



Dysbiosis and relapse-related microbiome in inflammatory bowel disease: A shotgun metagenomic approach



Gerard Serrano-Gómez^a, Luis Mayorga^a, Iñigo Oyarzun^a, Joaquim Roca^b, Natalia Borrueal^c, Francesc Casellas^c, Encarna Varela^a, Marta Pozuelo^a, Kathleen Machiels^d, Francisco Guarner^a, Severine Vermeire^{d,e}, Chaysavanh Manichanh^{a,f,g,*}

^a Gut Microbiome Lab, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain

^b Molecular Biology Institute of Barcelona (IBMB), Spanish National Research Council (CSIC), Barcelona, Spain

^c Crohn-Colitis Care Unit, Digestive Diseases, Hospital Universitari Vall d'Hebron, Spain

^d Translational Research Center for Gastrointestinal Disorders (TARGID), Department of Chronic Diseases Metabolism and Ageing, KU Leuven, Leuven, Belgium

^e Department of Gastroenterology and Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium

^f Departament de Medicina, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

^g Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Spain

ARTICLE INFO

Article history:

Received 1 August 2021

Received in revised form 18 October 2021

Accepted 27 November 2021

Available online 2 December 2021

Keywords:

Crohn's disease

Ulcerative colitis

Flare

Shotgun metagenomics

ABSTRACT

Crohn's disease (CD) and ulcerative colitis (UC), the two main forms of inflammatory bowel disease (IBD), affect several million people worldwide. CD and UC are characterized by periods of clinical remission and relapse. Although IBD patients present chronic alterations of the gut microbiome, called dysbiosis, little attention has been devoted to the relapse-related microbiome. To address this gap, we generated shotgun metagenomic data from the stools of two European cohorts—134 Spanish (followed up for one year) and 49 Belgian (followed up for 6 months) subjects—to characterize the microbial taxonomic and metabolic profiles present. To assess the predictive value of microbiome data, we added the taxonomic profiles generated from a previous study of 130 Americans. Our results revealed that CD was more dysbiotic than UC compared to healthy controls (HC) and that strategies for energy extraction and propionate production were different in CD compared to UC and HC. Remarkably, CD and UC relapses were not associated with alpha- or beta-diversity, or with a dysbiotic score. However, CD relapse was linked to alterations at the species and metabolic pathway levels, including those involved in propionate production. The random forest method using taxonomic profiles allowed the prediction of CD vs. non-CD with an AUC = 0.938, UC vs. HC with an AUC = 0.646, and CD relapse vs. remission with an AUC = 0.769. Our study validates previous taxonomic findings, points to different relapse-related growth and defence mechanisms in CD compared to UC and HC and provides biomarkers to discriminate IBD subtypes and predict disease activity.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Crohn's disease (CD) and ulcerative colitis (UC), the two major subtypes of inflammatory bowel disease (IBD), are chronic intestinal disorders characterized by periods of relapse and remission. Distinguishing CD from non-CD and UC from non-UC remains a challenge as this process is traditionally based on a combination

of clinical, endoscopic, histological, and radiological criteria [1]. However, proper distinction between CD and UC can still not be achieved in up to 10% of patients [2].

Although the exact cause of CD and UC remains unknown, studies have shown that gut microbiota plays a key role in IBD development, with patients who suffer from this disorder showing an imbalance in the microbial community, known as dysbiosis [3–6]. This dysbiosis is characterized by reduced microbial diversity, a shift in the abundance of Firmicutes phylum bacteria towards Proteobacteria [7–9], and disruption of metabolites, including acyl-carnitines, bile acids, and short-chain fatty acids (SCFAs) [10]. SCFAs have often been associated with gut inflammatory processes

* Corresponding author at: Gut Microbiome Group, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain.

E-mail address: cmanicha@gmail.com (C. Manichanh).

and IBD [11,12], as they can modulate the host's immune response [13,14].

Although these findings are highly relevant to explain differences between healthy individuals and IBD patients, little attention has been given to predicting clinical relapse in IBD. CD and UC patients have a 44% and 38% risk of relapse, respectively, after discontinuation of medication such as anti-TNFs [15]. An attempt using clinical, genetic, endoscopic, histological, serological and faecal markers to predict relapse has not yet provided a reliable prognostic tool [16]. Furthermore, approximately 75% of patients with CD will ultimately require surgery due to disease complications or unsuccessful treatment [17]. This surgery involves the resection of the inflamed section of the intestine but it does not provide a cure. Indeed, most of these patients will suffer a relapse known as postoperative recurrence (POR), making it necessary for them to undergo another surgical intervention [18]. Although many clinical risk factors, such as active smoking, perforating disease, shorter disease duration, and CD development at a younger age, have been associated with POR, these indicators alone are insufficient to predict the outcome of patients both after a remission period and after surgery [17].

Few studies have evaluated microbiome data in predicting relapse in IBD or, more specifically, in POR. Microbiome data based on bacterial taxonomic profiling [19,20] and bacterial and fungal loads [21] have provided promising predictive tools for relapse in POR and in a follow-up study, respectively. However, as most of these studies were performed with 16S rRNA data, further insights into the metabolic pathways involved in relapse or POR could not be achieved.

In the present study, we performed a shotgun sequencing metagenomics analysis of faecal samples collected from healthy individuals, patients of the two major IBD subtypes, and CD patients who needed surgical resection due to disease severity. The objectives were to find an association between potential microbial taxonomic (species level) and metabolic pathway profiles and disease course, and to propose a tool to discriminate CD from non-CD, and UC from non-UC, and to predict relapse.

2. Methods

2.1. Study cohorts

To evaluate microbiome changes related to disease activity, we recruited two discovery cohorts of patients in a longitudinal setting. A Spanish cohort recruited at the Hospital Vall d'Hebron (Barcelona, Spain) and a Belgian cohort recruited at the University Hospital Leuven (Belgium), both with the approval of their respective local Ethics Committees. To evaluate the predictive value of microbiome data, we also included an American cohort retrieved from the Inflammatory Bowel Disease Multi'omics Database (<https://ibdmdb.org/>).

The Spanish cohort consisted of CD (n = 34) and UC (n = 33) patients who were clinically in remission at baseline and were followed up for one year (Supplementary Table 1). This cohort was part of a previous study and clinical data of patients could be found in Pascal et al. [5]. They all provided a faecal sample at baseline and at relapse or at month 12 if they remained in remission. Clinical diagnosis of patients from this cohort was based on the Harvey-Bradshaw score for CD and the Colitis Activity Index for UC. Non-IBD subjects (n = 67) were included as healthy controls (HC), and these subjects provided faecal samples only at baseline. For all participants, the use of antibiotics during the two months before entering the study was an exclusion criterion. A total of 212 faecal samples were processed for DNA shotgun sequencing.

The Belgian cohort consisted of CD patients (n = 49) who needed surgical intervention due to therapy failure and/or development of penetrating and/or stricturing complications of the disease. Patients' characteristics can be found in Supplementary Table 2. These patients were at risk of developing POR, which was defined by a Rutgeerts score \geq i2b. This cohort was also part of a previous study [19]. A total of 98 faecal samples were provided before surgery and at month 6 post-surgery and were processed for DNA shotgun sequencing.

The American cohort consisted of 1638 stool samples (750 CD, 459 UC, 429 non-IBD) collected from 65 CD patients, 38 UC patients and 27 non-IBD subjects. Illumina sequencing libraries were prepared using the Nextera XT DNA Library Preparation kit (Illumina), following the protocol recommended by the manufacturer. Libraries were sequenced on HiSeq 2x101 to yield ~ 10 million PE reads/sample.

2.2. DNA shotgun sequencing

For the Spanish and Belgian cohorts, a frozen aliquot (200 mg) of each faecal sample was suspended in 250 μ l of guanidine thiocyanate–0.1 M Tris (pH 7.5) and 40 μ l of 10% N-lauroyl sarcosine. DNA was then extracted as described in [21].

For the Spanish cohort, the DNA shotgun library was prepared as described in Qin et al. [22]. The sequencing procedure was performed by the Beijing Genomics Institute (BGI) using the Illumina HiSeq platform and following Illumina standards [22]. The sequencing process provided an average of 47 million paired-end sequence reads per sample.

For the Belgian cohort, the DNA shotgun library was prepared using the Nextera XT DNA Library Prep Kit and the Illumina HiSeq sequencing platform. The sequencing process provided an average of 95 million paired-end sequence reads per sample. Sequence data will be deposited in the NCBI database following publication.

2.3. Upstream sequence analysis: Quality control, decontamination and profiling

The KneadData v0.7.4 pipeline (<https://huttenhower.sph.harvard.edu/kneaddata>) was used to pre-process and decontaminate the sequence reads. KneadData performed a quality filtering of the reads using trimmomatic [23] and then mapped them against a human reference genome database using bowtie 2 [24]. Reads with lengths below 50% of the total input read length and also those that mapped with the human genome were discarded from further analysis.

The HumanN3 pipeline [25] was then used to map the sequences against the UniRef90 database with default parameters to obtain taxonomic and functional profiles. Taxonomic profiles were provided by the MetaPhlan's intermediary output file in the HumanN3 pipeline and functional profiles from the final output. Since HumanN3 does not natively support paired-end reads as input, the corresponding FASTQ files were concatenated before sequence processing. All feature abundances that did not exceed 0.1% of the data with a minimum prevalence of 10% of the total samples were excluded from further analysis.

2.4. Downstream sequence and statistical analysis

Taxonomic profiles, output of HumanN3, were generated in stratified relative abundance, from phylum to species level. For this reason, no normalization was applied, but the stratified relative abundances were extracted according to the taxonomic level of interest (species level). The Chao1 and Shannon indexes were calculated to estimate microbial alpha-diversity. Weighted and

unweighted UniFrac distances were calculated to estimate beta-diversity [26].

A dysbiosis estimator or score, as proposed by Lloyd-Price et al. [10], was calculated using the unweighted UniFrac distances. This score was defined as the median unweighted UniFrac distance of any given sample with a reference set that was built with 67 samples from HC. Two-sided Mann-Whitney U tests were performed to compare dysbiosis scores between groups of samples with a target false discovery rate (FDR) < 0.05.

Since MetaCyc pathways (default pathways output of HumanN3) were not gut-specific and lacked some relevant metabolic routes for gut homeostasis, gene families were regrouped (summed) into KO (KEGG Orthology) categories and then mapped against gut metabolic modules using the omixer-rpmR package in R [27]. The resulting modules were regrouped by identifier and then normalized using the trimmed mean of M-values (TMM) of the edgeR package in R [28], and they were also log-transformed and sum-normalized to copies per million (CoPM) instead of counts per million (CPM) since the HumanN3 output data were presented in reads per kilobase (RPK). CoPM could be considered as transcripts per million (TPM) in RNA-Seq.

Statistical analyses were performed in R (version 4.0.3). Principal coordinate analysis (PCoA) was performed using the phyloseq package [29] based on unweighted and weighted UniFrac distances for taxonomical profiles and on Bray-Curtis distances [30] for functional profiles. The contribution of clinical factors to the effect size was calculated employing permutational analysis of variance (PERMANOVA) using the adonis2 function from R's vegan package (<https://cran.r-project.org/web/packages/vegan/index.html>). Differential abundances were assessed using the limma-voom method [31]. Limma is a Bioconductor R package that allows the building of linear multivariable models to test for differentially abundant features. Limma has a feature that allows adjustment for mean-variance between samples. This feature, known as voom, enables the inclusion of covariates in the model to account for confounding variables and reduce type I and type II errors. This method also allows adjustment for unknown variables (surrogate variables) using the sva package. The models did not consider any covariates such as patients' characteristics and treatments, apart from patient identifiers on paired analyses because they were not found to be significantly different after the PERMANOVA test (Supplementary Table 3). Significance tests were performed using the eBayes function provided by the limma package. This function computed moderated t-statistics, F-statistics and log-odds of differential abundance by empirical Bayes moderation of the standard errors towards a global value. P-values were subjected to multiple hypothesis testing correction using the Benjamini-Hochberg method with an FDR threshold of 0.15. The statistical power of the models was assessed by plotting the distribution of p-values of the model without including covariates, the model including known covariates, and the model including known and unknown covariates. Surrogate variables were included in the model only when they improved the statistical power of the test. For interpretation purposes, we considered the results most relevant as those with an FDR below 0.05 and a log₂FC cut-off of 0.58 (fold-change of approximately 1.5).

2.5. Construction of predictive models using random forest

Random forest is a supervised learning algorithm that constructs a model with multiple decision trees. At each tree split, the model uses a given number of variables that are bootstrapped randomly. The output of the random forest is the average of the prediction of all the decision trees. We used random forest to predict disease vs. non-disease or disease activity based on the microbial signatures discovered in the differential abundance analysis,

and as implemented in the R package caret (Max Kuhn 2021; R package version 6.0–86. <https://CRAN.R-project.org/package=caret>). The default number of trees (500) was applied and the number of variables at each split (mtry) that presented a better ROC (receiver operating characteristic) was selected. The model was trained with 2/3 of the training set using 10-fold cross-validation. The test set consisted of 1/3 of the cohort of interest. A validation set was also included.

3. Results

3.1. Alpha- and beta-diversity, and dysbiosis scores in IBD subtypes and activity.

To confirm previous reports on diversity in the different IBD subtypes and in remission [5,8,10], we calculated the Chao1 and Shannon indexes and the dysbiosis scores based on the taxonomic profiles issued from the DNA shotgun sequencing data of the Spanish cohort. Our findings mirrored previous 16S rRNA-based observations showing that CD was associated with a significant reduction in richness (Chao1 index) and evenness (Shannon index) compared to HC and UC, while UC did not display lower diversity compared to HC (Fig. 1A and B). Beta-diversity analysis using weighted and unweighted UniFrac distances also confirmed previous results showing that, at the global microbial composition level, the microbial community of CD patients formed a separate cluster from that of HC and UC (PERMANOVA, $p < 0.001$), while the UC microbiome was similar to that of HC (PERMANOVA, $p = 0.25$) (Fig. 1C and D).

To evaluate the level of dysbiosis, we calculated the dysbiosis score based on unweighted UniFrac dissimilarities to non-IBD metagenomes, as described by Lloyd-Price et al. [10]. This analysis revealed a clear alteration in CD compared to HC (two-sided Mann-Whitney U test, $p = 3.2E-13$) and UC (Mann-Whitney U test, $p = 1.7E-12$) (Fig. 1I). CD and UC patients were followed up for 12 months and faecal samples were collected and analysed around the relapse period (using the Harvey-Bradshaw Index for CD and the Colitis Activity Index for UC) or at month 12 if they remained in remission. Findings showed that neither alpha-diversity nor dysbiosis scores were successful in differentiating disease activity status (relapse vs. remission) for either CD or UC (Fig. 1E, F, G, H, J, and K).

The Belgian cohort showed similarity with the Spanish cohort when patients were in remission but not when they underwent a flare, as assessed by beta-diversity ($p < 0.05$ for weighted and unweighted UniFrac distances) analysis (Supplementary Fig. 1). This observation could be explained by the fact that most of the Belgian CD patients needed surgical intervention due to therapy failure and/or the development of penetrating and/or stricturing complications of the disease, and, therefore, they presented a more severe disease phenotype than the patients in the Spanish cohort. Like the Spanish cohort, recurrence vs. remission status in the Belgian cohort did not show significant differences in alpha- or beta-diversity (Supplementary Fig. 2).

3.2. Differentially abundant species in IBD subtypes and disease activity

In the Spanish cohort, differential abundant (DA) species analysis showed an enrichment of 12 and a depletion of 23 species in CD samples ($n = 33$) compared to UC ($n = 33$) at baseline (i.e., patients in remission) and HC ($n = 67$), respectively (Fig. 2A, Supplementary Tables 4 and 5). The most significant results corroborated previous reports on the alteration of the microbiome composition of CD compared to non-CD subjects. These results included the increase

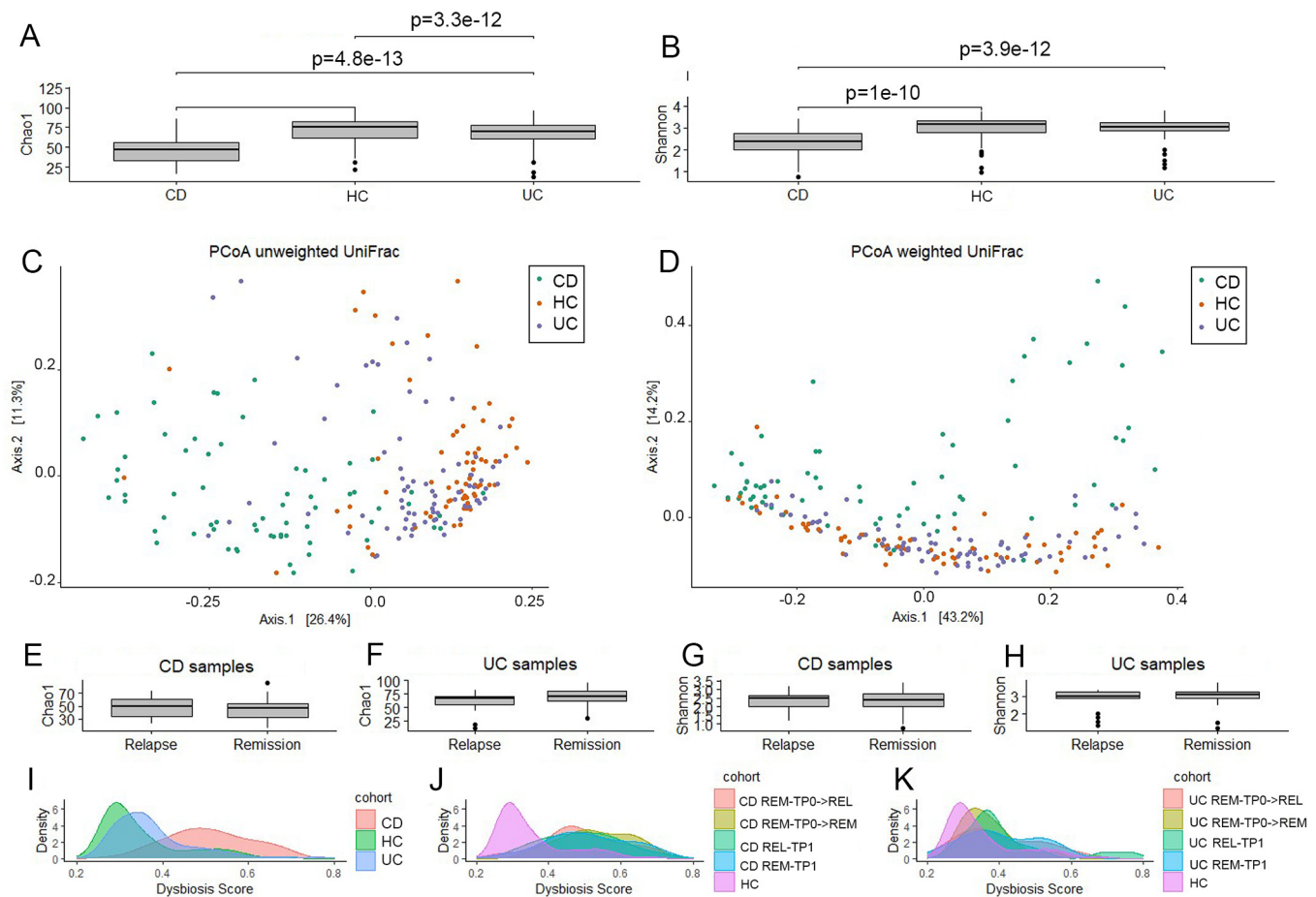


Fig. 1. Diversity and dysbiosis analyses based on the taxonomic profile of samples from the Spanish cohort. Alpha-diversity analyses using the Chao1 (A) and Shannon (B) indexes show significant differences between CD, UC and HC. Beta-diversity analyses using unweighted (C) and weighted (D) UniFrac Principal Coordinate Analysis (PCoA) clustered CD away from UC and HC. The Chao1 and Shannon indexes could not differentiate between relapse and remission in samples from CD (E and G) or UC (F and H). Dysbiosis score analyses also found significant differences between CD and UC and between CD and HC (I), but not between relapse and remission in CD (J) and UC (K). Crohn's disease (CD), ulcerative colitis (UC), and healthy control (HC).

in relative abundance of *Escherichia coli* in CD (FDR = 3.1×10^{-8} against UC, FDR = 0.0002 against HC), *Ruminococcus gnavus* (FDR = 3.4×10^{-8} against UC, FDR = 1.26×10^{-7} against HC), and *Clostridium clostridioforme* (FDR = 2.9×10^{-8} against UC, FDR = 6.9×10^{-7} against HC) in CD compared to UC, and a decrease in *Faecalibacterium prausnitzii* in CD compared to UC and HC [3,32,33]. Remarkably, *Veillonella parvula* was one of the most significantly enriched species in CD compared to UC and HC (FDR = 1.3×10^{-9} against UC, FDR = 3.8×10^{-13} against HC).

To evaluate alterations in the microbiome related to disease severity in CD, we performed various comparisons: baseline (REM-TP0->REL, samples collected from patients in remission at baseline who relapsed during the follow-up period, $n = 13$) vs. final relapse samples (REL-TP1, $n = 13$), and baseline (REM-TP0->REM, samples collected from patients in remission at baseline who remained in remission after one year of follow-up, $n = 20$) vs. final remission samples (REM-TP1, $n = 22$) in a pairwise manner, and we also compared remission (REM-TP1, $n = 22$) vs. relapse samples (REL-TP1, $n = 13$) during the follow-up. Relapse samples (REL-TP1) were depleted of 10 microbial species compared to remission samples (REM-TP1) (Supplementary Table 6). The shift from remission (REM-TP0) to relapse (REL-TP1) was not associated with significant differences. To uncover potential biomarkers of relapse, we also compared baseline samples of CD patients who remained in remission after one year (REM-TP0->REM) with those who

relapsed during the follow-up year (REM-TP0->REL). As a result, patients who relapsed (REM-TP0->REL) were enriched in three bacterial species (*Ruminococcus torques*, FDR = 2.3×10^{-7} ; *Clostridium bolteae*, FDR = 0.0004; *Fusicatenibacter saccharivorans*, FDR = 3.11×10^{-6}) compared to those who remained in remission (REM-TP0->REM) (Supplementary Table 7) (Fig. 2C).

Compared to HC ($n = 67$), UC patients ($n = 33$) were depleted in two species including *Methanobrevibacter smithii* (FDR = 0.03) (Supplementary Table 8). No species was altered in baseline samples of patients who relapsed (REL-TP0->REL, $n = 17$) when compared to relapse samples (REL-TP1, $n = 17$), but when comparing to baseline samples of patients who remained in remission (REM-TP0->REM, $n = 16$), two species were enriched (Supplementary Table 9).

In the Belgian cohort, CD patients who underwent POR ($n = 21$) presented an enrichment of *Dialister invisus*, *Ruthenibacterium lactatiformis*, *Parabacteroides distasonis* and *Bacteroides stercoris*, among others (FDR < 0.05), but a depletion of *Ruminococcus lactaris* (FDR = 0.048) compared to those who remained in remission ($n = 28$) (Supplementary Table 10). CD patients in remission in both the Spanish ($n = 33$) and Belgian ($n = 28$) cohorts did not present significant microbial differences, as reflected by the PERMANOVA test on weighted UniFrac distances, while patients who underwent a flare showed significant differences (PERMANOVA, $p = 0.038$, $n = 13$ for the Spanish and $n = 70$ for the Belgian cohort).

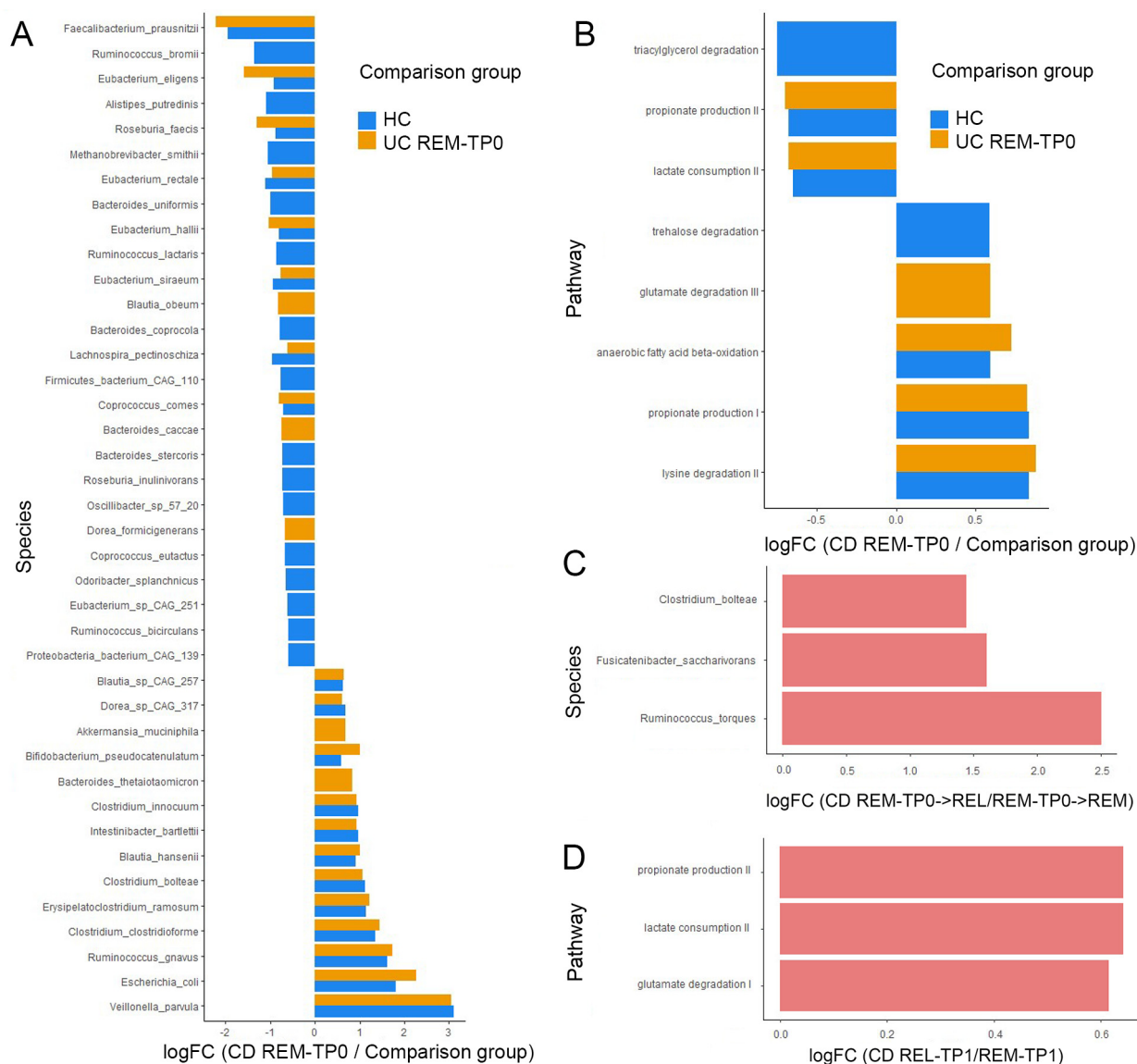


Fig. 2. Differential abundance analysis of taxonomic and metabolic pathway profiles. Log2FC of significant differentially abundant species (A) and pathways (B) between baseline samples of CD and UC/Hc (positive logFC refers to enrichment in CD). (C) Log2FC of significant differentially abundant species between CD REM-TP0->REL samples and REM-TP0->REM samples (positive logFC refers to enrichment in REM-TP0->REL samples). (D) Log2FC of significant differentially abundant pathways between REL-TP1 samples and REM-TP1 samples (positive logFC refers to increase in relapse). Threshold for significance: $q < 0.05$ and $\log_2FC > 0.58$. Crohn's disease (CD), ulcerative colitis (UC), healthy control (HC); REM-TP0->REL: samples collected from patients in remission at baseline who relapsed during the follow-up period; REM-TP0->REM: samples collected from patients in remission at baseline who remained in remission after one year of follow-up; REL-TP1: samples collected from patients who relapsed; REM-TP1: samples collected from patients who remained in remission. Threshold for significance: $q < 0.05$ and $\log_2FC > 0.58$.

3.3. Differentially abundant pathways in IBD subtypes and disease activity

In the Spanish cohort, DA pathway analysis of samples collected at baseline (i.e., patients in remission) showed that four Gut Metabolic Modules (GMMs) were enriched and two were depleted in CD ($n = 33$) compared to UC ($n = 33$), while four were enriched and three were depleted in CD ($n = 33$) compared to HC ($n = 67$). Among the implicated pathways, propionate production I, lysine degradation II, and anaerobic fatty acid beta-oxidation were enriched in CD compared to both UC and HC ($FDR < 1E-04$). The trehalose degradation and glutamate degradation III pathways were found to be depleted in HC and UC, respectively, compared to CD. Propionate production II and lactate consumption II were depleted in CD compared to HC and UC ($FDR < 1E-04$) (Fig. 2B, Supplementary Tables

11 and 12). Interestingly, *Escherichia coli* was the main contributor of the three enriched pathways and *Anaerostipes hadrus* was the main contributor of the two depleted pathways (Fig. 3). No GMMs were found differentially abundant in UC samples at baseline compared to HC.

Samples from CD patients who relapsed (REL-TP1, $n = 13$) were enriched in three GMMs, including propionate production II ($FDR = 0.009$), lactate consumption II ($FDR = 0.009$) and glutamate degradation I ($FDR = 0.02$), compared to those from CD patients in remission (REM-TP1, $n = 22$), with *Anaerostipes hadrus* as the main contributor for the two first pathways and *Acidaminococcus intestinalis* for the last one (Fig. 2D; Supplementary Table 13). However, in UC, a shift from remission (REM-TP0->REL, $n = 17$) to relapse (REL-TP1, $n = 17$) was not associated with any significant changes in the pathways. No microbial markers of UC relapse were found

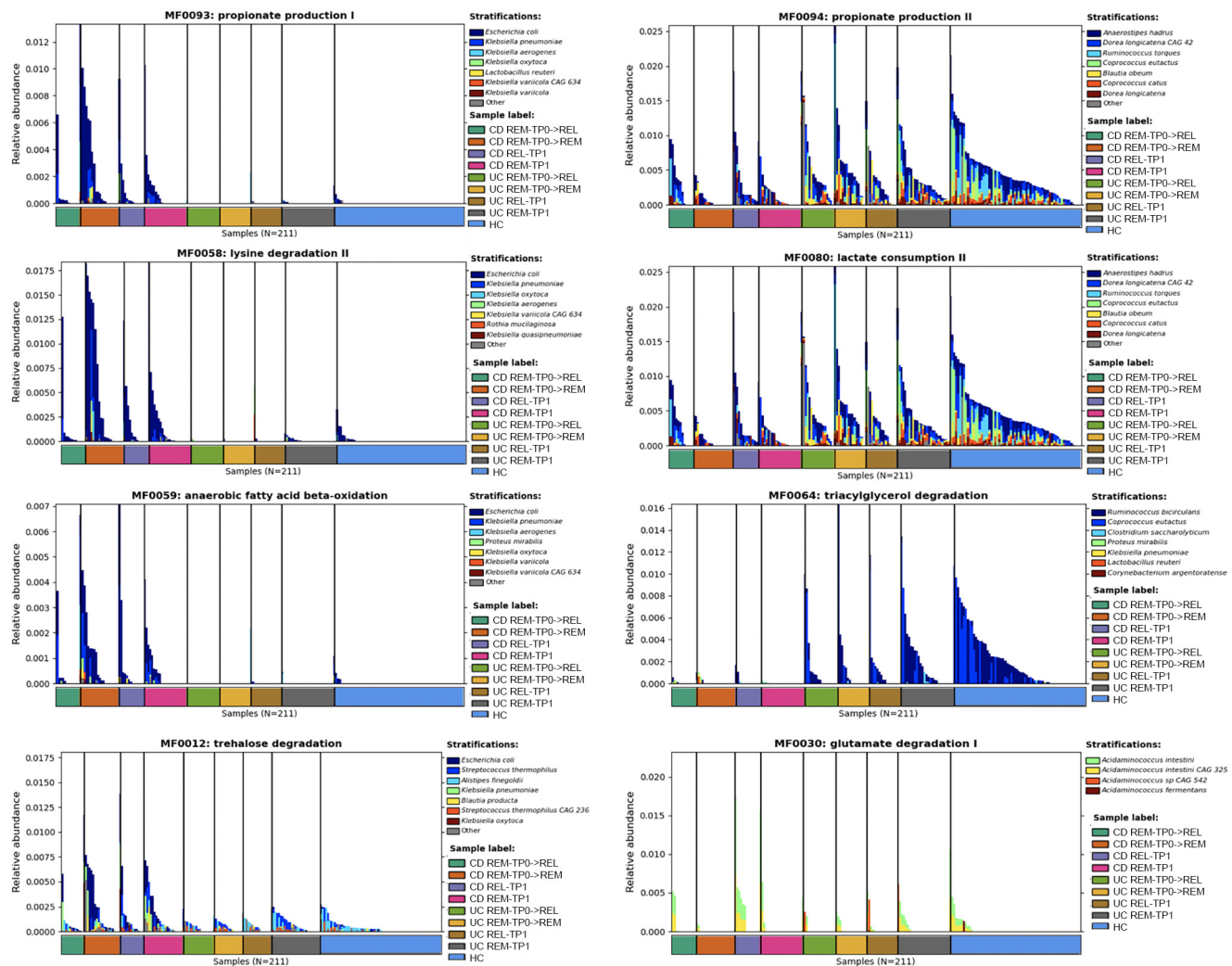


Fig. 3. Contribution of species to different metabolic pathways. *Escherichia coli* was the most important contributor to the propionate production I, lysine degradation II and anaerobic fatty acid beta-oxidation pathways, which were enriched in CD compared to UC and HC, and to the trehalose degradation pathway, enriched in CD compared to UC alone. *Anaerostipes hadrus* was the main contributor to the propionate production II and lactate consumption II pathways, which were depleted in CD compared to UC and HC. The propionate production II, lactate consumption II and glutamate degradation I pathways were also enriched in CD REL-TP1 compared to CD REM-TP1 samples, *Acidaminococcus intestini* being the main contributor of the latter. REM-TP0->REL: samples collected from patients in remission at baseline who relapsed during the follow-up period; REM-TP0->REM: samples collected from patients in remission at baseline who remained in remission after one year of follow-up; REL-TP1: samples collected from patients who relapsed; REM-TP1: samples collected from patients who remained in remission. Threshold for significance: $q < 0.05$ and $\log_2FC > 0.58$.

(REM-TP1, $n = 27$ vs. REL-TP1, $n = 17$) and no predictive functional biomarkers of relapse could be recovered from UC baseline samples (REM-TP0->REM vs. REM-TP0->REL).

In the CD Belgian cohort, no pathway was found to be significantly associated with relapse. Only one pathway in samples at baseline (patients with active disease, $n = 21$) showed a trend towards enrichment (lactate consumption II, FDR = 0.13) compared to samples from patients who remained in remission 6 months after surgery ($n = 28$).

3.4. Predictive value of the microbiome data to discriminate IBD from non-IBD, and relapse from remission

To assess the predictive value of the microbial signatures discovered and their consistency between cohorts, we built random forest models based on the abundance of microbiome species to predict health status and disease activity. To this end, in addition to the Spanish cohort, which allowed the discovery of potential biomarkers of IBD vs. non-IBD and IBD subtypes, and the Belgian

cohort, we included another large and longitudinal cohort (American cohort) retrieved from the Inflammatory Bowel Disease Multi-omics Database (<http://ibdmdb.org>). The American cohort consisted of 1638 stool samples (CD, $n = 750$; UC, $n = 459$; non-IBD, $n = 429$) collected from 65CD patients, 38 UC patients and 27 non-IBD subjects.

First, to distinguish CD from HC and UC, we built a model using the 26 microbial species found altered between CD-TP0 and HC and/or between CD-TP0 and UC-TP0 as predictor variables (*Bacteroides caccae*, *Bifidobacterium pseudocatenulatum*, *Blautia hansenii*, *Blautia obeum*, *Blautia sp* CAG 257, *Clostridium bolteae*, *Clostridium clostridioforme*, *Clostridium innocuum*, *Coprococcus comes*, *Dorea formicigenerans*, *Dorea sp* CAG 317, *Eggerthella lenta*, *Erysipelatoclostridium ramosum*, *Escherichia coli*, *Eubacterium eligens*, *Eubacterium hallii*, *Eubacterium rectale*, *Eubacterium siraeum*, *Faecalibacterium prausnitzii*, *Intestinibacter bartlettii*, *Lachnospira pectinoschiza*, *Roseburia faecis*, *Ruminococcus gnavus*, *Ruminococcus lactaris*, *Ruthenibacterium lactatiformans* and *Veillonella parvula*). The American cohort was used as a training ($\frac{2}{3}$) and test ($\frac{1}{3}$) data-

set because of its large size, and the European (Spanish + Belgian) cohort as a validation dataset. The model achieved an area under the receiver operating characteristic (ROC) curve (AUC) of 0.868 on the training set (CD = 500, non-CD = 592) using a 10-fold cross-validation, 0.908 on the testing set (CD = 250, non-CD = 296), and 0.938 on the validation European cohort (CD = 166, non-CD = 144) (Fig. 4A).

Second, to discriminate UC from HC, we built a random forest model using the abundance of the 9 differentially abundant species uncovered between UC-TP0 samples and HC (*Bifidobacterium pseudocatenulatum*, *Blautia hydrogenotrophica*, *Clostridium* sp. CAG 242, *Eubacterium* sp. CAG 38, *Firmicutes bacterium* CAG 110, *Gemmiger formicilis*, *Methanobrevibacter smithii*, *Oscillibacter* sp.57 20 and *Paraprevotella xylaniphila*). The American cohort was used as a training ($\frac{2}{3}$) and test ($\frac{1}{3}$) dataset and the Spanish cohort as a validation dataset. This model achieved a ROC AUC of 0.736 in the training set (UC, n = 306; HC, n = 286), 0.765 on the testing set (UC, n = 153; HC, n = 143) and 0.646 on the validation cohort (UC, n = 77; HC, n = 67) (Fig. 4B).

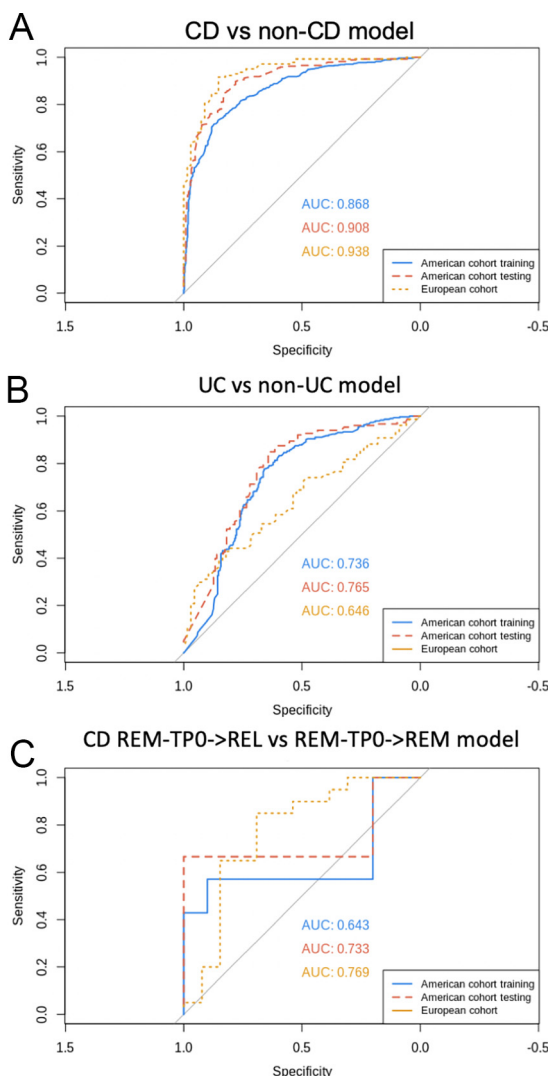


Fig. 4. Area under the receiver operating characteristic (ROC) curve (AUC) of the random forest models built from the abundance of predictor species. (A) The CD vs. non-CD model was built using the American cohort as the training ($\frac{2}{3}$) and test ($\frac{1}{3}$) dataset because of its large size, and the European (Spanish + Belgian) cohort as the validation dataset. (B) The UC vs. non-IBD model and (C) CD baseline remission vs. relapse model were built using the American cohort as the training ($\frac{2}{3}$) and test ($\frac{1}{3}$) dataset, and the Spanish cohort as the validation dataset.

Third, to predict relapse in CD patients, we used the 11 most differentially abundant species between REM-TP0->REM and REM-TP0->REL samples (*Ruminococcus torques*, *Anaerostipes hadrus*, *Fusicatenibacter saccharivorans*, *Clostridium bolteae*, *Dorea longicatena*, *Roseburia intestinalis*, *Prevotella copri*, *Phascolarctobacterium succinatutens*, *Ruthenibacterium lactatiformans*, *Streptococcus parasanguinis* and *Bifidobacterium pseudocatenulatum*) as predictor variables. The American cohort was used as a training ($\frac{2}{3}$) and test ($\frac{1}{3}$) dataset, and the Spanish cohort as a validation dataset. In this case, the ROC AUC was 0.643 on the training set (REM-TP0->REM, n = 7; REM-TP0->REL, n = 10), 0.733 on the testing set (REM-TP0->REM, n = 3; REM-TP0->REL, n = 5) and 0.769 on the validation cohort (REM-TP0->REM, n = 20; REM-TP0->REL, n = 13) (Fig. 4C).

4. Discussion

This study sought to characterize changes in the human microbiome related to IBD relapse and their associated metabolic pathways using a shotgun metagenomics approach. To this end, the preliminary analysis evaluated dysbiosis in IBD subgroups by calculating alpha- and beta-diversity and dysbiosis scores. The results showed evidence of microbiome dysbiosis in IBD, in particular in CD compared to HC, following the trend of previous studies based on the 16S rRNA approach [5,8,10]. Indeed, our finding related to a higher dysbiosis score in CD compared to UC is in line with previous works [3,8]. However, none of these studies uncovered the differentiation between relapse and remission states in IBD subgroups. In the present work, by performing DA analysis on taxonomic and functional profiles, we were able to uncover microbial species associated with IBD and also relapse or remission, in particular in CD.

The enrichment of some species such as *Escherichia coli*, *Ruminococcus gnavus*, and *Clostridium clostridioforme* and the depletion of *Faecalibacterium prausnitzii* in CD compared to UC and HC confirmed previous findings. However, the enrichment of *Veillonella parvula* in CD is a novel finding. Indeed, *V. parvula* belongs to the genus *Veillonella*, considered a commensal found in the oral and intestinal tract, but which has occasionally been identified as a pathogen in cases of osteomyelitis [34], meningitis [35], and periodontitis [36], and could therefore improve prediction models.

SCFAs are key metabolites in the maintenance of gut homeostasis, as some exert anti-inflammatory activity and help strengthen the intestinal barrier [37,38]. These SCFAs can be produced and degraded by many pathways. This is the case of propionate, which could be synthesized by different metabolic routes in CD patients and in UC/HC. Species that contributed to these pathways also clearly differed, with *Anaerostipes hadrus* (Firmicutes), one of the main contributors to the propionate production II pathway, enriched in HC and UC, and with species from the Enterobacterales order (Proteobacteria) contributing to the propionate production I pathway, which was enriched in CD. These findings confirm the shift from Firmicutes towards Proteobacteria in CD compared to non-CD reported in previous studies [39] and suggest that the dysbiosis is related to SCFA metabolism.

E. coli, one of the most enriched species in CD, was also the main contributor to the four most enriched and the two depleted metabolic pathways in CD compared to non-CD. However, this bacterium was almost undetected in UC samples (Supplementary Fig. 3). There is some controversy regarding the role of *E. coli* in both CD and UC. Some studies have addressed the presence of *E. coli* in UC [40,41] and others reported its implication exclusively in CD, in particular in ileal biopsies [7,42,43]. Our findings not only support the presence of *E. coli* almost exclusively in CD but also reveal the potential metabolic pathways by which *E. coli* is associ-

ated with CD or may contribute to the pathogenesis of this condition.

Our understanding is that, in CD, trehalose degradation, a carbohydrate degradation pathway, and anaerobic fatty acid beta-oxidation, a lipid degradation pathway, are used to generate energy for growth, with *E. coli* being the main contributor to these pathways. Instead, in HC and UC, the triacylglycerol degradation pathway is used to generate energy, with *Ruminococcus bicirculans* and *Coprococcus eutactus* appearing as the main contributors. Propionate production in CD patients is via the lysine degradation pathway, which generates succinate, a known intermediate of propionate via the propionate production I pathway, with *E. coli* as the main contributor of these pathways [44]. In contrast, HC and UC patients use the lactate consumption II pathway, which is known to provide an intermediate of propionate in the propionate production II pathway, with *Anaerostipes hadrus* being the dominant contributor of these pathways. The choice of one or the other metabolic pathway to produce a source of carbon or energy could be determined by the accumulation of a particular intermediate or to the higher prevalence of a particular species. In the scenario in which succinate accumulates in CD patients, we hypothesize that succinate may trigger inflammation. Indeed, succinate, known to be formed by the reversal of partial tricarboxylic acid (TCA) cycle reactions or produced during bacterial fermentation of dietary fiber [45], has recently emerged as an important signal in immunity and inflammation; however, its precise role in immunity needs to be explored further.

As in previous studies, we did not find significant differences at the species level between UC and HC, or between relapse and remission status [5]. In POR-related analyses, the enrichment of several bacterial species and the depletion of *Ruminococcus lactaris* in flares were not consistent with previous findings using the 16S rRNA approach [19]. This discrepancy could be explained by the different meta-omics approach used and the different number of samples analysed in each approach ($n = 189$ in 16S analysis and $n = 98$ in shotgun DNA analysis). At the time the proposal was drafted, the cost of shotgun sequencing was too high to allow the sequencing of the 189 samples used in our previous study [19]. Therefore, further validation would be needed to relate microbiome changes and POR.

Relapse-related taxonomic and functional analyses, based on the Spanish CD cohort, uncovered considerable alteration towards the loss of beneficial microbial species in relapse compared to remission samples. This result suggests that flare is associated with a loss of beneficial bacteria rather than a gain of potential pathogens. However, the enrichment of three bacterial species, two of which, *Ruminococcus Torques* and *Clostridium bolteae* [46,47], associated with autism, in samples of patients who were in remission at baseline and who relapsed during the one-year follow-up may suggest a predisposition to relapse.

One of the significant metabolic pathways used by CD patients in flare to generate energy is the glutamate degradation I pathway. Degradation of glutamate usually produces butyrate or its precursors, but the glutamate degradation I pathway, which was significantly increased in CD relapse compared to CD remission, only produces 2-oxoglutarate. This compound can be incorporated into the TCA, but it can also be used to produce more glutamate. These considerations may indicate that patients who underwent a flare produced butyrate and its precursors in a less efficient way than those who remained in remission, thus decreasing their protection against a more severe disease state.

To predict the IBD subtype, we built several random forest models based on the species found in the statistical analyses. Previous studies [19,21] have shown that the combination of clinical and microbiota data enhances the performance of this kind of model. However, due to the unavailability of common relevant clinical

data between the train/test cohort and the validation cohort, we used species abundance as the only predictor variable in the models. Despite expecting reduced performance due to the lack of clinical data, the CD vs. non-CD model (AUC = 0.908 on test set, AUC = 0.938 on validation cohort) performed better than models including bacterial and fungal loads alongside clinical data (AUC = 0.899) [21] and models including a combination of the whole metabolic and taxonomic profile of shotgun-metagenomic samples (AUC = 0.89 on validation cohort) [3]. The increased performance and consistency of our model across cohorts suggests that the 26 species found in this study may be highly relevant in CD development since they can be used to discriminate CD from non-CD patients.

The UC vs. non-IBD model did achieve the performance of our previous models [21]. The difference in performance between the train/test cohort and the validation cohort can lead to two main interpretations: model overfitting, or selection of inadequate predictor variables. Both explanations are plausible, but since this random forest was built with the abundance of species gathered from DA analysis with an FDR > 0.15, some of the predictors used could be false positives. The inability to find significant microbial signatures to discriminate UC patients from HC leads, once again, to the notion that the causality of UC is not as related to the microbiome as it is for CD. The relatively high performance of CD relapse prediction must also be interpreted with caution due to the low number of subjects used to build the models. This limitation points to the need for additional efforts to increase cohort size, in particular, to implement longitudinal studies that enable the inclusion of patients who may switch from remission to relapse or vice versa.

Our shotgun metagenomics study has provided valuable insight into CD and relapse-related microbiome dynamics. However, the impossibility to integrate the data of the two European cohorts due to the lack of a valid common disease activity indicator (Harvey-Bradshaw and Rutgeerts scores) could be considered a limitation. Although our findings are interesting and relevant, one should keep in mind that, given the DNA nature of our study, GMM information is based on potential metabolic pathways rather than expressed metabolic pathways. The approach described herein could be complemented with metatranscriptomics, which would provide information about the pathways expressed in the gut by the microbiota. It may also be possible to integrate both metagenomics and metatranscriptomics approaches to normalize RNA transcription by DNA gene copy number and obtain information about the pathways differentially expressed in distinct conditions, thus providing more detailed insights into the metabolic mechanisms of gut microbiota in IBD.

Notwithstanding these limitations, we can draw the following conclusions: 1) CD is more dysbiotic than UC both at the taxonomic and functional levels; 2) the production of SCFAs, such as propionate, is different in CD compared to UC/HC individuals. While this production is driven by Firmicutes phylum bacteria in HC and UC patients, it is driven by Proteobacteria in CD patients; and 3) the random forest method using microbiome taxonomic profiles allows good prediction of CD vs. non-CD and a satisfactory prediction of UC vs. HC, and relapse vs remission.

Funding

This study was supported by the Instituto de Salud Carlos III / FEDER, a government agency (grant numbers: PI17/00614; PI20/00130), and by the Crohn's & Colitis Foundation of America (Award ID: 514634).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.11.037>.

References

- Tontini GE, Vecchi M, Pastorelli L, Neurath MF, Neumann H. Differential diagnosis in inflammatory bowel disease colitis: state of the art and future perspectives. *World J Gastroenterol* 2015;21(1):21–46.
- Monteiro S, Castro FDD, Carvalho PB, Rosa B, Moreira MJ, Pinho R, et al. Essential role of small bowel capsule endoscopy in reclassification of colonic inflammatory bowel disease type unclassified. *World J Gastrointest Endosc* 2017;9(1):34. <https://doi.org/10.4253/wjge.v9.i1.34>.
- Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol* 2019;4(2):293–305.
- Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13(9):R79. <https://doi.org/10.1186/gb-2012-13-9-r79>.
- Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. *Gut* 2017;66(5):813–22.
- Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 2003;52:237–42.
- Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, Yantiss R, 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(5):403–18.
- Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* 2017;2(5). <https://doi.org/10.1038/nmicrobiol.2017.4>.
- Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;55:205–11.
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 2019;569(7758):655–62.
- Kumari R, Ahuja V, Paul J. Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India. *World J Gastroenterol* 2013;19:3404–14.
- Machiels K, Joossens M, Sabino J, De Preter V, Arijis I, Eeckhaut V, et al. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* 2014;63(8):1275–83.
- Haghikia A, Jörg S, Duscha A, Berg J, Manzel A, Waschbisch A, et al. Dietary Fatty Acids Directly Impact Central Nervous System Autoimmunity via the Small Intestine. *Immunity* 2015;43(4):817–29.
- Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* 2013;341(6145):569–73.
- Gisbert JP, Marín AC, Chaparro M. Systematic review: factors associated with relapse of inflammatory bowel disease after discontinuation of anti-TNF therapy. *Aliment Pharmacol Ther* 2015;42(4):391–405.
- Liverani E, Scafoli E, Digby RJ, Bellanova M, Belluzzi A. How to predict clinical relapse in inflammatory bowel disease patients. *World J Gastroenterol* 2016;22:1017–33.
- De Cruz P, Kamm MA, Prideaux L, Allen PB, Desmond PV. Postoperative recurrent luminal Crohn's disease: a systematic review. *Inflamm Bowel Dis* 2012;18:758–77.
- Assche GV, Rutgeerts P. Medical management of postoperative recurrence in Crohn's disease. *Gastroenterol Clin North Am* 2004;33(2):347–60.
- Machiels K, Pozuelo del Río M, Martínez-De la Torre A, Xie Z, Pascal Andreu V, Sabino J, et al. Early Postoperative Endoscopic Recurrence in Crohn's Disease Is Characterised by Distinct Microbiota Recolonisation. *J Crohns Colitis* 2020;14(11):1535–46.
- Sokol H, Brot L, Stefanescu C, Auzolle C, Barnich N, Buisson A, et al. Prominence of ileal mucosa-associated microbiota to predict postoperative endoscopic recurrence in Crohn's disease. *Gut* 2020;69(3):462–72.
- Sarrabayrouse G, Elias A, Yáñez F, Mayorga L, Varela E, Bartoli C, et al. Fungal and Bacterial Loads: Noninvasive Inflammatory Bowel Disease Biomarkers for the Clinical Setting. *mSystems* 2021;6(2). <https://doi.org/10.1128/mSystems.01277-20>.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464(7285):59–65.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114–20.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9.
- Franzosa EA, Mciver LJ, Rahnvar G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 2018;15(11):962–8.
- Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 2007;73:1576–85.
- Darzi Y, Falony G, Vieira-Silva S, Raes J. Towards biome-specific analysis of meta-omics data. *ISME J* 2016;10(5):1025–8.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139–40.
- McMurdie PJ, Holmes S, Watson M. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8(4):e61217.
- Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr* 1957;27(4):325–49.
- Ritchie ME, Phipson B, Wu Di, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47.
- Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. *Ruminococcus gnavus*, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci U S A* 2019;116(26):12672–7.
- Joossens M, Huys G, Cnockaert M, De Preter V, Verbeke K, Rutgeerts P, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 2011;60(5):631–7.
- Fisher RG, Denison MR. *Veillonella parvula* bacteremia without an underlying source. *J Clin Microbiol* 1996;34(12):3235–6.
- Bhatti MA, Frank MO. *Veillonella parvula* meningitis: case report and review of *Veillonella* infections. *Clin Infect Dis* 2000;31(3):839–40.
- Hoare A, Wang H, Meethil A, Abusleme L, Hong B-Y, Moutsopoulos NM, et al. A cross-species interaction with a symbiotic commensal enables cell-density-dependent growth and in vivo virulence of an oral pathogen. *ISME J* 2021;15(5):1490–504.
- Cox MA, Jackson J, Stanton M, Rojas-Triana A, Bober L, Lavery M, et al. Short-chain fatty acids act as antiinflammatory mediators by regulating prostaglandin E(2) and cytokines. *World J Gastroenterol* 2009;15:5549–57.
- Peng L, Li Z-R, Green RS, Holzman IR, Lin J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *J Nutr* 2009;139(9):1619–25.
- Vester-Andersen MK, Mirsepasi-Lauridsen HC, Prossberg MV, Mortensen CO, Træger C, Skovsen K, et al. Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Sci Rep* 2019;9(1). <https://doi.org/10.1038/s41598-019-49833-3>.
- Elliott TR, Hudspeth BN, Wu G, Cooley M, Parkes G, Quiñones B, et al. Quantification and characterization of mucosa-associated and intracellular *Escherichia coli* in inflammatory bowel disease. *Inflamm Bowel Dis* 2013;19(11):2326–38.
- Negróni A, Costanzo M, Vitali R, Superti F, Bertuccini L, Tinari A, et al. Characterization of adherent-invasive *Escherichia coli* isolated from pediatric patients with inflammatory bowel disease. *Inflamm Bowel Dis* 2012;18(5):913–24.
- Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser A-L, Barnich N, et al. High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. *Gastroenterology* 2004;127(2):412–21.
- Rhodes JM. The role of *Escherichia coli* in inflammatory bowel disease. *Gut* 2007;56(5):610–2.
- Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J* 2014;8(6):1323–35.
- Connors J, Dawe N, Van Limbergen J. The Role of Succinate in the Regulation of Intestinal Inflammation. *Nutrients* 2018;11(1):25. <https://doi.org/10.3390/nu11010025>.
- Pandit L, Cox LM, Malli C, D'Cunha A, Rooney T, Lokhande H, et al. Clostridium bolteae is elevated in neuromyelitis optica spectrum disorder in India and shares sequence similarity with AQP4. *Neurol Neuroimmunol Neuroinflamm* 2021;8(1):e907. <https://doi.org/10.1212/NXI.0000000000000907>.
- Wang Lv, Christophersen CT, Soric MJ, Gerber JP, Angley MT, Conlon MA. Increased abundance of *Sutterella* spp. and *Ruminococcus* torques in feces of children with autism spectrum disorder. *Mol. Autism* 2013;4(1). <https://doi.org/10.1186/2040-2392-4-42>.