

## University of Groningen

### Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease

Imhann, Floris; Vich Vila, Arnau ; Bonder, Marc Jan; Fu, Jingyuan; Gevers, Dirk; Visschedijk, Marijn C.; Spekhorst, Lieke M.; Alberts, Rudi; Franke, Lude; van Dullemen, Hendrik M.

*Published in:*  
Gut

*DOI:*  
[10.1136/gutjnl-2016-312135](https://doi.org/10.1136/gutjnl-2016-312135)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Imhann, F., Vich Vila, A., Bonder, M. J., Fu, J., Gevers, D., Visschedijk, M. C., Spekhorst, L. M., Alberts, R., Franke, L., van Dullemen, H. M., Ter Steege, R. W. F., Huttenhower, C., Dijkstra, G., Xavier, R. J., Festen, E. A. M., Wijmenga, C., Zhernakova, A., & Weersma, R. K. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*, *67*(1), 108-119. <https://doi.org/10.1136/gutjnl-2016-312135>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



OPEN ACCESS

ORIGINAL ARTICLE

# Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease

Floris Imhann,<sup>1,2</sup> Arnau Vich Vila,<sup>1,2</sup> Marc Jan Bonder,<sup>2</sup> Jingyuan Fu,<sup>3</sup> Dirk Gevers,<sup>4</sup> Marijn C Visschedijk,<sup>1,2</sup> Lieke M Spekhorst,<sup>1,2</sup> Rudi Alberts,<sup>1,2</sup> Lude Franke,<sup>2</sup> Hendrik M van Dullemen,<sup>1</sup> Rinze W F Ter Steege,<sup>1</sup> Curtis Huttenhower,<sup>4,6</sup> Gerard Dijkstra,<sup>1</sup> Ramnik J Xavier,<sup>4,5</sup> Eleonora A M Festen,<sup>1,2</sup> Cisca Wijmenga,<sup>2</sup> Alexandra Zhernakova,<sup>2</sup> Rinse K Weersma<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2016-312135>).

For numbered affiliations see end of article.

## Correspondence to

Dr Rinse K Weersma, Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, P.O. Box 30.001, Groningen 9700RB, The Netherlands; [r.k.weersma@umcg.nl](mailto:r.k.weersma@umcg.nl)

FI and AVV are shared first authors., AZ and RKW are shared last authors.

Received 24 April 2016  
Revised 15 August 2016  
Accepted 14 September 2016  
Published Online First  
10 October 2016

## ABSTRACT

**Objective** Patients with IBD display substantial heterogeneity in clinical characteristics. We hypothesise that individual differences in the complex interaction of the host genome and the gut microbiota can explain the onset and the heterogeneous presentation of IBD. Therefore, we performed a case–control analysis of the gut microbiota, the host genome and the clinical phenotypes of IBD.

**Design** Stool samples, peripheral blood and extensive phenotype data were collected from 313 patients with IBD and 582 truly healthy controls, selected from a population cohort. The gut microbiota composition was assessed by tag-sequencing the 16S rRNA gene. All participants were genotyped. We composed genetic risk scores from 11 functional genetic variants proven to be associated with IBD in genes that are directly involved in the bacterial handling in the gut: *NOD2*, *CARD9*, *ATG16L1*, *IRGM* and *FUT2*.

**Results** Strikingly, we observed significant alterations of the gut microbiota of healthy individuals with a high genetic risk for IBD: the IBD genetic risk score was significantly associated with a decrease in the genus *Roseburia* in healthy controls (false discovery rate 0.017). Moreover, disease location was a major determinant of the gut microbiota: the gut microbiota of patients with colonic Crohn's disease (CD) is different from that of patients with ileal CD, with a decrease in alpha diversity associated to ileal disease ( $p=3.28 \times 10^{-13}$ ).

**Conclusions** We show for the first time that genetic risk variants associated with IBD influence the gut microbiota in healthy individuals. *Roseburia spp* are acetate-to-butyrate converters, and a decrease has already been observed in patients with IBD.

## BACKGROUND AND AIMS

IBD, comprising Crohn's disease (CD) and ulcerative colitis (UC), is a chronic inflammatory disorder of the GI tract. In CD, inflammation can occur throughout the GI tract, whereas in UC, inflammation is confined to the mucosal layer of the colon. The clinical characteristics of IBD vary greatly between individuals with respect to disease location, disease activity and disease behaviour. The

## Significance of this study

### What is already known on this subject?

- The gut microbiota plays a key role in the pathogenesis of IBD.
- Known and presumed epidemiological risk factors for developing IBD such as mode of birth, breast feeding, smoking, hygiene, infections, antibiotics, diet and stress are all known to cause gut microbial perturbations.
- The large heterogeneity between patients with IBD is likely to result from individual differences in the complex interaction between the host genome and the gut microbiota.
- Discovering gene–microbiota interactions is difficult due to the large number of genomic markers as well as microbial taxa, requiring stringent multiple testing corrections.

### What are the new findings?

- Gut microbial changes could precede the onset of IBD. A high IBD genetic risk score is associated with a decrease in the genus *Roseburia* in the gut microbiota of healthy controls without gut complaints.
- Disease localisation is a major determinant of the IBD-associated gut microbiota composition.
- The use of a large well-phenotyped healthy control cohort next to an IBD cohort leads to an improved list of IBD-associated gut microbial differences.

### How might it impact on clinical practice in the foreseeable future?

- Better understanding of gene–microbiota interactions and proinflammatory gut microbial changes that precede the onset of IBD can lead to new IBD therapeutics, and perhaps even microbial prevention strategies.

origin of this heterogeneous clinical presentation remains poorly understood.<sup>1 2</sup>

The pathogenesis of IBD consists of an exaggerated immune response in a genetically susceptible



CrossMark

**To cite:** Imhann F, Vich Vila A, Bonder MJ, et al. *Gut* 2017;**67**:108–119.

host to the luminal microbial content of the gut. Driven by rapidly evolving genotyping and next-generation sequencing technologies, tremendous progress has been made in deciphering the host genomic landscape of IBD.<sup>3–4</sup> Systems biology approaches to genomic and biological data clearly show the importance of the interaction between the host genome and the microbial exposure in the gut.<sup>5</sup> Moreover, known and presumed epidemiological risk factors for developing IBD such as mode of birth (vaginal vs caesarean section), breast feeding, smoking, hygiene, infections, antibiotics, diet, stress and sleep pattern are all known to cause microbial perturbations, suggesting a key role for the gut microbiota in the pathogenesis of IBD.<sup>6–9</sup>

Previous studies have shown a reduced biodiversity in the gut microbial composition of patients with IBD, characterised by a reduction of known beneficial bacteria, such as *Faecalibacterium prausnitzii*, *Roseburia intestinalis* and other butyrate producers, and an increase of pathogens or pathobionts, for example, adherent-invasive *Escherichia coli* and *Shigella* species of the Enterobacteriaceae family. However, these studies used a relatively small number of controls, who were usually selected from the patient population of the gastroenterology department after excluding those with IBD.<sup>10</sup> Because recent gut microbiome research has shown significant effects of stool consistency and functional complaints on the gut microbiota,<sup>11–13</sup> previous results could have been influenced by their method of selection of controls.

While the main composition of the gut microbiota in CD has been studied extensively, the composition of the gut microbiota in patients with UC has received less attention.<sup>10–14–15</sup> Furthermore, the relationship between the gut microbiota and the clinical characteristics of IBD, including disease activity, disease duration and disease behaviour has only been studied in an exploratory manner.

Recent studies have begun to unravel the complex interaction of host genetics and the gut microbiota. These links between specific genetic variants and the abundance of specific bacteria are called microbiota quantitative trait loci (microbiotaQTLs). Twin studies show that the abundances of bacterial families Ruminococcaceae and Lachnospiraceae containing butyrate producers and acetate-to-butyrate converters are, to a certain degree, heritable.<sup>16–18</sup> Animal studies in mice specifically designed to discover microbiotaQTLs show the influence of genomic loci on several microbial genera.<sup>19</sup> Moreover, gut microbiota similarities in twins both concordant and discordant for IBD have been shown in several studies, further suggesting that host genetics can influence the gut microbiota.<sup>20–22</sup> Furthermore, preliminary data show that specific variants of the *NOD2* gene are associated with changes in the abundance of the Enterobacteriaceae family in patients with IBD.<sup>23</sup>

We hypothesise that the large heterogeneity between patients with IBD is likely to result from individual differences in the complex interaction between the host genome and the gut microbiota. Therefore, improving our knowledge of this interaction is crucial for our understanding of the pathogenesis of IBD.<sup>14</sup> So far, very few studies have been able to elucidate this interaction in an integrated manner. Here, we present a large single-centre case–control analysis of the luminal gut microbiota, the host genetics and clinical phenotypes of both CD and UC. To ensure optimal data quality, we adopted a rigorously standardised approach to collect and process fresh frozen faecal samples of 313 patients with IBD from a single hospital in the north of the Netherlands and 582 truly healthy controls from the same geographical area. For all individuals, extensive clinical data, laboratory and endoscopic findings were collected. In

addition, host genomic risk variants and risk scores were obtained in both the patients with IBD and the healthy controls to analyse host genomic influences on the gut microbial composition.

## METHODS

### Cohorts

In total, 357 patients with IBD were recruited from the specialised IBD outpatient clinic at the Department of Gastroenterology and Hepatology of the University Medical Center Groningen (UMCG) in Groningen, the Netherlands. All patients with IBD were diagnosed based on accepted radiological, endoscopic and histopathological evaluation. We excluded 44 patients with IBD who had a stoma, pouch or short bowel syndrome from further analyses. Healthy controls were selected from the 1174 participants of LifeLines DEEP, a cross-sectional general population cohort in the northern provinces of the Netherlands.<sup>24</sup> Data about medical history, medication use and gut complaints were meticulously reviewed by a medical doctor to ensure controls did not have any severe gut complaints or diseases, and did not use any medication that could confound our analysis of the gut microbiota. The selection process is described in detail in the online supplementary appendix. Pseudonymised data from patients with IBD and healthy controls were provided to the researchers. This study was approved by the Institutional Review Board of the UMCG (IRB number 2008.338). All participants signed an informed consent form.

### Clinical characteristics and medication use of patients with IBD

Extensive data on clinical characteristics and medication use were available for all patients with IBD at the time of stool sampling. Pseudonymised data were retrieved from the IBD-specific electronic patient records of the IBD Center at the Department of Gastroenterology and Hepatology of the UMCG. Disease activity at the time of sampling was determined by standardised and accepted clinical activity scores: the Harvey–Bradshaw index (HBI) for patients with CD and the Simple Clinical Colitis Activity Index (SCCAI) score for patients with UC. C reactive protein (CRP) and faecal calprotectin measurements were also available as indicators of disease activity. Disease localisation and behaviour were described according to the Montreal classification. Disease duration was determined as date of stool sampling in the study minus the date of diagnosis. IBD treatment at the time of sampling was scored (mesalazine, steroids, thiopurines, methotrexate, tumour necrosis factor  $\alpha$  (TNF- $\alpha$ ) inhibitors and other biologicals) as well as the use of other medications: proton pump inhibitors (PPIs), antidiarrhoeal medication (loperamide), bile salts, iron, minerals and vitamins at the time of sampling, and antibiotics use within the previous 3 months. Extraintestinal manifestations and complications of IBD were scored in several categories: (1) eye; (2) mouth; (3) skin; (4) joints; (5) Other (details in online supplementary appendix).

Serological measurements for antineutrophil cytoplasmic antibodies and anti-*Saccharomyces cerevisiae* antibodies were determined by immunofluorescence. Information on mode of birth, breast feeding during infancy and self-reported diets (see online supplementary appendix) was collected through questionnaires.

The association between a phenotype and the gut microbiota was only analysed if there were five or more patients with IBD with that phenotype. A list of all phenotypes can be found in the online supplementary appendix.

### Stool sample collection and faecal DNA extraction

Stool samples were collected for 313 cases with IBD and 582 controls. Identical protocols were used to collect and process all stool samples. All participants were asked to produce a stool sample at home. These were frozen by the participant within 15 min after stool production in the participant's home freezer. A research nurse visited each participant shortly after stool production to collect the sample on dry ice for transport to the UMCG at  $-80^{\circ}\text{C}$ . Samples were subsequently stored at  $-80^{\circ}\text{C}$  in the laboratory. All samples remained frozen until DNA isolation for which aliquots were made, and microbial DNA was isolated using the Qiagen AllPrep DNA/RNA Mini Kit cat # 80204 as previously described.<sup>10</sup>

### Host genotyping, variant selection and genetic risk modelling

Host DNA was available for all patients with IBD and healthy controls. Host DNA was isolated from peripheral blood as previously described.<sup>25</sup> Genotyping was performed using the ImmunoChip, an Illumina Infinium microarray comprising 196 524 single nucleotide variants (SNPs) and a small number of insertion/deletion markers, selected based on results from genome-wide association studies of 12 different immune-mediated diseases including IBD. Normalised intensities for all samples were called using the OptiCall clustering program.<sup>26</sup> The genotype prediction was improved via stringent calling with BeagleCall using recommended settings.<sup>27</sup> Marker and sample quality control were performed as previously described.<sup>3</sup> Human leucocyte antigen (HLA) imputation was performed using SNP2HLA. The Type 1 Diabetes Genetics Consortium genotype data were used as a reference panel for imputation. The SNP2HLA imputes the classical HLA alleles and amino acid sequences within the major histocompatibility complex (MHC) region on chromosome 6.<sup>28</sup>

To overcome statistical problems inherent to multiple testing when combining both genome-wide and 16S rRNA microbiota data, we adopted an approach of analysing a set of selected SNPs based on (i) their involvement in IBD, (ii) their predicted functional consequences and (iii) their role in bacterial sensing and signalling in the gut.<sup>23</sup>

Eleven known IBD genetic risk variants were selected for our genome-microbiota interaction analyses. We selected these risk variants ensuring that the selected IBD risk SNPs (as identified in the International IBD Genetics Consortium ImmunoChip analysis or targeted resequencing studies) are functional variants or are in strong linkage disequilibrium with functional variants that are implicated in the interaction of the host with the gut microbiota.<sup>3 29</sup> We included the following seven genetic variants in *NOD2*: rs104895431 (S431L), rs2066844 (R702W), rs5743277 (R703C), rs104895467 (N852S), rs2066845 (G908R), rs5743293 (fs1007insC) and rs104895444 (V793M). The variant rs10781499 in *CARD9* was selected because *Card9* has been shown to mediate intestinal epithelial cell restitution, T helper 17 responses and control of intestinal bacterial infection in mice.<sup>30</sup> Two variants in *FUT2*, rs516246 and rs1047781, were selected because these variants have been shown to influence colonic mucosa-associated microbiota in CD.<sup>31</sup> SNPs rs11741861 in *IRGM* and rs12994997 in *ATG16L1* were included because of their role in decreased selective autophagy that results in altered cytokine signalling and decreased antibacterial defence.<sup>32 33</sup>

In addition to these 11 genetic variants, we also created risk scores for all 200 known IBD risk variants.<sup>3 5</sup> We also analysed the influence of the HLA-DRB1\*01:03 haplotype on the gut

microbial composition in colonic disease, because this recently identified haplotype is associated with both UC and colonic CD and is suggested to be involved in appropriately controlling the immune response to colonic microbiota.<sup>34</sup>

### Determining the gut microbial composition

Illumina MiSeq paired-end sequencing was used to determine the bacterial composition of the stool samples. Forward primer 515F [GTGCCAGCMGCCGCGGTAA] and reverse primer 806R [GGACTACHVGGGTWTCTAAT] of hypervariable region V4 of the 16S rRNA gene were used. Custom scripts were used to remove the primer sequences and align the paired-end reads.<sup>10</sup>

### Operational taxonomic units: operational taxonomic unit-picking and filtering

The operational taxonomic unit (OTU) selection was performed using the QIIME reference optimal picking, using Usearch (V7.0.1090) to perform the clustering at 97% of similarity. Greengenes V.13.8 was used as a reference database. In all, 12 556 OTUs were identified. Samples with less than 10 000 counts were removed. OTUs that were not present in at least 1% of our samples or with a low abundance (<0.01% of the total counts) were filtered out.

### Function prediction

The functional imputation tools PICRUSt and HUMAnN were used to investigate the functional implications of the gut microbiota of patients with IBD. More information about the function prediction and the software can be found in the online supplementary appendix.

### Statistical analysis

The richness and the  $\beta$ -diversity of the microbiota dataset were analysed using QIIME.<sup>35</sup> The Shannon diversity index and the number of observed species per sample were used as  $\alpha$ -diversity metrics.  $\beta$ -diversity was calculated using unweighted Unifrac distances and represented in Principal Coordinate Analyses (PCoA). The Wilcoxon test and Spearman correlations were used to identify differences in the Shannon index and the relations between the principal coordinates.  $\chi^2$  tests, Fisher's exact tests, Spearman correlations and Wilcoxon-Mann-Whitney tests (WMW tests) were used to determine the differences in the clinical characteristics of patients with IBD. QIIMETOMAASLIN was used to convert the OTU counts into relative taxonomical abundance. OTUs representing identical taxonomies were aggregated, and higher taxon levels were added when multiple OTUs represented that taxon. Due to the limitations of the resolution on taxonomical classification using 16S gene sequencing, we restricted our analysis to the genus level and above. The initial 12 556 OTUs were classified into 250 taxonomical levels.

We used MaAsLin to identify differentially abundant taxa and pathways: (1) between patients with IBD and healthy controls, (2) between different IBD phenotypes and (3) between individuals with diverse amounts of IBD genetic risk variants.<sup>15</sup> MaAsLin performs boosted additive general linear models between metadata and microbial abundance data. The default settings of MaAsLin were used in all analyses. We used the Q-value package implemented in MaAsLin to correct for multiple testing. A false discovery rate (FDR) of 0.05 was used as the cut-off value for significance. The effect of the IBD diagnosis (CD or UC) on the gut microbiota composition was analysed by adding the IBD diagnosis versus healthy as a discrete predictor in the MaAsLin general linear mixed model analysis.



Unweighted genetic risk scores were calculated for every participant by summing up the risk alleles of the above-mentioned SNPs (risk allele=1; IBD protective allele=0).<sup>25</sup> Weighted genetic risk scores were calculated for every participant by summing up the log-normalised odds of the genetic variants of the same above-mentioned SNPs. Both risk scores were added as a predictor to the additive general linear model in MaAsLin. The analyses of the host genome and the microbiota composition were performed separately in patients with IBD and healthy controls.

### Correction for factors influencing the gut microbiota

Parameters that potentially influence the gut microbiota were identified by statistical analysis of cohort phenotypes, univariate MaAsLin analyses and literature search, and subsequently added as cofactors to the additive linear model. In every analysis, the parameters age, gender, body mass index, read-depth, PPI use, antibiotics use and IBD medication (mesalazine, steroids, thiopurines, methotrexate and TNF- $\alpha$  inhibitors) were added as covariates. Stool consistency also affects the gut microbiota. However, since stool consistency, mainly the occurrence of diarrhoea, is a key characteristic of increased IBD disease activity, stool consistency was not used as a covariate in all models. However, stool consistency was incorporated in the analyses, since the clinical disease activity scores used—the HBI for CD and the SCCAI—take the number of liquid stools per day (in the HBI) and the number of bowel movements during the day and during the night (in the SCCAI) into account.

## RESULTS

### The clinical characteristics of patients with IBD and the selection of healthy controls

The cohort consists of 313 patients with IBD (188 patients with CD, 107 patients with UC and 18 patients with IBD intermediate/IBD undetermined (IBDI/IBDU)) and 582 healthy controls selected from the population cohort LifeLines DEEP (selection criteria can be found in the online supplementary appendix).<sup>24</sup> Patients with CD were younger than healthy controls (41.3 vs 45.9 years;  $p=1\times 10^{-4}$ , WMW test), while patients with UC were not older than healthy controls ( $p=0.32$ , WMW test). At the time of sampling, 81 patients with IBD (25.8%) had active disease, defined as an HBI of higher than 4 in patients with CD or an SCCAI score higher than 2.5 in patients with UC. Of the patients with IBD, 23.7% had used antibiotics within the last 3 months. PPI use was more frequent in patients with IBD (24.5%) than in healthy controls (4.7%) ( $p<0.001$ ,  $\chi^2$  test). Extensive information on all clinical characteristics and medication use is presented in [table 1](#).

### Overall composition of the gut microbiota in patients with IBD and healthy controls

The predominant phyla in both patients with IBD and healthy controls were Firmicutes (73% in patients with IBD, 75% in healthy controls), Actinobacteria (9% in patients with IBD, 13% in healthy controls) and Bacteroidetes (14% in patients with IBD, 8% in healthy controls). Clostridia was the most abundant class (64% in patients with IBD, 68% in healthy controls). An overview of the abundances at all taxonomic levels can be found in online supplementary table S1.

#### Alpha diversity

A statistically significant decrease in the Shannon index was observed in patients with IBD compared with healthy controls as depicted in online supplementary figure S1 ( $p=5.61\times 10^{-14}$ , Wilcoxon test) and [figure 1](#).

#### Principal coordinate analysis

The differences in gut microbial composition between patients with IBD and healthy controls were also observed in the PCoA analysis. Statistically significant differences were found in the first three components (PCoA1  $p=2.62\times 10^{-68}$ , PCoA2  $p=0.033$ , PCoA3  $p=1.50\times 10^{-10}$ , Wilcoxon test). The gut microbiota of healthy controls clustered together, while the gut microbiota of patients with IBD were more heterogeneous, partially overlapping the healthy controls. The shape of the PCoA plot is mainly explained by the disease location and the Shannon index (see results below), as depicted in [figure 2A–D](#).

### IBD genetic risk variants are associated to unfavourable gut microbiota changes in healthy controls

The role of 11 functional genomic variants associated to IBD in the genes *NOD2*, *CARD9*, *ATG16L1*, *IRGM* and *FUT2* was investigated. In the unweighted analysis in healthy controls, a higher number of IBD risk alleles was associated with a decrease in the abundance of the genus *Roseburia* of the phylum Firmicutes (FDR=0.017) as depicted in [figure 3](#). In patients with IBD as well as subsets of patients with IBD (patients with CD, patients with UC, patients with ileal CD, patients with ileocolonic CD and patients with colonic CD), neither the single genetic risk variants, the HLA-DRB1\*01:03 haplotype, nor the weighted or unweighted composite scores of genetic risk alleles showed any statistically significant effect on the gut microbiota composition. All results of the analyses with the risk scores of 11 SNPs can be found in online supplementary table S2. Risk scores including all 200 IBD risk SNPs did not show any significant relations with the gut microbiota composition.

### Dysbiosis in patients with CD and UC: new associations Crohn's disease

Compared with healthy controls, 69 taxa were statistically significantly altered in patients with CD (genus and above; 28%; FDR<0.05). These alterations are presented in [table 2](#) and depicted in the cladogram in online supplementary figure S2A. The phyla Bacteroidetes (FDR=1.12 $\times 10^{-14}$ ) and Proteobacteria (FDR=2.71 $\times 10^{-22}$ ) were increased, while the phyla Actinobacteria (FDR=7.15 $\times 10^{-10}$ ) and Tenericutes (FDR=1.90 $\times 10^{-12}$ ) were decreased. Within the phylum Bacteroidetes, the order Bacteroidales was increased (FDR=1.12 $\times 10^{-14}$ ) as well as the genus *Parabacteroides* within the family Porphyromonadaceae (FDR=0.0016). Within the order Clostridiales of the phylum Firmicutes, seven families were decreased: Mogibacteriaceae, Christensenellaceae, Clostridiaceae, Dehalobacteriaceae, Peptococcaceae, Peptostreptococcaceae and Ruminococcaceae (FDR<0.05). The family Enterobacteriaceae of the phylum Proteobacteria, containing many known gut pathogens, was increased (FDR=0.0020). The genera *Bifidobacterium*, *Ruminococcus* and *Faecalibacterium* were also decreased in patients with CD (FDR=2.16 $\times 10^{-6}$ , 4.70 $\times 10^{-5}$  and 7.82 $\times 10^{-23}$ , respectively).

The changes in relative abundance of the statistically significantly altered families are depicted in [figure 4](#). The complete list of increased and decreased taxa including direction, coefficient and FDR values is presented in online supplementary table S3.

#### Ulcerative colitis

In patients with UC, 38 of the taxa were statistically significantly altered compared with healthy controls (genus and above; 12%; FDR<0.05). These alterations are presented in [table 3](#) and depicted in a cladogram in online supplementary figure S2B.

Table 1 Clinical characteristics of patients with IBD and healthy controls

Average (SD) or count (%)	CD only	Ileal CD only	Colonic CD only	Ileocolonic CD only	UC only	IBDU/IBDI only	IBD	Healthy controls
Number of samples	188	68	36	78	107	18	313	582
Sequence read depth (SD)	47 730 (37 278)	44 610 (36 541)	49 870 (39 818)	48 060 (37 803)	49 090 (37 050)	66 653 (43 157)	48 820 (37 539)	48 740 (29 705)
Demographics								
Age (SD)	41.3 (14.5)	42.54 (14.15)	42.39 (14.5)	39.9 (14.7)	47.3 (14.6)	44.1 (16.8)	43.6 (14.8)	45.9 (13.7)
Gender (M/F) (%)	62/126 (33/67%)	23/45 (33/66%)	12/24 (33/66%)	25/53 (32/67%)	52/55 (48/51%)	7/11 (39/61%)	122/191 (39/61%)	302/280 (52/48%)
Weight and BMI								
Weight (SD)	75.7 (16.2)	77.5 (17.2)	74.4 (14.3)	74.9 (16.3)	81.27 (16.1)	84.9 (27.1)	78.2 (17.1)	77.4 (13.3)
BMI (SD)	24.9 (4.6)	25.0 (4.9)	25.1 (4.6)	24.7 (4.7)	26.46 (4.4)	27.9 (8.3)	25.4 (4.9)	24.9 (3.7)
Disease location								
Ileum (%)	68 (36%)	68 (100%)	NA	NA	NA	NA	68(4%)	NA
Colon (%)	36 (19%)	NA	36 (100%)	NA	106 (99%)	8 (44%)	152 (48%)	NA
Both (%)	78 (41%)	NA	NA	78 (100%)	NA	2 (11%)	80 (25%)	NA
Disease activity								
CRP (SD)	10.7 (16.63)	11.1 (21.3)	15.0 (18.8)	8.8 (9.4)	6.2 (7.3)	7.29 (8.6)	8.9 (13.7)	NA
Faecal calprotectin (SD)	390.1 (535.1)	296.9 (533.3)	445.6 (693.2)	432.4 (437.6)	776.6 (1986.8)	870.2 (1166.4)	531.9 (1220.3)	NA
Harvey-Bradshaw index (SD)	3.45 (3.86)	3.25 (3.18)	3.8 (4.7)	3.6 (4.13)	NA	NA	3.45 (3.86)	NA
SCCAI (SD)	NA	NA	NA	NA	1.8 (2.2)	1.4 (2.0)	1.8 (2.2)	NA
Disease duration and age at diagnosis								
Disease duration in years (SD)	12.36 (9.13)	12.73 (9.0)	12.14 (8.6)	12.14 (9.7)	11.21 (8.48)	10.5 (8.93)	11.8 (8.8)	NA
Age at diagnosis (years) (SD)	28.9 (12.4)	29.9 (11.3)	30.2 (15.3)	27.8 (11.8)	36.08 (14.4)	32.5 (17.2)	31.8 (13.7)	NA
Disease behaviour Crohn's disease								
Montreal classification B1	104 (55%)	33 (48%)	28 (77%)	41 (52%)	NA	NA	104 (55%)	NA
Montreal classification B2	59 (31%)	26 (38%)	5 (13%)	25 (32%)	NA	NA	59 (31%)	NA
Montreal classification B3	25 (13%)	9 (13%)	3 (8%)	12 (15%)	NA	NA	25 (13%)	NA
Disease severity UC								
Montreal classification S1	NA	NA	NA	NA	6 (5%)	NA	6 (5%)	NA
Montreal classification S2	NA	NA	NA	NA	39 (36%)	NA	39 (36%)	NA
Montreal classification S3	NA	NA	NA	NA	44 (41%)	NA	44 (41%)	NA
Montreal classification S4	NA	NA	NA	NA	17 (15%)	NA	17 (15%)	NA
Serology								
ANCA pos/neg (%)	55/127 (30/67%)	18/49 (26/72%)	14/19 (39/52%)	18/49 (72/26%)	53/49 (50/45%)	8/10 (44/56%)	116/186 (37/69%)	NA
ASCA pos/neg (%)	85/92 (44/49%)	38/28 (56/41%)	8/24 (22/66%)	37/38 (47/49%)	16/85 (80/14%)	2/16 (11/88%)	102/194 (32/62%)	NA
Birth and breast feeding								
Vaginal birth	160 (85%)	58 (85%)	31 (86%)	66 (84%)	93 (87%)	18 (100%)	269 (85%)	NA
Caesarian section	5 (2%)	1 (1%)	1 (2%)	3 (3%)	2 (1%)	0 (0%)	7 (2%)	NA
Breast fed	96 (51%)	36 (53%)	22 (61%)	36 (46%)	62 (58%)	11 (61%)	169 (54%)	NA
Smoking								
Current smokers	58 (31%)	24 (35%)	12 (33%)	24 (35%)	15 (14%)	4 (22%)	77 (25%)	98 (17%)
IBD medication								
Mesalazine	12 (6%)	6 (9%)	1 (2%)	5 (6%)	87 (81%)	3 (17%)	113 (36%)	0 (0%)
Steroids	40 (21%)	8 (11%)	9 (25%)	20 (26%)	18 (17%)	5 (28%)	60 (19%)	0 (0%)
Thiopurines	67 (36%)	27 (40%)	15 (42%)	23 (30%)	32 (30%)	6 (33%)	104 (33%)	0 (0%)

Continued

Table 1 Continued

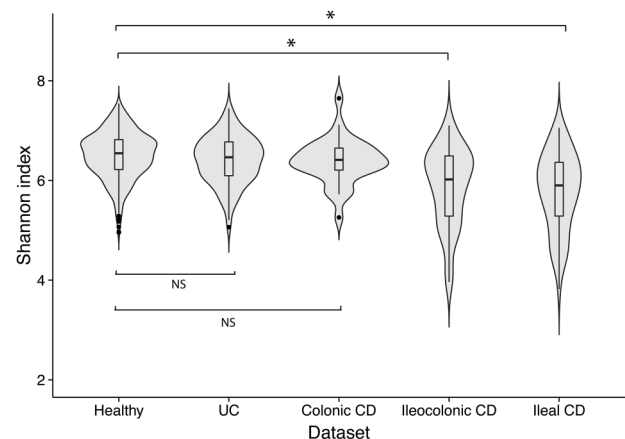
	CD only	Ileal CD only	Colonic CD only	Ileocolonic CD only	UC only	IBDU/IBDI only	IBD	Healthy controls
Average (SD) or count (%)								
Methotrexate	22 (11%)	6 (9%)	3 (8%)	12 (15%)	1 (1%)	0 (0%)	23 (7%)	0 (0%)
Anti-TNF $\alpha$	79 (42%)	26 (38%)	18 (50%)	34 (44%)	10 (9%)	3 (17%)	92 (29%)	0 (0%)
Other medication								
Antibiotics	41 (21%)	13 (19%)	11 (39%)	17 (21%)	15 (14%)	4 (22%)	59 (19%)	0 (0%)
Proton pump inhibitors	44 (23%)	16 (24%)	6 (17%)	21 (27%)	13 (12%)	2 (11%)	60 (19%)	26 (4%)
Antidiarrhoeal	29 (16%)	14 (3%)	2 (5%)	13 (17%)	4 (3%)	1 (5%)	33 (11%)	0 (0%)
Bile salts	3 (1%)	2 (3%)	0 (0%)	1 (1%)	3 (2%)	1 (5%)	7 (2%)	0 (0%)
Immunosuppressants	92 (51%)	32 (51%)	19 (55%)	38 (50%)	38 (36%)	5 (30%)	135 (44%)	0 (0%)
Mineral	5 (2%)	1 (1%)	3 (8%)	1 (1%)	2 (1%)	0 (0%)	7 (2%)	0 (0%)
Osteoporosis medication	5 (2%)	2 (3%)	1 (2%)	2 (3%)	1 (1%)	0 (0%)	6 (2%)	NA
Vitamins	74 (41%)	34 (52%)	5 (14%)	32 (42%)	2 (1%)	2 (11%)	78 (25%)	0 (0%)
Self-reported diets								
Diabetes diet	2 (1%)	0 (0%)	1 (2%)	1 (1%)	4 (4%)	0 (0%)	6 (2%)	0 (0%)
Fat-limited diet	6 (3%)	2 (3%)	1 (2%)	2 (3%)	4 (4%)	1 (5%)	11 (4%)	9 (2%)
Vegetarian diet	9 (5%)	1 (1%)	3 (8%)	5 (7%)	6 (6%)	1 (5%)	15 (5%)	39 (7%)
Other diet	18 (10%)	6 (9%)	4 (11%)	8 (11%)	10 (10%)	0 (0%)	28 (9%)	23 (4%)

ANCA, antineutrophil cytoplasmic antibodies; ASCA, anti-Saccharomyces cerevisiae antibodies; BMI, body mass index; CD, Crohn's disease; CRP, C reactive protein; IBDI, IBD intermediate; IBDU, IBD undetermined; SCCA1, Simple Clinical Colitis Activity Index; TNF- $\alpha$ , tumour necrosis factor  $\alpha$ .

Similar to the patients with CD, the abundances of the phyla Bacteroidetes (FDR=8.87 $\times 10^{-13}$ ) and Proteobacteria (FDR=4.06 $\times 10^{-5}$ ) were increased, while the phylum Firmicutes (FDR=0.0079) was decreased in patients with UC. Within the phylum Bacteroidetes, the order Bacteroidales (FDR=8.87 $\times 10^{-13}$ ), the family Rikenellaceae (FDR=0.025) and the genus *Bacteroides* (FDR=1.72 $\times 10^{-18}$ ) were all increased compared with healthy controls. *Lachnobacterium* and *Roseburia*, genera in the order Clostridiales of the phylum Firmicutes, were also increased in patients with UC (FDR=0.023 and FDR=0.00056, respectively). The changes in relative abundance of the altered families are depicted in figure 4 (FDR<0.05). The complete list of increased and decreased taxa, including direction, coefficient and FDR values, is presented in online supplementary table S3.

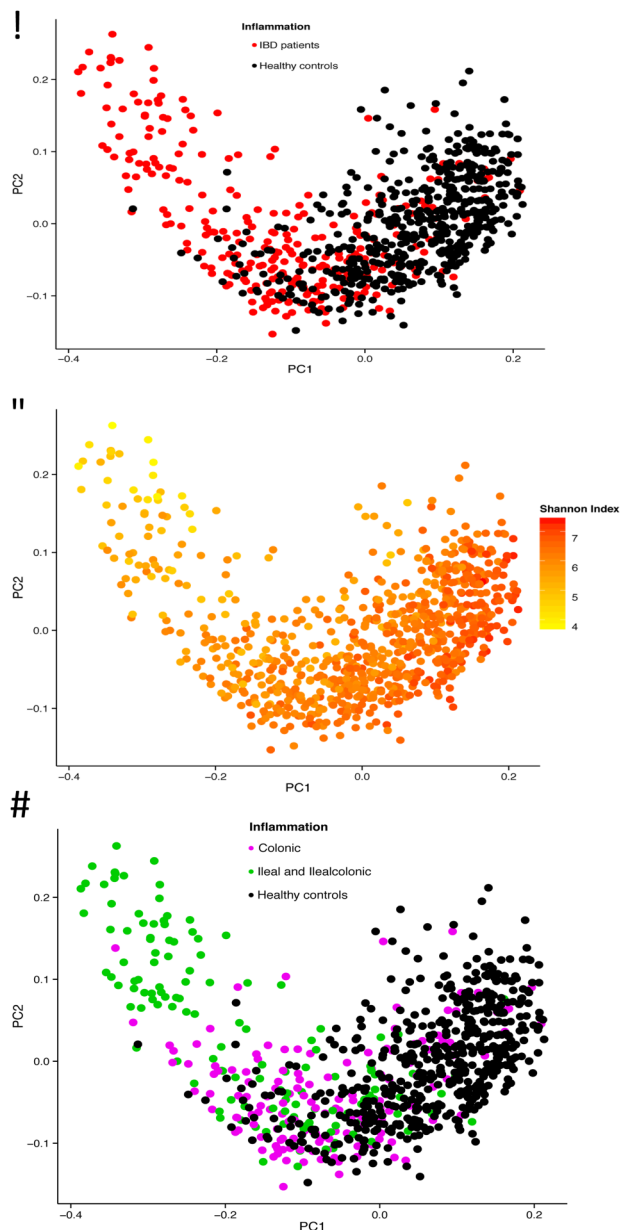
### Disease location is a major determinant of the gut microbiota in patients with IBD

The PCoA depicted in figure 2C shows the difference between the gut microbiota of patients with colonic disease (colonic CD and UC combined) and patients with ileal disease (ileal CD and ileocolonic CD combined). There is overlap between healthy controls and patients with colonic disease, while in concordance with the  $\alpha$ -diversity analysis in figure 1, the gut microbiota of patients with ileal disease deviates more from healthy controls. The statistical analysis of the PCoA supports this result: the first component is related to disease location (PCoA1  $r=0.63$ ,  $p=7.39\times 10^{-91}$ , Spearman correlation), and patients with colonic CD differ from patients with ileal CD ( $p=5.42\times 10^{-9}$ ). The  $\alpha$ -diversity analysis shows similar results: the gut microbiota of patients with IBD with colonic disease is not statistically significantly decreased compared with healthy controls (Shannon index in patients with UC=6.41 vs Shannon index in healthy controls=6.50,  $p=0.06$ ; Shannon index in patients with colonic CD=6.38 vs Shannon index in healthy controls=6.50,  $p=0.08$ , Wilcoxon test). On the contrary, patients with IBD with ileal disease show a statistically significant decrease in  $\alpha$ -diversity (patients with ileal CD vs healthy controls  $p=3.28\times 10^{-13}$  and patients with ileocolonic CD vs healthy controls  $p=3.11\times 10^{-11}$ , Wilcoxon test), as depicted in figure 1.



**Figure 1**  $\alpha$ -diversity (Shannon index) of the gut microbiota of healthy controls, patients with UC, patients with colonic Crohn's disease (CD), patients with ileocolonic CD and patients with ileal CD.  $\alpha$ -diversity is not decreased in colonic disease (UC and colonic CD) compared with healthy controls. In contrast, in patients with ileal and ileocolonic CD, the  $\alpha$ -diversity is statistically significantly decreased (patients with ileal CD vs healthy controls  $p=3.28\times 10^{-13}$  and patients with ileocolonic CD vs healthy controls  $p=3.11\times 10^{-11}$ , Wilcoxon test).

## Gut microbiota

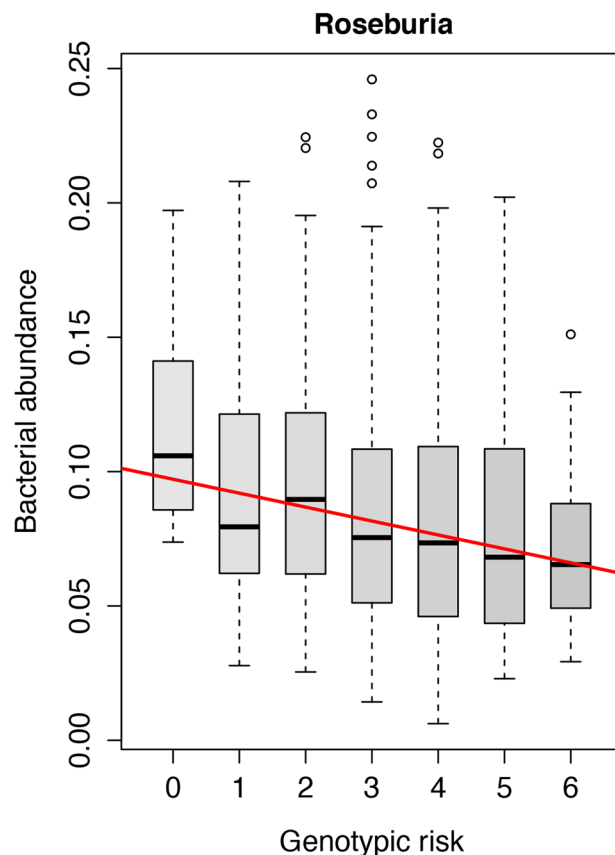


**Figure 2** Principal coordinate analysis (PCoA) of stool samples of 313 patients with IBD and 582 healthy controls. (A) The gut microbiota of patients with IBD is different from the gut microbiota of healthy controls, with only partial overlap. (B) The first component is related to the Shannon index. (C and D) There is more overlap between colonic disease (UC and colonic Crohn's disease (CD) combined) and healthy controls than between ileal disease (ileal CD and ileocolonic CD combined) and healthy controls. The first component is related to disease location (PCoA1  $r=0.63$ ,  $p=7.39 \times 10^{-91}$ , Spearman correlation) and patients with colonic CD differ from patients with ileal CD ( $p=5.42 \times 10^{-9}$ ).

Whether the IBD genetic risk was associated with disease location was also tested. The genetic risk could not explain the disease location (colonic IBD vs ileal involved IBD; unweighted genetic risk score using 200 SNPs; Spearman correlation;  $r=0.045$ ;  $p=0.47$ ). The taxonomy analysis of disease location is presented in the online supplementary appendix.

### Effects of IBD disease activity on the gut microbiota

We analysed several read-outs for disease activity at the time of sample collection: the clinical HBI scores for patients with CD



**Figure 3** Increased risk score of 11 IBD-related genetic variants in gut bacterial handling genes (*NOD2*, *CARD9*, *IRGM*, *ATG16L1* and *FUT2*) is statistically significantly associated to decreased abundance of *Roseburia* spp. in healthy controls (false discovery rate=0.017).

and SCCAI scores for patients with UC, as well as CRP and faecal calprotectin level measurements for all patients with IBD. A higher HBI was associated with an increase of the family Enterobacteriaceae in patients with CD (FDR=0.036). No significant associations were found between the gut microbiota and the SCCAI in patients with UC. Neither CRP nor faecal calprotectin was statistically significantly associated with altered bacterial abundances in the gut. Details of the disease activity analyses can be found in online supplementary tables S4 and S5.

### Effects of IBD disease duration on the gut microbiota

The disease duration in patients with IBD was measured from the date of diagnosis up to the date of sample collection. A longer duration of the disease, corrected for age, was associated with a higher abundance of the phylum Proteobacteria (FDR=0.045) (see online supplementary table S6).

### Analysis of other IBD subphenotypes

Other gut microbial associations with other IBD subphenotypes including medication, smoking behaviour and extraintestinal manifestations can be found in the Results section of the online supplementary appendix.

### Pathway prediction and gut microbiota function changes in patients with IBD

Multiple metabolic pathways including butyrate metabolism, endotoxin metabolism and antibiotic resistance pathways were differentially expressed between patients with IBD, UC, CD, ileal CD, ileocolonic CD and colonic CD as compared with



**Table 2** Comparison of altered taxa in patients with Crohn's disease (CD) compared with healthy controls: family level and above**Gut microbiota alterations in patients with CD (current study: FDR<0.05)**

Taxon (family and above)	Phylum (or kingdom)	Current study*	Gevers <i>et al</i> †	Morgan <i>et al</i> ‡	Willing <i>et al</i> §
f__Methanobacteriaceae	Archea (kingdom)	Down	Not reported	Not reported	Not reported
p__Actinobacteria		Down	Down	Not reported	Up in colonic CD
c__Actinobacteria	Actinobacteria	Down	Down	Not reported	Up in colonic CD
f__Micrococcaceae	Actinobacteria	Not reported	Up	Not reported	Not reported
f__Bifidobacteriaceae	Actinobacteria	Down	Down	Down, in lower taxonomic levels	Up in colonic CD
f__Coriobacteriaceae	Actinobacteria	Down	Down	Not reported	Up in colonic CD
p__Bacteroidetes		Up	Down	Not reported	Not reported
o__Bacteroidales	Bacteroidetes	Up	Down	Not reported	Not reported
f__Porphyromonadaceae	Bacteroidetes	Up	Down	Down, in lower taxonomic levels	Unknown genus in this family: down in ileal CD
p__Firmicutes		Down, in lower taxonomic levels	Down	Down	Up in colonic CD
c__Bacilli	Firmicutes	Up, in lower taxonomic levels	Up	Associated to ileal involvement	Up in ileal CD
f__Aerococcaceae	Firmicutes	Up	Not reported	Not reported	Not reported
f__Enterococcaceae	Firmicutes	Up	Not reported	Not reported	Not reported
o__Gemellales	Firmicutes	Not reported	Up	Not reported	Not reported
f__Gemellaceae	Firmicutes	Not reported	Up	Not reported	Not reported
f__Streptococcaceae	Firmicutes	Not reported	Up	Not reported	Not reported
c__Clostridia	Firmicutes	Down	Down	Down	Down in ileal CD
o__Clostridiales	Firmicutes	Down	Down	Down	Down in ileal CD
f__Mogibacteriaceae	Firmicutes	Down	Not reported	Not reported	Not reported
f__Christensenellaceae	Firmicutes	Down	Down	Not reported	Not reported
f__Clostridiaceae	Firmicutes	Down	Down	Not reported	Not reported
f__Dehalobacteriaceae	Firmicutes	Down	Not reported	Not reported	Not reported
f__Lachnospiraceae	Firmicutes	Up, but genera in lower levels both going up and down	Down	Down, in lower taxonomic levels	Down, in lower taxonomic levels
f__Peptococcaceae	Firmicutes	Down	Not reported	Not reported	Down in ileal CD
f__Peptostreptococcaceae	Firmicutes	Down	Not reported	Not reported	
f__Ruminococcaceae	Firmicutes	Down	Down	Down	Down in ileal CD
f__Veillonellaceae	Firmicutes	Not reported	Up	Up	Up in lower taxonomic levels in ileal CD
f__Erysipelotrichaceae	Firmicutes	Down	Down	Associated to ileal involvement	Not reported
p__Fusobacteria		Not reported	Not reported	Not reported	Up in ileal CD
o__Fusobacteriales	Fusobacteria	Not reported	Up	Not reported	Up in ileal CD
f__Fusobacteriaceae	Fusobacteria	Not reported	Up	Not reported	Up in ileal CD
p__Proteobacteria		Up	Up	Up	Up in ileal CD
c__Betaproteobacteria	Proteobacteria	Up	Up	Not reported	Not reported
o__Burkholderiales	Proteobacteria	Up	Up	Not reported	Not reported
f__Neisseriaceae	Proteobacteria	Not reported	Up	Not reported	Not reported
c__Gammaproteobacteria	Proteobacteria	Up	Up	Up	Up in ileal CD
f__Aeromonadaceae	Proteobacteria	Not reported	Not reported	Not reported	Up in ileal CD
o__Campylobacteriales	Proteobacteria	Not reported	Up	Not reported	Not reported
f__Enterobacteriaceae	Proteobacteria	Up	Up	Up	Up in ileal CD
f__Pasteurellaceae	Proteobacteria	Not reported	Up	Not reported	
p__Tenericutes		Down	Not reported	Not reported	
c__Mollicutes	Tenericutes	Down	Not reported	Not reported	Down in ileal CD, up in colonic CD
f__Anaeroplasmataceae	Tenericutes	Not reported	Not reported	Not reported	Down in ileal CD, up in colonic CD
f__Verrucomicrobiaceae	Verrucomicrobia	Not reported	Down	Not reported	Not reported

k\_\_, kingdom; p\_\_, phylum; c\_\_, class; o\_\_, order; f\_\_, family.

\*313 patients with IBD including 188 patients with CD; 582 healthy controls; stool only.

†*Cell Host Microbe* 2014; 447 patients with CD; 221 controls; stool and biopsy.‡*Genome Biology* 2012; 204 patients with IBD including 121 patients with CD and 27 controls; stool and biopsy.§*Gastroenterology* 2010; 40 twin pairs concordant or discordant for CD/UC (23 CD pairs, 15 UC pairs, 2 healthy pairs).

FDR, false discovery rate.

## Gut microbiota

healthy controls. These altered Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways are presented in online supplementary figure S3 and table S7. The metabolism of short chain fatty acids was decreased in patients with IBD, as indicated by the decrease of the propanoate (also known as propionate) metabolism in patients with CD and UC (ko00640; CD:  $FDR=2.74 \times 10^{-11}$  and UC:  $FDR=3.59 \times 10^{-5}$ ), the decrease of the butanoate (also known as butyrate) metabolism in patients with CD (ko00650;  $FDR=5.31 \times 10^{-9}$ ) and the decreased fatty acid metabolism in patients with CD (ko00071;  $FDR=4.28 \times 10^{-18}$ ). Lipopolysaccharide or endotoxin biosynthesis was increased in both patients with CD and UC (ko00540; CD:  $FDR=4.69 \times 10^{-7}$  and UC:  $FDR=0.027$ ).  $\beta$ -lactam resistance metabolism was increased in patients with CD (ko00312;  $FDR=4.69 \times 10^{-7}$ ). There were no significant pathway increases or decreases related to the clinical disease activity score, the HBI, for patients with CD (see online supplementary table S8). More detailed information on the predicted pathways can be found in the Results section of the online supplementary appendix.

## CONCLUSIONS

By performing this extensive integrated case-control analysis of the gut microbiota, the host genome and the clinical characteristics of IBD, we have identified new gut microbial associations with IBD and are now able to refine our understanding of the findings of previous studies. We found a relation between host genetic IBD susceptibility variants and the gut microbiota composition in healthy individuals and observed the effect of disease location on the gut microbiota. Moreover, we report microbial associations with multiple IBD subphenotypes.

## Onset of IBD: genetic risk factors for IBD associated with proinflammatory gut microbiota alterations in healthy individuals

Discovering gene-microbiota interactions is difficult due to the large number of genomic markers as well as microbial taxa, requiring stringent multiple testing correction, thus limiting the possibility of finding statistically significant results. To resolve this issue, we created risk scores of known functional IBD risk variants proven to be involved in the bacterial handling in the gut. This hypothesis-based gene-microbiota approach limits the

number of tests that need to be done and has proven to be successful.

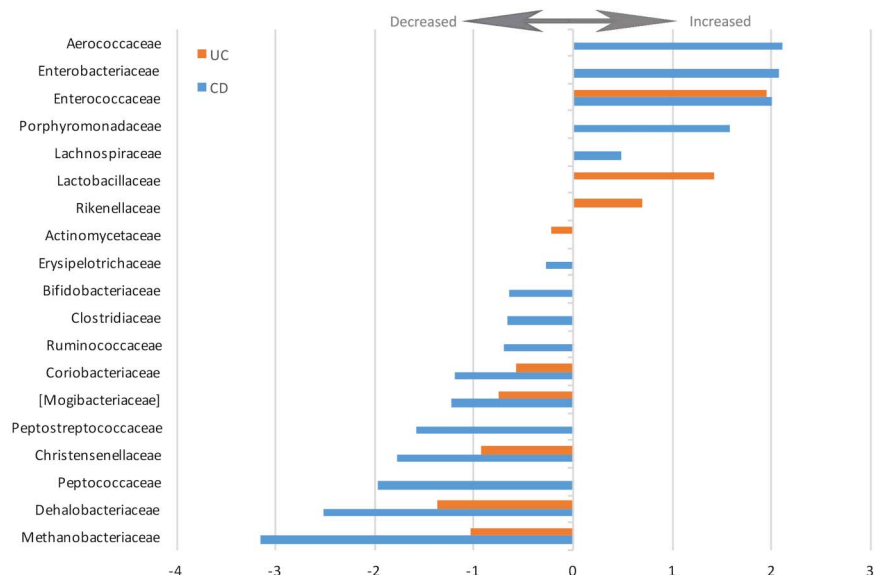
The gut microbiota interacts with the intestinal epithelium and the host immune system.<sup>18 36-39</sup> Recently, it was hypothesised that the interaction of the immune system with the gut microbiota goes two ways: 'good' gut microbiota can ameliorate immune responses, but the gut immune system can also 'farm' good bacteria in order to maintain immune-microbe homeostasis.<sup>36 37</sup> We can show support for this hypothesis: in healthy individuals, an increased genetic burden in functional variants in genes involved in bacterial handling (*NOD2*, *IRGM*, *ATG16L1*, *CARD9* and *FUT2*) is associated with a decrease of the acetate-to-butyrate converter *Roseburia spp.*

The species *Roseburia intestinalis* is one of the 20 most abundant species in the gut microbiota.<sup>40</sup> Importantly, a decrease in *Roseburia spp.* is already associated to the gut microbiota of patients with IBD.<sup>10 15</sup> In an in vitro model, *Roseburia spp.* specifically colonised the mucins, which govern mucosal butyrate production.<sup>41</sup> Butyrate derived from Clostridium clusters IV, VIII and XIVa to which *Roseburia spp.* belong has been shown to induce  $T_{reg}$  cells, preventing or ameliorating intestinal inflammation.<sup>38 39</sup> The abundances within the family Lachnospiraceae, to which *Roseburia spp.* belong, are significantly more similar in monozygotic twins than in dizygotic twins.<sup>17</sup> Moreover, unaffected siblings of patients with CD share a decrease in *Roseburia spp.*<sup>22</sup>

This finding in healthy individuals carrying IBD genetic risk variants has implications for our understanding of the onset of IBD. We hypothesise that genetic risk factors of the gut immune system lead to 'farming' of a more proinflammatory gut microbiota and increased susceptibility to IBD. Subsequent unfavourable microbial perturbations due to environmental risk factors could further disturb the immune-microbe homeostasis in the gut, eventually leading to IBD.

In addition to our genetic risk score based on specific functions, analyses using genetic risk scores of all 200 known IBD susceptibility variants, many of whose function is unknown, did not yield any statistically significant results in either patients with IBD or in healthy controls. We could not detect any gene-microbiota interactions in patients with IBD, probably due to the already well-established dysbiosis as a consequence of the inflammation in the gut. Another complication is the

**Figure 4** Log<sub>2</sub>-fold change of increased and decreased bacterial families in patients with UC and Crohn's disease versus healthy controls (false discovery rate<0.05).



**Table 3** Comparison of significant taxa associations in patients with UC: family level and above

Gut microbiota alterations in patients with UC (current study: FDR&lt;0.05)

Taxon (family and above)	Phylum (or kingdom)	Current study*	Gevers <i>et al</i> †	Morgan <i>et al</i> ‡
f__Methanobacteriaceae	Archea	Down	Not reported	Not reported
f__Actinomycetaceae	Actinobacteria	Down	Not reported	Not reported
f__Coriobacteriaceae	Actinobacteria	Down	Not reported	Not reported
p__Bacteroidetes		Up	Not reported	Not reported
o__Bacteroidales	Bacteroidetes	Up	Not reported	Not reported
f__Porphyromonadaceae	Bacteroidetes	Not reported	Not reported	Up
p__Firmicutes		Down	Down	Not reported
f__Enterococcaceae	Firmicutes	Up	Not reported	Not reported
f__Lactobacillaceae	Firmicutes	Up	Not reported	Not reported
c__Clostridia	Firmicutes	Down, in lower taxonomic levels	Down	Not reported
o__Clostridiales	Firmicutes	Down, in lower taxonomic levels	Down	Not reported
f__Mogibacteriaceae	Firmicutes	Down	Not reported	Not reported
f__Christensenellaceae	Firmicutes	Down	Not reported	Not reported
f__Clostridiaceae	Firmicutes	Down, in lower taxonomic levels	Down, in lower taxonomic levels	Not reported
f__Dehalobacteriaceae	Firmicutes	Down	Not reported	Not reported
f__Lachnospiraceae	Firmicutes	Within the family genera, both going up and down	Down, in lower taxonomic levels	Not reported
f__Ruminococcaceae	Firmicutes	Down, in lower taxonomic levels	Down	Not reported
f__Veillonellaceae	Firmicutes	Not reported	Up	Not reported
f__Erysipelotrichaceae	Firmicutes	Down, in lower taxonomic levels	Not reported	Down, in lower taxonomic levels
f__Streptococcaceae	Firmicutes	Not reported	Not reported	Down
p__Proteobacteria		Up	Not reported	Not reported
c__Betaproteobacteria	Proteobacteria	Up	Not reported	Not reported
o__Burkholderiales	Proteobacteria	Up	Not reported	Not reported
p__Tenericutes		Down	Not reported	Down
c__Mollicutes	Tenericutes	Down	Not reported	Down
f__Anaeroplasmataceae	Tenericutes	Not reported	Not reported	Down
f__Verrucomicrobiaceae	Verrucomicrobia	Down, in lower taxonomic levels	Not reported	Not reported

k\_\_, kingdom; p\_\_, phylum; c\_\_, class; o\_\_, order; f\_\_, family.

\*313 patients with IBD including 188 patients with Crohn's disease (CD); 582 healthy controls; stool only.

†*Cell Host Microbe* 2014; 447 patients with CD; 221 controls; stool and biopsy.‡*Genome Biology* 2012; 204 patients with IBD including 121 patients with CD and 27 controls; stool and biopsy.§*Gastroenterology* 2010; 40 twin pairs concordant or discordant for CD/UC (23 CD pairs, 15 UC pairs, 2 healthy pairs).

FDR, false discovery rate.

interrelatedness of the genotypes and phenotypes in IBD. For example, *NOD2* risk variants are known to be associated with ileal CD, and we show that ileal CD has a specific microbial signature. After correction for treatment, disease activity and disease location, we could not find any statistically significant genome–microbiota relations in patients with IBD.

### Dysbiosis in patients with CD and UC: new associations identified, previous associations corrected

The dysbiosis of the gut microbiota in patients with IBD is profound: the abundances of 69 taxa in patients with CD and 38 taxa in patients with UC were altered compared with healthy individuals (FDR<0.05). We compared our results on the phylum, class, order and family levels with two previous studies looking into the gut microbiota of patients with IBD.<sup>10 15 20</sup> This comparison is presented in [tables 2](#) (patients with CD) and [3](#) (patients with UC). An important new finding of our study is the increase in the phylum Bacteroidetes in both patients with CD and UC. Increased levels of Bacteroidetes have recently been discovered in patients with IBS.<sup>13</sup> Since the control groups used in the previous IBD studies also had functional GI complaints (ie, IBS), this would have confounded any comparisons

between Bacteroidetes levels in patients with IBD and controls, masking any meaningful enrichment in IBD.

The genus *Bacteroides* within the phylum Bacteroidetes is increased in our patients with UC. The involvement of *Bacteroides spp* in the pathogenesis of IBD has been implied in animal studies. In *NOD2* knockout mice, the exaggerated inflammatory response in the small intestine was dependent on *Bacteroides vulgatus*.<sup>42</sup> *Bacteroides thetaiotaomicron* induced colitis in HLA-B27 transgenic rats.<sup>43</sup> Another study looking into the effects of the vitamin D receptor in mice found increased levels of *Bacteroides spp* in colitis and increased levels of *Bacteroides fragilis* in colon biopsies of patients with UC.<sup>44</sup>

Increased abundance of the families Streptococcaceae, Micrococcaceae and Veillonellaceae, previously associated with IBD, is now associated to PPI use in our study. PPI use is overrepresented in patients with IBD.<sup>45</sup> Since previous studies did not correct for PPI use, we assume that alterations in the abundances of these taxa were wrongly assigned to the effect of IBD.

Our study is the largest gut microbiota study in patients with UC to date, and within it we can now begin to resolve the landscape of the UC gut microbiota. We were able to find many new associations, including the association with a decreased

abundance of phylum Tenericutes, which we also find to be associated with more extensive UC.

### Disease location is a major determinant of the gut microbial composition in IBD

We showed the importance of disease location for the composition of the gut microbiota in patients with IBD. In our PCoA, the gut microbiota of patients with colonic CD is more similar to the microbiota of patients with UC than to that of patients with ileal CD. While different clusters of gut microbiota samples are also observed in recent IBD metagenomics research, we have been able to relate these clusters to the disease location phenotype.<sup>46</sup> The importance of disease location also matches recent insights into host genetics, in which, based on genetic risk scores, colonic CD lies between UC and ileal CD.<sup>4</sup> We found that the gut microbiota composition in stool could explain the differences in IBD disease location, while the genetic risk variants in our cohort could not. Moreover, there is important overlap in the clinical presentation of colonic CD and UC, for example, the risk of developing colorectal carcinoma in colonic CD is similar to UC, but different from ileal CD.<sup>47</sup> Based on both the previous genetic findings and our current microbiota findings, it is becoming more apparent that colonic CD and ileal CD are different diseases within the IBD spectrum.

Through careful selection of healthy controls, meticulous standardisation of stool collection, extensive phenotyping and host genotyping, we were able to successfully perform analyses and gain insight into the gut microbiota as the key mediator of the IBD pathogenesis. For the first time, we find evidence for the role of the gut microbiota in the onset of IBD: healthy individuals with a high genetic risk load for IBD also have unfavourable changes in their gut microbiota. This relationship warrants further investigation as it might be both a potential target for treatment and a possibility for prevention of IBD in genetically susceptible hosts or their families.

#### Author affiliations

<sup>1</sup>Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>2</sup>Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>3</sup>Department of Pediatrics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>4</sup>Broad Institute of Harvard and MIT, Boston, Massachusetts, USA

<sup>5</sup>Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>6</sup>Biostatistics Department, Harvard School of Public Health, Boston, Massachusetts, USA

**Acknowledgements** The authors thank all the participants of the UR-IBD and LifeLines DEEP cohorts for contributing stool samples; Dianne Jansen, Jacqueline Mooibroek, Anneke Diekstra, Brecht Wedman, Rina Doorn, Astrid Maatman, Tiffany Poon, Wilma Westerhuis, Daan Wiersum, Debbie van Dussen, Martine Hesselink, Ettje Tigchelaar, Soesma A. Jankipersadings, Maria Carmen Cenit and Jackie Dekens for logistics support, laboratory support, data collection and data management; the research group of Morris Swertz for providing the high-performance computing infrastructure including the Calculon cluster computer; the Parelsnoer Institute for supporting the IBD biobank infrastructure; Timothy Tickle, Curtis Huttenhower, Alexandra Sirota, Chengwei Luo and Aleksander Kostic for their help in training the first and second authors; Marten Hofker and Eelke Brandsma for contributing to the scientific discussion. This article was edited for language and formatting by Kate McIntyre, Associate Scientific Editor in the Department of Genetics, University Medical Center Groningen.

**Contributors** RKW, DG, AZ, GD, CH and RJX designed the study. FI, MCV, LMS, HMVD, RWFTS, GD and RKW collected the data. FI, AVV, MJB, RA and JF analysed the data. FI, AVV, EAMF and RKW drafted the manuscript. CW, JF, EAMF, LF, DG, AZ, GD, CH, RJX and RKW critically reviewed the manuscript.

**Funding** RKW, JF and LF are supported by VIDI grants (016.136.308, 864.13.013 and 917.14.374) from the Netherlands Organization for Scientific Research (NWO). EAMF is funded by a career development grant from the Dutch Digestive Foundation

(MLDS) (No. CDG-014). Sequencing of the LifeLines DEEP cohort was funded by a Top Institute Food and Nutrition grant GH001 to CW. CW is further supported by an ERC advanced grant (ERC-671274). AZ holds a Rosalind Franklin fellowship (University of Groningen) and a CardioVascular Onderzoek Nederland grant (CVON 2012-03).

**Competing interests** None declared.

**Ethics approval** Institutional Review Board of the University Medical Center Groningen.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

#### REFERENCES

- 1 Abraham C, Cho JH. Inflammatory bowel disease. *N Engl J Med* 2009;361:2066–78.
- 2 Ordás I, Eckmann L, Talamini M, *et al*. Ulcerative colitis. *Lancet* 2012;380:1606–19.
- 3 Liu JZ, van Sommeren S, Huang H, *et al*. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979–86.
- 4 Cleyne I, Boucher G, Jostins L, *et al*. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 2015;6736:1–12.
- 5 Jostins L, Ripke S, Weersma RK, *et al*. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–24.
- 6 Ananthkrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol* 2015;12:205–17.
- 7 Jost T, Lacroix C, Braegger CP, *et al*. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ Microbiol* 2014;16:2891–904.
- 8 Thaiss CA, Zeevi D, Levy M, *et al*. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell* 2014;159:514–29.
- 9 David LA, Maurice CF, Carmody RN, *et al*. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014;505:559–63.
- 10 Gevers D, Kugathasan S, Denson LA, *et al*. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15:382–92.
- 11 Tigchelaar EF, Bonder MJ, Jankipersadings SA, *et al*. Gut microbiota composition associated with stool consistency. *Gut* 2016;65:540–2.
- 12 Dupont HL. Review article: evidence for the role of gut microbiota in irritable bowel syndrome and its potential influence on therapeutic targets. *Aliment Pharmacol Ther* 2014;39:1033–42.
- 13 Chung CS, Chang PF, Liao CH, *et al*. Differences of microbiota in small bowel and faeces between irritable bowel syndrome patients and healthy subjects. *Scand J Gastroenterol* 2016;51:410–19.
- 14 Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel diseases: current status and the future ahead. *Gastroenterology* 2014;146:1489–99.
- 15 Morgan XC, Tickle TL, Sokol H, *et al*. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79.
- 16 Blekhan R, Goodrich JK, Huang K, *et al*. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol* 2015;16:191.
- 17 Goodrich JK, Waters JL, Poole AC, *et al*. Human genetics shape the gut microbiome. *Cell* 2014;159:789–99.
- 18 Brestoff JR, Artis D. Commensal bacteria at the interface of host metabolism and the immune system. *Nat Immunol* 2013;14:676–84.
- 19 Leamy LJ, Kelly SA, Niefeldt J, *et al*. Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol* 2014;15:552.
- 20 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;139:1844–54.e1.
- 21 Joossens M, Huys G, Cnockaert M, *et al*. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 2011;60:631–7.
- 22 Hedin CR, McCarthy NE, Louis P, *et al*. Altered intestinal microbiota and blood T cell phenotype are shared by patients with Crohn's disease and their unaffected siblings. *Gut* 2014;63:1578–86.
- 23 Knights D, Silverberg MS, Weersma RK, *et al*. Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med* 2014;6:107.
- 24 Tigchelaar EF, Zhernakova A, Dekens JA, *et al*. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* 2015;5:e006772.



- 25 Festen EA, Stokkers PC, van Diemen CC, *et al.* Genetic analysis in a Dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk. *Am J Gastroenterol* 2010;105:395–402.
- 26 Shah TS, Liu JZ, Floyd JAB, *et al.* optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* 2012;28:1598–603.
- 27 Browning BL, Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am J Hum Genet* 2009;85:847–61.
- 28 Jia X, Han B, Onengut-Gumuscu S, *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* 2013;8:e64683.
- 29 Rivas MA, Beaudoin M, Gardet A, *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;43:1066–73.
- 30 Sokol H, Conway KL, Zhang M, *et al.* Card9 mediates intestinal epithelial cell restitution, T-helper 17 responses, and control of bacterial infection in mice. *Gastroenterology* 2013;145:591–601.e3.
- 31 Rausch P, Rehman A, Künzel S, *et al.* Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci USA* 2011;108:19030–5.
- 32 McCarroll SA, Huett A, Kuballa P, *et al.* Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* 2008;40:1107–12.
- 33 Sadaghian Sadabad M, Regeling A, de Goffau MC, *et al.* The ATG16L1-T300A allele impairs clearance of pathosymbionts in the inflamed ileal mucosa of Crohn's disease patients. *Gut* 2015;64:1546–52.
- 34 Goyette P, Boucher G, Mallon D, *et al.* High-density mapping of the MHC identifies a shared role for HLA-DRB1\*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet* 2015;47:172–9.
- 35 Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6.
- 36 Ley RE. The Gene-Microbe Link. *Nature* 2015;518:57.
- 37 Velasquez-Manoff M. Gut microbiome: the peacekeepers. *Nature* 2015;518:S3–11.
- 38 Furusawa Y, Obata Y, Fukuda S, *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 2013;504:446–50.
- 39 Atarashi K, Tanoue T, Oshima K, *et al.* Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* 2013;500:232–6.
- 40 Qin J, Li R, Raes J, *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65.
- 41 Van den Abbeele P, Belzer C, Goossens M, *et al.* Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J* 2013;7:949–61.
- 42 Ramanan D, Tang MS, Bowcutt R, *et al.* Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal bacteroides vulgatus. *Immunity* 2014;41:311–24.
- 43 Hansen JJ, Huang Y, Peterson DA, *et al.* The colitis-associated transcriptional profile of commensal Bacteroides thetaiotaomicron enhances adaptive immune responses to a bacterial antigen. *PLoS ONE* 2012;7:1–10.
- 44 Wu S, Zhang YG, Lu R, *et al.* Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis. *Gut* 2015;64:1082–94.
- 45 Imhann F, Bonder MJ, Vich Vila A, *et al.* Proton pump inhibitors affect the gut microbiome. *Gut* 2016;65:740–8.
- 46 Lewis JD, Chen EZ, Baldassano RN, *et al.* Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* 2015;18:489–500.
- 47 Averboukh F, Ziv Y, Kariv Y, *et al.* Colorectal carcinoma in inflammatory bowel disease: a comparison between Crohn's and ulcerative colitis. *Color Dis* 2011;13:1230–5.



## Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease

Floris Imhann, Arnau Vich Vila, Marc Jan Bonder, Jingyuan Fu, Dirk Gevers, Marijn C Visschedijk, Lieke M Spekhorst, Rudi Alberts, Lude Franke, Hendrik M van Dullemen, Rinze W F Ter Steege, Curtis Huttenhower, Gerard Dijkstra, Ramnik J Xavier, Eleonora A M Festen, Cisca Wijmenga, Alexandra Zhernakova and Rinse K Weersma

*Gut* 2018 67: 108-119 originally published online October 8, 2016  
doi: 10.1136/gutjnl-2016-312135

---

Updated information and services can be found at:  
<http://gut.bmj.com/content/67/1/108>

---

*These include:*

### References

This article cites 47 articles, 8 of which you can access for free at:  
<http://gut.bmj.com/content/67/1/108#BIBL>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Topic Collections

Articles on similar topics can be found in the following collections

[Open access](#) (404)  
[Crohn's disease](#) (932)

---

### Notes

---

To request permissions go to:  
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:  
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:  
<http://group.bmj.com/subscribe/>