiles including reduced *Parasutterella* were associated with clinical disease progression over a mean follow-up of 3.7 years. Random forest classifiers using MLI bacterial abundances could distinguish disease state (area under the curve (AUC) 0.93), stricturing or penetrating Crohn's disease behavior (AUC 0.82), and future clinical disease progression (AUC 0.74).

<sup>†</sup>Dermot P.B. McGovern and Jonathan Braun are co-senior authors.

\*Correspondence: [jjacobs@mednet.ucla.edu](mailto:jjacobs@mednet.ucla.edu)

<sup>2</sup>Division of Gastroenterology, Hepatology and Parenteral Nutrition, Veterans Administration Greater Los Angeles Healthcare System, Los Angeles, USA  
Full list of author information is available at the end of the article



CD patients showed alterations in the MLI metabolome including increased cholate:deoxycholate ratio compared to controls.

**Conclusions:** Obesity, CD in endoscopic remission, and high CD genetic risk have overlapping colonic mucosal-luminal interface (MLI) microbiome features, suggesting a shared microbiome contribution to CD and obesity which may be influenced by genetic factors. Microbial profiling during endoscopic remission predicted Crohn's disease behavior and progression, supporting that MLI sampling could offer unique insight into CD pathogenesis and provide novel prognostic biomarkers.

**Keywords:** Crohn's disease, Obesity, Microbiome, Mucosal-luminal interface, Disease behavior, Disease progression, Genetic risk score

## Background

Crohn's disease (CD) is a chronic inflammatory bowel disease (IBD) of the digestive tract with substantial morbidity and mortality due to intestinal ulceration, obstruction, and perforation leading to malnutrition, infections, and debilitating symptoms [1]. Prevalence is substantial (~0.3%) but stabilizing in the US and other industrialized countries, while incidence is rising globally across newly westernizing countries [2]. The pathophysiology of CD involves a combination of genetic susceptibility (concordance in twins of 30–50%, and over 200 genetic loci of genome-wide significance [3]) and environmental risk factors [2, 4]. The intestinal microbiome was initially implicated in disease development based on animal models (e.g., [5–7]). When extended to studies of human microbiome (see foundational studies [8–16]; reviewed in [17]), CD patients compared to healthy populations are distinguished by lower microbial diversity, expansion, and (more notably) depletion of >30 microbial taxa, temporal instability [18–20], and altered production of many classes of metabolites and proteins with emerging documentation of their ecologic and host bioactivity [21–23]. These findings strongly support the hypothesis that CD involves a perturbation of the microbiome, termed dysbiosis, that synergizes with environmental and genetic risk factors to incite intestinal inflammation [24, 25].

Existing studies have provided insight into the link between the microbiome and CD but many questions remain. First, while both host genetics and gut microbiome composition are associated with CD, genome-wide analysis has uncovered very limited impact of host genetics on microbial composition [26–29]. This has led to the view that mechanisms of CD genetic risk act in parallel rather than upstream of microbiome disease traits. However, the advent of genome-wide CD polygenic risk score methodologies (GRS) offers a novel and more statistically powerful method to reconsider the relationship of host genetics to disease-associated traits such as microbiome composition, by testing them in healthy populations independent of CD disease state [30–32]. Second, CD is heterogeneous in phenotype, including diverse

location patterns as well as clinically distinct categories of disease behavior summarized under the Montreal classification as stricturing (B2), internal penetrating (B3), and non-stricturing/non-penetrating (B1) [33]. The contributions of the microbiome to disease behavior have received limited attention to date [20, 34, 35]. Third, most studies of healthy control populations have used feces to characterize the microbiome. The mucosal microbiome of non-IBD individuals and its relation to other microbiome-associated conditions such as obesity and metabolic disorders [36, 37] or to genetic factors remain largely uncharacterized. This is an important limitation as studies comparing fecal and mucosal samples have found that the fecal microbiome is not representative of the mucosal microbiome, and more robust microbial differences have been found in the tissue microbiome than the fecal microbiome when comparing CD to controls [11, 38]. Fourth, existing moderate- to large-sized studies of the CD microbiome have included patients with active disease or defined remission by clinical parameters, which have been shown in CD to correlate poorly with inflammation measured by biomarkers and endoscopic appearance [11–14, 39, 40]. As inflammation itself appears to alter the gut microbiome as seen in animal models and in human studies associating severity of inflammation and therapeutic response to microbial profiles, it is unclear to what extent the reported dysbiosis associated with CD truly reflects an underlying disease association as opposed to microbial markers of ongoing intestinal inflammation [11, 12, 41].

We undertook a study to address these gaps in our understanding of the colonic mucosal microbiome of CD in remission. In deciding which specimen types to include, we focused on the mucosal-luminal interface (MLI). This represents colonic mucus and adherent microbes present after bowel preparation which can be released by lavage of the colonic surface during colonoscopy [42]. This provides insight into microbes in close proximity to the mucosa that have been shown to be distinct from those in the lumen and that may be most biologically pertinent for regulating host mucosal

phenotype. A study comparing colonic lavage samples and colonic biopsies found that they had roughly comparable microbial profiles [43]. Unlike biopsies, which predominantly contain human tissue, MLI sampling yields a mix of bacterial and human material that we and others have shown is amenable for metabolomics, proteomics, shotgun metagenomics, and viromics [42, 44–49]. MLI proteomics, in particular, has been demonstrated to yield panels of proteins that can differentiate IBD from non-IBD as well as IBD phenotypes [45, 47, 48]. Microbiome sequencing of MLI samples has been utilized to identify microbial associations with type 2 diabetes, systemic sclerosis, and colorectal cancer [50–52].

In this study, we performed cross-sectional sampling of the MLI microbiome of CD patients in endoscopic remission and non-IBD controls, stratified by obesity status. This allowed for assessment of mucosal microbiome profiles in obesity and adjustment for the confounding effects of obesity in comparisons of CD to controls.

**Methods**

**Cohort recruitment and sample collection**

Eighty-eight CD patients undergoing colonoscopy and 110 controls without IBD undergoing screening colonoscopy were recruited from endoscopy suites at Cedars-Sinai Medical Center between 6/29/2011 and 3/19/2014. The demographic details of this cohort are summarized in Table 1. Details of cohort recruitment strategy and sampling methodology were previously reported [42, 44, 45]. For convenience, we highlight the following details. All CD patients were reported by their gastroenterologists to be in clinical remission at the time of collection. The participating study endoscopists, who were all experienced IBD specialists, confirmed that study patients were in endoscopic remission at the time of sampling based upon the absence of visible signs of active inflammation such as linear or aphthous ulceration or cobblestoning. Controls undergoing screening colonoscopy were validated by normal endoscopic appearance of the colon during colonoscopy. The subjects underwent lavage of the sigmoid colon and cecum at sites without visible pooled luminal content with 30 mL of sterile 0.9% saline passed through the endoscope channel. The lavaged content was aspirated by vacuum suction into a collection trap, typically yielding over 20 mL, then immediately transferred to ice. Within the same day, the lavaged samples were centrifuged at 4000g for 30 min to separate the sample into a pellet and supernatant, which were stored at -80°C until future microbiome and metabolomics analysis, respectively. A clinical research coordinator collected relevant metadata including age, gender, body mass index (BMI), and Montreal classification (disease behavior and disease location). Clinical chart review for longitudinal

**Table 1** Cohort demographics. Continuous variables are shown as median (interquartile range). Significance of demographic data was determined by Fisher’s exact test for categorical data and the Wilcoxon rank-sum test for continuous data

	CD (n=88)	Control (n=110)	p-value
Gender			
Male	47%	70%	0.001
Female	53%	30%	
Race/ethnicity			
Caucasian	91%	95%	0.78
Hispanic	3%	0%	
African-American	2%	1%	
Asian/Pacific Islander	3%	2%	
Non-white Hispanic	3%	3%	
Age	41 (29–53)	64 (57–73)	2 × 10 <sup>-16</sup>
BMI	23.5 (21.0–26.2)	25.8 (23.6–27.8)	0.0001
GRS	0.35 (-0.22–1.0)	-0.07 (-0.80–0.68)	0.003
Age at diagnosis	24 (19–33)		
Duration (years)	11 (6–17)		
CD location			
L1 = Ileal	15%		
L2 = Colonic	11%		
L3 = Ileocolonic	74%		
Upper GI involvement	8%		
Perianal disease	26%		
CD behavior			
B1 = Non-stricturing	34%		
B2 = Stricturing	32%		
B3 = Penetrating	34%		
Medication			
5-aminosalicylate	47%		
Immunomodulator	29%		
Biologic (anti-TNF)	57%		

outcomes was performed in October 2017. Peripheral blood was collected from all patients for genetic analysis. The Cedars-Sinai Medical Center Institutional Review Board approved this research and the protocols governing participants (IRB #3358). All subjects provided informed consent to participate. The following datasets were generated for this study: 16S rRNA gene sequencing (CD—88 sigmoid lavage samples, 88 cecum lavage samples; non-IBD—110 sigmoid lavage samples, 109 cecum lavage samples), global untargeted metabolomics (CD—87 sigmoid lavage samples, 86 cecum lavage samples; non-IBD—108 sigmoid lavage samples, 105 cecum lavage samples), and genetic risk score (75 CD blood samples

and 97 non-IBD blood samples). Details of data production are provided in the “Methods” subsections below.

#### Genetic risk score

DNA was extracted from peripheral blood and applied to ImmunoChip, a custom platform containing nearly 200,000 single-nucleotide polymorphisms (SNPs) near genes related to immune function and inflammatory disease [30]. Quality control for genotype data was performed as previously described [30]. Gene risk scores (GRS) were calculated using the weighted sum of 186 SNPs associated with CD, including SNPs that are also associated with ulcerative colitis [30–32]. The SNPs were assumed to be independently associated with risk. For each SNP, an additive genetic model was calculated then the log odds ratio was multiplied by the number of corresponding risk alleles (0, 1, or 2). The GRS was calculated by summing these products across all genes according to the following formula:  $GRS = \sum_i \beta_i G_i$ .

#### 16S rRNA gene sequencing

Genomic DNA was extracted from the 395 pelleted samples (a single cecal lavage pellet was lost during handling) using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) with a 30-s beat-beating step in a Mini-Beadbeater-16 (BioSpec Products, Bartlesville, OK, USA) [53]. Polymerase chain reaction amplification of bacterial 16S rRNA genes was performed using PCR primers (F515/R806) targeting the V4 hypervariable region, with the reverse primers including a Golay barcode [54]. PCR products were purified using the MinElute 96 UF PCR Purification Kit (Qiagen, Valencia, CA, USA). DNA sequencing (100 bp reads) was performed using an Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA, USA) as previously described [55]. Raw sequence data was demultiplexed and filtered in QIIME v1.9.1 using `split_libraries_fastq` with  $q=19$  [56]. Deblur v1.1.0 was used with default parameters and min-reads 10 to denoise the data into amplicon sequence variants (ASVs) [57]. Taxonomy was assigned using the RDP classifier implemented in the `assignTaxonomy` function of the R package `dada2` v1.16.0 and the Silva v138 database [58, 59]. Three samples with fewer than 50,000 sequences were excluded from the analysis. The sequence depth of the remaining samples ranged from 50,862 to 821,276, with a mean depth of 411,244.

#### Microbiome data analysis

Alpha diversity was assessed using Chao1 and Shannon index with data rarefied to 50,862 sequences using the `estimate_richness` function of `Phyloseq` v1.32.0 in R v4.0.2 [60]. Statistical significance was assessed using multivariate ANOVA with post hoc Tukey implemented

with the `aov` and `TukeyHSD` functions in R v4.0.2. Beta diversity analysis was performed with the `vegdist` function of the R package `vegan` v2.5-6 using Bray-Curtis dissimilarity [61]. These results were visualized by principal coordinates analysis (`pcor` function in the R package `ape` v5-4.1); ellipses representing 95% confidence intervals were added using the `stat_ellipse` function of `ggplot2` v3.3.5 in R. Permutational multivariate analysis of variance using distance matrices was performed with the `adonis` function in `vegan` v2.5-6 with 100,000 permutations to determine statistical significance of differences in beta diversity [62].

Differential abundance testing was performed using non-rarefied 16S rRNA sequence data filtered to remove ASVs present in less than 25% of samples. The resulting filtered datasets were analyzed using DESeq2 v1.28.1 implemented through `Phyloseq` v1.32.0 in R [60, 63]. This algorithm performs normalization using size factors estimated by the median-of-ratios method, employs an empirical Bayesian approach to shrink dispersion, and fits the data to multivariate negative binomial models [64].  $p$ -values for variables in the linear models (e.g., IBD status) were converted to  $q$ -values using `qvalue` v2.20.0 in R v4.0.2 to correct for multiple hypothesis testing [65]. ASVs with  $q$ -values below 0.05 or 0.1 (for GRS analyses) and mean normalized relative abundance  $> 10^{-5}$  were considered significant. Similar analyses were also performed with the 16S rRNA sequence data summarized at the phylum level. Dysbiosis indices represent the log of the sum of relative abundances of taxa significantly enriched in CD by DESeq2 models divided by the sum of relative abundances of taxa depleted in CD.

To assess functional capacity of the microbiome for bile acid metabolism, the metagenome was predicted as relative abundances of KEGG orthologies using the `PICRUSt2` function of `QIIME2` v2019.10 [66]. Bacterial genes involved in bile acid 7 $\alpha$ -dehydroxylation were identified based on the `MetaCyc` annotation for this pathway and matched to the corresponding KEGG orthologies: *baiB* (K15868), *baiA* (K15869), *baiCD* (K15870), *baiF* (K15871), *baiE* (K15872), and *baiN* (K07007) [67]. Relative abundances of these genes were added to generate an overall relative abundance for the bile acid 7 $\alpha$ -dehydroxylation pathway.

#### Ultra-performance liquid chromatography-mass spectrometry (LC-MS) untargeted metabolomics

Frozen aliquots of MLI supernatant from 386 samples underwent untargeted metabolomics analysis. Oasis MCX solid-phase extraction sorbents (Waters) were used to remove any potential interfering clinical materials such as polyethylene glycol. The aliquots were then sonicated at 37 °C for 90 s, chilled on ice, mixed in 150

$\mu\text{L}$  of Optima acetonitrile containing internal standards, 4-nitrobenzoic acid and debrisoquine, then centrifuged. Supernatant was placed in a new tube, dried under a gentle stream of  $\text{N}_2$ , and resuspended in solvent A (98% water, 2% acetonitrile, and 0.1% formic acid) for LC-MS. The MS analysis was performed by injecting 5  $\mu\text{L}$  of each sample into a reverse-phase 50  $\times$  2.1 mm H-class ultra-performance liquid chromatography Acquity 1.7- $\mu\text{M}$  BEH C18 column (Waters) coupled to a time-of-flight mass spectrometry. The mobile phase consisted of solvent A and 100% acetonitrile containing 0.1% formic acid (solvent B). The Premier Q-TOF Xevo G2-S mass spectrometer (Waters) was operated in the positive (ESI+) and negative (ESI-) electrospray ionization modes scanning a 50–1200  $m/z$  range. The following 13-min gradient was used: 95%/5% solvent A/solvent B at 0.45 ml/min for 8 min, 50%/50% solvent A/solvent B for 2 min, 2%/98% solvent A/solvent B for 2 min, and 95%/5% solvent A/solvent B for the remaining 1 min. The lock-spray consisted of leucine-enkephalin (556.2771  $[\text{M}+\text{H}]^+$  and 554.2615  $[\text{M}-\text{H}]^-$ ). The MS data were acquired in centroid mode and processed using MassLynx software (Waters Corp, Milford, MA, USA) to construct a data matrix consisting of the retention time,  $m/z$ , and intensity (via the peak area normalized to protein concentration) for each ion. A total of 4441 ions were detected in the two acquisition modes. Our in-house statistical analysis program was used to putatively identify ions, utilizing the Human Metabolome Database (HMDB), LipidMaps, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and BioCyc allowing for the following adducts within a 20 parts per million (ppm) mass error window:  $\text{H}^+$ ,  $\text{Na}^+$ , and/or  $\text{NH}_4^+$  in the ESI+ mode;  $\text{H}^-$  and  $\text{Cl}^-$  in the ESI- mode [68].

#### Metabolomics data analysis

Raw peak intensities for spectral features present in greater than 10% of samples underwent KNN imputation in R with  $k=10$  using an R script provided with a publication [69]. The resulting imputed datasets underwent further analysis using the MetaboAnalyst R package v3.0.3 [70]. Filtering was performed by the “iqr” (interquartile range) criteria, data was log-transformed, and data underwent quantile normalization. Batch effect correction was performed using Combat as implemented in the PerformBatchCorrection function of MetaboAnalyst v3.0.3 [71]. Data was visualized by principal coordinates analysis of Euclidean distances. Differential abundance of spectral features was assessed by multivariate general linear models implemented in the limma package v3.44.3 in R v4.0.2 incorporating gender and obesity.  $p$ -values obtained from limma were processed using the mummichog function of MetaboAnalyst v3.0.3 for identification

of metabolic pathways enriched in differentially abundant features [72]. Pathway enrichment  $p$ -values were estimated using 100 permutations with gamma modeling of the permutations. Microbe-metabolite correlations were calculated with the cor function in R v4.0.2 using Spearman’s correlation of residuals from multivariate models implemented in limma (metabolites) and DESeq2 (microbes).

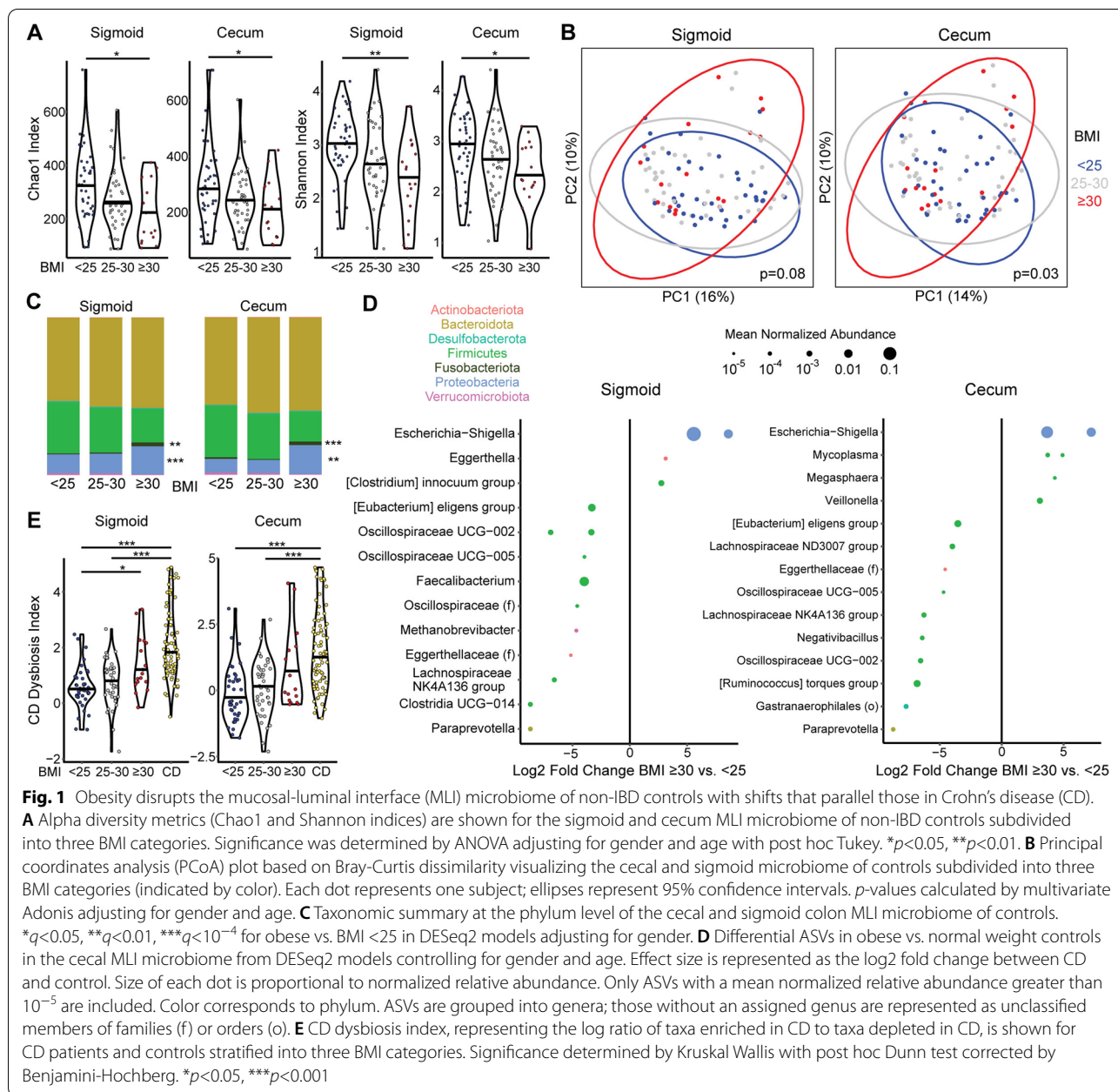
#### Random forests classifiers

Microbiome and metabolomics datasets were split 60%/40% into training and test subsets, respectively, using the createDataPartition function of the caret package v6.0-88 in R v4.0.2 [73]. Random forests classifiers were created using the train function of caret with  $mtry=2$  and 1001 trees [74]. Features were inputted into the algorithm if they were significantly associated with the trait of interest in multivariate models. An initial classifier was created then all features with an importance score greater than 2 in the preliminary classifier were used to construct a refined classifier with fewer features. The accuracy of the final random forests classifiers was assessed using the test subset of the data with confidence intervals determined by bootstrapping. This was performed with the roc.test function of the pROC package v1.16.2 in R [75].

## Results

### Obesity is associated with reduced bacterial diversity and pathobiont expansion in the MLI microbiome

Individuals without IBD undergoing screening colonoscopy were recruited as a control group for this study. The majority were overweight or obese, presenting an opportunity to investigate mucosal microbiome profiles of obesity to build on the existing literature on fecal microbiome alterations and investigate whether obesity is a confounding factor in CD vs. non-IBD comparisons [36]. Obesity status was represented by BMI categories based on cutoffs of 25 and 30 for overweight and obese, respectively. This approach was taken to increase applicability to the CD cohort, which includes both overweight/obese subjects as well as a significant subset who are underweight as a consequence of disease. This would render invalid any linear relationships of the microbiome with BMI derived from an overweight population control cohort. Bacterial diversity was found to be significantly reduced in the sigmoid and cecum of obese non-IBD subjects compared to those with  $\text{BMI}<25$  by metrics of richness alone (Chao1 index) and richness combined with evenness (Shannon index) (Fig. 1A). Overweight subjects ( $\text{BMI } 25\text{--}30$ ) had intermediate bacterial diversity that did not significantly differ from either the obese or the  $\text{BMI}<25$  groups. Differences in bacterial composition



across samples were assessed by Bray-Curtis dissimilarity and visualized by principal coordinate analysis (PCoA) (Fig. 1B). Obesity category (BMI <25, BMI 25–30, or BMI >30) was significantly associated with variation in Bray-Curtis dissimilarity across non-IBD samples by adonis after adjusting for gender and age in the cecum but did not reach significance in the sigmoid. Analysis of only subjects with BMI >30 vs. those with BMI <25 demonstrated significant associations of obesity category with microbial composition at both sites ( $p = 0.007$  in the cecum,  $p = 0.005$  in the sigmoid). At the phylum level,

obese subjects had increased abundance of Proteobacteria and Fusobacteriota in both the sigmoid and cecum as compared with subjects with BMI <25 (Fig. 1C). Differentially abundant amplicon sequence variants (ASVs) were identified using DESeq2, an algorithm that employs an empirical Bayesian approach to shrink dispersion and fits the data to negative binomial models. Obesity as compared with BMI <25 was associated with marked expansion of a highly abundant *Escherichia-Shigella* ASV in both the sigmoid and cecal MLI as well as expansion of *Mycoplasma* ASVs in the cecum, with depletion of ASVs

largely belonging to genera within the Firmicutes phylum such as *Faecalibacterium* (Fig. 1D).

### The MLI microbiome of CD patients in endoscopic remission is characterized by lower bacterial diversity, pathobiont enrichment, and depletion of anti-inflammatory microbes

All CD subjects included in this study were confirmed to be in endoscopic remission at the time of sample collection. There were equal numbers of each of the three disease behaviors: 34% with non-stricturing, non-penetrating CD (B1), 32% with stricturing disease (B2), and 34% with internal penetrating (B3) disease (Table 1). Most CD subjects (74%) had ileocolonic (L2) disease with 15% having ileal (L1) and the remaining 11% having colonic (L3) disease. The majority were on treatment with a biologic (57%; all of these were receiving an anti-TNF agent including infliximab, adalimumab, or certolizumab), with 47% of patients receiving 5-aminosalicylates (47%) and 29% an immunomodulatory agent (thiopurine or methotrexate).

MLI samples from the cecum and sigmoid colon underwent high-depth 16S rRNA gene sequencing (mean 411,244 sequences per sample) to detect low abundance taxa that may distinguish CD from controls as well as demarcate phenotypic subsets of CD. All analyses were controlled for obesity category (non-overweight, overweight, or obese) given its association with MLI bacterial diversity and composition in controls. The cecal and sigmoid MLI microbiota were found to have lower bacterial diversity in CD patients compared to controls by the Chao1 index of richness ( $p=6 \times 10^{-9}$  for sigmoid,  $p=5 \times 10^{-8}$  for cecum; adjusted for age, gender, and obesity status) and the Shannon index of richness and evenness ( $p=9 \times 10^{-8}$ ,  $p=1 \times 10^{-6}$  adjusted for age, gender, and obesity) (Fig. 2A). Beta diversity analysis demonstrated a highly significant association of CD status with bacterial composition in both the sigmoid and cecum MLI adjusting for gender, age, and obesity status (Fig. 2B).

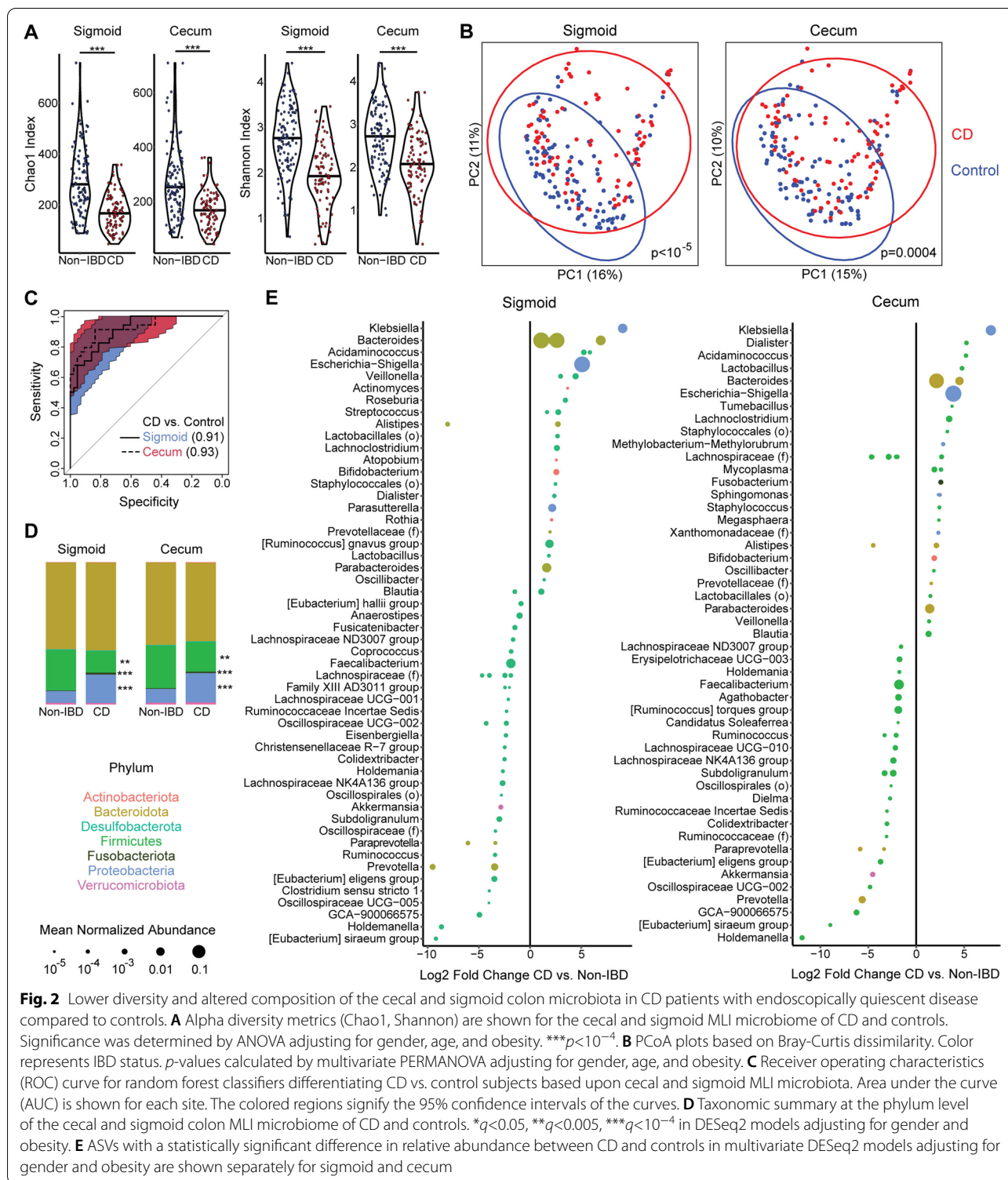
To assess whether MLI bacterial profiles could be used to predict CD vs. control status, the cohort was divided into training and test subsets and random forest classifiers were created using the training data. Classifiers based on sigmoid and cecal MLI bacterial profiles had high accuracy to differentiate CD vs. controls when applied to the test data, with area under the curve of 0.91 (95% confidence interval (CI) 0.85–0.96) and 0.93 (95% CI 0.86–0.98), respectively (Fig. 2C). At the phylum level, CD was characterized by expansion in both the sigmoid and cecal MLI of Proteobacteria ( $p=3 \times 10^{-22}$ ,  $p=4 \times 10^{-7}$ ) and Fusobacteriota ( $p=3 \times 10^{-15}$ ,  $p=4 \times 10^{-6}$ ) with concomitant reduction in Firmicutes ( $p=0.002$ ,  $p=0.009$ ) (Fig. 2D). At the ASV level, CD was associated

with enrichment of 31 ASVs in the sigmoid and 31 ASVs in the cecum, with depletion of 52 ASVs in the sigmoid and 41 ASVs in the cecum (Fig. 2E). The most strongly enriched ASV in CD in both the sigmoid and cecum was identified as a member of the *Klebsiella* genus. There was also strong enrichment of a highly abundant ASV identified as belonging to *Escherichia/Shigella* and several abundant ASVs identified as *Bacteroides*. Additional potential pathobionts were enriched including *Fusobacterium*, *Staphylococcus*, *Streptococcus*, *Rothia*, and *Mycoplasma* spp. Conversely, there was depletion of members of *Akkermansia*, *Prevotella*, and many genera within the Firmicutes phylum including *Faecalibacterium*. These taxonomic shifts in the CD MLI microbiome were summarized by CD dysbiosis indices, generated by calculating the log ratio of relative abundances of taxa enriched in CD to abundances of taxa depleted in CD. CD patients had highly significant increases in cecal and sigmoid dysbiosis indices relative to controls with BMI <25 or 25–30 but were not statistically different from obese controls (Fig. 1E). Interestingly, CD dysbiosis index was significantly greater in obese subjects compared to those with BMI <25 in the sigmoid ( $p=0.01$ ) and trended higher in the cecum ( $p=0.07$ ), consistent with taxonomic shifts in obesity paralleling those in CD.

### Biogeographic differences between the cecal and sigmoid colon MLI microbiome are disrupted in CD

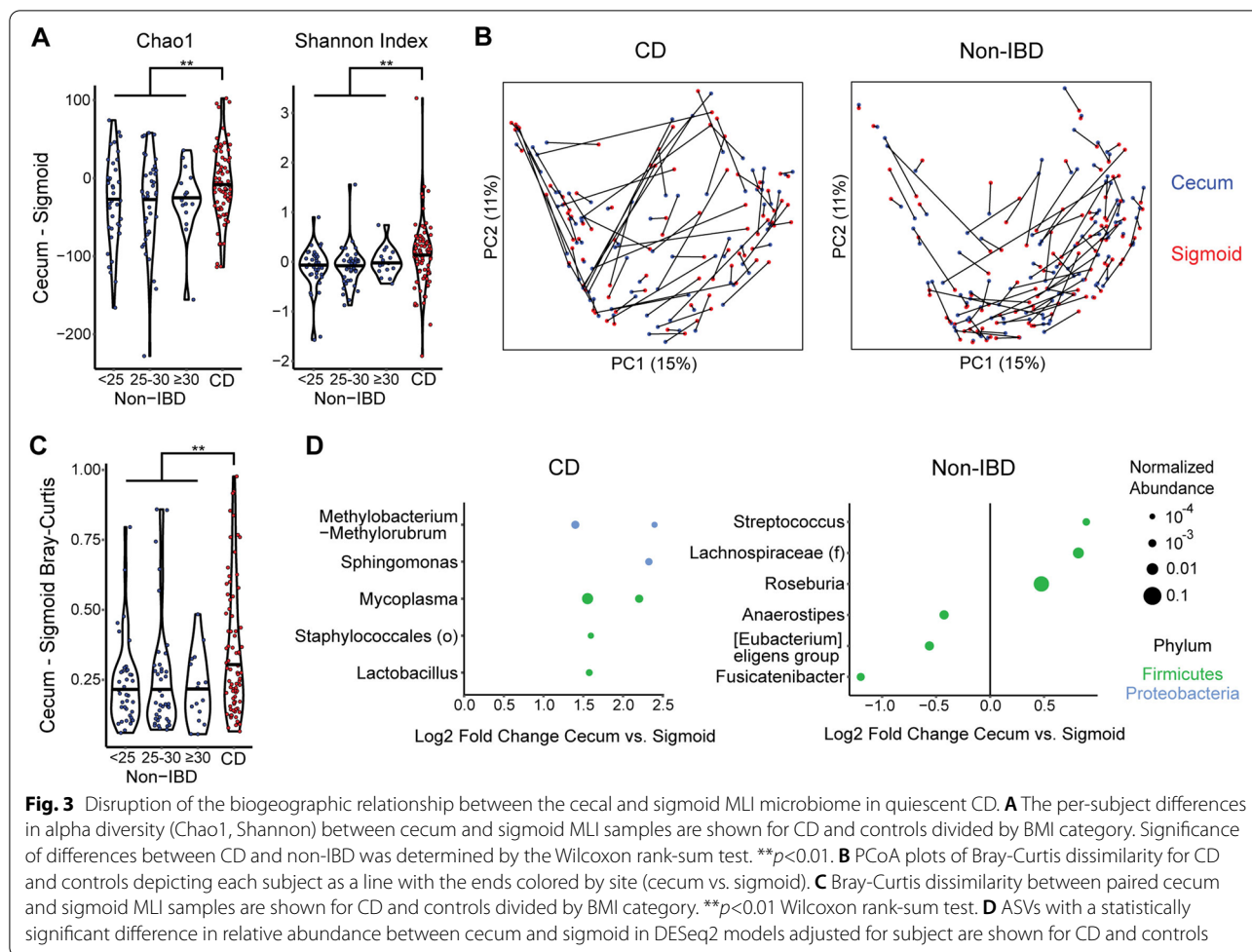
We next investigated what MLI microbial features differentiated the cecum and sigmoid colon and whether these biogeographic differences were affected by CD. First, we evaluated differences in alpha diversity between cecal and sigmoid MLI samples in a paired analysis. The cecum of non-IBD controls had reduced richness compared to the sigmoid by the Chao1 index ( $p=9 \times 10^{-7}$ ) and a non-significant trend towards reduced Shannon index ( $p=0.08$ ). CD subjects, in contrast, had no significant difference in microbial diversity between the cecum and sigmoid in paired analysis. Moreover, the pairwise differences in cecum vs. sigmoid bacterial diversity of CD subjects were significantly higher than those of non-IBD controls by both the Chao1 ( $p=0.008$ ,  $-6.5$  vs.  $-28.8$ ) and Shannon indices ( $p=0.006$ ,  $0.13$  vs.  $-0.05$ ) (Fig. 3A). This indicates increased relative bacterial diversity in the cecum compared to the sigmoid in CD even though CD was characterized overall by reduced diversity compared to controls. There were no differences across BMI categories within the non-IBD controls. An analysis was then performed of pairwise differences in beta diversity between the cecum and sigmoid colon using Bray-Curtis dissimilarity. Substantial intra-individual variation was seen between cecal and sigmoid bacterial composition, which was significantly greater in CD patients than in controls ( $p=0.005$ )





(Fig. 3B, C). This suggests that Crohn's disease, even in remission, results in reconfiguration of the biogeographic relationship to not only reverse the alpha diversity gradient seen in controls but also increase the difference in

bacterial composition between the cecum and sigmoid colon. In comparison, no difference was seen across BMI categories within the non-IBD controls. In non-IBD controls, 7 ASVs significantly differed between cecum and



sigmoid adjusting for subject, including cecal enrichment of *Streptococcus* and *Roseburia* ASVs and depletion of an *Anaerostipes* ASV (Fig. 3D). CD subjects had 6 differential ASVs between sigmoid and cecum that did not overlap with the ASVs seen in non-IBD and included enrichment of *Mycoplasma*, *Sphingomonas*, and *Lactobacillus* spp. in the cecum.

**CD genetic risk is associated with alterations in MLI microbiota**

The inclusion in this cohort of both CD in endoscopic remission and non-IBD controls provided an opportunity to investigate the effects of CD-associated genetic risk factors on the MLI microbiome. As this study was not powered to individually assess the greater than 200 loci genetically linked to CD, we instead used the strategy of summarizing genetic risk in subjects with a polygenic risk score. This score was calculated by summing the log odds ratios from 186 single-nucleotide polymorphisms associated with CD. As anticipated, this genetic risk score (GRS) was significantly higher in the CD subjects than

non-IBD controls (Table 1). GRS was also significantly higher in complicated CD than uncomplicated CD; uncomplicated CD and non-IBD controls had equivalent GRS (Fig. 4A). All GRS analyses involving CD subjects were therefore adjusted for disease behavior. There was no significant association of GRS with alpha diversity in non-IBD and CD subjects analyzed separately, adjusting for gender, age, and CD disease behavior (for CD subjects) (Fig. 4B). GRS was significantly associated with cecal but not sigmoid MLI microbial composition in CD subjects ( $p=0.04$ ) when evaluated as a continuous variable in adonis analysis of Bray-Curtis dissimilarity adjusted for gender, age, and CD behavior (Fig. 4C). There was no association of GRS with variation in Bray-Curtis dissimilarity in adonis analyses of non-IBD controls. Similarly, GRS was associated with CD dysbiosis index in the cecum of CD subjects ( $p=0.02$ ) but the trend did not reach significance in the sigmoid and there was no association in non-IBD controls (Fig. 4D). Differential abundance testing demonstrated an association of higher GRS with taxonomic shifts including enrichment

of a *Fusobacterium* ASV in the cecum of CD subjects and both the cecum and sigmoid of non-IBD controls, as well as depletion of a *Prevotella* ASV in the sigmoid and cecum of both CD and non-IBD (Fig. 4E). Interestingly, an *Akkermansia* ASV was positively associated with GRS in both the cecum and sigmoid of non-IBD controls but was negatively associated with GRS in the cecum of CD subjects. In CD patients, ASVs associated with higher GRS significantly overlapped with ASVs that were enriched or depleted in CD compared to non-IBD controls in both the cecum and sigmoid (Fig. 4F). In non-IBD controls, ASVs enriched with higher GRS significantly overlapped with ASVs enriched in CD compared to controls in the cecum.

#### CD complicated by penetrating or stricturing disease is characterized by increased MLI dysbiosis

Having confirmed that CD is characterized by distinct MLI bacterial composition compared to controls during endoscopic remission, we proceeded to investigate whether MLI bacterial composition and diversity were associated with disease behavior. We compared CD patients with complicated disease based upon the Montreal classification (B2=stricturing or B3=internal penetrating) to those with uncomplicated disease (B1). Lower bacterial diversity was seen in the sigmoid and cecum in patients with complicated CD (B2/B3). This reached significance for the Chao1 index ( $p=0.004$ ,  $p=0.04$ ) but not the Shannon index ( $p=0.1$ ,  $p=0.1$ ) after adjusting for gender, age, and obesity (Fig. 5A). Beta diversity analysis using Bray-Curtis dissimilarity did not identify statistically significant differences in bacterial composition by disease behavior adjusting for gender, age, and obesity, though there was a trend towards significance in the cecum ( $p=0.07$ ) (Fig. 5B). Nevertheless, taxa level differences were sufficient to construct random forests classifiers to differentiate complicated (B2 or B3) vs. uncomplicated (B1) disease. Classifiers were trained using data from 60% of CD patients and tested on the remaining 40% of CD patients. In the test subset, sigmoid and cecum MLI classifiers had AUC of 0.81 (95% CI 0.61–0.96) and 0.82 (95% CI 0.63–0.96) (Fig. 5C).

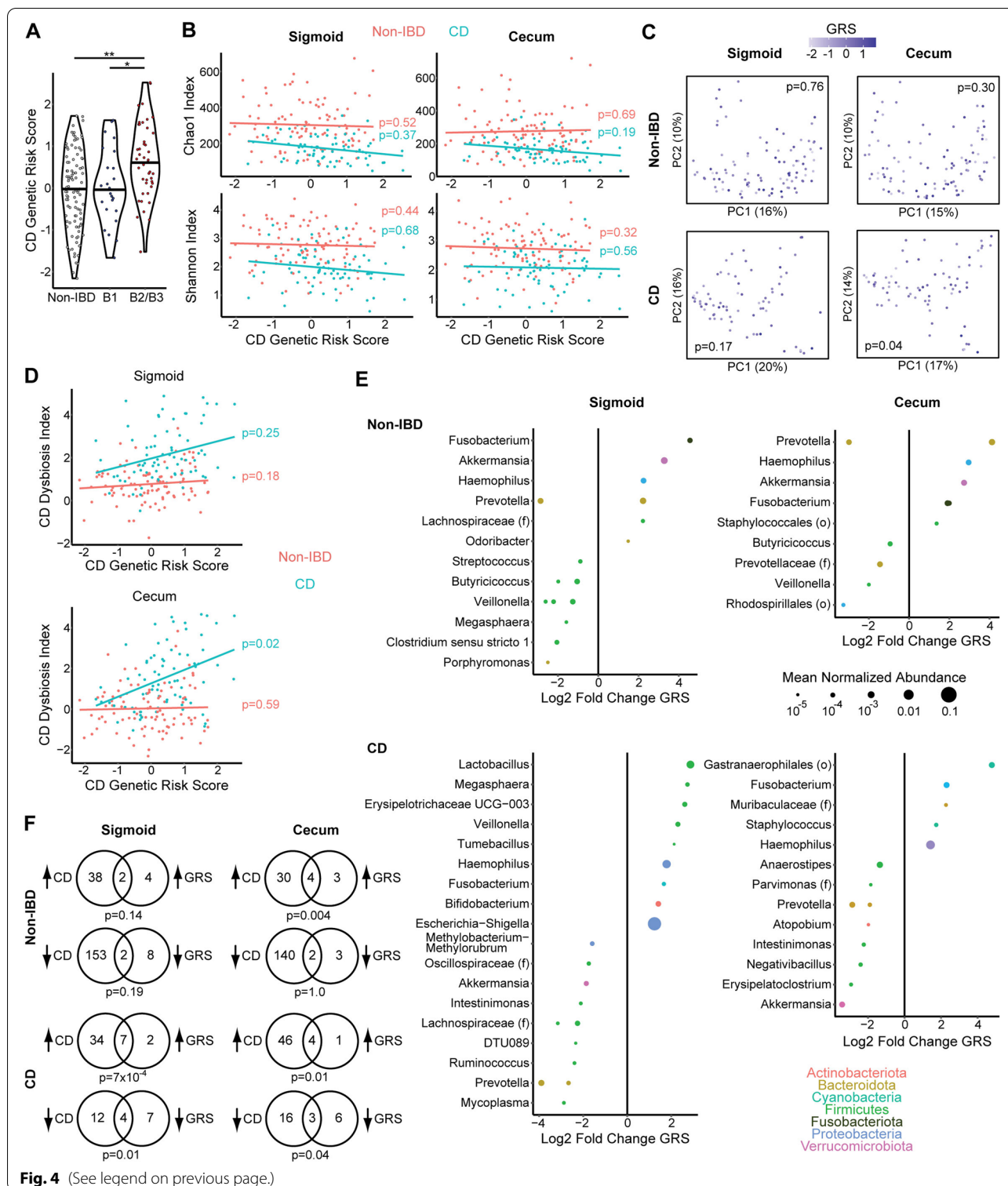
Moreover, complicated CD was associated with statistically significant increases in CD dysbiosis index compared to uncomplicated CD in both the sigmoid and cecum (Fig. 5D). Differential abundance testing demonstrated enrichment of *Fusobacterium*, *Enterococcus*, *Actinomyces*, [*Ruminococcus*] *gnavus* group, and *Akkermansia* ASVs in subjects with complicated CD (Fig. 5E). Conversely, uncomplicated CD was associated with higher relative abundance of many ASVs in genera within the Firmicutes phylum, including those associated with short-chain fatty acid production such as *Faecalibacterium* and *Ruminococcus* that were depleted in CD compared to non-IBD controls (Figs. 2E and 5E).

#### MLI microbiome predicts future risk of CD progression

Given the association of the MLI microbiome with CD disease behavior, we investigated whether MLI microbial profiles could also predict future CD disease course. Follow-up clinical assessment was available for 72 CD patients, with 32 (44%) showing disease progression after MLI sampling. Progressors were identified by clinician chart review as having any of the following during the period between the index colonoscopy and chart review: hospitalization for CD, episode of bowel obstruction, surgery to treat CD, or requirement to change IBD medication. Follow-up time after MLI sampling was equivalent between progressors and non-progressors (mean 3.63 years vs. 3.76 years,  $p=0.64$ ). There was no significant difference in sigmoid or cecal alpha diversity between CD patients who progressed vs. non-progressors by the Chao1 index ( $p=0.10$ ,  $p=0.33$ ) and the Shannon index ( $p=0.94$ ,  $p=0.73$ ) adjusting for gender, age, obesity, and disease behavior. There were also no significant differences in sigmoid or cecal MLI beta diversity by CD progression status adjusting for gender, age, obesity, and disease behavior ( $p=0.49$ ,  $p=0.58$ ). Sigmoid and cecal CD dysbiosis indices did not significantly differ between progressors and non-progressors adjusting for gender, age, obesity, and disease behavior ( $p=0.99$ ,  $p=1.0$ ). However, progressors had differential relative abundances of multiple ASVs compared to non-responders, including depletion of a *Parasutterella* ASV in both sigmoid

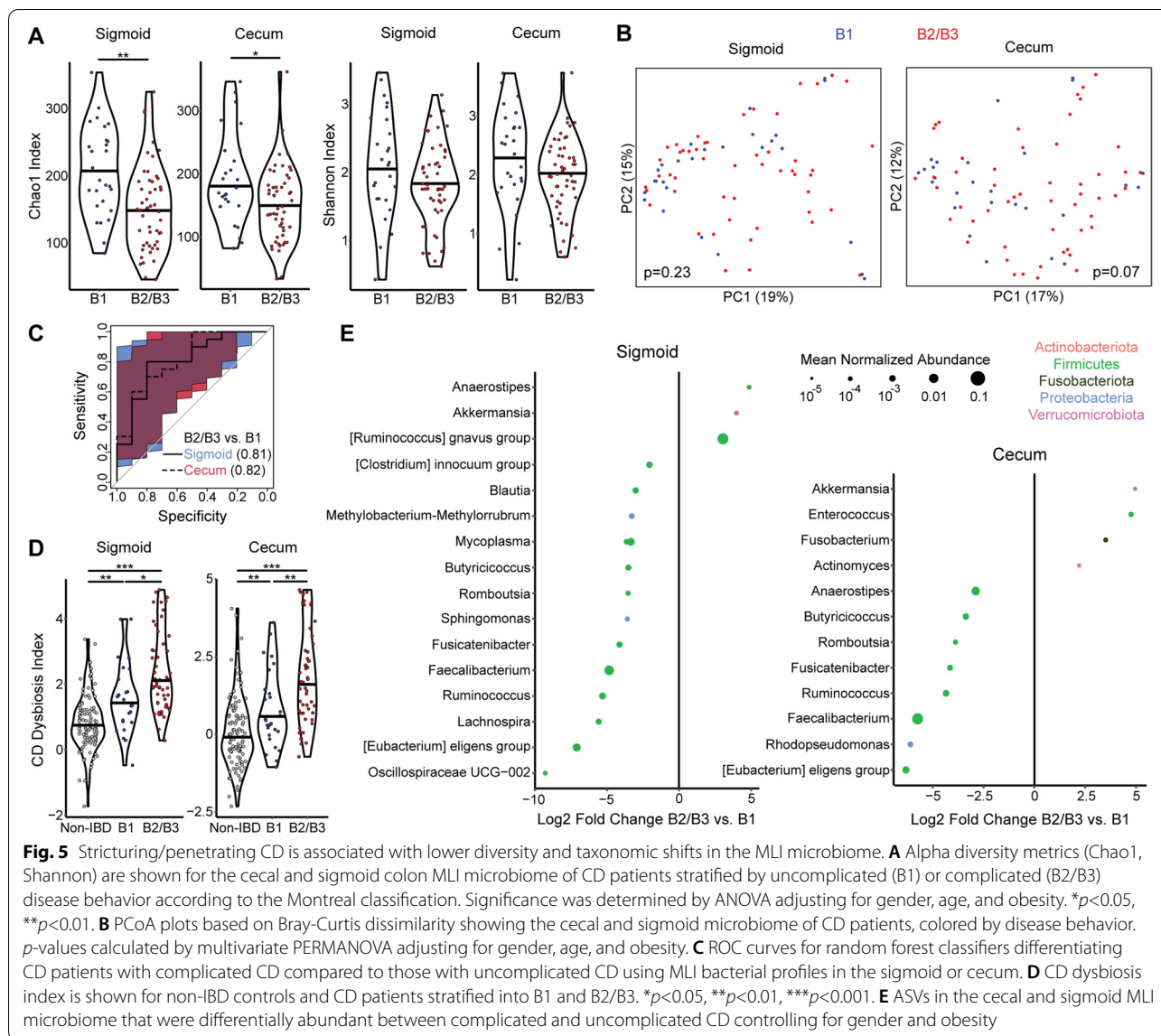
(See figure on next page.)

**Fig. 4** CD genetic risk is associated with taxonomic shifts in the MLI microbiome of CD patients and non-IBD controls. **A** CD genetic risk score for non-IBD controls and CD patients divided by disease behavior (B1 and B2/B3). **B** Alpha diversity metrics are shown by CD genetic risk score for non-IBD controls and CD patients, differentiated by color with separate regression lines. **C** PCoA plots visualizing the cecal and sigmoid microbiome of CD patients, with color scales representing CD genetic risk score (GRS).  $p$ -values calculated by multivariate PERMANOVA adjusting for gender, age, and obesity. **D** CD dysbiosis index is shown by CD genetic risk score for non-IBD controls and CD patients. Significance in panels **A**, **B**, and **D** was determined by ANOVA adjusted for gender, age, obesity, and disease behavior (for CD patients). **E** ASVs in the cecal and sigmoid MLI microbiome that were significantly associated with CD GRS ( $q<0.1$ ) adjusting for gender, obesity, and CD disease behavior (for CD patients). **F** Venn diagrams depicting overlap of GRS-associated ASVs with ASVs that were enriched or depleted in CD vs. controls in the same region (sigmoid or cecum). Analyses were performed separately in non-IBD controls and CD patients. Significance of overlap was determined by Fisher's exact test



and cecum (Fig. 6A). These taxa were used to construct random forests classifiers for CD progression which had AUC of 0.74 (95% CI 0.52–0.93) and 0.70 (95% CI 0.48–0.89) using sigmoid or cecal MLI profiles (Fig. 6B).

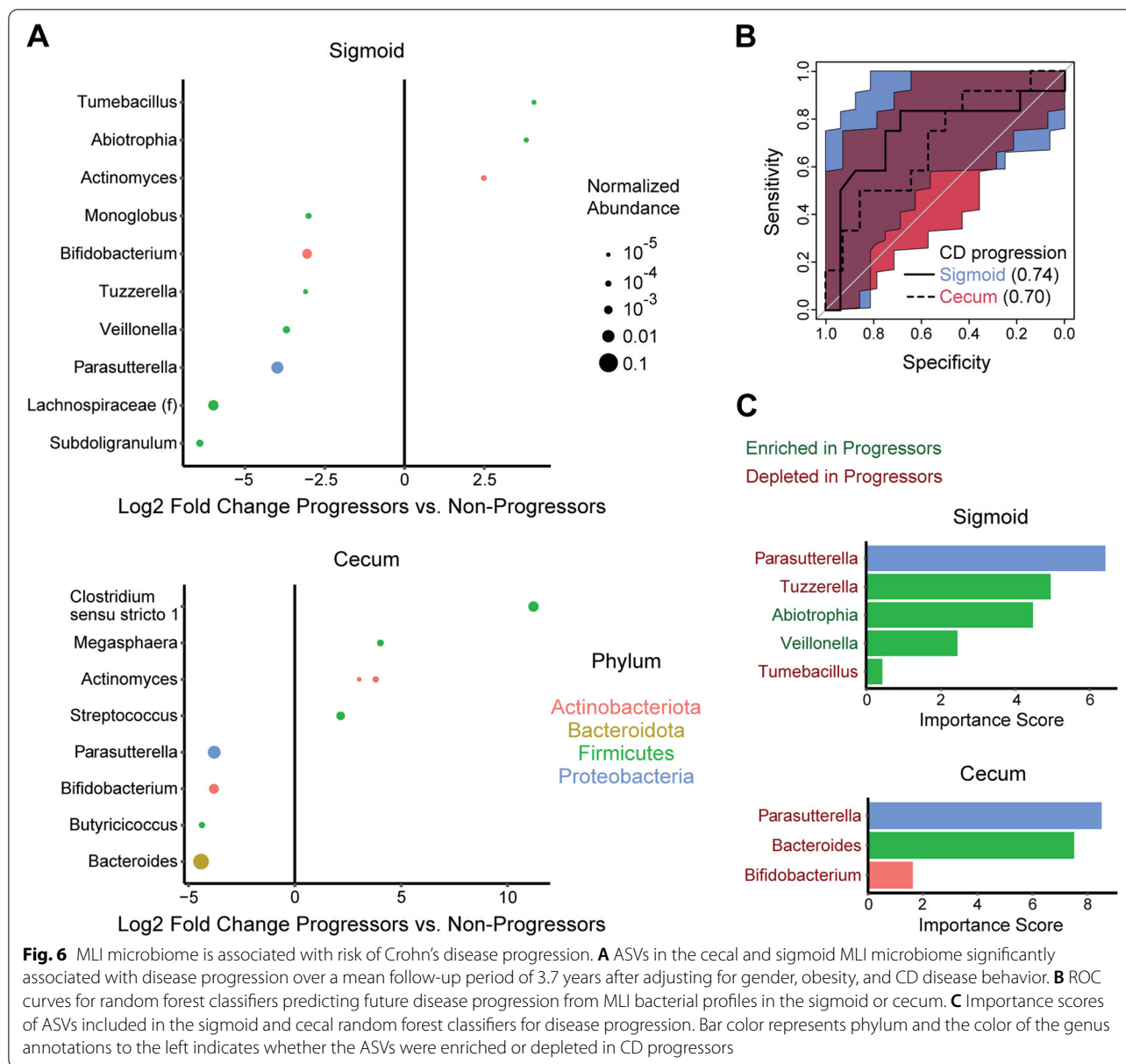
In both random forest classifiers, the *Parasutterella* ASV made the greatest contribution to classifier accuracy, followed closely in the cecal classifier by a *Bacteroides* ASV that was depleted in progressors (Fig. 6C).



**CD is characterized by shifts in the MLI metabolome including increased cholate to deoxycholate ratio**

LC-MS untargeted metabolomics analysis was performed on the MLI samples (Additional files 1 and 2, for positive and negative ESI mode metabolomics, respectively). The global MLI metabolomics profiles of CD patients significantly differed from that of controls in the cecum ( $p=0.03$ ) and trended towards significance in the sigmoid colon ( $p=0.08$ ) (Fig. 7A). Random forests classifiers based on metabolomics data had moderate accuracy for differentiating CD vs controls with AUC of 0.77 (0.65–0.88) and 0.76 (0.62–0.85) for sigmoid and cecum, respectively (Fig. 7B). Mummichog analysis of the list of putatively identified differentially abundant spectral features demonstrated that the

following pathways were significantly enriched in the sigmoid: bile acid biosynthesis, de novo fatty acid biosynthesis, fatty acid activation, fatty acid metabolism, and lysine metabolism (Fig. 7C). Of these, only bile acid biosynthesis was also enriched in cecum. Spectral features assigned by mummichog to the bile acid biosynthesis pathway included two with single annotations for cholate and deoxycholate. In both the sigmoid and cecum, CD showed increased cholate (primary bile acid) and decreased deoxycholate (secondary bile acid) relative to non-IBD controls, resulting in highly significant differences in the log ratio of cholate:deoxycholate (Fig. 7D). Correlation analysis was then performed to identify microbes associated with levels of these two bile acids after adjusting for disease status, gender, and

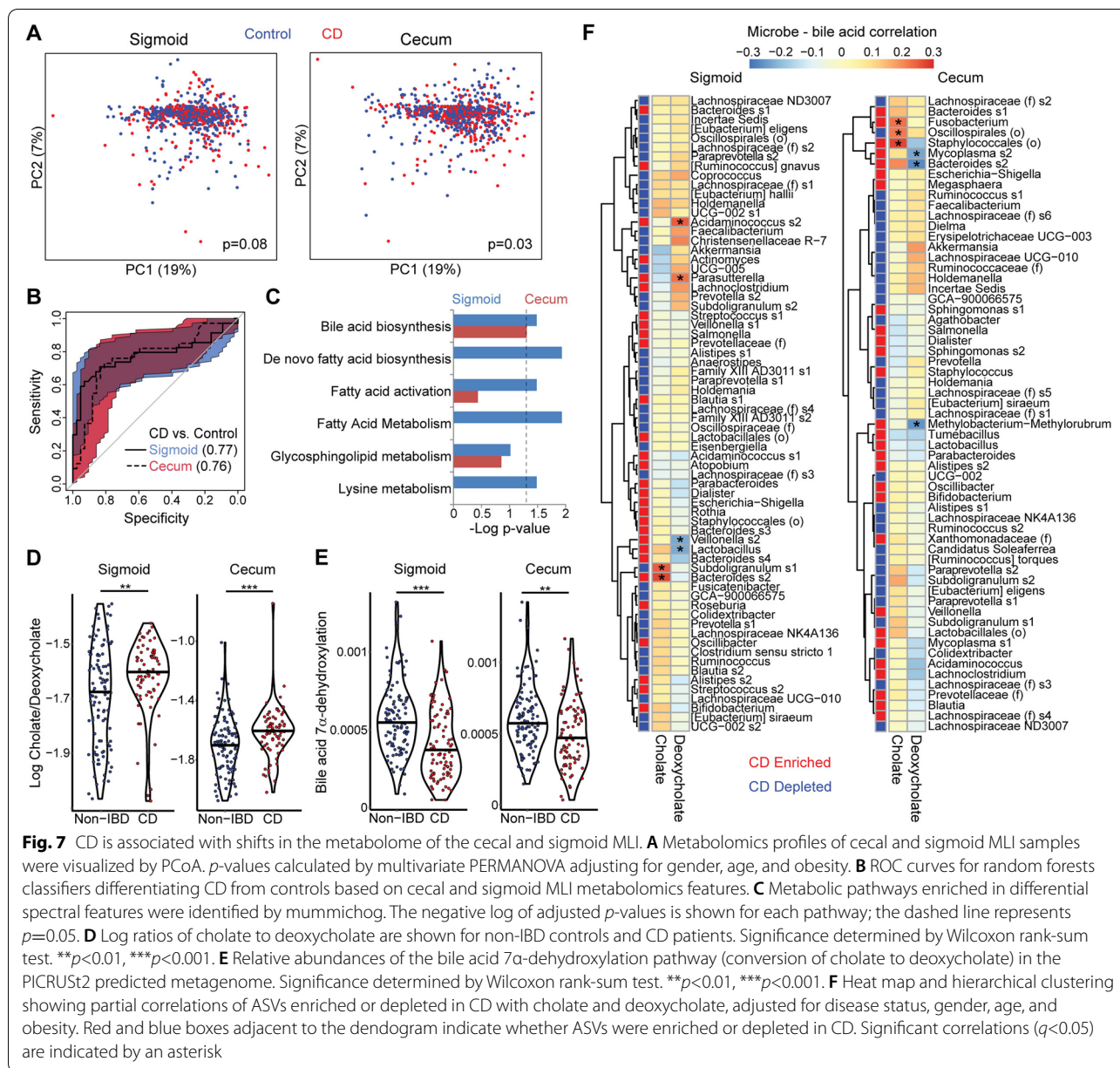


obesity. Six significant microbe-bile acid correlations were identified in the sigmoid and six in the cecum (Fig. 7F). Microbes that were enriched in CD and positively associated with cholate included a *Fusobacterium* ASV (cecum), an ASV within the Staphylococcales order (cecum), and a *Bacteroides* ASV (sigmoid). ASVs that were enriched in CD and negatively associated with deoxycholate included *Veillonella* (sigmoid), *Lactobacillus* (sigmoid), *Mycoplasma* (cecum), *Methylobacterium-Methylorubrum*, and the same *Bacteroides* ASV that were positively correlated with cholate (cecum). Given the association of microbes with bile acid levels, we investigated whether CD was associated

with a shift in predicted abundance of genes involved in bile acid 7 $\alpha$ -dehydroxylation, which mediates bacterial conversion of cholate to deoxycholate. CD was associated with decreased abundance of this pathway in both sigmoid and cecum MLI, consistent with the increased cholate:deoxycholate ratio at both sites (Fig. 7E).

### Discussion

We report the findings of the first moderate- to large-sized study characterizing the colonic mucosal microbiome in Crohn's disease during endoscopic remission compared to that of non-IBD controls. Similar to prior studies of the fecal and tissue microbiome, CD patients



had lower MLI bacterial diversity and altered composition compared to controls [8–15]. One of the major caveats of these past studies has been the potential contribution of active intestinal inflammation to the observed microbial profiles, as animal model studies have demonstrated that induction of intestinal inflammation by itself triggers characteristic shifts in the microbiome including expansion of aerotolerant bacteria and human studies have found that anti-inflammatory therapies reduce dysbiosis [12, 41, 76]. In this study, we detected a robust CD bacterial signature even in the absence of endoscopic inflammation, supporting that

bacterial changes are a distinctive feature of CD pathophysiology separate from secondary effects of intestinal inflammation on the microbiome. However, as histological or biochemical parameters of inflammation were not assessed, it remains possible that low-grade inflammation that was not evident endoscopically could contribute to the observed microbiome alterations. Of note, similar results were obtained in a recent study which assessed the microbiome of colonic tissue biopsies taken from 15 non-inflamed CD patients and 16 non-IBD controls [77]. Interestingly, CD not only affected MLI bacterial profiles in the cecum and sigmoid colon considered separately,

but also changed the relationship between these two intestinal regions. CD patients showed reduced intra-individual similarity between the sigmoid and cecum as well as increased bacterial diversity in the cecum relative to the sigmoid. While IBD-associated shifts in microbial abundances can vary by colonic region, to our knowledge this is the first study to assess for perturbation of biogeographic relationships across colonic sites [10].

Our study demonstrated altered abundance of a wide range of taxa in the CD MLI microbiome during endoscopic remission compared to that of controls, including marked expansion of *Escherichia-Shigella* and *Klebsiella* spp. This is consistent with prior reports that the Enterobacteriaceae family, which includes *Escherichia-Shigella* and *Klebsiella*, is highly enriched in biopsies from CD patients compared to controls [10–12, 16]. There is an extensive literature demonstrating that adherent-invasive strains of *E. coli* isolated from CD patients can promote colitis and enteritis in animal models [78]. In our study, *Klebsiella* showed an even more prominent enrichment than *Escherichia-Shigella* in CD compared to controls. *Klebsiella* enrichment has been previously reported in the fecal and ileal microbiome of CD patients; this study would be the first to report colonic enrichment [12, 77]. *Klebsiella*, including strains of human-derived oral *Klebsiella pneumoniae*, has been reported to promote animal models of colitis [79, 80]. The strong colonic mucosal *Klebsiella* signature identified here supports the human relevance of these preclinical studies and the possibility that the MLI is an important ecologic niche for oral-derived pathobionts in the gastrointestinal tract. CD was also characterized by expansion of several highly abundant *Bacteroides* ASVs, matching the findings of a large fecal shotgun metagenomics study which reported enrichment of *B. fragilis* and *B. vulgatus* in CD [16]. *B. vulgatus* and enterotoxigenic strains of *B. fragilis* have been shown to promote experimental models of colitis [81, 82]. The MLI CD microbiome was also characterized by global depletion of Firmicutes, which encompassed members of a wide range of genera including *Faecalibacterium*, *Ruminococcus*, *Anaerostipes*, *Coprococcus*, and *Clostridium*, all of which have been previously reported to be depleted in the fecal and/or colonic tissue microbiome of CD patients [12, 14, 34]. Of these, the most consistent association in the literature has been depletion of *Faecalibacterium prausnitzii*, which has been shown to reduce severity of experimental colitis through production of anti-inflammatory products such as butyrate and a novel protein, MAM [83, 84]. Collectively, bacterial taxa that were enriched or depleted in CD were used to construct a CD dysbiosis index which could be used to summarize MLI bacterial shifts associated with CD.

This study also provided insight into the mucosal microbiome of obesity. Despite the compelling preclinical data supporting a role for the microbiome in obesity, specific taxonomic signatures of human obesity have been elusive due to variability across fecal microbiome studies [85–90]. The potential importance of the mucosal microbiota in obesity is supported by reports that obese individuals have increased intestinal permeability and translocation of gut microbes into the mesenteric adipose tissue [91, 92]. Strikingly, we found that obese non-IBD controls had similar MLI bacterial characteristics as CD, including reduced diversity, expansion of pathobionts such as *Escherichia-Shigella*, and depletion of beneficial microbes such as *Faecalibacterium*. This was reflected as a significant elevation in CD dysbiosis index compared to controls with BMI <25. Among the obesity-associated shifts in the MLI microbiome, enrichment of *Escherichia-Shigella* and *Megasphaera* and depletion of *Faecalibacterium*, *Paraprevotella*, and *Methanobrevibacter* have been reported in at least one prior fecal microbiome study [93]. These findings indicate that mucosal dysbiosis is a common thread between CD and obesity—both diseases rising rapidly in incidence in the Western world [85]. This is consistent with the current understanding of obesity as a low-grade inflammatory state and suggests that mucosal dysbiosis could contribute to inflammation in the settings of both obesity and IBD.

Complicated CD with either stricturing or penetrating disease behavior was associated with reduced MLI microbial diversity and increased CD dysbiosis index relative to uncomplicated CD. This is consistent with a recent study of colon tissue biopsies from 15 CD patients which observed reduced diversity in complicated CD (B2 and B3) compared to uncomplicated CD [77]. There is little existing literature on specific colonic mucosal taxa that distinguish CD disease behaviors. One study of newly diagnosed pediatric CD patients reported that those who went on to progress to B2 or B3 had increased levels of *Ruminococcus* and *Collinsella* and reduced levels of *Rothia* and *Veillonella* in rectal tissue biopsies obtained at initial presentation [34]. These taxa do not overlap with those identified in this study, potentially reflecting the effects of active inflammation in the prior study as well as differences in the microbes that are associated with future disease behavior as opposed to established disease behavior. In this study, *Akkermansia* enrichment and *Faecalibacterium* depletion were observed in both the sigmoid and cecum of complicated CD. The association of reduced *Faecalibacterium* with CD complications supports the concept that the anti-inflammatory effects of these microbes protect against development of disease complications. Relative expansion of *Akkermansia* sp. in complicated CD was surprising as this microbe was



reduced overall in CD compared to controls. This suggests a complex role for *Akkermansia* spp. in CD, consistent with preclinical studies which have found that *Akkermansia* can reduce or exacerbate intestinal inflammation depending on the animal model [94, 95]. Significant enrichment of a *Fusobacterium* ASV in complicated CD was seen in the cecum but not the sigmoid. This suggests that *Fusobacterium*—oral pathobionts implicated in colorectal cancer that can also exacerbate intestinal inflammation—may contribute to disease processes in the proximal colon that promote stricturing and/or penetrating disease [96, 97].

MLI microbial profiles were also found to predict future CD disease progression. To date, much of the literature on microbial biomarkers of CD disease course has centered on the risk of post-operative reoccurrence after ileal or ileocecal resection. These studies have demonstrated that ileal mucosal microbial profiles at the time of resection are highly predictive of recurrence whereas fecal microbial profiles have comparatively weak accuracy [83, 98, 99]. One study of longitudinally collected fecal samples from a cohort of 45 CD patients in clinical remission followed for 2 years identified 17 differential taxa between the 12 patients who flared and those who did not, including *Sutterella*, *S24.7*, Gemellaceae, and Christensenellaceae [20]. In our study with 72 CD patients in endoscopic remission, the most consistent signals were depletion of *Parasutterella* and *Bifidobacterium* spp. in those who had disease progression. The signature taxa in this study differed from that of the prior study likely due to the use of mucosal rather than fecal samples, differences between cross-sectional and longitudinal assessment, and exclusion of patients in clinical remission with endoscopic signs of inflammation in our study. Among the microbes associated with disease progression in this study, *Parasutterella* made the strongest contribution to classifier accuracy. Interestingly, *Parasutterella* spp. have been previously reported to be greatly enriched in the ileal submucosa of CD patients, suggesting that these microbes have invasive properties [100]. This would be the first report of expansion of these microbes in the CD colon and, if validated, warrants investigation into mechanisms by which these and other mucosal pathobionts can modulate disease course in CD.

Our study also contributes new insights into genetic factors that influence the intestinal microbiome. Existing gene-microbe association studies have largely focused on the fecal microbiome of large healthy cohorts, and analyzed for association of taxa enriched with individual locus variants at genome-wide significance [26, 29, 101–104]. However, it is notable that while host traits reflect the combinatorial effect of genetic loci, few studies have evaluated the relationship of summary measures

of IBD genetic risk with the microbiome. We are aware of only one such study that reported an association of a genetic risk score derived from 11 loci with reduced fecal *Roseburia* in healthy controls but not IBD patients [105]. In the present study, we found that a comprehensive genetic risk score comprising 186 known variants was significantly associated with microbial composition and higher CD dysbiosis index in the cecum of CD patients. This relationship did not reach significance in the sigmoid, suggesting greater impact of genetic burden on the microbiome in the proximal colon. A strength of this study is the assessment of a summary metric of IBD genetic risk on microbiome composition in non-IBD subjects, which isolates their microbiome impact from that related to disease state. Controls showed taxonomic changes with higher GRS including enrichment of *Fusobacterium* but did not show an association of GRS with CD dysbiosis index. This suggests that in the absence of disease, only specific mucosal taxa are responsive to global genetic risk for CD and that genetic modulation of mucosal responses in CD may contribute to the association of GRS with dysbiosis in CD.

MLI sampling also offered insight into metabolite shifts in the colonic mucosa of CD patients. In particular, we observed an increased ratio of a primary bile acid, cholate, relative to its corresponding secondary bile acid, deoxycholate. These findings suggest a reduction in bacterial transformation of primary to secondary bile acids and are consistent with a recent fecal metabolomics study reporting increased cholate and decreased deoxycholate in CD patients with dysbiosis as well as findings from preclinical models of intestinal injury [19, 106]. The shift towards greater primary bile acid in the CD MLI was associated with alterations of specific MLI taxa which could reflect either a contribution of these microbes to bile acid transformation—consistent with reduced predicted abundance of genes in the bile acid 7 $\alpha$ -dehydroxylation pathway—or downstream effects of altered bile acid metabolism on these microbes. The latter possibility is consistent with the known wide-ranging effects of bile acid receptors such as FXR on intestinal physiology including production of antimicrobial products [107]. Of note, these correlations were based on microbial relative abundances and different relationships may be observed between metabolite levels and absolute abundances of mucosal bacteria. These findings highlight a potential role for altered bile acid metabolism in mucosal dysbiosis in IBD and demonstrate that metabolomics analysis of MLI samples can be used to characterize mucosal host-microbe interactions in diseases such as CD.

Beyond the insights that can be gained from taxonomic and metabolite associations with CD disease

phenotypes, this study also suggests the potential clinical application of mucosal microbial profiles as biomarkers for Crohn's disease. The largest published CD study using colonic mucosal samples reported a classifier for distinguishing CD from controls that had AUC 0.78 [11]. The current study builds on this existing literature by demonstrating that colonic MLI microbial biomarkers not only robustly separate CD from controls (AUC 0.91, 0.93), but also distinguish CD disease behaviors with high accuracy even during endoscopic remission (AUC 0.81, 0.82) and can predict future disease progression (AUC 0.74, 0.70). In clinical practice, such biomarker panels could be used to identify CD patients in remission who are at higher risk for flare or disease complications. This could guide clinical decisions on therapies and disease monitoring, a challenge in CD as active disease may not manifest in clinical symptoms until substantial disease progression has occurred. While this study provides significant insights into mucosal microbial markers of CD disease phenotype, it has limitations which include the demographic differences between controls and CD patients, lack of longitudinal sampling, moderate sample size of disease subgroups such as complicated CD, and absence of an independent validation cohort to assess classifier accuracy. Future studies are warranted to address the clinical potential of MLI-based biomarkers for CD clinical management and to investigate individual MLI microbes that may be instigators of inflammation and disease progression.

## Conclusions

During endoscopic remission, Crohn's disease patients show robust differences in colonic mucosal microbiome diversity, composition, and biogeography compared to unaffected controls. This indicates that microbial changes are a distinctive feature of CD pathophysiology separate from secondary effects of intestinal inflammation on the microbiome. Elements of this mucosal microbiome signature also predict CD behavior and future disease progression, representing potential biomarkers for prognostication and interception therapies. Specific taxa in both CD patients and non-IBD controls were associated with CD genetic risk, providing evidence that pre-disease risk may be mediated by selective microbial shifts in the mucosa. Obese controls showed changes in colonic mucosal microbial composition that paralleled those of CD, suggesting a shared microbial contribution to these two diseases that are both rising in incidence in the Western world.

## Abbreviations

ASV: Amplicon sequence variant; AUC: Area under the curve; BMI: Body mass index; CD: Crohn's disease; CI: Confidence interval; GRS: Gene risk scores; HMDB: Human Metabolome Database; IBD: Inflammatory bowel disease; IQR: Interquartile range; KEGG: Kyoto Encyclopedia of Genes and Genomes; MLI: Mucosal-luminal interface; PCoA: Principal coordinate analysis; ppm: Parts per million; SNP: Single-nucleotide polymorphism.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-022-01099-7>.

**Additional file 1.** Positive ESI mode metabolomics: This file contains the positive ESI mode metabolomics data. Each spectral feature is labeled by its m/z and retention time.

**Additional file 2.** Negative ESI mode metabolomics: This file contains the negative ESI mode metabolomics data. Each spectral feature is labeled by its m/z and retention time.

## Acknowledgements

We are grateful to the Crohn's disease and non-IBD volunteer subjects whose selfless participation in this project made this study possible.

## Authors' contributions

J.P.J., J.Br., and D.P.B.M. designed the study. D.P.B.M., P.F., E.V., A.I., G.M., D.S., and S.T. recruited subjects. M.T. processed samples. M.G. and A.F. performed metabolomics. P.R. and J.Bo. performed 16S rRNA gene sequencing. D.L., T.H., and D.P.B.M. performed genetic analysis. J.P.J., M.G., V.L., and T.M. performed bioinformatics analysis. J.P.J. and J.Br. prepared the manuscript. All authors read and approved the final manuscript.

## Funding

This study was funded by The Crohn's and Colitis Foundation of America, USPHS grant P01DK046763, the NIH/National Center for Advancing Translational Science (NCATS) UL1TR000124, and the Cedars-Sinai F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute. J.P.J. was supported by NIH/NIDDK T32DK07180-39 and VA CDA2 IK2CX001717. D.P.B.M. was supported by The European Union, The Joshua L. and Lisa Z. Greer Chair in IBD Genetics, and grants DK062413, DK046763-19, AI067068, and HS021747.

## Availability of data and materials

The sequencing data supporting the conclusions of this article are available in the NCBI Bioproject repository, PRJNA737297 (<https://www.ncbi.nlm.nih.gov/bioproject/737297>) [108]. The metabolomics data have been deposited in the Metabolomics Workbench repository under ID 3353 (<https://www.metabolomicsworkbench.org>) and are included as Additional files 1 and 2, representing positive and negative ESI mode metabolomics, respectively [109]. The R code used for bioinformatics analyses are provided in a Github repository (<https://github.com/jjgithub650/Mucosal-luminal-interface-microbiome-in-Crohns-disease>) [110].

## Declarations

### Ethics approval and consent to participate

The Cedars-Sinai Medical Center Institutional Review Board approved the protocols governing participants (IRB #3358). All subjects provided informed consent to participate in this study. The research conformed to the principles of the Helsinki Declaration.

### Consent for publication

Not applicable.

### Competing interests

Cedars-Sinai Medical Center, Dr. Stephan Targan, and Dr. Dermot McGovern have financial interests in Prometheus Biosciences, Inc., a company which has access to the data and specimens in Cedars-Sinai's MIRIAD Biobank (including the data and specimens used in this study). Prometheus Biosciences seeks to

develop commercial products. The remaining authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Vatche and Tamar Manoukian Division of Digestive Diseases, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095-6949, USA. <sup>2</sup>Division of Gastroenterology, Hepatology and Parenteral Nutrition, Veterans Administration Greater Los Angeles Healthcare System, Los Angeles, USA. <sup>3</sup>Lerner Research Institute, Cleveland Clinic, Cleveland, USA. <sup>4</sup>F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, USA. <sup>5</sup>National Institute of Standards and Technology, Gaithersburg, USA. <sup>6</sup>Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA. <sup>7</sup>Department of Plant Pathology and Microbiology, University of California Riverside, Riverside, USA. <sup>8</sup>Department of Medicine, Keck School of Medicine of USC, Los Angeles, USA. <sup>9</sup>Department of Biochemistry and Molecular & Cellular Biology, Georgetown University, Washington, USA.

Received: 13 January 2022 Accepted: 2 August 2022

Published online: 15 August 2022

#### References

- Bernstein CN, Loftus EV, Ng SC, Lakatos PL, Moum B. Hospitalisations and surgery in Crohn's disease. *Gut*. 2012;61:622–9.
- Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2018;390:2769–78.
- Huang H, Fang M, Jostins L, Umicevic Mirkov M, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*. 2017;547:173–8.
- Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol*. 2015;12:205–17.
- Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E, et al. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infect Immun*. 1998;66:5224–31.
- Garrett WS, Lord GM, Punit S, Lugo-Villarino G, Mazmanian SK, Ito S, et al. Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell*. 2007;131:33–45.
- Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*. 2011;145:745–57.
- Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A*. 2007;104:13780–5.
- Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*. 2010;139:1844–1854 e1841.
- Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*. 2012;13:R79.
- Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15:382–92.
- Lewis JD, Chen EZ, Baldassano RN, Otlej AR, Griffiths AM, Lee D, et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe*. 2015;18:489–500.
- Casen C, Vebo HC, Sekelja M, Hegge FT, Karlsson MK, Ciemniewska E, et al. Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Aliment Pharmacol Ther*. 2015;42:71–83.
- Pascal V, Pozuelo M, Borrue N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. *Gut*. 2017;66:813–22.
- Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*. 2006;55:205–11.
- Vich Vila A, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Sci Transl Med*. 2018;10:eap8914.
- Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory bowel disease. *Nat Rev Microbiol*. 2019;17:497–511.
- Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol*. 2017;2:17004.
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569:655–62.
- Braun T, Di Segni A, Ben-Shoshan M, Neuman S, Levhar N, Bubis M, et al. Individualized dynamics in the gut microbiota precede Crohn's disease flares. *Am J Gastroenterol*. 2019;114:1142–51.
- Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol*. 2018;3:337–46.
- Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4:293–305.
- Mills RH, Vazquez-Baeza Y, Zhu Q, Jiang L, Gaffney J, Humphrey G, et al. Evaluating metagenomic prediction of the metaproteome in a 4.5-year study of a patient with Crohn's disease. *mSystems*. 2019;4:e00337-18.
- Jacobs JP, Braun J. Immune and genetic gardening of the intestinal microbiome. *FEBS Lett*. 2014;588:4102–11.
- Graham DB, Xavier RJ. From genetics of inflammatory bowel disease towards mechanistic insights. *Trends Immunol*. 2013;34:371–8.
- Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. 2021;53:156–65.
- Scepanovic P, Hodel F, Mondot S, Partula V, Byrd A, Hammer C, et al. A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome*. 2019;7:130.
- Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect of host genetics on the gut microbiome. *Nat Genet*. 2016;48:1407–12.
- Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekman R, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159:789–99.
- Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–24.
- Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47:979–86.
- Chen GB, Lee SH, Montgomery GW, Wray NR, Visscher PM, Geary RB, et al. Performance of risk prediction for inflammatory bowel disease based on genotyping platform and genomic risk score method. *BMC Med Genet*. 2017;18:94.
- Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut*. 2006;55:749–53.
- Kugathasan S, Denson LA, Walters TD, Kim MO, Marigorta UM, Schirmer M, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet*. 2017;389:1710–8.
- Morgan XC, Kabackchiev B, Waldron L, Tyler AD, Tickle TL, Milgrom R, et al. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol*. 2015;16:67.
- Bouter KE, van Raalte DH, Groen AK, Nieuwdorp M. Role of the gut microbiome in the pathogenesis of obesity and obesity-related metabolic dysfunction. *Gastroenterology*. 2017;152:1671–8.

37. Agus A, Clement K, Sokol H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut*. 2021;70:1174–82.
38. Vasapolli R, Schutte K, Schulz C, Vital M, Schomburg D, Pieper DH, et al. Analysis of transcriptionally active bacteria throughout the gastrointestinal tract of healthy individuals. *Gastroenterology*. 2019;157:1081–1092. e1083.
39. Targownik LE, Sexton KA, Bernstein MT, Beatie B, Sargent M, Walker JR, et al. The relationship among perceived stress, symptoms, and inflammation in persons with inflammatory bowel disease. *Am J Gastroenterol*. 2015;110:1001–12 quiz 1013.
40. Gracie DJ, Williams CJ, Sood R, Mumtaz S, Bholah MH, Hamlin PJ, et al. Poor correlation between clinical disease activity and mucosal inflammation, and the role of psychological comorbidity, in inflammatory bowel disease. *Am J Gastroenterol*. 2016;111:541–51.
41. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe*. 2007;2:119–29.
42. Li X, LeBlanc J, Truong A, Vuthoori R, Chen SS, Lustgarten JL, et al. A metaproteomic approach to study human-microbial ecosystems at the mucosal luminal interface. *PLoS One*. 2011;6:e26542.
43. Watt E, Gemmell MR, Berry S, Glaire M, Farquharson F, Louis P, et al. Extending colonic mucosal microbiome analysis—assessment of colonic lavage as a proxy for endoscopic colonic biopsies. *Microbiome*. 2016;4:61.
44. Presley LL, Ye J, Li X, Leblanc J, Zhang Z, Ruegger PM, et al. Host-microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal-luminal interface. *Inflamm Bowel Dis*. 2012;18:409–17.
45. Li X, LeBlanc J, Elashoff D, McHardy I, Tong M, Roth B, et al. Microgeographic proteomic networks of the human colonic mucosa and their association with inflammatory bowel disease. *Cell Mol Gastroenterol Hepatol*. 2016;2:567–83.
46. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*. 2013;1:17.
47. Deeke SA, Starr AE, Ning Z, Ahmadi S, Zhang X, Mayne J, et al. Mucosal-luminal interface proteomics reveals biomarkers of pediatric inflammatory bowel disease-associated colitis. *Am J Gastroenterol*. 2018;113:713–24.
48. Zhang X, Deeke SA, Ning Z, Starr AE, Butcher J, Li J, et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat Commun*. 2018;9:2873.
49. Yan A, Butcher J, Mack D, Stintzi A. Virome sequencing of the human intestinal mucosal-luminal interface. *Front Cell Infect Microbiol*. 2020;10:582187.
50. Wang C, Zhang H, Liu H, Zhang H, Bao Y, Di J, et al. The genus *Sutterella* is a potential contributor to glucose metabolism improvement after Roux-en-Y gastric bypass surgery in T2D. *Diabetes Res Clin Pract*. 2020;162:108116.
51. Volkmann ER, Chang YL, Barroso N, Furst DE, Clements PJ, Gorn AH, et al. Association of systemic sclerosis with a unique colonic microbial consortium. *Arthritis Rheumatol*. 2016;68:1483–92.
52. Zhang H, Chang Y, Zheng Q, Zhang R, Hu C, Jia W. Altered intestinal microbiota associated with colorectal cancer. *Front Med*. 2019;13:461–70.
53. Tong M, Jacobs JP, McHardy IH, Braun J. Sampling of intestinal microbiota and targeted amplification of bacterial 16S rRNA genes for microbial ecologic analysis. *Curr Protoc Immunol*. 2014;107:7.41.41–47.41.11.
54. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6:1621–4.
55. Jacobs JP, Goudarzi M, Singh N, Tong M, McHardy IH, Ruegger P, et al. A disease-associated microbial and metabolomics state in relatives of pediatric inflammatory bowel disease patients. *Cell Mol Gastroenterol Hepatol*. 2016;2:750–66.
56. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
57. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech XZ, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems*. 2017;2:e00191–16.
58. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
59. Quast C, Pruesse E, Yilmaz P, Gerken J, Schwier T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590–6.
60. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
61. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. *vegan: Community Ecology Package*. R package version 2.5-6 edition; 2019.
62. McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*. 2001;82:290–7.
63. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol*. 2014;10:e1003531.
64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
65. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003;100:9440–5.
66. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38:685–8.
67. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res*. 2018;46:D633–9.
68. Mak TD, Laiakis EC, Goudarzi M, Fornace AJ Jr. *MetaboLyzer*: a novel statistical workflow for analyzing Postprocessed LC-MS metabolomics data. *Anal Chem*. 2014;86:506–13.
69. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, et al. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics*. 2018;14:128.
70. Pang Z, Chong J, Li S, Xia J. *MetaboAnalystR 3.0*: toward an optimized workflow for global metabolomics. *Metabolites*. 2020;10:186.
71. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–27.
72. Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, et al. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol*. 2013;9:e1003123.
73. Kuhn M. *Building Predictive Models in R Using the caret Package*. *J Stat Softw*. 2008;28:1–26.
74. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
75. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
76. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol*. 2017;14:573–84.
77. Shahir NM, Wang JR, Wolber EA, Schaner MS, Frank DN, Ir D, et al. Crohn's disease differentially affects region-specific composition and aerotolerance profiles of mucosally adherent bacteria. *Inflamm Bowel Dis*. 2020;26:1843–55.
78. Palmela C, Chevarin C, Xu Z, Torres J, Sevrin G, Hirten R, et al. Adherent-invasive *Escherichia coli* in inflammatory bowel disease. *Gut*. 2018;67:574–87.
79. Atarashi K, Suda W, Luo C, Kawaguchi T, Motoo I, Narushima S, et al. Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science*. 2017;358:359–65.
80. Garrett WS, Gallini CA, Yatsunenkov T, Michaud M, DuBois A, Delaney ML, et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe*. 2010;8:292–300.

81. Ramanan D, Tang MS, Bowcutt R, Loke P, Cadwell K. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity*. 2014;41:311–24.
82. Sears CL, Geis AL, Housseau F. *Bacteroides fragilis* subverts mucosal biology: from symbiont to colon carcinogenesis. *J Clin Invest*. 2014;124:4166–72.
83. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A*. 2008;105:16731–6.
84. Quevrain E, Maubert MA, Michon C, Chain F, Marquant R, Tailhades J, et al. Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut*. 2016;65:415–25.
85. Maruvada P, Leone V, Kaplan LM, Chang EB. The human microbiome and obesity: moving beyond associations. *Cell Host Microbe*. 2017;22:589–99.
86. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JL. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444:1027–31.
87. Turnbaugh PJ, Backhed F, Fulton L, Gordon JL. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe*. 2008;3:213–23.
88. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*. 2013;341:1241214.
89. Thaiss CA, Itav S, Rothschild D, Meijer M, Levy M, Moresi C, et al. Persistent microbiome alterations modulate the rate of post-dieting weight regain. *Nature*. 2016;540:544–51.
90. Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio*. 2016;7:e01018-16.
91. Gummesson A, Carlsson LM, Störlién LH, Backhed F, Lundin P, Lofgren L, et al. Intestinal permeability is associated with visceral adiposity in healthy women. *Obesity (Silver Spring)*. 2011;19:2280–2.
92. Massier L, Chakaroun R, Tabei S, Crane A, Didt KD, Fallmann J, et al. Adipose tissue derived bacteria are associated with inflammation in obesity and type 2 diabetes. *Gut*. 2020;69:1796–806.
93. Pinart M, Dotsch A, Schlicht K, Laudes M, Bouwman J, Forslund SK, et al. Gut microbiome composition in obese and non-obese persons: a systematic review and meta-analysis. *Nutrients*. 2021;14:12.
94. Ganesh BP, Klopfeisch R, Loh G, Blaut M. Commensal *Akkermansia muciniphila* exacerbates gut inflammation in *Salmonella Typhimurium*-infected gnotobiotic mice. *PLoS One*. 2013;8:e74963.
95. Zhai R, Xue X, Zhang L, Yang X, Zhao L, Zhang C. Strain-specific anti-inflammatory properties of two *Akkermansia muciniphila* strains on chronic colitis in mice. *Front Cell Infect Microbiol*. 2019;9:239.
96. Abed J, Emgard JE, Zamir G, Faroja M, Almogy G, Grenov A, et al. Fap2 mediates *Fusobacterium nucleatum* colorectal adenocarcinoma enrichment by binding to tumor-expressed Gal-GalNAc. *Cell Host Microbe*. 2016;20:215–25.
97. Liu H, Hong XL, Sun TT, Huang XW, Wang JL, Xiong H. *Fusobacterium nucleatum* exacerbates colitis by damaging epithelial barriers and inducing aberrant inflammation. *J Dig Dis*. 2020;21:385–98.
98. Sokol H, Brot L, Stefanescu C, Auzolle C, Barnich N, Buisson A, et al. Prominence of ileal mucosa-associated microbiota to predict postoperative endoscopic recurrence in Crohn's disease. *Gut*. 2020;69:462.
99. Machiels K, Pozuelo Del Rio M, Martinez-De la Torre A, Xie Z, Pascal Andreu V, Sabino J, et al. Early postoperative endoscopic recurrence in Crohn's disease is characterised by distinct microbiota recolonisation. *J Crohns Colitis*. 2020;14:1535–46.
100. Chiodini RJ, Dowd SE, Chamberlin WM, Galandiuk S, Davis B, Glassing A. Microbial population differentials between mucosal and submucosal intestinal tissues in advanced Crohn's disease of the ileum. *PLoS One*. 2015;10:e0134382.
101. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018;555:210–5.
102. Aschard H, Laville V, Tchetgen ET, Knights D, Imhann F, Seksik P, et al. Genetic effects on the commensal microbiota in inflammatory bowel disease patients. *PLoS Genet*. 2019;15:e1008018.
103. Knights D, Silverberg MS, Weersma RK, Gevers D, Dijkstra G, Huang H, et al. Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med*. 2014;6:107.
104. Li D, Achkar JP, Haritunians T, Jacobs JP, Hui KY, D'Amato M, et al. A pleiotropic missense variant in *SLC39A8* is associated with Crohn's disease and human gut microbiome composition. *Gastroenterology*. 2016;151:724–32.
105. Imhann F, Vich Vila A, Bonder MJ, Fu J, Gevers D, Visschedijk MC, et al. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*. 2018;67:108–19.
106. Goudarzi M, Mak TD, Jacobs JP, Moon BH, Strawn SJ, Braun J, et al. An integrated multi-omic approach to assess radiation injury on the host-microbiome axis. *Radiat Res*. 2016;186:219–34.
107. Lavelle A, Sokol H. Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol*. 2020;17:223–37.
108. Jacobs JP, Goudarzi M, Lagishetty V, Li D, Mak T, Tong M, Ruegger P, Haritunians T, Landers C, Fleshner P, et al. Colonic mucosal-luminal interface microbiome in Crohn's disease. PRJNA737297, NCBI Bioproject. 2022. <https://www.ncbi.nlm.nih.gov/bioproject/737297>. Accessed 18 July 2022.
109. Jacobs JP, Goudarzi M, Lagishetty V, Li D, Mak T, Tong M, Ruegger P, Haritunians T, Landers C, Fleshner P, et al. Colonic mucosal-luminal interface microbiome in Crohn's disease. Submission ID 3353, Metabolomics Workbench. 2022. <https://www.metabolomicsworkbench.org>. Accessed 18 July 2022.
110. Jacobs JP, Goudarzi M, Lagishetty V, Li D, Mak T, Tong M, Ruegger P, Haritunians T, Landers C, Fleshner P, et al. Mucosal-luminal-interface-microbiome-in-Crohns-disease. Github repository. 2022. <https://github.com/jjgithub650/Mucosal-luminal-interface-microbiome-in-Crohns-disease>. Accessed 18 July 2022.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

