Mazzoni *et al. Microbiome* (2026) 14:104
https://doi.org/10.1186/s40168-026-02344-6

Microbiome

RESEARCH

Open Access

Human DNA levels in feces reflect gut inflammation and associate with presence of gut species in IBD patients across the age spectrum

Chiara Mazzoni[1†], Bracha-Lea Ochana[2†], Esther Orlanski-Meyer[3], Ami Ben Ya'acov[3], Gili Focht[3], Esty Harpenas[1], Shimrit Shmorak[1], Oren Ledder[3], Raffi Lev-Tzion[3], Ruth Shemer[2], Eyal Shteyer[3,4*], Yuval Dor[2*] and Moran Yassour[1,5*]

Abstract

Background Feces represent a complex biological matrix that provides valuable information about intestinal physiology and gut microbial activity. Comprehensive fecal DNA sequencing is mostly utilized as a non-invasive way to profile the gut microbiome, and both clinical practice and research on inflammatory bowel diseases (IBD) would greatly benefit from accurate and non-invasive methods to monitor gut inflammation in IBD patients. In IBD, excessive immune cell recruitment and epithelial cell shedding in the gut increase the amount of human DNA in feces, making fecal DNA profiling a desirable approach to monitor gut inflammation dynamics.

Methods We used a combination of sequencing techniques to comprehensively characterize the fecal DNA diversity in a newly established cohort of pediatric IBD patients and controls (Pediatric cohort, $N=134$ children, Israel). We performed methylation-based human cell-specific profiling together with shotgun metagenomics to characterize the human and the microbial DNA content in feces, respectively. Moreover, we included a large complementary external cohort including adult IBD patients and controls (Adult cohort, $N=689$ adults, the Netherlands), not only to compare microbial patterns across the age spectrum, but also to extend our findings from the methylation-based profiling to the more broadly-available quantification of human DNA in metagenomic sequencing.

Results We found that neutrophil DNA dominates fecal human DNA content in IBD patients, and our measurements were highly correlated with fecal calprotectin levels. Combining neutrophil and other cell type DNA fractions in one metric was able to distinguish between remissive and active cases of IBD. Human reads percentage by metagenomics was well correlated with disease severity and species richness, which had distinct trends in CD and UC over time. We used a combination of species richness, human DNA percentage, and microbiome composition data to predict IBD and distinguish CD from UC in both adult and pediatric IBD cohorts.

†Chiara Mazzoni and Bracha-Lea Ochana contributed equally to this study.

*Correspondence:
Eyal Shteyer
eyals@szmc.org.il
Yuval Dor
yuvald@ekmd.huji.ac.il
Moran Yassour
moranya@mail.huji.ac.il
Full list of author information is available at the end of the article

Conclusions The comprehensive characterization of human and microbiome fecal DNA is a useful approach to track immune response level and investigate the interaction that the immune system has with gut microbiome richness and composition over time, enriching opportunities for better disease monitoring and thus better treatment of IBD patients.

Keywords Human DNA, Neutrophils, Microbiome, IBD, Prediction

Background

Human feces contain DNA, not only from the trillions of microorganisms inhabiting the gut (microbiome), but also from human cells, which are shed from the epithelium into the lumen as part of their physiological turnover [1–4]. In healthy individuals, human DNA content in feces is negligible, usually not exceeding 10% of the total fecal DNA [5–7]. However, in case of chronic inflammatory states of the gut, such as in inflammatory bowel diseases (IBD), gut epithelial cells are shed at a faster rate into the lumen [3, 8], and several immune cell types migrate to the gut mucosa in response to pro-inflammatory stimuli [9]. Several studies already reported that in these and other scenarios, human DNA content in feces increases considerably [6, 10, 11]; nevertheless, human DNA quantification in feces has been hardly considered as a possible biomarker of intestinal inflammation so far. In metagenomic pipelines for the analysis of the human microbiome, human DNA is commonly considered an unwanted byproduct of shotgun metagenomics. Human DNA reads are removed early on in the analysis, to reduce computational burden, and most importantly, to protect the identity of participants prior to publication in data repositories [5]. Many have also proposed techniques for reducing human DNA content in DNA samples, especially where a considerable amount, if not the majority of the reads are expected to be human, such as in skin, saliva and vaginal swabs [12–14]. In this context, most microbiome studies report the number of human reads merely as part of the technical assessment of the sequencing process alone.

Although gut microbiome studies have so far ignored human fecal DNA in their analyses, assessing its quantity and origin could complement and enrich microbiome research, especially studies focusing on IBD. The association between IBD and gut microbiome has been heavily researched since the inception of the microbiome field [15–27], yet finding a nexus between gut microbiome composition and IBD onset and development is still an open question [28–30]. Gut microbiome composition changes based on IBD subtype, Crohn's disease (CD), and Ulcerative colitis (UC) being the two major ones, and often has been reported to reflect states of disease flare [31–33]. Moreover, gut microbiome variability across IBD patients has been reported to reach over 60%, largely above the variability explained by factors such as disease location or dysbiosis levels [34], hence searching for the factors that might drive this variability is imperative to better understand, diagnose, and treat IBD. Diagnosing IBD is often challenging, since its symptoms can overlap with other gastrointestinal disorders. The diagnosis process is usually lengthy and dependent on multiple clinical assessments, and delayed diagnosis has been associated with aggravating patients' symptoms [35, 36]. Even after diagnosis, IBD patients are characterized by a large spectrum of inflammatory states that fluctuate over time and require individualized management with diverse pharmacological regimes [37, 38]. As a result, there is an established need for less time-consuming and less invasive diagnostic tools that can alleviate patient burden, allowing for earlier detection of disease, and facilitate more responsive monitoring of treatment efficacy [28]. For example, analysis of the human DNA found in feces could elucidate not only on the general inflammatory status of the patients' gut, but also on tissue involvement, such as colon and small intestinal injury, as well as immune response level and response to medical treatment. To perform such an in-depth assessment, repeatedly over time on chronically ill IBD patients, human cell profiling needs to be cost-effective and non-invasive. In this regard, quantification of total human DNA is straightforward, since computational alignment of the sequencing reads to a human genome reference is routinely used to discard human reads. Conversely, characterization of the origin of human fecal DNA is more complex, but several approaches are available. The most cost-effective and less invasive ones were initially developed for the analysis of cell-free DNA (cfDNA) in plasma. Indeed, plasma cfDNA sequencing and analysis has been proven to effectively detect tissue involvement in various processes in the body, even remote ones, most often of inflammatory nature [39–43]. One of the most recent approaches to the analysis of cfDNA composition is methylation profiling, for which we, and others, have compiled large cell methylation-marker atlases [44, 45].

In this study, we performed a combination of human and microbial fecal DNA profiling of IBD pediatric patients and controls from a newly established cohort, in order to assess a wide variety of IBD-associated features, such as tissue injury, inflammatory immune response,

and luminal microbiome composition. We profiled the microbial community structure together with total human DNA percentage and tissue of origin (using methylation-based profiling). As nearly all publicly-available microbiome cohorts do not include methylation profiles of the human DNA, we explored the generalizability of our findings by looking at the total human reads percentage as retrieved by metagenomic sequencing. For this purpose, we added to our analysis a large external microbiome cohort from the Netherlands ("Methods" section) to examine whether total human reads percentage could serve as a proxy for the more in-depth methylation profiling we performed on our cohort. Multifaceted genomic approaches, such as this one, have the potential to elucidate on novel aspects of host-microbe interaction in the context of chronic gut inflammation, and potentially solve unexplained variability that is observed in the clinical setting.

Methods

Patient enrollment and sample collection

Subjects (children, up to the age of 18 years old) were enrolled in the SZ cohort at the Pediatric Gastroenterology Institute at Shaare Zedek Medical Center, Jerusalem, Israel. Subjects were included in the study according to one of the following criteria: (i) subjects who were already diagnosed with IBD, undergoing a follow-up visit, (ii) subjects who received a new diagnosis of IBD by colonoscopy, (iii) subjects who underwent a colonoscopy for different indications, but had normal examination results, or (iv) subjects otherwise healthy (recruited for the purpose of the study at the orthopedic clinic at Shaare Zedek Medical Center). Fecal samples were collected in ethanol by the subjects and kept in a home freezer till the sample could be transferred (within 24 h) to the hospital and stored in a − 80 °C freezer. Samples included in the study amount to 134 in total (Control = 29, CD = 61, UC = 44), with 17 individuals sampled twice, and 8 individuals sampled three times.

For the majority of cases where IBD was unclassified (IBDU, *N* patients = 6/7, *N* samples = 10/11), patients were included in the analysis as UC for three reasons: (i) at a later time point the patient was re-diagnosed as having UC, (ii) the Shaare Zedek clinical team routinely assesses IBDU cases with Pediatric Ulcerative Colitis Activity Index (pUCAI), (iii) after performing a principal component analysis (PCoA) on the microbiome compositions of the samples (only Shaare Zedek samples, not shown) IBDU centroid overlapped with the UC centroid, confirming that at the population level, IBDU cases were indistinguishable from UC patients. Only one IBDU sample (*N* = 1/11, GB_543_1) was included in the analysis as CD given that at a later time point the subject was clinically re-diagnosed as a CD patient. Metadata pertaining to the cohort is provided in Supplementary Table 5.

DNA extraction

Samples were subjected to DNA extraction in several batches, due to low throughput of the DNA extraction protocol used. The protocol was optimized with custom modifications altering the use of the Power Soil Pro Kit (QIAGEN) protocol, in order to maximize DNA yield, fragment length, and microbial lysis [46]. The modifications increased the number of steps and volume of reagents required to isolate DNA from each sample, but importantly increased the yield and quality of the DNA required for the multiple sequencing procedures used in this study and more (e.g., Oxford Nanopore sequencing). Briefly, 200 mg of aliquoted feces were used as input to the DNA extraction protocol, characterized by three separate lysis steps. The three lysis steps include (i) a 10% SDS-based lysis, (ii) a ProteinaseK-based lysis, and (iii) a bead-beating-based lysis, with optimized power and time settings. In between lysis steps, high-speed centrifugation collects the unlysed pellet at the bottom, leaving the supernatant containing DNA to be collected in a separate tube. At the end of the lysis steps, three separate tubes containing DNA in a supernatant are processed separately with a column-based DNA isolation procedure, using the Power Soil Pro Kit, finally eluting the DNA in 70 µL of buffer. Performing only one DNA extraction, the protocol yields ~ 40 mg (median) of total DNA/sample, ensuring enough DNA for sequencing and other DNA-based procedures. DNA was quantified using dsDNA High Sensitivity kit (Denovix). Each DNA extraction batch included a no-sample control that was processed as a sample.

Methylation profiling of human DNA

Cell-specific methylation marker development

To be able to quantify the DNA fraction coming from small intestinal and colon cells in fecal DNA samples, cell type-specific methylation markers were developed as previously described [47, 48]. Thanks to comparative methylome analysis previously performed, using data from Illumina 450 k methylation arrays spanning multiple tissues, small intestine- and colon-specific markers were identified as regions with at least 5 CpG sites in a 150 bp long minimal window, with an average methylation value lower than 0.4 in the specific cell type targeted. Primers for the methylation markers were designed to amplify a ~ 100 bp window surrounding the informative CpGs, and the identified markers were validated in vitro by being tested against corresponding

cell/tissue type genomic DNA samples (SupplementaryFig. 2a), and in-silico against all previously established markers.

Sample processing and marker sequencing

To quantify the fraction of different cell types in patients' fecal DNA, DNA samples were treated with bisulfite using EZ DNA Methylation-Gold™ (Zymo Research), according to the manufacturer's instructions, and eluted in 20 µL. After bisulfite conversion, CpG bases that were originally methylated would be "CG", while unmethylated bases would be "TG." Bisulfite-treated DNA was then amplified in a two-step multiplex PCR reaction with primers specific to the cell type of interest. A first PCR reaction is run with short adapters and several combinations of primer sequences (30 pairs maximum), followed by an exonuclease step and a second PCR reaction using primers specific to the adapters. This final PCR reaction adds sequencing barcodes, hence the PCR products can be pooled together, run on a 3% agarose gel with ethidium bromide staining, and extracted by a Zymo gel recovery kit. Finally, the pooled PCR products were sequenced on a NextSeq sequencer with a read length of 150 bp, for a total depth of 10 K reads/sample.

Cell type fraction quantification

Sequenced reads were demultiplexed and DNA methylation patterns were aligned to the corresponding cell-specific methylation marker sequences developed for this study (small intestine and colon markers) as well as others previously identified [43, 44, 48]. Alignment was performed using Bismark [49] and a computational pipeline available on github (https://github.com/Joshmoss11/btseq). Reads were filtered out when having < 80% similarity to a marker sequence. Proper bisulfite conversion was assessed checking the expected CpG sites. Markers that were mapped by less than 1000 reads in a specific sample were excluded from downstream analyses and reported as missing. Finally, the fraction of tissue-specific DNA in the sample was calculated as the fraction of molecules in which all CpG sites were unmethylated. After calculating the sum of cell type fractions, a sample was considered successfully processed if the sum was either > 5% or < 150%. Samples exceeding 150% were re-amplified and sequenced while samples whose sum was < 5% were discarded. This led to the exclusion of five samples (Control = 1, CD = 2, UC = 2).

Absolute quantification of human DNA

To quantify the number of human DNA molecules, QX200™ EvaGreen ddPCR Supermix (BIORAD) was used according to manufacturer instructions. Primers were designed for the human SFPTC-1 gene and validated to yield comparable copy numbers to the human TERT gene, commercially used for the same purpose (TaqMan™ Copy Number Reference Assay). Each ddPCR assay was run on 5 ng of fecal DNA sample, a no-template control, and a positive control to validate the efficacy of the reactions. Manually set cutoff thresholds were used for each sample, according to acceptance criteria defined during the optimization of each reaction (QuantaSoft™ software version 1.7.4). Percentage of human DNA per ng of DNA was then inferred, assuming that each copy of human SFPTC-1 identified in stool represented one genome equivalent, or 3.3 picograms of human DNA. The per sample-percentage of human DNA was then used to normalize celltype-specific percentages.

The primers for the ddPCR reaction were:

SFTPC1: 5′-AGC AAA GAG GTC CTG ATG GAG A-3′ (forward), 5′-GCA GGG CCC ATC.

ACA CAC AT-3′ (reverse).

Shotgun metagenomic sequencing, preprocessing and taxonomic classification

Shotgun metagenomic sequencing libraries were prepared using Nextera DNA Library Prep kit by employing half of the reagents' volume and required input DNA. For each library preparation, a no-template control was included, together with all the no-sample controls from the corresponding DNA extraction batches. Both types of controls were processed as samples. Libraries were sequenced at a target depth of 8 M reads/sample with 150 bp single-end reads on a Miseq or Nextseq machine in six separate batches. Reads were quality filtered using fastq-mcf (-q 10 -l 75 –qual-mean 20; https://github.com/ExpressionAnalysis/ea-utils; v1.05), human reads were mapped against the human genome assembly (GRCh38.p13) by using bowtie2 [50] (v2.5.0) and samtools [51] (-f 4; v1.16.1), then quantified and removed. Non-human reads were taxonomically classified using MetaPhlAn4 [52] (database v. mpa_vJan21_CHOCOPhlAnSGB_202103). All analyses on microbiome composition were performed at the species-level (MetaPhlAn4 SGB IDs). A minority of the no-sample and no-template controls were positive to bacterial DNA classifiable by MetaPhlAn4. Contamination was assessed using *decontam* R package [53] (v1.13.0) by combining the *frequency* and *prevalence* methods accounting for the DNA extraction batch structure. Taxa classified as likely contaminants—that were found to be prevalent in the publicly available datasets used as external cohorts in this study—were not deemed to be contaminants from the kit reagents, but rather cross-contamination from the biological samples. We then checked whether sequencing batches could explain variability in the microbiome relative abundance profiles by Permanova using *adonis2()* function in the *vegan* R

package (v2.7–1). Given sequencing batches could significantly explain 8% of the variability, we proceeded to adjust the microbiome relative abundances for sequencing batches using the *adjust_batch()* function from the *MMUPHin* [54] R package (v1.20.0), using disease type as covariate.

Clinical information processing

Pediatric cohort

Clinical information for all subjects in the SZ cohort was obtained from RedCap. Pediatric-specific disease activity scores, such as Pediatric Ulcerative Colitis Activity Index (pUCAI) and Pediatric Crohn's Disease Activity Index (pCDAI), were used to assess the severity of symptoms. These are numerical scores that are categorized in levels by using standard conversion tables into *remission*, *mild*, *moderate*, and *severe*. For the purpose of this study, all categories other than *remission* were combined into the new category *active*. Disease location, as assessed by Paris classification (CD) and endoscopy (CD and UC), was converted into a categorical variable reading "ileum," "ileocolon," or "colon," with presence of "upper_GI" disease in a separate logical variable (Supplementary Table 5).

Adult cohort

For both 1000IBD and LLDeep samples, we were provided with some basic metadata about the subjects, including disease activity scores such as Harvey Bradshaw index (for CD) and the Simple Clinical Colitis Activity index (for UC). For harmonization purposes, these indices were first converted from numerical to categorical following standard tables, and then divided into *remission* and *active* as indicated above. Disease location indication was harmonized with what reported for the Pediatric cohort, hence it was converted from Montreal classification (CD and UC) to a categorical variable reading "ileum," "ileocolon," or "colon," with presence of "upper_GI" disease in a separate logical variable.

All cohorts

All samples had the total number of reads and the number of human reads quantified to calculate the percentage of human reads by metagenomics, used throughout. Fecal calprotectin values were capped at 2100 mgc/g whenever the value was above this threshold.

Statistical analyses

The statistical analyses (all performed in R) aimed at characterizing the differences between control and IBD subtypes and the associative relationship between microbiome composition and human DNA percentage.

Exploratory principal coordinate analysis (PCoA) was performed using robust Aitchison distance on unfiltered relative abundance data to account for compositionality, using *vegdist()* and *prcomp()* functions.

Species alpha diversity was assessed using the Shannon index as implemented in *diversity*(*index* = "*shannon*") within the *vegan* package (v2.7–1).

Correlation analyses were throughout performed using Spearman's rank correlation coefficients, testing the monotonic relationships between variables, using *stat_cor()* within the *ggpubr* package (v0.6.1). When specified, cohort-adjusted Spearman correlation coefficients were calculated in order to account for cohort-specific effects. These were computed by fitting a linear mixed model for each one of the correlative variables (fecal calprotectin or human reads %) using *cohort*, *disease activity level*, and *age* as explanatory variables and *subject* as a random variable. Then Spearman correlation coefficients were computed by correlating the two models' residuals with each other.

Differential statistical inference was performed throughout using non-parametric Mann–Whitney rank-sum tests, using *stat_compare_means*() within the *ggpubr* package (v0.6.1) or *wilcox_test*() by *rstatix* package (0.7.2). When specified, in order to calculate differences in species relative abundance across cohorts, differential means were calculated as least-square means (*emmeans* R package; v1.11.2–8) after fitting the following linear mixed model: $species\ asin(relative\ abundance) \sim human\ DNA\ percentage + disease\ type + age + cohort + (1|subject)$. From here we focused on species that had all least-square means > 0 across study groups and ranked them based on their respective least-square means ($N = 520$, Fig. 3a–c). We further classified these species in one of six categories (IBD/CD/UC-lost, IBD/CD/UC-expanded; bottom Fig. 3a), based on their regression coefficient ("*expanded*" when coefficient > 0, "*lost*" when coefficient < 0) and p value (p value ≤ 0.05), corrected for multiple hypotheses testing using Benjamini–Hochberg procedure. Species with p value > 0.05 were classified as "*no change*."

Linear mixed models

All models are reported in Supplementary Table 1.

We modeled fecal calprotectin by running *lmer()* function by the R package *lmerTest* (v3.1–3) as following: $FecalCalprotectin \sim human\ DNA\ percentage + disease\ activity\ level + treatment\ advancement_{\ only\ for\ Pediatric} + age + (1|subject)_{\ only\ for\ Pediatric}$. Samples having fecal calprotectin reaching 2100 mgc/g were excluded from the model, given these were artificially capped at this value.

Treatment advancement, as included in the model, was conceived as a numerical variable apt to capture the clinical treatment tier the patient had been treated with at the time of sampling, assigning numbers to the treatments categories, ordered by well-known stratified treatment regimes: 0 = "None," 1 = "ASA (oral and rectal), antibiotics or dietary supplement, 2 = "steroids,", 3 = "immunomodulators," 4 = "immunosuppressants or biologics."

After having identified neutrophils as the main source of human DNA, we remodeled fecal calprotectin only for the Pediatric cohort as following: *fecal calprotectin ~ neutrophils DNA percentage + disease activity level + treatment advancement + (1|subject) + (1|age).*

Finally, we modeled species count using *age* separately for each cohort (Pediatric and Adult) distinguishing by disease type and adjusting for multiple samples for the same subject, as follows: *species count ~ age * cohort + disease type + (1 | subject).*

XGBoost models

All XGBoost models built using cross-validation, for both relative abundance and presence-absence microbiome data (CV, Fig. 4a), were run using *xgb.cv()* function by *xgboost R* package (v1.7.11.1) on the Adult cohort using the following parameters (nfold = 5, stratified = TRUE, nrounds = 200, objective = "binary:logistic," eval_metric = "auc," eta = 0.05, gamma = 1,lambda = 3, nthread = 3, max_depth = 10, min_child_weight = 1, subsample = 0.8, colsample_bytree = 0.8, prediction = T). ROC curves were plotted using *ggroc()* function by *pROC* R package (v1.19.0.1).

Feature importance according to *Gain*, *Cover*, and *Frequency* for all models was extracted by rerunning each model at its best round with xgb.train() function and then using xgb.importance() function. Additionally, we performed SHAP analysis by using *SHAPforxgboost* R package (v0.1.3). First, we used *shap.values()* function to retrieve mean SHAP values, and then *shap.prep()* function for retrieving sample-specific values.

Validation of the CV models was run on the Pediatric cohort using *predict()* function and SHAP analysis was run as previously. Age and sex were excluded from the training variables in all models, given that the CD group in the Pediatric cohort is both younger and lower in male subjects (not shown).

Results

Fecal human DNA quantification by different methods correlates with standard metrics of inflammation

To explore the relationship between gut microbiome and human DNA levels in feces, we performed comprehensive microbial and human DNA profiling for a pediatric IBD cohort established at Shaare Zedek Medical Center in Jerusalem (Pediatric cohort, SZ). For this study, we sequenced 134 fecal samples from 101 children (median age = 15), including Crohn's disease patients (CD; 40 children), Ulcerative colitis patients (UC; 32 children), and control subjects not affected by gut inflammatory conditions (29 children; Fig. 1a; "Methods" section). CD and UC number of patients include unclassified IBD patients (IBDU; 7 children) that were included in the study as either CD or UC for later re-diagnosis or other reasons (see "Methods" section). Samples were collected together with detailed clinical information and pediatric-specific disease scores evaluating disease activity (pCDAI and pUCAI; "Methods" section). We combined shotgun metagenomic sequencing with two independent methods for human DNA-specific quantification and profiling (Fig. 1a): (i) a multiplex digital droplet PCR (ddPCR) to quantify the total human DNA content in feces [43], and (ii) a methylation-based marker approach for the quantification of several different human cell types and tissues in feces [44, 48] ("Methods" section). In addition, the results from the ddPCR assay were used as a normalization factor for the methylation-based method.

In order to contrast and compare our findings against larger, already published microbiome cohorts, we included the analysis of another published cohort of Dutch adults (Adult cohort), including the 1000IBD cohort [55] (IBD = 352) and the LLDeep-followup cohort [56] (Control = 337; Fig. 1a). By including the Adult cohort in our analyses, we could test whether the observations from the Pediatric cohort could generalize to the Adult cohort, while drawing the dynamic arc of microbial and human DNA composition in feces across the age spectrum. Moreover, we wanted to check whether our findings from the human DNA methylation profiling could be recapitulated by using human reads percentage by shotgun metagenomics, which comes at no additional cost and is available in most microbiome-specific sequencing studies.

We started with comparing the different human DNA quantification methods to evaluate whether they performed consistently across fecal samples, and found that human reads percentage by shotgun metagenomic sequencing was largely well-correlated to the human DNA percentage measured by ddPCR (Fig. 1b). Then, using pCDAI and pUCAI scores, we checked if disease activity had an impact on the amount of human DNA in feces, and found that, on average, human DNA was higher in active cases of both CD and UC compared to controls and remission IBD, with quite large variability among patients, especially in active UC cases (Fig. 1c, d). The fraction of patients with human DNA over 1% was the highest in active cases of both CD and UC (28% for CD and 47% in UC), modest in remission cases (15% in

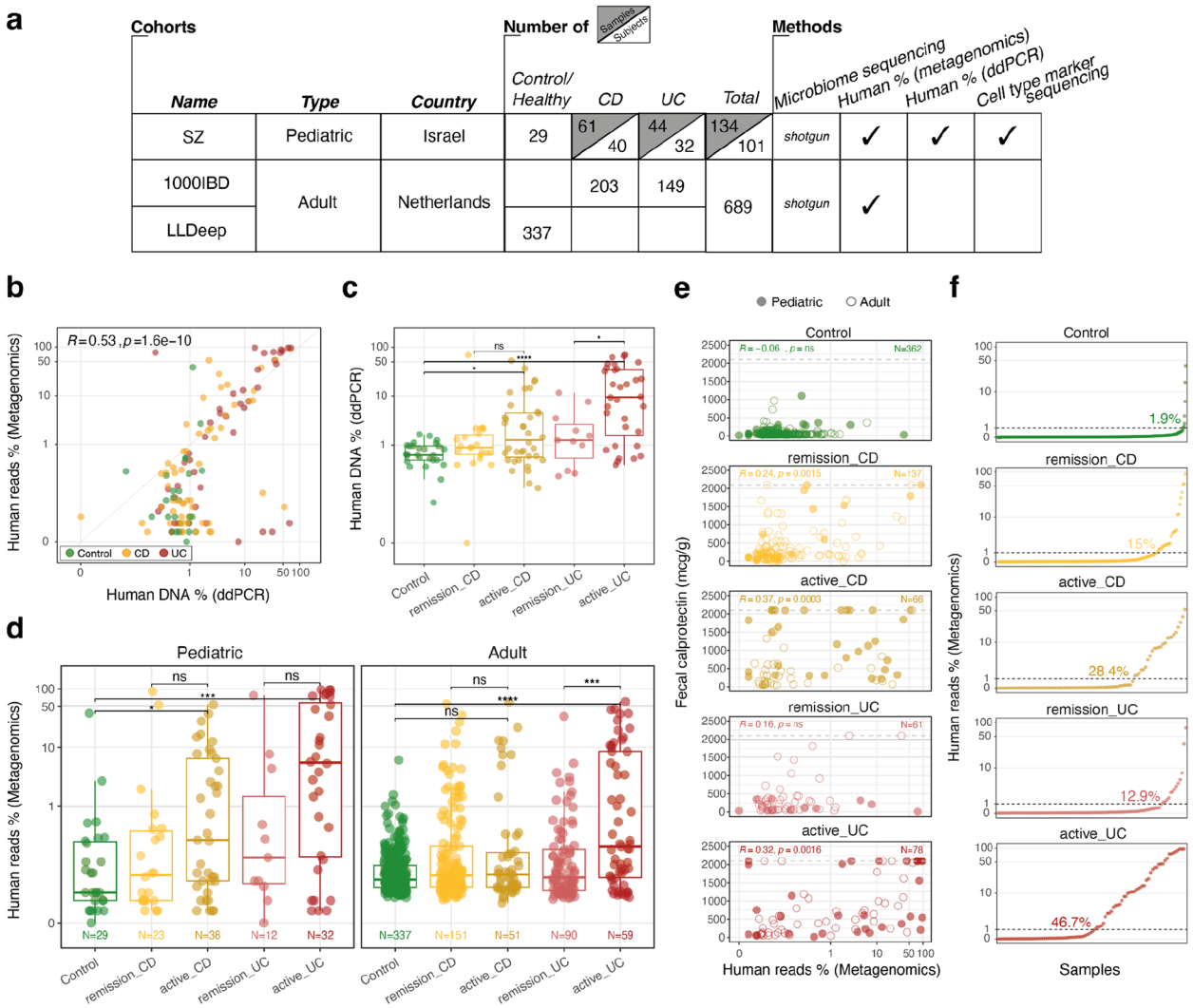

Fig. 1 Human DNA content in feces from IBD patients is a proxy for disease activity level. **a** Overview of cohorts, number of samples and subjects, and methods included in the study. **b** Human DNA percentage as calculated by two methods and their correlation (Spearman): percentage of human DNA as evaluated by ddPCR (Methods; x-axis) against percentage of shotgun metagenomic reads mapping to the human genome (GRCh38.p13 assembly; y-axis). **c** Human DNA percentage (ddPCR) across study groups. **d** Human reads percentage (Metagenomics) across study groups by cohort. **e** Cohort-adjusted correlation (Spearman, Methods) between human reads percentage (Metagenomics) and fecal calprotectin. **f** Fraction of samples having > 1% human reads percentage (Metagenomics) across disease activity levels (both Pediatric and Adult data). All comparisons are statistically significant at alpha = 0.05 by Mann–Whitney rank-sum test: $p < 0.05$: '*', $p < 0.01$: '**', $p < 0.001$: '***', $p <$ 1e-04: '****'

CD, 13% in UC) and negligible in controls (2%; Fig. 1f). Since our pediatric cohort included longitudinal sampling for some of the IBD patients, we also compared samples from the same patient, and checked whether human reads percentage specifically changed in parallel to a reported change in activity level of the disease (CD = 8, UC = 5; Supplementary Fig. 1). For all patients (except one) experiencing a transition from remissive to active IBD (or vice versa), human reads percentage in remission corresponded to either higher or equal levels of human reads percentage in flare, across both CD and UC cases. This also highlighted that human DNA content in some subjects stayed low, irrespective of the disease activity level.

As human DNA amount in feces correlated with disease activity, we hypothesized that human DNA could also be correlated with fecal calprotectin, a standard biomarker of intestinal inflammation [57]. While the two measures correlated better in active IBD cases compared to remissive cases (Fig. 1e), we proceeded to test whether human reads percentage by shotgun metagenomics could predict fecal calprotectin values across the different study groups. For this, we employed linear mixed models, one for each age cohort independently (Pediatric and

Adult), including disease activity and treatment advancement (when available), while adjusting for age and subject identity (for the longitudinal Pediatric cohort). In our models, human reads percentage and disease activity alone explained a fair amount of the variation in fecal calprotectin values (Pediatric $R^2 = 0.33$, Adult $R^2 = 0.29$), but when treatment advancement was included in the model for the Pediatric cohort ("Methods" section), the proportion of variance explained by the model increased significantly ($R^2 = 0.53$). In the model, treatment advancement was negatively associated with fecal calprotectin, as indicated by its negative coefficient estimate ($\beta = -159.04$, $p < < 0.05$, Supplementary Table 1; "Methods" section), highlighting how much clinical treatment based on biological therapy or immunosuppressants might influence immune response, specifically fecal calprotectin levels. Overall, human DNA percentage in feces could be a useful additional parameter to assess inflammatory status, although factors such as clinical treatment can determine large variance observed within and across the study groups.

Fecal DNA origin informs on IBD inflammation levels

As we found more fecal human DNA in active cases of IBD, we imagined that investigating the origin of human DNA in feces would elucidate how different tissues and cell populations are involved in the gut inflammatory process. For this purpose, we identified novel methylation-based markers for small intestinal and colon cells (see "Methods" section; Supplementary Table 2, SupplementaryFig. 2a) as previously developed for immune cells such as leukocytes (neutrophils and monocytes) and lymphocytes (B and T cells) [43] and used the entire marker set to estimate these cell percentages across fecal samples (Fig. 2a, "Methods" section). Among the cell populations profiled, neutrophils were the most abundant in feces of IBD patients, and they could differentiate not only controls from IBD, but also remissive from active cases of UC (Fig. 2a, b, SupplementaryFig. 2b). Neutrophil levels correlated almost perfectly with the total human DNA amount, especially where both measures were above 1%, reaching up to 50% of the total human DNA in some cases (Spearman $R = 0.88$, $p < < 0.001$ for all samples together; Fig. 2d). Neutrophil DNA percentage also correlated well with human reads percentage as inferred by metagenomic sequencing (overall Spearman $R = 0.64$, $p < < 0.001$; Supplementary Fig. 2c).

As calprotectin is the most abundant cytosolic protein in neutrophils and gets released into the lumen after cell burst [58, 59], we next wondered whether our measured neutrophil percentage could reliably predict fecal calprotectin levels, better than just human DNA percentage, as examined in the previous section. We evaluated this with a linear mixed model only for the Pediatric cohort, for which neutrophil DNA quantification was available ("Methods" section, Supplementary Table 2), and added similar clinical variables as before. Notably, this model clearly outperformed the previous model and explained the majority of the variance within the data ($R^2 = 0.68$). Conversely, when we flipped the model, and used fecal calprotectin measurements to predict neutrophil levels, our model explained almost all the variance within the data ($R^2 = 0.90$). While neutrophil DNA was highest in active IBD, colon DNA was either the first or second most abundant component in controls and remissive IBD, but it could not differentiate controls from IBD patients, since its values mainly ranged between 0 and 10% of total human DNA across all patients (Fig. 2c–e, Supplementary Fig. 2d).

To check whether we could detect complex dynamics involving several cell compartments in the gut, we tried to combine the different tissues and cell types measurements into one single metric. For this purpose, we first calculated neutrophil–lymphocyte ratio (NLR), as it has been previously suggested to be differentiating endoscopic activity in IBD patients when measured by cell counts from peripheral blood [60–62] (Fig. 2f). Next, we calculated neutrophil-epithelial ratio (NER), exploiting the unique capability of our approach to quantify epithelial cells (Fig. 2g). For both CD and UC, NER outperformed NLR in its ability to differentiate remissive from active IBD cases.

To make sure our measurement of neutrophil DNA in feces was not mainly driven by neutrophils presence in blood as a result of gut mucosal bleeding, we checked whether neutrophil DNA (or human reads percentage by shotgun metagenomics) correlated with clinical assessment of blood in feces (Supplementary Fig. 2e, f). The presence of blood in feces is accounted for in the pUCAI score for UC patients, and although this is a patient-reported metric, possibly lacking precision, we did not find blood in feces could explain the high levels of human DNA content in UC patients. Moreover, control subjects, who were recruited in the study because of blood presence in feces, did not exhibit higher levels of human DNA content compared to other controls who did not report blood in feces (all Mann–Whitney rank-sum test p values = ns, SupplementaryFig. 2 g, h). Indeed, mucosal bleeding does not always imply chronic mucosal inflammation, although we expect to observe at least some amount of mucosal bleeding whenever there is chronic mucosal inflammation. Although blood in feces could potentially contribute to some of the neutrophil DNA we measured, we suggest this is not a major driver, indicating gut mucosal storming of neutrophils is the largest contributor for neutrophil DNA content in feces.

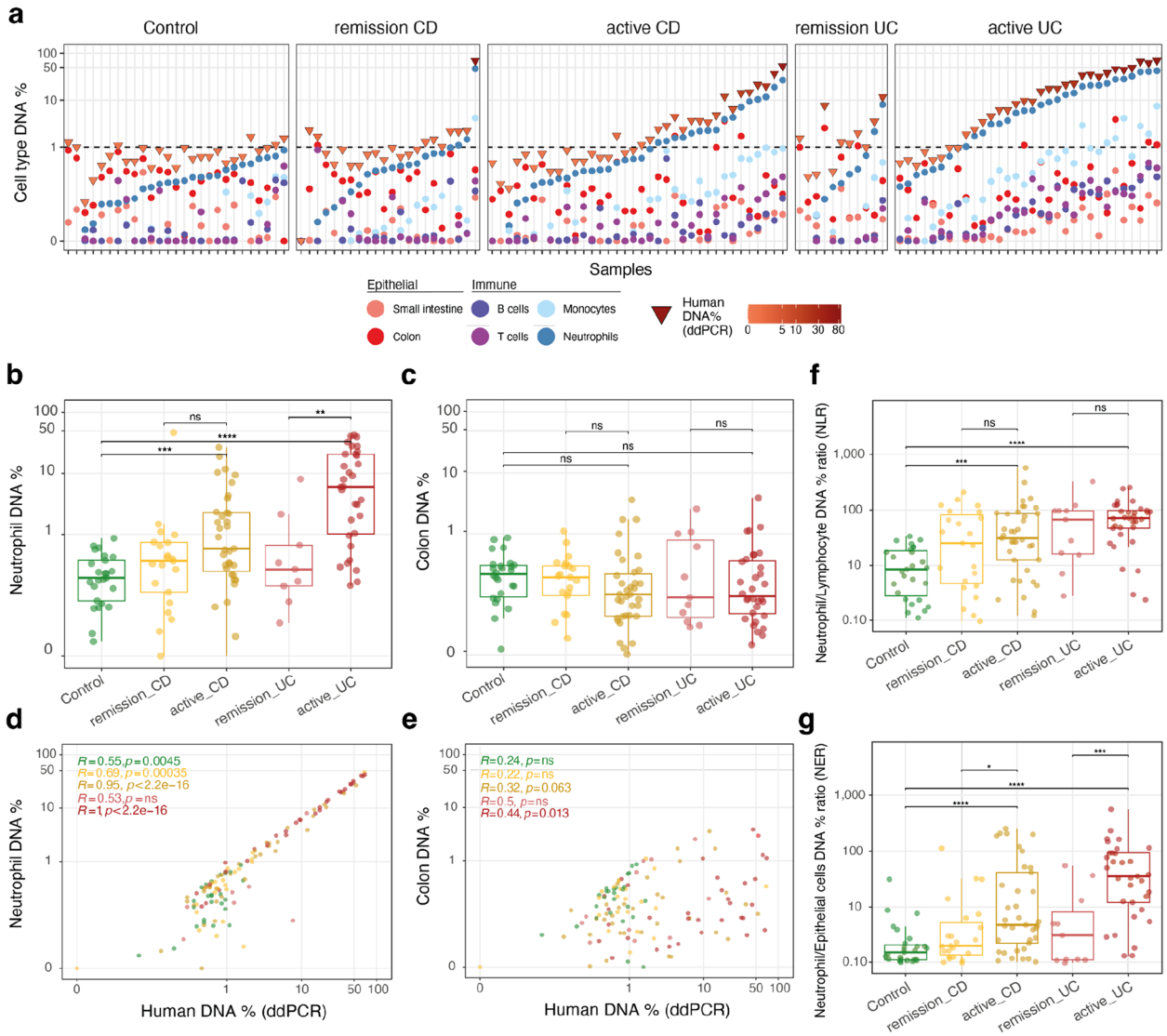

Fig. 2 Cell type profiling of feces from IBD patients reveals the majority of human DNA comes from neutrophils. **a** Methylation-based cell- and tissue-specific DNA percentage across samples in different disease activity groups. **b** Neutrophil DNA and **c** colon DNA percentage across study groups, **d** neutrophil DNA and **e** colon DNA percentage correlation (Spearman) with total human DNA percentage (ddPCR). **f** Neutrophil/Lymphocyte (B and T cells) DNA percentage ratio (NLR) from feces. **g** Neutrophil/Epithelial cells (small intestine and colon) DNA percentage ratio (NER) as an alternative to NLR. All comparisons are statistically significant at alpha = 0.05 by Mann–Whitney rank-sum test: $p < 0.05$: '*', $p < 0.01$: '**', $p < 0.001$: '***', $p <$ 1e-04: '****'

In summary, methylation-based profiling exposed neutrophil DNA dominance in the human DNA fraction in feces, and comprehensively allowed us to characterize several cell fractions in an integrative way. Evaluating the levels of different cell populations in the gut, as they are released in- or recruited to- the lumen, allowed us to distinguish the different study groups, along the inflammatory spectrum.

Species count drives sample variance in IBD and across age ranges

With respect to gut microbiome composition, it has been well-established that reduced microbial diversity is a hallmark of gut inflammation [63–65]. Here, we wanted to assess whether microbial diversity in IBD was correlated with human DNA levels as measured and described in the previous sections, and whether we could find similar trends in both pediatric and adult IBD patients. To characterize in detail the differences in microbial

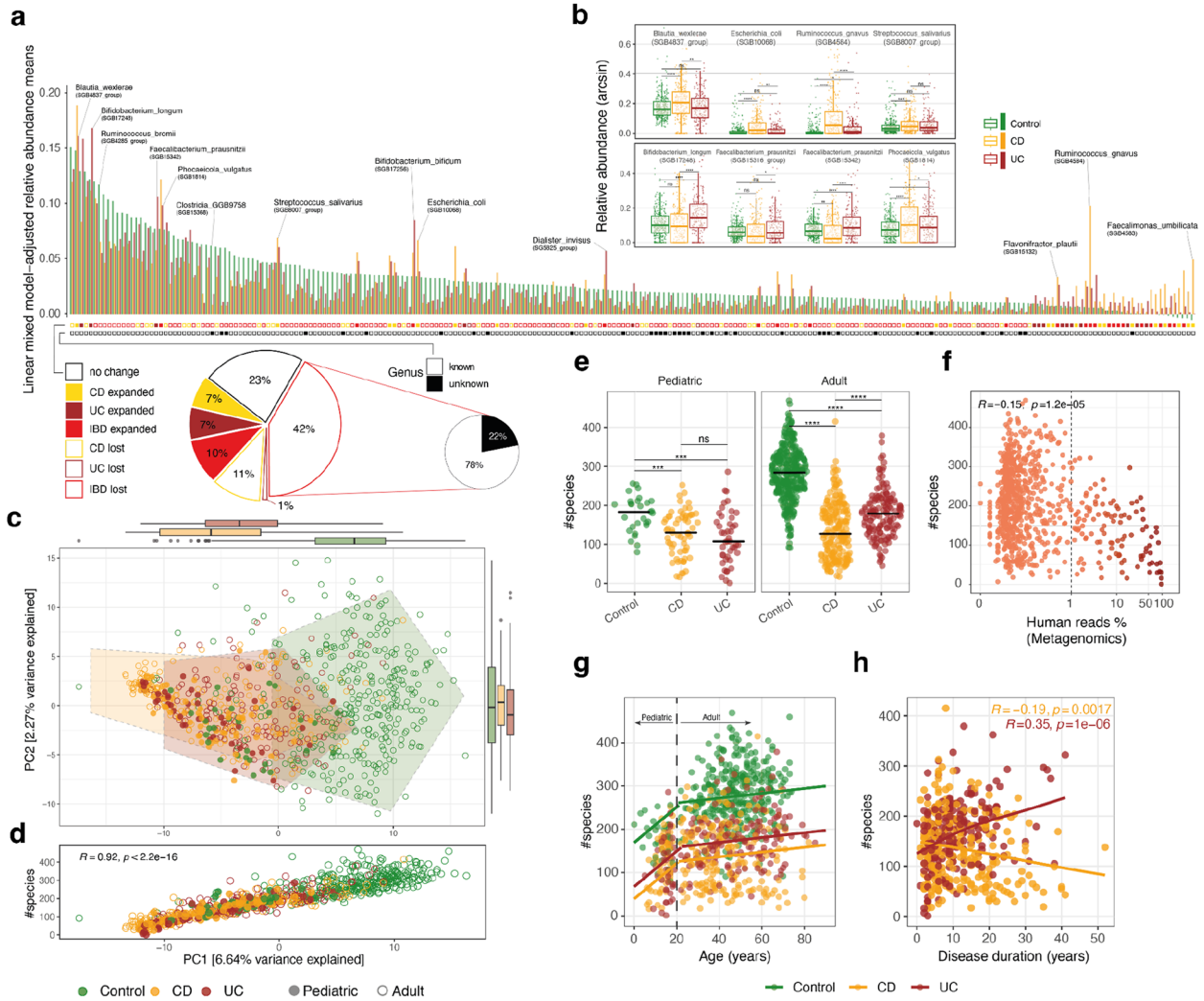

Fig. 3 Gut microbiome richness drives variance across IBD and Control subjects and differentially correlates with disease duration in CD and UC. **a** (top) Species least-square mean relative abundance ($N = 241/520$ shown; arcsine-transformed) by study group, taking into account human DNA percentage and age, and adjusting for cohort and subject (see Methods); several species of interest are named and highlighted; (bottom) breakdown of the 520 species in six different categories, based on their linear mixed model coefficients in CD, UC, or both (IBD) (Methods). **b** Arcsine-transformed relative abundance across study groups for species previously known to be increased in IBD patients (top) and species reported in this study to be increased in IBD patients (bottom; Benjamini–Hochberg corrected p values for cohort-adjusted least-square means (derived from the linear mixed model, Methods). **c** Principal coordinate analysis (PCoA) on species-level gut microbiome composition data (CLR-transformed relative abundance) as profiled by Metaphlan (v.4.1; Methods) with full points for the Pediatric cohort and empty points for the Adult cohort. **d** Species number correlation (Spearman) with the first principal component (PC1) of the PCoA. **e** Species number by study group across the Pediatric and Adult cohorts (Mann–Whitney rank-sum test). **f** Species number correlation (Spearman) with human reads percentage (Metagenomics; both Pediatric and Adult data). **g** Species number across ages and **h** over disease duration (both Pediatric and Adult data) with fitted regression lines (see Supplementary Table 1, *species_age.combined* and *species_disease_duration.combined* respectively) and Spearman correlation coefficients. All comparisons are statistically significant at alpha = 0.05 by Mann–Whitney rank-sum test: $p < 0.05$: '*' $p < 0.01$: '**' $p < 0.001$: '***' $p < 1e\text{-}04$: '****'

profiles between IBD patients and controls, we ran a linear mixed model to explain species relative abundance across groups taking into account human DNA percentage and age, adjusting for cohort differences and multiple-subject sampling. We then collected species-specific residuals and estimated least-square means, which we ordered by control values ($N = 520$, Fig. 3a, top arcsine-transformed, Supplementary Table 3). Then, based on the model's coefficients and p values, we categorized the species into six groups, based on whether they were positively (*expanded*) or negatively (*lost*) associated with both or either one of the IBD groups (Fig. 3a, bottom). For

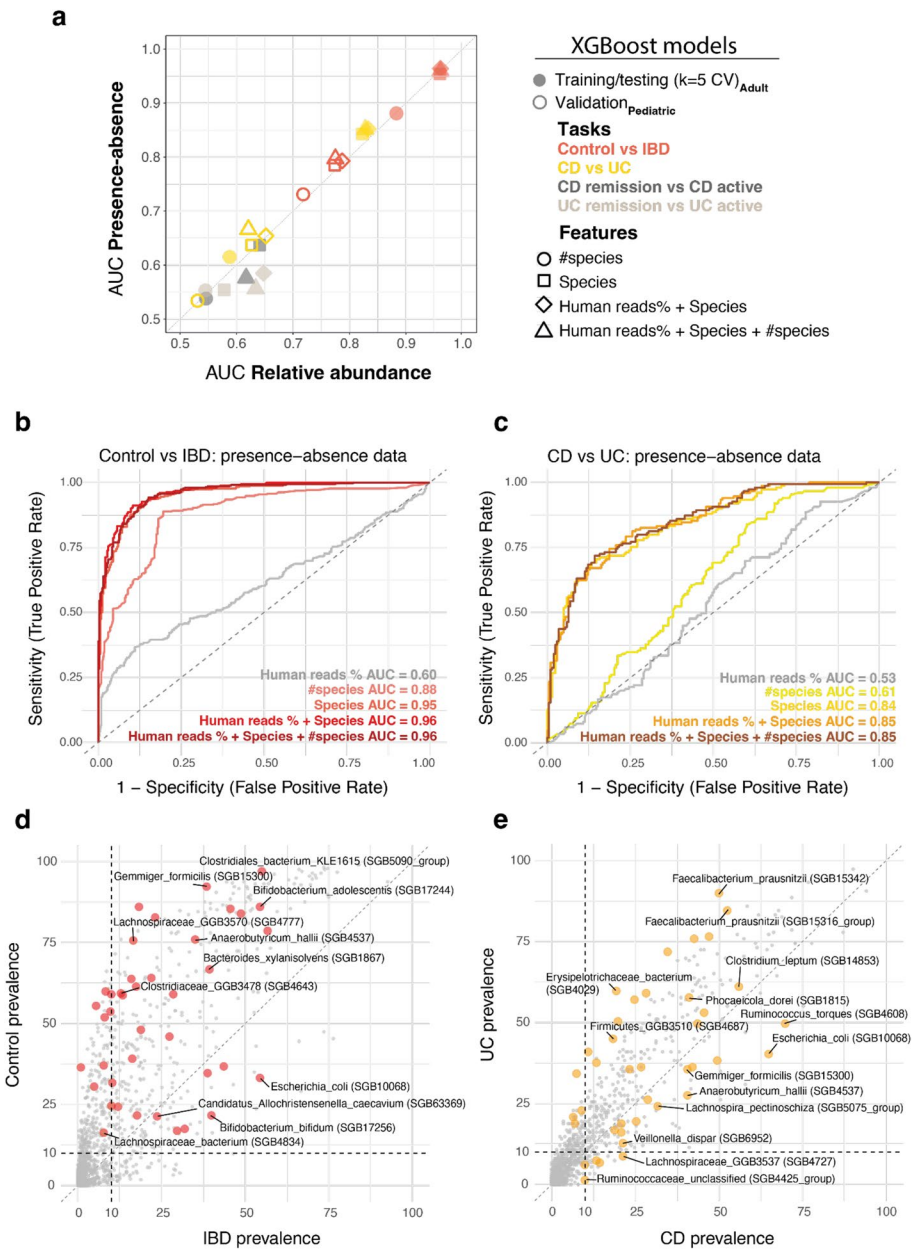

Fig. 4 XGBoost models trained on presence-absence species data and human reads percentage accurately predict adult IBD and moderately translate to pediatric patients. **a** Performance of XGBoost tree-based machine learning models (AUC values) when using relative-abundance data (x-axis) versus presence-absence data (y-axis). Each point represents a model trained on a specific combination of Features (shapes) and for different Tasks (colors). Performance comes from either fivefold Cross-Validation (full, on the Adult cohort) or external validation (empty, on the Pediatric cohort). **b** ROC curves of presence-absence data models, color coded by feature combination, predicting Control vs. IBD and **c.** CD vs. UC. **d** Species prevalence across IBD subjects (x-axis) vs. control subjects (y-axis) and **e** species prevalence across CD subjects (x-axis) vs. UC subjects (y-axis) in the Adult cohort only. The top twenty species with the highest importance across prediction models are colored

example, *Bifidobacterium longum* and *Faecalibacterium prausnitzii* (SGB15342) were preferentially expanded in UC, while *Blautia wexlerae* and *Streptococcus salivarius* were uniquely expanded in CD (Fig. 3b). Notably, the single species that had the highest adjusted mean relative abundance in controls and the highest relative drop in both CD and UC was *Ruminococcus bromii,* followed by *Candidatus Cibiobacter qucibialis* (SGB15286) (7th and 8th highest in control, respectively). Among the species with the lowest adjusted mean relative abundance in controls, we found species such as *Ruminococcus gnavus*, *Flavonifractor plautii*, and *Faecalimonas umbilicata*

(Fig. 3a, right). Species lost in either UC or CD accounted for 15% of the total (62 out of 520), while species lost in both CD and UC accounted for 42% of the species pool (216 out of 520). Notably, 22% of the species lost in both CD and UC were of unknown genera, to date solely identified computationally and described by metagenomic-derived gene markers [52] (GGB-SGBs). In particular, the highest ranking among them was Clostridia_GGB9758_SGB15368, located at the 36th place from the top of the distribution (Fig. 3a, middle).

To compare pediatric and adult IBD patients according to their gut microbiome composition, we then performed a principal coordinate analysis (PCoA, Aitchison distance on relative abundance data), to visualize sample clustering and dispersion. Visually, pediatric patients, adult patients, and control subjects formed a continuum, where most IBD patients clustered furthest away from control subjects (Fig. 3c). Permutational analysis of variance (PERMANOVA) of the gut microbiome composition confirmed that IBD patients were statistically different from controls (adonis $p < 0.001$), but also revealed that pediatric patients and adult patients could be distinguished from each other (adonis $p = 0.001$). This was not surprising, since the gut microbiome in childhood is still in development, and early IBD onset could disturb the natural process of microbiome maturation [66, 67]. To understand how much of the observed variation could be explained by microbial richness, we calculated the correlation between the first principal component in the PCoA (PC1) and species count, calculated as number of species with median relative abundance > 0% (Fig. 3d, also done with Shannon index, shown in Supplementary Fig. 3a). Notably, species count was better correlated with PC1 than Shannon index (Spearman R = 0.92 and R = 0.84, respectively), indicating that presence and absence of species might drive the majority of the variability across control and IBD gut microbiome samples. Indeed, when examining species count distributions, we found that species count in controls was roughly double the species count in IBD patients (overall: $\text{median}_{\text{Control}} = 274.5$, $\text{median}_{\text{CD}} = 128.5$, $\text{median}_{\text{UC}} = 168.0$; (Fig. 3e) irrespective of disease severity (Supplementary Fig. 3b), but the difference was more pronounced in adults than in children, also when looking at Shannon diversity (Fig. 3e, Supplementary Fig. 3c). Moreover, in pediatric IBD patients, species count differences between CD and UC were not statistically significant, possibly driven by the fact that children have a less mature microbiome. Interestingly, pediatric CD patients had similar species count to adult CD patients, unlike UC patients and control subjects (Supplementary Fig. 3d), which indicates that CD exerts a stronger stunting effect on the gut microbiome richness compared to UC, which persists in time. In parallel to the drop in species count, IBD patients were also characterized by increased human DNA amount (above 1%) corresponding to species count below 200 (Fig. 3f, Supplementary Fig. 3e). Species count was also negatively correlated to fecal calprotectin levels and NER, as calculated in the previous section (Supplementary Fig. 3f, g). Lastly, species count was found to be positively correlated with the total number of non-human reads in the samples (Supplementary Fig. 4a). To investigate whether the reduction in species count observed in IBD patients could be solely explained by differences in the number of non-human reads, we subsampled all samples to 1 Million non-human reads and recalculated the gut microbiome profile and species count (Supplementary Fig. 4b). The results confirmed that the number of non-human reads did not artificially inflate the species count difference between IBD patients and controls, highlighting that the high content of human DNA and reduced species richness are both biological features of the IBD-associated gut contents.

Finally, we looked at the role of age and disease duration in determining species count levels. We ran a linear mixed model to explain species count levels using age, but separately for each cohort and disease type (Fig. 3g, Supplementary Table 1—*species_age.combined*, "Methods" section). In accordance with the notion that children have a maturing gut microbiome [68], the linear mixed model confirmed a steeper increase of species count over time in pediatric subjects compared to adult subjects. Species count was stunted by IBD diagnosis, where CD resulted in a larger negative contribution than UC, suggesting that the gut microbiome in these cases is more severely impaired, as previously mentioned. When looking at species count since IBD diagnosis (Fig. 3h, Supplementary Table 1—*species_disease_duration.combined*), species count in UC had a positive correlation and linear mixed model coefficient with disease duration, while CD had negative ones, highlighting there could be a cumulative negative effect of inflammation on species richness over time, especially in CD patients. Overall, we assessed the spectrum of gut microbiome species changes between IBD patients and controls, highlighted known and unknown species that are either expanded or lost in IBD, and the power of species count in recapitulating Control-to-IBD gut microbiome sample variance across age ranges and disease duration.

Presence and absence of species predicts and characterizes IBD subtypes

We proceeded to examine whether we could use the microbiome data together with the human reads percentage to predict sample phenotypes. Specifically, we built

tree-based machine learning models using extreme gradient boosting (XGBoost) to either predict IBD, classify the IBD subtypes (CD or UC), or within these subtypes, distinguish between remissive versus active states of the disease. As input, we used species count (*#species*) together with percentage of human reads (*Human reads %*) and species composition data (*Species*; presence-absence or relative abundance; Fig. 4). We trained and tested the XGBoost models in five-fold cross validation (CV) on the Adult cohort, to exploit the sample size advantage. Then we validated the models' performance on the Pediatric cohort, despite knowing their performance would be suboptimal, given the overall differences between adult and pediatric subjects mentioned in the previous section. Yet, we deemed informative investigating how much the adult microbiome-based models would generalize to the pediatric microbiome samples. To our surprise, XGBoost models trained on either relative abundance or presence-absence data of the species had practically the same performance (Fig. 4a), with few exceptions for models in the Remissive*vs*Active task, which performed better using relative abundance data, but overall underperforming with AUCs < 0.66. Since the performance of models using only presence-absence data was as good as the one using relative abundance microbial composition, from here on we continued only with presence-absence models. Across CV and independent validation, species count alone had lower AUC values compared to the other more complex models including full species data, although its performance exceeded our expectations in the Control*vs*IBD task (AUC = 0.88; Fig. 4a). The models including more than species count alone performed comparably, if not better than previously reported IBD predictive models (AUC ≥ 0.95; Fig. 4b, c) [23, 69], and had very similar performances to each other, with little contribution by human reads percentage in both Control*vs*IBD and CD*vs*UC tasks. As a reference, models trained only on human reads percentage performed poorly with AUCs equal to 0.60 and 0.53 for Control*vs*IBD and CD*vs*UC respectively. However, human reads percentage was often among the most important features according to SHAP analysis values (Supplementary Fig. 5a–d), which highlights how human DNA content is most informative in combination with microbiome data. As expected, performance in the independent validation on pediatric microbiome samples was lower than in the CV, especially for the CD*vs*UC task (Control*vs*IBD AUC = 0.8, CD*vs*UC AUC = 0.68; both *Human reads %* + *Species* + *#species* models; Supplementary Fig. 5e,g). As lower predictive performance on the validation data was expected, we appreciated how the model could generalize for the Control*vs*IBD on such a different cohort, geographically and demographically.

To identify the features that were most discriminative in each task, we performed SHAP analysis across the different CV and validation models (Supplementary Fig. 5a–d,f,h). Not surprisingly, the most predictive features in the CV and validation models were found to have at least 10% prevalence across controls, IBD, CD, and UC alike (Adult cohort, Fig. 4d, e, Supplementary Table 4). To mention a few, *Bifidobacterium bifidum* and *Escherichia coli* were more prevalent in IBD cases than in Controls, while *Faecalibacterium prausnitzii* species (*SGB15342* and *SGB15316*) were more prevalent in UC compared to CD. Using these predictive models for IBD, we showed that species count alone is quite informative and correlates with the level of inflammation as assessed by human DNA content in feces. We appreciated how species presence-absence data was as predictive as relative abundance data, highlighting how, in IBD microbiome profiling accurate detection of species is more important that precise estimation of relative abundance.

Discussion

Fecal contents are considered a reflection of the physiological processes taking place along the gastrointestinal tract, on both the host and the gut microbiome side. On the host side, analyses of feces provide information on food consumption, absorption, and metabolism [70–72], as well as immune system status and response to transient or chronic infection [73]. On the microbiome side, microorganisms secrete metabolites that can influence a plethora of host-related processes, even beyond the gut, hence studying their composition and dynamics provides an important key to interpret their role in human physiology [74–83]. In this study, we combined analyses on both the host and gut microbiome side, in order to investigate their interaction in the context of chronic inflammatory bowel diseases (IBD).

On the host side, we applied methylation-based profiling of human fecal DNA in IBD patients and controls to explore the value of identifying human DNA tissue-of-origin in the characterization of the immune status of individuals. Genomics-based approaches, such as methylation-based human DNA profiling, can represent an alternative to traditional immunodetection methods for the characterization of cell-level and tissue-level processes in the body. For example, fecal calprotectin is released from dying neutrophil cells [58, 59], and it is commonly quantified as a non-specific biomarker of gut inflammation [84–86]. Here, we used the methylation-based approach to directly quantify neutrophil cell death, which yielded effective and sensitive measurements, especially in cases of high-inflammation background, such as severe IBD patients, where the standard

fecal calprotectin quantification method is limited by its detection range.

Our methylation-based profiles indeed confirmed neutrophil storming to the gut mucosa and lumen to be a dominant process during inflammation [87, 88], which aligns with fecal calprotectin values, as we showed with linear regression analysis. However, we cannot exclude that part of the neutrophil DNA signal we measured could possibly come from neutrophils present in the blood released in the feces, as part of mucosal bleeding. Moreover, we highlighted that neutrophil DNA prevails over colon DNA in feces, despite the fact that feces are formed along the colon and rectum and cell epithelial shedding is a known response to gut inflammation [89, 90]. Nevertheless, we cannot exclude that other processes, such as DNA degradation in the gut environment during inflammation, might specifically reduce epithelial DNA concentration in feces. In the absence of any element for physical protection, epithelial DNA could be degraded [91], or even consumed by microorganisms scavenging for nutrients [92, 93]. Conversely, neutrophil DNA can be particularly well shielded from degradation, given that it is extruded from neutrophils together with other proteins, creating antimicrobial web-like structures called neutrophil extracellular traps (NETs) [94, 95]. Our methylation-based profiling could be expanded to more cell types of the gut epithelium, opening avenues for a more rapid and cost-effective alternative to single-cell approaches such as cell sorting [96–100]. Methylation-based biomarkers could be developed, not only for less abundant cells, such as goblet and paneth cells, but also for specific functional immune subtypes that uniquely expand in case of inflammation, such as aged neutrophils [101], inflammatory macrophages [102], myeloid-derived immunosuppressive cells or exhausted T cells [103–108].

On the microbiome side, we investigated how species richness relates to disease severity, human fecal DNA level, and disease course over time, highlighting species that expand or drop in IBD. Among the species we highlighted as expanded in IBD, we identified previously reported taxa, such as *E. coli*, *R. gnavus*, *P. vulgatus*, and *F. plautii* [17, 23, 34, 109]. Additionally, we found *B. longum* and *B. bifidum* as expanded and more prevalent in UC patients, respectively. Both *B. longum* and *B. bifidum* are among the most well-established pioneer species of the healthy infant gut microbiome [110–113], and as such we do not expect them to be implicated in the pathogenesis of IBD. We mentioned *R. gnavus* which is both an IBD-associated species [114, 115] and a known infant gut colonizer, with recently reported infant-specific clades [116, 117]. These observations prompt us to hypothesize that higher abundance and prevalence in IBD do not necessarily associate with strictly pathogenic traits. The presence of these species might, however, be explained by their great capacity of metabolizing glycans, HMOs in the infant gut [118–121], and mucin-derived glycans in the inflamed gut [122–124]. Indeed, in the context of gut inflammation, where survival conditions are at the hardest and species are in a race for survival, species presence correlates strongly with their metabolic capacity to endure challenging conditions [125, 126], and although some studies have already suggested this, there is the need to expand research in this direction.

Finally, in predicting IBD, we showed presence-absence of species to be as predictive as relative abundance data, in line with other reports [127]. This should encourage the use of presence-absence data in microbiome analysis over relative abundance data, especially in cases where relative abundance is less reliable. For instance, relative abundance can be misleading in disease contexts where microbial load drastically changes across samples, due to frequent bowel movements causing DNA content dilution [128]. For the same reason, several studies have suggested the use of absolute abundance over relative abundance [129–132], which can be estimated by using metagenomic host read fraction, as reported recently [133]. Moreover, differential analysis of relative abundance has been reported to yield very different results based on which method was used to perform it, hence clearer consensus on the most appropriate methods to use would be also needed to minimize reporting spurious results [134].

As shown in the validation of our XGBoost models, training on adult microbiome data only partially generalized to the pediatric test data, performing moderately well in the ControlvsIBD task (best AUC = 0.8), while doing poorly in the CDvsUC task (best AUC = 0.68; Fig. 4a, Supplementary Fig. 4 g–h). This highlights how different the adult IBD microbiome is compared to the pediatric IBD microbiome and suggests the need for developing pediatric-specific diagnostic tools in the context of larger pediatric IBD studies, including both CD and UC patients. Ultimately, this study approaches the important aspect of non-invasive, microbiome-based diagnostics [135], which other studies have also recently addressed, with various predictive learning approaches [69, 136, 137].

To date, most microbiome studies on IBD have investigated pediatric and adult IBD cohorts separately, without much emphasis on characterizing the IBD-associated microbiome as it relates to age [34, 109, 138]. Although we understand the value of keeping the age of the target study groups as narrow as possible, here we decided to incorporate a large set of external microbiome samples to cover the entire age spectrum and to be able to compare the pediatric to the adult IBD-associated gut microbiome.

This age-inclusive approach gave us the opportunity to identify overarching characteristics of IBD, such as changes in species richness across ages, which have been largely overlooked by other microbiome studies, but are an important topic of discussion in the clinical setting [139]. Our comparative analysis was accompanied by a range of challenges and limitations, such as (i) small number of control pediatric subjects; (ii) heterogeneity of the patients with respect to clinical variables [140]; (iii) differential availability of metadata, including medications, detailed disease location, IBD family history, and disease activity scores; and finally (iv) disregard to viral gut microbiome diversity [141–145].

On the technical side, our small cohort of DNA samples was processed in multiple DNA extraction and sequencing runs, due to the low throughput of our optimized DNA extraction protocol. Batch effects in the microbiome and human DNA profiles were analyzed and accounted for, but we cannot exclude that they might have played a role in the results of our analyses. Despite the limitations, our analyses addressed some important aspects related to IBD progression over time, such as the interaction between species richness and disease duration, which are likely to be generalizable to other cohorts.

Our work highlights human DNA contents in feces as informative and predictive of IBD, especially when combined with gut microbiome profiling. When using microbiome shotgun metagenomic sequencing, our approach makes use of human DNA information that is otherwise disregarded, yet we imagine that a human DNA-specific PCR assay on fecal DNA could extend existing microbiome-specific approaches to diagnose IBD [69]. These comprehensive analytical approaches can help identify how the gut microbiome responds and contributes to gut inflammation, and how it can be treated or modified to ameliorate symptoms and reduce life-long complications in IBD patients [146, 147].

Conclusions

This study integrates comprehensive microbiome profiling with detailed human DNA quantification and cell-type characterization, providing a dual perspective on host-microbiome interactions in the context of inflammatory bowel diseases (IBD). We employ a methylation-based approach to settle previous speculations regarding the origin of human DNA in feces during gut inflammation, and we combine it with shotgun metagenomic sequencing to identify overarching gut microbiome characteristics across pediatric and adult IBD patients. We showed that human DNA content in feces primarily comes from neutrophils and correlates with IBD activity level and fecal calprotectin measurements. We described the expansion and loss of IBD-associated bacterial species, but at the same time we highlighted how presence-absence microbiome data alone can predict IBD as well as relative abundance data. We believe the findings of this study could have applications in the non-invasive monitoring of patients' inflammatory status, and that our methylation-based profiling could be further tailored to experimentally investigate the expansion of more recently discovered cellular subpopulations during inflammation.

Abbreviations

IBD	Inflammatory bowel disease
CD	Crohn's disease
UC	Ulcerative colitis
NLR	Neutrophil/Lymphocyte ratio
NER	Neutrophil/Epithelial ratio
XGBoost	Extreme gradient boosting
NET	Neutrophil extracellular trap

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-026-02344-6.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank the participants for joining the research, Dr Abed Nasereddin and Dr Idit Shiff from the Genomics Applications Laboratory at the Faculty of Medicine of the Hebrew University for their support with DNA sequencing. We thank Dr Ranko Gacesa and Prof. Dr. R.K. Weersma at the University Medical Center Groningen for the support provided in metadata retrieval and curation for the 1000IBD and LLDeep cohorts.

Authors' contributions

MY, YD, RS and ES conceived the project. CM and MY designed the computational analyses. CM performed the computational and statistical analyses and generated the figures, MY supervised the computational analyses. CM, SS and EH performed DNA extraction and sequencing library preparation. BO performed the ddPCR assays and methylation-based profiling of the samples. ES, EOM and ABY provided the patients' samples. CM, GF and EOM curated the patients' clinical information. CM and MY wrote the manuscript. MY, YD, RS and ES co-supervised the project.

Funding

This research was funded by a grant from the Helmsley Charitable Trust (to YD, ES and MY). Research in YD lab is supported by grants from the Israel Science Foundation, Israel's Ministry of Science and Technology (MOST), Israel's Innovation authority, the Waldholtz / Pakula family, the Robert M. and Marilyn Sternberg Family Charitable Foundation, Horizon Europe (PANCAID consortium) and NCI (2U01CA210171-06). MY holds the Rosalind, Paul and Robin Berlin Faculty Development Chair in Perinatal Research. YD holds the Walter and Greta Stiel Chair and Research grant in Heart studies.

Data availability

The human-filtered metagenomic sequencing data for the Pediatric cohort generated in this study is deposited in the SRA database under BioProject PRJNA1265906. Metadata pertaining to the Pediatric cohort is provided in **Supplementary Table 5**. The use of the Adult cohort (LLDeep and 1000IBD cohorts) in this study respects the form agreed with Lifelines and the UMCG Department of Genetics. Metagenomic sequencing data and basic metadata for the LLDeep cohort was provided under a Data Access Agreement with the UMCG Department of Genetics and access to the data was granted through the following EGA Dataset Accession Number: EGAD00001006959. Metagenomic sequencing data and clinical metadata for the 1000IBD cohort was

provided under a Data Transfer Agreement with the University Medical Center Groningen and Prof. Dr. R.K. Weersma, and access was granted to metadata at EGAD00001003991 and metagenomic sequencing data at EGAD00001004194.

Declarations

Ethics approval and consent to participate

The experimental protocol was approved by the Shaare Zedek Medical Center Institutional Review Board (SZMC-21–0165). Informed consent was provided by the legal guardian and assent by adolescents older than 16 years old.

Competing interests

The authors declare no competing interests.

Author details

[1]Microbiology and Molecular Genetics Department, Institute for Medical Research Israel-Canada, Hebrew University of Jerusalem, Jerusalem, Israel. [2]Developmental Biology and Cancer Research Department, Institute for Medical Research Israel-Canada, Hebrew University of Jerusalem, Jerusalem, Israel. [3]Juliet Keidan Institute of Pediatric Gastroenterology Hepatology and Nutrition, Shaare Zedek Medical Center, The Eisenberg R&D Authority, Jerusalem, Israel. [4]Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel. [5]The Rachel and Selim Benin, School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem, Israel.

Received: 3 June 2025 Accepted: 7 January 2026
Published: 25 February 2026

References

1. Kai Y. Intestinal villus structure contributes to even shedding of epithelial cells. Biophys J. 2021Feb 16;120(4):699–710.
2. Creamer B, Shorter RG, Bamforth J. The turnover and shedding of epithelial cells. I. The turnover in the gastro-intestinal tract. Gut. 1961;2(2):110–8.
3. Jiang Y, Chen J, Du Y, Fan M, Shen L. Immune modulation for the patterns of epithelial cell death in inflammatory bowel disease. Int Immunopharmacol. 2025May;8(154):114462.
4. Sender R, Milo R. The distribution of cellular turnover in the human body. Nat Med. 2021Jan 11;27(1):45–8.
5. Tomofuji Y, Sonehara K, Kishikawa T, Maeda Y, Ogawa K, Kawabata S, et al. Reconstruction of the personal information from human genome reads in gut metagenome sequencing data. Nat Microbiol. 2023May 15;8(6):1079–94.
6. Vincent C, Mehrotra S, Loo VG, Dewar K, Manges AR. Excretion of host DNA in feces is associated with risk of Clostridium difficile infection. J Immunol Res. 2015May;18(2015):246203.
7. The Human Microbiome Project Consortium. A framework for human microbiome research. Nature. 2012Jun 13;486(7402):215–21.
8. Bahar Halpern K, Korem Kohanim Y, Biram A, Harnik Y, Egozi A, Yakubovsky O, et al. The cellular states and fates of shed intestinal cells. Nat Metab. 2023;5(11):1858–69.
9. Luster AD, Alon R, von Andrian UH. Immune cell migration in inflammation: present and future therapeutic targets. Nat Immunol. 2005Nov 17;6(12):1182–90.
10. Casellas F, Antolín M, Varela E, García-Lafuente A, Guarner F, Borruel N, et al. Fecal excretion of human deoxyribonucleic acid as an index of inflammatory activity in ulcerative colitis. Clin Gastroenterol Hepatol. 2004;2(8):683–9.
11. Jiang P, Lai S, Wu S, Zhao XM, Chen WH. Host DNA contents in fecal metagenomics as a biomarker for intestinal diseases and effective treatment. BMC Genomics. 2020May 11;21(1):1–11.
12. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome. 2018Feb 27;6(1):42.
13. Marquet M, Zöllkau J, Pastuschek J, Viehweger A, Schleußner E, Makarewicz O, et al. Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. Sci Rep. 2022Mar 7;12(1):1–10.
14. Ahannach S, Delanghe L, Spacova I, Wittouck S, Van Beeck W, De Boeck I, et al. Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples. iScience. 2021;24(11):103306.
15. Huttenhower C, Kostic AD, Xavier RJ. Inflammatory bowel disease as a model for translating the microbiome. Immunity. 2014 Jun 19 [cited 2025 May 10];40(6). Available from: https://pubmed.ncbi.nlm.nih.gov/24950204/
16. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012Sep 26;13(9):1–18.
17. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol. 2018Jan 8;3(3):337–46.
18. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014Mar 12;15(3):382–92.
19. Sokol H, Lay C, Seksik P, Tannock GW. Analysis of bacterial bowel communities of IBD patients: What has it revealed? Inflamm Bowel Dis. 2008Feb 14;14(6):858–67.
20. Fischbach MA. A meta-analysis of the gut microbiome in inflammatory bowel disease patients identifies disease-associated small molecules. Cell Host Microbe. 2025Feb 12;33(2):218-34.e12.
21. Mehta RS, Mayers JR, Zhang Y, Bhosle A, Glasser NR, Nguyen LH, et al. Gut microbial metabolism of 5-ASA diminishes its clinical efficacy in inflammatory bowel disease. Nat Med. 2023Feb 23;29(3):700–9.
22. Zorea A, Pellow D, Levin L, Pilosof S, Friedman J, Shamir R, et al. Plasmids in the human gut reveal neutral dispersal and recombination that is overpowered by inflammatory diseases. Nat Commun. 2024 Apr 11 [cited 2025 May 14];15(1). Available from: https://pubmed.ncbi.nlm.nih.gov/38605009/
23. Vila AV, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. Sci Translat Med. 2018 Dec 19 [cited 2025 May 14]; Available from: https://www.science.org/doi/https://doi.org/10.1126/scitranslmed.aap8914
24. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017Feb 13;2(5):1–7.
25. Lee JWJ, Plichta D, Hogstrom L, Borren NZ, Lau H, Gregory SM, et al. Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. Cell Host Microbe. 2021Aug 11;29(8):1294-304.e4.
26. Carasso S, Zaatry R, Hajjo H, Kadosh-Kariti D, Ben-Assa N, Naddaf R, et al. Inflammation and bacteriophages affect DNA inversion states and functionality of the gut microbiota. Cell Host Microbe. 2024Mar 13;32(3):322-34.e9.
27. Federici S, Kredo-Russo S, Valdés-Mas R, Kviatcovsky D, Weinstock E, Matiuhin Y, et al. Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. Cell. 2022Aug 4;185(16):2879-98.e24.
28. Solitano V, Bernstein CN, Dotan I, Dignass A, Domilici R, Dubinsky MC, et al. Shaping the future of inflammatory bowel disease: a global research agenda for better management and public health response. Nat Rev Gastroenterol Hepatol. 2025Apr;22:1–15.
29. Sartor RB. Beyond random fecal microbial transplants: next generation personalized approaches to normalize dysbiotic microbiota for treating IBD. Gastroenterology clinics of North America. 2025 Jun [cited 2025 May 14];54(2). Available from: https://pubmed.ncbi.nlm.nih.gov/40348491/
30. Nagayama M, Gogokhia L, Longman RS. Precision microbiota therapy for IBD: premise and promise. Gut microbes. 2025 Dec [cited 2025 May 14];17(1). Available from: https://pubmed.ncbi.nlm.nih.gov/40190259/
31. Schirmer M, Denson L, Vlamakis H, Franzosa EA, Thomas S, Gotman NM, et al. Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. Cell Host Microbe. 2018Oct 10;24(4):600-10.e4.
32. Tamburini FB, Tripathi A, Gold MP, Yang JC, Biancalani T, McBride JM, et al. Gut microbial species and endotypes associate with remission in

ulcerative colitis patients treated with anti-TNF or anti-integrin therapy. J Crohns Colitis. 2024Nov 4;18(11):1819–31.
33. Caenepeel C, Falony G, Machiels K, Verstockt B, Goncalves PJ, Ferrante M, et al. Dysbiosis and associated stool features improve prediction of response to biological therapy in inflammatory bowel disease. Gastroenterology. 2024Mar 1;166(3):483–95.
34. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655–62.
35. Hong SM, Baek DH. Diagnostic procedures for inflammatory bowel disease: laboratory, endoscopy, pathology, imaging, and beyond. Diagnostics. 2024Jun 28;14(13):1384.
36. Gallinger Z, Ungaro R, Colombel JF, Sandler RS, Chen W. P030 delayed diagnosis of Crohn's disease is common and associated with an increased risk of disease complications. Inflamm Bowel Dis. 2019;25(Supplement_1):S14–5.
37. Vieujean S, Jairath V, Peyrin-Biroulet L, Dubinsky M, Iacucci M, Magro F, et al. Understanding the therapeutic toolkit for inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2025Jan;31:1–24.
38. Duan R, Mei W, Lei M, Chen D, Pan T, Kong F, et al. Care needs profiles of Crohn's disease patients and their associations with symptom clusters, post-traumatic growth, and family function: a latent profile analysis. BMC gastroenterology. 2025 May 9 [cited 2025 May 14];25(1). Available from: https://pubmed.ncbi.nlm.nih.gov/40346452/
39. De Vlaminck I, Valantine HA, Snyder TM, Strehl C, Cohen G, Luikart H, et al. Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. Sci Transl Med. 2014;6(241):241ra77.
40. Lubotzky A, Zemmour H, Neiman D, Gotkine M, Loyfer N, Piyanzin S, et al. Liquid biopsy reveals collateral tissue damage in cancer. JCI Insight. 2022. https://doi.org/10.1172/jci.insight.153559.
41. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell. 2016Jan 14;164(1–2):57–68.
42. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019May 29;570(7761):385–9.
43. Fox-Fisher I, Piyanzin S, Ochana BL, Klochendler A, Magenheim J, Peretz A, et al. Remote immune processes revealed by immune-derived circulating cell-free DNA. Elife. 2021Nov;29(10):e70520.
44. Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, et al. A DNA methylation atlas of normal human cell types. Nature. 2023;613(7943):355–64.
45. Zhu T, Liu J, Beck S, Pan S, Capper D, Lechner M, et al. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. Nat Methods. 2022Mar 11;19(3):296–306.
46. Mazzoni C. Modified three peaks protocol - power soil DNA isolation kit v1. 2025. Available from: https://www.protocols.io/view/modified-three-peaks-protocol-power-soil-dna-isola-haqsb2dwf
47. Neiman D, Gillis D, Piyanzin S, Cohen D, Fridlich O, Moss J, et al. Multiplexing DNA methylation markers to detect circulating cell-free DNA derived from human pancreatic β cells. JCI Insight. 2020. https://doi.org/10.1172/jci.insight.136579.
48. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun. 2018Nov 29;9(1):5068.
49. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011Apr 14;27(11):1571.
50. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009Mar 4;10(3):1–10.
51. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021. https://doi.org/10.1093/gigascience/giab008.
52. Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol. 2023;41(11):1633–44.
53. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018Dec 17;6(1):226.
54. Ma S, Shungin D, Mallick H, Schirmer M, Nguyen LH, Kolde R, et al. Population structure discovery in meta-analyzed microbial communities and inflammatory bowel disease using MMUPHin. Genome Biol. 2022Oct 3;23(1):1–31.
55. Imhann F, Van der Velde KJ, Barbieri R, Alberts R, Voskuil MD, Vich Vila A, et al. The 1000IBD project: multi-omics data of 1000 inflammatory bowel disease patients; data release 1. BMC Gastroenterol. 2019Jan 8;19(1):1–10.
56. Website. Available from: https://doi.org/10.1136/bmjopen-2014-006772
57. Lundberg JO, Hellström PM, Fagerhol MK, Weitzberg E, Roseth AG. Technology insight: calprotectin, lactoferrin and nitric oxide as novel markers of inflammatory bowel disease. Nat Clin Pract Gastroenterol Hepatol. 2005;2(2):96–102.
58. Røseth AG, Fagerhol MK, Aadland E, Schjønsby H. Assessment of the neutrophil dominating protein calprotectin in feces. A methodologic study. Scand J Gastroenterol. 1992;27(9):793–8.
59. Teng TS, Ji AL, Ji XY, Li YZ. Neutrophils and Immunity: From Bactericidal Action to Being Conquered. J Immunol Res. 2017Jan 1;2017(1):9671604.
60. Celikbilek M, Dogan S, Ozbakır O, Zararsız G, Kücük H, Gürsoy S, et al. Neutrophil-lymphocyte ratio as a predictor of disease severity in ulcerative colitis: neutrophil-to-lymphocyte ratio and Ülcerative colitis. J Clin Lab Anal. 2013Jan;27(1):72–6.
61. Fu W, Fu H, Ye W, Han Y, Liu X, Zhu S, et al. Peripheral blood neutrophil-to-lymphocyte ratio in inflammatory bowel disease and disease activity: a meta-analysis. Int Immunopharmacol. 2021Dec;101(Pt B):108235.
62. Langley BO, Guedry SE, Goldenberg JZ, Hanes DA, Beardsley JA, Ryan JJ. Inflammatory bowel disease and neutrophil-lymphocyte ratio: a systematic scoping review. J Clin Med. 2021Sep 17;10(18):4219.
63. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun. 2020Oct 15;11(1):5206.
64. Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017;15(10):630–8.
65. Lamont RJ, Hajishengallis G. Polymicrobial synergy and dysbiosis in inflammatory disease. Trends in molecular medicine. 2015 Mar [cited 2025 May 14];21(3). Available from: https://pubmed.ncbi.nlm.nih.gov/25498392/
66. Davis EC, Monaco CL, Insel R, Järvinen KM. Gut microbiome in the first 1000 days and risk for childhood food allergy. Ann Allergy Asthma Immunol. 2024;133(3):252–61.
67. Laue HE, Willis AD, Wang F, MacDougall MC, Xu Y, Karagas MR, et al. Early-life and concurrent predictors of the healthy adolescent microbiome in a cohort study. Genome med. 2025 May 8 [cited 2025 May 14];17(1). Available from: https://pubmed.ncbi.nlm.nih.gov/40340756/
68. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. The human microbiome and child growth - first 1000 days and beyond. Trends Microbiol. 2019;27(2):131–47.
69. Zheng J, Sun Q, Zhang M, Liu C, Su Q, Zhang L, et al. Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. Nat Med. 2024Oct 4;30(12):3555–67.
70. Diener C, Holscher HD, Filek K, Corbin KD, Moissl-Eichinger C, Gibbons SM. Metagenomic estimation of dietary intake from human stool. Nat Metab. 2025Feb;18:1–14.
71. Sillner N, Walker A, Lucio M, Maier TV, Bazanella M, Rychlik M, et al. Longitudinal profiles of dietary and microbial metabolites in formula- and breastfed infants. Front Mol Biosci. 2021May;28(8):660456.
72. Lagoumintzis G, Patrinos GP. Triangulating nutrigenomics, metabolomics and microbiomics toward personalized nutrition and healthy living. Hum Genomics. 2023Dec 8;17(1):109.
73. Wu M, Zheng W, Song X, Bao B, Wang Y, Ramanan D, et al. Gut complement induced by the microbiota combats pathogens and spares commensals. Cell. 2024Feb 15;187(4):897-913.e18.
74. Nie Q, Luo X, Wang K, Ding Y, Jia S, Zhao Q, et al. Gut symbionts alleviate MASH through a secondary bile acid biosynthetic pathway. Cell. 2024 May 23 [cited 2025 May 14];187(11). Available from: https://pubmed.ncbi.nlm.nih.gov/38653239/
75. De Vadder F, Kovatcheva-Datchary P, Goncalves D, Vinera J, Zitoun C, Duchampt A, et al. Microbiota-generated metabolites promote

metabolic benefits via gut-brain neural circuits. Cell. 2014 Jan 16 [cited 2025 May 14];156(1–2). Available from: https://pubmed.ncbi.nlm.nih.gov/24412651/
76. Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. Cell. 2016 May 5 [cited 2025 May 14];165(4). Available from: https://pubmed.ncbi.nlm.nih.gov/27133167/
77. Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun. 2017 Oct 10 [cited 2025 May 14];8(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29018189/
78. Hu H, Shao W, Liu Q, Liu N, Wang Q, Xu J, et al. Gut microbiota promotes cholesterol gallstone formation by modulating bile acid composition and biliary cholesterol secretion. Nat Commun. 2022 Jan 11 [cited 2025 May 14];13(1). Available from: https://pubmed.ncbi.nlm.nih.gov/35017486/
79. Tian P, Yang W, Guo X, Wang T, Tan S, Sun R, et al. Early life gut microbiota sustains liver-resident natural killer cells maturation via the butyrate-IL-18 axis. Nat Commun. 2023 Mar 27 [cited 2025 May 14];14(1). Available from: https://pubmed.ncbi.nlm.nih.gov/36973277/
80. Park EM, Chelvanambi M, Bhutiani N, Kroemer G, Zitvogel L, Wargo JA. Targeting the gut and tumor microbiota in cancer. Nat Med. 2022Apr 19;28(4):690–703.
81. Benakis C, Brea D, Caballero S, Faraco G, Moore J, Murphy M, et al. Commensal microbiota affects ischemic stroke outcome by regulating intestinal γδ T cells. Nat Med. 2016 May [cited 2025 May 14];22(5). Available from: https://pubmed.ncbi.nlm.nih.gov/27019327/
82. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nat Med. 2016 Oct [cited 2025 May 14];22(10). Available from: https://pubmed.ncbi.nlm.nih.gov/27711063/
83. Wu G, Xu T, Zhao N, Lam YY, Ding X, Wei D, et al. A core microbiome signature as an indicator of health. Cell. 2024Nov 14;187(23):6550-65.e11.
84. Bunn SK, Bisset WM, Main MJ, Golden BE. Fecal calprotectin as a measure of disease activity in childhood inflammatory bowel disease. J Pediatr Gastroenterol Nutr. 2001;32(2):171–7.
85. Ágreda FM, Origüen J, Rodriguez-Goncer I, San JR, López-Medrano F, Parra P, et al. Predictive value of fecal calprotectin and lactoferrin levels for negative outcomes in Clostridioides difficile infection. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2024 Feb [cited 2025 May 10];43(2). Available from: https://pubmed.ncbi.nlm.nih.gov/38072880/
86. Shastri YM, Bergis D, Povse N, Schäfer V, Shastri S, Weindel M, et al. Prospective multicenter study evaluating fecal calprotectin in adult acute bacterial diarrhea. Am J Med. 2008;121(12):1099–106.
87. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. Nat Rev Immunol. 2013;13(3):159–75.
88. Danne C, Skerniskyte J, Marteyn B, Sokol H. Neutrophils: from IBD to the gut microbiota. Nat Rev Gastroenterol Hepatol. 2023Dec 18;21(3):184–97.
89. Martínez-Sánchez LDC, Ngo PA, Pradhan R, Becker LS, Boehringer D, Soteriou D, et al. Epithelial RAC1-dependent cytoskeleton dynamics controls cell mechanics, cell shedding and barrier integrity in intestinal inflammation. Gut. 2023 Feb [cited 2025 May 14];72(2). Available from: https://pubmed.ncbi.nlm.nih.gov/35241625/
90. Liebing E, Krug SM, Neurath MF, Siegmund B, Becker C. Wall of resilience: how the intestinal epithelium prevents inflammatory onslaught in the gut. Cell Mol Gastroenterol Hepatol. 2025Jan 1;19(2):101423.
91. Maturin L Sr, Curtiss R III. Degradation of DNA by nucleases in intestinal tract of rats. Science. 1977 Apr 8 [cited 2025 May 10]; Available from: https://www.science.org/doi/https://doi.org/10.1126/science.322286
92. Wasmund K, Pelikan C, Schintlmeister A, Wagner M, Watzka M, Richter A, et al. Genomic insights into diverse bacterial taxa that degrade extracellular DNA in marine sediments. Nat Microbiol. 2021Jun 14;6(7):885–98.
93. Anderson CJ, Medina CB, Barron BJ, Karvelyte L, Aaes TL, Lambertz I, et al. Microbes exploit death-induced nutrient release by gut epithelial cells. Nature. 2021Aug 4;596(7871):262–7.
94. Neuenfeldt F, Schumacher JC, Grieshaber-Bouyer R, Habicht J, Schröder-Braunstein J, Gauss A, et al. Inflammation induces pro-NETotic neutrophils via TNFR2 signaling. Cell Rep. 2022Apr 19;39(3):110710.
95. Thiam HR, Wong SL, Wagner DD, Waterman CM. Cellular Mechanisms of NETosis. Ann rev cell develop biol. 2020 Oct 6 [cited 2025 May 14];36. Available from: https://pubmed.ncbi.nlm.nih.gov/32663035/
96. Li J, Simmons AJ, Hawkins CV, Chiron S, Ramirez-Solano MA, Tasneem N, et al. Identification and multimodal characterization of a specialized epithelial cell type associated with Crohn's disease. Nat Commun. 2024Aug 22;15(1):1–19.
97. Jaeger N, Gamini R, Cella M, Schettini JL, Bugatti M, Zhao S, et al. Single-cell analyses of Crohn's disease tissues reveal intestinal intraepithelial T cells heterogeneity and altered subset distributions. Nat Commun. 2021Mar 26;12(1):1–12.
98. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell. 2019 Jul 25 [cited 2025 May 14];178(3). Available from: https://pubmed.ncbi.nlm.nih.gov/31348891/
99. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell survey of the small intestinal epithelium. Nature. 2017 Nov 16 [cited 2025 May 14];551(7680). Available from: https://pubmed.ncbi.nlm.nih.gov/29144463/
100. Mitsialis V, Wall S, Liu P, Ordovas-Montanes J, Parmet T, Vukovic M, et al. Single-cell analyses of colon and blood reveal distinct immune cell signatures of ulcerative colitis and Crohn's disease. Gastroenterology. 2020Aug 1;159(2):591-608.e10.
101. Zhang D, Chen G, Manwani D, Mortha A, Xu C, Faith JJ, et al. Neutrophil ageing is regulated by the microbiome. Nature. 2015Sep 16;525(7570):528–32.
102. Hegarty LM, Jones GR, Bain CC. Macrophages in intestinal homeostasis and inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2023Apr 17;20(8):538–53.
103. Garrido-Trigo A, Corraliza AM, Veny M, Dotti I, Melón-Ardanaz E, Rill A, et al. Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. Nat Commun. 2023Jul 26;14(1):4506.
104. Youngblood B, Oestreich KJ, Ha SJ, Duraiswamy J, Akondy RS, West EE, et al. Chronic virus infection enforces demethylation of the locus that encodes PD-1 in antigen-specific CD8(+) T cells. Immunity. 2011Sep 23;35(3):400–12.
105. Sen DR, Kaminski J, Barnitz RA, Kurachi M, Gerdemann U, Yates KB, et al. The epigenetic landscape of T cell exhaustion. Science. 2016Dec 2;354(6316):1165–9.
106. Hollen MK, Stortz JA, Darden D, Dirain ML, Nacionales DC, Hawkins RB, et al. Myeloid-derived suppressor cell function and epigenetic expression evolves over time after surgical sepsis. Crit Care. 2019Nov 13;23(1):355.
107. Gabrilovich DI, Nagaraj S. Myeloid-derived suppressor cells as regulators of the immune system. Nat Rev Immunol. 2009;9(3):162–74.
108. Ashkenazi-Preiser H, Reuven O, Uzan-Yulzari A, Komisarov S, Cirkin R, Turjeman S, et al. The cross-talk between intestinal microbiota and mdscs fuels colitis-associated cancer development. Cancer Research Communications. 2024Apr 15;4(4):1063–81.
109. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019;4(2):293–305.
110. Akiyama S, Nishijima S, Kojima Y, Kimura M, Ohsugi M, Ueki K, et al. Multi-biome analysis identifies distinct gut microbial signatures and their crosstalk in ulcerative colitis and Crohn's disease. Nat Commun. 2024Nov 27;15(1):1–14.
111. Ennis D, Shmorak S, Jantscher-Krenn E, Yassour M. Longitudinal quantification of Bifidobacterium longum subsp. infantis reveals late colonization in the infant gut independent of maternal milk HMO composition. Nature Communications. 2024 Jan 30;15(1):1–12.
112. Pasolli E, Mauriello IE, Avagliano M, Cavaliere S, De Filippis F, Ercolini D. Bifidobacteriaceae diversity in the human microbiome from a large-scale genome-wide analysis. Cell Rep. 2024Dec 24;43(12):115027.
113. Henrick BM, Rodriguez L, Lakshmikanth T, Pou C, Henckel E, Arzoomand A, et al. Bifidobacteria-mediated immune system imprinting early in life. Cell. 2021Jul 22;184(15):3884-98.e11.

114. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Med. 2017 Nov 28 [cited 2025 May 10];9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29183332/
115. Zhai L, Huang C, Ning Z, Zhang Y, Zhuang M, Yang W, et al. Ruminococcus gnavus plays a pathogenic role in diarrhea-predominant irritable bowel syndrome by increasing serotonin biosynthesis. Cell Host Microbe. 2023Jan 11;31(1):33-44.e5.
116. Lou YC, Rubin BE, Schoelmerich MC, DiMarco KS, Borges AL, Rovinsky R, et al. Infant microbiome cultivation and metagenomic analysis reveal Bifidobacterium 2'-fucosyllactose utilization can be facilitated by coexisting species. Nat Commun. 2023Nov 16;14(1):1–14.
117. Nooij S, Plomp N, Sanders IMJG, Schout L, van der Meulen AE, Terveer EM, et al. Metagenomic global survey and in-depth genomic analyses of Ruminococcus gnavus reveal differences across host lifestyle and health status. Nat Commun. 2025Jan 30;16(1):1–14.
118. Fehr K, Moossavi S, Sbihi H, Boutin RCT, Bode L, Robertson B, et al. Breastmilk feeding practices are associated with the co-occurrence of bacteria in mothers' milk and the infant gut: the CHILD cohort study. Cell host & microbe. 2020 Aug 12 [cited 2025 May 14];28(2). Available from: https://pubmed.ncbi.nlm.nih.gov/32652062/
119. Kijner S, Kolodny O, Yassour M. Human milk oligosaccharides and the infant gut microbiome from an eco-evolutionary perspective. Curr Opinion Microbiol. 2022 Aug [cited 2025 May 14];68. Available from: https://pubmed.ncbi.nlm.nih.gov/35598464/
120. Kiely LJ, Busca K, Lane JA, van Sinderen D, Hickey RM. Molecular strategies for the utilisation of human milk oligosaccharides by infant gut-associated bacteria. FEMS microbiology reviews. 2023 Nov 1 [cited 2025 May 14];47(6). Available from: https://pubmed.ncbi.nlm.nih.gov/37793834/
121. Ho NT, Li F, Lee-Sarwar KA, Tun HM, Brown BP, Pannaraj PS, et al. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. Nat Commun. 2018 Oct 9 [cited 2025 May 14];9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/30301893/
122. Ducarmon QR, Karcher N, Tytgat HLP, Delannoy-Bruno O, Pekel S, Springer F, et al. Large-scale computational analyses of gut microbial CAZyme repertoires enabled by Cayman. bioRxiv. 2024 [cited 2025 May 10]. p. 2024.01.08.574624. Available from: https://www.biorxiv.org/content/https://doi.org/10.1101/2024.01.08.574624v1.abstract
123. Katoh T, Ojima MN, Sakanaka M, Ashida H, Gotoh A, Katayama T. Enzymatic adaptation of Bifidobacterium bifidum to host glycans, viewed from glycoside hydrolyases and carbohydrate-binding modules. Microorganisms. 2020Mar 28;8(4):481.
124. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell. 2016Nov 17;167(5):1339-53.e21.
125. Veseli I, Chen YT, Schechter MS, Vanni C, Fogarty EC, Watson AR, et al. Microbes with higher metabolic independence are enriched in human gut microbiomes under stress. eLife. 2024 Nov 20 [cited 2025 May 9];12. Available from: https://elifesciences.org/reviewed-preprints/89862
126. Watson AR, Füssel J, Veseli I, DeLongchamp JZ, Silva M, Trigodet F, et al. Metabolic independence drives gut microbial colonization and resilience in health and disease. Genome Biol. 2023Apr 17;24(1):1–21.
127. Giliberti R, Cavaliere S, Mauriello IE, Ercolini D, Pasolli E. Host phenotype classification from human microbiome data is mainly driven by the presence of microbial taxa. PLoS Comput Biol. 2022Apr 21;18(4):e1010066.
128. Tito RY, Verbandt S, Aguirre Vazquez M, Lahti L, Verspecht C, Lloréns-Rico V, et al. Microbiome confounders and quantitative profiling challenge predicted microbial targets in colorectal cancer development. Nat Med. 2024Apr 30;30(5):1339–48.
129. Maghini DG, Dvorak M, Dahlen A, Roos M, Doyle B, Kuersten S, et al. Quantifying bias introduced by sample collection in relative and absolute microbiome measurements. Nat Biotechnol. 2024 Feb [cited 2025 May 10];42(2). Available from: https://pubmed.ncbi.nlm.nih.gov/37106038/
130. Barlow JT, Bogatyrev SR, Ismagilov RF. A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. Nat Commun. 2020May 22;11(1):1–13.
131. Vandeputte D, Kathagen G, D'hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature. 2017 Nov 1;551(7681):507–11.
132. Nishijima S, Stankevic E, Aasmets O, Schmidt TSB, Nagata N, Keller MI, et al. Fecal microbial load is a major determinant of gut microbiome variation and a confounder for disease associations. Cell. 2025Jan 9;188(1):222-36.e15.
133. Tang G, Carr AV, Perez C, Ramos Sarmiento K, Levy L, Lampe JW, et al. Metagenomic estimation of absolute bacterial biomass in the mammalian gut through host-derived read normalization. mSystems. 2025 Aug 19;10(8):e0098425.
134. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. Nat Commun. 2022Jan 17;13(1):1–16.
135. Yzet C, Meudjo E, Brazier F, Hautefeuille V, Moreau C, Robert C, et al. Intestinal ultrasound, fecal calprotectin, and their combination to predict endoscopic mucosal healing in ulcerative colitis: a real-life cross-sectional study. Inflamm Bowel Dis. 2024Jul 18;31(5):1231–6.
136. Chang D, Gupta VK, Hur B, Cobo-López S, Cunningham KY, Han NS, et al. Gut Microbiome Wellness Index 2 enhances health status prediction from gut microbiome taxonomic profiles. Nat Commun. 2024Aug 28;15(1):1–14.
137. Ha SM, Lee K, Kim GH, Hurych J, Cinek O, Shim JO. Gut-microbiota-based ensemble model predicts prognosis of pediatric inflammatory bowel disease. iScience. 2024 Nov 22 [cited 2025 May 11];27(12). Available from: https://pubmed.ncbi.nlm.nih.gov/39691780/
138. Agrawal M, Allin KH, Petralia F, Colombel JF, Jess T. Multiomics to elucidate inflammatory bowel disease risk factors and pathways. Nat Rev Gastroenterol Hepatol. 2022Mar 17;19(6):399–409.
139. Ruel J, Ruane D, Mehandru S, Gower-Rousseau C, Colombel JF. IBD across the age spectrum—is it the same disease? Nat Rev Gastroenterol Hepatol. 2013Dec 17;11(2):88–98.
140. Graham DB, Xavier RJ. Pathway paradigms revealed from the genetics of inflammatory bowel disease. Nature. 2020Feb 26;578(7796):527–39.
141. Adiliaghdam F, Amatullah H, Digumarthi S, Saunders TL, Rahman RU, Wong LP, et al. Human enteric viruses autonomously shape inflammatory bowel disease phenotype through divergent innate immunomodulation. Sci Immunol. 2022 Apr 8 [cited 2025 May 11]; Available from: https://www.science.org/doi/https://doi.org/10.1126/sciimmunol.abn6660
142. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell. 2015Jan 29;160(3):447–60.
143. Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. Cell Host Microbe. 2019Dec 11;26(6):764-78.e5.
144. Tian X, Li S, Wang C, Zhang Y, Feng X, Yan Q, et al. Gut virome-wide association analysis identifies cross-population viral signatures for inflammatory bowel disease. Microbiome. 2024Jul 18;12(1):1–20.
145. Ungaro F, Massimino L, Furfaro F, Rimoldi V, Peyrin-Biroulet L, D'Alessio S, et al. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. Gut Microbes. 2018Sep 25;10(2):149.
146. Peyrin-Biroulet L, Loftus EV Jr, Colombel JF, Sandborn WJ. Long-term complications, extraintestinal manifestations, and mortality in adult Crohn's disease in population-based cohorts. Inflamm Bowel Dis. 2011;17(1):471–8.
147. Manser CN, Borovicka J, Seibold F, Vavricka SR, Lakatos PL, Fried M, et al. Risk factors for complications in patients with ulcerative colitis. United Eur Gastroenterol J. 2016;4(2):281–7.

Publisher's Note