GUT MICROBIOTA

Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome

Arnau Vich Vila[1,2]*, Floris Imhann[1,2]*, Valerie Collij[1,2]*, Soesma A. Jankipersadsing[2†], Thomas Gurry[3,4†], Zlatan Mujagic[5†], Alexander Kurilshikov[2†], Marc Jan Bonder[2], Xiaofang Jiang[3,4], Ettje F. Tigchelaar[2], Jackie Dekens[2], Vera Peters[1], Michiel D. Voskuil[1,2], Marijn C. Visschedijk[1,2], Hendrik M. van Dullemen[1], Daniel Keszthelyi[5], Morris A. Swertz[2], Lude Franke[2], Rudi Alberts[1,2], Eleonora A. M. Festen[1,2], Gerard Dijkstra[1], Ad A. M. Masclee[5], Marten H. Hofker[6‡], Ramnik J. Xavier[3,4], Eric J. Alm[3,4], Jingyuan Fu[2,6], Cisca Wijmenga[2§], Daisy M. A. E. Jonkers[5§], Alexandra Zhernakova[2§], Rinse K. Weersma[1§‖]

Changes in the gut microbiota have been associated with two of the most common gastrointestinal diseases, inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS). Here, we performed a case-control analysis using shotgun metagenomic sequencing of stool samples from 1792 individuals with IBD and IBS compared with control individuals in the general population. Despite substantial overlap between the gut microbiome of patients with IBD and IBS compared with control individuals, we were able to use gut microbiota composition differences to distinguish patients with IBD from those with IBS. By combining species-level profiles and strain-level profiles with bacterial growth rates, metabolic functions, antibiotic resistance, and virulence factor analyses, we identified key bacterial species that may be involved in two common gastrointestinal diseases.

INTRODUCTION

Inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS) are two of the most common gastrointestinal (GI) disorders, affecting 0.3 to 0.5% and 7 to 21% of the worldwide population, respectively. Both disorders impose a large burden on patients, impairing their quality of life as well as their ability to work and function socially (*1*, *2*). In addition, the economic burden of these disorders in the United States and Europe exceeds 10 billion dollars a year in direct health care costs and indirect economic costs (*2*, *3*).

IBD, comprising Crohn's disease (CD) and ulcerative colitis (UC), is a chronic intermittent disorder characterized by intestinal inflammation. IBS is defined as a combination of GI symptoms, including abdominal pain, constipation, or diarrhea (*4*). Patients with IBD and IBS may have similar symptoms, but whereas the pathogenesis of IBD consists of mucosal inflammation, the pathogenesis of IBS remains poorly understood, and there is no causative anatomical or biochemical abnormality that can be used to diagnose IBS (*2*).

The gut microbiota is presumed to play a large role in both IBD and IBS (*5*, *6*). However, thus far, large-scale gut microbiome sequencing associated with IBD and IBS compared with healthy controls has only been accomplished using low-resolution 16*S* ribosomal RNA (rRNA) marker gene sequencing (*7*–*9*). Functional studies have so far only focused on single bacterial species or strains in the gut. Here, we aimed to bridge the gap between previous 16*S* rRNA sequencing studies and functional studies by identifying complete gut microbiome profiles using high-resolution shotgun metagenomic sequencing and looking at both the species level and strain level in stool samples from individuals with IBS or IBD. We also aimed to identify potential targets for microbiome-targeted therapy by analyzing microbial pathways, antibiotic resistance, and virulence factors in the gut microbiota of patients with IBS and IBD compared with control individuals in the general population.

We undertook high-resolution shotgun metagenomic sequencing of stool samples from three well-phenotyped Dutch cohorts: LifeLines DEEP, a general population cohort, the University Medical Center of Groningen IBD (UMCG IBD) cohort, and the Maastricht IBS (MIBS) case-control cohort. In total, we analyzed stool samples from 1792 participants: 355 patients with IBD, 412 patients with IBS, and 1025 controls (table S1).

RESULTS

Species-level and strain-level identification shows microbiome signatures in stool samples from patients with IBD or IBS

Species-level and strain-level identification of the gut microbiome was necessary to identify potential disease-associated microbes that could be cultured and then investigated in functional studies. First, we assessed the overall composition (Fig. 1) and the microbial alpha diversity (fig. S1) of the gut microbiome of stool samples from control individuals and those with IBS or IBD. Next, we performed association analyses of the relative taxonomy abundance for each group of individuals (table S2), correcting for 26 previously identified confounding factors (table S3) (*10*). In total, 219 of the 477 identified

[1]University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology, Groningen, Netherlands. [2]University of Groningen and University Medical Center Groningen, Department of Genetics, Groningen, Netherlands. [3]Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [5]Maastricht University Medical Center+, Division Gastroenterology-Hepatology, NUTRIM School for Nutrition, and Translational Research in Metabolism, Maastricht, Netherlands. [6]University of Groningen and University Medical Center Groningen, Department of Pediatrics, Groningen, Netherlands.
*Co-first authors.
†Co-second authors.
‡Deceased.
§Co-last authors.
‖Corresponding author. Email: r.k.weersma@umcg.nl

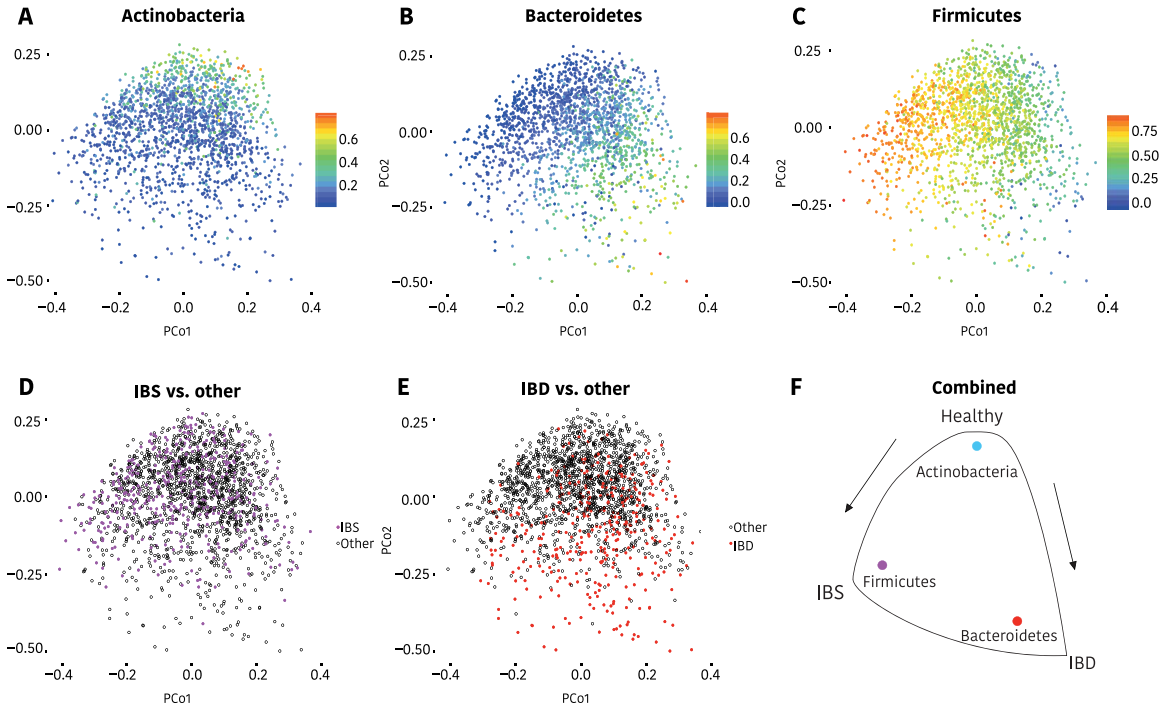

Fig. 1. Principal coordinate analysis of Bray-Curtis dissimilarities showing the gut microbiome spectrum of 1792 human fecal metagenomes. Bray-Curtis dissimilarities were calculated from taxonomic end points. End points were defined as the lowest nonredundant taxonomic level. The first principal coordinate is represented by the *x* axis, and the second principal coordinate is represented by the *y* axis. The relative abundance of the three most abundant bacterial phyla—Actinobacteria (**A**), Bacteroidetes (**B**), and Firmicutes (**C**)—underlies the first two principal coordinates (PCos). The metagenomes of patients with IBS (**D**) or IBD (**E**) differed from those of the population controls (IBD versus control PCo1, $P = 1.20 \times 10^{-5}$; PCo2, $P = 2.20 \times 10^{-16}$; IBS versus control PCo1, $P = 8.05 \times 10^{-6}$; PCo2 $P = 6.72 \times 10^{-7}$; two-sided unpaired Wilcoxon rank-sum test) and from each other (PCo1, $P = 2.22 \times 10^{-7}$; PCo2, $P = 5.06 \times 10^{-12}$). On average, as schematically depicted (**F**), controls had more Actinobacteria in their stool than did patients with IBD or IBS. Patients with IBS had more Firmicutes and less Bacteroidetes than did controls. In contrast, patients with IBD had less Firmicutes and more Bacteroidetes than did controls.

nonredundant taxa were associated with patients with CD (table S4), 102 taxa with patients with UC (table S5), and 66 taxa with patients with IBS who had been diagnosed by a gastroenterologist (IBS-GE; table S6) [significance threshold for all associations, false discovery rate (FDR) < 0.01]. Patients with CD or UC showed similar dysbiotic gut microbiome profiles. Of the 102 UC-associated bacterial taxa, 87 were also found to be associated with the gut microbiome profiles of patients with CD. However, we also identified 15 UC-specific associations, including the species *Bacteroides uniformis* (FDR = 8.31×10^{-5}; table S5) and *Bifidobacterium bifidum* (FDR = 6.78×10^{-7}; table S5). Compared with controls, patients with IBD and patients with IBS-GE showed substantial overlap in the increase and decrease in the relative abundance of bacterial species in their gut microbiome. In total, 24 taxa were associated with both IBD and IBS (table S7 and fig. S2). These associations included a decrease in several butyrate-producing bacteria, including *Faecalibacterium prausnitzii*, a known beneficial bacterium with anti-inflammatory properties that was lower in individuals with CD or IBS-GE (FDR = 1.85×10^{-34} or 7.30×10^{-6}, respectively; table S9). No significant differences were observed in patients with UC compared with controls (FDR = 0.93; table S9), although a trend toward lower *F. prausnitzii* was observed in patients with UC with active disease, defined as Simple Clinical Colitis Activity Index values above 2.5 ($P = 0.05$, FDR = 0.39; table S9). In addition to the 24 overlapping associations, we also found disease-specific associations. The abundance of *Bacteroides* species, for example, was only increased in patients with IBD but not in those with IBS (table S9). *Bacteroides* are typically symbionts but can also be opportunistic pathogens (*11*). In this study, observed increases in patients with CD or UC included *Bacteroides fragilis* ($FDR_{CD} = 1.33 \times 10^{-5}$, $FDR_{UC} = 0.0039$; table S9), previously linked to impaired bacterial tolerance handling by CD-associated genetic variation in the genes *NOD2* and *ATG16L1*, and *Bacteroides vulgatus* ($FDR_{CD} = 1.00 \times 10^{-9}$; table S4), linked to pathogenesis of CD and *NOD2* host genetic variants (*12, 13*). An increase in species of the Enterobacteriaceae family was observed only in patients with CD (table S9), including increases in *Escherichia/Shigella* species, which are known to invade the gut mucosal epithelium, cause bloody diarrhea, and ulceration of the colon (*14*). Moreover, the abundance of species such as *Bifidobacterium longum* that are capable of resisting enteric infections by *Shigella* species was lower in patients with CD ($FDR_{CD} = 6.13 \times 10^{-6}$; table S4) (*15*). IBS-GE was associated with an increase in several *Streptococcus* species (table S6). In contrast, there were no significant alterations in the gut microbiome associated with an IBS diagnosis based on questionnaire responses (IBS-POP; table S8). However, when a looser significance threshold was applied, the decreased abundance of *F. prausnitzii* and the increase in *Streptococcus* species could be replicated (FDR < 0.1; table S8). Figure 2 gives an overview of the gut microbiome associations identified in CD, UC, and IBS-GE, depicting the numbers of increased and decreased species per family. Detailed results of the case-control taxonomy analyses including all disease cohorts versus control data are shown in table S9.

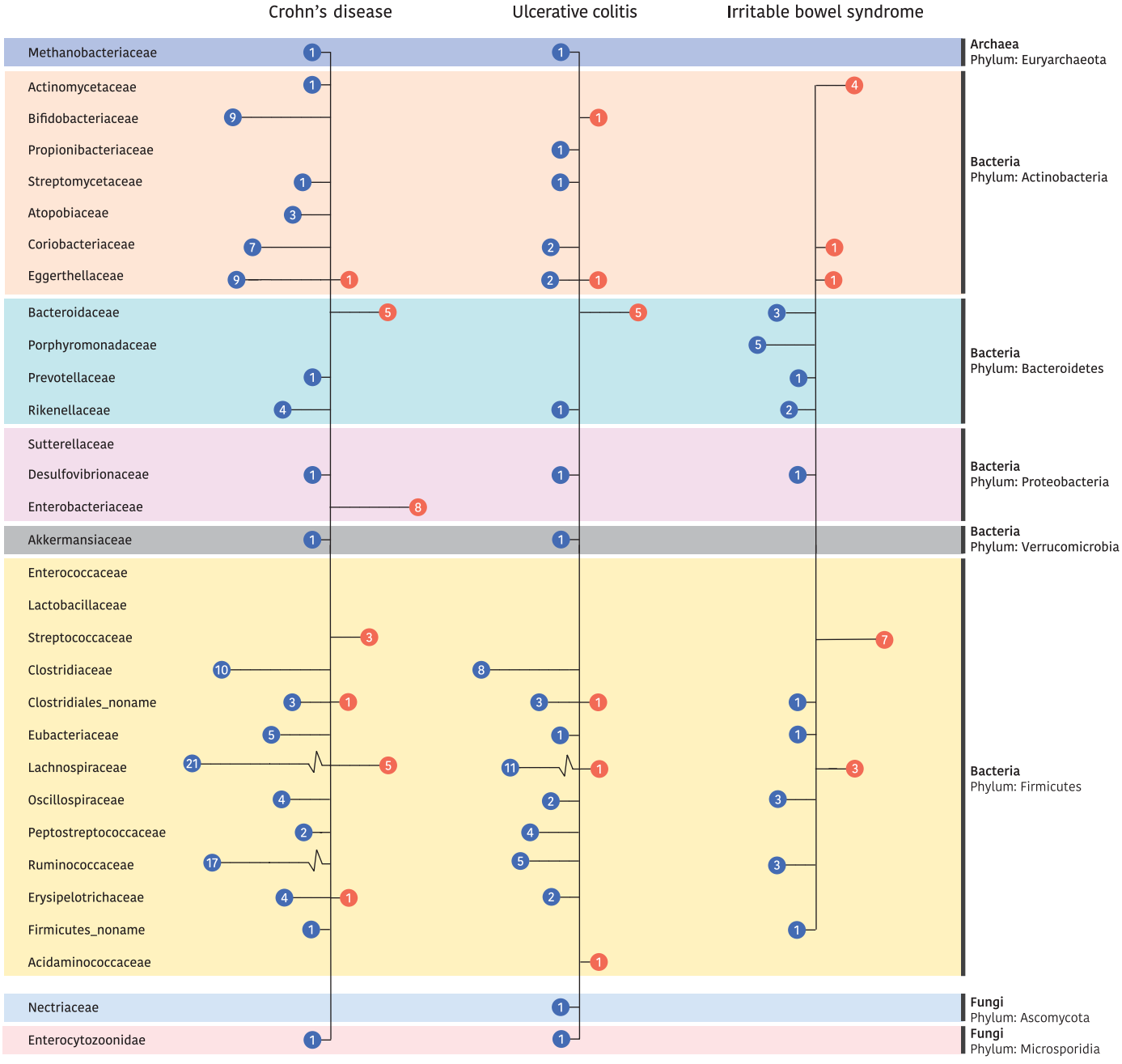

Fig. 2. Gut microbiota species associated with CD, UC, and IBS-GE compared with controls. Statistically significant results (FDR < 0.01) of the case-control multivariate model analyses are depicted. Per microbial family, the number of species that were increased (orange) or decreased (blue) is shown including 134 species in CD belonging to 24 families, 58 species in UC belonging to 21 families, and 37 species in IBS-GE belonging to 15 families.

We next asked how disease state affected strain-level diversity. We hypothesized that if conditions favored the growth of pathogenic bacteria, then the strain diversity of those organisms may increase compared with diversity values in healthy individuals. Conversely, for beneficial microbes, if these organisms were more likely to be lost from the gut or to suffer from generally reduced population sizes, then population bottlenecks may reduce diversity. We investigated bacterial strain diversity in stool samples from patients with IBD or IBS by assessing the genetic heterozygosity in a set of marker genes. We consistently found increased strain diversity in likely pathogenic species and reduced strain diversity in beneficial species in stool samples from patients with IBD or IBS compared with controls. In total, we found that strain diversity of 21, 15, or 1 bacterial species was altered in patients with CD, UC, and IBS-GE, respectively

(FDR < 0.01; table S10). For example, in patients with CD, UC, and IBS-GE, the strain diversity of the beneficial bacterium *F. prausnitzii* ($FDR_{CD} = 1.34 \times 10^{-13}$, $FDR_{UC} = 1.87 \times 10^{-7}$, $FDR_{IBS\text{-}GE} = 3.56 \times 10^{-5}$, and $FDR_{IBS\text{-}POP} = 0.03$) was decreased (table S10). In stool samples from patients with CD or UC, the strain abundance of *Roseburia intestinalis* decreased ($FDR_{CD} = 3.30 \times 10^{-13}$, $FDR_{UC} = 2.56 \times 10^{-9}$; table S10). *Roseburia* species are acetate-to-butyrate converters that reside in the intestinal mucus layer, where they have anti-inflammatory effects. For some bacteria, e.g., *F. prausnitzii*, both the abundance and the strain diversity were decreased in IBD or IBS-GE (tables S9 and S10). However, for other bacteria, e.g., *R. intestinalis*, the abundance was not altered in the disease, whereas the strain diversity did decrease (Fig. 3 and tables S9 and S10).

Different bacterial growth dynamics are observed in stool samples from patients with IBD or IBS

Cross-sectional studies provide an overview of the relative abundance of bacterial taxa at a single time point and therefore do not capture the complex dynamics of the microbial ecosystems in the gut of patients with IBD or IBS. Recently, it has been shown that bacterial growth dynamics could be inferred from a single metagenomic sample by studying the pattern of sequencing read coverage [peak-to-trough ratio (PTR)] across the gut bacterial genomes (*16*). The assessment of disease-associated growth rate differences could help to identify actively growing bacteria and, hence, could help to prioritize disease-associated taxonomy results. In our dataset, bacterial growth rates could be determined for 40 species and were altered in four species in patients with CD, five species in patients with UC, and one species in patients with IBS-GE, compared with control individuals (FDR < 0.01) (table S11). In patients with CD, the bacterial growth rates of *B. fragilis* ($FDR_{CD} = 0.005$) and *Escherichia coli* ($FDR_{CD} = 0.0004$) were increased compared with controls (table S11).

Gut microbiota composition can be used to distinguish IBD from IBS-GE

Given the observed differences in gut microbiome between patients with IBD and IBS-GE, we investigated the use of microbial taxonomy markers as potential predictors of disease. Because of the substantial overlap in clinical presentation, it can be difficult for a general practitioner or gastroenterologist to distinguish between IBD and IBS, and colonoscopies are performed in a large number of patients to reach the correct diagnosis. We applied a machine learning technique based on generalized linear models with penalized maximum likelihoods to our gut microbiome data. To overcome the lack of an independent replication cohort, the prediction accuracy was estimated by performing a 10-fold cross-validation, dividing the disease cohort into a 90% training set and a 10% discovery set. The microbial composition showed a better prediction accuracy [area under the curve $(AUC)_{mean} = 0.91$ (0.81 to 0.99)] than the currently used fecal inflammation biomarker calprotectin [$AUC_{mean} = 0.80$ (0.71 to 0.88); $P = 0.002$, two-sided paired Wilcoxon rank-sum test; table S12]. Only minor differences in the ability to discriminate between IBD and IBS were observed when using either the microbial taxonomy data or the microbial pathways or both datasets combined (table S12). Next, a selection of the top 20 taxonomies (table S13) with the largest effect size in the prediction model was tested, resulting in an AUC_{mean} of 0.90. The use of the top five taxonomies also led to a similar prediction accuracy as fecal calprotectin measurements (milligrams per kilogram; top five taxa, $AUC_{mean} = 0.81$ and $AUC_{calprotectin} = 0.80$; tables S12 and S13). When we combined the fecal calprotectin measurements with the top 20 selected taxonomies, the model reached the highest prediction accuracy ($AUC_{mean} = 0.93$; Fig. 4 and table S13).

Metagenomic analysis reveals functional changes in the gut microbiota in stool samples from patients with IBD and IBS

Metagenomic sequencing enabled the determination of the functional capacity of the gut microbiome from patients with CD, UC, or IBS-GE. In stool samples from patients with CD, UC, or IBS-GE, a number of microbial pathways were altered compared with those of controls (175, 61, or 38 altered pathways, respectively; FDR < 0.01; tables S14 and S15). We identified both overlap and differences in microbial functions that included the synthesis of amino acids, neurotransmitters, and vitamins, as well as the regulation of mineral absorption and the degradation of complex carbohydrates (table S15). The fermentation of pyruvate to butanoate, a butyrate precursor, was decreased in stool samples from patients with IBD and IBS-GE (CENTFERM_PWY, $FDR_{IBD} = 6.10 \times 10^{-10}$, $FDR_{IBS\text{-}GE} = 6.57 \times 10^{-5}$; table S15). In patients with CD, the decreased fermentation pathways, the higher sugar degradation, and the increased biosynthesis of quinones formed a microbial environment characteristic of inflammation (table S15). In patients with UC, pathways producing butyrate and acetate were decreased (e.g., PWY_5676, $FDR_{UC} = 0.0029$), and pathways producing lactate were increased (ANAEROFRUCAT_PWY, $FDR_{UC} = 0.0004$; P122_PWY, $FDR_{UC} = 0.0001$; table S15). However, in patients with IBS-GE, the metabolic signatures were characterized by increased fermentation (e.g., FERMENTATION_PWY, $FDR_{IBS\text{-}GE} = 6.24 \times 10^{-7}$) and carbohydrate degradation pathways (e.g., LACTOSECAT_PWY, $FDR_{IBS\text{-}GE} = 0.0016$; table S15).

We found alterations in several microbial L-arginine pathways, suggesting that there may be depletion of L-arginine in patients with CD. Three microbial L-arginine biosynthesis pathways were decreased in patients with CD (PWY_7400, $FDR_{CD} = 0.0007$; ARGSYN_PWY, $FDR_{CD} = 0.0003$; ARGSYNBSUB_PWY, $FDR_{CD} = 1.01 \times 10^{-9}$; table S15). Vitamins can act as antioxidants, one example being vitamin B2 or riboflavin. Several flavin pathways were decreased in patients with CD (PWY_6167, $FDR_{CD} = 2.29 \times 10^{-6}$; PWY_6168, $FDR_{CD} = 1.47 \times 10^{-6}$; RIBOSYN2_PWY, $FDR_{CD} = 0.0003$; table S15) and UC (PWY_6167, $FDR_{UC} = 0.01$; table S15).

Patients with IBD or IBS show increased abundance of virulence factors in their gut microbiota

Virulence factors contribute to the pathogenic potential of bacteria through several mechanisms, including increased adhesion of bacteria to the gut mucosa, immune system evasion, or suppression of the host immune response. We assessed the homology between our metagenomic reads and the protein sequences from the Virulence Factor Database (VFDB). Among patients with CD, UC, IBS-GE, or IBS-POP, the relative abundance of 262 virulence factors was increased compared with controls (FDR < 0.01; table S16). In patients with CD, the abundance of 216 virulence factors was increased (table S16). Proteins belonging to different iron uptake pathways were increased, including the yersiniabactins ybt ($FDR_{ybt\text{-}a} = 0.002$, $FDR_{ybt\text{-}s} = 3.40 \times 10^{-7}$, $FDR_{ybt\text{-}t} = 5.12 \times 10^{-7}$, $FDR_{ybt\text{-}u} = 4.20 \times 10^{-7}$, and $FDR_{ybt\text{-}x} = 2.95 \times 10^{-7}$) usually found in *Yersinia pestis* and the enterobactin proteins entA-F ($FDR < 6.78 \times 10^{-5}$) and entS ($FDR = 2.30 \times 10^{-8}$) usually found in *E. coli* (table S16). The abundance of enterobactins correlated with the relative abundance of Enterobacteriales (Spearman coefficient, rho = 0.8; FDR < 0.01; table S16).

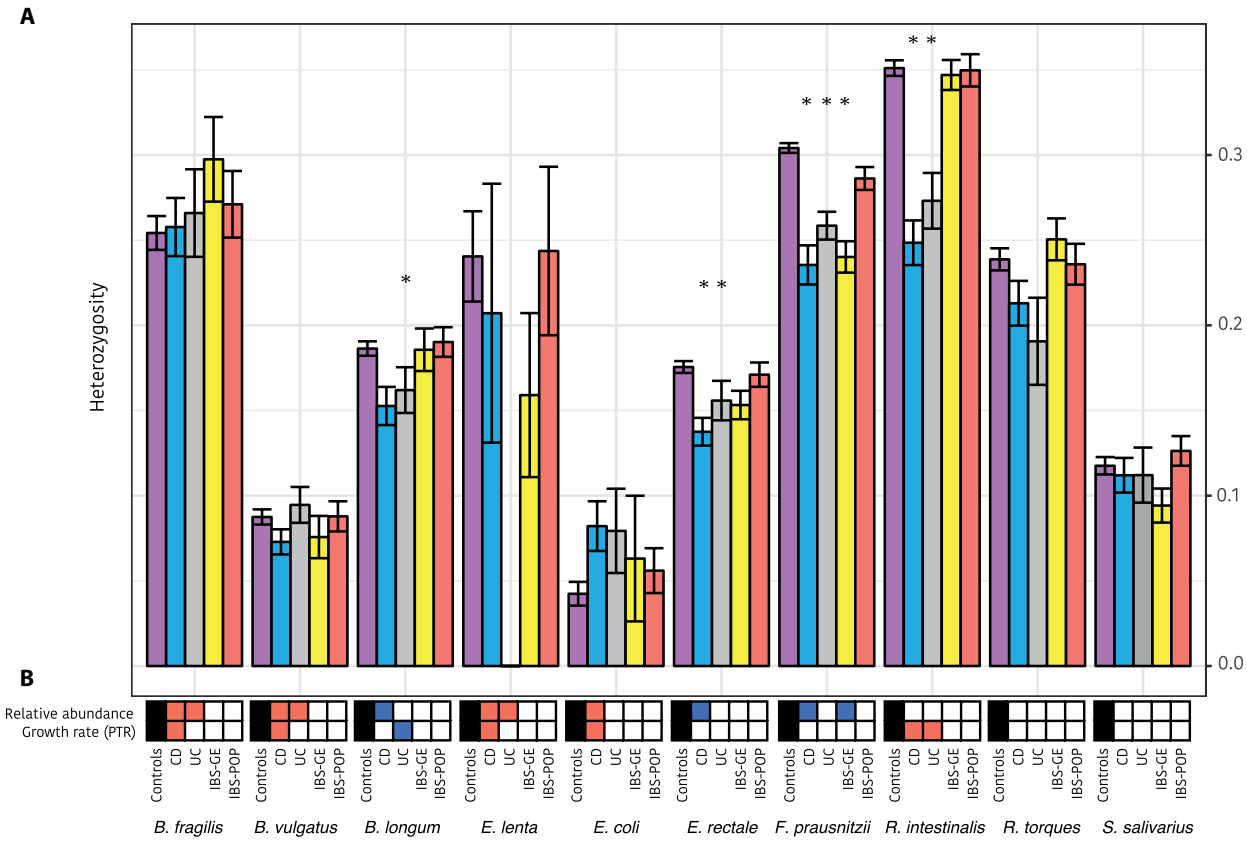

Fig. 3. Differences in bacterial abundance, bacterial strain diversity, and bacterial growth rates of key species in patients with IBD and IBS and controls. (**A**) Bar plots representing the heterozygosity values within bacterial species are shown; SEs are indicated. Heterozygosity is used as an estimation of the strain diversity within a species. Higher heterozygosity values indicate the presence of multiple strains of the same species. Each bar represents a cohort: Controls are depicted in purple, patients with CD in blue, patients with UC in gray, patients with IBS-GE in yellow, and patients with IBS-POP in red. Each asterisk indicates significant differences when comparing to controls (FDR < 0.01). (**B**) Heatmaps indicate significant changes in relative abundance and growth rates [peak-to-trough ratio algorithm (PTR)] of each bacterial species in disease cohorts compared with controls. Red boxes indicate a significant increase, and blue boxes a significant decrease (FDR < 0.01).

This increase in virulence factors was also reflected in an increase in the enterobactin pathway in patients with CD (ENTBACSYN_PWY, FDR = 0.006; table S15). Many pathogens have acquired efficient iron uptake mechanisms that give them a survival advantage in low iron environments (*17*–*20*). This was reflected in alterations in several microbial iron uptake pathways in patients with CD (HEME-BIOSYNTHESIS-II, PWY-5918, and PWY-5920, FDR < 0.01; table S15). In patients with UC, 35 virulence factors were increased, for example, the relative abundance of MU-toxin and its transport protein complex containing nagI, nagJ, and nagL were increased ($FDR_{nagI} = 3.56 \times 10^{-5}$, $FDR_{nagJ} = 4.59 \times 10^{-13}$, and $FDR_{nagL} = 9.11 \times 10^{-9}$; table S16).

Changes in the microbiome composition in patients with IBD and IBS have an impact in the antibiotic resistance load

Metagenomic sequencing provides the opportunity to study the resistome of patients with IBD or IBS on a large scale. To see whether increases in antibiotic resistance were present in the gut microbiota of patients with IBD or IBS, we assessed the homology between metagenomic reads and protein sequences from the antibiotic resistance database, Comprehensive Antibiotic Resistance Database (CARD). Subsequently, to identify the microbes that potentially harbored the antibiotic resistance proteins, we correlated the abundance of antibiotic resistance genes with taxonomy abundance. In patients with CD, the abundance of 142 genes encoding antibiotic resistance proteins was higher than that in controls. Of these antibiotic resistance proteins, 63 were components of efflux complexes that remove antibiotics from the bacteria, thereby preventing the antibiotics from working effectively (table S17). These efflux complexes consist of three proteins that span the inner membrane, the periplasm, and the outer membrane of bacteria. Some efflux pumps can only transport one specific type of antibiotic, whereas other efflux pumps, called multidrug efflux pumps, can transport several types of antibiotics. The antibiotic resistance protein TolC, which was increased in patients with CD ($FDR = 5.26 \times 10^{-6}$; table S17), is an outer membrane protein comprising several multidrug efflux pumps. TolC is often combined with other inner membrane and periplasmic efflux proteins including AcrA, AcrB, MdtA/B/C, MdtE/F, emrA/B, and emrK/Y. The abundance of these proteins was also increased in patients with CD ($FDR_{crA} = 1.41 \times 10^{-9}$, $FDR_{AcrB} = 4.60 \times 10^{-11}$, $FDR_{MdtA} = 4.75 \times 10^{-5}$, $FDR_{MdtB} = 0.002$, $FDR_{MdtC} = 2.28 \times 10^{-15}$, $FDR_{MdtE} = 0.005$, $FDR_{MdtF} = 0.0001$, $FDR_{emrA} = 1.23 \times 10^{-5}$, $FDR_{emrB} = 2.99 \times 10^{-8}$, $FDR_{emrK} = 2.54 \times 10^{-8}$, and $FDR_{emrY} = 8.83 \times 10^{-9}$; table S17). The abundance of *TolC* in patients with CD correlated with taxonomy abundance of the genus *Escherichia* that was also increased (Spearman coefficient, rho = 0.80; $FDR < 1.0 \times 10^{-16}$; table S17). In patients with UC, the abundance of 66 genes encoding antibiotic resistance

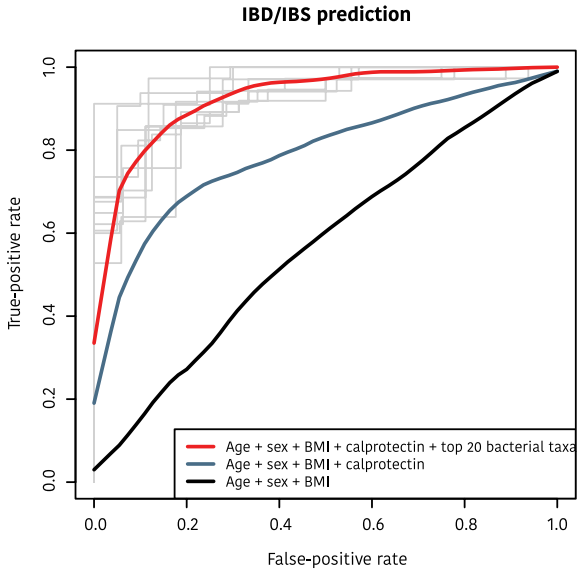

Fig. 4. Prediction model to distinguish IBD from IBS. Shown is a receiver operating characteristic curve (ROC) describing the prediction accuracy of three different models calculated using a 10-fold cross-validation. The black line represents the prediction accuracy when using age, sex, and body mass index of each participant to discriminate between patients with IBD or IBS-GE. When adding fecal calprotectin measurements to the model (blue line), the AUC achieved a mean value of 0.80. Adding the relative abundance of the top 20 most discriminating bacterial taxa (red line) improved the classification accuracy power (AUC = 0.90).

proteins was higher than that in controls. One of the highest differentially abundant antibiotic resistance proteins in patients with UC was cepA (FDR = 4.85×10^{-12}; table S17). This antibiotic resistance protein is a β-lactamase, an enzyme mediating resistance to β-lactam antibiotics, including the frequently prescribed antibiotics amoxicillin and penicillin (*21*). The abundance of the antibiotic resistance gene *cepA* correlated with the abundance of the genus *Bacteroides*, which was increased in patients with UC and CD (Spearman coefficient, rho = 0.86; FDR < 1.0×10^{-16}; table S17). Several genes encoding for antibiotic resistance proteins were increased in patients with IBS, and the abundance of 32 antibiotic resistance genes was increased in patients with IBS-GE compared with controls. One of most increased antibiotic resistance proteins in patients with IBS-GE was mecB (FDR = 0.0001; table S17), which is involved in resistance to methicillin. This protein is usually found in species belonging to the *Macrococcus* genus, which is closely related to the *Staphylococcus* genus (*22*). In patients with IBS-POP, the abundance of 13 genes encoding for antibiotic resistance proteins was increased compared with controls, including PBP2x (FDR = 0.0056; table S17), a penicillin-binding protein. PBP2x, usually found in *Streptococcus pneumoniae* (*23*), was highly correlated with the taxonomy abundance of the genus *Streptococcus* (Spearman coefficient, rho = 0.91; FDR < 1.0×10^{-16}; table S17) in our gut microbiome data. We investigated whether current antibiotic use correlated with the presence of antibiotic resistance genes, but only a few individuals were taking antibiotics, and no statistically significant associations were found.

Gut microbiota changes are associated with disease-specific factors and disease subphenotypes

Previous studies have established that the composition of the gut microbiota is influenced by over 100 intrinsic and extrinsic factors (e.g., dietary factors, medications, disease, and anthropometric factors) in the general population (*10, 24*). However, in IBD and IBS, both the gut microbiota composition and various phenotypes (e.g., defecation frequency, medication use, and previously performed GI surgical interventions) may be altered. Therefore, we recalculated the relation between intrinsic and extrinsic factors and the overall microbial composition (Bray-Curtis dissimilarities), alpha diversity (Shannon index), and gene richness (tables S18 to S21 and Fig. 5). These results, together with the correlations of the intrinsic and extrinsic factors (tables S22 to S25), resulted in the lists of factors that were included in subsequent association analyses (table S26). Univariate and multivariate within-cases association analyses were performed on taxonomy (tables S27 to S34) and microbial pathways (tables S35 to S43). In CD, only 1% of the microbial variance could be explained by inflammatory disease activity (FDR = 0.077; table S18). In contrast, ileocecal resection in patients with CD resulting in the removal of the ileocecal valve was the factor that explained 5% of the variance (FDR = 0.00159; table S18). The absence of the ileocecal valve was associated with a decrease in microbial and gene richness, specifically with decreases in the beneficial bacterium *F. prausnitzii* ($FDR_{CD\text{-}ileal} = 8.01 \times 10^{-10}$; table S27) and the Ruminococcaceae family ($FDR_{CD\text{-}ileal} = 4.63 \times 10^{-10}$; table S27) and an increase in *Fusobacterium* ($FDR_{CD\text{-}ileal} = 0.002$; table S27). This suggested that removing the ileocecal valve had negative consequences for the gut microbiota of patients with IBD. Vitamin D supplementation in patients with CD was associated with a decreased abundance of *Akkermansia muciniphila* ($FDR_{CD} = 0.19$; table S27), a mucin-degrading bacterium that grows in a low-fiber environment (*25*).

DISCUSSION

The use of shotgun metagenomic sequencing data allowed us to explore the complexity of the gut microbial ecosystem with high resolution. We were also able to describe some important characteristics of the microbial community, including the strain diversity, the growth dynamics, and the presence of genes involved in bacterial virulence and in antibiotic resistance mechanisms that can provide an adaptive advantage to opportunistic and pathogenic microbes. We also explored the changes in microbial pathway profiles, providing relevant information on the functional consequences of microbiome dysbiosis. The integration of these datasets allowed us to pinpoint key species as targets for functional studies in IBD and IBS (Fig. 3) and to connect knowledge of the etiology and pathogenesis of IBD and IBS with the gut microbiome to provide potential new targets for treatment.

Before our results can be translated into clinical practice, much more additional evidence is required to overcome the limitations of our study. The relevance of the microbial pathways described in this study needs to be supported by metatranscriptomics and metabolomics data, as well as functional experiments. We have described the resistome and virulence factor abundance in the gut microbiota of patients with IBD or IBS. However, to identify the relevant mechanisms associated with GI disease, experiments based on culturomics and whole-genome sequencing of specific bacterial strains are needed. In addition, replication in independent cohorts, including in patients with other GI disorders or prediagnostic groups, will be needed to validate the sensitivity and specificity of our prediction model. In this study, we made use of two cohorts consisting of patients already diagnosed with IBD or IBS. Therefore, our prediction model does

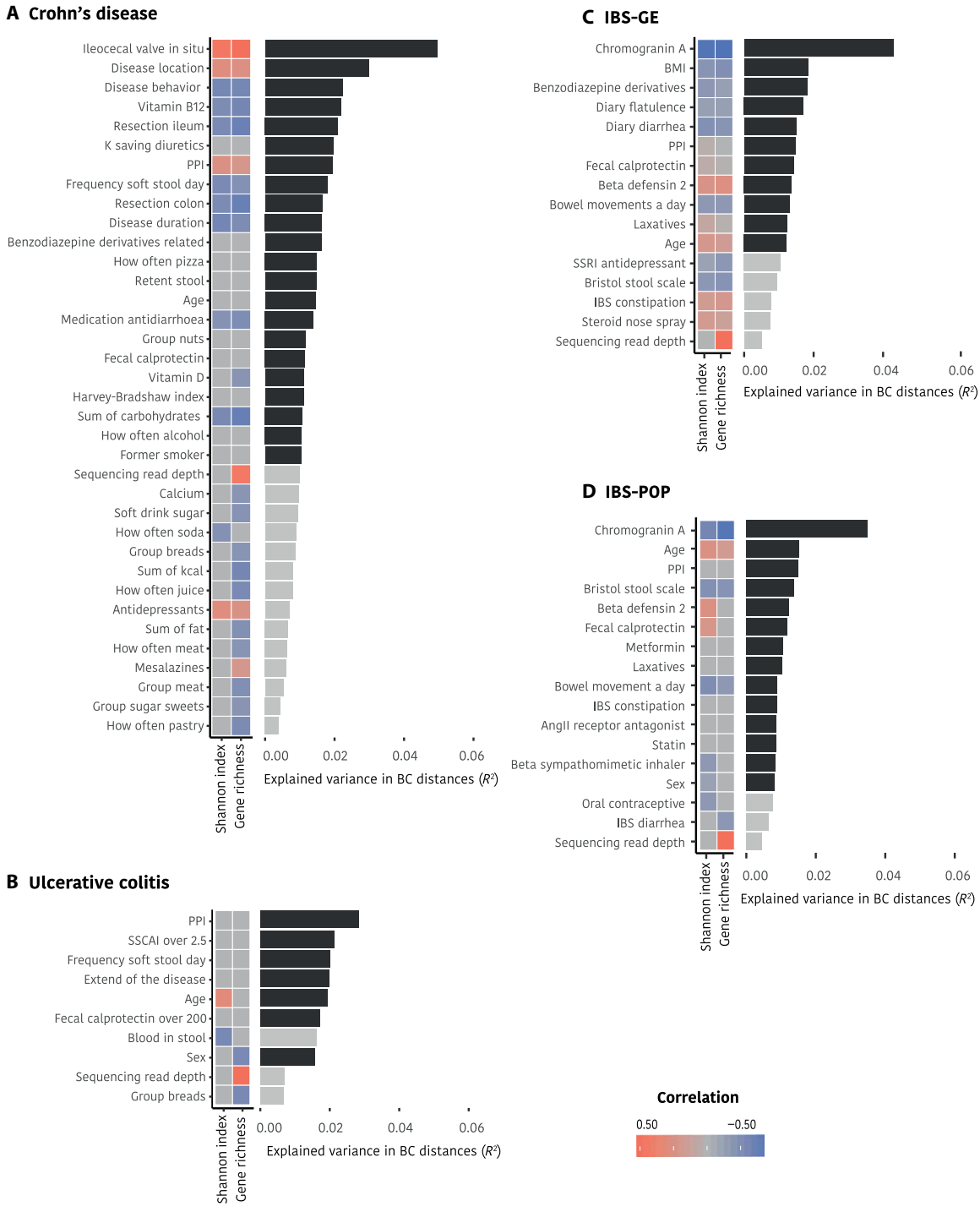

Fig. 5. Associated phenotypes for microbial richness and gut microbiota composition. Shown are associated phenotypes for microbial richness and gut microbiota composition in four disease cohorts: (**A**) CD, (**B**) UC, (**C**) IBS-GE, (**D**) IBS-POP. In the bar plots, the *x* axis represents the explained variance of each phenotype on gut microbiota composition expressed as Bray-Curtis (BC) dissimilarities. Black bars indicate statistical significance (FDR < 0.1). The heatmap indicates significant positive correlations (red) or negative correlations (blue) between phenotypes and microbial richness (Shannon index) and bacterial gene richness (the number of different microbial gene families per sample). PPI, proton pump inhibitors; SSCAI, Simple Clinical Colitis Activity Index; SSRI, Selective Serotonin Reuptake Inhibitor.

not reflect the clinical situation where treatment-naïve patients or patients with other comorbidities could present with different microbiome characteristics. Moreover, variations in laboratory protocols, sequencing techniques, or geographical origin of samples may also influence the accuracy of our model. Cross-sectional cohorts of patients with established disease allowed us to find the influence of many different subphenotypes; however, these cohorts can only provide limited insight into the mechanisms underlying the onset

of IBD or IBS. Longitudinal studies will help to determine the dynamics of the disease, as well as distinguishing the microbial features that are causal from those that are consequences of disease. Another limitation of this study was the relatively low numbers of well-defined patients with IBS. Therefore, we could not perform an in-depth characterization of the IBS subphenotypes such as patients with constipation or diarrhea.

The availability of many phenotypic characteristics, e.g., medication use or lifestyle, for each participant in our study enabled us to perform a strict case-control analysis while taking important confounding factors into account. The use of well-characterized cohorts should become a common practice when studying the microbiome in a disease context. The use of drugs such as proton pump inhibitors or laxatives, which are more often used by patients with IBD or IBS, has a large impact on the gut microbiota composition. Considering these effects, correction for these medications is essential for identifying disease-associated microbial features and avoiding false-positive associations due to changes in GI acidity or bowel mobility. In addition, our study provides new information about the effects of lifestyle and medication on microbiome composition and function in patients with IBD or IBS and finds associations between microbial signatures and the subphenotypes of IBD and IBS. Whereas disease activity explains a large proportion of the variation in microbial composition in patients with UC, disease location and gut resections have a large impact on the gut ecosystem in patients with CD. This fact highlights the importance of collecting and considering disease-specific phenotypes when analyzing the microbial composition of patients with IBD or IBS.

Dysbiosis of the gut microbiota was observed in patients with IBD. The two main subtypes of IBD (CD and UC) showed substantial overlap in their gut microbial signatures. These shared signatures could be an indicator of gut inflammation. However, when compared with controls, the microbial changes in patients with CD were larger than those in patients with UC. This is concordant with previous studies that identified inflammation of the ileum as one of the main drivers of differential microbiome signatures between CD and UC (*7*, *26*). Furthermore, in patients with CD, the removal of the ileocecal valve was found to be associated with a reduction in microbiome richness (Fig. 5) and a decrease in pathways involved in the degradation of primary bile acids (table S35). These findings are consistent with clinical observations of bile acid malabsorption in patients with IBD (*27*). In addition, absence of the ileocecal valve was related to a decrease in the relative abundance of *F. prausnitzii*. *F. prausnitzii* is an anaerobic bacterium that is sensitive to small changes in bile salt concentrations (*28*). Oxidative stress produced by inflammation in the gut, together with a decrease in antioxidant biosynthesis pathways and changes in bile acid metabolism, could explain the observed reduction in *F. prausnitzii* solely in the CD subtype of IBD.

A moderate decrease in *F. prausnitzii* accompanied by an increase in the abundance of *Streptococcus* species was the main characteristic of the gut microbiota of participants with IBS symptoms based on ROME-III criteria; this was consistent with similar changes observed in the clinical IBS cohort. Larger changes in gut microbiota composition were observed in the IBS cohort defined by a gastroenterologist, including a decrease in butyrate-producing bacteria and an increase in taxa belonging to the *Actinomyces*, *Streptococcus*, and *Blautia* genera. Although no significant differences were observed between the gut microbiotas of IBS subtypes, when comparing patients with IBS with diarrhea to controls, an increase in the relative abundance of *Eggerthella lenta* and a decrease in the sulfate-reducing bacterial family Desulfovibrionaceae were observed.

Although the gut microbiota composition has been described as stable across individuals in different population cohorts even in the presence of high interindividual taxonomic variation (*10*), a large number of microbial pathways were shown to be disrupted in patients with IBD or IBS. Our comprehensive analyses of microbial pathways provide relevant information that can help in the design of better therapeutics aimed at restoring the microbial ecosystem in patients with IBD or IBS. Thus far, the results of prebiotic, probiotic, dietary, and fecal transplantation interventions meant to invoke beneficial changes in the gut microbiome in IBD and IBS have been disappointing. However, focusing on interventions that change the functions of the gut microbiota could be more successful. For example, combining antioxidant vitamin supplementation with fecal microbiome transplantation or *F. prausnitzii* probiotics could protect anaerobic bacteria from oxidative stress during intestinal inflammation; providing L-arginine supplements to patients with CD could enhance wound healing in the damaged gut.

Our study also found more evidence for mechanisms implicated in the maintenance of gut health. For example, in patients with IBD, we found a reduction in the methanogenesis pathway (table S15). This pathway is strongly correlated with the presence of Methanobacteria, of which *Methanobrevibacter smithii* is the most abundant species (*29*). Another example is our observed reduction in pathways that produce hydrogen sulfide in patients with IBD (e.g., SO4ASSIM-PWY and PWY-821; table S15). Although the effect of changes in concentrations of hydrogen sulfide is still being debated, several studies have shown that this molecule could have antioxidant and immune-regulatory properties (*30*).

Virulence factors are key features for the selective advantage of potentially pathogenic bacteria over common members of the healthy gut microbiota. Mechanisms that alter the mucosal composition or increase bacterial adhesion, secretion of toxins, or competition with the host for resources could contribute to IBD and IBS pathogenesis. So far, studies of virulence mechanisms in the context of GI diseases have focused on specific groups of bacteria such as adherent-invasive *E. coli* (*31*) and microbial proteases (*32*). By exploring the pathogenic potential of the gut microbiota community in IBD and IBS, we were able to identify other potential targets such as Mu-toxin in patients with UC. Although these findings still need to be validated by targeted approaches or transcriptomic analyses, the virulence factor associations we present provide a better understanding of the pathogenesis of both disorders.

The changes we identified in gut microbiota composition and functional potential in patients with IBD and IBS could lead to new tools that assist diagnosis in clinical practice. Sophisticated models that include a combination of different blood or stool biomarkers and that have been validated in a replication cohort are required to design new diagnostic tests. Our results suggest that in the future the use of probes directed at key bacterial species could complement fecal calprotectin measurements in distinguishing the diagnosis of IBS and IBD.

MATERIALS AND METHODS

Study design

The aims of this cross-sectional study were to describe the features of the gut microbiota of patients with IBD or IBS and to compare

them to those of control individuals from the general population. We analyzed fecal metagenomes of 1792 individuals. We combined species-level profiles and strain-level profiles with bacterial growth rates, metabolic function, antibiotic resistance, and virulence factor analyses to identify key bacterial species that may be involved in GI diseases.

The following three cohorts from The Netherlands were used: LifeLines DEEP, UMCG IBD cohort, and MIBS cohort. IBD was diagnosed by a gastroenterologist based on accepted radiological, endoscopic, and histopathological evaluation. Of the 355 patients with IBD, 208 patients were diagnosed with CD, 126 patients with UC, and 21 patients with IBD-unclassified/indeterminate. We included two groups of IBS patients: The IBS-GE group consisted of 181 patients with IBS who were diagnosed by a gastroenterologist or other physician, and the IBS-POP group consisted of 231 patients with IBS from the general population whose IBS was determined on the basis of self-reported ROME-III diagnostic criteria. The control group was defined as individuals from the LifeLine Deep cohort ($n = 893$) and MIBS ($n = 132$) without GI complaints. Extensive phenotypic data were prospectively collected for both patients with IBD and patients with IBS-GE. In addition, multiple questionnaires were sent out to all participants in all cohorts to collect a wide range of uniformly processed phenotypes including disease activity, disease complaints, diet, and medication use. Each participant signed an informed consent form before participation in the cohort according to the UMCG Institutional Review Board (IRB; #M12.113965, 2008.338) and the Maastricht University Medical Center (MUMC+) IRB (#MEC 08-2.066.7/pl).

Sample collection and metagenomic sequencing

Each participant collected a single stool sample at home, which was frozen or refrigerated immediately after stool production. All the samples were then processed after the same pipeline in one laboratory (UMCG, Groningen). Fecal DNA was isolated, and metagenomic shotgun sequencing was performed as previously described using the Illumina HiSeq (*10*), generating on average 30 million reads (~3 Gb) per sample. After filtering for quality, 1792 gut metagenomes were used in all subsequent analyses.

Microbiome characterization

All metagenomic sequencing data were processed using the same extensive processing pipeline: (i) bacterial, viral, and micro-eukaryote abundances were determined using Kraken (*33*); (ii) strain diversity was determined by computing the heterozygosity of polymorphic loci within bacterial species; (iii) bacterial growth rates were estimated using a previously published peak-to-trough ratio algorithm (*16*); (iv) microbial genes and pathways were determined using the HUMAnN2 software and the MetaCyc reference (*34*); and (v) the abundances of antibiotic resistance proteins and virulence factors were identified by aligning the metagenomic reads to protein sequences in the CARD (*35*) and VFDB (*36*), respectively.

Statistical analyses

All statistical analyses were conducted in R (v 3.3.2). To compare the collected phenotypes of the disease cohort with the population controls, a χ^2 test was used for binary data. Categorical data were tested using either the two-sided unpaired *t* test for normally distributed data or the two-sided unpaired Wilcoxon rank-sum test for non-normally distributed data. The Spearman coefficient was used to evaluate the correlation between phenotypes and the correlation between microbiome features. The proportion of explained variance of each phenotype on the microbial composition dissimilarities was evaluated using a PERMANOVA test implemented in the adonis function in the vegan R package (v.2.4-1). The association between microbiome features and disease phenotypes was tested using linear models with Maaslin R library (v.0.0.4). Disease phenotype prediction tests based on microbiome features were constructed using elastic net linear models from glmnet R package (v.2.0-10), and the comparison between the goodness of fit of each model was tested using the two-sided paired Wilcoxon rank-sum test. The Benjamini and Hochberg procedure was used to adjust *P* values for multiple comparisons. An FDR <0.01 was considered statistically significant. A detailed description of the methods can be found in Supplementary Materials and Methods and figs. S3 to S6.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/10/472/eaap8914/DC1

Materials and Methods

Fig. S1. Comparison of microbial richness between cohorts.

Fig. S2. Venn diagram of overlapping taxa between IBD and clinical IBS.

Fig. S3. Cohorts, sample collection, and sample processing algorithm.

Fig. S4. Principal coordinate analysis plot on Bray-Curtis dissimilarities of controls.

Fig. S5. Phenotype data processing algorithm.

Fig. S6. Metagenomic sequencing data pipeline.

Fig. S7. Overview of statistical analyses.

Fig. S8. Prediction model to distinguish cohort of origin in disease.

Fig. S9. Prediction model to distinguish cohort of origin in controls.

Table S1. Summary statistics of phenotypes.

Table S2. Summary statistics of gut microbiome taxonomy.

Table S3. Variables included in linear models case-control analyses.

Table S4. Taxonomy results of CD versus controls.

Table S5. Taxonomy results of UC versus controls.

Table S6. Taxonomy results of IBS-GE versus controls.

Table S7. Taxonomy results in the overlap of IBD and IBS-GE.

Table S8. Taxonomy results of IBS-POP versus controls.

Table S9. Taxonomy results of all diseases versus controls.

Table S10. Strain diversity results of all diseases versus controls.

Table S11. Bacterial growth rate results of all diseases versus controls.

Table S12. Prediction accuracy of all prediction models.

Table S13. Top 20 gut microbiome features in the prediction model.

Table S14. Summary statistics of gut microbiome MetaCyc function.

Table S15. Pathway results of all diseases versus controls.

Table S16. Virulence factor results of all diseases versus controls.

Table S17. Antibiotic resistance gene results of all diseases versus controls.

Table S18. Associated phenotypes on gene richness, gut microbiome composition, and Shannon index in CD.

Table S19. Associated phenotypes on gene richness, gut microbiome composition, and Shannon index in UC.

Table S20. Associated phenotypes on gene richness, gut microbiome composition, and Shannon index in IBS-GE.

Table S21. Associated phenotypes on gene richness, gut microbiome composition, and Shannon index in IBS-POP.

Table S22. Correlation phenotypic factors with an FDR <0.1 from the Adonis analysis in CD.

Table S23. Correlation phenotypic factors with an FDR <0.1 from the Adonis analysis in UC.

Table S24. Correlation phenotypic factors with an FDR <0.1 from the Adonis analysis in IBS-GE.

Table S25. Correlation phenotypic factors with an FDR <0.1 from the Adonis analysis in IBS-POP.

Table S26. Variables included in multivariate linear models within disease cohorts.

Table S27. Taxonomy results within the CD univariate model.

Table S28. Taxonomy results within the CD multivariate model.

Table S29. Taxonomy results within the UC univariate model.

Table S30. Taxonomy results within the UC multivariate model.

Table S31. Taxonomy results within the IBS-GE univariate model.

Table S32. Taxonomy results within the IBS-GE multivariate model.

Table S33. Taxonomy results within the IBS-POP univariate model.

Table S34. Taxonomy results within the IBS-POP multivariate model.

Table S35. Pathway results within the CD univariate model.

Table S36. Pathway results within the CD multivariate model.

Table S37. Pathway results within the UC univariate model.
Table S38. Pathway results within the UC multivariate model.
Table S39. Pathway results within the IBS-GE univariate model.
Table S40. Pathway results within the IBS-GE multivariate model.
Table S41. Pathway results within the IBS-POP univariate model.
Table S42. Pathway results within the IBS-POP multivariate model.
Table S43. Cohort-associated taxa and IBD versus IBS taxonomical associations.
References (*37*–*51*)

REFERENCES AND NOTES

1. C. Abraham, J. H. Cho, Inflammatory bowel disease. *N. Engl. J. Med.* **361**, 2066–2078 (2009).
2. W. D. Chey, J. Kurlander, S. Eswaran, Irritable bowel syndrome. *JAMA* **313**, 949–958 (2015).
3. G. G. Kaplan, The global burden of IBD: From 2015 to 2025. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 720–727 (2015).
4. A. C. Ford, P. Bercik, D. G. Morgan, C. Bolino, M. I. Pintos-Sanchez, P. Moayyedi, Validation of the Rome III criteria for the diagnosis of irritable bowel syndrome in secondary care. *Gastroenterology* **145**, 1262–1270.e1 (2013).
5. G. De Palma, M. D. J. Lynch, J. Lu, V. T. Dang, Y. Deng, J. Jury, G. Umeh, P. M. Miranda, M. Pigrau Pastor, S. Sidani, M. I. Pinto-Sanchez, V. Philip, P. G. McLean, M.-G. Hagelsieb, M. G. Surette, G. E. Bergonzelli, E. F. Verdu, P. Britz-McKibbin, J. D. Neufeld, S. M. Collins, P. Bercik, Transplantation of fecal microbiota from patients with irritable bowel syndrome alters gut function and behavior in recipient mice. *Sci. Transl. Med.* **9**, eaaf6397 (2017).
6. M. Schaubeck, T. Clavel, J. Calasan, I. Lagkouvardos, S. B. Haange, N. Jehmlich, M. Basic, A. Dupont, M. Hornef, M. von Bergen, A. Bleich, D. Haller, Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* **65**, 225–237 (2016).
7. F. Imhann, A. Vich Vila, M. J. Bonder, J. Fu, D. Gevers, M. C. Visschedijk, L. M. Spekhorst, R. Alberts, L. Franke, H. M. van Dullemen, R. W. F. Ter Steege, C. Huttenhower, G. Dijkstra, R. J. Xavier, E. A. M. Festen, C. Wijmenga, A. Zhernakova, R. K. Weersma, Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* **67**, 108–119 (2018).
8. D. Gevers, S. Kugathasan, L. A. Denson, Y. Vázquez-Baeza, W. Van Treuren, B. Ren, E. Schwager, D. Knights, S. J. Song, M. Yassour, X. C. Morgan, A. D. Kostic, C. Luo, A. González, D. McDonald, Y. Haberman, T. Walters, S. Baker, J. Rosh, M. Stephens, M. Heyman, J. Markowitz, R. Baldassano, A. Griffiths, F. Sylvester, D. Mack, S. Kim, W. Crandall, J. Hyams, C. Huttenhower, R. Knight, R. J. Xavier, The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).
9. J. Tap, M. Derrien, H. Törnblom, R. Brazeilles, S. Cools-Portier, J. Doré, S. Störsrud, B. Le Nevé, L. Öhman, M. Simrén, Identification of an intestinal microbiota signature associated with severity of irritable bowel syndrome. *Gastroenterology* **152**, 111–123.e8 (2017).
10. A. Zhernakova, A. Kurilshikov, M. J. Bonder, E. F. Tigchelaar, M. Schirmer, T. Vatanen, Z. Mujagic, A. V. Vila, G. Falony, S. Vieira-Silva, J. Wang, F. Imhann, E. Brandsma, S. A. Jankipersadsing, M. Joossens, M. C. Cenit, P. Deelen, M. A. Swertz; LifeLines cohort study, R. K. Weersma, E. J. M. Feskens, M. G. Netea, D. Gevers, D. Jonkers, L. Franke, Y. S. Aulchenko, C. Huttenhower, J. Raes, M. H. Hofker, R. J. Xavier, C. Wijmenga, J. Fu, Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
11. V. Kumar, J. Sivaraman, Structural characterization of BVU_3255, a methyltransferase from human intestine antibiotic resistant pathogen *Bacteroides vulgatus*. *J. Struct. Biol.* **176**, 409–413 (2011).
12. H. Chu, A. Khosravi, I. P. Kusumawardhani, A. H. K. Kwon, A. C. Vasconcelos, L. D. Cunha, A. E. Mayer, Y. Shen, W.-L. Wu, A. Kambal, S. R. Targan, R. J. Xavier, P. B. Ernst, D. R. Green, D. P. B. McGovern, H. W. Virgin, S. K. Mazmanian, Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science* **352**, 1116–1120 (2016).
13. D. Ramanan, M. S. Tang, R. Bowcutt, P. Loke, K. Cadwell, Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity* **41**, 311–324 (2014).
14. M. Anderson, P. J. Sansonetti, B. S. Marteyn, Shigella diversity and changing landscape: Insights for the twenty-first century. *Front. Cell. Infect. Microbiol.* **6**, 45 (2016).
15. C. G. Buffie, E. G. Pamer, Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* **13**, 790–801 (2013).
16. T. Korem, D. Zeevi, J. Suez, A. Weinberger, T. Avnit-Sagi, M. Pompan-Lotan, E. Matot, G. Jona, A. Harmelin, N. Cohen, A. Sirota-Madi, C. A. Thaiss, M. Pevsner-Fischer, R. Sorek, R. J. Xavier, E. Elinav, E. Segal, Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
17. M. Miethke, M. A. Marahiel, Siderophore-based iron acquisition and pathogen control. *Microbiol. Mol. Biol. Rev.* **71**, 413–451 (2007).
18. L. Ma, A. Terwilliger, A. W. Maresso, Iron and zinc exploitation during bacterial pathogenesis. *Metallomics* **7**, 1541–1554 (2015).
19. V. Singh, B. S. Yeoh, X. Xiao, M. Kumar, M. Bachman, N. Borregaard, B. Joe, M. Vijay-Kumar, Interplay between enterobactin, myeloperoxidase and lipocalin 2 regulates *E. coli* survival in the inflamed gut. *Nat. Commun.* **6**, 7113 (2015).
20. A. J. Bäumler, V. Sperandio, Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).
21. T. V. Sydenham, J. Sóki, H. Hasman, M. Wang, U. S. Justesen, Identification of antimicrobial resistance genes in multidrug-resistant clinical *Bacteroides fragilis* isolates by whole genome shotgun sequencing. *Anaerobe* **31**, 59–64 (2015).
22. K. Becker, B. Ballhausen, R. Köck, A. Kriegeskorte, Methicillin resistance in *Staphylococcus* isolates: The "*mec* alphabet" with specific consideration of *mecC*, a *mec* homolog associated with zoonotic *S. aureus* lineages. *Int. J. Med. Microbiol.* **304**, 794–804 (2014).
23. M. van der Linden, J. Otten, C. Bergmann, C. Latorre, J. Liñares, R. Hakenbeck, Insight into the diversity of penicillin-binding protein 2x alleles and mutation in viridans streptococci. *Antimicrob. Agents Chemother.* **61**, e02646–16 (2017).
24. G. Falony, M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov, M. J. Bonder, M. Valles-Colomer, D. Vandeputte, R. Y. Tito, S. Chaffron, L. Rymenans, C. Verspecht, L. De Sutter, G. Lima-Mendez, K. D'hoe, K. Jonckheere, D. Homola, R. Garcia, E. F. Tigchelaar, L. Eeckhaudt, J. Fu, L. Henckaerts, A. Zhernakova, C. Wijmenga, J. Raes, Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
25. M. S. Desai, A. M. Seekatz, N. M. Koropatkin, N. Kamada, C. A. Hickey, M. Wolter, N. A. Pudlo, S. Kitamoto, N. Terrapon, A. Muller, V. B. Young, B. Henrissat, P. Wilmes, T. S. Stappenbeck, G. Núñez, E. C. Martens, A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
26. J. Halfvarson, C. J. Brislawn, R. Lamendella, Y. Vázquez-Baeza, W. A. Walters, L. M. Bramer, M. D'Amato, F. Bonfiglio, D. McDonald, A. Gonzalez, E. E. McClure, M. F. Dunklebarger, R. Knight, J. K. Jansson, Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2**, 17004 (2017).
27. S. Massironi, A. Zilli, F. Cavalcoli, D. Conte, M. Peracchi, Chromogranin a and other enteroendocrine markers in inflammatory bowel disease. *Neuropeptides* **58**, 127–134 (2016).
28. M. Lopez-Siles, S. H. Duncan, L. J. Garcia-Gil, M. Martinez-Medina, *Faecalibacterium prausnitzii*: From microbiology to diagnostics and prognostics. *ISME J.* **11**, 841–852 (2017).
29. M. Pimentel, R. P. Gunsalus, S. S. C. Rao, H. Zhang, Methanogens in human health and disease. *Am. J. Gastroenterol. Suppl.* **1**, 28–33 (2012).
30. F.-F. Guo, T.-C. Yu, J. Hong, J.-Y. Fang, Emerging roles of hydrogen sulfide in inflammatory and neoplastic colonic diseases. *Front. Physiol.* **7**, 156 (2016).
31. Y. Yang, Y. Liao, Y. Ma, W. Gong, G. Zhu, The role of major virulence factors of AIEC involved in inflammatory bowl disease—A mini-review. *Appl. Microbiol. Biotechnol.* **101**, 7781–7787 (2017).
32. N. Vergnolle, Protease inhibition as new therapeutic strategy for GI diseases. *Gut* **65**, 1215–1224 (2016).
33. D. E. Wood, S. L. Salzberg, Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
34. E. A. Franzosa, L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, K. S. Lipson, R. Knight, J. G. Caporaso, N. Segata, C. Huttenhower, Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
35. A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K. Bhullar, M. J. Canova, G. De Pascale, L. Ejim, L. Kalan, A. M. King, K. Koteva, M. Morar, M. R. Mulvey, J. S. O'Brien, A. C. Pawlowski, L. J. V. Piddock, P. Spanogiannopoulos, A. D. Sutherland, I. Tang, P. L. Taylor, M. Thaker, W. Wang, M. Yan, T. Yu, G. D. Wright, The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
36. L. Chen, D. Zheng, B. Liu, J. Yang, Q. Jin, VFDB 2016: Hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).
37. S. Scholtens, N. Smidt, M. A. Swertz, S. J. L. Bakker, A. Dotinga, J. M. Vonk, F. van Dijk, S. K. R. van Zon, C. Wijmenga, B. H. R. Wolffenbuttel, R. P. Stolk, Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.* **44**, 1172–1180 (2015).
38. E. F. Tigchelaar, A. Zhernakova, J. A. M. Dekens, G. Hermes, A. Baranska, Z. Mujagic, M. A. Swertz, A. M. Muñoz, P. Deelen, M. C. Cénit, L. Franke, S. Scholtens, R. P. Stolk, C. Wijmenga, E. J. M. Feskens, Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: Study design and baseline characteristics. *BMJ Open* **5**, e006772 (2015).
39. Z. Mujagic, E. F. Tigchelaar, A. Zhernakova, T. Ludwig, J. Ramiro-Garcia, A. Baranska, M. A. Swertz, A. A. M. Masclee, C. Wijmenga, F. J. Van Schooten, A. Smolinska, D. M. A. E. Jonkers, A novel biomarker panel for irritable bowel syndrome and the application in the general population. *Sci. Rep.* **6**, 26420 (2016).
40. A. Ersryd, I. Posserud, H. Abrahamsson, M. Simrén, Subtyping the irritable bowel syndrome by predominant bowel habit: Rome II versus Rome III. *Aliment. Pharmacol. Ther.* **26**, 953–961 (2007).
41. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

42. I. Siddiqui, H. Majid, S. Abid, Update on clinical and research application of fecal biomarkers for gastrointestinal diseases. *World J. Gastrointest. Pharmacol. Ther.* **8**, 39–46 (2017).
43. G. T.-J. Huang, H.-B. Zhang, D. Kim, L. Liu, T. Ganz, A model for antimicrobial gene therapy: Demonstration of human β-defensin 2 antimicrobial activities in vivo. *Hum. Gene Ther.* **13**, 2017–2025 (2002).
44. S. K. Ames, S. N. Gardner, J. M. Marti, T. R. Slezak, M. B. Gokhale, J. E. Allen, Using populations of human and microbial genomes for organism detection in metagenomes. *Genome Res.* **25**, 1056–1067 (2015).
45. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
46. R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, P. D. Karp, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **44**, D471–D480 (2016).
47. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
48. J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, vegan: Community ecology package (2016).
49. W. Revelle, psych: Procedures for psychological, psychometric and personality research (2017).
50. R Core Team, R: A language and environment for statistical computing (2016).
51. X. C. Morgan, T. L. Tickle, H. Sokol, D. Gevers, K. L. Devaney, D. V. Ward, J. A. Reyes, S. A. Shah, N. LeLeiko, S. B. Snapper, A. Bousvaros, J. Korzenik, B. E. Sands, R. J. Xavier, C. Huttenhower, Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79 (2012).

Acknowledgments: We thank all of the participants of the LifeLines DEEP, the UMCG IBD cohort, and the Maastricht University Medical Center IBS cohort for contributing stool samples and phenotypes; D. Jansen, the IBD nurses at the UMCG IBD clinic; J. Arends, M. Platteel, A. Maatman, T. Poon, W. Westerhuis, M. Klaassen, L. Bolte, and M. Hesselink for logistics and laboratory support, data collection, and data management; R. Gómez Aldudo and S. Khan for the main figures; the research group of M. Swertz for providing the high-performance computing infrastructure including the Calculon cluster computer; and the Parelsnoer Institute for supporting the IBD biobank infrastructure. We also thank J. Senior and K. McIntyre for editing the manuscript. **Funding:** R.K.W., J.F., and L.F. are supported by VIDI grants (nos. 016.136.308, 864.13.013, and 917.14.374) from the Netherlands Organization for Scientific Research (NWO). R.K.W. is further supported by a Diagnostics Grant from the Dutch Digestive Foundation (no. D16-14). A.Z. holds a Rosalind Franklin fellowship from the University of Groningen. Sequencing of the control cohort was funded by a grant from the Netherlands' Top Institute Food and Nutrition grant no. GH001 to C.W., who is further supported by an ERC advanced grant (ERC-671274) and a Spinoza award (NWO SPI 92-266). A.Z. and L.F. are supported by ERC starting grants (ERC-715772 and ERC-637640). J.F. and A.Z. are supported by a CardioVasculair Onderzoek Nederland grant (CVON 2012-03). E.A.M.F. is supported by a Career Development grant from the Dutch Digestive Organization (MLDS; CDG-014). M.C.V. is supported by an AGIKO grant (92.003.577) from the Netherlands Organization for Scientific Research (NWO). Z.M. holds a Niels Stensen fellowship (from Amsterdam, The Netherlands).

Author contributions: R.K.W., E.J.A., A.Z., D.M.A.E.J., M.H.H., C.W., A.V.V., and F.I. designed the study. F.I., V.C., M.C.V., M.D.V., and V.P. collected and processed the phenotypic data from the UMCG IBD cohort. S.A.J., Z.M., and D.K. collected and processed the phenotypic data from the MIBS cohort. E.F.T., S.A.J., and A.Z. collected and processed the phenotypic data from the LifeLines DEEP cohort. G.D., E.A.M.F., H.M.v.D., and R.K.W. participated in the patient inclusion and the sample collection in the IBD cohort. Z.M. and A.A.M.M. participated in the patient inclusion and sample collection in the MIBS cohort. J.D. participated in the inclusion of individuals in the LifeLines DEEP cohort. A.V.V., J.F., and M.J.B. processed the metagenomic sequencing reads. A.K., V.C., and A.V.V. designed the prediction models. T.G. and X.J. performed the prediction of virulence factors, antibiotic resistance genes, and strain richness calculations. A.V.V., V.C., and A.K. performed the statistical analyses. F.I., T.G., V.C., and A.V.V. wrote the manuscript. S.A.J., Z.M., A.K., M.J.B., X.J., E.F.T., J.D., V.P., M.D.V., M.C.V., H.M.v.D., D.K., M.A.S., L.F., R.A., E.A.M.F., G.D., A.A.M.M., R.J.X., E.J.A., J.F., C.W., D.M.A.E.J., A.Z., and R.K.W. critically assessed the manuscript. **Competing interests:** F.I. has received speaker fees from AbbVie and was a shareholder of the health care IT company Aceso B.V. and of Floris Medical Holding B.V. G.D. declares unrestricted research grants from AbbVie, Takeda, and Ferring Pharmaceuticals, is on the advisory boards for Mundipharma and Pharmacosmos, and has received speakers' fees from Takeda and Janssen Pharmaceuticals. R.K.W. declares consulting work for Takeda. The other authors declare that they have no competing interests. **Data and materials availability:** The raw metagenomic sequencing data for the LifeLines DEEP cohort and the age and gender information per sample are available upon request from the European Genome-Phenome Archive (EGA; https://www.ebi.ac.uk/ega/home) at accession number EGAS00001001704. The raw metagenomic sequences for the Groningen cohort can be requested with the accession number EGAD00001004194. The raw metagenomic sequences for the Maastricht cohort are available from EGA with the accession number EGAS00001001924 under a material transfer agreement with the MIBS cohort Project Data Access Committee. Other phenotypic data can be requested from the LifeLines cohort study (https://lifelines.nl/lifelines-research/access-to-lifelines) following the standard protocol for data access.

Submitted 6 September 2017
Resubmitted 6 April 2018
Accepted 16 July 2018
Published 19 December 2018
10.1126/scitranslmed.aap8914

Citation: A. Vich Vila, F. Imhann, V. Collij, S. A. Jankipersadsing, T. Gurry, Z. Mujagic, A. Kurilshikov, M. J. Bonder, X. Jiang, E. F. Tigchelaar, J. Dekens, V. Peters, M. D. Voskuil, M. C. Visschedijk, H. M. van Dullemen, D. Keszthelyi, M. A. Swertz, L. Franke, R. Alberts, E. A. M. Festen, G. Dijkstra, A. A. M. Masclee, M. H. Hofker, R. J. Xavier, E. J. Alm, J. Fu, C. Wijmenga, D. M. A. E. Jonkers, A. Zhernakova, R. K. Weersma, Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Sci. Transl. Med.* **10**, eaap8914 (2018).