

# Host Genetic and Gut Microbial Signatures in Familial Inflammatory Bowel Disease

Yoo Min Park, MD<sup>1</sup>, Eunji Ha, BSc<sup>2</sup>, Ki-Nam Gu, MSc<sup>2</sup>, Ga Young Shin, MD<sup>1</sup>, Chang Kyun Lee, MD, PhD<sup>1</sup>, Kwangwoo Kim, PhD<sup>2</sup> and Hyo Jong Kim, MD, PhD<sup>1</sup>

**INTRODUCTION:** The family history of inflammatory bowel disease (IBD) has been strongly associated with risk of developing IBD. This study aimed to identify the host genetic and gut microbial signatures in familial IBD.

**METHODS:** Genetic analyses using genome-wide single nucleotide polymorphism genotyping and whole exome sequencing were performed to calculate weighted genetic risk scores from known IBD-associated common variants and to identify rare deleterious protein-altering variants specific to patients with familial IBD in 8 Korean families that each included more than 2 affected first-degree relatives (FDRs) and their unaffected FDR(s). In parallel, gut microbial community was analyzed by 16S rRNA sequencing of stools from the sample individuals.

**RESULTS:** The risk of familial IBD was not well explained by the genetic burden from common IBD-risk variants, suggesting the presence of family-shared genetic and environmental disease-risk factors. We identified 17 genes (*AC113554.1*, *ACE*, *AKAP17A*, *AKAP9*, *ANK2*, *ASB16*, *ASIC3*, *DNP1*, *DUS3L*, *FAM200A*, *FZD10*, *LAMA5*, *NUTM2F*, *PKN1*, *PRR26*, *WDR66*, and *ZC3H4*) that each contained rare, potentially deleterious variants transmitted to the affected FDRs in multiple families. In addition, metagenomic analyses revealed significantly different diversity of gut microbiota and identified a number of differentially abundant taxa in affected FDRs, highlighting 22 novel familial disease-associated taxa with large abundance changes and the previously reported gut dysbiosis including low alpha diversity in IBD and 16 known IBD-specific taxa.

**DISCUSSION:** This study identified familial IBD-associated rare deleterious variants and gut microbial dysbiosis in familial IBD.

**SUPPLEMENTARY MATERIAL** accompanies this paper at <http://links.lww.com/CTG/A324>

*Clinical and Translational Gastroenterology* 2020;11:e00213. <https://doi.org/10.14309/ctg.0000000000000213>

## INTRODUCTION

Inflammatory bowel disease (IBD), including Crohn disease (CD) and ulcerative colitis (UC), is characterized by chronic and devastating inflammation primarily in the gastrointestinal tract. The prevalence of IBD is dramatically increasing in developing and developed countries, and IBD has evolved into a global disease in recent years (1). The etiology of IBD is generally explained by inappropriate activation of the intestinal mucosal immune system in response to commensal bacteria in genetically susceptible individuals exposed to environmental risk factors. However, many of the risk factors remain unknown (2).

Since the landmark study by Kirsner et al. (3), a family history of IBD has been established as the strongest risk factor for IBD (4) in both the West (5) and East (6). For example, first-degree relatives (FDRs) of affected probands in a Danish population had a ~15-fold

higher life-time risk of IBD compared with individuals in the general population (7). In Korea, where the incidence of IBD is still low, the FDRs of affected probands with IBD have a similar risk of future IBD, approximately 13.8-fold higher than the general population (6). These findings suggest that family-specific rare genetic variants with large effect sizes and family-specific environmental factors are involved in familial aggregation of IBD, providing a rationale to investigate the genetic and environmental signatures in familial IBD.

Genome-wide association studies (GWAS) using unrelated IBD cases and controls have identified more than 240 IBD-risk loci that are explained mostly by common risk variants with modest effect sizes (8,9) that collectively impair the immune system in the development of IBD (10). However, the genetic heritability estimated by the common GWAS variants in IBD-risk loci accounts for only a

<sup>1</sup>Center for Crohn's and Colitis, Department of Gastroenterology, Kyung Hee University College of Medicine, Seoul, Republic of Korea; <sup>2</sup>Department of Biology and Department of Life and Nanopharmaceutical Sciences, Kyung Hee University, Seoul, Republic of Korea. **Correspondence:** Hyo Jong Kim, MD, PhD. E-mail: [hjkim@khmc.or.kr](mailto:hjkim@khmc.or.kr). Kwangwoo Kim, PhD. E-mail: [kkim@khu.ac.kr](mailto:kkim@khu.ac.kr).

Received January 27, 2020; accepted June 18, 2020; published online July 21, 2020

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of The American College of Gastroenterology

small fraction of IBD risk (11). The large proportion of the missing heritability of complex diseases such as IBD have been suspected to be hidden behind large-effect rare variants whose disease association significance is generally weak because of a low statistical power in GWAS using unrelated subjects (12).

Gut dysbiosis has also been well documented to be critically associated with the pathogenesis of IBD (13). As the gut microbial community is affected by many factors including family-shared factors such as diet and a complex gene-microbiota interaction (14–17), FDRs tend to show a higher degree of similarity in gut microbial composition (18). Thus, the comparison of gut microbiota between familial patients with IBD and their unaffected FDRs with similar gut microbial communities provides a unique opportunity to pinpoint the IBD-risk gut microbial signature in less statistical noise.

Taken together, previous observations have illustrated the potential advantage of using familial IBD in identifying disease-risk signatures regarding host genetic variants and gut microbiome, although clinical and biological data of patients with familial IBD and their unaffected FDRs are scarce. In our study, we recruited 8 families with  $\geq 2$  IBD-affected FDRs and  $\geq 1$  unaffected FDR, obtaining their genomic DNAs in whole blood, gut bacterial DNA in stool samples, and clinical metadata. Genomic analysis of both common and rare variants and metagenomic analysis of gut microbiota revealed the landscape of disease-specific signatures in familial IBD.

## METHODS

### Study cohort

Patients with both a definite diagnosis of IBD (CD or UC) and a family history of IBD were eligible for this study. Unaffected FDRs (parents, full siblings, or children) of the patients with familial IBD were recruited to participate in this study. All patients were affiliated with a prospective single-center IBD cohort at Kyung Hee University Medical Center, and they met the diagnostic criteria of CD or UC based on clinical and endoscopic findings using histopathology and radiologic results (19,20). We excluded patients with IBD-unclassified or IBD-indeterminate type.

We identified 23 families with  $\geq 2$  FDRs of patients among the IBD cohort ( $n = 960$ ). After excluding 15 families because of refusal for consent by the family members, we finally recruited 8 families including 16 patients with IBD (including 10 CD and 6 UC patients) and 9 unaffected FDRs (Figure 1). The pedigree of the study cohort is illustrated in Figure 2. Twelve patients with IBD had concordant type (4 CD families and 2 UC families), and the remaining 4 patients had discordant type (2 CD/UC families). All study participants provided written informed consent for this work. The study was approved by the Institutional Review Board of the Kyung Hee University Medical Center.

All the study participants provided whole blood for genetic analysis, serum for serologic tests, and/or fresh stool for microbiome analysis. Within 12 weeks before stool sampling, all participants were strictly withheld from treatment with antibiotics or probiotics. We collected basic demographic features from all participants and the clinical information from the patients (Table 1). All the study patients were in remission (defined as the CD activity index  $< 150$  or the Mayo score  $< 3$ ) and did not use proton-pump inhibitors at the time of stool collection. Serum C-reactive proteins, perinuclear antineutrophil cytoplasmic antibodies, and anti-*Saccharomyces cerevisiae* antibodies were measured in patients with familial IBD. About 1 g of stool was collected using the Stool Nucleic Acid Collection and Preservation Tube (NORGEN Bio-Tek) and immediately stored at  $-80^{\circ}\text{C}$  until DNA extraction.

### Genome-wide variant analysis and calculation of weighted genetic risk scores

Genome-wide data of single nucleotide polymorphisms (SNPs) were generated from patients' genomic DNA using the Korea biobank array (Affymetrix Axiom KORV1.1) (21) to calculate individual-level genetic risk scores for IBD, CD, and/or UC. Raw genotype data (with 827,783 variants) was filtered by a general quality control (QC) procedure (see Table S1, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>), retaining 406,422 variants with a call rate  $\geq 95\%$ , minor allele frequency  $\geq 1\%$ , and a  $P$  value for Hardy-Weinberg equilibrium  $\geq 1 \times 10^{-7}$  in the 25 individuals. Whole genome imputation was then performed by the prephaser SHAPEIT2 (22,23) and the imputator Minimac3 (24) with the imputation reference panel of the 1,000 Genomes Project phase 3. The imputation of variants in non-pseudoautosomal region in chromosome X was conducted in the same pipeline for male and female samples separately. SNP-level imputation QC retained  $\sim 8$  million variants with INFO values  $\geq 0.4$ .

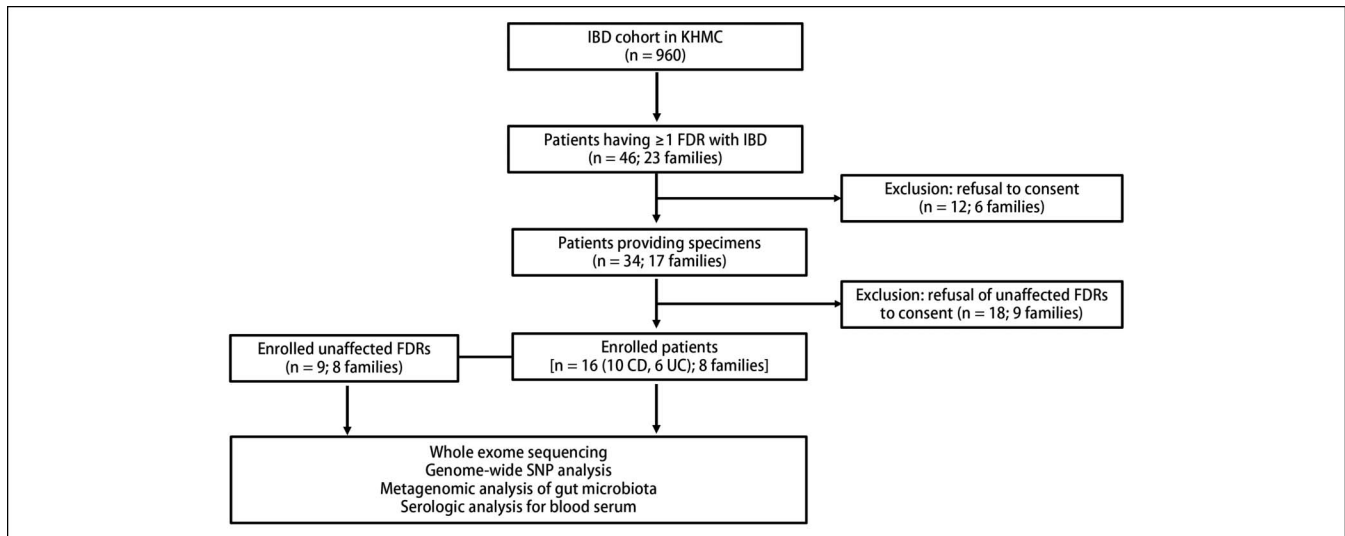
The weighted genetic risk scores (wGRSs) value of each individual was calculated as a linear summation of the number of disease-associated alleles weighted by the effect estimates reported in previous large GWAS of IBD, CD, or UC (see Table S2, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>) (8,9,25–35). The wGRS distribution in a general population was estimated using 100,000 random wGRS simulated based on the Korean allele frequencies.

### Whole exome sequencing and variant calling

The exome library was constructed from blood genomic DNA in the study subjects using the SureSelect V6 target enrichment system and was sequenced on an Illumina platform with 101-bp paired-end reads. Read alignment was performed using the Bowtie2 program with the reference human genome (hg38) (36). All procedures for QC and variant calling followed the GATK4 best practice pipeline (37) with a joint calling option (38). Common variants in dbSNP146 (minor allele frequency  $\geq 1\%$ ) were excluded from subsequent analysis to identify rare exonic variants that were shared between the affected individuals (39). Deleterious effect on protein function was predicted based on the SIFT and PolyPhen-2 algorithms using the variant effect predictor program (40,41).

### Sequencing of gut microbiome

The V3 and V4 amplicon of the 16S rRNA gene in stool DNA and sequencing library for the Illumina paired-end sequencing platform were prepared using Herculase II Fusion DNA Polymerase and Nextera XT library prep kit v2, respectively. Reads with poor base call quality or short insert sizes (excluding Nextera adapter sequence) were removed by using a Trimmomatic (42). Sequencing fragments merged from QC-passed paired-end reads of the V3 and V4 regions were further filtered according to the expected length (438–469bp) (43) of the target region. Operational taxonomic unit (OTU) clustering was performed using the QIIME2 program (44) with the open reference OTU picking method at 97% sequence similarity using the vsearch plug-in (45), excluding chimeric or rare OTUs. Taxonomic classification and phylogenetic annotation of OTUs were carried out using reference sequences (gg\_13\_8\_97.fasta) from Greengenes (46) with classify-sklearn and q2-phylogeny plugins in QIIME2 (47). The quality-filtered reads ( $n = 2,208,488$ ) were clustered into 1,895 OTUs at 97% sequence similarity.



**Figure 1.** Flow chart of the study. CD, Crohn disease; FDR, first-degree relative; IBD, inflammatory bowel disease; SNP, single nucleotide polymorphisms; UC, ulcerative colitis.

Alpha diversity (richness) of the gut microbial community in each individual was quantified by the Shannon index, and the association with disease status was tested using a Kruskal-Wallis test. Beta diversity (interindividual dissimilarity) of gut microbial composition in the study population was estimated using a principal coordinate analysis with the weighted UniFrac algorithm. Statistical significance for the difference in measured beta diversity among phenotypic groups was assessed by a permutational multivariate analysis of variance with adjustment for family and sex (48). Differentially abundant gut microbiota between case and control groups were identified using DESeq2 (49) in various taxonomic levels, adjusting for family and sex.

## RESULTS

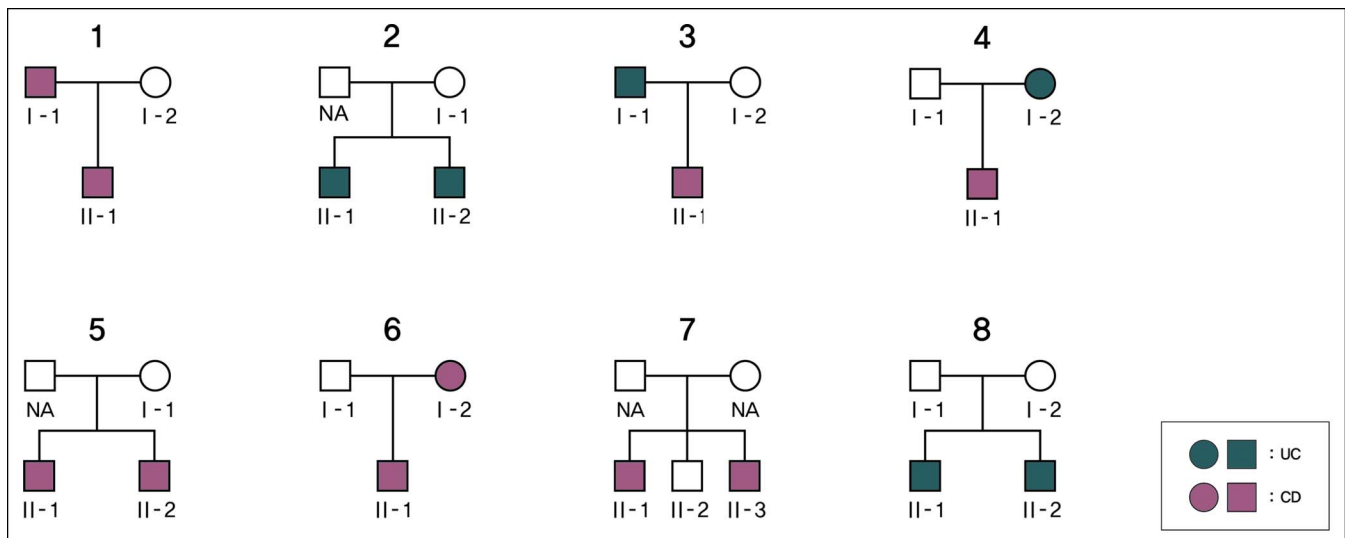
### Characteristics of the study cohort

The basic characteristics of the study cohort are listed in Table 1. There were no significant differences in demographic factors that could affect gut microbial composition, including the Bristol stool form scale (50), alcohol/cigarette consumption history, and breastfeeding in infancy.

The enrolled nuclear families with familial IBD tended to have more unaffected FDRs in the parental generation. At the time of stool and blood sampling, all patients with familial IBD were in clinical remission (Mayo score =  $1.0 \pm 1.0$  in UC; CD activity index =  $87.5 \pm 77.9$  in CD). The relatedness of family members was genetically confirmed using genome-wide SNP array data and whole exome sequencing data.

### Known common disease variants did not well explain risk of familial IBD

We investigated whether the known common variants associated with risk of IBD, CD, or UC were enriched in patients with familial IBD by calculating individual-level wGRS based on reported effect estimates (see Table S2, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>). Statistical significance of the increase in average wGRS in 2 affected members of each family was assessed by positioning the observed average into the distribution of the 100,000 wGRS of unaffected unrelated individuals simulated based on the known risk allele frequencies in the Korean population.



**Figure 2.** Pedigree of the study families. NA indicates samples and clinical data not available. CD, Crohn disease; UC, ulcerative colitis.

**Table 1. Characteristics of the study cohort**

Characteristic	Probands (n = 16)	Unaffected FDR (n = 9)	P Value <sup>a</sup>
<b>Demographics</b>			
Generation, n (%)			
First	4 (25.0)	8 (88.9)	
Second	12 (75.0)	1 (11.1)	
Age, yr, mean (SD)			
First generation	49.5 (6.2)	54.5 (12.0)	0.366
Second generation	31.7 (12.5)	48.0 (N/A)	
Male, n (%)	13 (81.2)	4 (44.4)	0.087
Bristol stool form scale, mean (SD)	4.5 (1.4)	3.7 (1.0)	0.210
Alcohol consumption, n (%)	6 (37.5)	4 (44.4)	1.000
Current smoker, n (%)	0 (0.0)	0 (0.0)	1.000
Cesarean section, n (%)			
Parents	0 (0.0)	0 (0.0)	1.000
Children	7 (58.3)	0 (0.0)	0.462
Breastfed, n (%)	14 (87.5)	9 (100.0)	0.520
<b>Disease characteristics</b>			
Type of IBD	UC	CD	
n (%)	6 (37.5)	10 (62.5)	
Duration of IBD, mo, mean (SD)	58.0 (51.7)	136.8 (81.2)	N/A
Disease activity of IBD <sup>b</sup> , mean (SD)	1.0 (1.0)	87.5 (77.9)	N/A
hs-CRP (mg/L), mean (SD)	16.2 (52.3)	0.4 (0.7)	
Fecal calprotectin (μg/g), mean (SD)	29.7 (35.4)	13.6 (9.3)	
Positivity for pANCA, n (%)	2 (10.5)	1 (6.6)	N/A
Positivity for ASCA, n (%)			
Anti-ASCA IgA	3 (15.7)	1 (6.6)	N/A
Anti-ASCA IgG	2 (10.5)	5 (33.3)	N/A
Bowel resection, n (%)	0 (0.0)	1 (10.0)	N/A
Perianal disease, n (%)	0 (0.0)	5 (50.0)	N/A
Current use of biologics, n (%)	2 (33.3)	6 (60.0)	N/A
Current use of immunomodulators, n (%)	2 (33.3)	3 (30.0)	N/A

ASCA, anti-Saccharomyces cerevisiae antibodies; CD, Crohn disease; FDR, first-degree relative; hs-CRP, high-sensitive serum C-reactive protein; IBD, inflammatory bowel disease; N/A, not available; pANCA, perinuclear anti-neutrophil cytoplasmic antibody; UC, ulcerative colitis.

<sup>a</sup>P values were calculated by Fisher exact tests for categorical variables and *t* tests for numeric variables.

<sup>b</sup>Mayo scores for UC or Crohn disease activity index for CD.

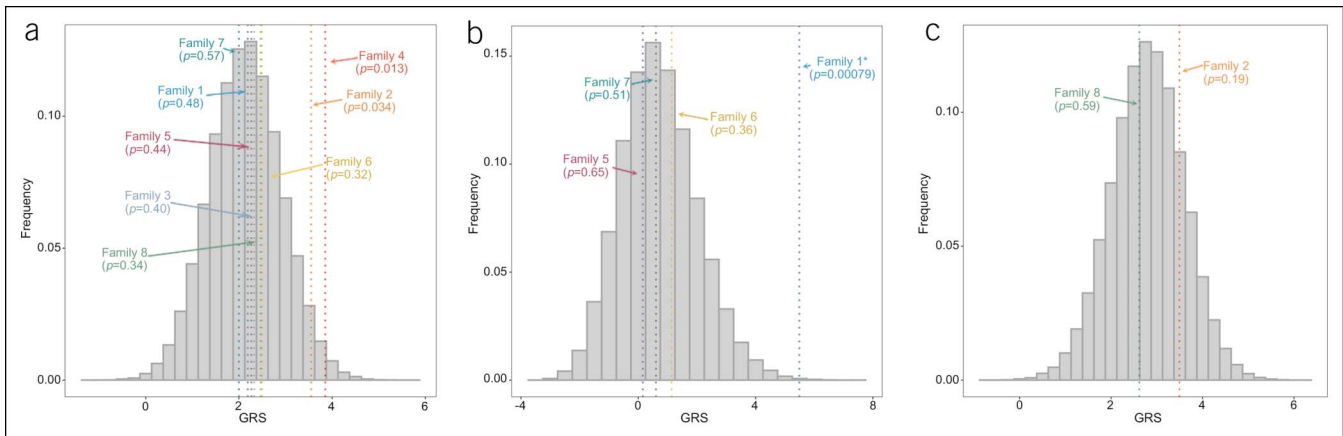
A single family with familial CD showed a significantly higher burden of common CD-risk variants than the expected distribution of unaffected individuals at a false-discovery rate of 5% ( $P$  value =  $7.9 \times 10^{-4}$ , Figure 3). There was, however, no significant increase in wGRS in the other families, raising the possibility that family-specific rare variants with large effects may drive risk of disease.

#### Whole exome sequencing identified rare exonic variants that were transmitted to patients with familial IBD

We sequenced exonic regions to profile rare missense variants in patients and their unaffected FDRs, achieving an average coverages of  $\geq 92.3\%$  for the protein-coding regions at 10x per-base depth on all samples (see Table S3, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>).

A series of filtration procedures narrowed down all detected variants ( $n = 154,319$ ) to 28,184 by retaining rare, potentially deleterious protein-altering variants that were transmitted preferentially to patients in a recessive or dominant genetic model.

We identified 17 genes (*AC113554.1*, *ACE*, *AKAP17A*, *AKAP9*, *ANK2*, *ASB16*, *ASIC3*, *DNPH1*, *DUS3L*, *FAM200A*, *FZD10*, *LAMA5*, *NUTM2F*, *PKN1*, *PRR26*, *WDR66*, and *ZC3H4*) that each contained rare damaging missense variant(s) in at least 2 families. All the identified variants followed dominant inheritance. Damaging effects of the observed variants on the gene products were predicted in silico and described in Table 2. For example, the missense variant in *LAMA5* resulting in a change from valine to methionine at amino acid position 3,111 (transcript ENST00000252999) in 2 independent families was



**Figure 3.** Weighted genetic risk scores (wGRS) in patients with familial IBD. A rank-based nonparametric approach was used to estimate the levels of significance for the increase in disease-specific GRS for (a) IBD, (b) CD, and (c) UC in patients in each family based on the rank of observed wGRS within the simulated wGRS distribution of a general population ( $n = 100,000$ ; indicated by gray histograms). At a false-discovery rate of 5%, a statistical significance for the wGRS inflation was determined at a nominal  $P$  value of  $6.25 \times 10^{-3}$  in IBD,  $1.25 \times 10^{-2}$  in CD, and  $2.5 \times 10^{-2}$  in UC, meaning that wGRS values over 4.06 in IBD, 3.94 in CD, and 4.33 in UC are, respectively, considered significantly higher than expected. Patients in a single CD family (indicated by \*) had a significantly high wGRS. CD, Crohn disease; IBD, inflammatory bowel disease; UC, ulcerative colitis.

one of the most deleterious variants according to the prediction methods (predicted as deleterious in SIFT and as probably damaging in PolyPhen-2). Several genes with multiple rare damaging variants were discovered in a disease subtype analysis (*AC113554.1*, *ACE*, *AKAP9*, *ANK2*, *ASIC3*, *DUS3L*, *NUTM2F*, and *ZC3H4* in CD; none in UC; Table S4, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>). Most of the listed genes excluding *LAMA5* and *ACE* have been examined in few or no studies regarding their relationship with IBD or even their own function. However, interestingly, the gene expression levels of most genes were detectable (or relatively high) in disease-relevant tissues such as colon or small intestine and immune tissue (see Table S5, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>) (51).

### Signatures of gut microbial profiles

Gut microbial community in all the subjects was profiled by 16S rRNA sequencing. We observed that the study patients in remission mostly had formed stool at the time of fecal sample collection. The microbial richness estimated by the Shannon diversity index showed a significantly lower alpha diversity in the patients than the unaffected FDRs ( $P$  value = 0.022), highlighting the poor microbial richness in familial CD ( $P$  value = 0.0062), but not UC ( $P$  value = 0.23) compared with the unaffected FDRs (Figure 4a). Similarly, the beta diversity analysis using the weighted UniFrac algorithm did not reveal a significant difference in gut microbial composition between familial CD and controls ( $P$  value > 0.05; Figure 4b).

Bacterial abundance was normalized for the sequencing library size and composition across the individuals to investigate the differential abundance of bacterial taxa in gut in patients with familial IBD. A total of 22 taxa at various taxonomic levels showed significant differences in relative abundance between the patients with familial IBD and the unaffected FDRs at a false discovery rate threshold of 5% (Figure 5 and see Table S6, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>). For example, familial IBD patients had significantly higher levels of the class *Gammaproteobacteria*; the orders *Turicibacterales* and *Pasteurellales*; the families *Turicibacteraceae*, *Veillonellaceae*, *Pasteurellaceae*, and *Peptostreptococcaceae*; and the genera *Veillonella*, *Turicibacter*, and *Haemophilus*, whereas

the patients had significantly lower levels of the class *Deltaproteobacteria*, the order *Desulfovibrionales*, and the family *Desulfovibrionaceae*. Of the 22 detected differentially abundant taxa, 11 have been reported for their IBD-specific abundance in previous studies (17,52–56). The previously reported directions of bacterial abundance in patients with IBD were consistent with those in our analysis, except for 2 taxa with relatively weak fold changes in abundance. In addition, there were the other 11 familial IBD-specific taxa such as an unclassified species of *Bacteroides* and *B. uniformis* with >10-fold decreased abundance in patients with familial IBD.

We also compared the gut microbial profiles between IBD subtypes and unaffected FDRs. We identified 16 familial CD- and UC-specific taxa, 5 of which were reported to be deficient and abundant in patients in other studies (17,52–54,56), consistent with our analysis results. The other 11 unreported taxa were considered novel disease markers under a false discovery rate of 5% (Figure 5, see Tables S7 and S8 Supplementary Digital Content 1, <http://links.lww.com/CTG/A324> and <http://links.lww.com/CTG/A324>).

### DISCUSSION

This study first investigated both the host genetic and gut microbial signatures in familial IBD through an intensive analysis of common and rare variants in host genomic DNA and gut microbial bacteria composition in patients and unaffected FDRs.

Our study yielded 3 important observations. First, known common disease-risk variants were not enriched in patients with familial IBD. We found a little evidence of the inflated genetic burden of IBD in patients with familial IBD because we detected a single trio (out of 8 families) in which 2 patients with CD had significantly higher wGRS based on GWAS-identified variants than expected. Similarly, a previous large-scale study that compared wGRS in patients with familial and sporadic IBD showed a statistically significant but very slight increase in wGRS in familial IBD (57). These 2 consistent results on the weak association of wGRS and familial IBD strongly indicate that the surprisingly high risk of IBD in individuals with a family history of IBD could not be simply explained by a genetic burden of common IBD-risk variants with small effect sizes, which may suggest the presence of



**Table 2.** Genes with rare deleterious alleles that were detected in multiple families with familial IBD and shared among affected FDRs only

Gene	No. of families with any variant	Variant name	Variation location (hg38)	Allele (frequency)	Most affected transcript	Amino acid position	Amino acids	SIFT	PolyPhen-2
AC113554.1	2	rs373319603	chr17: 63494457	A (0.00%)	ENST00000290866	1,123	V/M	Deleterious	Probably damaging
		rs538659872	chr17: 63496827	T (0.02%)	ENST00000290866	1,178	P/L	Deleterious	Possibly damaging
ACE	2	rs373319603	chr17: 63494457	A (0.00%)	ENST00000577647	549	V/M	Deleterious	Possibly damaging
		rs538659872	chr17: 63496827	T (0.02%)	ENST00000577647	604	P/L	Deleterious	Possibly damaging
AKAP17A	2	chrX_1599467_G/A	chrX: 1599467	A (0.00%)	ENST00000381261	396	R/H	Deleterious low confidence	Probably damaging
		chrX_1601004_C/G	chrX: 1601004	G (0.00%)	ENST00000313871	500	P/A	Tolerated low confidence	Benign
		chrX_1601134_C/T	chrX: 1601134	T (0.00%)	ENST00000313871	543	P/L	Deleterious low confidence	Benign
AKAP9	2	rs201048693	chr7: 92096722	G (0.04%)	ENST00000394534	1,101	R/G	Deleterious	Benign
		rs777536948	chr7: 92058368	C (0.00%)	ENST00000491695	67	I/T	Deleterious	Benign
ANK2	2	chr4_113335848_C/G	chr4: 113335848	G (0.00%)	ENST00000357077	1,128	L/V	Deleterious	Probably damaging
		rs759647520	chr4: 113323785	G (0.00%)	ENST00000506722	955	D/G	Deleterious low confidence	Probably damaging
ASB16	2	chr17_44176917_C/T	chr17: 44176917	T (0.00%)	ENST00000293414	250	A/V	Deleterious	Probably damaging
		rs771202055	chr17: 44172127	A (0.00%)	ENST00000293414	128	R/Q	Deleterious	Benign
ASIC3	2	rs777980557	chr7: 151051247	G (0.00%)	ENST00000490540	12	M/R	Deleterious	Probably damaging
DNPH1	2	rs201350596	chr6: 43229300	C (0.02%)	ENST00000361727	172	I/T	Deleterious	Benign
		rs773395453	chr6: 43226051	A (0.00%)	ENST00000230431	120	R/C	Deleterious	Probably damaging
DUS3L	2	rs375298428	chr19: 5785332	A (0.02%)	ENST00000590343	112	T/M	Deleterious low confidence	Benign
		rs757677248	chr19: 5791086	G (0.02%)	ENST00000590110	19	G/E	Deleterious low confidence	Probably damaging
FAM200A	2	chr7_99548059_G/C	chr7: 99548059	C (0.00%)	ENST00000449309	117	R/G	Deleterious	Probably damaging
		rs749157767	chr7: 99548011	A (0.00%)	ENST00000408938	133	R/W	Deleterious	Probably damaging
FZD10	2	rs199663032	chr12: 130163216	T (0.10%)	ENST00000229030	92	A/S	Deleterious	Probably damaging
		rs61758967	chr12: 130164446	A (0.10%)	ENST00000229030	502	E/K	Tolerated	Probably damaging
LAMA5	2	rs199902568	chr20: 62312429	T (0.00%)	ENST00000252999	3,111	V/M	Deleterious	Possibly damaging

Table 2. (continued)

Gene	No. of families with any variant	Variant name	Variation location (hg38)	Allele (frequency)	Most affected transcript	Amino acid position	Amino acids	SIFT	PolyPhen-2
NUTM2F	2	rs12237784	chr9: 94318485	G (0.00%)	ENST00000253262	751	K/Q	Deleterious	Benign
		rs773143235	chr9: 94318511	A (0.00%)	ENST00000253262	742	D/V	Deleterious	Possibly damaging
PKN1	2	rs200208099	chr19: 14469964	A (0.20%)	ENST00000591461	161	A/T	Tolerated low confidence	Probably damaging
		rs368909367	chr19: 14443533	T (0.00%)	ENST00000586557	141	A/V	Tolerated	Probably damaging
PRR26	2	chr10_662817_G/A	chr10: 662817	A (0.00%)	ENST00000381489	151	R/Q	Deleterious low confidence	Benign
WDR66	2	rs201018883	chr12: 121999823	C (0.10%)	ENST00000288912	1,038	K/N	Deleterious	Benign
ZC3H4	2	rs144221330	chr19: 47086476	T (0.69%)	ENST00000253048	260	G/S	Tolerated	Probably damaging
		rs373777449	chr19: 47066888	A (0.00%)	ENST00000601973	734	A/V	Deleterious	Probably damaging

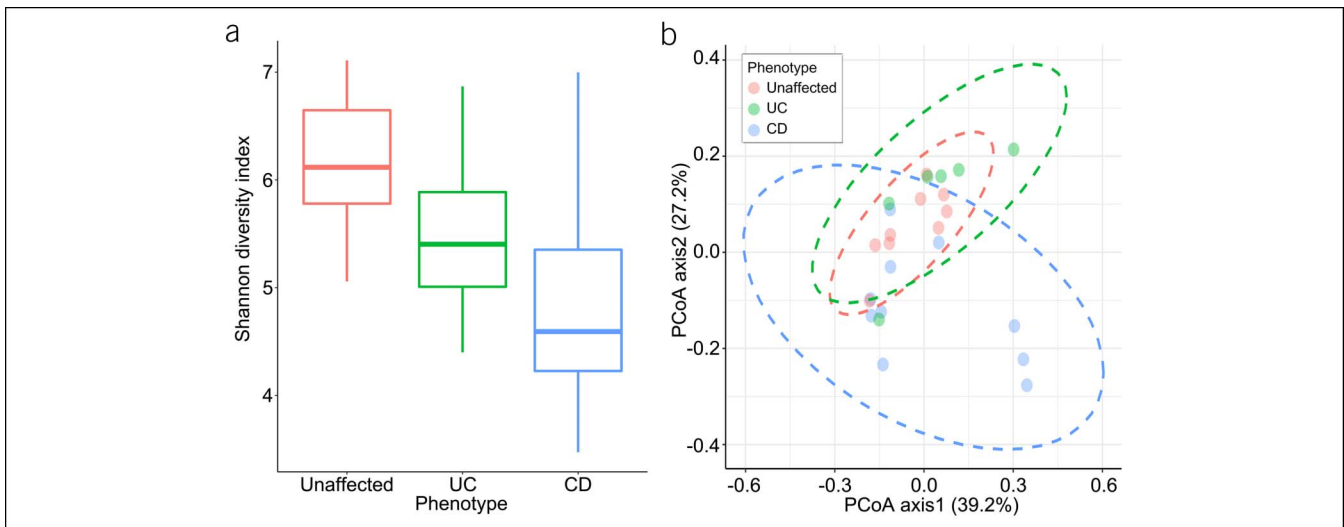
The effect of the amino acid substitution on protein function was predicted using SIFT and PolyPhen-2. In Polyphen-2, the predicted damaging impact increases in the order of "benign," "possibly damaging," and "probably damaging." The population-level frequencies of potentially deleterious alleles were retrieved from the data in East Asians or all individuals of the 1,000 Genomes projects phase 3. (Higher frequencies were written.)

FDR, first-degree relative; IBD, inflammatory bowel disease.

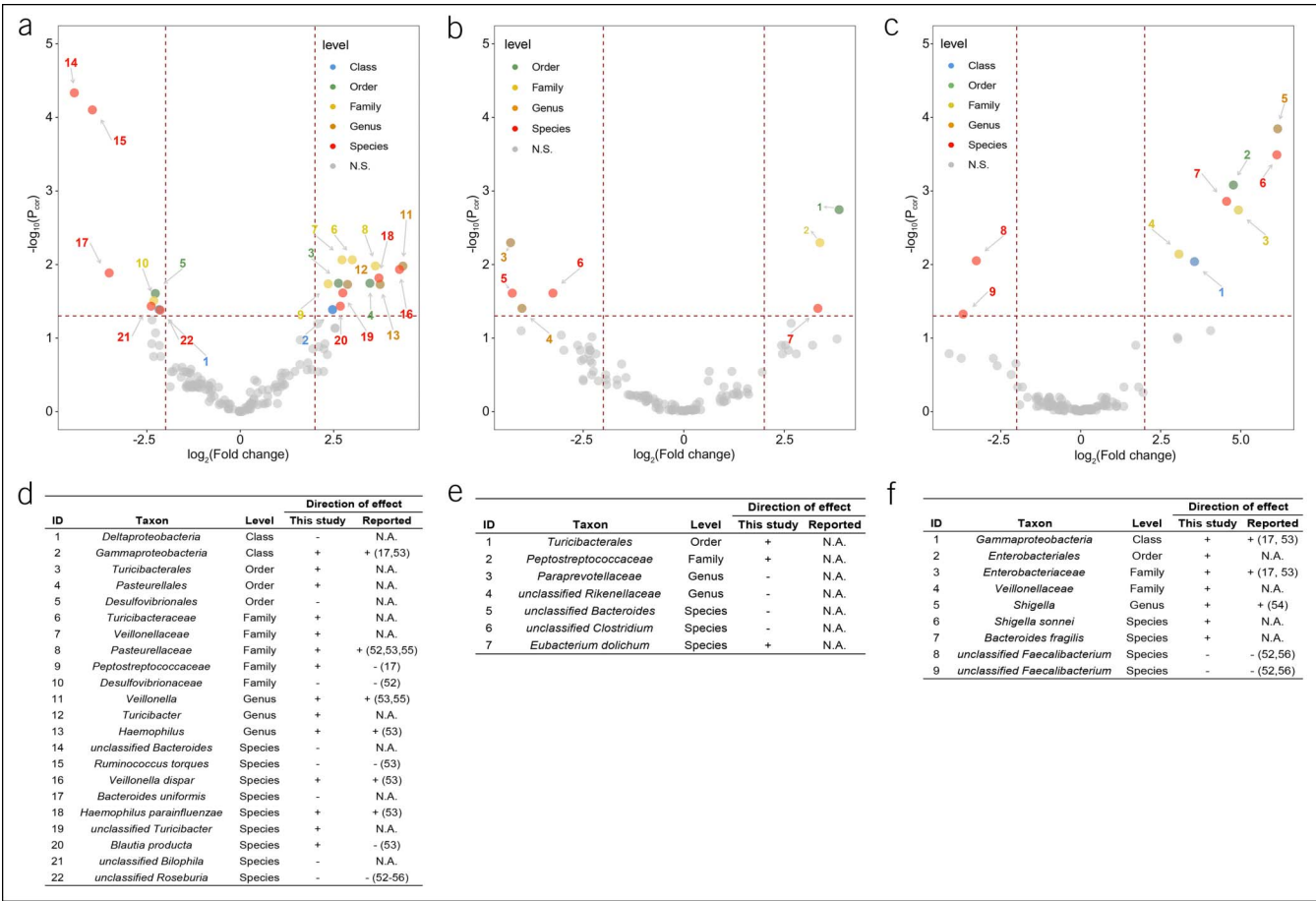
rare family-private variants with large effect sizes and/or family-shared environmental factors.

Second, we detected gene-wise recurrent family-specific missense variants that appeared among patients with familial IBD in 7 of 8

families. Those variants may explain a large portion of IBD risk, although these findings need to be replicated in independent families or validated in experimental approaches to link the disease-associated genes/variants to IBD. Interestingly, the expression



**Figure 4.** Taxonomic alpha and beta diversities in the gut microbial community in families with familial inflammatory bowel disease. **(a)** the alpha-diversity of gut microbial composition in each study subject was estimated based on Shannon index and box-plotted according the individual phenotypes—unaffected FDRs (denoted as unaffected) and patients with ulcerative colitis and Crohn disease. There was a significant difference in alpha diversity between inflammatory bowel disease patients and unaffected FDRs ( $P$  value = 0.022; Kruskal-Wallis test). **(b)** the beta-diversity of bacterial composition between individuals was measured using a principal coordinate analysis (PCoA) with the weighted UniFrac algorithm and plotted based on PCoA axis 1 and 2 that, respectively, explained 39.2% and 27.2% of the variance in bacterial community of the study subjects. The significant differences in bacterial community was not observed between the patients and unaffected FDRs ( $P$  value > 0.05; PERMANOVA). CD, Crohn disease; FDR, first-degree relative; PERMANOVA, permutational multivariate analysis of variance; UC, ulcerative colitis.



**Figure 5.** Differentially abundant taxa in the gut microbial community in patients with familial inflammatory bowel disease (**a**), ulcerative colitis (**b**), and Crohn's disease (**c**). All the taxa with false discovery rate-corrected  $P$  values ( $P_{adj}$ )  $\leq 0.05$  and  $\geq 4$ -fold abundance changes were numbered to provide detailed information in separate tables (**d–f**) including taxa names, level, and the direction of effect (+, increased in patients; -, decreased in patients) in this study and previous large studies. Reference numbers were provided with the direction of effect. N.A., not available; N.S., not significant.

levels of the identified genes with deleterious missense variants were detectable or even relatively high in gastrointestinal tissues or immune cells, suggesting that they play pathological roles in disease-relevant tissues. For example, *LAMA5*, one of the identified genes, encodes a subunit of the extracellular matrix protein called laminin that is a key component in maintaining integrity of the intestinal basement membrane layer. A recent GWAS suggested the genetic association of common variants within the *LAMA5* gene with pediatric IBD (58). Another identified gene *ACE* is highly expressed in small intestine (see Tables S5, Supplementary Digital Content 1, <http://links.lww.com/CTG/A324>), especially in patients with IBD (59). Recent studies described the potentially therapeutic effect of ACE inhibitors on IBD (60,61).

Third, we found distinct dysbiosis of gut microbiota at various taxonomic levels and reduced alpha diversity in the patients with familial IBD, compared with their unaffected FDRs. Our analysis confirmed several previously known IBD-specific taxa and the lowered alpha diversity in patients, indicating that there is some confidence in our samples and analysis. In addition, we discovered a large number of familial IBD-specific taxa that have not yet been suggested in sporadic patients with IBD. The newly identified taxa had large differences in abundance between patients and controls, which may shed light on the missing liability of familial IBD and may

facilitate study on the interplay between their metabolites and host factors that requires much larger sample sizes and experimental validation.

All the affected participants with familial IBD in the study were registered in a single prospective IBD cohort characterized in Kyung Hee Medical Center, which helped to prepare the well-characterized clinical information and samples, eliminating potential recall and selection bias. However, although the collection of both genetic and gut microbial data from familial IBD families were highly challenging, we note that the major limitations in our study are the lack of a replication study using an independent cohort of familial IBD and the potentially low statistical power because of the relatively small sample size that might cause a large proportion of true disease signatures to be missing.

In summary, this study identified familial IBD-associated rare deleterious variants and gut microbial dysbiosis in Korean families with familial IBD, highlighting the presence of known and novel, large-effect factors and suggesting a complex interplay between host and gut microbiota involved in the immune response in familial IBD. The present results may have important clinical implications on prediction of IBD risk in asymptomatic relatives of patients with familial IBD. The unaffected relatives with family-specific deleterious genetic variants and/or gut



dysbiosis may require closer follow-up with longitudinal profiling of gut microbiota for early diagnosis of IBD.

## CONFLICTS OF INTEREST

**Guarantor of the article:** Hyo Jong Kim, MD, PhD, and Kwangwoo Kim, PhD.

**Specific author contributions:** Yoo Min Park, MD and Eunji Ha, BSc contributed equally to this work. Chang Kyun Lee, MD, PhD, Kwangwoo Kim, PhD, and Hyo Jong Kim, MD, PhD, contributed equally to this work. Study conception and design: C.K.L, K.K, and H.J.K. Recruiting study subjects and characterizing patients with IBD: Y.M.P., G.Y.S., and C.K.L. Generating biological data: Y.M.P., G.Y.S., C.K.L., and H.J.K. Analyzing the data and interpreting the results: Y.M.P., E.H., K.N.G., C.K.L., K.K., and H.J.K. Writing the manuscript: Y.M.P., E.H., C.K.L., K.K., and H.J.K.

**Financial support:** This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017R1A5A2014768).

**Potential competing interests:** None to report.

**Ethics approval and consent to participate:** All study participants provided written informed consent for this work. The study was approved by the Institutional Review Board of Kyung Hee University Hospital (KHUH-2018-03-009).

## Study Highlights

### WHAT IS KNOWN

- ✓ The family history of IBD is one of the most powerful predictors of developing IBD.
- ✓ Both genetic variants and gut dysbiosis confer the susceptibility to IBD.

### WHAT IS NEW HERE

- ✓ Genomic analyses identifies several genes that each has rare, deleterious patient-specific variants in multiple families with familial IBD, but the high genetic load of common IBD-risk variants is not found in patients with familial IBD.
- ✓ The significant differences in gut microbial composition between patients with familial IBD and their unaffected first-degree relatives suggest the potential roles of large-effect novel and known IBD-risk taxa in the development of familial IBD.

### TRANSLATIONAL IMPACT

- ✓ This study provides a catalogue of familial IBD-specific biomarkers including human genes with rare, deleterious variants, and gut microbial taxa with differential abundance in patients with familial IBD, which can be used to predict risk of familial IBD.
- ✓ This study will contribute to revealing the pathological mechanism by which the newly identified biomarkers increase a risk for developing IBD and presenting clinical symptoms.

## REFERENCES

1. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: A systematic review of population-based studies. *Lancet* 2018;390(10114):2769–78.
2. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;474(7351):307–17.
3. Kirsner JB, Spencer JA. Family occurrences of ulcerative colitis, regional enteritis, and ileocolitis. *Ann Intern Med* 1963;59:133–44.
4. Torres J, Burisch J, Riddle M, et al. Preclinical disease and preventive strategies in IBD: Perspectives, challenges and opportunities. *Gut* 2016; 65(7):1061–9.
5. Lashner BA, Evans AA, Kirsner JB, et al. Prevalence and incidence of inflammatory bowel disease in family members. *Gastroenterology* 1986; 91(6):1396–400.
6. Park JB, Yang SK, Byeon JS, et al. Familial occurrence of inflammatory bowel disease in Korea. *Inflamm Bowel Dis* 2006;12(12):1146–51.
7. Moller FT, Andersen V, Wohlfahrt J, et al. Familial risk of inflammatory bowel disease: A population-based cohort study 1977–2011. *Am J Gastroenterol* 2015;110(4):564–71.
8. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017;49(2):256–61.
9. Yang SK, Hong M, Zhao W, et al. Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut* 2014;63(1):80–7.
10. Mirkov MU, Verstockt B, Cleynen I. Genetics of inflammatory bowel disease: Beyond NOD2. *Lancet Gastroenterol Hepatol* 2017;2(3):224–34.
11. Park SC, Jeon YT. Genetic studies of inflammatory bowel disease-focusing on Asian patients. *Cells* 2019;8(5):404.
12. Stahl EA, Wegmann D, Trynka G, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012;44(5):483–9.
13. Somineni HK, Kugathasan S. The microbiome in patients with inflammatory diseases. *Clin Gastroenterol Hepatol* 2019;17(2):243–55.
14. Chu H, Khosravi A, Kusumawardhani IP, et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science* 2016;352(6289):1116–20.
15. Knights D, Silverberg MS, Weersma RK, et al. Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med* 2014;6(12):107.
16. Scott KP, Gratz SW, Sheridan PO, et al. The influence of diet on the gut microbiota. *Pharmacol Res* 2013;69(1):52–60.
17. Imhann F, Vich Vila A, Bonder MJ, et al. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* 2018;67(1):108–19.
18. Dill-McFarland KA, Tang ZZ, Kemis JH, et al. Close social relationships correlate with human gut microbiota composition. *Sci Rep* 2019;9(1):703.
19. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019;13(2):144–64.
20. Sturm A, Maaser C, Calabrese E, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD Part 2: IBD scores and general principles and technical aspects. *J Crohns Colitis* 2019;13(3):273–84.
21. Moon S, Kim YJ, Han S, et al. The Korea biobank array: Design and identification of coding variants associated with blood biochemical traits. *Sci Rep* 2019;9(1):1382.
22. Delaneau O, Howie B, Cox AJ, et al. Haplotype estimation using sequencing reads. *Am J Hum Genet* 2013;93(4):687–96.
23. Delaneau O, Marchini J, Genomes Project C, et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* 2014;5:3934.
24. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48(10):1284–7.
25. Yang SK, Hong M, Oh H, et al. Identification of loci at 1q21 and 16q23 that affect susceptibility to inflammatory bowel disease in Koreans. *Gastroenterology* 2016;151(6):1096–9.e1094.
26. Hong M, Ye BD, Yang SK, et al. Immunochip meta-analysis of inflammatory bowel disease identifies three novel loci and four novel associations in previously reported loci. *J Crohns Colitis* 2018;12(6): 730–41.
27. Lee HS, Lee SB, Kim BM, et al. Association of CDKN2A/CDKN2B with inflammatory bowel disease in Koreans. *J Gastroenterol Hepatol* 2018; 33(4):887–93.
28. Kakuta Y, Kawai Y, Naito T, et al. A genome-wide association study identifying RAP1A as a novel susceptibility gene for Crohn's disease in Japanese individuals. *J Crohns Colitis* 2019;13(5):648–58.
29. Jung ES, Cheon JH, Lee JH, et al. HLA-C\*01 is a risk factor for Crohn's disease. *Inflamm Bowel Dis* 2016;22(4):796–806.

30. Paige E, Barrett J, Stevens D, et al. Landmark models for optimizing the use of repeated measurements of risk factors in electronic health records to predict future disease risk. *Am J Epidemiol* 2018;187(7):1530–8.
31. Yang SK, Hong M, Choi H, et al. Immunochip analysis identification of 6 additional susceptibility loci for Crohn's disease in Koreans. *Inflamm Bowel Dis* 2015;21(1):1–7.
32. Yamazaki K, Umeno J, Takahashi A, et al. A genome-wide association study identifies 2 susceptibility Loci for Crohn's disease in a Japanese population. *Gastroenterology* 2013;144(4):781–8.
33. Asano K, Matsushita T, Umeno J, et al. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat Genet* 2009;41(12):1325–9.
34. Yang SK, Hong M, Zhao W, et al. Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic susceptibility with Caucasians. *Inflamm Bowel Dis* 2013;19(5):954–66.
35. Moon CM, Kim SW, Ahn JB, et al. Deep resequencing of ulcerative colitis-associated genes identifies novel variants in Candidate genes in the Korean population. *Inflamm Bowel Dis* 2018;24(8):1706–17.
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9.
37. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;43:11 10 11–33.
38. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 2018:201178.
39. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308–11.
40. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol* 2016;17(1):122.
41. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6(2):80–92.
42. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114–20.
43. Whon TW, Chung WH, Lim MY, et al. The effects of sequencing platforms on phylogenetic resolution in 16 S rRNA gene profiling of human feces. *Sci Data* 2018;5:180068.
44. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37(8):852–7.
45. Rognes T, Flouri T, Nichols B, et al. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 2016;4:e2584.
46. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72(7):5069–72.
47. Bokulich NA, Kaehler BD, Rideout JR, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;6(1):90.
48. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;26(1):32–46.
49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550.
50. O'Donnell L, Virjee J, Heaton KW. Detection of pseudodiarrhoea by simple clinical assessment of intestinal transit rate. *BMJ: Br Med J* 1990;300(6722):439.
51. Jain A, Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* 2019;35(11):1966–7.
52. Pascal V, Pozuelo M, Borruel N, et al. A microbial signature for Crohn's disease. *Gut* 2017;66(5):813–22.
53. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15(3):382–92.
54. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13(9):R79.
55. Lewis JD, Chen EZ, Baldassano RN, et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* 2015;18(4):489–500.
56. Erickson AR, Cantarel BL, Lamendella R, et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS One* 2012;7(11):e49138.
57. Borren NZ, Conway G, Garber JJ, et al. Differences in clinical course, genetics, and the microbiome between familial and sporadic inflammatory bowel diseases. *J Crohns Colitis* 2018;12(5):525–31.
58. Shaw KA, Cutler DJ, Okou D, et al. Genetic variants and pathways implicated in a pediatric inflammatory bowel disease cohort. *Genes Immun* 2019;20(2):131–42.
59. Ning L, Shan G, Sun Z, et al. Quantitative proteomic analysis reveals the deregulation of nicotinamide adenine dinucleotide metabolism and CD38 in inflammatory bowel disease. *Biomed Res Int* 2019;2019:3950628.
60. Fairbrass KM, Hoshen D, Gracie DJ, et al. Effect of ACE inhibitors and angiotensin II receptor blockers on disease outcomes in inflammatory bowel disease. *Gut* 2020 [Epub ahead of print April 2, 2020.] (doi:10.1136/gutjnl-2020-321186).
61. Garg M, Royce SG, Tikellis C, et al. Imbalance of the renin-angiotensin system may contribute to inflammation and fibrosis in IBD: A novel therapeutic target? *Gut* 2020;69(5):841–51.

**Open Access** This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work, provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.