



# Development of a Novel Metagenomic Biomarker for Prediction of Upper Gastrointestinal Tract Involvement in Patients With Crohn's Disease

Min Seob Kwak\*, Jae Myung Cha, Hyun Phil Shin, Jung Won Jeon and Jin Young Yoon

Department of Internal Medicine, Kyung Hee University Hospital at Gangdong, College of Medicine, Kyung Hee University, Seoul, South Korea

## OPEN ACCESS

### Edited by:

Hyundoo Hwang,  
BBB Inc., South Korea

### Reviewed by:

Luiz Gustavo Gardinassi,  
Universidade Federal de Goiás  
(IPTSP – UFG), Brazil  
Guillaume Sarrabayrouse,  
Vall d'Hebron Research Institute  
(VHIR), Spain

### \*Correspondence:

Min Seob Kwak  
kwac63@khu.ac.kr

### Specialty section:

This article was submitted to  
Microbial Immunology,  
a section of the journal  
Frontiers in Microbiology

Received: 19 March 2020

Accepted: 06 May 2020

Published: 03 June 2020

### Citation:

Kwak MS, Cha JM, Shin HP,  
Jeon JW and Yoon JY (2020)  
Development of a Novel Metagenomic  
Biomarker for Prediction of Upper  
Gastrointestinal Tract Involvement  
in Patients With Crohn's Disease.  
*Front. Microbiol.* 11:1162.  
doi: 10.3389/fmicb.2020.01162

The human gut microbiota is an important component in the pathogenesis of Crohn's disease (CD), promoting host–microbe imbalances and disturbing intestinal and immune homeostasis. We aimed to assess the potential clinical usefulness of the colonic tissue microbiome for obtaining biomarkers for upper gastrointestinal (UGI) tract involvement in CD. We analyzed colonic tissue samples from 26 CD patients (13 with and 13 without UGI involvement at diagnosis) from the Inflammatory Bowel Disease Multi-Omics Database. QIIME1, DiTaxa, linear discriminant analysis effect size (LEfSe), and PICRUST2 methods were used to examine microbial dysbiosis. Linear support vector machine (SVM) and random forest classifier (RF) algorithms were used to identify the UGI tract involvement-associated biomarkers. There were no statistically significant differences in community richness, phylogenetic diversity, and phylogenetic distance between the two groups of CD patients. DiTaxa analysis predicted significant association of the species *Ruminococcus torques* with UGI involvement, which was confirmed by the LEfSe analysis ( $P = 0.025$ ). For the feature ranking method in both linear SVM and RF models, the species *R. torques* and age at diagnosis contributed to the combined models. The L-methionine biosynthesis III ( $P = 0.038$ ) and palmitate biosynthesis II ( $P = 0.050$ ) were under-represented in CD with UGI involvement. These findings suggest that *R. torques* might serve as a novel potential biomarker for UGI involvement in CD and its correlations, in addition to a range of bacterial species. The mechanisms of interaction between hosts and *R. torques* should be further investigated.

**Keywords:** microbiome, 16S rRNA, Crohn's disease, upper gastrointestinal tract, biomarker

## INTRODUCTION

Crohn's disease (CD) is a heterogeneous disorder with a multifactorial etiology, including genetic factors, host immune system, environmental factors, and gut microbiota, and is characterized by chronic relapsing transmural inflammation which can affect the gastrointestinal tract (Strober et al., 2007). It may affect any part of the gastrointestinal tract, from the mouth to the perianal area, although the terminal ileum and the right colon are the most commonly affected sites

(Strober et al., 2007; Gajendran et al., 2018). Previous studies estimated the prevalence of CD patients affected with an upper gastrointestinal (UGI) tract involvement at 16–34% for adults (Lenaerts et al., 1989; Cameron, 1991; Kefalas, 2003; Castellaneta et al., 2004; Van Limbergen et al., 2008) and 26–54% for children (Annunziata et al., 2012; Diaz et al., 2015). The UGI tract involvement in CD represents a risk of complications, such as stricturing and fistulizing phenotypes (Bernell et al., 2000), recurrence (Wolters et al., 2006), further hospitalization (Chow et al., 2009), and surgery (Bernell et al., 2000; Wolters et al., 2006; Henriksen et al., 2007; Lazarev et al., 2013; Davis, 2015; Gomollon et al., 2017).

Accordingly, the European Crohn's and Colitis Organization consensus guideline recommends that UGI endoscopy and radiology, such as magnetic resonance imaging, computed tomography, and small bowel capsule endoscopy, should be performed in all CD patients where UGI tract involvement is suspected (Annese et al., 2013). However, UGI tract involvement is a diagnostically challenging presentation in CD, due to a lack of specific clinical symptoms, and thus, there is a heavier reliance on imaging modalities in practice.

Chronic inflammation in CD patients is related to altered interactions between the host and the microbiota, and microbial imbalance (Frank et al., 2007; Xavier and Podolsky, 2007; Sartor, 2008; Halfvarson et al., 2017). Currently, the human microbiome, comprising of the entire microbial complement related with human hosts, is a critical and emerging area for biomarker discovery (Pascal et al., 2017; Douglas et al., 2018; Mills et al., 2019). The identification of microbial biomarkers and their use for the prediction of the disease provide valuable information for predictions in a wide range of applications.

Hence, the aims of this study were to compare the metagenomic profile in CD patients with and without UGI involvement at diagnosis, and to identify the metagenomic biomarkers predicting its development.

## MATERIALS AND METHODS

### Data Sources and Processing

We used the data from the Inflammatory Bowel Disease Multi'omics Database<sup>1</sup> for the most comprehensive description to date of host and microbial activities in inflammatory bowel diseases. Tissue samples gathered during the initial screening colonoscopy at diagnosis were collected according to a standardized protocol, and the V4 region of the 16S rRNA gene was PCR-amplified and sequenced in the MiSeq platform (Illumina) (for detailed protocols see <http://ibdmdb.org/protocols>). We divided the subjects into two groups, “nonL4” versus “L4” -where nonL4 are CD patients without UGI tract involvement and L4 are those with UGI tract involvement in disease extent.

<sup>1</sup><http://ibdmdb.org>

### Community Analysis

The obtained raw data were analyzed using Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.0, a software that performs microbial community analysis and taxonomic classification of microbial genomes (Navas-Molina et al., 2013). Sequences were assigned to operational taxonomic units (OTUs) with a 97% similarity threshold and subsequently picked by UCLUST against a closed reference table, the latest version of the Greengenes OTU database (Edgar, 2010). For diversity analysis, samples were normalized so all the samples could be compared. Alpha diversity of OTU libraries was described using the Chao1, phylogenetic diversity (PD) whole tree, and observed species, and were compared using a Student's *t*-test. Distance matrices were constructed using the unweighted and weighted UniFrac algorithms in QIIME from the whole community phylogenetic tree. Significant differences between the predefined groups were analyzed using one-way analysis of similarities (ANOSIM) with 999 permutations with their corresponding Global-R statistics.

### Biomarker Detection and Functional Analysis

To determine the potential biomarker OTUs, linear discriminant analysis effect size (LEfSe) analysis was performed with a linear discriminant analysis (LDA) score threshold of > 1.0 to detect features significantly different in abundance between the groups (Fisher, 1936; Segata et al., 2011).

In addition, we conducted subsequence-based 16S rRNA data processing using the DiTaxa software, which substitutes standard OTU-clustering method by segmenting 16S rRNA reads into the most frequent variable-length subsequences, for sequence phenotype classification and biomarker detection (Asgari et al., 2018). The linear support vector machine (SVM) and random forest classifier (RF) algorithms were used to build a predictive model and to calculate the importance of all variables and rank them accordingly. For linear SVM, we set the cost to the value of 1 and use RF classifier in the default settings.

PICRUSt2 was used to predict microbial content from each sample's data and functionally annotate the data (Langille et al., 2013). The results were further subjected to statistical analysis of taxonomic and functional profiles (STAMP v2.1.1) software (Parks et al., 2014). To investigate the metabolic network of the predicted organism, we used MetaCyc database<sup>2</sup>, which contains data regarding chemical compounds, reactions, enzymes, and metabolic pathways that have been experimentally validated and reported in the scientific literature (Caspi et al., 2016). The statistical analyses were performed using R version 3.5.1 (R Core Team, 2017; Venables and Smith, 2020). All significant thresholds were set at a two-sided *p*-value of 0.05.

## RESULTS

### Baseline Characteristics

Among the 37 potential CD patients, four patients with insufficient data on disease extent and seven patients who were

<sup>2</sup><https://metacyc.org>

not receiving tissue samples at the time of diagnosis were excluded, leaving 26 patients for analysis. Patients of the L4 group were diagnosed at a younger median age of 13.0 years (IQR 10.5–15.5 years) compared to 19.0 years (IQR 14.5–28.0 years) for the patients in nonL4 group ( $P = 0.005$ ), and the male to female ratio was 2.3 and 1.2 in the L4 and nonL4 groups, respectively ( $P = 0.650$ ) (Table 1). The baseline C-reactive protein (CRP) score and simple endoscopic score for Crohn's disease (SES-CD) did not differ significantly between the groups ( $P = 0.711$  and  $P = 0.056$  for L4 and nonL4, respectively) (Table 1). However, the CD patients with UGI tract involvement had higher erythrocyte sedimentation rate (ESR) than those without UGI tract involvement ( $P = 0.033$ ) (Table 1). All tissue samples were obtained from the rectum and ileum (Table 1). None of the patients were on any active medication, such as corticosteroids, immunomodulators, or biological agents at the time of sample collection. The detailed demographic and clinical characteristics are summarized in Table 1.

## Taxonomic Characterization

We analyzed the intestinal microbiota diversity of the two groups and tested whether intestinal microbiota diversity could be related to disease extent. The alpha diversity indices of Chao1, PD whole tree, and observed species diversity are shown in Supplementary Figure S1. All three diversity indices were higher in nonL4 compared to L4, but there were no significant differences between the two groups ( $P = 0.522$ ,  $P = 0.503$ , and  $P = 0.275$  for Chao1, PD whole tree, and observed species diversity, respectively; Supplementary Figure S1). Beta diversity was further evaluated using weighted-UniFrac analysis, which showed similar bacterial communities in patients of both groups (Supplementary Figure S1). Furthermore, an unweighted UniFrac-based principal coordinate analysis (PCoA) showed that samples were clustered by subject (ANOSIM:  $R = -0.010$ ;

$P = 0.477$ ) (Supplementary Figure S2). We also performed a weighted-UniFrac PCoA analysis with ANOSIM ( $R = -0.043$ ;  $P = 0.898$ ) (Supplementary Figure S2).

## Bacterial Abundance and Distribution

Subsequently, we analyzed the intestinal microbiota abundance and distribution in the two groups and tested whether they could be related to UGI tract involvement in CD. At the genus level, bacteria from *Akkermansia* (0.3% vs. 1.8%), *Haemophilus* (0.2% vs. 1.7%), *Oscillospira* (1.0% vs. 1.3%), *Parabacteroides* (0.8% vs. 0.9%), *Clostridium* (0.1% vs. 0.9%), *Dialister* (0.7% vs. 0.8%), *Lachnospira* (0.1% vs. 0.7%), *Streptococcus* (0.3% vs. 0.5%), *Coprococcus* (0.4% vs. 0.5%), and *Ruminococcus* [f\_Ruminococcaceae] (0.3% vs. 0.4%) were less abundant, whereas those from *Bacteroides* (34.9% vs. 33.0%), *Faecalibacterium* (13.7% vs. 11.1%), *Ruminococcus* [f\_Lachnospiraceae] (7.4% vs. 5.4%), *Prevotella* (3.5% vs. 0.3%), *Fusobacterium* (3.1% vs. 2.6%), *Sutterella* (2.4% vs. 2.2%), *Blautia* (1.3% vs. 0.8%), *Veillonella* (1.2% vs. 1.1%), *Dorea* (0.7% vs. 0.5%), *Bilophila* (0.6% vs. 0.3%), *Phascolarctobacterium* (0.3% vs. 0.1%), and *Odoribacter* (0.2% vs. 0.1%) were more abundant in L4 compared to nonL4 (Figure 1).

## Metagenomic Biomarker Discovery

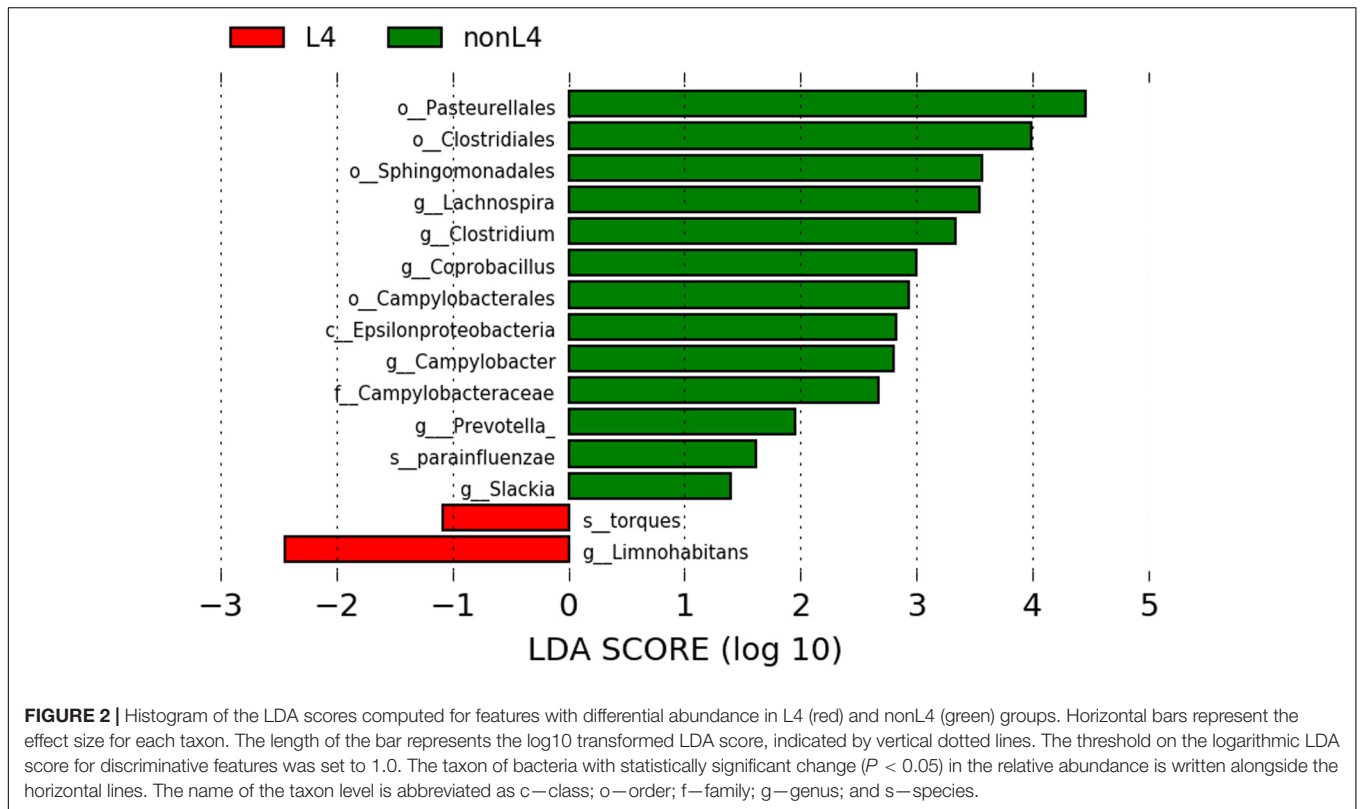
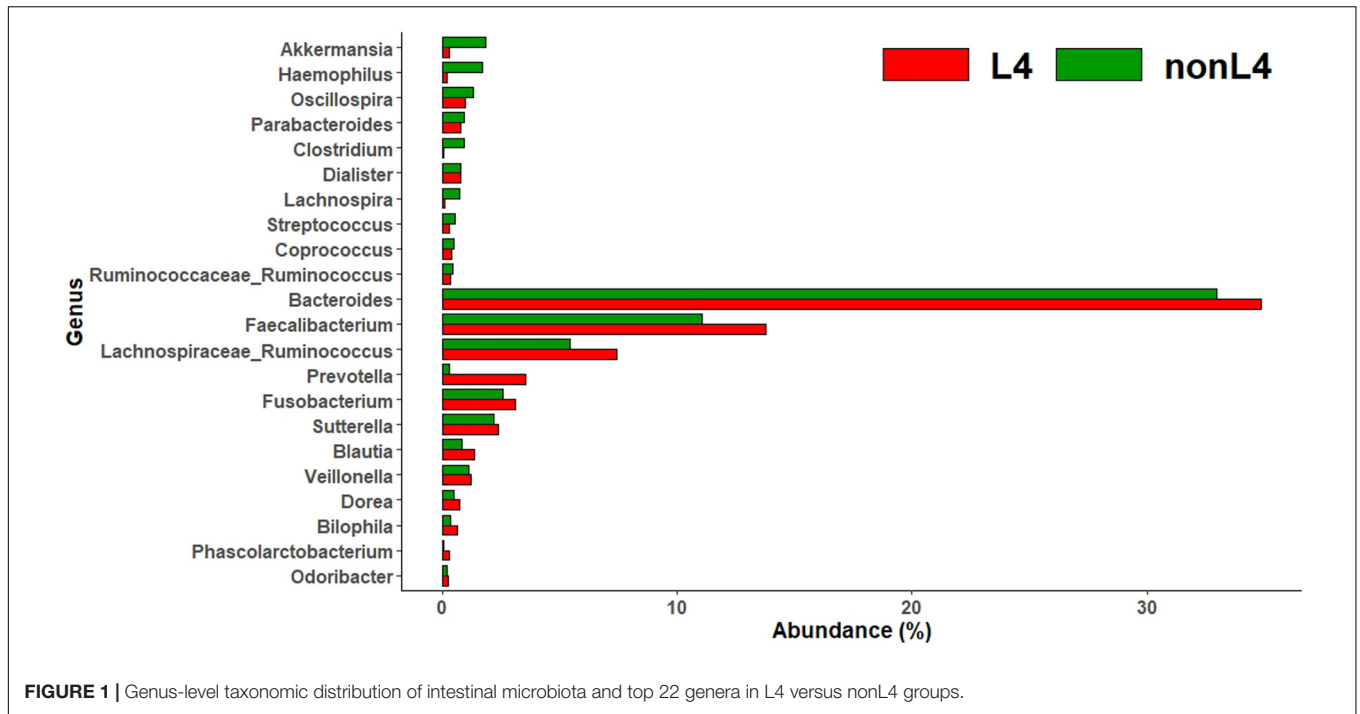
We found significant differences in the community compositions between the two groups by LEfSe analysis. As shown in Figure 2, the microbial composition was also significantly different at the order level among groups. The Pasteurellales ( $P = 0.042$ ), Sphingomonadales ( $P = 0.045$ ), Campylobacterales ( $P = 0.024$ ), and Clostridiales ( $P = 0.043$ ) exhibited a relatively higher abundance in nonL4 group (Figure 2). The patients in nonL4 group had members of the class Epsilonproteobacteria ( $P = 0.024$ ) and the family *Campylobacteraceae* ( $P = 0.024$ ) that were significantly dominant than those in L4 group patients (Figure 2). Furthermore, there were seven significantly different genera, composed of *Campylobacter* ( $P = 0.024$ ), *Prevotella* ( $P = 0.034$ ), *Clostridium* ( $P = 0.043$ ), *Coprobacillus* ( $P = 0.015$ ), *Slackia* ( $P = 0.034$ ), and *Lachnospira* ( $P = 0.015$ ) that were enriched in the nonL4 group, while *Limnohabitans* ( $P = 0.034$ ) was enriched in the L4 group (Figure 2). At the species level, significantly more *Haemophilus parainfluenzae* were detected in the patients in nonL4 group ( $P = 0.028$ ), while *Ruminococcus torques* were enriched in L4 group patients ( $P = 0.015$ ) (Figure 2).

Comparative taxonomic visualization of detected differentially expressed markers for DiTaxa and a common workflow are shown in Supplementary Figure S3 for samples from CD patients with UGI tract involvement versus those without UGI tract involvement. Taxa predicted by DiTaxa analysis for samples from L4 group patients versus those from nonL4 group exhibited *Ruminococcus faecis*, *Coprococcus comes*, *Dorea formicigenerans*, *Ruminococcus torques*, CCMM\_s, *Eubacterium hallii*, *Bilophila wadsworthia*, *Blautia faecis*, *Ruminococcus gnavus*, *Alistipes putredinis*, *Bacteroides fingoldii*, *Roseburia faecis*, *Faecalibacterium prausnitzii*, *Roseburia inulinivorans*, *Lachnospira pectinoschiza*, *Intestinibacter bartlettii*, *Clostridium symbiosum*, *Agathobacter rectalis*, *Roseburia*

TABLE 1 | Baseline characteristics of the patients.

Variable	L4 (n = 13)	nonL4 (n = 13)	P-value
Age at diagnosis, year (median, IQR)	13.0 (5.0)	19.0 (13.5)	0.005
Race (n, %)			1.000
White	12 (92.3)	12 (92.3)	
American, Indian, or Alaska Native	1 (7.7)	0 (0.0)	
Other	0 (0.0)	1 (7.7)	
Sex (n, %)			0.650
M	9 (69.2)	7 (53.8)	
F	4 (30.8)	6 (46.2)	
Biopsy location (n, %)			0.216
Ileum	6 (46.2)	3 (23.1)	
Rectum	7 (53.8)	10 (76.9)	
ESR (mm/h)	40.0 (27.0)	11.0 (28.0)	0.033
CRP (mg/L)	4.8 (3.3)	4.9 (12.6)	0.711
SES-CD score	8 (9.0)	2.5 (6.0)	0.056

IQR, interquartile range; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein; SES-CD, simple endoscopic score for Crohn's disease. The p-Values were obtained using Mann-Whitney test and chi-square test.



*intestinalis*, *Clostridium bolteae*, *Fusicatenibacter saccharivorans*, *Anaerostipes hadrus*, *Bacteroides caccae*, *Bacteroides uniformis*, *Flavonifractor plautii* as significantly associated (Table 2).

Of these, the following seven taxa were actually identified with UCLUST-based methods in QIIME (Table 2): *Dorea formicigenerans*, *Ruminococcus torques*, *Ruminococcus gnavus*,

**TABLE 2 |** Taxa predicted by DiTaxa analysis for UGI tract involvement in CD.

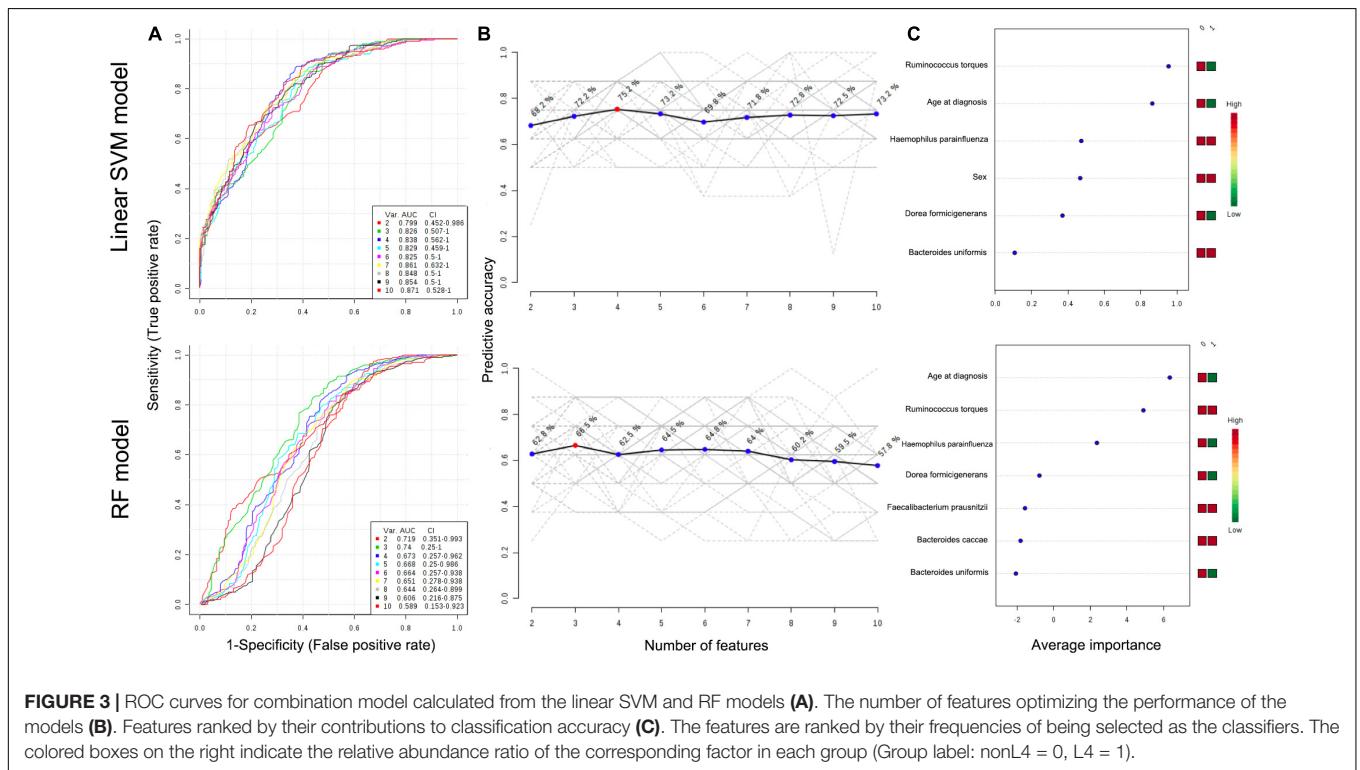
Direction	Taxonomy	Marker	P-value	Number of markers
+	<i>Ruminococcus faecis</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggagtggaagctggtgtgaaaacccggggctcaacccgggactgcattggaactgtcaatctagatagaccggagaggaagcggaaattcctagtgtagcggtgaaatcgtagatattaggaggaaacaccagtgccgaagcgcttactggacggttaactgacgttgaggctcgaagcgtggggagcaaacagg	0.014	2
+	<i>Coprococcus comes</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggctgtgtaagctgaaagcggggctcaaccccgggactgcttggaaactatgcagctagagtgctggagaggaagtggaattccagtgtagcggtgaaatcgtagatattgggaggaacaccagtgccgaagcgcttactggacggttaactgacgttgaggctcgaagcgtggggagcaaacagg	0.025	1
+	<i>Dorea formicigenerans*</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggctgtgcaagctgaaagcggatgggctcaacctgtggactgcttggaaactgtcagctagagtgctggagaggaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgatgactgacgttgaggctcgaagcgtggggagcaaacagg	0.025	1
+	<i>Ruminococcus torques*</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggagtggaagctgtagtgaacccggggctcaaccccgggactgcattggaactgttactctagagtgctggagaggaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgataactgacgttgaggctcgaagcgtggggagcaaacagg	0.025	1
+	CCMM s	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggcactaaggctgtggtgaaagatcgaagcttaactcggtaagccatggaaac	0.025	1
+	<i>Eubacterium hallii</i>	tgctcgctagatagacagggagcggcgaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcggcctgtgactgttactgacgctgagcgacgaaagcgtggggagcaaacagg	0.046	1
+	<i>Bilophila wadsworthia</i>	ccctagatattcggaggaacaccggtggcgaagcgccactggacggttaactgacgctgaggtgcgaagcgtgggtagcaaacagg	0.046	1
+	<i>Blautia faecis</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggcgcagcaagctgtgtagaaagcaggggcttaacccctggactgcattggaactcgtgtac	0.046	1
+	<i>Ruminococcus gnavus*</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgtagacggcctgcaagcagatgtgaaagccggggctcaaccccgggactgcattggaactgtcagctagagtgctggagaggaagcggaaattcctggtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttctggacgatgactgacgttgaggctcgaagcgtggggagcaaacagg	0.046	1
+	<i>Alistipes putredinis</i>	tacggaggtcaagcggttatccggattactgggtgtaaaaggagcgttagcgggttgataagtagagtgaaatcggggctcaacccctgaacgtgctctaactgttgactagagtagtgctggtaggcggaatgtatggtgtagcggtgaaatcgtagatattacagaacaccgattgcgaagcagcttaccaaactatacctgacgttgaggcacgaaagcgtggggagcaaacagg	0.046	1
+	<i>Bacteroides fingoldii</i>	tacggaggtccgagcgttatccggattactgggtgtaaaaggagcgttagtggtgtaagtcagttgaaagtttgcggctcaacccgtaaaattgcagttgatactggctgtctgtagtacagtagaggtggcgg	0.046	1
-	<i>Roseburia faecis*</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggctgaggcggcaagctctgatgtgaaagccggggctcaaccccggtactgcattggaactgtcgtactagagtgctggaggggaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgataactgacgctgaggctcgaagcgtggggagcaaacagg	0.005	1
-	<i>Faecalibacterium prausnitzii*</i>	aaggcaagttggaagtgaatccatggctcaaccctgaactccttcaaaactgttttctgtagtagcagaggtaggcggaaattcccggtgtagcggtgaaatcgtagatattcggaggaacaccagtgccgaagcgccactactggcaccactgacgctgaggctcgaagtggtgtaaacagg	0.014	1
-	<i>Roseburia inulinivorans</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggcgaagcctaaagctctgatgtgaaagccggggctcaaccccggtac	0.025	1
-	<i>Lachnospira pectinoschiza</i>	agaggcaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactgctggactgtaactgacactgaggctcgaagcgtggggagcaaacagg	0.025	1
-	<i>Intestinibacter bartlettii</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggctttaaagtcaggagtgaaagcctacggctcaacccgtagactcctgaaactggaggactgagtgacagagagagagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggactgtaactgacgctgaggcagcaagcgtggggagcaaacagg	0.035	1
-	<i>Roseburia inulinivorans</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggcgaagcctaaagctctgatgtgaaagccggggctcaaccccggtactgcattggaactggtcatctagatagtgctggaggggaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgataactgacgctgaggctcgaagcgtggggagcaaacagg	0.046	5
-	<i>Clostridium symbiosum</i>	tgtttaactggagtgctggagaggaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagggcactactggacgataactgacgttgaggctcgaagcgtggggagcaaacagg	0.046	1
-	<i>Agathobacter rectalis</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggctgaggcggcaagctctgatgtgaaagccggggctcaaccccggtactgcattggaactgtcgtactagagtgctggaggggaagcggaaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgataactgacactgaggctcgaagcgtggggagcaaacagg	0.046	1
-	<i>Roseburia intestinalis</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggctgaggcgaagctctgatgtgaaagccggggctcaaccccggtactgcattggaactgtcgtactgagtgctggaggggaagcggaaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgataactgacgttgaggctcgaagcgtggggagcaaacagg	0.046	1
-	<i>Clostridium bolteae</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggcgaagcctgaaagtgaaagccagggctcaacccctgggactgcttggaaactgtttctagagtgctggagaggaagtggaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccgaagcgcttactggacgataactgacgttgaggctcgaagcgtggggagcaaacagg	0.046	1
-	<i>Fusicatenibacter saccharivorans</i>	tacgtatggtgcaagcggttatccggattactgggtgtaaaaggagcgcagcggcgaagcctgctgatgtgaaacccagggcttaacccctgggactgcattggaactgtctgctcgtgctgagtgccggagaggaagcggaaattcctagtgtagcggtgaaatcgtagatattaggaggaacaccagtgccga	0.046	1

(Continued)

TABLE 2 | Continued

Direction	Taxonomy	Marker	P-value	Number of markers
-	<i>Anaerostipes hadrus</i>	tacgtagggggcaagcgttatccggaattactgggtgtaaaggggtcgtaggtggtatggcaagtcagaagtgaaccaccaggg cttaactctgggactgctttgaaactgcagactggagtgccaggagaggaagcgggaattcctagtgtagcgggtgaaatcgctagat attaggagg	0.046	1
-	<i>Bacteroides caccae</i> *	tacggcggtccgagcgttatccggattattgggttaaaggagcgtagcggattgtaagtcagttgtgaaagtttgcggctcaac cgtaaattgcagttgatactggcagcttgagtgtagtagaggtggcggaattcgtggtgtagcgggtgaaatgcttagatcacg aagaactccgattgcaagggcagctcactggagtgtaactgacgctgatgctcgaagtggtgatcaaacagg	0.046	1
-	<i>Bacteroides caccae</i>	tacggaggtccgagcgttatccggattattgggttaaaggagcgtagcggattgtaagtcagttgtgaaagtttgcggctcaa ccgtaaaattgcagttgatactggcagcttgagtgtagtagaggtggcggaattcgtggtgtagcgggtgaaatgcttagatcacga agaactccgattgcaagggcagctcactggagtgtaactgacgctgatgctcgaagtggtgatcaaacagg	0.046	1
-	<i>Bacteroides uniformis</i> *	tacggaggtccgagcgttatccggattattgggttaaaggagcgtagcggattgtaagtcagttgtgaaagtttgcggctcaa ccgtaaaattgcagttgatactgggtgcttgagtagtagaggtggcggaattcgtggtgtagcgggtgaaatgcttagat cacgaagaactccgattgcaagggcagctcactggagtgtaactgacgctgatgctcgaagtggtgatcaaaaagg	0.046	1
-	<i>Flavonifractor plautii</i>	taaaggcgtgtagcgggattgcaagtcagatgtaaaactggggctcaacctccagcctgattgaaactgtagtcc	0.046	1

UGI, upper gastrointestinal; CD, Crohn's disease; +, prediction for L4 group; -, prediction for nonL4 group; \*taxa identified with UCLUST-based methods in QIIME.



*Roseburia faecis*, *Faecalibacterium prausnitzii*, *Bacteroides caccae*, and *Bacteroides uniformis*.

### Metagenomic Biomarker Evaluation

To further characterize the predictive value of the eight identified taxa by LEfSe or DiTaxa methods, we performed ROC analysis with clinical variables (age at diagnosis and sex) using the machine learning models (Figure 3). A comparison of the average performance as a predictive model suggests the superiority of SVM: the average performance of SVM is >0.799 AUC and 68.2–75.2% accuracy, while that of RF is <0.740 AUC and 57.8–66.5% accuracy

(Figure 3). For the top performing model architecture, the addition of microbial features improves the predictive performance of linear SVM model; however, the performance in RF model tends to decrease. Notably, for the feature ranking method in both linear SVM and RF models, the top two factors -the species *Ruminococcus torques* and age at diagnosis- contributed to the combined models (Figure 3). Figure 3 also shows that the addition of the signature of the species *Haemophilus parainfluenzae* into the models enabled us to achieve the highest accuracy and to increase the diagnostic performance of UGI tract involvement in CD patients.

## Metagenomic Functional Analysis

In addition, the functional diversity of the different putative metagenomes was assessed using the PICRUSt2 software. Pathways displaying a significant difference in mean proportions between L4 and nonL4 groups were represented (Figure 4). The pathways, including thiazole biosynthesis II ( $p < 0.001$ ), superpathway of thiamine diphosphate biosynthesis II ( $P = 0.010$ ), and octane oxidation ( $P = 0.035$ ), were over-represented, whereas L-methionine biosynthesis III ( $P = 0.038$ ) and palmitate biosynthesis II ( $P = 0.050$ ) were under-represented in L4 (Figure 4). For selected pathways, we also examined the extent to which these pathways are linked with the species *Ruminococcus torques*. As shown in Figure 5, the two related MetaCyc pathways, L-methionine biosynthesis III and palmitate biosynthesis II, showed an association with *Ruminococcus torques*, which may play important roles in the intestinal integrity and barrier function (Figure 5).

## DISCUSSION

To our knowledge, this is the first study to identify a reliable metagenomic biomarker for UGI tract involvement in CD. The reported frequency of UGI tract involvement in CD largely varies. The main cause of the discrepancies regarding prevalence rates of UGI tract lesions is probably related to irregularly performing different diagnostic modalities for CD diagnosis, presumably because of the low reliability of mapping disease extent in clinical practice. However, to date, no data has been analyzed regarding a simple biomarker for CD patients with UGI tract involvement.

Our main hypothesis is that the possible differences in taxonomic composition might potentially be used as proxy biomarkers for UGI tract involvement in CD patients, since altered microbial communities have been demonstrated to be an essential factor in driving intestinal inflammation in CD (Tamboli et al., 2004; Sartor, 2008).

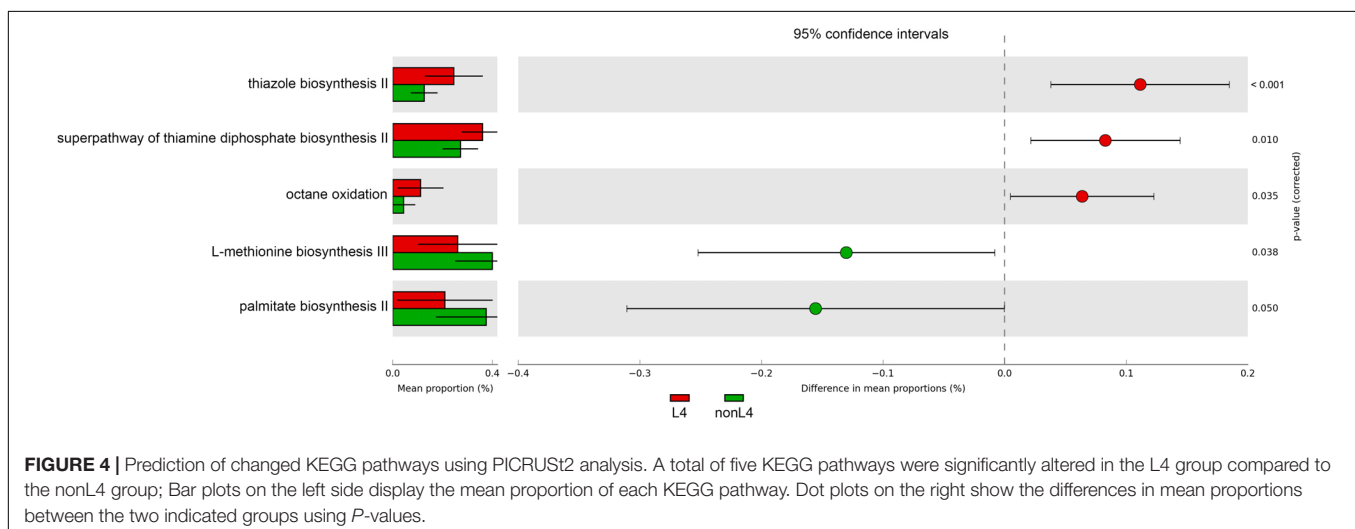
We analyzed the differences in the tissue microbial community of CD patients at the species level. In this study,

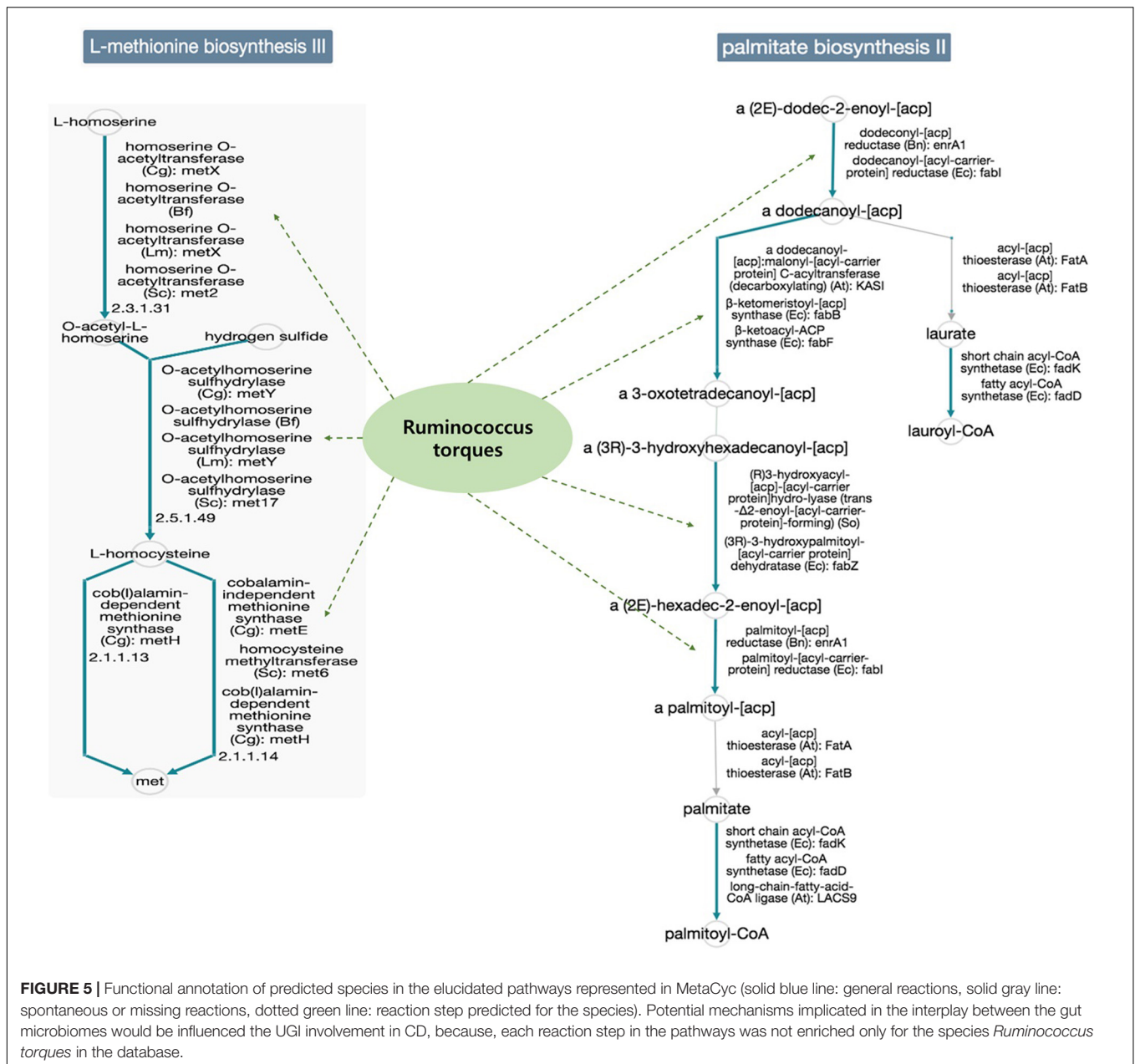
the species *Dorea formicigenerans*, *Ruminococcus torques*, *Ruminococcus gnavus*, *Roseburia faecis*, *Faecalibacterium prausnitzii*, *Bacteroides caccae*, *Bacteroides uniformis*, and *Haemophilus parainfluenza* were identified as predictive biomarkers by the LEfSe or DiTaxa programs. Interestingly, *Ruminococcus torques*, a butyrate-producing bacterial species, was the only one commonly identified by the two different algorithms.

The authors also examined whether the composition of the microbiota, with clinical predictors, could predict whether the patient would have UGI tract involvement or not using two different machine learning algorithms (linear SVM and RF). Modest predictive performances were achieved with a few features (eight taxa, age at diagnosis, and sex), especially in linear SVM. Notably, the most influential features for predicting disease extent were levels of the species *Ruminococcus torques* (positive correlation) and age at diagnosis (negative correlation). These consistent reproducible results for the species *Ruminococcus torques* present the possibility of using microbiota analysis as a screening tool to determine CD patients at high risk of UGI tract involvement.

Contrary to the inconsistent results regarding the signature of *Ruminococcus torques* in fecal samples of CD patients (Joossens et al., 2011; Gevers et al., 2014; Takahashi et al., 2016), most of the studies from tissue samples showed consistently high levels of the species in the mucosa of patients with CD compared to that in healthy subjects (Martinez-Medina et al., 2006; Png et al., 2010). Furthermore, our results show that its abundance remains significantly high in the CD patients with UGI tract involvement.

No role in CD pathophysiology has been suggested so far for the species *Ruminococcus torques*, belonging to the *Clostridium coccoides* group/cluster XIVa. They utilize MUC2, the main secreted mucin in the human intestine, as the sole carbon source and have a strong gastrointestinal mucin-degrading ability, providing further evidence of their adaptability in the human gut mucosal environment (Colina et al., 1996; Dethlefsen et al., 2006; Png et al., 2010). Therefore, it has been proposed that excessive mucin degradation by these bacteria may contribute to intestinal





disorders, as access of luminal antigens to the intestinal immune system is facilitated (Ganesh et al., 2013).

In addition, we found an inverse relationship for age at diagnosis with UGI involvement in CD. Present study showed that the median age of the patients with UGI involvement was significantly lower compared to those without UGI involvement. This is in accordance with a previous study by Thomas and colleagues who demonstrated a higher rate of younger patients ( $\leq 16$  years) suffering from UGI tract involvement compared to those without (9.4% versus 17.8%,  $P = 0.005$ ) (Greuter et al., 2018). Another study by Lopez-Siles et al. observed that CD patients below 16 years of age had a striking reduction in the population of *Akkermansia* sp. that possess a lower mucolytic

activity compared to those with disease onset at a later age, which is in agreement with our findings (Lopez-Siles et al., 2018).

Taken together, the mucus barrier dysfunction, due to the replacement of a less mucolytic bacteria, such as *Akkermansia* species by a more mucolytic one, such as *Ruminococcus torques*, at young age may influence the microbial community on the intestinal mucosa and be instrumental in the development of UGI tract involvement in CD.

To characterize the functional role of the microbiome in phenotype, we annotated the taxa by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. This analysis suggested that L-methionine biosynthesis III and palmitate biosynthesis II pathway were decreased, which are linked with



*Ruminococcus torques*, while the three KEGG pathways predicted to be increased in L4 group were not associated with the species. Currently, the role of methionine metabolism and its metabolites, and palmitate metabolic pathway in the pathogenesis of CD is poorly understood. Methionine is known to improve the integrity and barrier function of the small intestinal mucosa and villus morphology, and development in previous studies from animal models (Chen et al., 2014; Shen et al., 2014). A previous *in vivo* study by Wei et al. showed that palmitate plays a key role in the preservation of the gut barrier function by regulating the secretion and function of MUC2 (Wei et al., 2012). Another recent study also demonstrated that palmitate enhances MUC2 production in goblet cells of intestine, leading to the establishment of a thick mucus gel, thereby maintaining the integrity of the gut barrier (Benoit et al., 2015). These findings imply that there might be a decrease in the two protective pathways through the communication between intestinal cells and microbial community, especially the species *Ruminococcus torques* may induce excessive mucus degradation of small intestine in CD patients with UGI tract involvement.

The main strength of this study is that it evaluated the potential metagenomic biomarkers for prediction of UGI tract involvement in CD patients through various analyses. Further, we analyzed the CD patients with new-onset disease, before the commencement of treatment. Changes in microbiota community structure important for disease pathogenesis are likely to be more evident in new-onset and treatment-naive patients than those undergoing treatment. Lastly, the study focuses on the mucosa-associated microbiota samples, which may be more relevant to disease pathogenesis and diagnosis than fecal samples. However, our study was limited by the small sample size. Another limitation was that it comprised of predominately white patients, and thus, the findings may not be generalized to other racial populations. Finally, although our study was focused only on the microbial community, microbial metabolites also have great potential for improving diagnosis of CD and reflect the abnormalities of the host intestine microbiota. Therefore, new biomarkers for CD patients with UGI involvement could be developed by integrated analysis of metabolomics and metagenomics from a multinational and multicenter cohort.

## CONCLUSION

In conclusion, the species *Ruminococcus torques* in the tissue microbial community of CD patients might serve as a novel potential biomarker for UGI tract involvement. The UGI tract involvement in CD is higher in younger age group patients; therefore, it should be carefully monitored in them. The mechanisms of interactions between the host and *Ruminococcus torques* should be further investigated.

## REFERENCES

Annese, V., Daperno, M., Rutter, M. D., Amiot, A., Bossuyt, P., East, J., et al. (2013). European evidence based consensus for endoscopy in inflammatory bowel disease. *J. Crohn Colit.* 7, 982–1018.

## DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/Supplementary Material.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

MK designed the study. JC and HS analyzed and interpreted the data, and wrote the manuscript. JJ and JY supervised the project and revised the manuscript. All authors vouch for the data and analysis, have approved the final version, and agreed to publish the manuscript.

## FUNDING

This research was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), which is funded by the Korean Ministry of Science, ICT and Future Planning (grant number NRF-2019R1C1C1003524).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01162/full#supplementary-material>

**FIGURE S1** | Analysis of alpha diversity as predicted by Chao 1 estimator, PD whole tree, and observed species (A); and beta diversity measured by weighted-UniFrac distances in L4 versus nonL4 groups (B).

**FIGURE S2** | Principal coordinates analysis (PCoA) based on (A) unweighted and (B) weighted UniFrac distance; blue for the nonL4 and red for the L4 (ANOSIM:  $R = -0.010$ ,  $P = 0.477$ ;  $R = -0.043$ ,  $P = 0.898$ ).

**FIGURE S3** | Heat map of occurrence of markers by DiTaxa analysis between L4 and nonL4 groups. The rows are sorted based on the taxonomic marker assignments and the columns represent each group and are sorted firstly, based on their phenotype, and secondly, based on their pattern similarity.

Annunziata, M. L., Caviglia, R., Papparella, L. G., and Cicala, M. (2012). Upper gastrointestinal involvement of Crohn's disease: a prospective study on the role of upper endoscopy in the diagnostic work-up. *Dig. Dis. Sci.* 57, 1618–1623. doi: 10.1007/s10620-012-2072-0

- Asgari, E., Münch, P. C., Lesker, T. R., Mchardy, A. C., and Mofrad, M. R. K. (2018). DiTaxa: nucleotide-pair encoding of 16S rRNA for host phenotype and biomarker detection. *Bioinformatics* 35, 2498–2500. doi: 10.1093/bioinformatics/bty954
- Benoit, B., Bruno, J., Kayal, F., Estienne, M., Debard, C., Ducroc, R., et al. (2015). Saturated and unsaturated fatty acids differently modulate colonic goblet cells *in vitro* and in rat pups. *J. Nutr.* 145, 1754–1762. doi: 10.3945/jn.115.211441
- Bernell, O., Lapidus, A., and Hellers, G. (2000). Risk factors for surgery and postoperative recurrence in Crohn's disease. *Ann. Surg.* 231, 38–45.
- Cameron, D. J. (1991). Upper and lower gastrointestinal endoscopy in children and adolescents with Crohn's disease: a prospective study. *J. Gastroenterol. Hepatol.* 6, 355–358. doi: 10.1111/j.1440-1746.1991.tb00870.x
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480.
- Castellaneta, S. P., Afzal, N. A., Greenberg, M., Deere, H., Davies, S., Murch, S. H., et al. (2004). Diagnostic role of upper gastrointestinal endoscopy in pediatric inflammatory bowel disease. *J. Pediatr. Gastroenterol. Nutr.* 39, 257–261.
- Chen, Y., Li, D., Dai, Z., Piao, X., Wu, Z., Wang, B., et al. (2014). L-methionine supplementation maintains the integrity and barrier function of the small-intestinal mucosa in post-weaning piglets. *Amino Acids* 46, 1131–1142. doi: 10.1007/s00726-014-1675-5
- Chow, D. K., Sung, J. J., Wu, J. C., Tsoi, K. K., Leong, R. W., and Chan, F. K. (2009). Upper gastrointestinal tract phenotype of Crohn's disease is associated with early surgery and further hospitalization. *Inflamm. Bowel Dis.* 15, 551–557. doi: 10.1002/ibd.20804
- Colina, A. R., Aumont, F., Deslauriers, N., Belhumeur, P., and De Repentigny, L. (1996). Evidence for degradation of gastrointestinal mucin by *Candida albicans* secretory aspartyl proteinase. *Infect. Immun.* 64, 4514–4519. doi: 10.1128/iai.64.11.4514-4519.1996
- Davis, K. G. (2015). Crohn's disease of the foregut. *Surg. Clin. North Am.* 95, 1183–1193.
- Dethlefsen, L., Eckburg, P. B., Bik, E. M., and Relman, D. A. (2006). Assembly of the human intestinal microbiota. *Trends Ecol. Evol.* 21, 517–523. doi: 10.1016/j.tree.2006.06.013
- Diaz, L., Hernandez-Oquet, R. E., Deshpande, A. R., and Moshiree, B. (2015). Upper gastrointestinal involvement in Crohn disease: histopathologic and endoscopic findings. *South Med. J.* 108, 695–700. doi: 10.14423/smj.0000000000000373
- Douglas, G. M., Hansen, R., Jones, C. M. A., Dunn, K. A., Comeau, A. M., Bielawski, J. P., et al. (2018). Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. *Microbiome* 6:13.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7, 179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x
- Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13780–13785. doi: 10.1073/pnas.0706625104
- Gajendran, M., Loganathan, P., Catinella, A. P., and Hashash, J. G. (2018). A comprehensive review and update on Crohn's disease. *Dis. Mon.* 64, 20–57.
- Ganesh, B. P., Klopffleisch, R., Loh, G., and Blaut, M. (2013). Commensal *Akkermansia muciniphila* exacerbates gut inflammation in *Salmonella typhimurium*-infected gnotobiotic mice. *PLoS One* 8:e74963. doi: 10.1371/journal.pone.0074963
- Gevers, D., Kugathasan, S., Denson, L. A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microb.* 15, 382–392.
- Gomollon, F., Dignass, A., Anness, V., Tilg, H., Van Assche, G., Lindsay, J. O., et al. (2017). 3rd European evidence-based consensus on the diagnosis and management of Crohn's Disease 2016: part 1: diagnosis and medical management. *J. Crohns Colit.* 11, 3–25.
- Greuter, T., Piller, A., Fournier, N., Safroneeva, E., Straumann, A., Biedermann, L., et al. (2018). Upper gastrointestinal tract involvement in Crohn's disease: frequency, risk factors, and disease course. *J. Crohns Colit.* 12, 1399–1409. doi: 10.1093/ecco-jcc/jjy121
- Halfvarson, J., Brislawn, C. J., Lamendella, R., Vazquez-Baeza, Y., Walters, W. A., Bramer, L. M., et al. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* 2:17004.
- Henriksen, M., Jahnsen, J., Lygren, I., Aadland, E., Schulz, T., Vatn, M. H., et al. (2007). Clinical course in Crohn's disease: results of a five-year population-based follow-up study (the IBSEN study). *Scand. J. Gastroenterol.* 42, 602–610. doi: 10.1080/00365520601076124
- Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60, 631–637. doi: 10.1136/gut.2010.223263
- Kefalas, C. H. (2003). Gastrointestinal Crohn's disease. *Proc. Bayl. Univ. Med. Cent.* 16, 147–151.
- Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676
- Lazarev, M., Huang, C., Bitton, A., Cho, J. H., Duerr, R. H., McGovern, D. P., et al. (2013). Relationship between proximal Crohn's disease location and disease behavior and surgery: a cross-sectional study of the IBD Genetics Consortium. *Am. J. Gastroenterol.* 108, 106–112. doi: 10.1038/ajg.2012.389
- Lenaerts, C., Roy, C. C., Vaillancourt, M., Weber, A. M., Morin, C. L., and Seidman, E. (1989). High incidence of upper gastrointestinal tract involvement in children with Crohn disease. *Pediatrics* 83, 777–781.
- Lopez-Siles, M., Enrich-Capo, N., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., Garcia-Gil, L. J., et al. (2018). Alterations in the abundance and co-occurrence of *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* in the colonic mucosa of inflammatory bowel disease subjects. *Front. Cell Infect. Microbiol.* 8:281. doi: 10.3389/fcimb.2018.00281
- Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D., and Garcia-Gil, L. J. (2006). Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. *Inflamm. Bowel Dis.* 12, 1136–1145. doi: 10.1097/01.mib.0000235828.09305.0c
- Mills, R. H., Vazquez-Baeza, Y., Zhu, Q., Jiang, L., Gaffney, J., Humphrey, G., et al. (2019). Evaluating metagenomic prediction of the metaproteome in a 4.5-year study of a patient with Crohn's disease. *mSystems* 4:e0337-18.
- Navas-Molina, J. A., Peralta-Sanchez, J. M., Gonzalez, A., McMurdie, P. J., Vazquez-Baeza, Y., Xu, Z., et al. (2013). Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol.* 531, 371–444. doi: 10.1016/b978-0-12-407863-5.00019-8
- Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494
- Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut* 66, 813–822.
- Png, C. W., Linden, S. K., Gilshenan, K. S., Zoetendal, E. G., McSweeney, C. S., Sly, L. I., et al. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment *in vitro* utilization of mucin by other bacteria. *Am. J. Gastroenterol.* 105, 2420–2428. doi: 10.1038/ajg.2010.281
- R Core Team (2017). *R: A Language And Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Sartor, R. B. (2008). Microbial influences in inflammatory bowel diseases. *Gastroenterology* 134, 577–594. doi: 10.1053/j.gastro.2007.11.059
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60.
- Shen, Y. B., Weaver, A. C., and Kim, S. W. (2014). Effect of feed grade L-methionine on growth performance and gut health in nursery pigs compared with conventional DL-methionine. *J. Anim. Sci.* 92, 5530–5539. doi: 10.2527/jas.2014-7830
- Strober, W., Fuss, I., and Mannon, P. (2007). The fundamental basis of inflammatory bowel disease. *J. Clin. Invest.* 117, 514–521. doi: 10.1172/jci30587
- Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., et al. (2016). Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's Disease. *Digestion* 93, 59–65. doi: 10.1159/000441768
- Tamboli, C. P., Neut, C., Desreumaux, P., and Colombel, J. F. (2004). Dysbiosis as a prerequisite for IBD. *Gut* 53:1057.

- Van Limbergen, J., Russell, R. K., Drummond, H. E., Aldhous, M. C., Round, N. K., Nimmo, E. R., et al. (2008). Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. *Gastroenterology* 135, 1114–1122. doi: 10.1053/j.gastro.2008.06.081
- Venables, W., and Smith, D. (2020). *An Introduction to R*. Available online at: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
- Wei, X., Yang, Z., Rey, F. E., Ridaura, V. K., Davidson, N. O., Gordon, J. I., et al. (2012). Fatty acid synthase modulates intestinal barrier function through palmitoylation of mucin 2. *Cell Host Microb.* 11, 140–152. doi: 10.1016/j.chom.2011.12.006
- Wolters, F. L., Russel, M. G., Sijbrandij, J., Ambergen, T., Odes, S., Riis, L., et al. (2006). Phenotype at diagnosis predicts recurrence rates in Crohn's disease. *Gut* 55, 1124–1130. doi: 10.1136/gut.2005.084061
- Xavier, R. J., and Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 448, 427–434. doi: 10.1038/nature06005

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kwak, Cha, Shin, Jeon and Yoon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.