

Dynamics of the human gut microbiome in inflammatory bowel disease

Jonas Halfvarson¹, Colin J. Brislawn², Regina Lamendella³, Yoshiki Vázquez-Baeza⁴, William A. Walters⁵, Lisa M. Bramer⁶, Mauro D'Amato^{7,8}, Ferdinando Bonfiglio⁹, Daniel McDonald¹⁰, Antonio Gonzalez¹¹, Erin E. McClure³, Mitchell F. Dunklebarger³, Rob Knight^{4,10,11} and Janet K. Jansson^{2*}

Inflammatory bowel disease (IBD) is characterized by flares of inflammation with a periodic need for increased medication and sometimes even surgery. The aetiology of IBD is partly attributed to a deregulated immune response to gut microbiome dysbiosis. Cross-sectional studies have revealed microbial signatures for different IBD subtypes, including ulcerative colitis, colonic Crohn's disease and ileal Crohn's disease. Although IBD is dynamic, microbiome studies have primarily focused on single time points or a few individuals. Here, we dissect the long-term dynamic behaviour of the gut microbiome in IBD and differentiate this from normal variation. Microbiomes of IBD subjects fluctuate more than those of healthy individuals, based on deviation from a newly defined healthy plane (HP). Ileal Crohn's disease subjects deviated most from the HP, especially subjects with surgical resection. Intriguingly, the microbiomes of some IBD subjects periodically visited the HP then deviated away from it. Inflammation was not directly correlated with distance to the healthy plane, but there was some correlation between observed dramatic fluctuations in the gut microbiome and intensified medication due to a flare of the disease. These results will help guide therapies that will redirect the gut microbiome towards a healthy state and maintain remission in IBD.

Both the state and the dynamics of the human gut microbiome in healthy individuals are highly personalized^{1–7}. Although cross-sectional studies have revealed dysbiosis of the gut microbiome in inflammatory bowel disease (IBD)^{8–12}, little is known about the individual nature of microbiome dynamics in IBD, beyond a study of three ulcerative colitis (UC) patients before and after ileostomy and two small studies of IBD patients in remission or during changes in disease activity^{13–15}. Here, we study the long-term dynamics of the gut microbiome from an IBD cohort of 137 individuals (49 Crohn's disease (CD), 60 UC, 4 lymphocytic colitis (LC), 15 collagenous colitis (CC)) and 9 healthy controls (HC). We sampled at three-month intervals, collecting 1–10 samples per individual for a total of 683 samples (Supplementary Table 1). The microbiome composition in each sample was determined by sequencing the V4 region of the 16S rRNA gene for a total of 248 million 16S rRNA gene amplicons. To determine links between the gut microbiome and clinical factors, we collected clinical data, including faecal calprotectin (f-calprotectin) concentration and surgical resection status. To control sampling bias, we restricted our statistical analyses of volatility to a subset of the cohort that had sequence data from the first four time points and that had matching f-calprotectin concentrations. This yielded 276 samples from 69 patients (Supplementary Table 1 and Supplementary Section 1; results were similar when all subjects were considered). We also included patient genetic load scores (GLSs) based on 163 known

IBD risk loci for 29 patients to assess potential links between the host genetics, IBD and the microbiome¹⁶.

As expected from previous work^{10,11}, by principal coordinates analysis (PCoA) of unweighted UniFrac distances we found that HC and IBD subtypes formed distinct clusters, with ileal Crohn's disease (ICD) patients least similar to healthy controls (Supplementary Fig. 1, ADONIS stratified by time point, $P < 0.001$). As in previous studies^{10,12,13}, we found differences in the alpha and beta diversity of the microbiome according to IBD subtype and between subtypes and healthy controls, and identified several families that correlated with health or disease state, for example, Enterobacteriaceae with ICD and Ruminococcaceae with HC (Supplementary Figs 1 and 2). Individual taxa that are differentially abundant between IBD subtypes compared to HC are listed in Table 1 (DESeq2, log₂ fold change). The IBD microbiomes contained significantly lower abundances of putative beneficial microorganisms present in HC, as previously reported^{8–13}, including *Prevotella copri* and the butyrate-producing bacterium *Faecalibacterium prausnitzii* (Table 1 and Supplementary Fig. 3).

From an animated ordination of the samples (Supplementary Video 1), we observed that although microbiome samples from the healthy individuals varied over time, they were restricted to a small volume of the ordination space. In contrast, IBD subtypes traversed far more of the total volume, sporadically visiting the area where healthy samples resided. To summarize these dynamics, we identified a 'healthy plane' (HP). Briefly, this plane is calculated

¹Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro SE-701 82, Sweden. ²Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99354, USA. ³Juniata College, Huntingdon, Pennsylvania 16652, USA. ⁴Department of Computer Science and Engineering, University of California, San Diego, California 92093, USA. ⁵Max Planck Institute, 72076 Tübingen, Germany. ⁶National Security Directorate, Pacific Northwest National Laboratory, Richland, Washington 99354, USA. ⁷Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, SE-171 76 Stockholm, Sweden. ⁸BioDonostia Health Research Institute San Sebastian, and IKERBASQUE Basque Foundation for Science Bilbao 48011, Spain. ⁹Department of Biosciences and Nutrition, Karolinska Institutet, SE-171 77, Stockholm Sweden. ¹⁰Department of Pediatrics, University of California, San Diego, California 92093, USA. ¹¹Center for Microbiome Innovation, University of California, San Diego, California 92093, USA.

*e-mail: janet.jansson@pnnl.gov

Table 1 | Differential abundance in specific taxa according to disease phenotype comparisons (DESeq2).

Groups compared	BaseMean	log2(fold change)	padj	Taxonomic annotation
ICD-r over ICD-nr	13.32	-7.05	0.0000000	<i>Faecalibacterium prausnitzii</i>
	4.08	-5.57	0.0000011	Lachnospiraceae
	3.32	-5.08	0.0000001	Ruminococcaceae
	6.49	-5.40	0.0000009	Ruminococcaceae
	19.85	-6.11	0.0000005	Ruminococcaceae
	94.87	-5.87	0.0000001	Ruminococcaceae
	53.59	-8.10	0.0000012	Ruminococcaceae <i>Ruminococcus</i>
	72.13	-5.14	0.0000241	Clostridiales
ICD-r over HC	14.91	7.20	0.0000000	Alteromonadales [Chromatiaceae]
	13.32	-7.22	0.0000000	<i>Faecalibacterium prausnitzii</i>
	2.94	-5.34	0.0000007	Ruminococcaceae
	2.33	-5.62	0.0000000	Clostridiales
	10.41	-7.47	0.0000000	Lachnospiraceae
	15.12	-7.62	0.0000001	Lachnospiraceae <i>Coproccoccus</i>
	4.13	-8.43	0.0000000	Lachnospiraceae
	19.85	-7.15	0.0000000	Ruminococcaceae
	5.43	-8.72	0.0000000	Ruminococcaceae
	7.16	-6.98	0.0000151	Clostridiales
	3.78	-6.69	0.0000249	Clostridiales
	2.10	-6.53	0.0000000	Ruminococcaceae
	5.76	-7.85	0.0000000	Ruminococcaceae
	53.59	-8.64	0.0000000	Ruminococcaceae <i>Ruminococcus</i>
	72.13	-5.71	0.0000000	Clostridiales
	121.13	-9.95	0.0000000	<i>Prevotella copri</i>
	5.87	-7.58	0.0000000	<i>Methanobrevibacter</i>
ICD-nr over HC	14.91	6.47	0.0000185	Alteromonadales [Chromatiaceae]
	2.94	-7.00	0.0000756	Ruminococcaceae
	5.43	-8.17	0.0000682	Ruminococcaceae
	121.13	-7.82	0.0000185	<i>Prevotella copri</i>
CCD over HC	5.43	-8.65	0.0000000	Ruminococcaceae
	121.13	-7.94	0.0000000	<i>Prevotella copri</i>
UC over HC	6.53	6.32	0.0000978	<i>Alistipes massiliensis</i>

Criteria for inclusion: BaseMean > 1 and padj < 0.0001. Brackets indicate putative taxonomy based upon phylogenetic placement as given in the Greengenes taxonomy. BaseMean is the mean of normalized counts for all samples; padj is the Benjamini-Hochberg adjusted *P* value.

in a space derived from PCoA of unweighted UniFrac distances of healthy subjects (Supplementary Video 1). We constructed a model using the samples from the HC patients and fit them to a two-dimensional plane embedded in a three-dimensional space using the least-squares method (Fig. 1). The plane is then restricted to only span the three-dimensional ranges of the HC samples. This plane was used as a proxy to represent the normal microbial variation within healthy subjects and to summarize the abnormal, intermittent dysbiosis associated with IBD.

The procedure was as follows. Let S be a set of n samples $s_1, s_2, \dots, s_n$ corresponding to a group of trajectories, each trajectory pertaining to a subject with at least four samples collected at distinct points in time. Each sample is represented as a three-dimensional vector corresponding to sample coordinates in ordinated space, that is, $s_1 = (x_1, y_1, z_1)$; $s_2 = (x_2, y_2, z_2)$; ... $s_n = (x_n, y_n, z_n)$. We fit a linear model to S by the least-squares method to obtain coefficients for the equation of a three-dimensional surface T , then we restrict a segment of this surface to the ranges given by $[\min_x(S), \max_x(S)]$ and $[\min_y(S), \max_y(S)]$, and we define this to be a plane representative of S or P_s . When S is the set of samples from healthy subjects in an ordination space we refer to P_s as the healthy plane (HP). Finally, we define d_k to be the Euclidean distance from a sample k to the nearest point lying on P_s . After measuring d_k for all samples in our study, we group samples according to their diagnosis and compare the distributions of distances. Figure 1a–c demonstrates this procedure and Fig. 1d,e demonstrates the placement of the HP in the context of our full IBD data set. Samples from the HP are co-located with one another, whereas many samples from IBD patients are further away. To investigate whether this effect is due to ‘outlier’ groups of samples that are dominated by taxa that are

typically rare in healthy individuals, we excluded all Proteobacteria from the data set and compared the location of samples in unweighted UniFrac space with and without Proteobacteria, using Procrustes analysis. As shown in Fig. 1f,g, the omission of Proteobacteria is significant ($P < 0.01$), but this effect is largest in the already dysbiotic ICD-r patients (ICD patients that had previously undergone ileocaecal resection) and aligns with PC3. The effect on healthy controls is minimal and healthy samples are still located only near the HP (Fig. 1f,g).

Calculation of the mean Euclidean distance from each sample to the HP revealed that all subtypes of IBD significantly deviated from the HP (generalized linear mixed effects model (GLM), all $P < 0.00268$, Fig. 2a). Samples from patients with colonic Crohn’s disease (CCD) and UC were closer to the HP, and some samples did not differ significantly from healthy controls, whereas ICD samples were the most distant from the HP (Supplementary Video 1 and Fig. 2). The highest volatility was observed for ICD-r patients, followed by ICD patients without surgery (ICD-nr) based on UniFrac distances between successive samples (Fig. 2b). ICD-r patients also had low gut microbial richness (Supplementary Fig. 2). Ileocaecal resection is a major modifier of intestinal physiology and the observed pronounced volatility in ICD-r patients might be partly explained by removal of the ileocaecal valve *per se*. Ileal inflammation may also influence bile salt uptake and, consequently, colonic transit time and microbiome volatility. Interestingly, several IBD patients had complex trajectories that sporadically moved to and from the HP (Fig. 2c and Supplementary Video 1).

Although our study represents the largest longitudinal analysis of the IBD microbiome so far, the small number of patients in specific subgroups limited some statistical comparisons. Furthermore, the

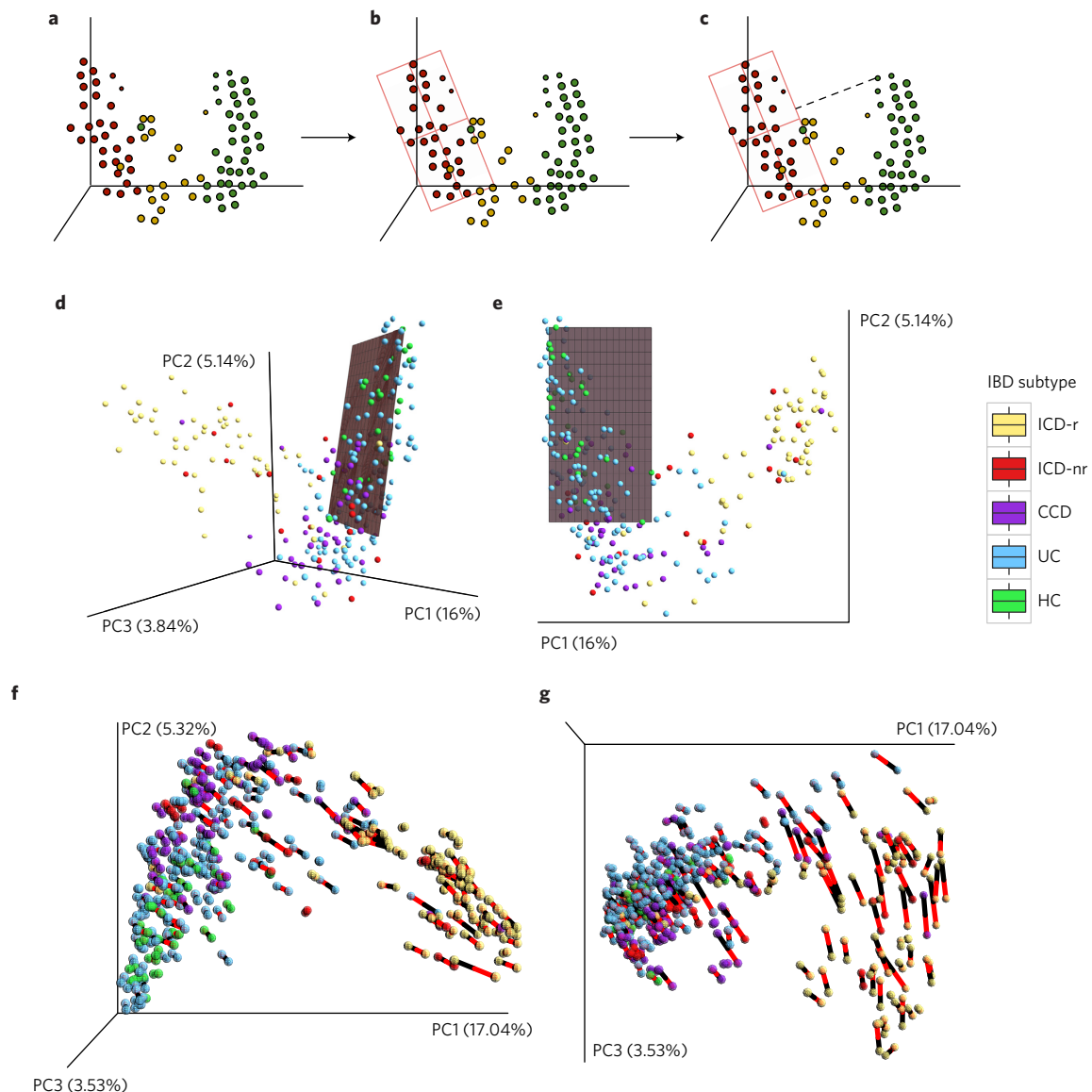


Figure 1 | Defining a healthy plane. **a–c**, Diagrams summarizing the procedure for creating a representative plane for a group of samples (S): sample selection (**a**), model fitting (**b**) and distance calculations (**c**) for all samples. Colours in **a–c** represent example groups used for fitting the plane. **d,e**, The healthy plane is then located in UniFrac space by fitting a line to the major axis of the points (**d**) and defining a least-squares fit to identify a plane that minimizes the sum of squares of distances to the nearest point on the plane (**e**). **f**, Verification that the position of the healthy plane is not driven by Proteobacteria-dominated outliers, with Procrustes analysis comparing original samples and those with Proteobacteria removed. A vector connects each original sample (red) with the same samples after Proteobacteria have been omitted (black). $P < 0.001$, sum of squared distances between objects after fitting, $M^2 = 0.018$, 999 permutations. **g**, The short length of most vectors indicates that the relative composition of most samples does not change when Proteobacteria are filtered out. Colours in **d–g** represent IBD subtype.

number of healthy controls was lower than the number of IBD patients. To address this limitation of unequal sampling of healthy individuals and IBD patients, we compared the volatility of our healthy controls to healthy participants in two published studies, the Student Microbiome Project (SMP) and the Moving Pictures (MP) data sets^{6,17}. There was less variability over time in healthy individuals across all three cohorts compared to those with IBD (Fig. 2b). This result emphasizes that IBD is characterized by volatile dysbiosis not found in healthy people and confirms earlier preliminary results and meta-analysis of much smaller studies^{13–15,18}.

To extend our understanding of the mechanisms underlying the microbiome dynamics, we explored the correlation between the dynamics and inflammatory activity in each sample using f-calprotectin $>150 \mu\text{g g}^{-1}$ as a surrogate for inflammatory activity. The concentration of f-calprotectin in stool samples has previously

been correlated with endoscopic and histopathological activity and is used in daily clinical practice because it is non-invasive¹⁹. We observed that concentrations of f-calprotectin were higher in all IBD subtypes than in healthy controls (Fig. 3). However, we did not observe a significant correlation between f-calprotectin and distance from the HP (Fig. 3, GLM $P = 0.275$). Although recent microarray analyses of the gut microbiome in a cohort of anti-tumour necrosis factor-treated paediatric IBD patients²⁰ and experiments with gnotobiotic faecal transplants suggest that microbial composition and function are causally associated with inflammatory activity²¹, our differential abundance testing revealed only weak trends and no specific taxa that varied significantly with active inflammation, using f-calprotectin $>150 \mu\text{g g}^{-1}$ as cutoff. However, the use of f-calprotectin as a proxy for inflammatory activity might have introduced bias, because f-calprotectin is a less accurate

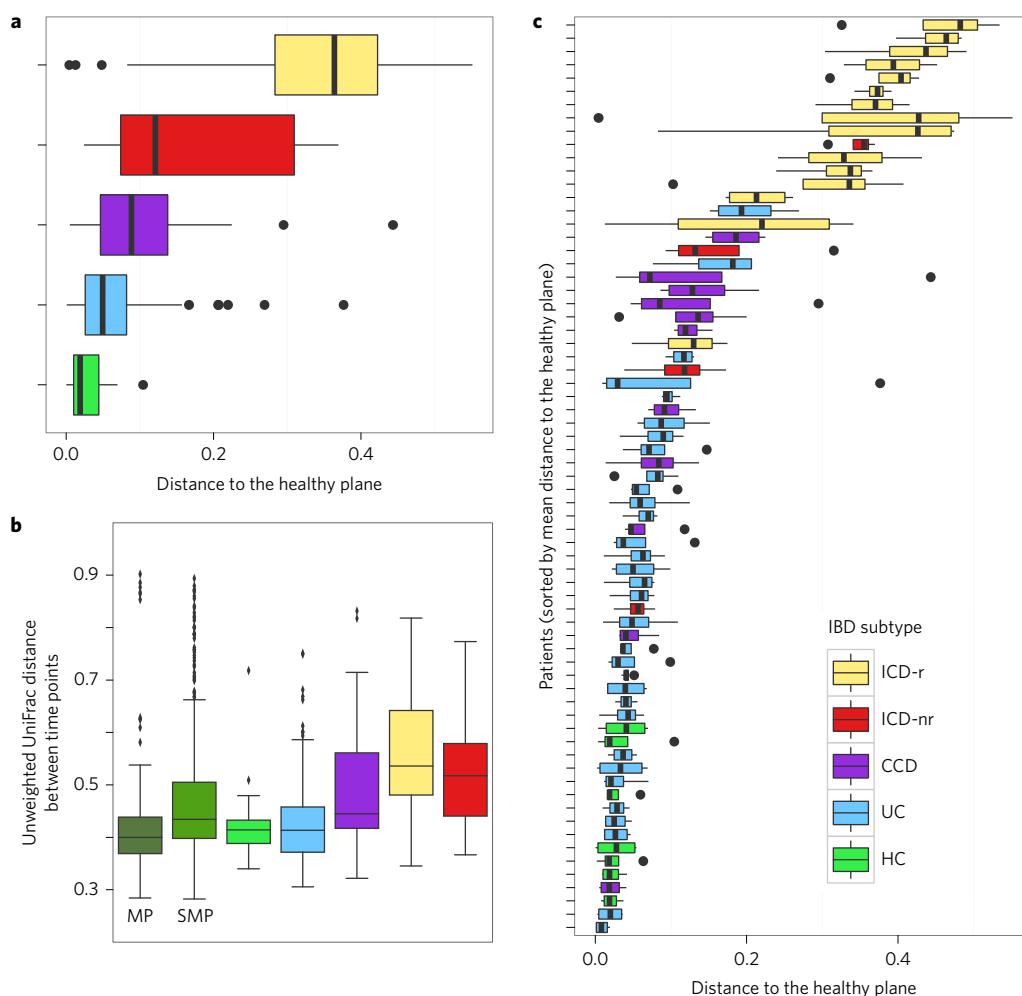


Figure 2 | The gut microbiomes of different IBD subtypes display different distributions relative to a healthy plane. **a**, Median distances from healthy plane (HP) for each IBD subtype. All IBD subtypes were significantly different from healthy controls (GLM, all $P < 0.00268$). **b**, UniFrac distances between subsequent samples. **c**, Distance to HP for each individual patient. HP was defined using data shown in Supplementary Video 1. Supplementary Table 1 shows the composition of the downstream analysis cohort. Boxes show interquartile range (IQR). Whiskers denote the lowest and highest values within $2.5 \times$ IQR of the median. Circles represent outliers. SMP, Student Microbiome Project; MP, Moving Pictures. Distances to the healthy plane are euclidean distances based on the ordinated UniFrac matrix.

marker of ileal than colonic inflammation²², and ICD patients displayed the greatest distance from the HP.

Examples of the microbiome dynamics for one representative HC and one from each IBD subgroup are shown in Fig. 4, in addition to changes in f-calprotectin concentrations, distance to HP and Shannon diversity (all individual profiles are shown in Supplementary Fig. 3), illustrating the more stable dynamics over time for HC and UC compared to the other clinical phenotypes of IBD, with the most fluctuations occurring for patient 69, who had undergone surgical resection. The f-calprotectin levels were low and relatively stable for the HC compared to the IBD patient (Fig. 4). In these examples, there were also substantial fluctuations in diversity for the HC, ICD-r and CCD patients compared to UC and ICD-nr patients.

We also explored the individual dynamics of the gut microbiome in IBD patients who experienced increased clinical disease activity according to the physician's global assessment (Supplementary Fig. 4). Recently, the short-term (6 weeks) dynamics of the microbiome in paediatric patients with active IBD treated with anti-TNF indicated that initiation of medical treatment changes the microbial composition at the genus level²⁰. Because we had few anti-TNF-exposed patients and our patients received a course of corticosteroids at flare as first-line therapy, we explored how corticosteroid administration influenced microbiome dynamics. Our data demonstrate that

a change in medication influences the volatility of the microbiome (Supplementary Fig. 4). Patients receiving a course of oral corticosteroids ($n = 7$) had more microbiome fluctuations than patients on stable medication ($n = 49$), based on calculations of the unweighted UniFrac distance between time points (Wilcoxon signed-rank test, $P = 0.04$). Our dynamic model suggests that beyond the association with IBD subtype and the weak correlation with inflammation, the dynamics of microbiome composition are influenced by changes in medication. The extent to which other factors, such as dietary changes and smoking, may have influenced the observed volatility remains speculative, because the collected information was insufficiently detailed to include these factors as covariates.

To evaluate the microbiome as a predictive tool, we combined the microbial and clinical data and used a supervised learning Random Forests model to predict IBD subtypes^{23,24}. To avoid overfitting, our models were built using abundances of OTUs (operational taxonomic units) from the first time points only, together with clinical metadata (body mass index (BMI), f-calprotectin concentrations, sex, and distance to the HP). Accuracy was evaluated using the remaining three time points, which were not used to train the model. Using this model, the IBD subtypes were discriminated from healthy controls and correctly predicted for 66.6% of samples (Supplementary Table 2), consistent with the findings

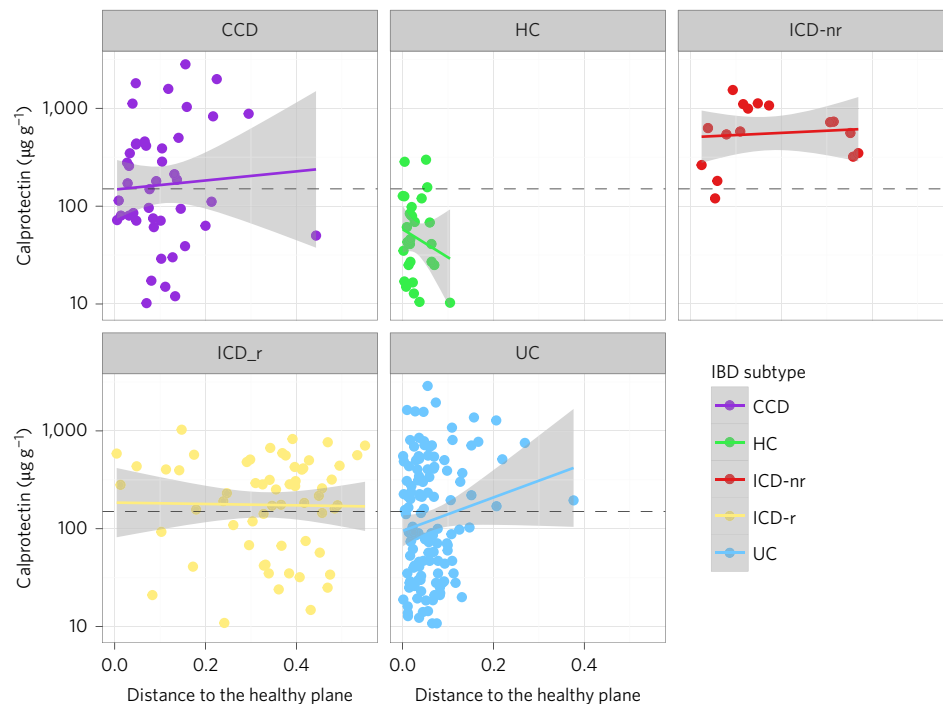


Figure 3 | Correlation between faecal calprotectin concentrations and distance to a defined HP in three-dimensional ordination space. Data represent a correlation of f-calprotectin levels and distance to the here defined HP in three-dimensional ordination space (Supplementary Video 1) for each individual and time point for different IBD subtypes. To compare the relationship between f-calprotectin and the HP, a GLM model was fit, with a conditional gamma distribution, using f-calprotectin and disease type as fixed effects and including a random subject effect. f-Calprotectin was not significant ($P = 0.27501$).

reported in ref. 11. Feature importance scores from this model revealed several potential microbial indicators of IBD subtypes, including OTUs matching to *Lachnospira*, *Clostridium*, *Oscillospira* and many unidentified Ruminococcaceae (Supplementary Table 3). Intriguingly, the accuracy of the model increased slightly if f-calprotectin concentrations were omitted, but decreased by at least 10% if the distance to the HP was removed (Supplementary Tables 2 and 3), suggesting that the HP is a more important factor in the model. Comparable levels of accuracy were previously achieved using rectal samples¹¹, but here we show that the same can be achieved with faecal samples, which are easier to collect. When immunochip data were included for a subset of 29 IBD individuals and the Random Forests model was repeated, the samples were still classified into the four IBD subtypes (UC, CCD, ICD-r and ICD-nr) (Supplementary Tables 2 and 3). Although distance to the HP remained the single most important feature for classification, GLSs were more predictive than sex, f-calprotectin or BMI when included in the model (Supplementary Table 2). However, including GLS only increased the overall accuracy of the model by about 2%, demonstrating the predictive potential of the microbiome. Interestingly, in a recent study of obesity covering 339,224 individuals, the 97 risk loci for obesity accounted for only 2.7% of BMI variation²⁵, whereas the microbiome classified lean from obese individuals with 90% accuracy²⁶, providing a precedent for the predictive value of the gut microbiome over human genetics in chronic disease.

In summary, by analysing faecal samples collected every three months from a large IBD cohort, we determined the long-term volatility of the gut microbiome in IBD. Our data revealed that although the microbiome of healthy individuals varied, it was only within a newly defined HP, whereas there was considerable volatility away from the HP for several of the IBD cohorts. Devising improved methods to detect the healthy state with non-invasive sampling, to predict when the healthy state will be departed, and to sustain the microbiome in this healthy state by erecting barriers that prevent the slip back into dysbiosis will be an important focus of future work.

Methods

Cohort demographics. Patients with CD or UC, the two major forms of IBD, attending outpatient clinic were consequently invited to take part. After obtaining written consent, BMI was recorded and patients were asked to provide a faecal sample and to fill in a questionnaire with clinical disease activity, present medication, dietary habits, use of antibiotics and use of non-steroidal anti-inflammatory drugs (NSAIDs). Disease phenotype was classified according to the Montreal classification²⁷. Individuals were then followed prospectively, and asked to provide faecal samples and fill in the questionnaire every third month for a two-year period. If a patient did not provide a faecal sample at any of the three-month periods, a reminder letter was sent. In total, 109 patients with IBD (CD, $n = 49$; UC, $n = 60$) took part. Nine additional individuals with no IBD or any other gastrointestinal conditions were recruited as HC, as well as 19 patients with other chronic inflammatory gastrointestinal diseases (4 lymphocytic colitis (LC) and 15 collagenous colitis (CC)). All 137 individuals were Caucasians, and together they provided 683 faecal samples during the two-year period (Supplementary Table 4). The study was approved by the Ethical Committee of the Medical Faculty, Uppsala University (2007/291).

Sample collection. Faecal samples were self-collected in sterile plastic containers and stored at -80°C until shipping on dry ice and processing.

Faecal calprotectin. To assess the degree of inflammatory activity at the collection of each faecal sample, the concentration of f-calprotectin was assessed by commercially available enzyme-linked immunosorbent assay (ELISA; Calprotectin Elisa Buhlmann Laboratories), according to the manufacturer's protocol.

DNA extraction and amplification. Genomic DNA was extracted from 0.25 g of faecal material from each sample using the Earth Microbiome DNA extraction protocol²⁸. Briefly, DNA was extracted using the 96-well format MoBio Powersoil DNA kit on an EpMotion 5075 robot with vacuum (Eppendorf). DNA was quantified with a Qubit 2.0 fluorometer (Invitrogen) according to the manufacturer's instructions.

PCR amplification and library preparation were performed similarly to the protocol described by Caporaso and co-authors²⁹. 515F/806R Illumina primers with unique reverse primer barcodes were used to target the V4 region of the 16S rRNA gene. Samples were amplified in triplicate and cleaned using the MO BIO 69 htp PCR cleanup kit. Each PCR reaction included 1× PCR buffer, 10 µM each forward and reverse primer, 200 µM dNTPs, 1 U ml⁻¹ Taq polymerase, 15 ng template DNA and PCR grade water, with a total reaction volume of 25 µl. Reactions were kept at 94 °C for 3 min for denaturation to occur. Amplification was performed by 25 cycles of 94 °C for 45 s, 58 °C for 60 s and 72 °C for 90 s. The V4 amplicons were

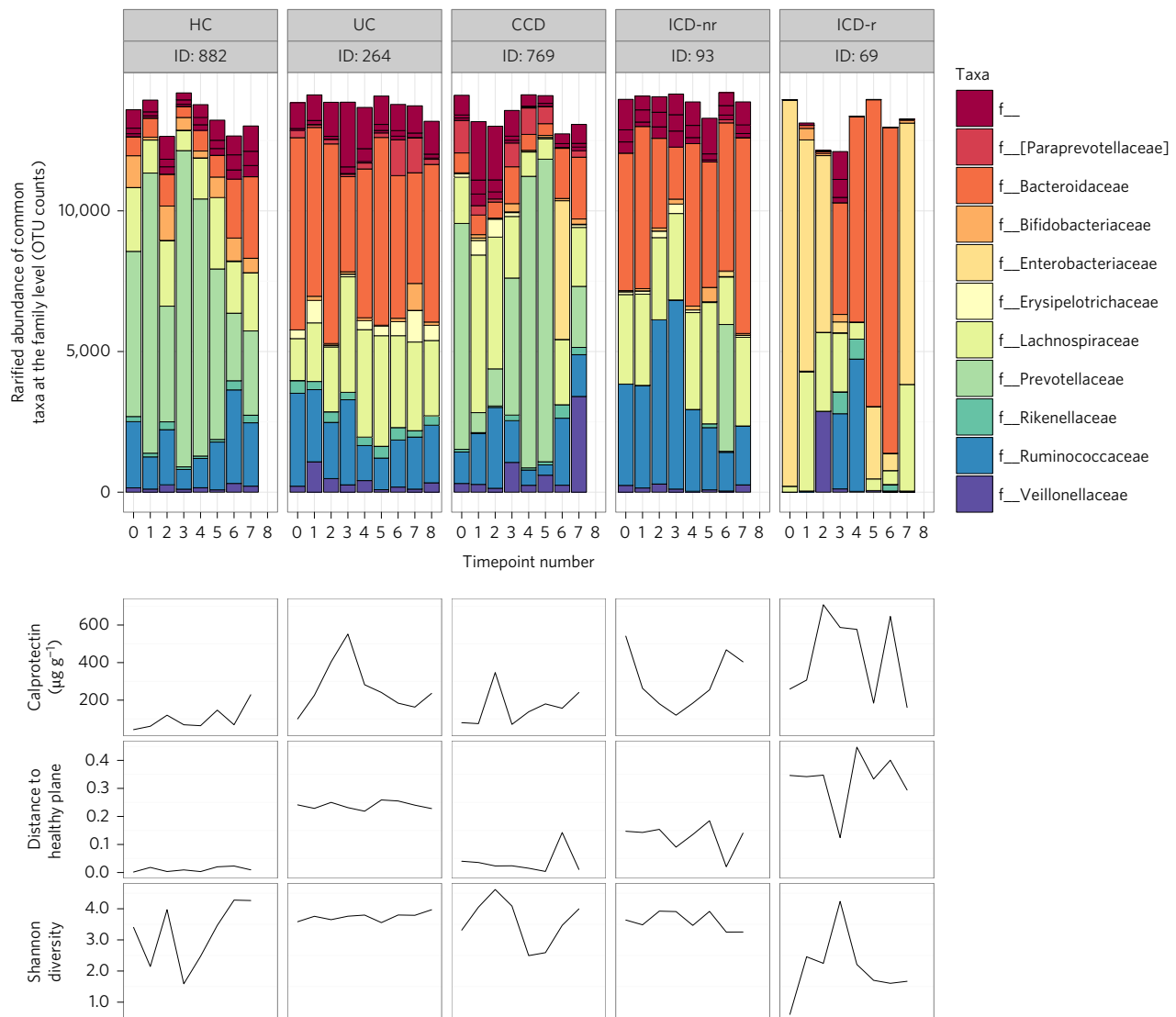


Figure 4 | Microbiome dynamics of selected individuals from each IBD subtype and a healthy control. From each IBD subtype and healthy control (HC) group, representative individuals sampled over the most time points and with complete clinical and sequence data were selected. Data represent f-calprotectin values, distance to the HP and Shannon diversity and rarefied abundances of most common taxa at the family level. Note that taxa unclassified at the family level are represented in the 'f_' category.

sequenced on the Illumina HiSeq 2000 platform, yielding single-end, 100-base-pair reads. Sequencing and quality assessment were performed at the Yale Center for Genome Analysis.

Phylogenetic analysis. Sequence data were processed using QIIME 1.9.0-dev through the online platform QIITA (<https://qiita.ucsd.edu>)³⁰. Four HiSeq lanes of data were demultiplexed with default quality filtering settings and subsequently combined, resulting in 248,547,926 total sequences. These sequences were clustered using SortMeRNA at 97% identity against the Greengenes rRNA reference database May 2013 release^{31,32}. Sequences that failed to match the database were discarded. A total of 237,653,256, or approximately 95.6% of the sequences, clustered against the Greengenes reference data set. Even sampling was performed at 14,553 sequences per sample for beta diversity and supervised learning analyses. The beta diversity principal coordinates plot of unweighted UniFrac distances was constructed using the same rarefied OTU table and visualized in Emperor^{33,34}. Clinical matched metadata, including f-calprotectin concentrations, were included when available. From each patient, the first four microbiome samples with matched f-calprotectin concentrations were subselected for use in downstream analysis.

Statistical analysis. The R package phyloseq was used to import and graph data, and the packages DESeq2, Random Forest and vegan were used to perform differential abundance testing and supervised learning^{24,35–38}. Statistical significance of unweighted UniFrac distance matrices comparing healthy controls and IBD subtypes was assessed using the ADONIS test. The variation of the microbial community over time was calculated with vector lengths produced by summing the total distance

between each subject's time points over the first three PCoA axes of unweighted UniFrac space. The Random Forests model was constructed using the first time point from each patient in the downstream analysis cohort, and prediction accuracy was measured using the subsequent three time points (CCD = 11 patients, ICD-nr = 4 patients, ICD-r = 15 patients, UC = 30 patients and HC = 7 patients). The following metadata categories were used as features for supervised learning: BMI, f-calprotectin (continuous), sex and distance to the healthy plane. For classification on the subset of samples with immunochip data, the above process was used while adding GLSs. We used a data set of 29 IBD samples (25 CD, 4 UC) with available immunochip data to estimate a GLS for each sample. In particular, GLS was calculated summing the number (counts) of risk alleles (0,1,2) at the lead single nucleotide polymorphism from each IBD risk locus ($n = 163$) according to Jostins and co-authors¹⁶.

After performing PCoA of the unweighted UniFrac distance matrix, HC samples were used to fit using the least-squares method on the first three principal coordinates. The distance from each sample to this plane was measured and added to the sample metadata both for weighted and unweighted UniFrac. To compare the distribution of samples relative to this HP, a GLM was fit, with a conditional gamma distribution, using disease type as a fixed effect and including a random subject effect.

To assess the variation between samples in an orderly manner (as described by the time point), we measured the UniFrac (weighted and unweighted) distance between samples that occur sequentially for any given subject. Samples representing the final collection point do not have a value in these columns, as there is no subsequent sample for comparison.

We examined the power of differential abundance tests for the 16S data. DESeq2 (ref. 24) assumes that counts can be modelled as a negative binomial distribution

with a mean parameter, allowing for size factors and a dispersion parameter. The test for differential abundances fits a generalized linear model with a negative binomial family and a log link function.

A power analysis was conducted by using samples from the downstream analysis cohort (Supplementary Table 4). The power of the differential abundance test is dependent on the sample size of groups, difference in mean counts, type 1 error rate and the dispersion value. We used a conventional type 1 error rate of 0.05 and assumed that we had two groups with sample sizes of $n_1 = 25$ and $n_2 = 80$. A total of 1,000 OTUs were randomly selected from the data, and dispersion parameters were estimated for each OTU. Data were then simulated from a negative binomial distribution, as specified by DESeq2, for each estimated dispersion value with means, giving twofold, 1.5-fold and 1.25-fold differences. For each mean fold difference value and dispersion value, a total of 5,000 data simulations were performed and a Wald test for difference in means using a generalized linear model was conducted. Based on these simulations, where data are generated with true differences, the fraction of times that the null hypothesis is correctly rejected (the power) was calculated, and our comparisons were well powered given our subset sample sizes (Supplementary Table 4).

Data availability. Microbiome data from this study are available on Qiita under study ID 1629 (<https://qiita.ucsd.edu/study/description/1629>) and using EBI accession no. ERP020401. Patient clinical information is available on Qiita and in Supplementary Section 1.

Code availability. Our analysis methods make use of standard, open-source software. R software packages are available on CRAN (<https://cran.r-project.org/>), bioconductor (<https://www.bioconductor.org>) and GitHub (<https://github.com/joey711/phyloseq>). Python software is available in the bioconda and biocore Conda channels, and is maintained on GitHub (<https://github.com/biocore>, <https://github.com/ElDeveloper/reference-plane>).

Received 12 September 2016; accepted 5 January 2017;
published 13 February 2017

References

- Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
- Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
- Human Microbiome Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- Martínez, I. *et al.* Gut microbiome composition is linked to whole grain-induced immunological improvements. *ISME J.* **7**, 269–280 (2013).
- Flores, G. E. *et al.* Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15**, 531 (2014).
- Zaura, E. *et al.* Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces. *mBio* **6**, e01693-15 (2015).
- 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).
- Willing, B. *et al.* Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm. Bowel Dis.* **15**, 653–660 (2009).
- 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 (2010).
- Gevers, D. *et al.* The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).
- Rajca, S. *et al.* Alterations in the intestinal microbiome (dysbiosis) as a predictor of relapse after infliximab withdrawal in Crohn's disease. *Inflamm. Bowel Dis.* **20**, 978–986 (2014).
- Wills, E. S. *et al.* Fecal microbial composition of ulcerative colitis and Crohn's disease patients in remission and subsequent exacerbation. *PLoS ONE* **9**, e90981 (2014).
- Young, V. B. *et al.* Multiphasic analysis of the temporal development of the distal gut microbiota in patients following ileal pouch anal anastomosis. *Microbiome* **1**, 9 (2013).
- Martínez, C. *et al.* Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *Am. J. Gastroenterol.* **103**, 643–648 (2008).
- Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
- Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
- Manichanh, C., Borruel, N., Casellas, F. & Guarner, F. The gut microbiota in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 599–608 (2012).
- Lewis, J. D. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. *Gastroenterology* **140**, 1817–1826 (2011).
- Kolho, K.-L. *et al.* Fecal microbiota in pediatric inflammatory bowel disease and its relation to inflammation. *Am. J. Gastroenterol.* **110**, 921–930 (2015).
- 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).
- Sipponen, T. *et al.* Correlation of faecal calprotectin and lactoferrin with an endoscopic score for Crohn's disease and histological findings. *Aliment. Pharmacol. Ther.* **28**, 1221–1229 (2008).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
- Knight, D., Parfrey, L. W., Zaneveld, J., Lozupone, C. & Knight, R. Human-associated microbial signatures: examining their predictive value. *Cell Host Microbe* **10**, 292–296 (2011).
- Silverberg, M. S. *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a working party of the 2005 Montreal World Congress of Gastroenterology. *Can. J. Gastroenterol.* **19** (Suppl. A), 5A–36A (2005).
- DNA Extraction Protocol (EMP, 2011); <http://www.earthmicrobiome.org/emp-standard-protocols/dna-extraction-protocol/>
- Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
- McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
- Lozupone, C. & Knight, R. Unifrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
- Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPERor: a tool for visualizing high-throughput microbial community data. *Gigascience* **2**, 16 (2013).
- Ikahara, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).
- McMurdie, P. J. & Holmes, S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
- Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).
- Oksanen, J. *et al.* vegan: Community Ecology Package. R package v. 2.4 (2016); <https://cran.r-project.org/web/packages/vegan>

Acknowledgements

This research was partially supported by the United States National Institutes of Health (grant NIH IU54DE023798-01), the Pacific Northwest National Laboratory (contract DE-AC05-76RL01830) and by the Crohn's and Colitis Foundation of America. Financial support was also provided by the Örebro University Hospital Research Foundation (grant OLL-507001), the Swedish Foundation For Strategic Research (grant RB13-0160) and the Swedish Research Council (521-2011-2764). Additional support was provided by a grant to Juniata College from the Howard Hughes Medical Institute through the Precollege and Undergraduate Science Education Program and the National Science Foundation (NSF award no. DBI-1248096). R.K. was a Howard Hughes Institute Early Career Scientist for part of the duration of this project.

Author contributions

J.H. and J.K.J. designed the study. J.H. carried out the clinical study and collected samples and clinical data. C.J.B., Y.V.-B., W.A.W., D.M., M.D'A., F.B., A.G., R.L., E.E.M., L.M.B. and M.E.D. performed bioinformatics and statistical analyses of the data. C.J.B., Y.V.-B., W.A.W. and A.G. produced the figures. J.H., C.J.B., Y.V.-B., J.K.J. and R.K. wrote the majority of the text with input from all co-authors.

Additional information

Supplementary information is available for this paper.

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

Correspondence and requests for materials should be addressed to J.K.J.

How to cite this article: Halfvarson, J. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2**, 17004 (2017).

Competing interests

The authors declare no competing financial interests.