Distinct Microbiotas are Associated with Ileum-Restricted and Colon-Involving Crohn's Disease

Timna Naftali, MD, Leah Reshef, PhD,† Amir Kovacs, MSc,† Ron Porat, MD,* Itay Amir, PhD,† Fred M. Konikoff, MD, MSc,* and Uri Gophna, PhD†*

Background: The etiology of inflammatory bowel disease is believed to involve a shift in the microbiota toward more proinflammatory species. Crohn's disease (CD) usually manifests as one of three phenotypes, involving inflammation of the terminal ileum, the colon, or both. However, what determines the particular phenotype and the level of disease activity remains unknown. In this study, we aim to characterize the intestinal microbiota associated with different CD phenotypes.

Methods: DNA was extracted from biopsies of 31 patients with ileal, ileocolic, or colon-restricted CD, and also from 5 non–inflammatory bowel disease control subjects, and analyzed by 16S rRNA gene amplicon pyrosequencing. Data were processed using the Quantitative Insights Into Microbial Ecology pipeline and analyzed using linear discriminant analysis with effect size estimation and PICRUSt algorithms. Two additional recently published cohorts were also analyzed in this study.

Results: Highly significant separation was observed between bacterial composition of ileal CD compared with CD with colonic involvement (genus level Bray–Curtis $P = 0.005$, $R = 20\%$). This separation was unaffected by the biopsy's location or its inflammatory state, or by the patients' condition (remission or relapse). *Faecalibacterium* was strongly reduced in ileal CD compared with CD with colonic involvement, whereas *Enterobacteriaceae* were more abundant in the former. *Fusobacterium* relative abundance was strongly correlated with disease activity in patients with ileal-involving, but not in colon-involving, CD.

Conclusions: Ileal and colon-involving CD sustain distinct microbiotas, suggesting that different mechanisms underlie the two major manifestations of CD. The potential contribution of *Fusobacterium* to inflammation in ileal CD should be further investigated.

(Inflamm Bowel Dis 2016;22:293–302)

Key Words: Crohn's disease, microbiota, dysbiosis, MDI

Multiple lines of evidence support microbial involvement in the etiology of inflammatory bowel disease (IBD). Several rodent models of IBD, such as interleukin-10 knock-out mice or transgenic HLA-B27 rats, do not develop experimental colitis in the absence of microbiota.[1,2] The products of genes associated with IBD susceptibility are involved in recognition of bacterial motifs[3,4] or in host response against intracellular pathogens.[5,6] Microbiota analysis studies have repeatedly shown association of IBD with microbial dysbiosis, a shift in the intestinal microbiota toward more proinflammatory microbial species.[7–9] Dysbiosis in Crohn's Disease (CD) has been particularly well characterized. Multiple studies have shown a reduction in key taxa of the *Ruminococcaceae* and *Lachnospiriceae* families and an increase in the *Enterobacteriaceae* and *Fusobacteriaceae*.[10–12] A recent study that has analyzed a large cohort of new-onset patients with CD has defined a Microbial Dysbiosis Index (MDI), intended to measure the extent of dysbiosis in a given sample in an IBD context.[13]

CD may occur anywhere along the digestive tract, yet most patients display involvement restricted to certain parts. Up to one-half of patients have disease affecting both ileum and colon (icCD), about a third have involvement of the small intestine, primarily the terminal ileum (iCD), and others have isolated colonic disease (cCD). Involvement of the proximal gastrointestinal tract is much less common. Differences in serological,[14] clinical,[15] and genetic[16–18] markers between iCD and cCD have been demonstrated, yet the mechanisms underlying these different location phenotypes remain unknown. Most importantly, in a study correlating gene expression patterns and response to infliximab, the researchers noted a far better treatment response, defined as complete endoscopic and histologic healing within 6

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication August 9, 2015; Accepted October 7, 2015.

From the *Department of Gastroenterology and Hepatology, Meir Medical Center, Kfar Saba, Israel; and †Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.

Supported in part by a grant from the chief scientist of the Israeli Ministry of Health.

Fred M. Konikoff, MD, MSc, was supported in part by the The Josefina Maus and Gabriela Cesarman Chair for Reaserch in Liver Disease.

The authors have no conflicts of interest to disclose.

T. Naftali and L. Reshef contributed equally to this work.

Reprints: Uri Gophna, PhD, Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: urigo@tauex.tau.ac.il).

DOI 10.1097/MIB.0000000000000662

Published online 11 January 2016.

weeks, among patients with cCD (63%) than among those with iCD (5%). Infliximab response strongly correlated with gene expression in patients with cCD, but not in patients with iCD.[19]

A number of studies have pointed to specific microbial characteristics of iCD. Adherent Invasive *Escherichia coli* strains, common in patients with CD, are especially prevalent in iCD[20] whereas *Faecalibacterium* is markedly reduced.[12] Willing et al. (2010) showed considerable difference between microbiotas of iCD and cCD in a twin cohort of patients in remission. Nevertheless, in practice, many researchers and physicians still regard and treat different CD phenotypes as if they were the same disease.

In this article, we characterize the microbiota associated with CD location phenotypes. We analyze inflamed and healthy biopsies, taken from dual locations, to assess the relationship between local inflammation, microbial composition, and the underlying CD phenotype. Both active and quiescent patients were included to enable investigation of disease activity effects within different location phenotypes.

MATERIALS AND METHODS

Patients and Controls

Thirty-one patients with CD treated at the Department of Gastroenterology and Hepatology at Meir Medical Center participated in this study. Biopsy samples were collected from the colon, terminal ileum (TI), or both, during routine endoscopy as part of the patients' follow-up. Crohn's Activity Disease Index (CDAI) was used to assess disease activity at the time of sampling. Additionally, biopsy samples were taken from the colon of five healthy control subjects undergoing colorectal cancer surveillance endoscopy. This cohort of 36 subjects is hereafter called as the "MEIR" cohort. All patients gave informed consent and the study was approved by the local ethics committee. Determination of the CD location phenotype was based on the detection of macroscopic inflammation during endoscopy. With the exception of 2 newly diagnosed patients, all participants had established disease; the duration since diagnosis ranged from 3 to 26 years. None of the patients had received antibiotic treatment in the month proceeding sampling; however, 21 patients received 5-aminosalicylic acid, steroid or azathioprine treatments (Table 1). Clinical evaluation data per patient are summarized in Supplementary Table S1 (Supplemental Digital Content, http://links.lww.com/IBD/B178).

Sample Collection

During endoscopy, clean biopsy forceps that did not touch formalin were used. Biopsies were taken from the TI, right, transverse, and left colon. If a certain segment had inflamed and not inflamed areas, the biopsies were taken from the inflamed area. Biopsies were immediately frozen at −20°C.

Sample Processing

Biopsies were thawed at room temperature and DNA was extracted using the Mo-Bio PowerSoil DNA Isolation Kit (Mo-Bio, Carlsbad, CA). Samples were homogenized using the Mini-Beadbeater (BioSpec Products, Bartlesville, OK) at maximum speed for 1 minute and then processed according to the manufacturer's protocol. Extracted DNA samples were stored at −20°C. The presence of bacterial DNA was validated by PCR, with the following primers: 27F 5'-AGAGTTTGATCMTGGCTCAG-3', 338R 5'-GCTGCCTCCCGTAGGAGT-3'.

16srRNA Pyrosequencing and Analysis

The 16S region ranging from position 28 to 519 was sequenced (Research and Testing Laboratory, Lubbock, TX) by the bacterial 16S-based tag-encoded FLX amplicon pyrosequencing method, using primers 28F 5'-GAGTTTGATCNTGGCTCAG-3', 519R 5'-GTNTTACNGCGGCKGCTG-3'. Average sequencing depth was 3163 seqs/sample. The resulting sequences

TABLE 1. Patient Characteristics

	Ileal Crohn's (iCD)	Ileocolic Crohn's (icCD)	Crohn's colitis (cCD)	All
Sex (female/male)	6/9	2/6	5/3	13/18
Age (yr, average ± SD)	39 ± 12.5	46 ± 16.8	37 ± 5.9	40.6 ± 8
Active/inactive[a]	7/8	2/6	4/4	13/18
CDAI (average, range)	106, 34–272	94, 28–199	146, 35–214	113, 28–272
Medication at the time of sampling				
Thiopurines	2	2	0	4
5-aminosalicylic acid	5	3	3	11
Methotrexate	1	0	0	1
5-aminosalicylic acid + thiopurines	1	0	0	1
Steroids + thiopurines	1	1	1	3
None	5	2	4	11

[a]Active disease defined as CDAI >150.

were processed by the Quantitative Insights Into Microbial Ecology (QIIME) package.[21] UCLUST was used to pick operational taxonomic units at a 0.97 similarity level, and taxonomy was assigned by BLAST (E-value < 0.001) using Greengenes database of February 2011 as a reference. Chimeric sequences were identified by ChimeraSlayer and removed. All sequences are freely accessible through the MG-RAST website, project ID: 15781.

Statistical Analysis

UniFrac distances were calculated using QIIME.[21] The analysis of similarity between groups (ANOSIM) probability test, Shannon index of diversity, and the Bray–Curtis similarity index were calculated using PAST.[22] The non-parametric Mann–Whitney probability test (2-tailed) was used for 2-group nonpaired comparisons. Linear discriminant analysis with effect size estimation (LEfSe)[23] was performed on genus level OTU tables using the on-line analysis tool available from http://huttenhower.sph.harvard.edu/galaxy/. The LEfSe algorithm allows identification of taxa that most strongly differentiate 1 group from another (i.e., biomarkers). A P value <0.05 was considered significant. Spearman's correlation method was used to correlate genus or family level OTU tables to clinical parameters. Taxa which were present in less than 10% of samples or taxa whose abundance across all samples was below 0.5% were removed before performing correlations. For functional analysis, the PICRUSt[24] algorithm was used to predict gene abundances associated with each sample. The HUMAnN[25] package was then used to summarize gene abundances into KEGG modules. Finally, LEfSe was again used to compare predicted KEGG modules of iCD samples to those of colon-involving CD (both icCD and cCD). Whenever necessary, multiple hypotheses were corrected using Benjamini and Hochberg's false discovery rate[26] method.

MDI Calculation

MDI is based on two groups of taxa that were previously found to be over or underrepresented in patients with CD when compared with normal controls.[13] Taxa in the first group (increased in CD) are *Enterobacteriaceae*, *Pasteurellaceae*, *Neisseriaceae*, *Gemellaceae*, *Fusobacteriaceae*, and *Veillonellaceae*. Taxa in the second (decreased in CD) group are *Erysiopelotrichaceae*, *Bifidobacteriaceae*, all Bacteroidales, and all Clostridiales excluding the *Veillonellaceae*. MDI is calculated using the following formula: log {(sum relative abundances [RAs] of taxa increased in CD)/(sum RAs of taxa decreased in CD)}.

RESULTS

The MEIR cohort includes thirty-one patients with CD, presenting with ileal Crohn's (iCD, n = 15), colon-restricted Crohn's (cCD, n = 8), and ileocolic Crohn's (icCD, n = 8) (Table 1). In 11 patients (5 with iCD, 4 with cCD, and 2 with icCD), the disease was active at the time of sampling, with an average CDAI of 199. Samples of both healthy and inflamed tissue were collected from eleven patients, and samples from both ileum and colon were collected from twelve patients. Colon biopsy samples were also collected from 5 control subjects with no signs of IBD. In all, 60 biopsy samples were analyzed by 16S pyrosequencing. Average sequence depth was 3163 seqs/sample; rarefaction analysis shows most of the samples approach saturation at a depth of 1000 seqs/sample (Supplementary Fig. S2, Supplemental Digital Content, http://links.lww.com/IBD/B178). None of the patients had been exposed to antibiotics for at least 4 weeks before sampling, yet 35% of them did receive nonantibiotic medications (detailed in Table 1). However, in our cohort, we detected no significant effects of 5-aminosalicylic acid, thiopurine, or steroid medication on the microbiota (Supplementary Fig. S3, Supplemental Digital Content, http://links.lww.com/IBD/B178).

Microbial Composition of Patients with Colon-Involving CD Differs from that of Patients with Ileal CD

Principal coordinate analysis using Bray–Curtis index showed that the microbiota of biopsy samples in this study (n = 53) cluster according to the CD phenotype: samples from patients with colon-involving CD (either cCD or icCD) formed a cluster separate from those of iCD (Fig. 1A, ANOSIM P value: 0.0001, R = 20%). Two patients emerged as "outliers" (marked by arrows in Fig. 1); notably, both these patients had icCD and seemed as outliers in every analysis conducted. Similar results were obtained using weighted ($P = 0.0003$, R = 15.4%) and unweighted ($P = 0.0001$, R = 17.7%) Unifrac distance matrices. Since in many cases sequence data were obtained from more than 1 sample per patient, data had to be analyzed in a manner that avoids bias. We thus prepared a smaller data set composed exclusively of healthy colon tissue samples, one sample per patient. Seventy-five percentages of the patients were represented in this data set. The separation between iCD and c/icCD (colon-involving CD) remained distinct when using Bray–Curtis (Fig. 1B, $P = 0.005$, R = 20%) and weighted Unifrac ($P = 0.02$, R = 14%), but not with unweighted Unifrac ($P = 0.11$, R = 6%). As unweighted Unifrac relies on presence or absence of taxa only, whereas weighted Unifrac is also affected by each taxa's RA, this implies that iCD and c/icCD differ mostly in the RA of certain key taxa.

Both Site and Inflammatory State of Biopsies Have No Effect on Microbial Composition

We find that biopsy samples taken from the same individual tend to cluster together regardless of their site of origin (colon or TI) or inflammatory state (inflamed or normal tissues); the effect of CD phenotype (iCD or colon-involving) on clustering is far stronger than any within-patient variation (Fig. 2). This is in keeping with previous studies showing consistent microbial composition from ileum to rectum.[12,13] Although most microbiome studies also report little or no effect of local tissue inflammation on the microbiota, this aspect remains controversial.[27,28] Because this issue is vital for elucidating the role of bacteria in IBD, we investigated it extensively using several approaches (Supplementary Figures S4–S6, Supplemental Digital Content, http://links.lww.com/IBD/B178). Principal coordinate analysis and ANOSIM

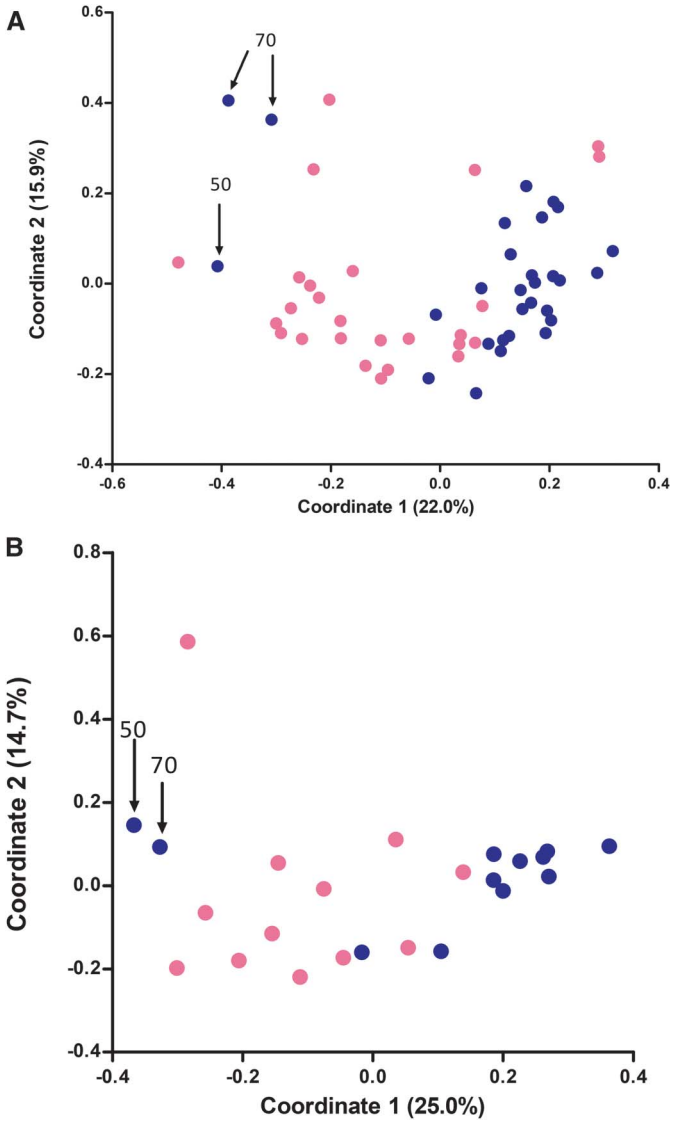

FIGURE 1. Principal coordinate analysis of samples from iCD (purple) versus colon-involving (both cCD and icCD, blue) patients. Bray–Curtis similarities (genus level) were used. A, All samples in the study (n = 53). B, One sample of healthy colon tissue per patient (n = 23). Individual patients with outlier microbiotas are marked with their patient number.

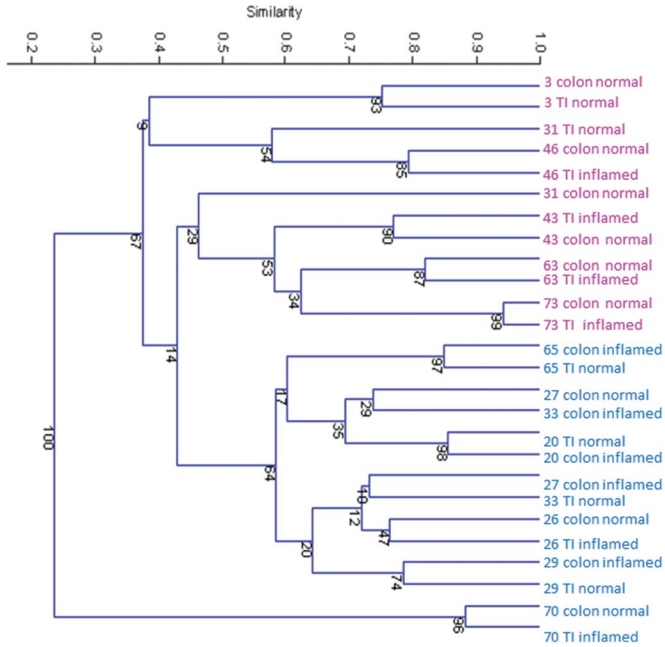

FIGURE 2. Clustering of paired samples. Samples from patients with iCD are in pink; samples from colon-involving (both cCD and icCD) patients in blue. Numbers on leaves denote specific patients; colon/TI refers to the biopsy origin and normal/inflamed refers to the biopsy's inflammatory state. Numbers on branches represent % bootstrap support.

were used to compare the overall similarity between inflamed and normal tissue samples; LEfSe was used to search for specific taxa, which may discriminate inflamed tissues from normal ones; and Wilcoxon matched-pairs test was used to look for differences in abundance of key specific taxa, previously shown to be inflammation-associated, between inflamed and normal biopsies obtained from the same patient (n = 10). To account for possible confounding effects of location phenotype or disease activity, whenever possible those tests were performed across the entire cohort and on subsets stratified by disease location or by activity. Importantly, none of these approaches detected robust statistical associations between microbiota and local tissue inflammation.

Since in our data, neither biopsy site nor biopsy inflammatory status had a significant effect on sample distribution, in the forthcoming analyses 1 sample per patient was arbitrarily chosen.

Specific Taxa Differentiating Between CD Location Phenotypes

To identify which taxa account for the difference between iCD and colon-involving (cCD and icCD) CD, we applied LEfSe,[23] an algorithm that identifies taxa differentially distributed between 2 groups. We found that iCD samples were richer in *Escherichia*, whereas colon-involving CD had higher levels of *Faecalibacterium* and 2 unidentified genera of the Clostridiales and Ruminococceaea (Supplementary Fig. S7, Supplemental Digital Content, http://links.lww.com/IBD/B178). As a reduction in *Faecalibacterium* is considered a hallmark of all kinds of IBD, it was of interest to compare its level in both CD groups to that of control patients with non-IBD. Median abundance of *Faecalibacterium* in the controls was 0.42, significantly higher than that of both colon-involving CD (0.16, $P = 0.02$) and iCD (0.002, $P = 0.01$), supporting the view that a reduction in *Faecalibacterium* is associated with all types of CD. To investigate whether *Faecalibacterium* abundance was related to local inflammation, we focused on its RA exclusively in biopsies taken from the colon. Colon biopsies of patients with iCD

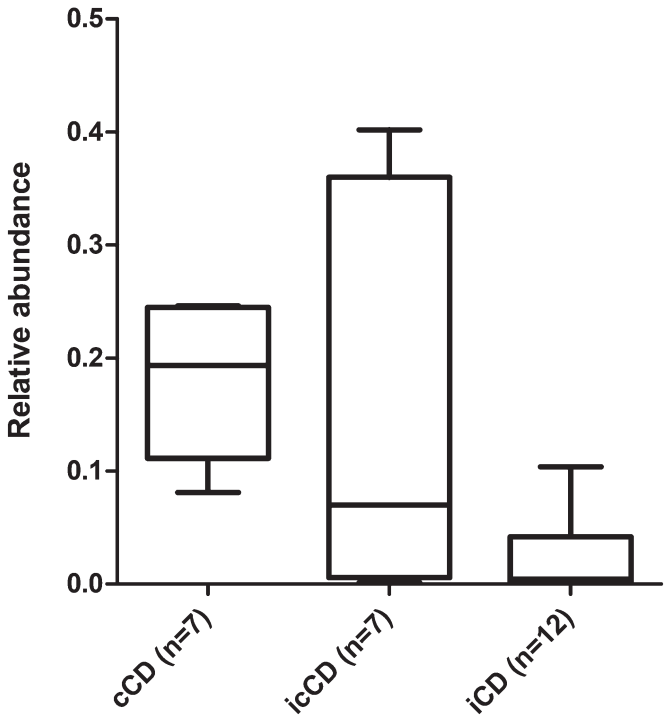

FIGURE 3. RA of *Faecalibacterium* across the different CD phenotypes. Each patient is represented by a single colon biopsy. Boxes represent the second and third quartiles, the inner line being the median, and whiskers show the 10th and 90th percentiles.

are, by definition, noninflamed; yet, *Faecalibacterium* was almost entirely absent in those samples (median = 0.004, Fig. 3). Conversely, 85% of colon biopsies from patients with cCD were of inflamed tissue, but the RA of *Faecalibacterium* in those samples was more than 20-fold higher than in iCD (median = 0.19). *Faecalibacterium* abundance in colon biopsies from patients with icCD, 14% of which were inflamed was significantly ($P = 0.02$) higher than in patients with iCD, yet still lower than in cCD; this difference, however, was not statistically significant ($P = 0.3$).

Functional Profiling of Microbiotas Associated with iCD and Colon-Involving Disease

The gene content of most human gut microbes is either known or can be inferred through comparison with closely related species. Thus, taxonomic composition may also be roughly translated to functional, mostly metabolic capabilities, using the PICRUSt[24] algorithm, which predicts gene abundances that are associated with each sample. Gene abundance predictions were collapsed into KEGG modules using the HUMAnN[25] package, and LEfSe was again used to compare the inferred functional profiles of samples from patients with iCD to those with colon-involving CD. Forty-one KEGG metabolic modules were identified as discriminating by the LEfSe tool (Supplementary Table S8, Supplemental Digital Content, http://links.lww.com/IBD/B178). To focus on the most meaningful of those, we calculated for each module the ratio between its average RA in the iCD group and the average RA in the colon-involving group, thereafter concentrating on those modules for which the ratio was larger than 2 or lower than 0.5 (Table 2). Notably, nineteen of the iCD-enriched modules fell within this range, but none of the features enriched in colon-involving CD did—the lowest [iCD]/[colon-involving CD] average RA ratio was 0.69. In keeping with previously published data,[12] iCD microbiotas were primarily characterized by an overrepresentation of amino acid transporters, especially of histidine, lysine, and arginine; by sugar transporters, especially the PTS system; by fermentation/respiration linked pathways and by modules which may be directly attributed to the overrepresentation of the Gram-negative *Escherichia* in iCD; these include LPS biosynthesis and AI-2 quorum sensing signaling. In keeping with the large number of functions overrepresented in iCD, we found that a slightly higher diversity of metabolic capabilities was associated with the iCD samples than with the colon-involving ones (Supplementary Fig. S9, Supplemental Digital Content, http://links.lww.com/IBD/B178). Interestingly, no difference in alpha-diversity between ileal and colon-involving CD was identified when analyzing OTU RA tables (which contain taxonomic rather than functional information), suggesting that the iCD-associated microbiotas display a particularly wide range of metabolic functions.

Although the main aim of this study was to compare CD phenotypes, it was also of interest to see to what extent the phenotype-discriminating features are represented in non-IBD controls. To this end, we calculated also the (iCD)/(healthy) and (colon-involving CD)/(healthy) average RA ratios for all LEfSe-identified phenotype-distinguishing features. As might be expected, all metabolic features distinguishing iCD from colon-involving CD were also considerably underrepresented in healthy controls. Of note, many (52%) of these features were also increased in colon-involving CD relative to controls, but to a lesser degree (see italic values in Table 2). This is in agreement with previously published data showing metabolic perturbation, relative to controls, exists in non-ileal and ileal CD.[12]

Ileal Disease is Associated with a Higher MDI

The MDI is calculated as the log of (total abundance of taxa increased in CD/total abundance of taxa decreased in CD)[13] (for a complete list of taxa, see Materials and Methods). Higher MDI values reflect stronger dysbiosis. When applying MDI calculations to our data, we observed that MDI values of samples from iCD are significantly higher than those of colon-involving patients (median iCD = −0.7, n = 15, median c/icCD = −1.7, n = 16, $P = 0.003$). Within the colon-involving group, patients with icCD and cCD had similar MDI values (median icCD = −1.6, n = 8; median cCD = −1.7, n = 8; $P = 0.8$), illustrating the validity of grouping those two phenotypes together. Notably, the same two patients with icCD who seemed as outliers in Bray–Curtis and Unifrac analyzes (Fig. 1) also had extremely high MDI values.

TABLE 2. Predicted Functional Features Distinguishing Microbiota of iCD From Colon-Involving CD

	P (FDR-Corrected)	Ratio (iCD)/(Colon-Involving CD)	Ratio (iCD)/(non-IBD)	Ratio (Colon-Involving CD)/(Non-IBD)
Biosynthesis				
M00124: pyridoxal biosynthesis, erythrose-4P → pyridoxal-5P	0.0266	3.0067	2.3038	0.7662
M00019: leucine biosynthesis, pyruvate → 2-oxoisovalerate → leucine	0.0331	2.5076	2.2990	0.9168
M00096: C5 isoprenoid biosynthesis, non-mevalonate pathway	0.0219	2.5190	2.3260	0.9234
Glycolysis/respiration related				
M00008: Entner–Doudoroff pathway, glucose-6P → glyceraldehyde-3P + pyruvate	0.0323	2.5909	3.5543	1.3718
M00117: ubiquinone biosynthesis, prokaryotes, chorismate → ubiquinone	0.0266	2.6212	4.9626	*1.8933*
M00150: complex II (succinate dehydrogenase/fumarate reductase), fumarate reductase	0.0460	2.4231	4.6754	*1.9295*
Sugar transport systems				
M00287: PTS system, galactosamine-specific II component	0.0331	2.8068	34.9204	*12.4415*
M00198: sn-glycerol 3-phosphate transport system	0.1343	2.0138	2.7451	1.3632
Amino acid transport systems				
M00229: arginine transport system	0.0457	2.5853	5.6775	*2.1961*
M00225: lysine/arginine/ornithine transport system	0.0266	2.9285	5.4130	*1.8484*
M00226: histidine transport system	0.0323	2.9480	5.6475	*1.9157*
M00336: twin-arginine translocation system	0.0323	3.1834	4.9988	1.5703
Other transport systems				
M00317: manganese/iron transport system	0.0339	4.7065	5.7057	1.2123
M00349: microcin C transport system	0.0460	2.6746	5.6492	*2.1121*
M00324: dipeptide transport system	0.0323	2.6882	5.9554	*2.2154*
M00300: putrescine transport system	0.0914	2.3426	6.0022	*2.5622*
Gram-negative associated modules				
M00060: lipopolysaccharide biosynthesis, KDO2-lipid A	0.1074	2.0076	1.8625	0.9277
M00219: AI-2 transport system	0.1000	2.0046	8.4703	*4.2254*
Other				
M00335: Sec (secretion) system	0.0295	2.8030	3.3427	1.1925
M00260: DNA polymerase III complex, bacteria	0.0295	2.9335	3.5402	1.2068

Ratios obtained by comparing the average relative abundances of each feature in each group.
Italics denote a greater than 2-fold difference in relative abundance between colon-involving CD and non-IBD.
FDR, false discovery rate.

Association of Disease Activity and Microbiota

To investigate correlations between disease activity and microbiota composition, we used the CDAI, a numerical index summarizing the severity of symptoms presented by the patient. One patient with iCD, for whom the CDAI on the day of sampling was absent, was excluded from this analysis. The MDI values of either the entire patient cohort (n = 30) or the colon-involving patients alone (n = 16) showed no association with CDAI (Spearman's r = −0.01, $P = 0.9$; r = −0.03, $P = 0.88$, respectively). Conversely, MDIs of patients with iCD alone (n = 14) showed a strong positive correlation with CDAI (Spearman's r = 0.43, Fig. 4). This correlation however did not reach statistical significance ($P = 0.12$). Notably, a single patient in the iCD cohort, marked by an arrow in Figure 4, seemed as an extremely strong outlier. This patient was found to have consumed large quantities (1 cup per day) of nonpasteurized camel's milk, which is known for its high content of secretary antibodies and unique oligosaccharide composition,[29] for several months before sampling. No other subject in this cohort reported a similar dietary habit. If this patient is excluded from analysis, the correlation becomes significant ($P = 0.02$, r = 0.6).

Because microbiota composition seemed to be correlated with disease activity in iCD, but not in colon-involving CD, it was of interest to see if any specific bacterial taxa correlate with CDAI in those 2 groups. Using a nonbiased approach, we examined

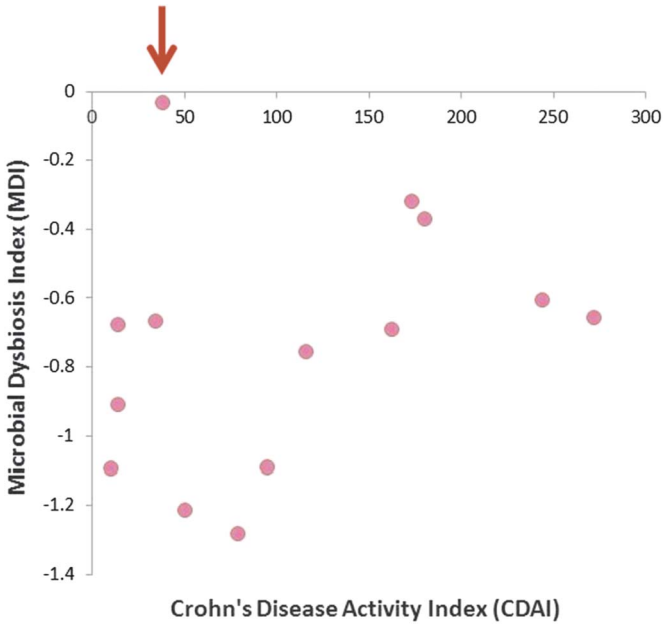

FIGURE 4. Correlation between MDI and CDAI, among patients with iCD only. A single outlier patient who was found to have a unique dietary habit is marked with red arrow.

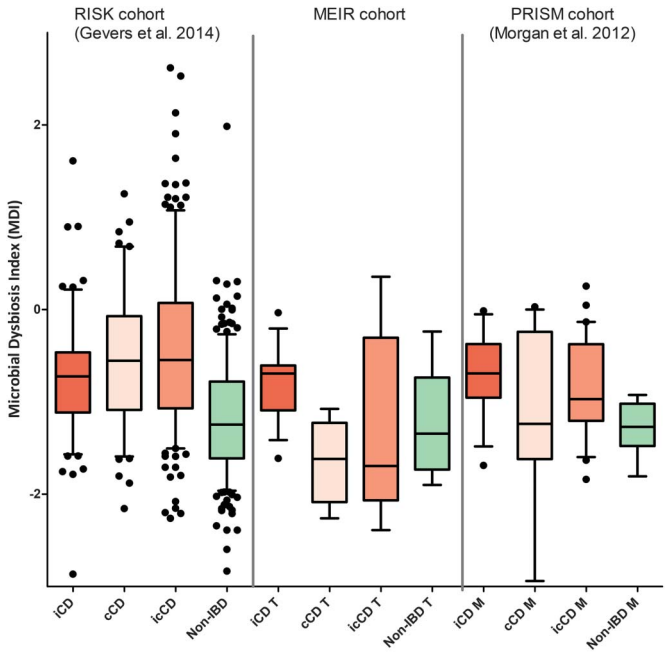

FIGURE 5. MDI values of iCD are significantly higher than of colon-involving CD in both MEIR and PRISM adult cohorts, but not in the new-onset pediatric RISK cohort. Boxes represent the second and third quartiles, the inner line being the median, and Whiskers show the 10th and 90th percentiles. For MDI, see also Materials and Methods.

correlations of CDAI to all genera present in the data set, excluding rare taxa that were filtered as described in Methods. *P* values were corrected by the Benjamini and Hochberg false discovery rate method[26] to account for multiple hypothesis testing. In the patient with iCD group, we found a strong and significant positive correlation between *Fusobacterium* and CDAI (false discovery rate–corrected $P = 0.038$, $R = 0.59$). Conversely, in the colon-involving CD data set, no taxa were significantly correlated with CDAI.

MDI Analysis in the Recently Published RISK and PRISM Cohorts

We next wished to investigate whether the trends we have observed were also evident in other previously studied cohorts. Direct comparisons between different sequencing projects are problematic because study-specific biases are almost inevitably introduced during the processing of samples, amplification, and sequencing. However, as the MDI comprised a ratio between the RAs of different taxa, we reasoned it may be used to compare different studies, even if conducted using different sequencing methods and primer sets. Two large sequencing projects analyzing samples of patients with IBD have been published in the last few years. The most recent of these, Gevers et al, 2014, had analyzed a new-onset pediatric cohort (RISK), whereas the PRISM cohort, analyzed by Morgan et al, 2012, comprised adult patients, mainly with established disease, resembling our own cohort. To avoid confounding effects, we used only the data derived from biopsy samples taken from patients who had not been exposed to antibiotics, similar to our own (MEIR) cohort. The MDI's of patients with iCD and of non-IBD controls were remarkably constant across the three studies (median MDI values in RISK, PRISM, and MEIR cohorts were −0.723, −0.691, −0.692, respectively, for iCD; and −1.25, −1.27, −1.34, for non-IBD). The MDIs of icCD, and especially of cCD, in the PRISM cohort were far lower than in iCD and were in fact indistinguishable from that of the control group (Fig. 5). This trend mirrors the one we had observed in our own cohort; however, owing to high interindividual variability observed in the PRISM cohort, it did not reach statistical significance in that data set. Additionally, iCD in the PRISM cohort displayed decreased *Faecalibacterium* and increased *Enterobacteriaceae*, relative to cCD, as expected by our results and others (Supplementary Fig. S10, Supplemental Digital Content, http://links.lww.com/IBD/B178). Conversely, the MDIs of icCD and cCD in the RISK new-onset pediatric cohort showed an opposite trend and were similar to, or higher than, those of iCD. Remarkably, even the most well-established characteristics of dysbiosis in iCD, such as increased *Enterobacteriaceae* and decreased *Faecalibacterium*, which had been shown in multiple studies,[10,20,30] were not observed in the RISK cohort (Supplementary Fig. S11, Supplemental Digital Content, http://links.lww.com/IBD/B178). Thus, phenotype-related microbial patterns may be characteristic of adult-onset disease or may develop during years of illness and treatments, or both.

We then went on to assess the association of MDI and disease activity in the PRISM cohort. Similar to the trends we had observed in the MEIR cohort, we found no overall difference in

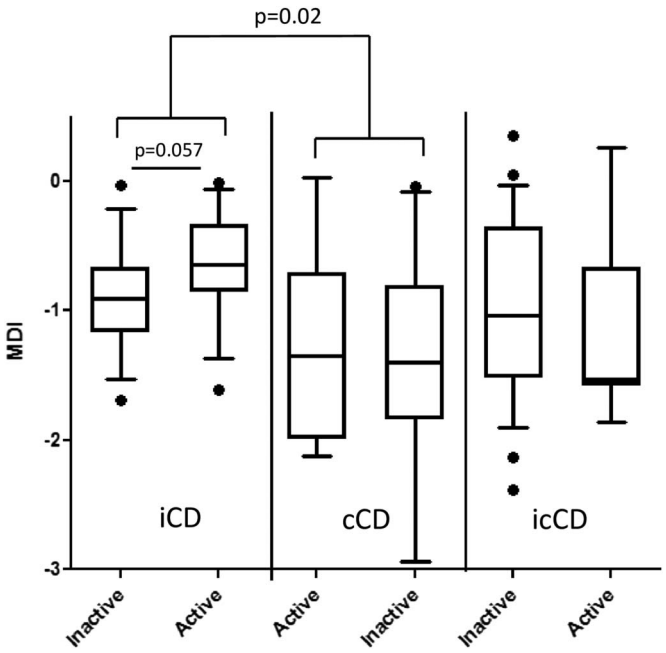

FIGURE 6. A slight increase in MDI is associated with active disease in iCD, but not in colon-involving CD (either cCD or icCD). Indexes shown are the combined values for both MEIR and PRISM cohorts. Boxes represent the second and third quartiles, the inner line being the median, and Whiskers show the 10th and 90th percentiles.

the MDIs of samples from active and inactive patients across the entire cohort, but did observe a small increase in MDI of samples from active patients compared with inactive ones, within the iCD group alone. However, splitting the PRISM iCD group (n = 13) by activity resulted in subsets too small to analyze with statistical probability tests. We thus combined the MDIs of samples from the PRISM and MEIR cohorts (Fig. 6). In this combined data set, samples from patients with active iCD had higher MDI values than from patients with inactive iCD ($P = 0.058$), whereas the MDI of samples from patients with active cCD or icCD was equal to, or lower than, that of samples from inactive patients.

DISCUSSION

With the progression of deep sequencing techniques, our ability to detect and characterize microbial communities has greatly improved. Yet, our understanding of the causes and effects of microbial dysbiosis in IBD remains incomplete. The simplest perception of dysbiosis in IBD suggests that it reflects nothing more than colonization of an inflammatory niche by bacteria best suited for it. This viewpoint is challenged by data showing dysbiosis also exists in remission-phase patients and in non-inflamed tissues.[11,31] Furthermore, we find that colon biopsies taken from patients with iCD carry the microbial signature typical to iCD, although the colon of patients with iCD is entirely free of inflammation. Likewise, biopsies taken from the TI of patients with cCD show dysbiosis characteristic to colon-involving CD, although the TI of those patients is not involved in inflammation. This point is further illustrated by the abundance pattern of the putatively protective genus *Faecalibacterium*. Although *Faecalibacterium* was reduced in all patients with CD relative to non-IBD controls, its abundance in colon biopsies of patients with cCD, most of which were inflamed, was over 20-fold higher than in colon biopsies of iCD, none of which were inflamed. It thus seems that microbial dysbiosis is inherent to the disease itself and reflects underlying mechanistic events. As iCD and cCD are associated with different forms of dysbiosis, this suggests different mechanisms underlie each of these 2 CD phenotypes.

Our comparative analyses of our data with 2 additional cohorts provided additional insights. In the PRISM cohort, as in ours, extreme dysbiosis was evident in iCD, whereas the microbiota of cCD was fairly close to normal, non-IBD, controls. Conversely, in the RISK new-onset pediatric cohort, all three CD phenotypes displayed similar extreme dysbiosis. Future studies are needed to reveal whether this phenomenon is due to mechanistic differences between pediatric and adult CD, or whether all CD cases start out with extreme dysbiosis, but the dysbiosis of cCD gradually declines over time. Elucidating this point may greatly further our understanding of CD development. Of note, evidence that microbial changes in IBD are independent of local mucosal inflammation has also been recently put forward in another study of new-onset pediatric IBD, performed on the RISK cohort, which had identified CD-unique gene expression and microbial patterns in tissues taken from noninflamed ileal mucosa.[32]

Although iCD and cCD form well-defined, discrete phenotypes, the classification of icCD is more challenging. Genetic studies of CD susceptibility markers indicate that for some markers, icCD performs as iCD, whereas for others, it appears as cCD.[16,17] The strength of genetic associations with icCD is often weaker than with the other CD phenotypes.[18] In our study, icCD strongly clustered with cCD and both phenotypes shared similar MDI values. Yet the 2 phenotypes were not identical, patients with icCD seemed more variable than colon-restricted ones, and *Faecalibacterium* abundance in patients with icCD was 2.7-fold lower than in colon-restricted ones (this difference however did not reach statistical significance). Thus in many regards, the icCD phenotype seems intermediate between iCD and cCD. Importantly, the prevalence and characteristics of icCD seem to vary between different nationalities and living styles. The "accepted" ratios for the three main CD location phenotypes are 50% ileocolic, 20% iCD, and 30% cCD. Conversely, CD epidemiology studies conducted in Israel throughout the past 20 years repeatedly report high (50%) prevalence of iCD, whereas ileocolic CD only accounts for approximately 25% of the cases.[33,34] Whereas iCD and cCD exhibited remarkably consistent microbial patterns in both the USA-based PRISM cohort and our Israeli one, ileocolic CD in the PRISM cohort displayed stronger dysbiosis than in our cohort; this was evident as a particularly low *Faecalibacterium* RA and as MDI values that were intermediate between the low MDI values of cCD and the high MDI values of iCD.

A crucial point to our understanding of the bacterial role in IBD is what, if any, changes occur in the microbiota during shifts

from remission to active disease. A number of studies have pointed to increased instability and variability of the microbiota during disease exacerbation.[35] However, owing to large heterogeneity between subjects and between studies, few uniform, reproducible associations between microbiota and disease activity have been described. In this study, we find that an association between the MDI and disease activity exists among patients with iCD with established disease. Importantly, this signal disappears if patients with iCD are analyzed together with colon-involving (cCD/icCD) ones. Furthermore, *Fusobacterium* in our data was strongly associated with disease activity in patients with iCD, but showed no correlation to disease activity in colon-involving patients. Future analyses of larger cohorts may uncover novel associations of microbiota and disease activity in cCD/icCD as well. However, we suggest that iCD and colon-involving CD may display different patterns of microbial association with disease activity and should thus preferentially be analyzed as separate groups. Mixing those 2 groups during analysis has an effect of "diluting" the signal, rendering identification of significant trends more difficult. This setback would be most influential in small-to-medium patient cohorts. In this regard, it is not surprising that the study which had most convincingly pointed to associations between microbial composition and disease activity[13] had been conducted on an especially large cohort (447 subjects).

The benefits of structuring cohorts according to CD phenotype are not necessarily restricted to microbial studies. It has long been known that the NOD2 mutation, the strongest genetic marker linked to CD, has no association with cCD[16]; conversely, HLA-DRB*0103 is strongly associated with colon-restricted disease,[17,18] but not with iCD. In a recently published study, JAK2 has been shown to be associated with ileum involving disease only.[36] Yet, most recent large-scale GWAS, gene expression, and proteomic studies do not stratify cohorts according to disease locations.[37–41]

One of the challenges in IBD microbial research lies in successful integration of highly variable data presented across many articles. Data variability stems not only from the high heterogeneity inherent to IBD cohorts but also from the use of varied processing, sequencing, and bioinformatic research techniques. The recently proposed MDI index[13] has the potential to serve as a standard, facilitating comparisons between different studies. Constructed on the order and family levels, it will not provide high sensitivity; but it should be fairly robust, allowing the identification of large-scale, overarching trends. In this study, we provide a first example (insofar as we know) of MDI usage. Our analysis of MDI values closely followed the trends shown by more traditional analysis methods, and, for the most part, we found remarkably similar MDI values across different cohorts, supporting the inclusion of this index in future microbiota articles.

We had demonstrated a strong association between *Fusobacterium* and disease activity in iCD. A number of recent studies have shown association of these taxa with various pathological conditions. These include not only CD,[13] particularly iCD,[11] but also ulcerative colitis,[42] pouchitis,[43,44] and colorectal cancer.[45] A culture-based study found bacteria belonging to the genus *Fusobacterium* to be more abundant in patients with active IBD compared with healthy non-IBD controls, and the strains isolated from patients with IBD showed higher cell-invasion capabilities than those isolated from the controls.[46] Thus, these bacteria have the potential to contribute to inflammation in patients with iCD.

Two treatment strategies are currently used in IBD treatment. The conventional approach targets the immune system, either by use of immunomodulatory drugs or by blocking proinflammatory signals such as TNF-alpha. An alternative approach attempts to manipulate the microbiota, whether by antibiotics, probiotics, or fecal matter transfer. The outcome of such microbiota-based manipulations probably depends on the initial microbial composition. Given that we and others have shown striking differences in composition between different IBD phenotypes, we propose that different treatment protocols should be considered for microbial manipulations of ileal and of colon-involving CD.

REFERENCES

1. Sellon RK, Tonkonogy S, Schultz M, et al. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infect Immun.* 1998;66:5224–5231.
2. Taurog JD. The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. *J Exp Med.* 1994;180:2359–2364.
3. Franchimont D. Deficient host-bacteria interactions in inflammatory bowel disease? the toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. *Gut*. 2004;53:987–992.
4. Bonen DK, Ogura Y, Nicolae DL, et al. Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology*. 2003;124:140–146.
5. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet.* 2007;39:207–211.
6. Yamazaki K, Onouchi Y, Takazoe M, et al. Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients. *J Hum Genet.* 2007;52:575–583.
7. Tamboli C, Neut C, Desreumaux P, et al. Dysbiosis as a prerequisite for IBD. *Gut*. 2004;53:1057–1057.
8. Gophna U, Sommerfeld K, Gophna S, et al. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J Clin Microbiol.* 2006;44:4136–4141.
9. Frank DN, Robertson CE, Hamm CM, et al. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis.* 2011;17:179–184.
10. Martinez-Medina M. Molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease. *Inflamm Bowel Dis.* 2009;15:872–882.
11. Willing BP, Dicksved J, Halfvarson J, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*. 2010;139:1844–1854.e1.
12. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 2012;13:R79.
13. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe.* 2014;15:382–392.
14. Ferrante M, Henckaerts L, Joossens M, et al. New serological markers in inflammatory bowel disease are associated with complicated disease behaviour. *Gut*. 2007;56:1394–1403.

15. Hancock L, Beckly J, Geremia A, et al. Clinical and molecular characteristics of isolated colonic Crohn's disease. *Inflamm Bowel Dis.* 2008;14:1667–1677.
16. Ahmad T, Armuzzi A, Bunce M, et al. The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology*. 2002;122: 854–866.
17. Silverberg MS, Mirea L, Bull SB, et al. A population- and family-based study of Canadian families reveals association of HLA DRB1*0103 with colonic involvement in inflammatory bowel disease. *Inflamm Bowel Dis.* 2003;9:1–9.
18. Fernandez L, Mendoza JL, Martinez A, et al. IBD1 and IBD3 determine location of Crohn's disease in the Spanish population. *Inflamm Bowel Dis.* 2004;10:715–722.
19. Arijs I, Quintens R, Van Lommel L, et al. Predictive value of epithelial gene expression profiles for response to infliximab in Crohn's disease. *Inflamm Bowel Dis.* 2010;16:2090–2098.
20. Baumgart M, Dogan B, Rishniw M, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME J.* 2007;1:403–418.
21. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7: 335–336.
22. Hammer Ø, Harper DAT, Ryan PD. Paleontological statistics software package for education and data analysis. *Palaeontol Electron.* 2001;4: 9–18.
23. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011;12:R60.
24. Langille MGI, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol.* 2013;31:814–821.
25. Abubucker S, Segata N, Goll J, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *Plos Comput Biol.* 2012;8.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc.* 1995;57:289–300.
27. Sepehri S, Kotlowski R. Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. *Inflamm Bowel Dis.* 2007;13:675–683.
28. Walker AW, Sanderson JD, Churcher C, et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol.* 2011;11:7.
29. Arab HH, Salama SA, Eid AH, et al. Camel's milk ameliorates TNBS-induced colitis in rats via downregulation of inflammatory cytokines and oxidative stress. *Food Chem Toxicol.* 2014;69:294–302.
30. Willing B, Halfvarson J. Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis.* 2009;15:653–660.
31. Tong M, Li X, Wegener Parfrey L, et al. A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *PLoS One*. 2013;8:e80702.
32. Haberman Y, Tickle TL, Dexheimer PJ, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest.* 2014;124:3617–3633.
33. Abu Freha N, Schwartz D, Elkrinawi J, et al. Inflammatory bowel disease among Bedouin Arabs in southern Israel: urbanization and increasing prevalence rates. *Eur J Gastroenterol Hepatol.* 2015;27:230–234.
34. Fidder HH, Avidan B, Lahav M, et al. Clinical and demographic characterization of Jewish Crohn's disease patients in Israel. *J Clin Gastroenterol.* 2003;36:8–12.
35. Wills ES, Jonkers DMAE, Savelkoul PH, et al. Fecal microbial composition of ulcerative colitis and Crohn's disease patients in remission and subsequent exacerbation. *PLoS One*. 2014;9:e90981.
36. Cleynen I, González JR, Figueroa C, et al. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut*. 2013;62: 1556–1565.
37. Juste C, Kreil DP, Beauvallet C, et al. Bacterial protein signals are associated with Crohn's disease. *Gut*. 2014;63:1566–1577.
38. Yamazaki K, Umeno J, Takahashi A, et al. A genome-wide association study identifies 2 susceptibility Loci for Crohn's disease in a Japanese population. *Gastroenterology*. 2013;144:781–788.
39. Julià A, Domènech E, Ricart E, et al. A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. *Gut*. 2013;62:1440–1445.
40. Ellinghaus D, Zhang H, Zeissig S, et al. Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. *Gastroenterology*. 2013;145:339–347.
41. Umeno J, Asano K, Matsushita T, et al. Meta-analysis of published studies identified eight additional common susceptibility loci for Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis.* 2011;17:2407–2415.
42. Andoh A, Sakata S, Koizumi Y, et al. Terminal restriction fragment length polymorphism analysis of the diversity of fecal microbiota in patients with ulcerative colitis. *Inflamm Bowel Dis.* 2007;13:955–962.
43. Komanduri S, Gillevet PM, Sikaroodi M, et al. Dysbiosis in pouchitis: evidence of unique microfloral patterns in pouch inflammation. *Clin Gastroenterol Hepatol.* 2007;5:352–360.
44. Reshef L, Kovacs A, Ofer A, et al. Pouch inflammation is associated with a decrease in specific bacterial taxa. *Gastroenterol*. 2015;149:718–727.
45. Tahara T, Yamamoto E, Suzuki H, et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res.* 2014;74:1311–1318.
46. Ohkusa T, Yoshida T, Sato N, et al. Commensal bacteria can enter colonic epithelial cells and induce proinflammatory cytokine secretion: a possible pathogenic mechanism of ulcerative colitis. *J Med Microbiol.* 2009;58: 535–545.