

Microbial network disturbances in relapsing refractory Crohn's disease

Bahtiyar Yilmaz^{1,2,9}, Pascal Juillerat^{1,2,9}, Ove Øyås³, Charlotte Ramon³, Francisco Damian Bravo², Yannick Franc⁴, Nicolas Fournier⁴, Pierre Michetti^{1,5,6}, Christoph Mueller^{1,7}, Markus Geuking¹, Valerie E. H. Pittet⁴, Michel H. Maillard^{5,6}, Gerhard Rogler⁸, Swiss IBD Cohort Investigators¹⁰, Reiner Wiest^{1,2}, Jörg Stelling³ and Andrew J. Macpherson^{1,2*}

Inflammatory bowel diseases (IBD) can be broadly divided into Crohn's disease (CD) and ulcerative colitis (UC) from their clinical phenotypes. Over 150 host susceptibility genes have been described, although most overlap between CD, UC and their subtypes, and they do not adequately account for the overall incidence or the highly variable severity of disease. Replicating key findings between two long-term IBD cohorts, we have defined distinct networks of taxa associations within intestinal biopsies of CD and UC patients. Disturbances in an association network containing taxa of the Lachnospiraceae and Ruminococcaceae families, typically producing short chain fatty acids, characterize frequently relapsing disease and poor responses to treatment with anti-TNF- α therapeutic antibodies. Alterations of taxa within this network also characterize risk of later disease recurrence of patients in remission after the active inflamed segment of CD has been surgically removed.

There is strong evidence indicating that the intestinal microbiota is responsible for triggering IBD: animal models of chronic intestinal inflammation are almost all mitigated when re-derived to germ-free status^{1,2}; humans with IBD have altered reactivity to their intestinal microbes³; and the genetic loci for human susceptibility to CD and/or UC mainly encode proteins required for innate or adaptive mucosal immune defense mechanisms and overlap with genes known to cause microbiota-sensitive intestinal inflammation in animal models⁴. While no persistent pathogen has been consistently identified as being responsible for IBD, strategies that reduce microbiota exposure, such as surgical fecal stream diversion⁵, elemental diets or antibiotic treatment, can have beneficial effects in particular IBD cases^{6–8}.

The intestinal microbiota stabilizes during childhood, but with inter-individual variations^{9–11} and fluctuations according to diet¹² or intestinal transit time^{13,14}. Microbial taxa compositions in IBD patients have been extensively studied (Extended Data Fig. 1a). These different studies have been carried out with different microbiological and sequencing techniques on various ethnic populations, with members of different ages and stages in the trajectory of disease, using fecal or tissue specimens. Alterations in the relative proportions of taxa within the Bacteroidetes, Firmicutes and Proteobacteria in CD patients compared with UC or non-IBD subjects are generally reported. Despite this, the overall picture remains very inconsistent (Extended Data Figs. 1 and 2).

Microbiota differences may account for (1) a gap in the IBD prevalence that is inadequately explained by known CD/UC genetic susceptibilities^{15–17}; (2) phenotypic differences in the severity and distribution of disease that are only partly explained from genetic

and immunological studies^{15,18–20} and (3) a significant proportion of patients that are poorly responsive to current selective or non-selective immunomodulatory treatments^{21–23}.

Within CD and UC there is immense variability in disease distribution, severity and responsiveness to treatment. These clinical disease behaviors are poorly predicted by host genetic susceptibility markers. Disease behavior, treatment responsiveness and environmental influences have not been comprehensively considered in studies of microbiota patterns in IBD so far (Extended Data Fig. 1b). Our hypothesis here was that microbiota composition and taxa networks underlie the phenotypic variability of established CD and UC that could be addressed with robust deep longitudinal information over many years for each patient. Given that important and potentially correctable differences should be reproducible in the same (central European) environment, we have determined microbiota changes according to disease phenotype that are robust to replication between two independent long-term longitudinal cohorts of patients. We show here microbiota networks that are reproducibly related to long-term disease type and severity, the extent and position of the involved intestinal segments and responsiveness to different treatments.

Results

Distinct microbial communities according to CD or UC diagnosis in IBD patients compared with controls. The microbial communities of IBD patients were assessed using 16S sequencing in biopsy samples obtained from a tissue biopsy biobank of the Swiss IBD cohort (Cohort 1). To validate the results a second independent cohort of patient biopsies (Cohort 2) including healthy,

¹Maurice Müller Laboratories, Department for Biomedical Research, University of Bern, Bern, Switzerland. ²Department of Visceral Surgery and Medicine, Bern University Hospital, University of Bern, Bern, Switzerland. ³Department of Biosystems Science and Engineering and SIB Swiss Institute of Bioinformatics, ETH Zurich, Basel, Switzerland. ⁴Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital, Lausanne, Switzerland. ⁵Gastroenterology La Source-Beaulieu, Lausanne, Switzerland. ⁶Service of Gastroenterology and Hepatology, Department of Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland. ⁷Division of Experimental Pathology, Institute of Pathology, University of Bern, Bern, Switzerland. ⁸Department of Gastroenterology and Hepatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland. ⁹These authors contributed equally to this work: Bahtiyar Yilmaz, Pascal Juillerat. ¹⁰A full list of members and affiliations appears at the end of the paper.

*e-mail: andrew.macpherson@dbmr.unibe.ch

non-IBD controls was collected from Bern gastroenterology clinics (Supplementary Table 1).

Matching clinical deep metadata sets were assembled prospectively and retrospectively from patient records for both cohorts. In addition to gender, age, marital status, family history, specifics of diagnosis, its timing and disease subtype at the outset, contemporaneously collected follow-up information for each patient included trajectories of disease evolution with blood parameters, environmental exposures (alcohol and tobacco use), daily exercise, body mass index (BMI), medication time-courses (antibiotics, corticosteroids, mesalamine and immunosuppressants), nutritional supplements, disease complications, surgical resections and inter-current comorbidities.

As expected from known taxa predilection in the human large intestine²⁴, the dominant bacterial phylotypes in the gastrointestinal tract of IBD patients from both cohorts and healthy controls were Bacteroidetes, Firmicutes and Proteobacteria, with a smaller proportions of Fusobacteria, Actinobacteria and Tenericutes (Supplementary Tables 1 and 2). Average Firmicutes proportions are known to be higher in comparison with Bacteroidetes and we observed this in all gastrointestinal segments of UC compared with CD in both cohorts, but with very wide variability across subjects and sites (Extended Data Fig. 3a,b). As also expected from previous studies²⁵, overall alpha diversity measured according to the Shannon and Simpson indices demonstrated significantly reduced diversity in CD samples compared with UC samples in both patient cohorts and compared with healthy, non-IBD controls in Cohort 2 (Extended Data Fig. 3c,d).

The distinctiveness of the CD microbiota was confirmed by beta diversity analysis, demonstrating a clustering of samples according to IBD diagnosis for both cohorts with the same homogeneity by Bray–Curtis dissimilarity metrics ($P < 0.001$) (Fig. 1a,b) as well as UniFrac distance metrics ($P < 0.001$) (Extended Data Fig. 3e–h).

We next evaluated specific microbial features associated with disease phenotypes that could be replicated at the genus level in both independent cohorts (Fig. 1c). Comparison of relative abundances in CD and UC patients revealed multiple concordant significant differences at phylum (Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria) and genus level (*Bifidobacterium*, *Coriobacteriaceae*•, *Barnesiellaceae*•••, Clostridiales genus••, *Prevotella* (Prevotellaceae family), *Prevotella*••• (Paraprevotellaceae family), *Lachnospira*, *Faecalibacterium*, *Coprococcus*, *Roseburia*, two *Lachnospiraceae* genera (• and ••), two *Ruminococcaceae* genera (• and ••), *Ruminococcus*, *Enterobacteriaceae*• and *Streptococcaceae*•• (Fig. 1c, Extended Data Fig. 3i,j and Supplementary Table 3). • denotes classified but unnamed genus, •• denotes unclassified genus and ••• denotes genus names (owing to polyphyly of the genus) usually proposed by the Greengenes database curators, and only some of these can be found in NCBI.

From the identification of taxa that are selectively associated with CD or UC or controls that are robust to replication in two independent cohorts from central European environment, these results were compared with an unsupervised meta-analysis of previously published studies carried out in different contexts (Extended Data Fig. 1a). In general, different genera within Firmicutes are reported to be reduced in CD compared with controls or in IBD (where the study concerned grouped CD+UC) compared with controls (Group 1 in Extended Data Fig. 1a); further genera within Proteobacteria are increased in CD or CD+UC compared with controls (Group 2 in Extended Data Fig. 1a). For other taxa, our meta-analysis showed results that were largely inconsistent between different studies (Group 3 in Extended Data Fig. 1a) regardless of age, gender, ethnicity, sample type or sequencing method. We now show in Extended Data Fig. 1a (bold taxa labels) that our results verify the previous data for 8 of 11 taxa in Group 1 and 3 of 4 taxa in Group 2. Our results also show 18 significant additional taxa

differences between CD and controls or between UC and controls that are robust to replication between the two cohorts in the heterogeneous Group 3. These changes that are seen between CD and UC are largely a consequence of differences between the CD microbiota and non-IBD controls, whereas UC microbial consortia are relatively similar to healthy subjects (Extended Data Fig. 4a,b).

Using our extended portfolio of emblematic taxa, machine learning enabled prediction of CD and UC clusters with 87.7% and 83.3% prediction accuracy for Cohort 1 and Cohort 2, respectively (Supplementary Table 4). Focusing on multiple predicted operational taxonomic units (OTUs) within specific genera, we confirmed that *Ruminococcus gnavus* was consistently increased in abundance in CD compared to UC and that *Faecalibacterium prausnitzii* was consistently reduced in both cohorts^{26,27} (Extended Data Fig. 4c,d).

We next used genome-scale metabolic models to analyze our sequencing data and elucidate how metabolism is modulated within microbial IBD communities²⁸. Principal component analysis (PCA) analysis illustrates robust data at the metabolic reaction level with consistent metabolic subsystems increased in CD belonging to B-vitamin and LPS biosynthesis, heparan sulfate and chondroitin sulfate degradation and fatty acid oxidation (Extended Data Fig. 4e–j).

We also carried out a meta-analysis comparing IBD animal models including dogs, cats and laboratory mice with the available human studies. In an unsupervised analysis, the microbiota clustered according to species (Extended Data Fig. 5a), and genetic models were the most representative of the changes seen in human IBD patients (Extended Data Fig. 5b,c).

In summary, the distinct CD and UC disease groups are characterized by altered bacterial composition and lower diversity in CD patients compared with UC patients and non-IBD subjects (Fig. 1a–c and Extended Data Figs. 3, 4a,b). Bacterial richness and diversity between UC and non-IBD were less pronounced (Extended Data Fig. 3c–f). Our results indicated that there is a greater net loss of potentially beneficial taxa in CD compared with UC (Fig. 1c and Extended Data Figs. 3i,j and 4a,b). These included *Faecalibacterium* capable of butyrate production; the secondary bile acid producer *Oscillospira*; the dietary carbohydrate utilizer *Bifidobacterium*; *Ruminococcus* that degrades mucus and the complex carbohydrate utilizer *Blautia*^{29,30}.

Longitudinal comparisons of microbial community compositions between CD and UC patients. Most of the known IBD genetic risk factors that are shared between CD and UC are uninformative about disease subtypes^{15,20}. Within distinctive clinical pathology and sites of gastrointestinal tract involvement that define CD or UC³¹, disease subtypes may shape the severity and treatment responsiveness for individual patients. While IBD is characterized by chronic relapsing–remitting intestinal inflammation, current published studies have characterized these diseases largely in generic terms that might be insensitive to the longitudinal disease course and/or treatment responsiveness (Extended Data Fig. 1b). Our study was therefore designed to combine deep longitudinal clinical phenotypes between CD and UC and the different disease subtypes, allowing us to replicate key findings using the two independent longitudinal cohorts.

In a subset of our patients, we had biopsies that were taken at different times in the clinical course of disease. In both cohorts, ordination analysis showed that microbiota profiles were personalized according to each individual, over periods ranging up to 9 years (Extended Data Fig. 6a). Individual taxa were generally stable over time (Extended Data Fig. 6b), even when there had been an inter-current change in disease activity defined according to clinical phenotyping (Extended Data Fig. 6c).

To address the question of variation in the microbiota compositions according to parameters of disease activity by alternative

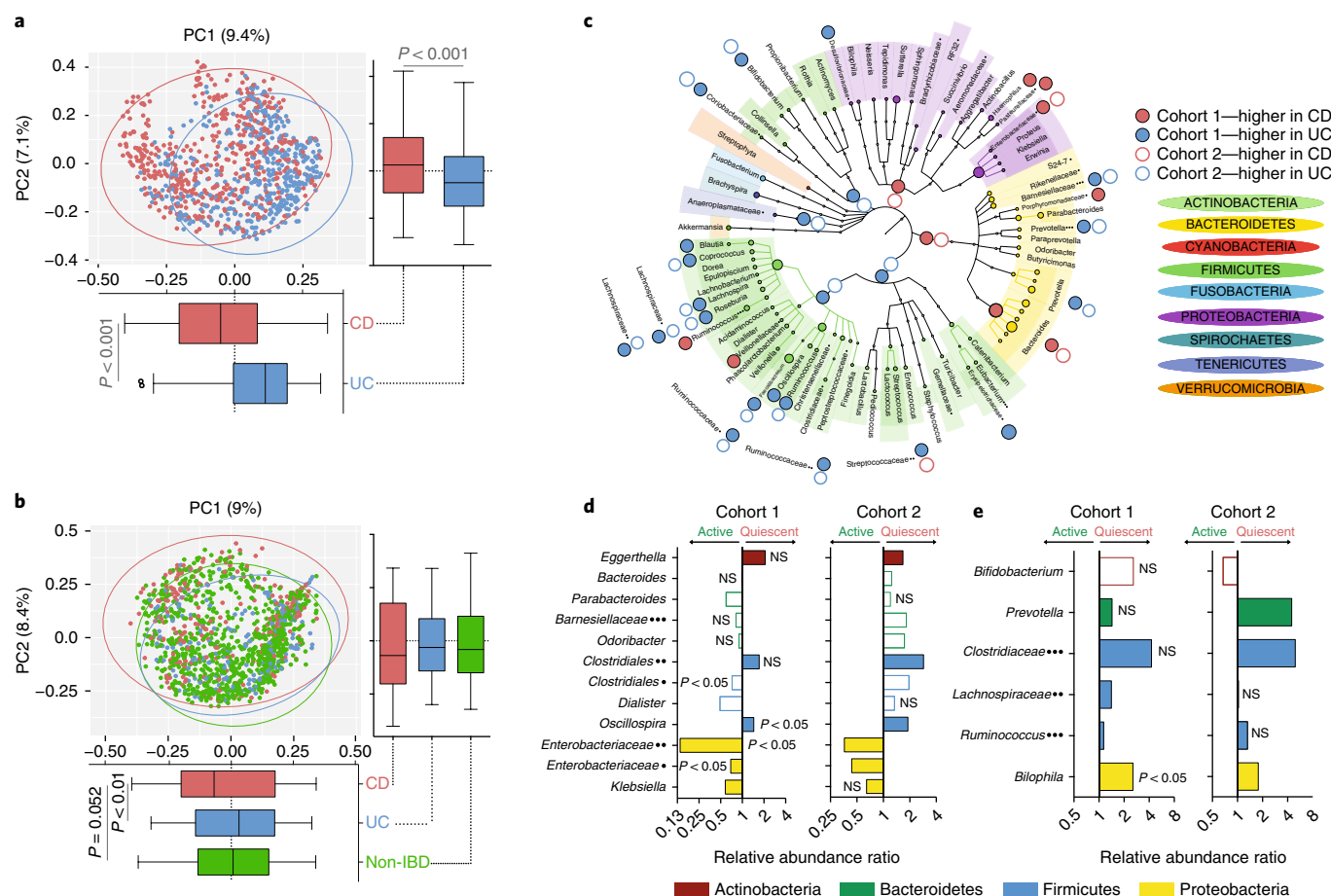


Fig. 1 | Distinct microbial features associated with IBD patients. **a,b**, Microbial clustering is shown based on Bray–Curtis dissimilarity principal coordinate analysis (PCoA) metrics using 2–4 intestinal biopsy samples per patient. There were 346 patients with IBD in Cohort 1 (**a**) and 156 patients with IBD and 227 non-IBD subjects in Cohort 2 (**b**) included in the study. Non-parametric analysis of variance (Adonis) was used to test for significant difference between groups on PCoA plot; in each case $P < 0.01$ for each disease group in each cohort, and a non-parametric two-sided Mann–Whitney U -test was used to analyze significant differences between ranks $\pm$ s.d. for PC1 and PC2. $P < 0.05$ was considered significant. Ellipsoids represent a 95% confidence interval surrounding each disease group. Box-and-whisker plots display first and third quartiles, and whiskers are from each quartile to the minimum or maximum. Significant comparisons are shown in **b**. Groups are not significantly different on the PC2 axis (**b**). **c**, Overall representation of overlapping significant taxa of disease groups between Cohort 1 and Cohort 2. **d,e**, Taxa significantly associated with ‘active’ or ‘quiescent’ clinical course in each cohort are plotted as relative abundance ratios in CD (**d**) and UC (**e**) disease groups. Analysis was performed on 494 CD and 447 UC samples in Cohort 1 (**d**) and 226 CD and 195 UC in Cohort 2 (**e**). Statistical analyses for taxonomy results were performed using the MaAsLin pipeline with Benjamini–Hochberg false-discovery rate (BH-FDR) (q value) (**c–e**). $q < 0.05$ was considered significant, and bars with $P < 0.05$ show that data were significant for individual taxa but did not reach significance after multiple-testing correction ($q < 0.2$) in **d,e**. Taxa with open bars show opposite relative abundance ratio between cohorts.

methods, we assessed the relationship of the microbiota composition to disease activity according to clinical indices of activity (CDAI for CD; MTWAI for UC) or fecal calprotectin measurements. Of 83 taxa analyzed, only *Enterobacteriaceae*• and *Klebsiella* in CD (Extended Data Fig. 6d), *Ruminococcus*••• and *Prevotella* in UC (Extended Data Fig. 6e) showed consistent compositional changes between the cohorts that were significantly aligned with clinical assessments of the disease activity (Fig. 1d,e) and according to the calprotectin biomarker (Extended Data Fig. 6d,e).

From these findings that microbiota composition profiles were mainly personalized according to the patient rather than the current disease activity, we proceeded to consider whether microbiotas within individuals could be related to their overall disease course or treatment responsiveness.

Co-occurrence pattern analysis of the intestinal ecosystems of IBD patients. We next used co-occurrence analysis to examine potential relationships between different taxa in either CD or UC

patients. The complex microbial ecosystem of the human intestinal tract is unevenly influenced by individual taxa within different microbial consortia^{10,11,24}; the ecological roles of individual community members are fundamentally important for overall biodiversity and are key to human health^{32,33}. Computationally derived co-occurrence associations are useful to infer influences between taxa in a range of ecosystems and are a starting point to examine the mechanisms of community construction and maintenance^{34,35}.

We first tested whether co-occurrence associations would differ strongly across different segments of the gastrointestinal tract. All patients in both cohorts of our study had received prior intestinal purging to remove fecal material and allow safe endoscopy with biopsies. Since purging involves substantial fluid flow along the length of the small and large intestine, as expected we did not see major microbial profile differences between different intestinal segments of individual patients (Extended Data Fig. 7a–f and Supplementary Table 5). Probably also because of the mixing effect of purging, we also found that the mucosal microbiota diversity of

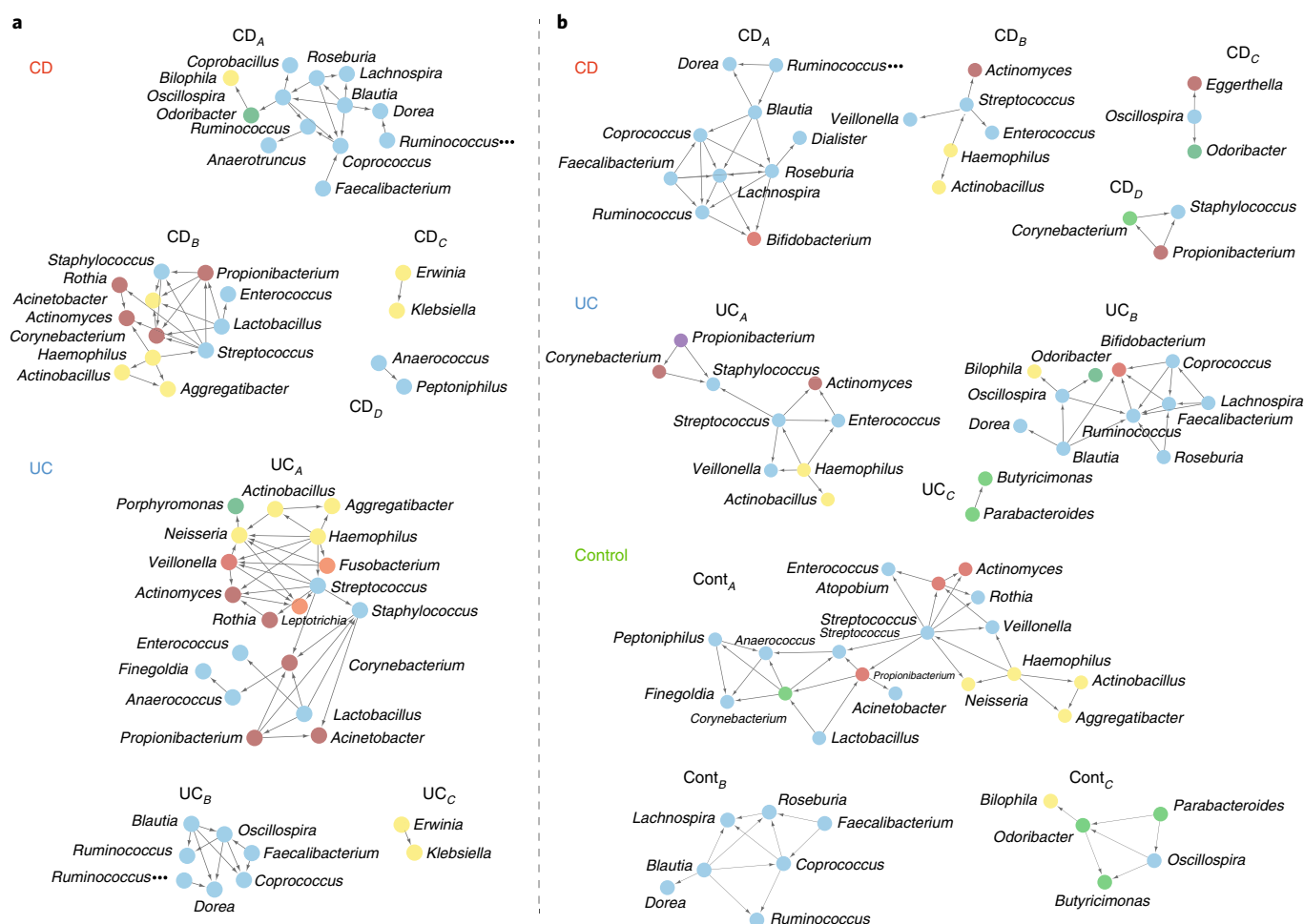


Fig. 2 | Microbial co-occurrence analysis of IBD and non-IBD subjects with gastrointestinal ecosystems. a,b, Ecosystem-specific co-occurrence patterns at genus level are visualized using network diagrams where microbial taxa represent nodes; the presence of a positive co-occurrence relationship based on correlation is represented by an edge for Cohort 1 (**a**) and Cohort 2 (**b**). Co-occurrence relationships with strong Spearman's correlation coefficient (ρ value > 0.5 and $P < 0.05$) are depicted with network diagram for each disease. Each cluster was assigned with a subscript to related disease or non-IBD control (for example, CD_A, UC_A, Cont_A).

inflamed and non-inflamed segments within the same patient were indistinguishable, independently of disease type (Extended Data Fig. 7g–j and Supplementary Table 5).

We then used all sequenced biopsies from different locations of the IBD patients' gut as replicate sampling groups for CD and UC independent ecosystems to address whether co-occurrence patterns among the microbial communities are the same in the different pathological contexts. We identified two consistent modules of co-occurring gut microbes for CD (CD_A and CD_B) and for UC (UC_A and UC_B) (Fig. 2 and Extended Data Fig. 8).

We found similar complexity and composition of the networks for both disease groups in each of the two cohorts (Supplementary Tables 6 and 7): the highest 'betweenness centrality scores' conceivably identifying keystone species within each ecosystem, showed that Firmicutes was a potential keystone phylum determining the network in both CD and UC disease groups of both cohorts (Extended Data Fig. 8a,b). However, Tenericutes for CD in Cohort 1 and Fusobacteria for UC in both cohorts were also probably important keystone phyla (Extended Data Fig. 8a,b). Measures of eigenvector and betweenness centrality showed in both cohorts that nodes contribute equivalently to the networks although there were more complex relationships of microbial taxa in CD (Extended Data Fig. 8c,d).

Based on betweenness centrality of each vertex, the top genera identified as keystone taxa in both cohorts with higher scores in CD compared to UC were *Streptococcus*, *Ruminococcus*, *Bifidobacterium*, *Propionibacterium*, *Corynebacterium*, *Coprococcus* and *Blautia* (Extended Data Fig. 8e,f). Further, the degree centrality measure helps identify the most prominent and influential taxa that we might miss with betweenness and Eigenvector centrality measures. Interestingly, *Faecalibacterium*, *Ruminococcus* of the Lachnospiraceae family, *Fusobacterium*, *Haemophilus* and *Parabacteroides* have some of the lowest betweenness scores although they show among the highest out-degree scores (number of edges originating in a node) in both disease groups of both cohorts, implying that *Faecalibacterium* and *Ruminococcus* are some of the most influential taxa in Cohort 1 (Extended Data Fig. 8g,h) and Cohort 2 (Extended Data Fig. 8i,j) independently of disease status. Overall, while there is similar complexity of the networks for both disease groups with relatively consistent modules of co-occurring microbial taxa within each ecosystem, the module labeled Cluster CD_A in both cohorts connects nodes of *Lachnospira*, *Blautia*, *Dorea*, *Coprococcus*, *Ruminococcus*, *Ruminococcus*..., *Faecalibacterium*, *Roseburia*, *Oscillospira* and *Bilophila* (Fig. 2): a similar Cluster UC_B is present in patients with UC, although it is less connected and overlaps with a Cluster Cont_B that is found in non-IBD controls

(Fig. 2). As shown subsequently in the paper, reductions in members of Cluster CD_A within the microbiota of CD patients generally characterize a worse outcome, less healthy lifestyles, poor responses to treatment and are associated with an increased risk of subsequent relapse in patients where the active segment of CD has been surgically removed.

Identification of critical variables in shaping the intestinal microbiota of IBD patients. Analyses of variability of the microbiota with chronic disease phenotype parameters suffer from the limitations that (1) correlations seen may be direct or indirect and (2) that the strength of these correlations may be disproportionate if taken in isolation from parameters of subject ageing, weight or other environmental influences. To allow a reasonable perspective of the relative influences of disease parameters or other personal variables, we then used the algorithms in the hierarchical all-against-all association (HALLA) tool for multi-resolution associations in high-dimensional, heterogeneous datasets with high power³⁶. Analysis performed at phylum (Fig. 3a,c) and genus rank (Fig. 3b,d) identified BMI and age at sampling/diagnosis as the most important variables correlating with gut microbiota of IBD patients. To a lesser extent, disease type, location and behavior and prior surgery were also important variables showing interactions with more than 60–70% of most abundant taxa of IBD patients. Overall lifestyles (including sport, smoking and alcohol consumption) were correlated with the gut microbiota profile in Cohort 1 (Fig. 3a–d): of these, smoking was the most significant factor. In contrast, biopsy location and inflammation status at the time of sampling were correlated with only a few taxa, which were largely inconsistent between cohorts.

Correlations with patient age and BMI in HALLA are consistent with published data showing strong effects of nutrition and intestinal physiological parameters on the microbiota^{12–14}. Although both CD and UC show age-dependent incidence, and active disease may cause weight loss, the greater strength of microbiota association with age at diagnosis than disease type or BMI compared with disease activity indicates that these are probably not dependent variables. Taken together with the similarities in taxa co-occurrence analysis at phylum and genus level between patients with CD or UC (Fig. 2), the relatively weaker correlations of the microbiota with disease parameters and responsiveness to treatment (Fig. 3) indicate that detailed minor compositional changes may be associated with disease type or subtype on a background of larger effects of body habitus or age.

The assessment of critical variables that may be related to the microbial profile including disease status, therapeutics and environmental factors. We next focused specifically on the relationship of the microbiota within the grouping of either CD or UC patients, according to the detailed disease phenotype or subtype, known trigger factors and treatment responsiveness using a sparse multivariate statistical approach.

We first analyzed whether the patient's lifestyle such as alcohol consumption, smoking habit or physical activity were associated with alterations of the gut microbiota composition (Supplementary Table 8). We only identified significant changes in microbial taxonomy (in every case corrected for multiple testing) in relation to host lifestyle in CD patients. Increasing exercise was positively correlated with the relative abundance of taxa from the Firmicutes and Bacteroidetes phyla, including *Clostridiales*•, *Blautia*, *Barnesiellaceae*•, *Faecalibacterium*, *Erysotrichaceae*•, *Ruminococcaceae*• genera and negatively correlated with *Veillonella* abundance (Extended Data Fig. 9a). Tobacco smoking or alcohol consumption *Parabacteroides* and *Clostridiales*•• were, respectively, associated within CD: both were negatively associated with *Sutterella* (Extended Data Fig. 9b,c). *Parabacteroides* was reduced in

familial CD patients (Extended Data Fig. 9d). Notably, exercise in CD patients, which is indicative of a healthier lifestyle²³, is associated with taxa including *Clostridiales*••, *Blautia*, *Faecalibacterium*, *Erysotrichaceae*• and *Ruminococcaceae*• that are represented in Cluster CD_A of the co-occurrence analysis (Fig. 2), although we cannot distinguish whether the alterations are either bias due to reduced disease activity in those able to participate in sport³⁷ or a beneficial alteration in microbiota composition.

Disease therapy including monoclonal antibody treatment against TNF- α and steroid use are effective treatments for many IBD patients^{38–40}. Unfortunately, not all patients are responsive to these therapies probably because of the variable redundancy of different triggering mechanisms of intestinal inflammation and the variable redundancy of TNF- α as an inflammatory cytokine⁴¹. We asked whether the responsiveness of IBD patients to different therapies could be correlated with differences in microbiota composition. Initial microbial analysis of CD and UC samples showed that there was no significant difference in alpha diversity for either cohort based on whether the patient had responded to treatment (Extended Data Fig. 9e,f). However, we did find significant differences in gut microbial composition between patients responding or not responding to TNF- α inhibitor therapy (Fig. 4a–c). Although there were no significant differences in the small numbers of UC patients that had been treated in both cohorts, comparison of the CD groups responding to treatment versus the CD patients failing to respond to treatment, showed and replicated increased *Bifidobacterium*, *Collinsella*, *Lachnospira*, *Lachnospiraceae*•, *Roseburia*, *Eggerthella* taxa and reduced *Phascolarctobacterium* associated with treatment success (Fig. 4b,c). *Bifidobacterium* are well recognized for their capacity for metabolism of oligosaccharides that may be resistant to host digestion and have been used as probiotics in IBD⁴². *Lachnospiraceae*• and *Roseburia* are both members of association CD_A (Fig. 2) and are characterized by production of short chain fatty acids that are a carbon source for intestinal epithelial cells and inducers of regulatory T cells^{43,44}. *Phascolarctobacterium* is common in the human gastrointestinal tract⁴⁵ and can produce acetate and propionate, although it has also been associated with inflammatory dysbiosis in animal models⁴⁶.

We asked whether the responsiveness of IBD patients to corticosteroid therapies could also be correlated with differences in microbiota composition. However, although significant differences in gut microbial taxa composition between groups responding to treatment versus groups failing to respond were observed in the PCoA analysis of both cohorts (Extended Data Fig. 9i,j), unlike the results with anti-TNF- α biological treatments, these were not replicated between the different cohorts (Fig. 4d and Extended Data Fig. 9k–m).

A major burden for patients with CD and UC is that the diseases follow a relapsing-remitting course of recurrent exacerbations and symptomatic improvements over many years, with unpredictable effects on morbidity, quality of life and the need for healthcare (Supplementary Table 9). The analysis of samples from patients with a quiescent or a relapsing course showed that while there were no significant differences in microbial alpha diversities either cohort, in CD, *Eggerthella*, *Clostridiales*•• and *Oscillospira* showed consistent replicated increases in relative abundance in patients with quiescent disease over time while *Enterobacteriaceae*•, *Enterobacteriaceae*•• and *Klebsiella* were associated with a more severe clinical course (Fig. 1d,e and Extended Data Fig. 9n–q). *Oscillospira* associated with the more benign course of CD are also members of association Cluster CD_A in Cohort 1 and in Cluster CD_C together with *Eggerthella* in Cohort 2 (Fig. 2).

Disease location is secondary determinant in gut microbiota profile of IBD patients. Separating the CD patients according to whether they had colonic disease or ileal CD disease (comprising ileal CD

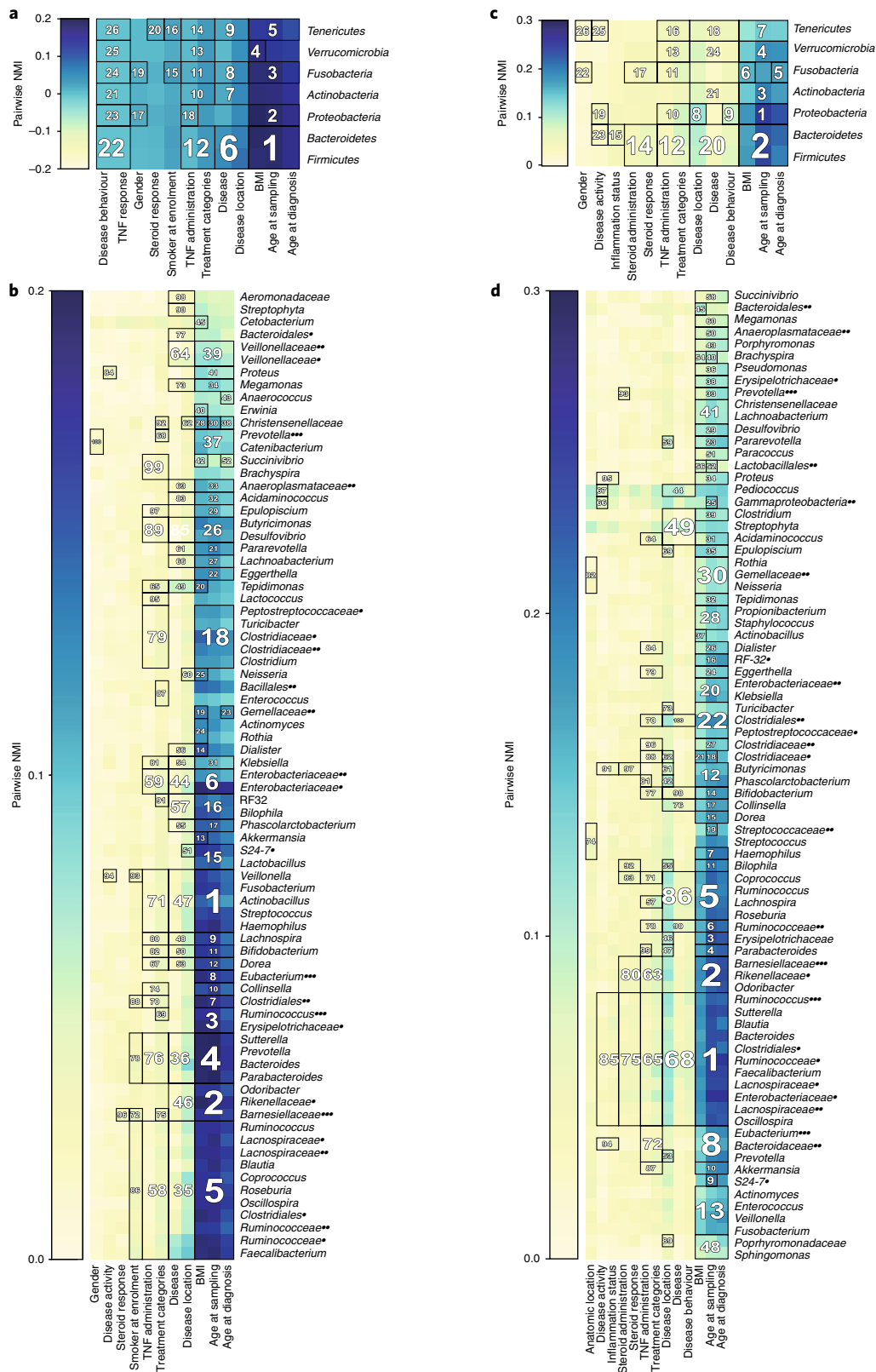


Fig. 3 | Clinical determinants correlate with intestinal microbiota composition of IBD patients. a–d, Hierarchical analysis performed using an end-to-end statistical method (HALLA) at phylum (**a,c**) and genus rank (**b,d**) for Cohort 1 (**a,b**) and Cohort 2 (**c,d**). Significant metadata among many clinical parameters were plotted after BH-FDR correction. $q < 0.05$ was considered significant (**a–d**). Association rank was sorted by high similarity score. The color intensity of the heatmap (pairwise normalized mutual information: NMI) and the numbers in the cells identify significant pairs of features (clinical determinants versus microbial composition) of patients with IBD.

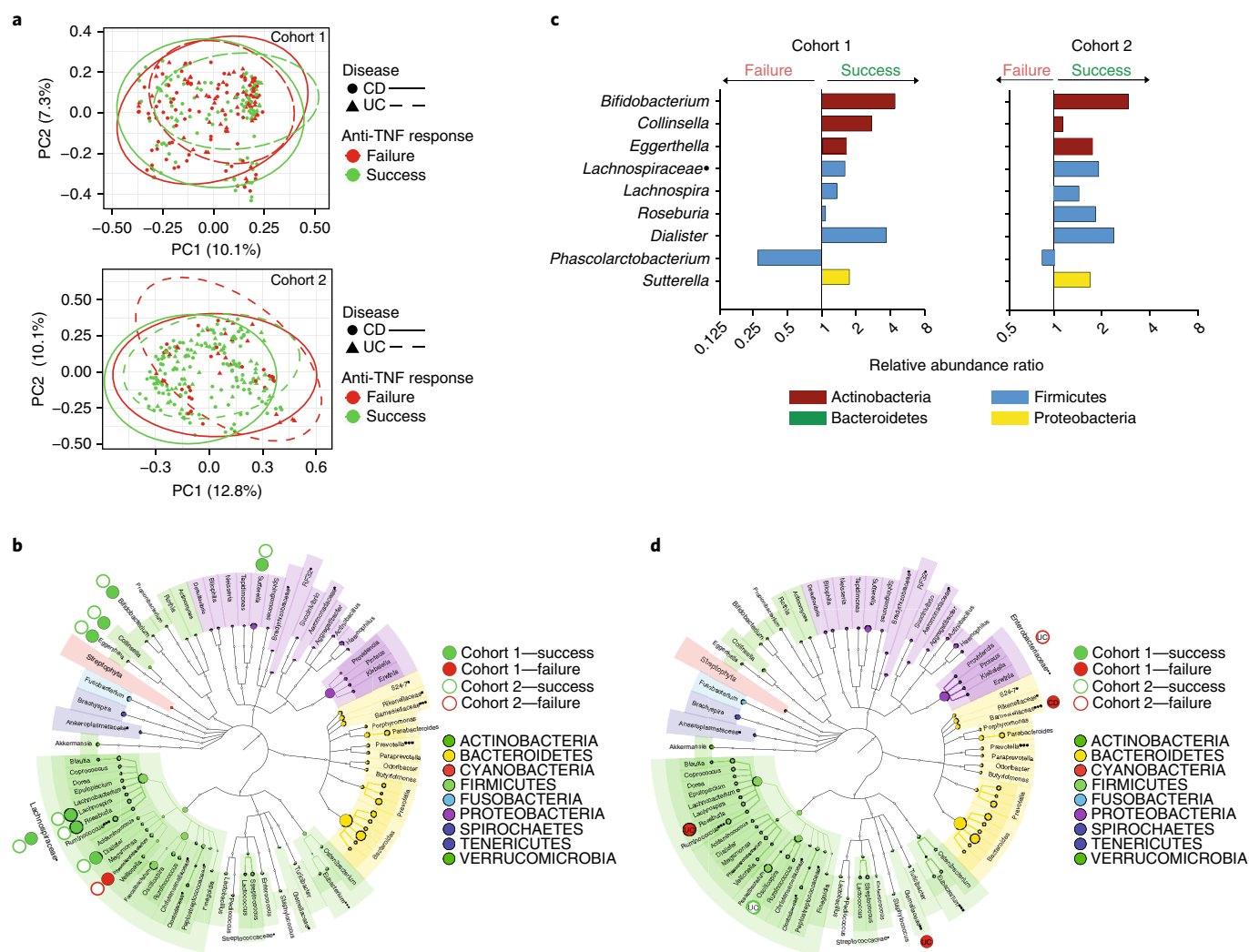


Fig. 4 | Intestinal microbiota composition of IBD patients with responsiveness to anti-TNF- α or corticosteroid therapies. **a, Microbial clustering is shown based on Bray-Curtis dissimilarity PCoA metrics of intestinal biopsy samples from IBD patients who responded or did not respond to anti-TNF- α therapy, showing reproducibility in both cohorts. CD (solid line) and UC (dashed line) are used to identify the disease groups in PCoA plot. Ellipsoids represent a 95% confidence interval surrounding each group. An Adonis test assessed the significant difference between groups in PCoA plot: $P < 0.01$ is for each compared group in both cohorts. **b**, Overall representation of overlapping significant taxa in responding and non-responding groups in each cohort are shown using phylogenetic trees for anti-TNF- α therapy. Closed circles correspond to Cohort 1, and open circles correspond to Cohort 2; green coloration represents higher relative abundances in responding groups, and red coloration represents higher relative abundance in non-responding groups. Significant taxa are plotted with respective circles. **c**, Significant unique microbial taxa identified as a signature of responding (success) and non-responding (failure) groups CD and UC are shown for Cohort 1 and Cohort 2 after using MaAsLin with BH-FDR (q value). **d**, Overall representation of overlapping significant taxa of responding and non-responding groups in each cohort are shown using phylogenetic trees for corticosteroid therapy. $q < 0.05$ was considered significant, and these are plotted (**b-d**). Analysis was performed on 88 success and 91 failure CD samples with 19 success and 48 failure UC samples in Cohort 1 and 129 success and 27 failure CD samples with 53 success and 18 failure UC samples in Cohort 2 (**a-d**).**

and ileo-colonic CD subtypes combined), we found no significant differences between non-IBD, colonic UC and colonic CD samples, although the gut microbiota composition of patients with colonic CD was closer to the microbiota of patients with UC than to that of patients with ileal CD, as previously reported³¹ (Fig. 5a–c). Metabolic reaction set analysis of IBD subtype microbiotas, which were consistently seen across the two patient cohorts, revealed that LPS biosynthesis, vitamin B2 and vitamin B7 metabolism were higher in colonic CD subtype compared with (colonic) UC (Fig. 5d), and LPS biosynthesis, peptide and folate metabolism, fatty acid synthesis, chondroitin sulfate and heparan sulfate degradation were functionally elevated in patients with ileal CD subtype compared with UC (Fig. 5e). Glycerophospholipid metabolism and cell wall biosynthesis were increased in UC microbiota compared with colonic CD or ileal CD, respectively. There was no

significant metabolic reaction set differences between ileal and colonic CD, although any distinction may be confounded by colonic purgatives prior to sampling as discussed previously.

Post-surgical CD patients have altered intestinal microbiota prior to disease recurrence. Although the outlook for IBD patients has been transformed by the advent of treatments that can be directed against selected pro-inflammatory pathways, surgery is still required in up to 40% of CD and 30% of UC patients at some point during their lifetime^{47,48}. UC surgery normally comprises removal of the entire colon and hence the source of the intestinal inflammation. Before the advent of current biological therapies, clinical studies established that removal of inflamed intestinal segments in CD cannot cure the

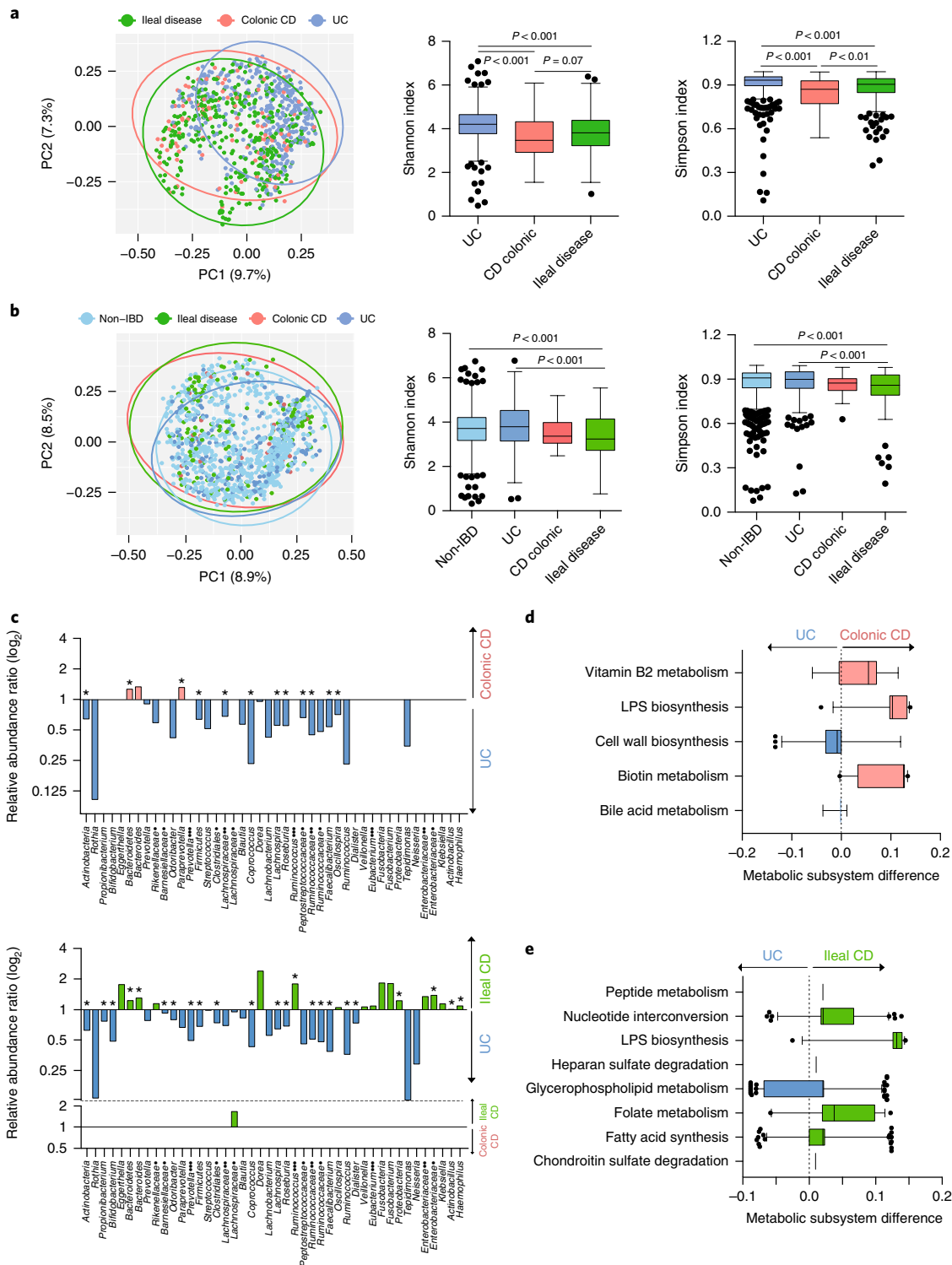


Fig. 5 | Disease locations are critically important determinants that correlate with intestinal microbiota of IBD patients. **a,b**, Microbial composition differences in IBD patients within different disease locations in Cohort 1 (**a**) and Cohort 2 (**b**) are plotted using beta diversity (Bray-Curtis dissimilarity PCoA) and alpha diversity (Shannon and Simpson indices). Ellipsoids in PCoA plots represent a 95% confidence interval surrounding each group. Box-and-whisker plots display first and third quartiles, and whiskers are from each quartile to the minimum or maximum. There is significant separation between groups on the PCoA plot ($P < 0.05$). Significant differences between groups in alpha diversity analysis were determined by an ordinary one-way ANOVA corrected for multiple comparisons using BH-FDR. **c**, Only significant bacterial abundance differences between colonic CD, ileal CD and colonic UC samples are plotted using the samples identified in **a** and **b**. Differences between tested groups were determined using MaAsLin with BH-FDR correction (q value). Asterisk indicates the similar significant taxonomic changes observed in Cohort 2. **d,e**, Only significantly different metabolic subsystems in colonic CD samples (**d**) and ileal involved CD samples (**e**) when compared with UC samples are shown. $q < 0.05$ was considered significant (**c-e**) and error bars on boxplots are standard deviations (**d,e**). The same samples described in **a** and **b** were used for analysis of metabolic subsystems (**d,e**). For analysis, 370 ileal, 110 colonic and 447 UC in Cohort 1 and 203 ileal, 28 colonic and 221 UC in Cohort 2 were used (**a-e**).

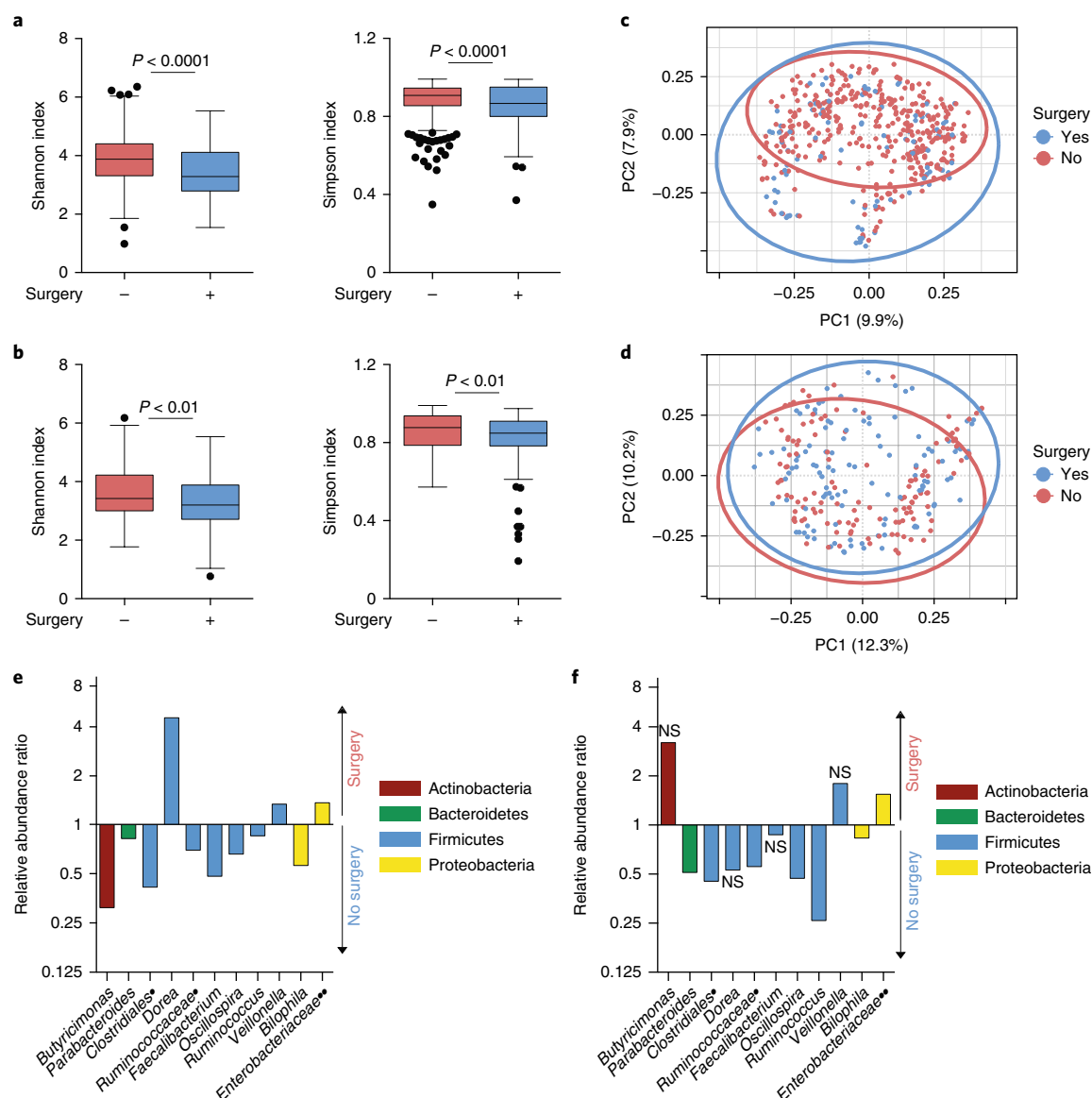


Fig. 6 | Altered microbiota profiles in inactive CD patients following surgical resection of active disease segments. **a–d**, Species richness between CD patients following surgery or without surgical resection was calculated using alpha diversity (Shannon and Simpson indices) for Cohort 1 (**a**) and Cohort 2 (**b**). Box-and-whisker plots display first and third quartiles, and whiskers are from each quartile to the minimum or maximum. Microbial composition differences between these samples was calculated with PCoA on Bray–Curtis distance for Cohort 1 (**c**) and Cohort 2 (**d**). Blue corresponds to CD patients who had undergone surgical resection, and red corresponds to CD patients without surgery. Ellipsoids represent a 95% confidence interval surrounding each group (**c,d**). Non-parametric two-sided Mann–Whitney *U*-test and Adonis test were used to identify the statistical significant differences between groups for alpha (**a,b**) and beta diversity (**c,d**), respectively. There is a significant separation on PCoA plots between compared groups analyzed in **c** and **d** ($P < 0.05$). (**e,f**) Only significant taxa ($q < 0.05$) associated with surgical resection status are plotted as relative abundance ratio for CD in Cohort 1 (**e**) and Cohort 2 (**f**). In Cohort 2, non-significant data ($q > 0.05$) are marked with NS (**f**). 105 (+) and 109 (–) samples in Cohort 1 and 96 (+) and 371 samples (–) in Cohort 2 were analyzed (**a–f**).

disease, because without ongoing treatment CD recurs in over 70% of patients at the anastomotic margin where previously uninfamed segments have been joined⁴⁹. Post-surgical patients represent an extremely interesting group because they usually do not have detectable Crohn's disease (since it has been surgically removed), but they will almost inevitably acquire it again unless they receive immunosuppressive medications. Thus, the microbiota in patients after surgery may be emblematic of the patient susceptible to CD (potential CD), without it being present at the time of sampling.

We therefore compared the post-operative microbiota in CD patients that had already undergone resection of the ileo-cecal

region or a segment or small intestine, with CD patients that had not needed an operation. The operated CD patients had distinct gut microbial profiles with reduced species richness in CD samples from both cohorts collected >3 months post-operatively (Fig. 6a,b). Further, composition differences of the intestinal microbiota were found between inactive CD patients with or without prior surgery (Fig. 6c,d). The taxa that were principally included replicated reductions in the following members of Cluster CD_A, *Ruminococcaceae**, *Oscillospira*, *Ruminococcus* and *Bifidobacterium*. In both cohorts *Parabacteroides* and *Clostridiales* were also reduced in inactive post-surgical patients, and *Enterobacteriaceae*** were significantly and reproducibly increased (Fig. 6e,f). These compositional

changes in post-surgical CD patients without active disease, but with an almost inevitable probability of relapse in a previously uninvolved intestinal segment if untreated (biased towards reductions in Cluster CD_A) may be capable of triggering CD relapse.

The implication that a disrupted composition of Cluster CD_A makes patients susceptible to CD relapse, even in post-surgical CD patients who temporarily have no activity because the active segment has been removed, implies that this cluster may be a biomarker of resistance to relapse of active disease independently of current inflammatory activity. It would follow that Cluster CD_A taxa are not themselves intrinsically associated with CD disease activity, unlike the *Enterobacteriaceae*• (Extended Data Fig. 6d). We found no significant changes in the Cluster CD_A taxa in longitudinally studied patients judged by alterations in disease severity status over time (Extended Data Fig. 10a). As an alternative approach, we also asked whether the calprotectin level could be a more sensitive luminal inflammatory biomarker for possible alterations in the association cluster CD_A with increased intestinal inflammation. Again, no CD_A taxon showed a significant correlation with calprotectin levels (Extended Data Fig. 10b). Overall, this data supports the idea of importance to this cluster of taxa to inhibit relapse, rather than being associated as a biomarker for the patients with the most inflammatory and immunologically active disease.

Discussion

In this paper, we report the relationship of intestinal biopsy microbiotas from patients with UC and CD obtained from two cohorts of patients over many years and capturing information about clinical phenotypes including responsiveness to different treatments, lifestyle differences and the severity of the relapsing-remitting clinical course. In our patient cohorts, relatively few taxa were shown to be correlated with the current activity of disease, whether assessed by clinical phenotyping or by the fecal calprotectin inflammatory biomarker.

We have confirmed previous reports of the microbiota differences between CD patients and controls, and UC patients and controls. In addition to the known associations of *Faecalibacterium* and the Firmicutes phylum with UC, and Bacteroidetes and *Ruminococcus*••• with CD, we have now shown (and replicated between cohorts) that other Firmicutes genera (including *Lachnospiraceae*•, *Clostridia*••, *Roseburia*, *Blautia*, *Ruminococcus* genus of the Ruminococcaceae family) are reduced in CD compared with UC, while a *Ruminococcus*••• genus of the Lachnospiraceae family was increased in CD compared with UC.

The notion that the dysbiosis we observed may be causal is supported by the special situation in patients in whom CD segments have been surgically removed, yet relapse in another previously uninvolved part of the intestine is extremely likely without immunosuppression.

In line with other studies that show major effects from nutrition and host intestinal physiology on the microbiota^{13,14}, we have found that the differences that are attributable to CD or UC, or to the different lifestyle or therapeutic factors, are far less than those engendered by age or body habitus. Since IBD are highly heterogeneous, we need to be cautious about associations between microbial gut profiles and clinical phenotypes with small but significant effect sizes. Nevertheless, a major biomedical problem is why the phenotypic behavior and therapeutic responsiveness of CD and UC and their subtypes are so different between patients. Here we could show and replicate between cohorts that increased *Bifidobacterium*, *Collinsella*, *Lachnospira*, *Lachnospiraceae*•, *Roseburia*, and *Eggerthella* taxa were related to the successful outcome of anti-TNF- α treatment in CD. This is consistent with the finding that *Lachnospiraceae*• and *Roseburia* taxa are reduced in CD and that these taxa produce short chain fatty acids as a carbon source for intestinal epithelial cells and induction of regulatory T cells^{43,44}. We

found that *Roseburia* was a probable keystone taxon in microbial consortial association analysis of CD but not UC. In murine and in vitro models, *Roseburia* can promote innate gut immunity of anti-microbial peptide up-regulation and improved gut barrier function in addition to regulatory T cell induction⁵⁰. Interestingly, reduction of two *Lachnospiraceae*, *Blautia* and *Roseburia* also predict pouchitis in patients that have undergone colectomy with reconstruction of a pouch of small intestine to substitute the rectum⁵¹. While reduced *Roseburia* has also been reported in UC compared with controls⁵², our findings are that the effect of the taxon on the clinical course is strongest in CD. We also found that *Clostridiales*•, *Blautia*, *Faecalibacterium*, *Eryselotrichaceae*• and *Ruminococcaceae*• were associated with regular exercise in CD. *Blautia*, *Faecalibacterium* and *Ruminococcus* are all probable keystone taxa in CD and UC, but with greater connectivity in the association analysis in CD. Many of the above taxa are connected in Cluster CD_A in the association analysis (Fig. 2).

Many of the taxa that are altered in association with a poor prognosis of disease course, therapeutic unresponsiveness, a less healthy lifestyle or likelihood of relapse in post-surgical patients are within the Cluster CD_A, identified through taxa association analysis. This contains taxa characterized by short chain fatty acid production through different metabolic pathways⁵³ that may increase the availability of intraluminal intestinal oxygen and allow facultative aerobes such as Enterobacteriaceae family to bloom through alterations in intestinal epithelial cell metabolism⁵⁴. We found no significant correlations between intestinal inflammation and relative abundance of CD_A taxa measured either according to clinical assessment of alterations in disease activity or in relation to the fecal calprotectin inflammatory biomarker. Taken together, this supports the idea of the importance to this cluster of taxa in inhibition of the disease, probably in a particular context of individual patient genetic susceptibility, rather than merely being associated as a biomarker for the patients with the most inflammatory and immunologically active disease.

Our data have shown a cluster of taxa including those characterized by their short chain fatty acid production that generally distinguish CD from UC, and are predictive of a quiescent course, a healthier lifestyle and responsiveness to anti-TNF- α therapy^{30,52,55–120}. These taxa have been replicated across two patient cohorts in the study and are deficient in patients after surgical resection that can be considered at very high risk for CD relapse despite removal of the earlier diseased intestinal segments. There is considerable interest in manipulation of the microbiota as an adjunctive therapy to improve the course and responsiveness of disease. Our results predict that fecal transplants or directed manipulation of the intestinal microbiota may be most successful where defects in Cluster CD_A can be corrected.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41591-018-0308-z>.

Received: 16 March 2018; Accepted: 19 November 2018;
Published online: 21 January 2019

References

1. Strober, W., Fuss, I. J. & Blumberg, R. S. The immunology of mucosal models of inflammation. *Annu. Rev. Immunol.* **20**, 495–549 (2002).
2. Blumberg, R. & Powrie, F. Microbiota, disease, and back to health: a metastable journey. *Sci. Transl. Med.* **4**, 137rv137 (2012).
3. Hegazy, A. N. et al. Circulating and tissue-resident CD4(+) T cells with reactivity to intestinal microbiota are abundant in healthy individuals and function is altered during inflammation. *Gastroenterology* **153**, 1320–1337 e1316 (2017).

4. Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298–306 (2011).
5. Rutgeerts, P. et al. Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet* **338**, 771–774 (1991).
6. Zachos, M., Tondeur, M. & Griffiths, A. M. Enteral nutritional therapy for induction of remission in Crohn's disease. *Cochrane Database Syst. Rev.* **4**, CD000542 (2007).
7. Cummings, J. H. & Kong, S. C. Probiotics, prebiotics and antibiotics in inflammatory bowel disease. *Novartis. Found. Symp.* **263**, 99–111 (2004). discussion 111–114, 211–118.
8. Gecse, K. B. et al. A global consensus on the classification, diagnosis and multidisciplinary treatment of perianal fistulising Crohn's disease. *Gut* **63**, 1381–1392 (2014).
9. Faith, J. J. et al. The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
10. Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
11. Schloissnig, S. et al. Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50 (2013).
12. Flint, H. J., Duncan, S. H. & Louis, P. The impact of nutrition on intestinal bacterial communities. *Curr. Opin. Microbiol.* **38**, 59–65 (2017).
13. Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
14. Vandeputte, D. et al. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* **65**, 57–62 (2016).
15. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
16. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
17. Uhlig, H. H. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. *Gut* **62**, 1795–1805 (2013).
18. Lee, J. C. & Lennard-Jones, J. E. Inflammatory bowel disease in 67 families each with three or more affected first-degree relatives. *Gastroenterology* **111**, 587–596 (1996).
19. Lee, J. C. et al. Gene expression profiling of CD4+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J. Clin. Invest.* **121**, 4170–4179 (2011).
20. Lee, J. C. et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat. Genet.* **49**, 262–268 (2017).
21. Park, S. H., Aniwan, S. & Loftus, E. V. Jr. Advances in the use of biologics and other novel drugs for managing inflammatory bowel disease. *Curr. Opin. Pharmacol.* **37**, 65–71 (2017).
22. Pillai, N., Dusheiko, M., Burnand, B. & Pittet, V. A systematic review of cost-effectiveness studies comparing conventional, biological and surgical interventions for inflammatory bowel disease. *PLoS ONE* **12**, e0185500 (2017).
23. Engels, M., Cross, R. K. & Long, M. D. Exercise in patients with inflammatory bowel diseases: current perspectives. *Clin. Exp. Gastroenterol.* **11**, 1–11 (2018).
24. Eckburg, P. B. et al. Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
25. Sartor, R. B. & Wu, G. D. Roles for intestinal bacteria, viruses, and fungi in pathogenesis of inflammatory bowel diseases and therapeutic approaches. *Gastroenterology* **152**, 327–339.e4 (2017).
26. Png, C. W. et al. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105**, 2420–2428 (2010).
27. Sokol, H. et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).
28. Magnusdottir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
29. Fujimoto, T. et al. Decreased abundance of *Faecalibacterium prausnitzii* in the gut microbiota of Crohn's disease. *J. Gastroenterol. Hepatol.* **28**, 613–619 (2013).
30. Lopez-Siles, M. et al. Mucosa-associated *Faecalibacterium prausnitzii* phylotype richness is reduced in patients with inflammatory bowel disease. *Appl. Environ. Microbiol.* **81**, 7582–7592 (2015).
31. Imhann, F. et al. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* **67**, 108–119 (2018).
32. Goodman, A. L. & Gordon, J. I. Our unindicted coconspirators: human metabolism from a microbial perspective. *Cell. Metab.* **12**, 111–116 (2010).
33. Muegge, B. D. et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
34. Mainali, K. P. et al. Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. *PLoS ONE* **12**, e0187132 (2017).
35. Williams, R. J., Howe, A. & Hofmockel, K. S. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Front. Microbiol.* **5**, 358 (2014).
36. Rahnavard, G. F. et al. HALLA: Hierarchical All-against-All significance testing. <http://huttenhower.sph.harvard.edu/halla> (2017).
37. Plevinsky, J. M., Wojtowicz, A. A., Pouloupoulos, N., Schneider, K. L. & Greenley, R. N. Perceived impairment in sports participation in adolescents with inflammatory bowel diseases: a preliminary examination. *J. Pediatr. Gastroenterol. Nutr.* **66**, 79–83 (2018).
38. Torres, J., Mehandru, S., Colombel, J. F. & Peyrin-Biroulet, L. Crohn's disease. *Lancet* **389**, 1741–1755 (2017).
39. Khanna, R. et al. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* **386**, 1825–1834 (2015).
40. Ungaro, R., Mehandru, S., Allen, P. B., Peyrin-Biroulet, L. & Colombel, J. F. Ulcerative colitis. *Lancet* **389**, 1756–1770 (2017).
41. Magnusson, M. K. et al. Anti-TNF therapy response in patients with ulcerative colitis is associated with colonic antimicrobial peptide expression and microbiota composition. *J. Crohns Colitis* **10**, 943–952 (2016).
42. Ghouri, Y. A. et al. Systematic review of randomized controlled trials of probiotics, prebiotics, and synbiotics in inflammatory bowel disease. *Clin. Exp. Gastroenterol.* **7**, 473–487 (2014).
43. Topping, D. L. & Clifton, P. M. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol. Rev.* **81**, 1031–1064 (2001).
44. Arpaia, N. et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451–455 (2013).
45. Wu, F. et al. *Phascolarctobacterium faecium* abundant colonization in human gastrointestinal tract. *Exp. Ther. Med.* **14**, 3122–3126 (2017).
46. Chen, J. et al. Dysbiosis of intestinal microbiota and decrease in paneth cell antimicrobial peptide level during acute necrotizing pancreatitis in rats. *PLoS ONE* **12**, e0176583 (2017).
47. Hancock, L. & Mortensen, N. J. How often do IBD patients require resection of their intestine? *Inflamm. Bowel Dis.* **14**, S68–S69 (2008).
48. Olivera, P., Spinelli, A., Gower-Rousseau, C., Danese, S. & Peyrin-Biroulet, L. Surgical rates in the era of biological therapy: up, down or unchanged? *Curr. Opin. Gastroenterol.* **33**, 246–253 (2017).
49. Rutgeerts, P. et al. Predictability of the postoperative course of Crohn's disease. *Gastroenterology* **99**, 956–963 (1990).
50. Patterson, A. M. et al. Human gut symbiont roseburia hominis promotes and regulates innate immunity. *Front. Immunol.* **8**, 1166 (2017).
51. Machiels, K. et al. Specific members of the predominant gut microbiota predict pouchitis following colectomy and IPAA in UC. *Gut* **66**, 79–88 (2017).
52. Machiels, K. et al. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* **63**, 1275–1283 (2014).
53. Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. Links between diet, gut microbiota composition and gut metabolism. *Proc. Nutr. Soc.* **74**, 13–22 (2015).
54. Byndloss, M. X. et al. Microbiota-activated PPAR-gamma signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570–575 (2017).
55. Liguori, G. et al. Fungal dysbiosis in mucosa-associated microbiota of Crohn's disease patients. *J. Crohns Colitis* **10**, 296–305 (2016).
56. Papa, E. et al. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS ONE* **7**, e39242 (2012).
57. Santoru, M. L. et al. Cross-sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. *Sci. Rep.* **7**, 9523 (2017).
58. Pedamallu, C. S. et al. Metagenomic characterization of microbial communities in situ within the deeper layers of the ileum in crohn's disease. *Cell. Mol. Gastroenterol. Hepatol.* **2**, 563–566.e565 (2016).
59. Kellermayer, R. et al. Microbiota separation and C-reactive protein elevation in treatment-naïve pediatric granulomatous Crohn disease. *J. Pediatr. Gastroenterol. Nutr.* **55**, 243–250 (2012).
60. Frank, D. N. et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).
61. Willing, B. P. et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854.e1841 (2010).
62. Wang, W. et al. Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *J. Clin. Microbiol.* **52**, 398–406 (2014).

63. Rajilic-Stojanovic, M., Shanahan, F., Guarner, F. & de Vos, W. M. Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm. Bowel Dis.* **19**, 481–488 (2013).
64. Seksik, P. et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* **52**, 237–242 (2003).
65. Halfvarson, J. et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2**, 17004 (2017).
66. Erickson, A. R. et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **7**, e49138 (2012).
67. Eun, C. S. et al. Does the intestinal microbial community of Korean Crohn's disease patients differ from that of western patients? *BMC. Gastroenterol.* **16**, 28 (2016).
68. Forbes, J. D., Van Domselaar, G. & Bernstein, C. N. Microbiome survey of the inflamed and noninflamed gut at different compartments within the gastrointestinal tract of inflammatory bowel disease patients. *Inflamm. Bowel Dis.* **22**, 817–825 (2016).
69. Andoh, A. et al. Multicenter analysis of fecal microbiota profiles in Japanese patients with Crohn's disease. *J. Gastroenterol.* **47**, 1298–1307 (2012).
70. Rehman, A. et al. Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. *J. Med. Microbiol.* **59**, 1114–1122 (2010).
71. Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D. & Garcia-Gil, L. J. Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction–denaturing gradient gel electrophoresis. *Inflamm. Bowel Dis.* **12**, 1136–1145 (2006).
72. Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
73. Ashton, J. J. et al. 16S sequencing and functional analysis of the fecal microbiome during treatment of newly diagnosed pediatric inflammatory bowel disease. *Medicine* **96**, e7347 (2017).
74. Kabeerdoss, J., Jayakanthan, P., Pugazhendhi, S. & Ramakrishna, B. S. Alterations of mucosal microbiota in the colon of patients with inflammatory bowel disease revealed by real time polymerase chain reaction amplification of 16S ribosomal ribonucleic acid. *Indian J. Med. Res.* **142**, 23–32 (2015).
75. Kolho, K. L. et al. Fecal microbiota in pediatric inflammatory bowel disease and its relation to inflammation. *Am. J. Gastroenterol.* **110**, 921–930 (2015).
76. Martinez-Medina, M. et al. Molecular diversity of *Escherichia coli* in the human gut: new ecological evidence supporting the role of adherent-invasive *E. coli* (AIEC) in Crohn's disease. *Inflamm. Bowel Dis.* **15**, 872–882 (2009).
77. Duboc, H. et al. Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. *Gut* **62**, 531–539 (2013).
78. Schwartz, A. et al. Microbiota in pediatric inflammatory bowel disease. *J. Pediatr.* **157**, 240–244 e241 (2010).
79. Kaakoush, N. O. et al. Microbial dysbiosis in pediatric patients with Crohn's disease. *J. Clin. Microbiol.* **50**, 3258–3266 (2012).
80. Walker, A. W. et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC. Microbiol.* **11**, 7 (2011).
81. Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F. & van Zanten, S. J. O. V. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J. Clin. Microbiol.* **44**, 4136–4141 (2006).
82. Knoll, R. L. et al. Gut microbiota differs between children with Inflammatory Bowel Disease and healthy siblings in taxonomic and functional composition: a metagenomic analysis. *Am. J. Physiol-Gastr. L.* **312**, G327–G339 (2017).
83. Hansen, R. et al. Microbiota of de-novo pediatric IBD: increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am. J. Gastr.* **107**, 1913–1922 (2012).
84. Jacobs, J. P. et al. A disease-associated microbial and metabolomics state in relatives of pediatric inflammatory bowel disease patients. *Cell. Mol. Gastroenterol. Hepatol.* **2**, 750–766 (2016).
85. Bajer, L. et al. Distinct gut microbiota profiles in patients with primary sclerosing cholangitis and ulcerative colitis. *World J. Gastr.* **23**, 4548–4558 (2017).
86. Baumgart, M. et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME. J.* **1**, 403–418 (2007).
87. Gevers, D. et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell. Host. Microbe.* **15**, 382–392 (2014).
88. Tong, M. et al. A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *PLoS ONE* **8**, e80702 (2013).
89. Pascal, V. et al. A microbial signature for Crohn's disease. *Gut* **66**, 813–822 (2017).
90. Mottawea, W. et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat. Commun.* **7**, 13419 (2016).
91. Morgan, X. C. et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome. Biol.* **13**, R79 (2012).
92. Verma, R., Verma, A. K., Ahuja, V. & Paul, J. Real-time analysis of mucosal flora in patients with inflammatory bowel disease in India. *J. Clin. Microbiol.* **48**, 4279–4282 (2010).
93. Arun Gupta, S. K., Wagner, Josef, Kirkwood, Carl, Morrison, Mark & McSweeney, Chris and finlay macrae. analysis of mucosal microbiota in inflammatory bowel disease using a custom phylogenetic microarray. *Austin J. Gastr.* **1**, 1–6 (2014).
94. Bibiloni, R., Mangold, M., Madsen, K. L., Fedorak, R. N. & Tannock, G. W. The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. *J. Med. Microbiol.* **55**, 1141–1149 (2006).
95. Sokol, H. et al. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm. Bowel Dis.* **15**, 1183–1189 (2009).
96. Andoh, A. et al. Comparison of the fecal microbiota profiles between ulcerative colitis and Crohn's disease using terminal restriction fragment length polymorphism analysis. *J. Gastroenterol.* **46**, 479–486 (2011).
97. Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M. & Doerffel, Y. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. *Inflamm. Bowel Dis.* **14**, 147–161 (2008).
98. Nishino, K. et al. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *J. Gastroenterol.* **53**, 95–106 (2017).
99. Chen, L. et al. Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease. *Medicine (Baltimore)* **93**, e51 (2014).
100. Suchodolski, J. S., Dowd, S. E., Wilke, V., Steiner, J. M. & Jergens, A. E. 16S rRNA gene pyrosequencing reveals bacterial dysbiosis in the duodenum of dogs with idiopathic inflammatory bowel disease. *PLoS ONE* **7**, e39333 (2012).
101. Xenoulis, P. G. et al. Molecular-phylogenetic characterization of microbial communities imbalances in the small intestine of dogs with inflammatory bowel disease. *FEMS Microbiol. Ecol.* **66**, 579–589 (2008).
102. Omori, M. et al. Fecal microbiome in dogs with inflammatory bowel disease and intestinal lymphoma. *J. Vet. Med. Sci.* **79**, 1840–1847 (2017).
103. Allenspach, K. et al. Evaluation of mucosal bacteria and histopathology, clinical disease activity and expression of Toll-like receptors in German shepherd dogs with chronic enteropathies. *Vet. Microbiol.* **146**, 326–335 (2010).
104. Suchodolski, J. S., Xenoulis, P. G., Paddock, C. G., Steiner, J. M. & Jergens, A. E. Molecular analysis of the bacterial microbiota in duodenal biopsies from dogs with idiopathic inflammatory bowel disease. *Vet. Microbiol.* **142**, 394–400 (2010).
105. Suchodolski, J. S. et al. The fecal microbiome in dogs with acute diarrhea and idiopathic inflammatory bowel disease. *PLoS ONE* **7**, e51907 (2012).
106. Rossi, G. et al. Comparison of microbiological, histological, and immunomodulatory parameters in response to treatment with either combination therapy with prednisone and metronidazole or probiotic VSL#3 strains in dogs with idiopathic inflammatory bowel disease. *PLoS ONE* **9**, e94699 (2014).
107. Minamoto, Y. et al. Alteration of the fecal microbiota and serum metabolite profiles in dogs with idiopathic inflammatory bowel disease. *Gut Microbes* **6**, 33–47 (2015).
108. Jergens, A. E. Feline idiopathic inflammatory bowel disease: what we know and what remains to be unraveled. *J. Feline. Med. Surg.* **14**, 445–458 (2012).
109. Vazquez-Baeza, Y., Hyde, E. R., Suchodolski, J. S. & Knight, R. Dog and human inflammatory bowel disease rely on overlapping yet distinct dysbiosis networks. *Nat. Microbiol.* **1**, 16177 (2016).
110. Inness, V. L., McCartney, A. L., Khoo, K. L. & Gibson, G. R. Molecular characterisation of the gut microflora of healthy and inflammatory bowel disease cats using fluorescence in situ hybridisation with special reference to *Desulfovibrio* spp. *J. Anim. Physiol. Anim. Nutr.* **91**, 48–53 (2007).
111. Elinav, E. et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).
112. Li, M. et al. Upregulation of intestinal barrier function in mice with DSS-induced colitis by a defined bacterial consortium is associated with expansion of IL-17A producing gamma delta T cells. *Front. Immunol.* **8**, 824 (2017).
113. Rooks, M. G. et al. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *ISME. J.* **8**, 1403–1417 (2014).

114. Simpson, K. W. et al. Adherent and invasive *Escherichia coli* is associated with granulomatous colitis in boxer dogs. *Infect. Immun.* **74**, 4778–4792 (2006).
115. Janeczko, S. et al. The relationship of mucosal bacteria to duodenal histopathology, cytokine mRNA, and clinical disease activity in cats with inflammatory bowel disease. *Vet. Microbiol.* **128**, 178–193 (2008).
116. Larmonier, C. B. et al. Reduced colonic microbial diversity is associated with colitis in NHE3-deficient mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **305**, G667–G677 (2013).
117. Spalinger, M. R. et al. PTPN2 controls differentiation of CD4(+) T cells and limits intestinal inflammation and intestinal dysbiosis. *Mucosal Immunol* **8**, 918–929 (2015).
118. Munyaka, P. M., Rabbi, M. F., Khafipour, E. & Ghia, J. E. Acute dextran sulfate sodium (DSS)-induced colitis promotes gut microbial dysbiosis in mice. *J. Basic Microbiol.* **56**, 986–998 (2016).
119. Okayasu, I. et al. A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* **98**, 694–702 (1990).
120. Robinson, A. M. et al. Alterations of colonic function in the Winnie mouse model of spontaneous chronic colitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **312**, G85–G102 (2017).

Acknowledgements

We thank all patients and the members of the Swiss IBD cohort and Bern cohort for their commitment. We also thank the staff of the University Hospital of Bern, Clinic of Visceral Medicine and Surgery, and the Bern City Hospitals led by F. Seibold and R. Tutuian for obtaining samples in Cohort 2. This research was supported by Systems X (GutX) to A.J.M. and J.S., and the Swiss IBD cohort (grant no. 33CS30-148422) to G.R., A.J.M. and C.M. The founding institutions had no role in the study design, analysis or interpretation of the results. We thank G. Rahnavard and C. Huttenhower (Department

of Biostatistics, Harvard T.H. Chan School of Public Health, USA) for their help in using the HALLA pipeline. We also thank J. Harrell Rieder, A. Suter, S. Brand, C. Mooser, W. Kwong Chung and J. Hugenschmidt for helping B.Y. during the process of sample preparation. We also thank G. Weingart (Department of Biostatistics, Harvard T.H. Chan School of Public Health, USA) for his enormous help in the optimization of MaAsLin running on the MacOS platform using R.

Author contributions

A.J.M. conceived, designed and supervised the study. B.Y. performed all the experiments, analyzed the data and wrote the manuscript with A.J.M. P.J. organized and collected the samples of the second cohort. P.J. E.D.B., Y.F., N.F. and M.G. were involved in data curation. O.O. and C.R. carried out metabolic reaction analysis, and J.S. supervised these analyses. A.J.M., P.J. P.M., C.M., V.E.H.P., M.H.M., G.R., R.W. and Swiss IBD cohort investigators acquired patient samples and detailed structured clinical phenotypes.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41591-018-0308-z>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-018-0308-z>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.J.M.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Swiss IBD Cohort Investigators

Karim Abdelrahman¹¹, Gentiana Ademi¹², Patrick Aepli¹³, Amman Thomas¹⁴, Claudia Anderegg¹⁵, Anca-Teodora Antonino¹⁶, Eva Archanioti⁶, Eviano Arrigoni¹⁷, Diana Bakker de Jong¹⁸, Bruno Balsiger¹⁹, Polat Bastürk¹⁸, Peter Bauerfeind²⁰, Andrea Becocci²¹, Dominique Belli²¹, José M. Bengoa¹⁷, Luc Biedermann⁸, Janek Binek²², Mirjam Blattmann⁸, Stephan Boehm²³, Tujana Boldanova²⁴, Jan Borovicka¹², Christian P. Braegger²⁵, Stephan Brand¹², Lukas Brügger², Simon Brunner¹⁸, Patrick Bühr²⁵, Bernard Burnand⁴, Sabine Burk⁸, Emanuel Burri²⁶, Sophie Buyse²⁷, Dahlia-Thao Cao²⁸, Ove Carstens², Dominique H. Cribiez¹³, Sophie Cunningham¹⁷, Fabrizia D'Angelo²⁹, Philippe de Saussure¹⁷, Lukas Degen¹⁸, Joakim Delarive³⁰, Christopher Doerig³¹, Barbara Dora⁸, Susan Drerup³², Mara Egger⁴, Ali El-Wafa³³, Matthias Engelmann³⁴, Jessica Ezri⁶, Christian Felley³⁵, Markus Fliegner³⁶, Nicolas Fournier⁴, Montserrat Fraga⁶, Yannick Franc⁴, Pascal Frei³⁷, Remus Frei¹², Michael Fried⁸, Florian Froehlich³⁸, Raoul Ivano Furlano³⁹, Luca Garzoni⁴⁰, Martin Geyer⁴¹, Laurent Girard¹⁷, Marc Girardin⁴², Delphine Golay⁴, Ignaz Good⁴³, Ulrike Graf Bigler², Beat Gysi⁴⁴, Johannes Haarer¹², Marcel Halama⁴⁵, Janine Haldemann¹⁹, Pius Heer⁴⁶, Benjamin Heimgartner², Beat Helbling³⁷, Peter Hengstler²², Denise Herzog⁴⁷, Cyrill Hess¹³, Roxane Hessler³¹, Klaas Heyland⁴⁸, Thomas Hinterleitner⁴⁹, Claudia Hirschi³⁴, Petr Hruz¹⁸, Pascal Juillerat², Carolina Khalid-de Bakker¹⁸, Stephan Kayser⁵⁰, Céline Keller⁵, Christina Knellwolf (-Grieger)¹², Christoph Knoblauch⁵¹, Henrik Köhler¹⁵, Rebekka Koller²⁵, Claudia Krieger(-Grübel)¹², Patrizia Künzler¹², Rachel Kusche¹⁵, Frank Serge Lehmann⁴⁶, Andrew J. Macpherson², Michel H. Maillard^{5,6}, Michael Manz⁴⁵, Astrid Marot⁴, Rémy Meier⁵², Christa Meyenberger¹², Pamela Meyer¹², Pierre Michetti^{5,6}, Benjamin Misselwitz⁸, Patrick Mosler⁵³, Christian Mottet⁵⁴, Christoph Müller⁷, Beat Müllhaupt⁸, Leilla Musso⁴, Michaela Neagu⁵⁵, Cristina Nichita⁵⁶, Jan Niess¹⁸, Andreas Nydegger^{5,6}, Nicole Obialo⁸, Diana Ollo³⁵, Cassandra Oropesa³⁵, Ulrich Peter⁴⁸, Daniel Peternac⁵⁷, Laetitia Marie Petit³⁵, Valérie Pittet⁴, Daniel Pohl⁸, Marc Porzner⁵⁸, Claudia Preissler⁵⁹, Nadia Raschle⁸, Ronald Rentsch⁶⁰, Alexandre Restellini⁴², Sophie Restellini³⁵, Jean-Pierre Richterich¹³, Frederic Ris³⁵, Branislav Risti⁶¹,

Marc Alain Ritz⁶², Gerhard Rogler⁸, Nina Röhrich¹², Jean-Benoît Rossel⁴, Vanessa Rueger²⁵, Monica Rusticeanu², Markus Sagmeister⁶³, Gaby Saner¹⁹, Bernhard Sauter⁶⁴, Mikael Sawatzki¹², Michael Scharl⁸, Martin Schelling¹², Susanne Schibli⁶⁵, Hugo Schlauri⁶⁶, Dominique Schluckebier³⁵, Daniela Schmid⁵¹, Sybille Schmid (-Uebelhart)², Jean-François Schnegg⁶⁷, Alain Schoepfer^{5,6}, Vivianne Seematter⁴, Frank Seibold¹⁹, Mariam Seirafi⁶⁸, Gian-Marco Semadeni¹², Arne Senning²⁵, Christiane Sokollik⁶⁵, Joachim Sommer⁴, Johannes Spalinger^{13,65}, Holger Spangenberg⁶⁹, Philippe Stadler⁷⁰, Peter Staub⁷¹, Dominic Staudenmann¹³, Volker Stenz⁷², Michael Steuerwald⁶², Alex Straumann⁴⁶, Bruno Strebel¹², Andreas Stulz¹³, Michael Sulz¹², Aurora Tatu², Michela Tempia-Caliera⁷³, Joël Thorens⁷⁴, Kaspar Truninger⁷⁵, Radu Tutuian², Patrick Urfer⁷⁶, Stephan Vavricka⁸, Francesco Viani⁷⁷, Jürg Vögtlin⁶², Roland Von Känel⁸, Dominique Vouillamoz⁷⁸, Rachel Vulliamy⁴, Paul Wiesel⁵⁶, Reiner Wiest², Stefanie Wöhrle¹³, Samuel Zamora⁴², Silvan Zander⁷⁹, Tina Wylie⁸⁰, Jonas Zeitz⁸ and Dorothee Zimmermann¹²

¹¹Clinique de Montchoisi, Lausanne, Switzerland. ¹²Kantonsspital St-Gallen, St-Gallen, Switzerland. ¹³Kantonsspital Luzern, Luzern, Switzerland. ¹⁴GI private practice, Waldkirch, St. Gallen, Switzerland. ¹⁵Kantonsspital Aarau, Klinik für Kinder und Jugendliche, Aarau, Switzerland. ¹⁶Hôpital Riviera—Site du Samaritain, Vevey, Vaud, Switzerland. ¹⁷GI private practice, Geneva, Switzerland. ¹⁸Department of Gastroenterology and Hepatology, Basel University Hospital, Basel, Switzerland. ¹⁹Gastroenterologische Praxis, Bern, Switzerland. ²⁰Department Gastroenterology and Hepatology, Stadtspital Triemli, Zurich, Switzerland. ²¹Department of Pediatric, Geneva University Hospital, Geneva, Switzerland. ²²Gastroenterologie am Rosenberg, St-Gallen, Switzerland. ²³Spital Bülach, Bülach, Zurich, Switzerland. ²⁴Department of Biomedicine, University of Basel, Basel, Switzerland. ²⁵University Children's Hospital, Zurich, Switzerland. ²⁶Department Gastroenterology, Kantonsspital Liestal, Liestal, Switzerland. ²⁷GI private practice, Yverdon-les-Bains, Switzerland. ²⁸Hôpital Neuchâtelois, La Chaux-de-fonds, Neuchâtel, Switzerland. ²⁹Department Gastroenterology and Hepatology, Geneva University Hospital, Geneva, Switzerland. ³⁰GI private practice, Lausanne, Switzerland. ³¹Clinique Cecil, Lausanne, Switzerland. ³²Schulthess Clinic, Zurich, Switzerland. ³³GI private practice, La Chaux-de-Fonds, Switzerland. ³⁴Gastropraxis Luzern, Luzern, Switzerland. ³⁵Centre de Gastroentérologie Beaulieu SA, Geneva, Switzerland. ³⁶Medical Center Sihlcity, Zurich, Switzerland. ³⁷Gastroenterologie Bethanien, Zurich, Switzerland. ³⁸Hospital of the Canton of Jura, Porrentruy And Delémont, Jura, Switzerland. ³⁹Universitäts-Kinderspital beider Basel (UKBB), Basel, Switzerland. ⁴⁰Clinique des Grangettes, Geneva University Hospital, Genève, Switzerland. ⁴¹GI private practice, Wetztingen, Aargau, Switzerland. ⁴²Groupe Médical d'Onex, Onex, Switzerland. ⁴³Spital Walenstadt, Walenstadt, St. Gallen, Switzerland. ⁴⁴GI private practice, Reinach, Switzerland. ⁴⁵Aerztehaus Fluntern, Zurich, Switzerland. ⁴⁶GI private practice, Olten, Switzerland. ⁴⁷HFR Hôpital fribourgeois—Pédiatrie, Fribourg, Switzerland. ⁴⁸KSW Kantonsspital Winterthur Kinderklinik, Winterthur, Switzerland. ⁴⁹GI private practice, Zurich, Switzerland. ⁵⁰GI private practice, Luzern, Switzerland. ⁵¹Kantonsspital Nidwalden, Stans, Nidwalden, Switzerland. ⁵²AMB - Arztpraxis MagenDarm Basel, Basel, Switzerland. ⁵³Kantonsspital Graubünden, Chur, Switzerland. ⁵⁴GI private practice, Sion, Switzerland. ⁵⁵Spital Tiefenau, Bern, Switzerland. ⁵⁶Centre médical d'Epalinges, Epalinges, Switzerland. ⁵⁷Spital Waid, Zurich, Switzerland. ⁵⁸Spital Lachen, Lachen, Switzerland. ⁵⁹Kantonsspital Olten, Olten, Switzerland. ⁶⁰GI private practice, St-Gallen, Switzerland. ⁶¹GI practice, Dietikon, Switzerland. ⁶²GI practice, Liestal, Switzerland. ⁶³GI private practice, Heerbrugg, Switzerland. ⁶⁴Klinik Hirslanden Zürich, Zurich, Switzerland. ⁶⁵Kinderklinik Bern, Bern University Hospital, Bern, Switzerland. ⁶⁶Derby Center, Wil, Switzerland. ⁶⁷GI private practice, Montreux, Switzerland. ⁶⁸Clinique La Colline, Geneva, Switzerland. ⁶⁹Kantonsspital Wolhusen, Wolhusen, Switzerland. ⁷⁰GI private practice, Payerne, Switzerland. ⁷¹Spital Heiden Appenzell Ausserrhoden, Heiden, Switzerland. ⁷²Kantonsspital Münstlingen, Münstlingen, Switzerland. ⁷³Clinique des Grangettes, Chêne-Bougeries, Switzerland. ⁷⁴GI private practice, Yverdon, Switzerland. ⁷⁵GI private practice, Langenthal, Switzerland. ⁷⁶Hirslanden Klinik Aarau, Gastro Zentrum, Aarau, Switzerland. ⁷⁷Private practice, Vevey, Switzerland. ⁷⁸Private practice, Pully, Switzerland. ⁷⁹Spital Limmattal, Schlieren, Switzerland. ⁸⁰Infirmière de Recherche chez CHUV Lausanne University Hospital, Lausanne, Switzerland.

Methods

Cohorts and study design. Biopsy samples in this study were collected either (1) in the Swiss IBD cohort study²¹ that has included more than 3,500 patients (Cohort 1; approximately 5,000 intestinal biopsies have been submitted to the Biobank Cohort) or (2) newly recruited IBD patients and non-IBD subjects in the framework of the human intestinal community project of Bern University Hospital (Cohort 2). Cohort 1 includes biopsy samples from teaching hospitals, community hospitals and office practices in Switzerland. For all patients included in the cohort, deep longitudinal clinical and investigational phenotyping for an average of 5.7 years (range 1–11 years) was sequentially taken over at least annual follow-ups. Unlike the Swiss IBD cohort, which had focused exclusively on inflammatory bowel disease, Cohort 2 included asymptomatic control patients undergoing screening colonoscopies without macroscopic or microscopic abnormalities, in whom all measured hematologic and clinical chemistry parameters were normal. Non-IBD biopsy samples were collected from subjects that were registered for a screening ileo-colonoscopy without any gastrointestinal symptoms, no suspected functional intestinal symptoms and where all other biochemical and hematological work-up was negative.

Inclusion criteria for Swiss IBD cohort patients were on established Lennard-Jones criteria²², confirmed by radiological, endoscopic and histological findings. Subject to informed consent for the longitudinal study, non-residents of Switzerland and patients unable or unwilling to provide blood samples were excluded. At inclusion, the patient's physician filled an enrollment questionnaire to make a clinical evaluation of the disease activity (CDAI, HBI for CD and MTWAI, SCAI for UC), the disease extent and location based on the Montréal classification, including extraintestinal manifestations, previous surgery (bowel resection or perianal), complications (fistula, stenosis, anemia, osteoporosis), family history, duration of disease, smoking status, previous treatment success/failure and current therapy and dosage, as well as laboratory data (C-reactive protein, hemoglobin, auto-antibodies and also stool calprotectin from 1,244 patients). Clinical treatment for IBD was scored (5-ASA preparations, steroids, thiopurines, methotrexate, tumor necrosis factor α (TNF- α) inhibitors or other biologicals) as well as the use of other concurrent medications and non-IBD surgeries. The clinical follow-up was on an annual basis and prospectively collected by the clinician or a study nurse with an additional questionnaire about their psychosocial situation (quality of life, depression and anxiety) and clinical IBD trajectory (disease activity, hospitalization, change in medication, adherence, surgeries). Hard copy data from the enrollment and clinical follow-up were then digitalized using Access 2010 (Microsoft).

The inclusion criteria and data collection for consenting patients for the local Bern cohort were similar, except that patients were diagnosed and followed up in the Bern University Hospital (Inselspital) IBD clinic and in the Bern City Hospitals.

Clinical data extraction, categorization and analysis. Structured clinical metadata was prospectively collected according to pre-determined standards containing variables of Cohort 1 and Cohort 2 (documented electronically using the Research Electronic Data Capture database²³). These metadata were imported and processed as data frame in R (<http://www.r-project.org>) using the xlsx package. Assessment of the microbiota composition from intestinal biopsies was then analyzed according to parameters including the type, course and severity of disease using the detailed clinical criteria summarized in Supplementary Table 1 for Cohort 1 and Supplementary Table 2 for Cohort 2. Statistical analyses were performed using Student's *t*-test, Wilcoxon's rank sum test and Pearson's chi-squared test to test normality between groups.

Ethics statement. Licensed gastroenterologists collected biopsy samples and clinical data of patients identified as CD or UC and non-IBD subjects. The SIBDC study (Cohort 1) protocol was approved by the Ethics Committee of the Kanton Zurich (reference KEK-ZH-number 2013-0284) and this cohort is also registered with Trials Registration Number NCT02849821. The Bern Human Intestinal Community project (Cohort 2) was approved by the Bern Cantonal Ethics Commission (Ref: KEK-BE: 251/14 and 336/14) with signed informed consent was obtained from all participants. Bern cohort data were anonymized and collected in the electronic data capture system (Research Electronic Data Capture)²³, activated for the trial after successfully passing formal quality controls. The Bern electronic data capture system and the database are hosted by the Clinical Trial Unit of Bern University. The study complies with all relevant ethical regulations.

DNA extraction from human biopsies. Intestinal endoscopic biopsies were initially collected into 2 ml microfuge tubes containing 500–600 μ l RNAlater (Sigma-Aldrich) and stored at -20°C prior to DNA extraction protocol⁸⁷. Total DNA was isolated using AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions. Briefly, 600 μ l of Buffer RLT Plus 2-mercaptoethanol and a metal bead were added into each tube. Samples were then homogenized with the Retsch Tissue Lyser (QIAGEN) at 30 Hz for 3 min, followed by 3 min centrifugation at 13,000g (Eppendorf). Supernatants were transferred into the AllPrep DNA mini spin column and centrifuged at 9,000g for 30 s. DNA attached to spin columns was washed/de-salted using 500 μ l of Buffer AW1 and Buffer AW2. Last, DNA samples were eluted with 25–30 μ l EB buffer into 1.5 ml microfuge

tubes. The concentration and purity of the isolated DNA samples were analyzed in NanoDrop (Thermo Scientific).

Microbial profiling of biopsies. 1255 samples from Cohort 1 and 1846 biopsy samples from Cohort 2 were used in this study. The V5/V6 region of 16S rRNA genes was amplified with Invitrogen Platinum *Taq* DNA polymerase from 1,000–2,000 ng of DNA using a range of oligonucleotide primers specific for the domains V5 and V6 of rDNA bacteria. Specifically, all forward core primers had been modified by the addition of a Personal Genome Machine (PGM) sequencing adaptor, a 'GT' spacer and unique barcode that allowed us to have up to different 96 barcodes. The expected product length was ~350 base pairs including adaptors and barcodes. Bacteria-specific primers (forward 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGC-barcode-ATTAGATACCCYGGTAGTCC-3' and reverse 5'-CCTCTCTATGGGAGTCGGTGATACGAGCTGACGACARCCATG-3') were used^{124,125}. PCR conditions consisted of an initial 5 min at 94°C denaturation step, followed by 35 cycles of 1 min denaturation at 94°C , 20 s annealing cycle at 46°C and 30 s extension cycle at 72°C , with a final extension for 7 min at 72°C . Samples were kept at 4°C until loading into an 1% agarose gel. Amplicons were purified using Gel Extraction Kit (Qiagen) after gel electrophoresis for 1 h. The concentrations of amplicons were then evaluated by Qubit 3.0 Fluorometer (ThermoFisher) prior to library preparation and used at 26 pM. To prepare template-positive Ion PGM Template OT2 400 Ion Sphere Particles containing clonally amplified DNA, we used the Ion OneTouch Instrument with the Ion PGM Template OT2 400 Kit (ThermoFisher). Sequencing was performed using the Ion PGM Sequencing 400 Kit and Ion 316 Chip V2 within the Ion PGM System (ThermoFisher)¹²⁶.

Raw sequences were first loaded into the QIIME 1.9.1 pipeline, as described¹²⁷ using custom analysis scripts for analysis on the UBELIX Linux cluster of the University of Bern. Only samples with more than 4,500 high-quality reads were then used for further analysis using R packages. Operational taxonomic units were picked using UCLUST with a 97% sequence identity threshold, using the default options as implemented in QIIME and followed by taxonomy assignment using the latest Greengenes database (greengenes.secondgenome.com). OTUs that were not present in at least 0.01% of our samples or with a low abundance ($<0.001\%$ of the total counts) were filtered out. In combination with the detailed clinical patient metadata (diagnosis, age, BMI, inflammation status, location of biopsy, disease severity, treatment and so on), PCA was performed to determine whether samples with certain properties cluster. Due to the limitations of the resolution on taxonomical classification using 16S gene sequencing technique, analyses were restricted to genus level. The OTU abundance *biom* file and mapping file were used for statistical analyses and data visualization in the R statistical programming environment package *phyloseq*¹²⁸. Calculation of the α -diversity (Observed OTUs, Simpson and Shannon indices), β -diversity (Bray–Curtis genus-level community dissimilarities, weighted and unweighted UniFrac-based PCoA), and statistical analysis of clustering using Mann–Whitney *U*-tests for alpha diversity and Adonis (PERMANOVA) for beta diversity to confirm that the strength and statistical significance of groups in the same distance metrics in the QIIME pipeline and *phyloseq* in R^{128,129}. Multivariate homogeneity of group dispersion was to calculate the average distance of the groups and to test if the dispersion of any group was significantly different from the others. Using *k*-nearest neighborhood networks created by thresholding the Jaccard dissimilarity matrix, the differences in microbial community diversities between biopsies collected from different locations (ileum, colon or rectum) for both cohorts were analyzed. Further analyses were performed using a Random Forest supervised learning algorithm with additional Boruta feature selection to identify the most important predictive taxa or community within microbial communities in R. HALLA³⁶ and multivariate analysis by linear models (MaAsLin) R package³¹ were used to find associations between clinical metadata (age, BMI, gender, smoking habits, medication, inflammation status, anatomic location and so on) and microbial community abundance. The Q value package was implemented in MaAsLin to correct for multiple testing (BH-FDR correction, *q* value) of 0.05. Taxa present in at least 30% of the samples and taxa that had more than 0.0001% of total abundance were set as the cut-off values for further analysis. After correction for a false discovery rate, $q < 0.05$ was considered significant. Notable near-significant differences ($P < 0.05$ and $0.05 < q < 0.2$) were also highlighted in the figures. Non-significant data were indicated in figures by 'NS'.

All the relevant codes for running the MaAsLin and HALLA pipeline are available in C. Huttenhower's group webpage (<https://huttenhower.sph.harvard.edu>). Plots were generated with ggplot2 using *phyloseq* object or Graphpad Prism v.7.0a.

Biopsies (1,255 samples in Cohort 1 and 1,846 in Cohort 2) from different pre-defined positions along the length of the colon and the distal small intestine (terminal ileum) from each patient generated 35,678,215 and 40,879,931 high-quality 16S rDNA reads with an average amplicon length of 293 bp (Supplementary Tables 1 and 2). Sequences were clustered into 359,010 and 392,170 taxa, respectively, based on their shared sequence similarity at a 97% threshold (3% sequence divergence)¹³⁰. The average estimated coverage for OTU characterization was 94.2 and 93.2%^{130,131}.

Microbial association network construction within disease status. Co-occurrence analyses were performed based on the organization of the data generated by *phyloseq* object in R. Network analyses were performed using the scripts reported by Williams et al.³⁵. Briefly, we considered the co-occurrence to be positive rank correlations identified using Spearman's correlation between identified taxa within each disease condition. Positive correlation (weak correlation: $\rho > 0.25$ and strong correlation: $\rho > 0.5$) helped us to identified even relatively less strong and highly correlated significant positive correlations between co-occurring taxa in each identified ecosystem, respectively. The OTUs that had more than 0.0001% total of relative abundance and taxa present in at least 30% of the samples were retained for further statistical analysis and illustration of the significant co-occurrence network between microbial taxa. We thus obtained a data frame that listed disease status, the originating environment of the samples, assigned taxa, Spearman's correlation coefficient, ρ value calculated using PERMANOVA and the abundances of each taxa. We calculated the modules detected using Vertex and edge betweenness centrality functions in *igraph* package and generated betweenness score for each microbial taxon. The 'betweenness centrality scores' for each vertex is the number of these shortest paths going through a node and 'eigenvector centrality' is a measure of node connectivity to important neighbors in a given network. Spearman's correlation coefficients calculated between the different co-occurring taxa within each environment were used for generating network plots. *Igraph* objects were transferred into Cytoscape using *RJSONIO* and *httr* packages in R. The community structure analysis of biological co-occurrence network carried out using GLay implemented into Cytoscape as a fast-greedy algorithm¹³². Network plots were generated using a customized organic layout of Cytoscape v.3.5.1.

Reaction-level analysis. OTUs were mapped to metabolic reactions via a previously published collection of genome-scale metabolic models (GSMMs)²⁸. We included OTUs that could be mapped to a taxonomic rank of family or lower and to at least one GSMM. The normalized abundance of a reaction i in a patient sample, $a_r(i)$ was calculated as:

$$a_r(i) = \sum_{j=1}^N a_{\text{otu}}(j) E(i, j) / \sum_{j=1}^N a_{\text{otu}}(j),$$

where $a_{\text{otu}}(j)$ is the abundance of OTU j in the patient sample, N is the number of detected OTUs and $E(i, j)$ is the expected probability (frequency of occurrences) of reaction i in the GSMMs associated with OTU j . To analyze metabolic reaction differences between patient groups, a two-sample t -test was used to compare the mean abundances of each reaction and each binary patient classification of interest (for example, UC versus CD) between the two groups. The Benjamini–Hochberg procedure¹³³ was applied to correct for multiple testing and adjusted $P \leq 0.05$ between groups was considered significant. The metabolic subsystem classification of reactions was obtained from the GSMMs and for each t -test comparing groups of patients and each subsystem, Fisher's exact test was performed to determine if the subsystem was overrepresented among the significantly different reactions; subsystems with $P \leq 0.05$ were considered enriched. All analyses were performed in MATLAB (R2018a).

Fecal calprotectin. Calprotectin is a member of the S100 family of calcium binding proteins found in the cytosol of neutrophils and thus, abundant in all body fluids in proportion to the degree of inflammation present. Fecal calprotectin was measured by a quantitative enzyme-linked immunosorbent assay after using Calprotectin–Stool Extraction with CALEX. The assays were carried out according to the instructions (The BÜHLMANN fecal Calprotectin Assays, BÜHLMANN Laboratories). All fecal samples were processed within 72 h after collection. The assay plates were read at an absorbance $A_{450\text{ nm}}$ as instructed in manufacturer's documentation. The calprotectin cut-off level representing a positive value was $\geq 50 \mu\text{g}$ calprotectin per gram of stool.

Search strategy and data extraction. We systematically searched the publications in PubMed (including MEDLINE) and Google Scholar. The electronic search algorithm consisted of terms relating to key concepts of 'inflammatory bowel disease', 'IBD', 'microbiome', 'gut microbiota', 'metagenomic', '16S rDNA', 'human', 'mice', 'cats', 'dogs', 'animal models', 'naturally induced colitis' and 'chemically induced colitis'^{35–120,134–187}. Reference lists of included articles, related reviews, and other relevant sources were manually searched. After initial screening of titles and abstracts, full text assessments were performed for eligibility of the data to include into the figure. Characteristics of each eligible study were extracted: disease information, species, race (ethnicity), gender, age (median year), number of samples used for sequencing, type of samples, method of characterization of the microbiota, model of inflammation induction. Results of included studies were then organized and corresponding heatmaps generated using Euclidean clustering of sample annotations (vertical) including race/ethnicity, gender, median age, patient number, sample type, model of inflammation induction, sequencing method and microbial taxa (horizontal) at different taxonomic ranks.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

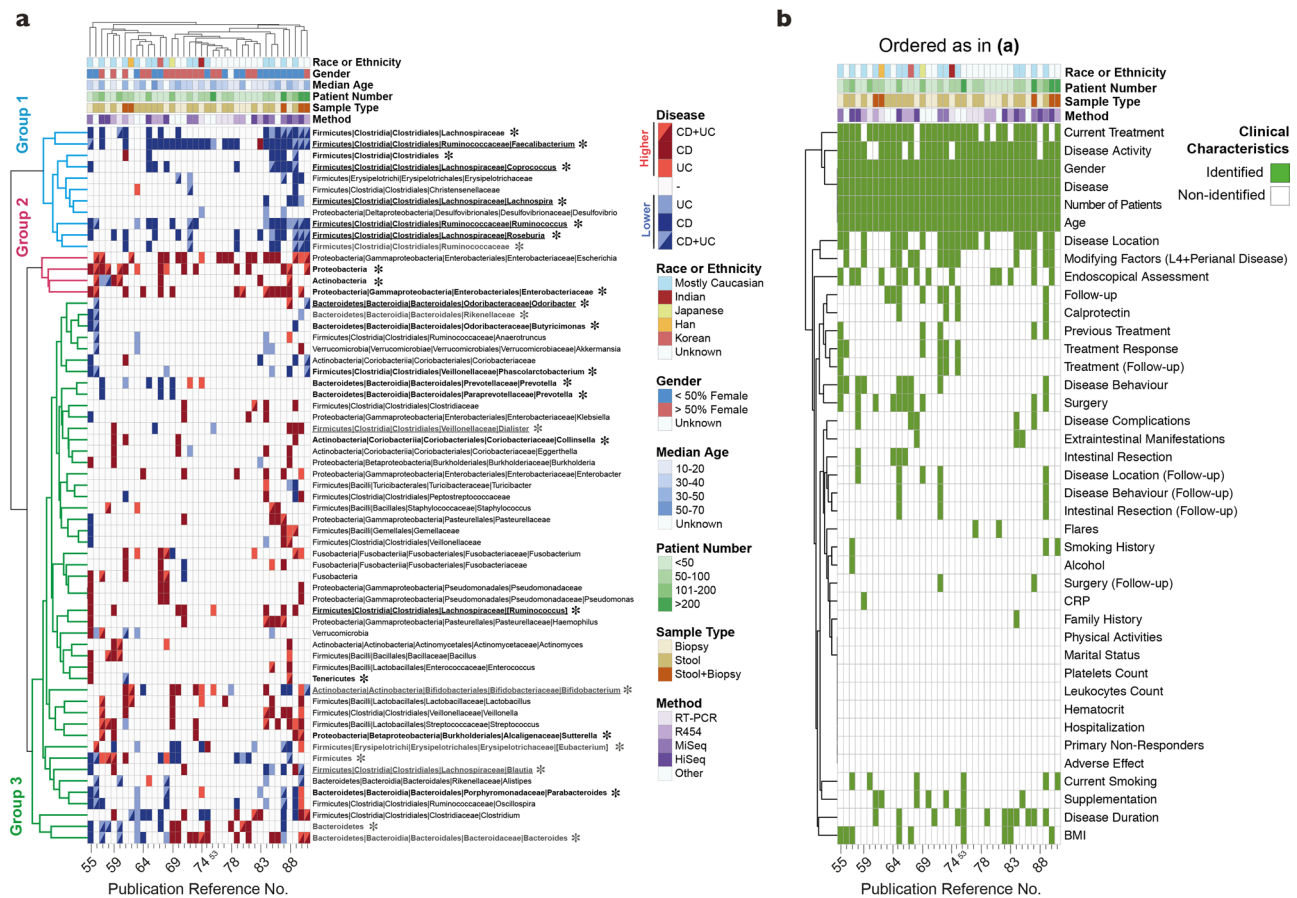
Data availability

All sequencing datasets from the current study have been deposited in a figshare repository and are publicly available. The Cohort 1 fasta file with the mapping file are available at <https://figshare.com/s/e9f2cfd0f0328ca5811> (<https://doi.org/10.6084/m9.figshare.7335068>) and the Cohort 2 fasta file with the mapping file are available at <https://figshare.com/s/bbdd5dfb01e29484efa1> (<https://doi.org/10.6084/m9.figshare.7335071>). Associated codes for the analysis using R packages and QIIME can be found in these depositories. The genome-scale metabolic model script and dataset are available at <https://figshare.com/s/a34f96698ca6fcd36ac2>.

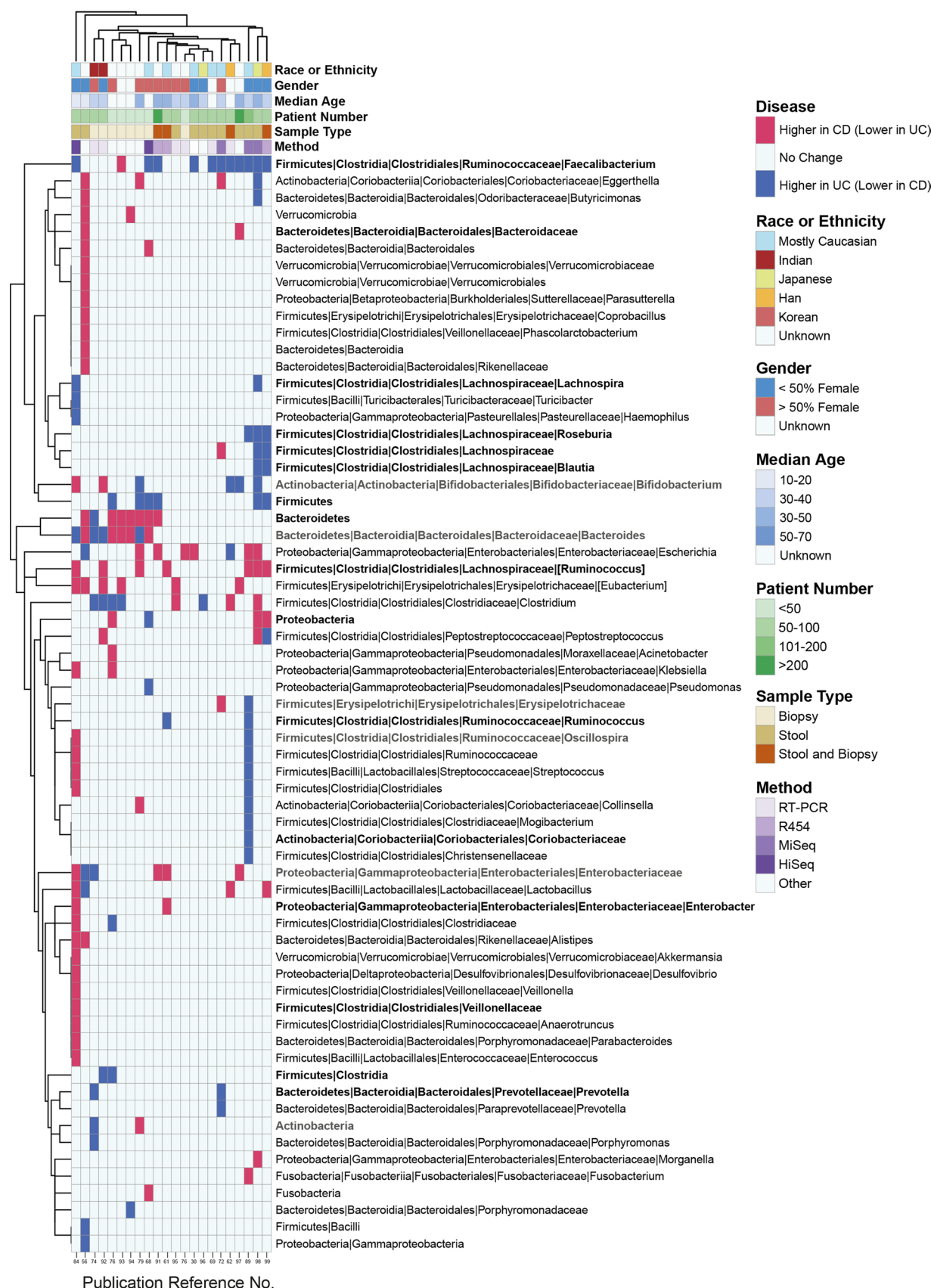
References

- Pittet, V. et al. Cohort profile: the Swiss Inflammatory Bowel Disease Cohort Study (SIBDCS). *Int. J. Epidemiol.* **38**, 922–931 (2009).
- Lennard-Jones, J. E. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol. Suppl.* **170**, 2–6, discussion 16–19 (1989).
- Harris, P. A. et al. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
- Sundquist, A. et al. Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol.* **7**, 108 (2007).
- Yilmaz, B. et al. The presence of genetic risk variants within PTPN2 and PTPN22 is associated with intestinal microbiota alterations in Swiss IBD cohort patients. *PLoS One* **13**, e0199664 (2018).
- Whiteley, A. S. et al. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J. Microbiol. Methods* **91**, 80–88 (2012).
- Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
- Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Res* **5**, 1492 (2016).
- Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
- Good, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–264 (1953).
- Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135–3137 (2010).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
- Yang, I. et al. Intestinal microbiota composition of interleukin-10 deficient C57BL/6J mice and susceptibility to *Helicobacter hepaticus*-induced colitis. *PLoS ONE* **8**, e70783 (2013).
- Ren, Y. et al. Polysaccharide of *Herichium erinaceus* attenuates colitis in C57BL/6 mice via regulation of oxidative stress, inflammation-related signaling pathways and modulating the composition of the gut microbiota. *J. Nutr. Biochem.* **57**, 67–76 (2018).
- Nunberg, M. et al. Interleukin 1 α -deficient mice have an altered gut microbiota leading to protection from dextran sodium sulfate-induced colitis. *mSystems* **3**, e00213–17 (2018).
- Schaubek, M. et al. Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* **65**, 225–237 (2016).
- Lupp, C. et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell. Host. Microbe* **2**, 204 (2007).
- Nones, K. et al. Multidrug resistance gene deficient (*mdr1a*–/–) mice have an altered caecal microbiota that precedes the onset of intestinal inflammation. *J. Appl. Microbiol.* **107**, 557–566 (2009).
- Manchester, A. C. et al. Association between granulomatous colitis in french bulldogs and invasive *Escherichia coli* and response to fluoroquinolone antimicrobials. *J. Vet. Intern. Med.* **27**, 56–61 (2013).
- Roy, U. et al. Distinct microbial communities trigger colitis development upon intestinal barrier damage via innate or adaptive immune cells. *Cell Rep* **21**, 994–1008 (2017).
- Johnston, D. G. W. et al. Loss of microRNA-21 influences the gut microbiota causing reduced susceptibility in a murine model of colitis. *J. Crohns Colitis* **12**, 835–848 (2018).
- Berry, D. et al. Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *ISME J.* **6**, 2091–2106 (2012).

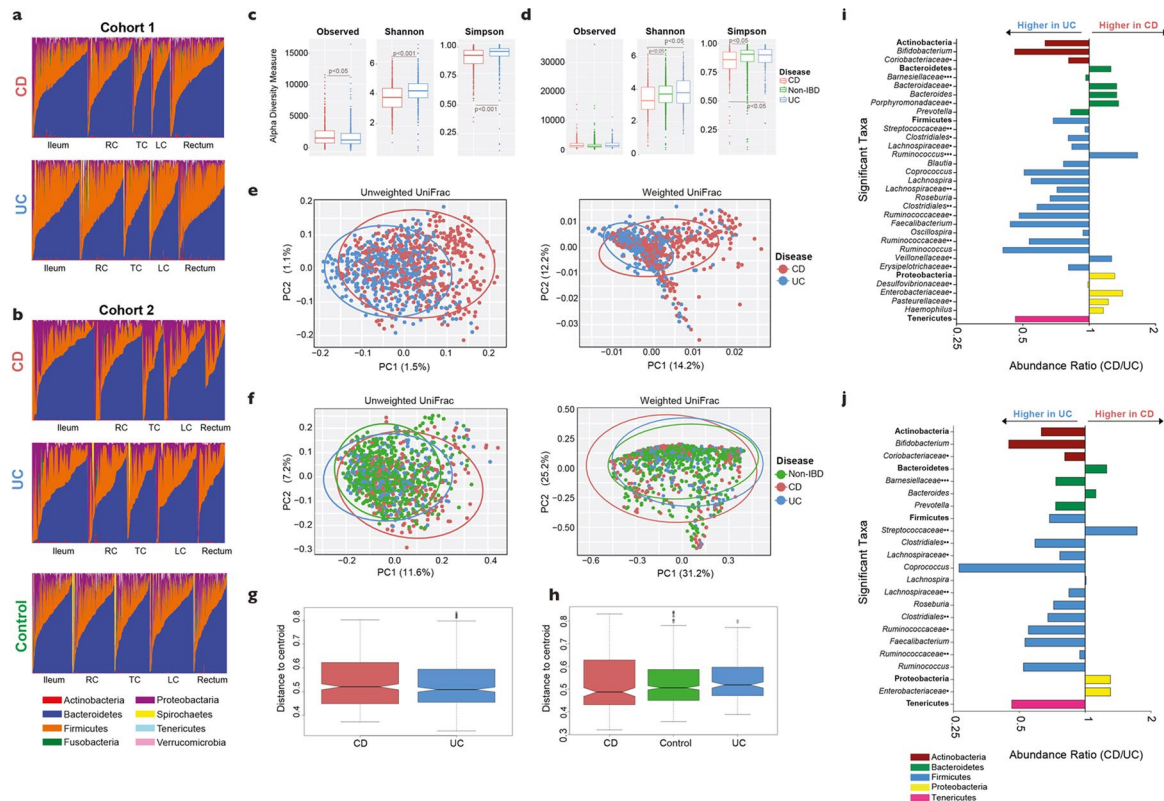
144. Zhang, Q. et al. Accelerated dysbiosis of gut microbiota during aggravation of DSS-induced colitis by a butyrate-producing bacterium. *Sci. Rep.* **6**, 27572 (2016).
145. Schwab, C. et al. Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *ISME J.* **8**, 1101–1114 (2014).
146. Zimmermann, J. et al. The intestinal microbiota determines the colitis-inducing potential of T-bet-deficient Th cells in mice. *Eur. J. Immunol.* **48**, 161–167 (2018).
147. Garrett, W. S. et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host. Microbe* **8**, 292–300 (2010).
148. Palm, N. W. et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).
149. Maharshak, N. et al. Altered enteric microbiota ecology in interleukin 10-deficient mice during development and progression of intestinal inflammation. *Gut Microbes* **4**, 316–324 (2013).
150. Chassaing, B. et al. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. *Nature* **519**, 92–96 (2015).
151. Osaka, T. et al. Meta-analysis of fecal microbiota and metabolites in experimental colitic mice during the inflammatory and healing phases. *Nutrients* **9**, 1329 (2017).
152. Perez-Munoz, M. E. et al. Discordance between changes in the gut microbiota and pathogenicity in a mouse model of spontaneous colitis. *Gut Microbes* **5**, 286–295 (2014).
153. Nagalingam, N. A., Kao, J. Y. & Young, V. B. Microbial ecology of the murine gut associated with the development of dextran sodium sulfate-induced colitis. *Inflamm. Bowel Dis.* **17**, 917–926 (2011).
154. Alkadhi, S., Kunde, D., Cheluvappa, R., Randall-Demllo, S. & Eri, R. The murine appendiceal microbiome is altered in spontaneous colitis and its pathological progression. *Gut Pathog.* **6**, 25 (2014).
155. Carvalho, F. A. et al. Interleukin-1beta (IL-1beta) promotes susceptibility of Toll-like receptor 5 (TLR5) deficient mice to colitis. *Gut* **61**, 373–384 (2012).
156. Seregin, S. S. et al. NLRP6 protects Il10(–/–) mice from colitis by limiting colonization of *Akkermansia muciniphila*. *Cell Rep.* **19**, 2174 (2017).
157. Selvanantham, T. et al. NKT cell-deficient mice harbor an altered microbiota that fuels intestinal inflammation during chemically induced colitis. *J. Immunol.* **197**, 4464–4472 (2016).
158. Lamas, B. et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat. Med.* **22**, 598 (2016).
159. Madsen, K. L. et al. Antibiotic therapy attenuates colitis in interleukin 10 gene-deficient mice. *Gastroenterology* **118**, 1094–1105 (2000).
160. Mitchell, J. et al. Colonic inhibition of phosphatase and tensin homolog increases colitogenic bacteria, causing development of colitis in Il10^{–/–} mice. *Inflamm. Bowel Dis.* **24**, 1718–1732 (2018).
161. Hakansson, A. et al. Immunological alteration and changes of gut microbiota after dextran sulfate sodium (DSS) administration in mice. *Clin. Exp. Med.* **15**, 107–120 (2015).
162. Heimesaat, M. M. et al. Shift towards pro-inflammatory intestinal bacteria aggravates acute murine colitis via Toll-like receptors 2 and 4. *PLoS ONE* **2**, e662 (2007).
163. Wohlgemuth, S., Haller, D., Blaut, M. & Loh, G. Reduced microbial diversity and high numbers of one single *Escherichia coli* strain in the intestine of colitic mice. *Environ. Microbiol.* **11**, 1562–1571 (2009).
164. Arthur, J. C. et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120–123 (2012).
165. Schuppler, M., Lotzsch, K., Waidmann, M. & Autenrieth, I. B. An abundance of *Escherichia coli* is harbored by the mucosa-associated bacterial flora of interleukin-2-deficient mice. *Infect. Immun.* **72**, 1983–1990 (2004).
166. Hoenjten, F. et al. Antibiotics with a selective aerobic or anaerobic spectrum have different therapeutic activities in various regions of the colon in interleukin 10 gene deficient mice. *Gut* **52**, 1721–1727 (2003).
167. Zenewicz, L. A. et al. IL-22 deficiency alters colonic microbiota to be transmissible and colitogenic. *J. Immunol.* **190**, 5306–5312 (2013).
168. Dennis, K. L. et al. Adenomatous polyps are driven by microbe-instigated focal inflammation and are controlled by IL-10-producing T cells. *Cancer Res.* **73**, 5905–5913 (2013).
169. Abecia, L. H. L., Khoo, C., Frantz, N. & McCartney, A. Effects of a novel galactooligosaccharide on the faecal microbiota of healthy and inflammatory bowel disease. *Int. J. Probiotic. Prebiotics* **5**, 61–68 (2010).
170. Xu, J. et al. Does canine inflammatory bowel disease influence gut microbial profile and host metabolism? *BMC Vet. Res.* **12**, 114 (2016).
171. Jones-Hall, Y. L., Kozik, A. & Nakatsu, C. Ablation of tumor necrosis factor is associated with decreased inflammation and alterations of the microbiota in a mouse model of inflammatory bowel disease. *PLoS ONE* **10**, e0119441 (2015).
172. Bel, S. et al. Reprogrammed and transmissible intestinal microbiota confer diminished susceptibility to induced colitis in TMF^{–/–} mice. *Proc. Natl Acad. Sci. USA* **111**, 4964–4969 (2014).
173. Bloom, S. M. et al. Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell. Host. Microbe* **9**, 390–403 (2011).
174. He, Q. et al. Dysbiosis of the fecal microbiota in the TNBS-induced Crohn's disease mouse model. *Appl. Microbiol. Biotechnol.* **100**, 4485–4494 (2016).
175. Couturier-Maillard, A. et al. NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. *J. Clin. Invest.* **123**, 700–711 (2013).
176. Ettreiki, C. et al. Juvenile ferric iron prevents microbiota dysbiosis and colitis in adult rodents. *W. J. Gastr.* **18**, 2619–2629 (2012).
177. Devkota, S. et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10^{–/–} mice. *Nature* **487**, 104–108 (2012).
178. Zhang, Z. et al. Chlorogenic acid ameliorates experimental colitis by promoting growth of *akkermansia* in mice. *Nutrients* **9**, 677 (2017).
179. Ghosh, S., Molcan, E., DeCoffe, D., Dai, C. & Gibson, D. L. Diets rich in n-6 PUFA induce intestinal microbial dysbiosis in aged mice. *Br. J. Nutr.* **110**, 515–523 (2013).
180. Ye, J. et al. Bacteria and bacterial rRNA genes associated with the development of colitis in IL-10(–/–) mice. *Inflamm. Bowel Dis.* **14**, 1041–1050 (2008).
181. Vijay-Kumar, M. et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
182. Zhu, W. et al. Precision editing of the gut microbiota ameliorates colitis. *Nature* **553**, 208–211 (2018).
183. Moon, C., Stupp, G. S., Su, A. I. & Wolan, D. W. Metaproteomics of colonic microbiota unveils discrete protein functions among colitic mice and control groups. *Proteomics* **18**, 1700391 (2018).
184. Du, Z. et al. Development of gut inflammation in mice colonized with mucosa-associated bacteria from patients with ulcerative colitis. *Gut Pathog.* **7**, 32 (2015).
185. Bibiloni, R., Simon, M. A., Albright, C., Sartor, B. & Tannock, G. W. Analysis of the large bowel microbiota of colitic mice using PCR/DGGE. *Lett. Appl. Microbiol.* **41**, 45–51 (2005).
186. Kudelka, M. R. et al. Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk. *Proc. Natl Acad. Sci. USA* **113**, 14787–14792 (2016).
187. Vereecke, L. et al. A20 controls intestinal homeostasis through cell-specific activities. *Nat. Commun.* **5**, 5103 (2014).



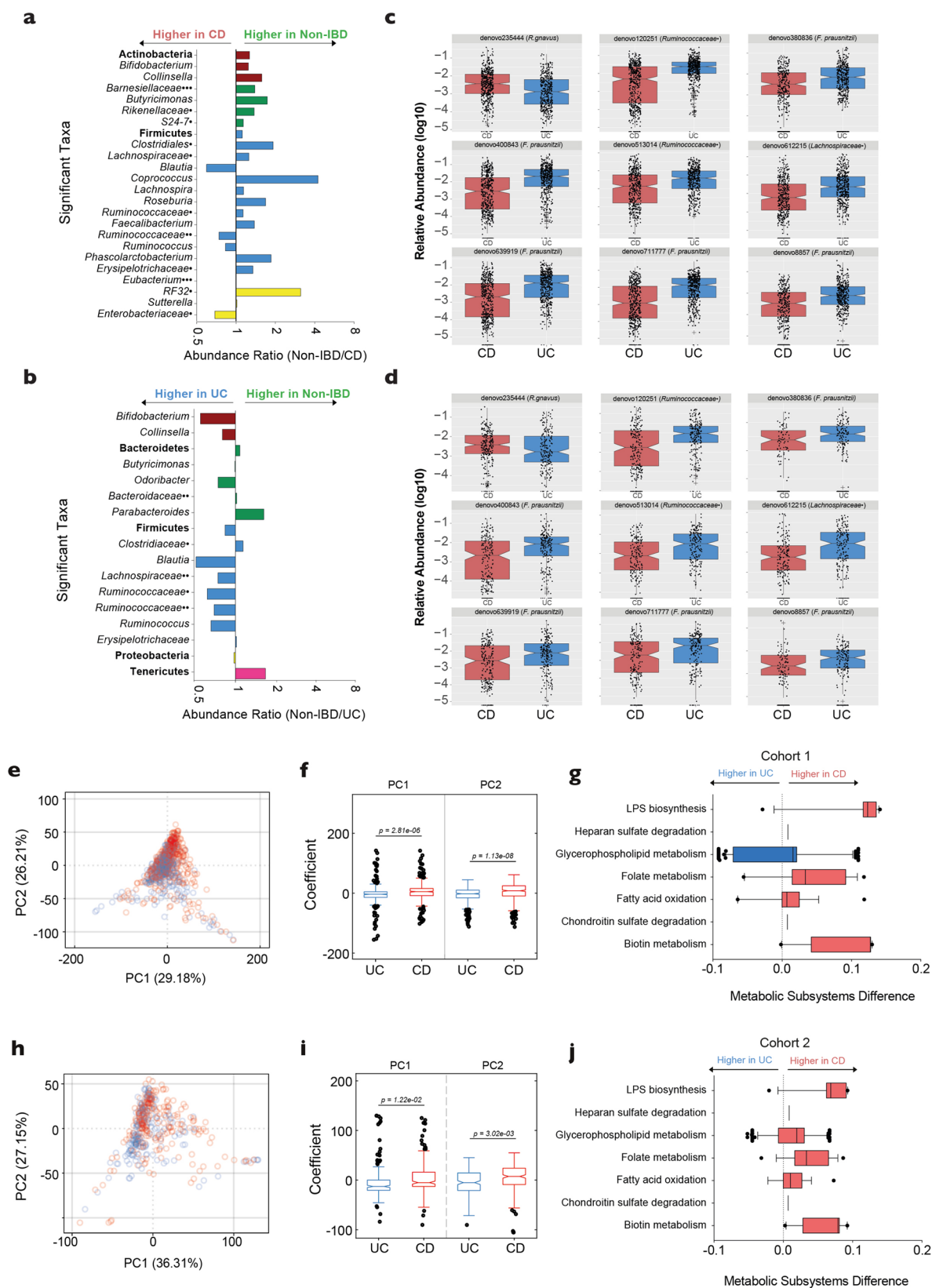
Extended Data Fig. 1 | Unique microbial taxa identified as IBD signatures using unsupervised meta-analysis of published human IBD studies. **a**, The heatmap shows significant microbial changes in CD, UC or IBD (CD + UC) disease groups compared to non-IBD subjects. Each disease group (CD, UC or CD + UC) was compared independently to non-IBD and each color code reports the direction of microbial changes in each respective disease status against non-IBD subjects. Euclidean clustering was performed for sample annotations (vertical) including race/ethnicity, gender, median age, patient number, sample type, sequencing method and microbial taxa (horizontal) at different taxonomic ranks. Taxa in black bold label with asterisk demonstrate those findings verified by a subset of the findings in Fig. 1. Taxa in gray bold label with gray asterisk demonstrate findings in some of the studies verified by the findings in Fig. 1 specifically, *Lachnospiraceae* family, *Lachnospira*, *Coprococcus*, Clostridiales order, *Faecalibacterium*, *Ruminococcus*, *Roseburia* and *Ruminococcaceae* family for Group 1, and *Actinobacteria*, *Proteobacteria* phyla and *Enterobacteriaceae* family for Group 2. 18 significant additional replicated taxa in the heterogeneous Group 3 are *Bacteroidetes* phylum and genera from this phylum including *Bacteroides*, *Odoribacter*, *Butyrivomona*, *Parabacteroides*, *Sutterella*, *Prevotella* (of *Prevotellaceae*), *Prevotella* (of *Paraprevotellaceae*) and *Rikenellaceae* family; also *Firmicutes* and genera from this phylum including *Phascolarctobacterium*, *Dialister*, *Eubacterium*••• and *Ruminococcus*•••; *Blautia*, *Collinsella* and *Bifidobacterium*; *Sutterella* from *Proteobacteria* and *Tenericutes*. Underlined taxa are matching with Cluster CD_A in Fig. 2. **b**, The heatmap (studies are in order as in **a**) shows clinically relevant information collected through the studies analyzed in **a**. Recorded clinical phenotyping information in a given study is shown in green color (Identified) and the lack of clinical data is represented in white color (non-identified).



Extended Data Fig. 2 | Microbial taxa comparison of CD with UC in published IBD studies. The heatmap shows the comparison of microbial changes between CD and UC. Taxa higher in CD (lower in UC) are in red, while taxa higher in UC (lower in CD) are in blue. Euclidean clustering was performed for sample annotations (vertical) including race/ethnicity, gender, median age, patient number, sample type and sequencing method and microbial taxa (horizontal) at different taxonomic ranks. Taxa names in bold demonstrate those findings verified by a subset of the findings in Fig. 1 and taxa names in gray bold shows the findings that partially match with our findings in Fig. 1.

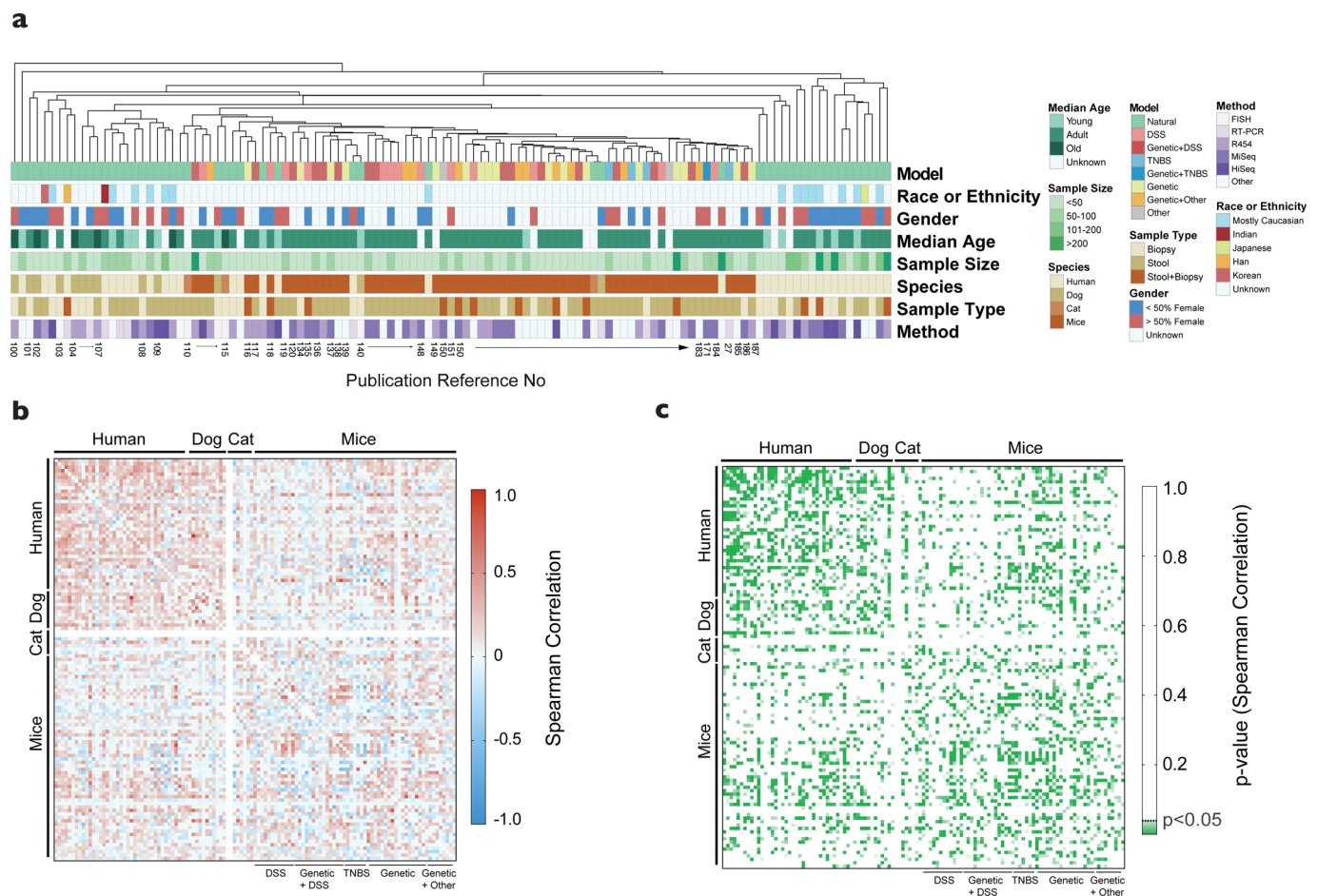


Extended Data Fig. 3 | The dominant bacterial phyla along the gastrointestinal tract of IBD patients and dysbiosis in IBD patients. **a,b**, Distribution of predominant bacterial phylotypes along the cephalocaudal axis of the gut in CD, UC and non-IBD subjects of Cohort 1 (**a**) and Cohort 2 (**b**) are depicted after stratification according to the relative abundance of Firmicutes at each sampling site. The dominant bacterial phylotypes are Bacteroidetes (51.6% IBD Cohort 1; 56% IBD Cohort 2 and 56% non-IBD), Firmicutes (34.9, 29 and 25.7%, respectively) and Proteobacteria (9.1, 17 and 14%); with a smaller proportion of Fusobacteria (0.8, 0.4 and 1.1%), Actinobacteria (0.79, 0.46 and 0.33%) and Tenericutes (0.2, 0.04 and 0.09%). **c-f**, Microbial composition differences between IBD patients and non-IBD subjects were identified by species richness (Observed OTUs, Shannon and Simpson indices) in Cohort 1 (**c**) and Cohort 2 (**d**) and microbiome clustering based on unweighted and weighted UniFrac PCoA metrics for Cohort 1 (**e**) and Cohort 2 (**f**). Box-and-whisker plots in **c** and **d** display first and third quartiles and whiskers are from each quartile to the minimum or maximum. **g,h**, Beta dispersion statistics were performed by analyzing the sampling distance to centroids for Cohort 1 (**g**) and Cohort 2 (**h**) and there is no significant differences between compared groups in **g** and **h**. **i,j**, Only significant taxa associated with CD or UC shown as relative abundance ratio in Cohort 1 (**i**) or Cohort 2 (**j**) were identified using MaAsLin pipeline with BH-FDR correction (q value). $q < 0.05$ was considered significant. Significant differences were determined by either non-parametric two-sided Mann-Whitney U -test (**c,d,g,h**) or Adonis test for multiple comparisons (**e,f**) and $P < 0.05$ was considered significant. Box-and-whisker plots in **c,d,g,h** display first and third quartiles and whiskers are from each quartile to the minimum or maximum. 494 CD and 447 UC samples in Cohort 1 and 230 CD, 195 UC and 770 non-IBD samples in Cohort 2 were used for analysis (**a-j**).

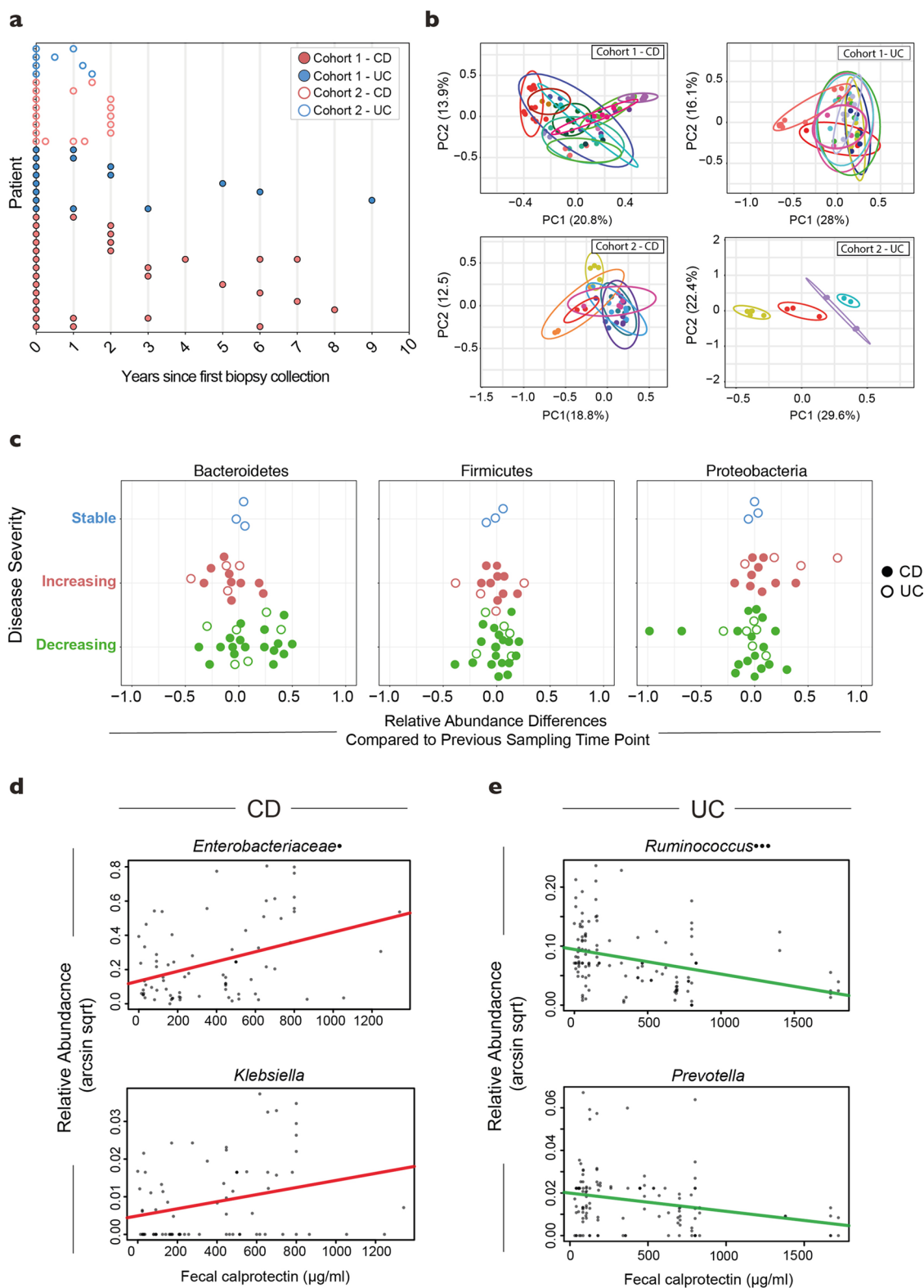


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Microbial taxa and functional metabolic subsystem differences in IBD patients. a,b, Significant taxonomic differences are depicted as relative abundance ratios between CD and non-IBD samples (**a**) and between UC and non-IBD samples (**b**) identified in MaAsLin pipeline with BH-FDR correction. $q < 0.05$ was considered significant. **c,d,** Relative abundance of the most important matching OTUs were identified using machine learning algorithm and are depicted for Cohort 1 (**c**) and Cohort 2 (**d**) using notched box whisker showing first and third quartiles with median value. Each dot represents a single sample (**c,d**). **e,f,h,i,** After mapping OTUs to metabolic reactions, calculated the metabolic distances between all pairs of patients based on raw reaction counts are shown on PCA plots based on L2 distance of total reaction counts between UC and CD and boxplots show the respective coefficients PC1 and PC2 axis in Cohort 1 (**e,f**) and Cohort 2 (**h,i**). PC1 and PC2 are the first two principal components. **g,j,** The principal component analysis (PCA) analysis illustrates robust data at the metabolic reaction level: (~60% variance explained by PC1/2 Blue indicates UC and red indicates CD patients and are shown using notched box whisker plots showing first and third quartiles with median value. Metabolic subsystems different between in CD and UC patients were identified in Cohort 1 (**g**) and in Cohort 2 (**j**). Similar significant metabolic pathway enrichment was detected in both cohorts. Red color represents the enrichment in CD and blue color represents the enrichment in UC. Box-and-whisker plots in **g,j** display first and third quartiles and whiskers are from each quartile to the minimum or maximum and possible outliers. Consistent metabolic subsystems increased in CD belonged to B-vitamin and LPS biosynthesis, heparan sulfate and chondroitin sulfate degradation and fatty acid oxidation. The BH-FDR was applied to correct for multiple testing and $q < 0.05$ between groups was considered significant (**f,i**). Fisher's exact test was performed to determine if the subsystem was overrepresented among the significantly different reactions; subsystems with $P < 0.05$ were considered enriched (**g,j**). 494 CD and 447 UC samples in Cohort 1 and 230 CD, 195 UC and 770 non-IBD samples in Cohort 2 were used for analysis (**a-d**).

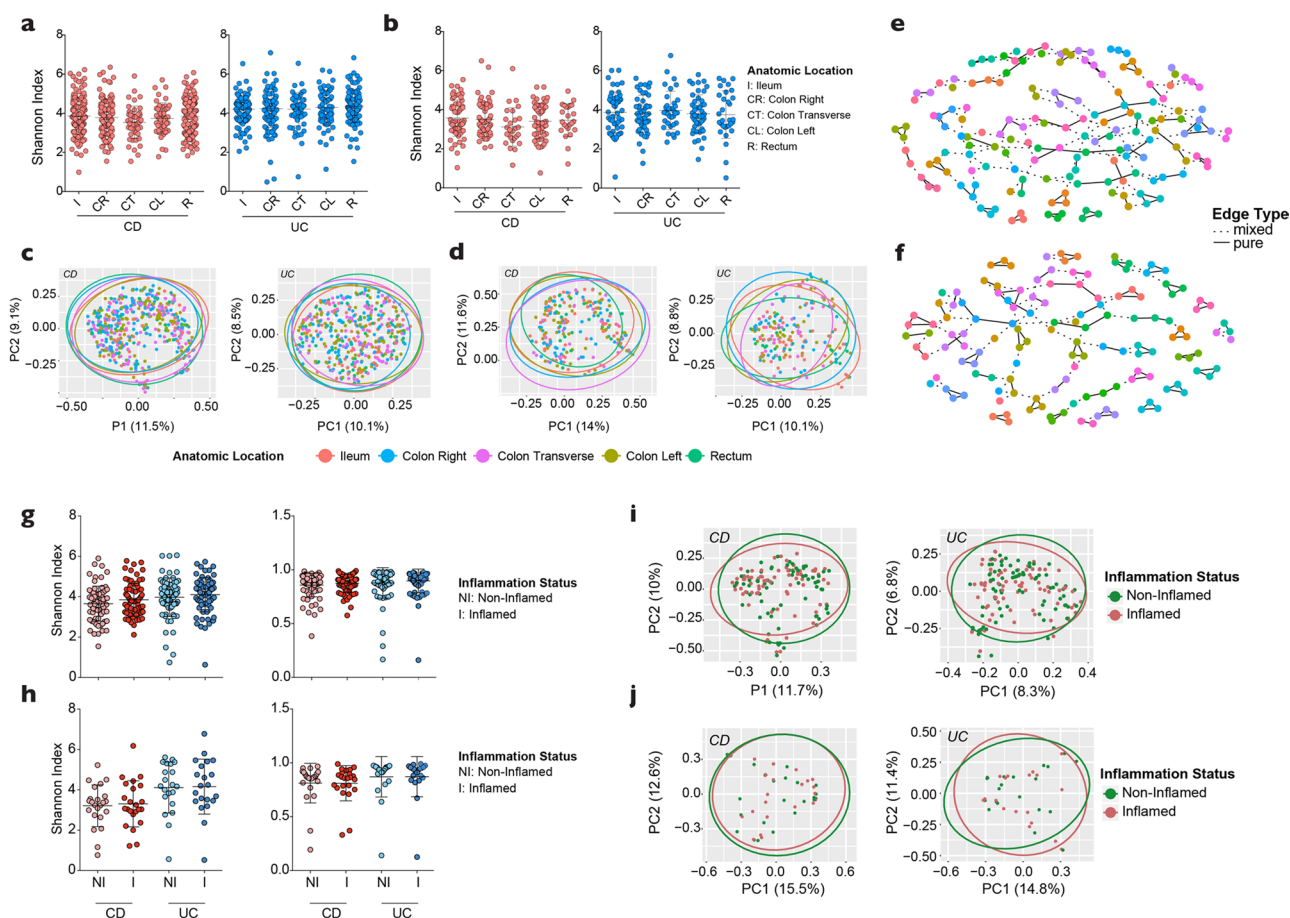


Extended Data Fig. 5 | Unique microbial taxa identified as IBD signatures across different species using unsupervised meta-analysis of the published IBD studies. **a**, The Euclidean clustering of IBD patients and animal models of IBD including dogs/cats diagnosed with IBD and mice with genetically and/or chemically (DSS or TNBS) induced colitis was performed using information of taxa identified significantly changing between disease groups and is plotted using the categorical information of the study models such as according to race/ethnicity, gender, median age, species, subject number, sample type, sample size, sequencing method and experimental model of IBD induction. **b**, The Spearman correlation heatmap shows the correlation between 123 different human and animal IBD studies based on identified 96 differentially abundant microbial taxa that are characterized in **a**. The correlation values ranging from 0 to 1 show positive correlation (in red) and the values ranging from -1 to 0 show negative correlation (in blue) between compared IBD studies. **c**, Statistical information of data of 123 independent human and animal IBD studies in total (**a**, **b**) is shown on the same matrix. Non-parametric two-tailed Spearman correlation test was performed and $P < 0.05$ was considered significant. Green color shows significant correlation between taxa plotted in **a**.

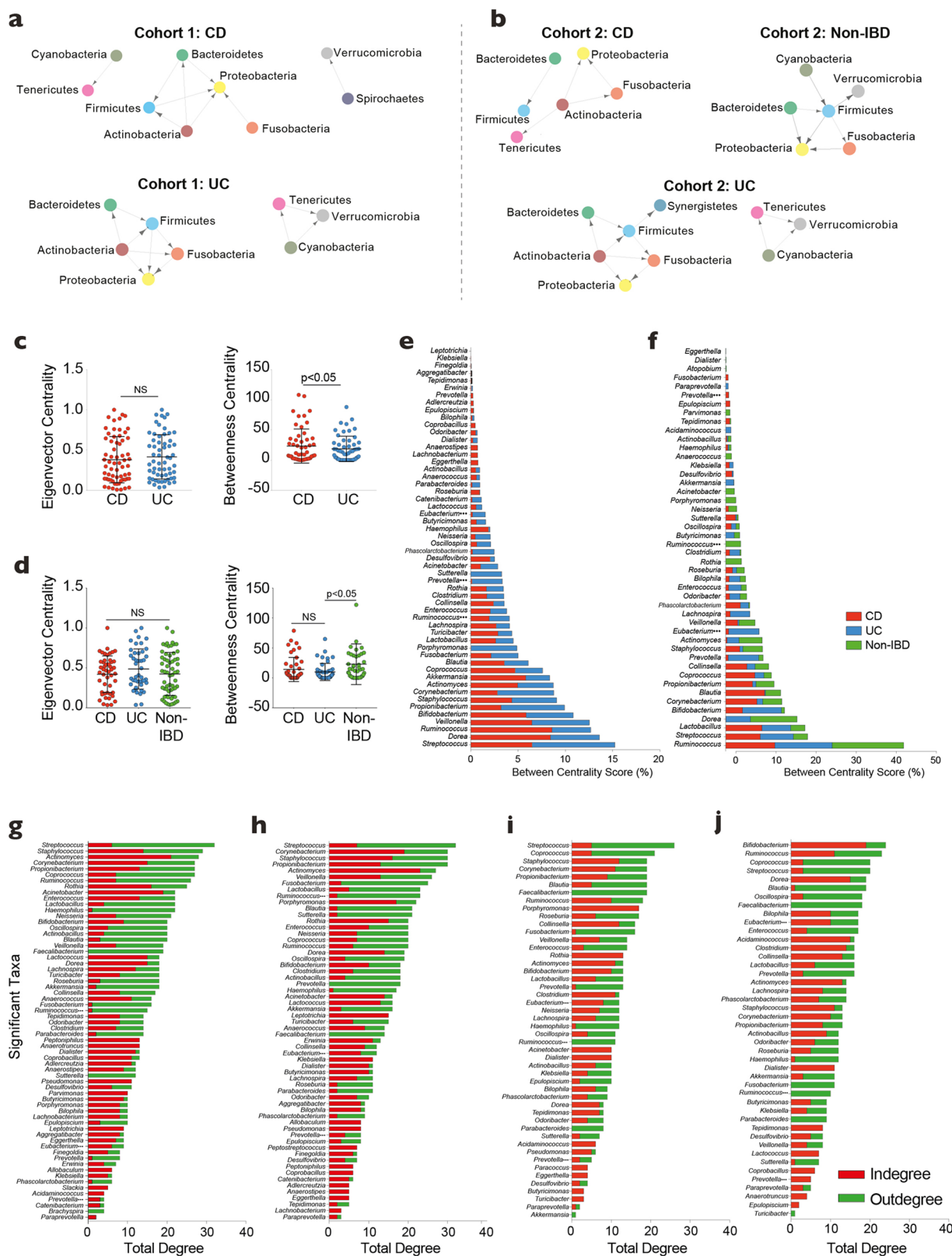


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Microbial stability over time in longitudinally studied IBD patients and correlation of intestinal inflammation with microbial abundance. **a**, Biopsies were collected from 22 individuals in Cohort 1 and 12 individuals in Cohort 2 over several years (1–9 years and 0.25–2 years, respectively). Each row corresponds to the time course of an individual patient. The resulting data comprised 176 biopsy samples. **b**, PCoA on Bray–Curtis dissimilarity distance matrix for longitudinally collected (as shown in **a**) 77 CD and 44 UC samples in Cohort 1 and 49 CD and 14 UC samples in Cohort 2 are plotted. Each color in represents an individual IBD patient. Ellipsoids represent a 95% confidence interval surrounding each disease group. **c**, The relative abundance changes for Bacteroides, Firmicutes and Proteobacteria phyla in IBD patients are plotted based on their disease severity changes over time. The x axis shows the relative abundance difference of a given phylum compared to previous sampling time point. A value higher than zero indicates that the phylum increases and a value lower than zero indicates that the phylum decreases compared to the previous sampling time point. Disease severity worsening over time is labeled as ‘decreasing’, improving over time is labeled as ‘increasing’ and stable disease severity is identified as ‘steady’ on the y axis. **d,e**, Fecal calprotectin that is positively correlated with *Enterobacteriaceae* and *Klebsiella* for 79 CD patients (**d**) and negatively correlated with *Ruminococcus* and *Prevotella* in 42 UC patients are shown on continuous data plot generated in MaAsLin pipeline with $q < 0.05$ (**e**). Spearman's rank correlation coefficient for taxa in CD: 0.284 and 0.147 and for taxa in UC: -0.322 and -0.2 , respectively. Adonis test was used to determine significant differences between the distance matrix of each group (**b**). Data shown in **b** was not significant when longitudinal samples were compared for individuals ($P > 0.05$). However, significant microbial differences were only observed between patients ($P < 0.05$). Taxa significantly associated with disease severity and fecal calprotectin were identified in MaAsLin pipeline (**c,d,e**) with BH-FDR correction and significant taxa are plotted. The $q < 0.05$ was considered significant.

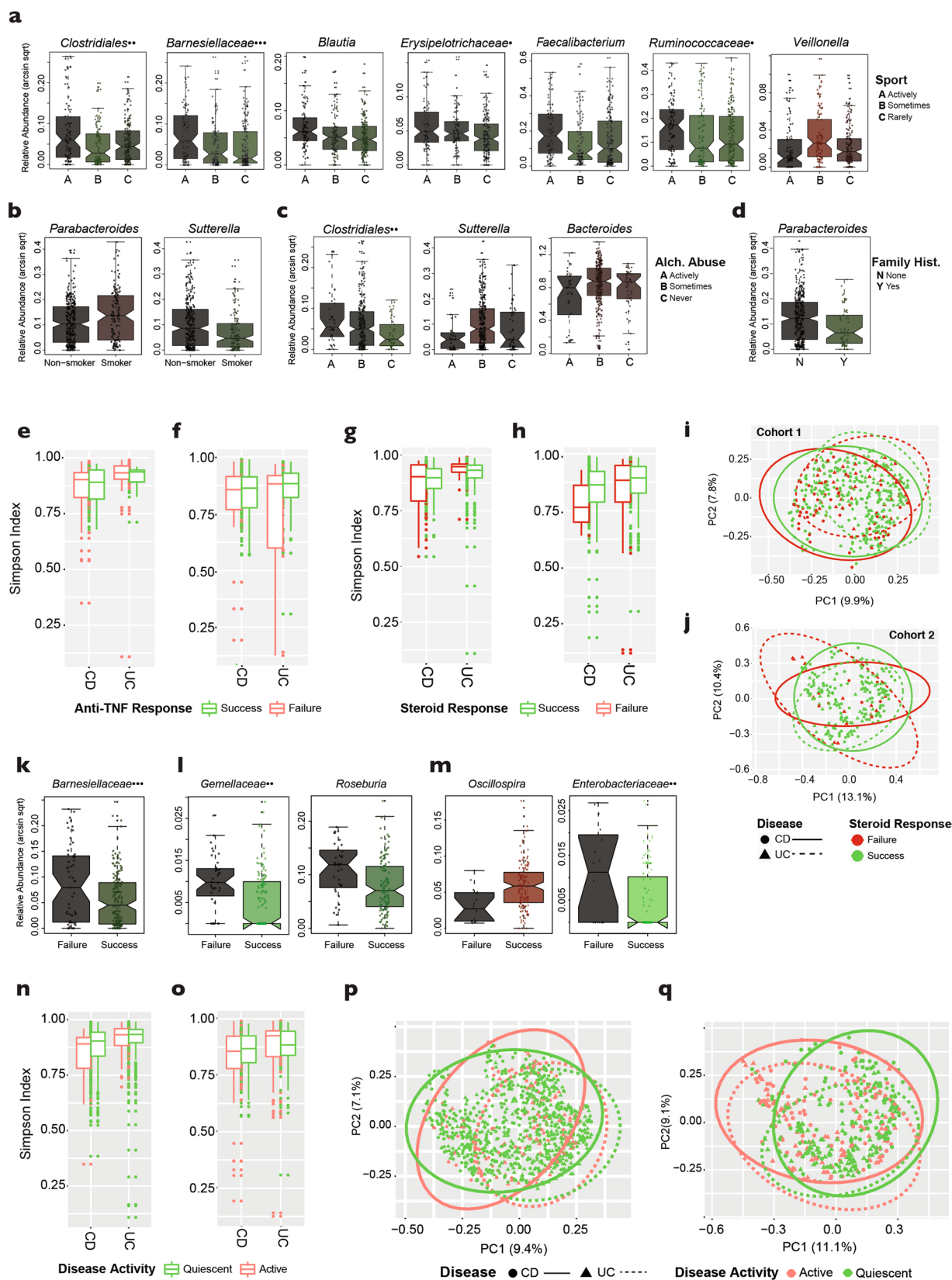


Extended Data Fig. 7 | Microbial profile along the gut in IBD patients. **a,b**, Species richness of samples collected along the gut including Ileum (I), right colon (RC), transverse colon (CT), left colon (CL) and rectum (R) were calculated with Shannon index for 494 CD and 447 UC samples shown in Cohort 1 (**a**) and 230 CD and 195 UC in Cohort 2 (**b**). **c,d**, Beta diversity of these samples are shown for Cohort 1 (**c**) and Cohort 2 (**d**). **e,f**, Samples collected from same patients cluster intra-individually, as depicted CD (**e**) and UC (**f**) patients. **g-j**, Species richness calculated with Shannon and Simpson indices for 494 CD and 447 UC samples shown in Cohort 1 and 230 CD and 195 UC in Cohort 2 with different inflammation status are shown in **g** for Cohort 1 and in **h** for Cohort 2. Beta diversity of these samples individually analyzed for CD and UC are shown for Cohort 1 (**i**) and Cohort 2 (**j**). Significant differences between groups were determined by one-way ANOVA corrected for multiple comparisons using BH-FDR and there is no significance between compared groups ($q > 0.05$) (**a,b,g,h**). Lines indicate mean values and error bars are standard deviations (**a,b,g,h**). Adonis test was used to determine significant differences between the dissimilarity distance matrix of each group and groups are not significantly different than each other (**c,d,i,j**). The edges strongly similar to each other are connected with a solid line (pure edge) and the edges partially similar to each other are connected with dashed lines (mixed edge) (**e,f**).



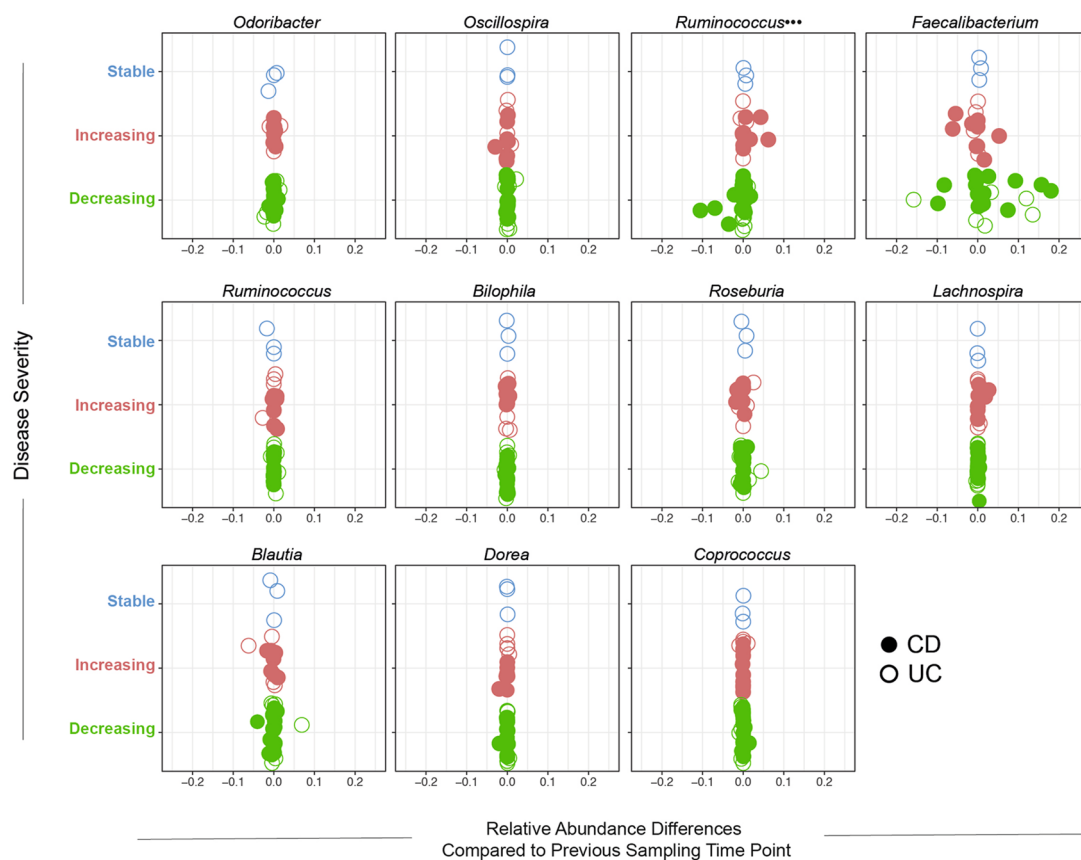
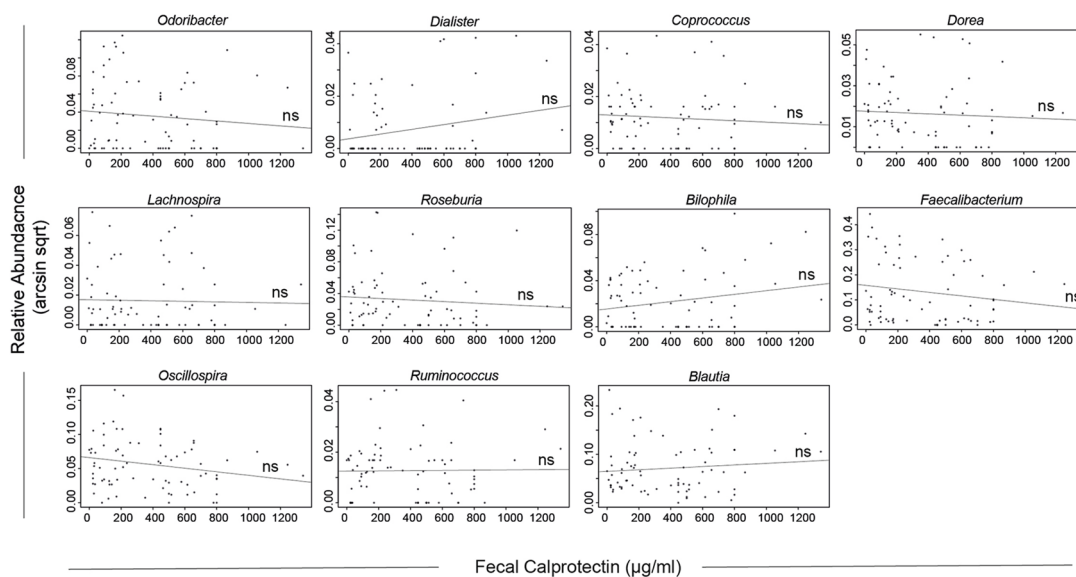
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Co-occurrence patterns and degree centrality scores identify the important components of IBD. **a,b**, Ecosystem-specific co-occurrence patterns are visualized using network diagrams where microbial phyla represent nodes and the presence of a positive co-occurrence relationship based on correlation is represented by an edge in Cohort 1 (**a**) and Cohort 2 (**b**). Co-occurrence relationships with less strong Spearman's correlation coefficients (ρ value > 0.25 and $P < 0.05$) are depicted with network diagram for each disease. **c,d**, The value of eigenvector and betweenness centralities for CD, UC and non-IBD samples were calculated in Cohort 1 (**c**) and in Cohort 2 (**d**). Significant differences between groups were determined by non-parametric two-sided Mann-Whitney U -test (**c**) and ordinary one-way ANOVA corrected for multiple comparisons using BH-FDR correction (**d**) and $P < 0.05$ was considered significant and significant results are shown on the plot. Lines indicate mean values and error bars are standard deviations. 65 CD and 61 UC taxa from Cohort 1 and 48 CD, 44 UC and 41 non-IBD taxa from Cohort 2 were used for analysis in **c** and **d**. **e,f**, Important genera based on their between centrality score for Cohort 1 (**e**) and Cohort 2 (**f**) are depicted for each cohort (CD in red, UC in blue and Non-IBD in green). **g-j**, Prominent and influential taxa were identified using in- and out-degree scores are shown for CD (**g,i**) and UC (**h,j**) for corresponding cohorts: Cohort 1 (**g,h**) and Cohort 2 (**i,j**). Taxa with ρ value > 0.25 and $P < 0.05$ are plotted (**e-j**). (Cohort 1, CD phyla 9 nodes (N) and 13 edges (E); UC phyla 8 N and 12E; CD genera 64 N, 473E; UC genera, 60 N, 440E and Cohort 2: CD phyla 6 N, 7E; UC 9 N, 11E; CD genera 40 N, 273E; UC genera 38 N, 276E).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Gut microbiota differences in IBD patients with different lifestyles and different responsiveness to disease. a–d. Major taxonomic changes were observed in 494 CD samples in Cohort 1 when samples were analyzed for sport activities (**a**), smoking status (**b**), alcohol abuse (**c**) and family history of disease (**d**). **e–g**, Species richness biopsy samples obtained from patients responsive (success) or unresponsive (failure) to anti-TNF- α therapies (**e,f**) and corticosteroid therapies (**g,h**) are shown in **e,g** for Cohort 1 and in **f,h** for Cohort 2. **i,j**, Microbial clustering of intestinal biopsy samples from IBD patients responding or non-responding to corticosteroids therapies for Cohort 1 (**i**) and Cohort 2 (**j**) is shown with PCoA on Bray–Curtis distance dissimilarity metrics. CD (solid line) and UC (dashed line) are used to identify the disease groups on PCoA plot. **k–m**, Unique microbial taxa identified as a signature of responding and non-responding groups in CD (182 with success and 47 with failure) (**k**) and in UC (**l**) are shown for Cohort 1 (131 with success and 36 with failure) and in UC (146 with success and 16 with failure) for Cohort 2 (**m**). **n,o**, Species richness (of biopsy samples obtained from patients with different disease activities, characterized by the frequency of exacerbations (active) and remissions (quiescent) are shown for Cohort 1 (**n**) and for Cohort 2 (**o**). Data is not significant in **n** and **o**. **p,q**, Microbial clustering of intestinal biopsy samples from IBD patients with different disease activities for Cohort 1 (**p**) and Cohort 2 (**q**) is shown with PCoA on Bray–Curtis distance dissimilarity metrics. CD (solid line) and UC (dashed line) are used to identify the disease groups on PCoA plot. 494 samples in CD and 447 samples in UC for Cohort 1 (**p**) and 226 samples in CD and 195 samples in UC for Cohort 2 (**q**) were analyzed. Microbial profiles were analyzed using MaAsLin pipeline with BH-FDR correction (q value) and $q < 0.05$ was considered significant (**a–d,k–m**). Significant taxa (**a–d,k–m**) are plotted using notched box whisker showing first and third quartiles and median value. (**a**) Sport: actively (several times per week), sometimes (once or twice a week) and rarely (less than once a week); (**c**) alcohol abuse; (**d**) family history: N (None), Y (Yes). Mann–Whitney U -test was used for statistical analysis of alpha diversity (**e–h,n,o**). No significant differences in species richness observed between groups ($P > 0.05$). Box-and-whisker plots display quartiles and range with standard deviations with possible outlier shown with dots in **e–h,n–o**. Adonis test assessed the significant difference differences between the dissimilarity distance matrix of each group (**i,j,p,q**). $P < 0.05$ for each compared group in each cohort (**i,j,p,q**).

a**b**

Extended Data Fig. 10 | The relative abundance changes in Cluster CD_A taxa with disease activity and intestinal inflammation. a, The relative abundance changes of taxa in CD_A cluster in longitudinally studied 34 IBD patients (77 CD and 44 UC samples in Cohort 1 and 49 CD and 14 UC samples in Cohort 2) based on the clinically defined changes in disease activity over time as described in Fig. 6c. **b**, The correlation between fecal calprotectin of 78 CD patients and Cluster CD_A taxa. Data was analyzed using MaAsLin with BH-FDR correction and data was not significant (NS; $q > 0.05$).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

► Experimental design

1. Sample size

Describe how sample size was determined.

Both in Swiss IBD Cohort (Cohort 1) and Bern Uni. Hosp. Cohort (Cohort 2), we have extracted the biopsy samples of patients who were sampled during the colonoscopy for 5 different intestinal locations (ileum, right colon, left colon, transverse colon and rectum). No sample size calculation was performed. This was a meta-analysis that used all available samples of participants with biopsies from 5 different locations.

2. Data exclusions

Describe any data exclusions.

In this study we have collected the biopsy samples from 5 different intestinal locations per patient. We processed all samples however, in some cases we didn't get enough DNA that was sufficient for 16S rRNA PCR samples and hence, those samples were excluded. After sequencing, samples with more than 4500 high quality reads were used for further analysis. We have established the cut-off criteria based on a rarefaction analysis and phylogenetic resolution that we obtained with alpha diversity analysis.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

We have run this study with two independent cohorts. Cohort 1 is Swiss IBD cohort that exists more than years. We have first obtained samples from this cohort and sequenced for 16S rRNA microbial profile. Then, as replica set we have the Cohort 2 that consists of newly recruited IBD patients and non-IBD subjects in the framework of the human intestinal community project of Bern University Hospital. We have analyzed the data for Cohort 2 with the settings we established for Cohort 1 and in our manuscript we reported both cohort results side by side. The experimental findings in Cohort 1 were reliably reproduced through repeated experiments using Cohort 2.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

IBD Patients were assigned with unique ID number according to their recruitment time and we processed/sequenced the sample by increasing number order and hence, patient with different disease status were randomized. Randomization is all dependent on the recruitment time of the patients.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Biopsy samples were assigned with unique patient ID number and hence, until we sequence biopsy samples for microbial profile we didn't have information of the clinical metadata information of the patients.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ A statement indicating how many times each experiment was replicated
- ☐ ☒ The statistical test(s) used and whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- ☐ ☒ Test values indicating whether an effect is present
*Provide confidence intervals or give results of significance tests (e.g. *P* values) as exact values whenever appropriate and with effect sizes noted.*
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

We initially analyzed the 16s rRNA data using QIIME v1.9.1. OTU picking step assigns OTUs by clustering sequences based on a user-defined similarity threshold, namely UCLUST (Edgar, RC 2010) that creates “seeds” of sequences which generate clusters based on percent identity. The otu_table.biom file, mapping file (prepared using Microsoft Excel 2016) and rep_set.tree file were loaded into phyloseq package in R version 3.3.0 for further analysis. Plots were generated using phyloseq object or GraphPad Prism v7. Taxonomy analysis were performed in Maaslin in R and all-to-all hierarchical clustering analysis was performed using HALLA (<https://huttenhower.sph.harvard.edu/halla>) from C. Huttenhower's group. Microbial association network construction within disease status were generated using script provided in manuscript Williams et. al., Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* 5. and network plots were generated using Cytoscape v3.5.1. All reaction level analysis were performed in MATLAB (R2018a, MathWorks)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

All 16S rRNA sequencing data / Clinical metadata of patients are readily available from the authors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No eukaryotic cell lines were used.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Study did not involve any mouse models.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Biopsy samples in this study were collected either a) in the Swiss IBD cohort study or b) newly recruited IBD patients and non-IBD subjects in the framework of the human intestinal community project of Bern University Hospital. Non-IBD biopsy samples were collected from subjects that were registered for a screening ileo-colonoscopy without any gastrointestinal symptoms, no suspected functional intestinal symptoms and where all other work-up was negative. Age range of the participants into this study is between 20-75 years and the gender (% female) value is CD: 56%, UC: 60% in Cohort 1 and CD: 42%, UC: 21% , Non-IBD: 53% in Cohort 2. More than 90% of the participants are Caucasian and mean disease duration in these patients in each cohort is around 10 years. The descriptions of Inclusion criteria of human participants have been provided in the "Cohorts and study design" method section. Briefly, inclusion criteria for Swiss IBD cohort patients were on established Lennard-Jones criteria, confirmed by radiological, endoscopic and histological findings. Additional characteristics, including disease activity, surgery status, disease location, remission status and therapy are detailed in Suppl. Table 1 for Cohort 1 and for Suppl. Table 2 for Cohort 2.

Licensed gastroenterologists collected biopsy samples and clinical data of patients identified as CD or UC and non-IBD subjects. The Swiss IBD Cohort study was approved by all local ethical committees of the participating hospitals: The Bern Human Intestinal Community project was approved by the Bern Cantonal Ethics Commission (Ref: KEK-BE: 251/14 and 336/14). This statement is included in the methods section.