이학박사 학위논문

한국인 염증성 장질환 감수성 유전자 발굴

Identification of susceptibility genes for inflammatory bowel disease in Koreans

울산대학교대학원

의 과 학 과

정 슬 기

한국인 염증성 장질환 감수성 유전자 발굴

지도교수 송 규 영

이 논문을 이학박사 학위 논문으로 제출함

2021 년 8 월

울산대학교대학원

의 과 학 과

정 슬 기

정슬기의 이학박사학위 논문을 인준함

심사위원 예 병 덕 (인)

심사위원 송 규 영 (인)

심사위원 이 종 극 (인)

심사위원 이 호 수 (인)

심사위원 한 범 (인)

울 산 대 학 교 대 학 원

2021 년 8 월

ABSTRACT

Inflammatory bowel disease (IBD), represented by Crohn's disease (CD) and ulcerative colitis (UC), is a chronic inflammatory disease of the gastrointestinal tract. CD is characterized by transmural granulomatous inflammation which affects any part of the gastrointestinal tract, whereas UC is characterized by mucosal inflammation limited to colon. IBD is thought to develop due to dysregulated mucosal immune responses to gut flora in genetically susceptible individuals.

Recent meta-analysis of genome-wide association studies (GWASs) of IBD performed in populations of European origin identified over 240 susceptibility loci, improving our understanding of IBD genetics. However, identified common variants account for only a fraction of IBD heritability. Moreover, despite of observed differences in clinical characteristics of IBD among different ethnicities, there have been limited studies in non-European populations.

To identify additional IBD susceptibility loci in Asians, we performed a GWAS using 1,726 IBD cases and 378 healthy controls genotyped using the Infinium Asian Screening Array-24 v1.0 (Illumina), and combined our previous GWAS dataset consisted of 1,469 IBD cases and 4,041 healthy controls using an inverse-variance fixed-effects meta-analysis in Korean population. We selected 10 novel candidate loci applying a threshold of $P_{meta} < 1 \times 10^{-6}$, and performed replication study using an additional 1,088 IBD cases and 845 controls. The meta-analysis of two GWAS datasets in Koreans identified 1 novel locus for ulcerative colitis at rs76227733 on 10q24 ($P_{combined} = 6.56 \times 10^{-9}$) and 2 novel loci for CD at rs2240751 on 19p13 ($P_{combined} = 3.03 \times 10^{-8}$) and rs6936629 in on 6q22 ($P_{combined} = 3.63 \times 10^{-8}$). Additionally, we examined 245 previously established loci in Europeans in our meta-analysis data. A total of 33 established loci were replicated in Korean population.

To gain insight into the potential functional roles of the identified loci, we performed RNA-sequencing using whole blood tissues of 101 Korean CD patients, and then built the eQTL database (http://asan.crohneqtl.com/). In the eQTL analysis, we identified 135,164 cis-eQTLs and 3,816 eGenes with the false discovery rate < 0.05. Integrated analysis of the extended GWAS and eQTL data revealed two target genes at two previously reported loci for IBD including *TNFSF15* at 9q32 and *GPR35* at 2q37. The IBD risk alleles from the two loci were associated with lower expression of *TNFSF15* or *GPR35* than protective alleles.

To compare biological pathways associated with CD or UC between Asians and Europeans, we performed pathway analysis using meta-analysis of two GWAS datasets in Koreans and summary statistics of GWAS in Europeans. In the case of CD, MHC and antigenic stimulus-related pathways were significant in Korean, whereas cytokine and transcription factor-related pathways were significant in European. In the case of UC, MHC and antigen binding-related pathways identified in the Korean population were also significant in the European population. We also estimated phenotypic variance of CD or UC based on the polygenic risk score (PRS). Variance explained by PRS derived from Korean data explained up to 14% of variance of CD, whereas those derived from European data explained 10%. For UC, variance explained by PRS_{EUR} of 12% was better than those explained by PRS_{KOR} of 7%.

We identified 3 novel susceptibility loci for IBD and replicated 33 previously reported loci, indicating distinct as well as common pathways associated with IBD in Europeans and Asians. The current study increased the number of IBD susceptibility loci in Koreans to 54. Our pathway analysis showed major differences in biological pathways associated with CD between East Asians and Europeans. In addition, PRS analysis showed that PRS of CD based on European data is less predictive in Koreans. These findings are consistent with our previous report that the effects for CD were more population-specific than for UC, emphasizing on diversity in genetic research.

Key words: inflammatory bowel disease; GWAS; eQTL; pathway analysis; polygenic risk scores

CONTENTS

LIST OF TABLES

LIST OF FUGURES

ABBREVIATIONS

CD	Crohn's disease
eQTL	Expression quantitative trait locus
GWAS	Genome-wide association study
HWE	Hardy-Weinberg Equilibrium
IBD	Inflammatory bowel disease
LD	Linkage disequilibrium
MAF	Minor allele frequency
MHC	Major histocompatibility complex
OR	Odds ratio
UC	Ulcerative colitis
PCA	Principal component analysis
PRS	Polygenic risk score
SNP	Single nucleotide polymorphism

1. INTRODUCTION

Inflammatory bowel disease (IBD) is a chronic relapsing inflammatory disorder of the gastrointestinal tract. Crohn's disease (CD) and ulcerative colitis (UC) are the two major subtypes of IBD. Although these two forms of IBD share similar clinical and pathological features including diarrhea, fever, rectal bleeding, and weight loss, there are differences in disease localization, histopathology and endoscopic features, suggesting differences in the underlying pathogenic mechanisms for each disease.[1,2] CD is characterized by segmental, transmural granulomatous inflammation that can affect anywhere in the intestinal tract from the mouth to the anus. The complications of CD include strictures, abscesses, and fistulas. The clinical phenotypes of CD with respect to disease location and occurrence of complications were defined by the Vienna classification.[3,4] Based on disease location, the major divisions are L1 (terminal ileum), L2 (colon), L3 (ileocolon), and L4 (upper GI). Disease behavior is classified as B1 (nonstricturing, nonpenetrating), B2 (stricturing), and B3 (penetrating). UC is limited to the mucosal layer of colonic tissue. Based on the anatomic extent of involvement, patients can be classified as having proctitis, left-sided colitis, and extensive colitis.[2]

The prevalence of IBD is lower in Asia than in the West; however, its incidence is rapidly increasing throughout Asia.[5-8] Epidemiological and clinical studies showed that the phenotype and clinical course of IBD differs between Asians and Europeans.[5,6,9,10] There is a male predominance of CD in Asia with a male-to-female ratio ranging from 1.67:1 to 2.9:1, while the ratio is known to be equal in Western countries.[5,9,10] Another difference of note is the disease location of CD. In the West, the proportions of CD patients having the disease locations of ileum alone, colon alone, and both the ileum and colon are approximately equal, whereas ileocolonic disease is predominant in Asia, accounting for around two-thirds of CD cases, and colonic and ileal disease account for 4~12% and 20~23%, respectively.[6,9,10] It is unclear whether these Asian-specific clinical characteristics of IBD are solely due to different environments between the East and West, which highlights the need for genetic studies of IBD in Asian population.

IBD is thought to arise by dysregulated mucosal immune responses to the gut flora in genetically susceptible individuals.[11] Family and twin studies showed that a positive family history is an important risk factor in both Korea and Western countries.[12,13] Previous genome-wide

association studies (GWASs) of European ancestry have greatly advanced our understanding in IBD genetics.[14-16] A meta-analysis by the International IBD Genetics Consortium, which combined GWAS and Immunochip data from 96,486 individuals with multiple ancestries including Asian samples, identified over 200 susceptibility loci for IBD and reported an overlap in the directionality of the odds ratios between European and Asian cohorts.[15] The latest genome-wide meta-analysis performed on populations of European ancestry reported 241 susceptibility loci for IBD.[16] However, identified common variants account for only a fraction of IBD. Moreover, despite of observed differences in clinical characteristics of IBD among different ethnicities, there have been limited studies in non-European populations.[17-25] Recent Asian GWAS of IBD identified a total of 46 susceptibility loci for IBD including 5 Asian-specific loci (*ATG16L2*, *SHC1*, *CDKN2A*, *ELF1*, *CDYL2*) and 41 established loci.[19-25] To identify additional susceptibility loci in Asians, we performed an extended GWAS by newly including 1,726 IBD cases (725 CD and 1001 UC) and 378 healthy controls genotyped using the Infinium Asian Screening Array-24 v1.0 (Illumina). We then conducted a GWAS meta-analysis of the two datasets, comprising a total of 3,195 cases and 4,419 controls of Korean ancestry. We used 1,088 additional cases (582 CD and 506 UC) and 845 additional healthy controls as a replication cohort.

The majority of the single nucleotide polymorphisms (SNPs) identified from GWASs are in the non-coding or intergenic regions of the genome with largely unknown regulatory functions, suggesting that the SNPs may affect the trait through regulation of gene expression. Pinpointing which genes are affected by the causal SNPs is essential to increase our insight into the biological mechanisms underlying causes of IBD. Regulatory elements can act over a long distance and in a cell-type specific manner, making the identification of the causal genes for a given pathologic condition and their roles extremely difficult. Expression quantitative trait locus (eQTL) studies associate genomic and transcriptomic data sets from the same individuals to identify loci that affect mRNA expression. By linking SNPs to changes in gene expression, eQTL can be useful for annotating GWAS variants. For the eQTL analysis, RNA sequencing (RNA-seq) is performed to quantify of mRNA expression level in biological samples. The workflow begins with RNA extraction, followed by ribosomal RNA depletion, cDNA synthesis, and preparation of an adaptor-ligated sequencing library. Next steps are aligning the sequencing reads to a reference genome, quantifying reads that overlap transcripts, filtering, and normalizing between samples. Previous

eQTL study using RNA-seq data of 280 intestinal mucosal biopsy samples from 165 IBD patients identified 172 target genes from the colocalization analysis with IBD GWAS in Europeans.[26] To determine the most functionally relevant genes at the IBD susceptibility loci identified in Asians, we performed RNA-seq using whole blood tissues of 101 Korean CD patients, and built the eQTL database (http://asan.crohneqtl.com/).

For various diseases, GWASs identified common SNPs with small effect sizes. Although a single SNP is not informative for assessing the disease risk, a combined effect size of all causal SNPs could explain substantial phenotypic variance of the disease.[27] The polygenic risk score (PRS) is calculated as a weighted sum of the number of risk alleles carried by an individual, where the risk alleles and their weights are defined by the loci and their measured effects from GWASs. The PRS was shown to have potential for broad-scale clinical uses including early detection and treatment of these diseases.[28] One of the most challenging aspects of moving PRS to the clinical use is ensuring that they are equally applicable to all health care users across ethnic groups.[29] Transferability of PRS across populations is limited, because PRS based on GWAS in one population usually provided attenuated predictive accuracy in other populations.[29-31] These findings highlight the need for large scale GWAS in diverse human populations to increase predictive accuracy of PRS.

We aimed to identify novel genetic variants associated with IBD in Koreans through an extended GWAS. Our meta-analyses identified 3 novel IBD loci and replicated 33 previously reported loci, increasing the number of IBD susceptibility loci in Koreans to 54. In addition, we performed eQTL analysis using RNA-seq of whole blood in a cohort of 101 Korean CD patients. As the eQTL database was constructed using CD patients and not healthy individuals, Korean CD eQTL datasets might provide a valuable resource for link between genetic variation and gene expression and regulation not only in Asians, but also in Europeans if new eQTL in patients become only evident when the gene is overexpressed as a result of modified inflammatory status. We estimated phenotypic variance explained by PRS of CD or UC based on GWAS in Koreans or Europeans, and compared the calculated variances between the two base files.

2. MATERIALS AND METHODS

2.1. Study subjects

Cohort I included 1,726 IBD cases (725 CD and 1,001 UC) and 378 healthy controls genotyped using the Infinium Asian Screening Array-24 v1.0 (Illumina). For discovery, we combined cohort I with cohort II consisted of 1,469 IBD patients (896 CD and 573 UC) and 4,041 controls genotyped using the OmniExpress and Omni1-Quad from our previously published GWAS[25] (Table 1). The replication cohort consisted of 1,088 individuals with IBD (582 CD and 506 UC) and 845 healthy controls. In total, 9,547 samples including 4,283 IBD patients (2,203 CD and 2,080 UC cases) and 5,264 controls were used for meta-analysis in the Korean population. The clinical characteristics of the patients are summarized in Table 2. All IBD patients were recruited from the IBD Clinic of Asan Medical Center.

2.2. Quality controls

Quality control (QC) was conducted for each dataset separately and the combined set of samples using a common approach. Standard QC procedures were performed using PLINK v1.9 (https://www.cog-genomics.org/plink2) and R 3.5.0 (http://www.r-project.org/). After removing gender mismatched 11 samples (3 CD, 5 UC, and 3 controls), all single nucleotide polymorphisms (SNPs) on the X, Y, and mitochondrial chromosomes were excluded. Insertion or deletion polymorphisms, SNPs with duplication, call rate < 98%, Hardy-Weinberg equilibrium (HWE) test $P < 1.0 \times 10^{-5}$ for controls, or minor allele frequency (MAF) < 0.01 were excluded. Five samples with a high proportion of missing genotypes (> 4%) (1 CD, 1 UC, and 3 controls) were removed. Nine samples (4 CD and 5 UC) were removed due to close genetic relatedness (PI_HAT > 0.2, IBS > 0.8). Subsequently, a total of 85 SNPs with $P < 1.0 \times 10^{-5}$ in differential missingness analysis were excluded. The principal-component analysis (PCA) was performed to detect population outliers and stratification by calculating first 10 PCs per individual using PLINK v1.9 after merging with 194 reference samples including European (CEU), Asian (CHB + JPT), and African (YRI) samples from the International HapMap Project. One CD case was removed following PCA (Figure 1A and B). After SNP and sample QC of cohort I data, 457,272 SNPs in 1,726 cases and 378 controls (average call rate of 99.99%) remained for further analyses (Table 3). We also applied the same QC procedures for cohort II[25] in this study (Table 4). Overlapping samples (13 CD and

Table 1. Study cohorts and genotyping platforms

Cohort	Platform	No. of samples			
		IBD	CD	UC	Controls for IBD/CD/UC
I	Asian Screening Array (Illumina)	1,726	725	1,001	378
II*	OmniExpress, Omni1-Quad (Illumina)	1,469	896	573	4,041
Combined		3,195	1,621	1,574	4,419
Replication	TaqMan genotyping assay (Thermo Fisher Scientific)	1,088	582	506	845

CD, Crohn's disease; IBD, inflammatory bowel disease; UC, ulcerative colitis.

*Cohort II: Previous our Korean GWAS data from Yang et al (ref. 25).

Table 2. Clinical characteristics of IBD patients

	Cohort I				Cohort II				Replication				Total			
	IBD	CD	UC	Control	IBD	CD	UC	Control	IBD	CD	UC	Control	IBD	CD	UC	Control
No. of samples	1,726	725	1,001	378	1,469	896	573	4,041	1,088	582	506	845	4,283	2,203	2,080	5,264
Male (%)	1,166 (67.6)	561 (77.4)	605 (60.4)	190 (50.3)	953 (64.9)	633 (70.6)	320 (55.9)	1,602 (39.6)	718 (66.1)	425 (73.0)	293 (58.0)	653 (77.3)	2,837 (66.3)	1,619 (73.5)	1,218 (58.6)	2,445 (46.4)
Mean age at sampling (yr)	35.5 ± 14.7	27.6 ± 9.2	41.1 ± 15.3	NA	31.2 ± 13.6	25.5 ± 9.1	40.1 ± 14.6	NA	34.5 ± 13.6	28.9 ± 9.9	40.9 ± 14.4	NA	33.7 ± 14.2	27.1 ± 9.4	40.8 ± 15.0	NA
Mean age at diagnosis (yr)	31.7 ± 14.0	24.2 ± 8.8	37.1 ± 14.5		27.6 ± 12.6	22.3 ± 8.2	36.0 ± 13.9		30.5 ± 12.9	25.3 ± 9.1	36.6 ± 13.9		30.0 ± 13.4	23.7 ± 8.7	36.7 ± 14.2	
Age group at diagnosis (%)																
≤ 16	141 (8.3)	104 (14.6)	37 (3.8)		288 (19.6)	237 (26.5)	51 (8.9)	NA	74 (6.8)	57 (9.8)	17 (3.4)	NA	503 (11.8)	398 (18.2)	105 (5.1)	NA
17~40	1,123 (66.1)	566 (79.5)	557 (56.5)		901 (61.4)	621 (69.3)	280 (49.0)	NA	771 (70.9)	478 (82.1)	293 (58.0)	NA	2,795 (65.7)	1,665 (76.0)	1,130 (54.8)	NA
≥ 40	434 (25.6)	42 (5.9)	392 (39.8)		279 (19.0)	38 (4.2)	241 (42.1)	NA	242 (22.3)	47 (8.1)	195 (38.6)	NA	955 (22.5)	127 (5.8)	828 (40.1)	NA
NA	28	13	15		1		1		1		1		30	13	17	
Location, no. (%)																
Ileum		106 (20.4)				158 (18.0)				147 (25.4)				411 (20.8)		
Colon		28 (5.4)				48 (5.5)				15 (2.6)				91 (4.6)		
Ileocolon		385 (74.2)				674 (76.6)				417 (72.0)				1,476 (74.6)		
NA		206				16				3				225		
Extent, no. (%)																
Proctitis			235 (34.3)				155 (27.2)				169 (33.9)				559 (31.9)	
Left-sided colitis			184 (26.8)				179 (31.5)				160 (32.1)				523 (29.8)	
Extensive colitis			267 (38.9)				235 (41.3)				169 (33.9)				671 (38.3)	
NA			315				4				8				327	
Behavior, no. (%)																
Inflammatory		267 (49.1)				343 (39.1)				249 (43.2)				859 (43.0)		
Stricturing		98 (18.0)				173 (19.7)				104 (18.0)				375 (18.8)		
Penetrating		179 (32.9)				362 (41.2)				224 (38.8)				765 (38.3)		
NA		181				18				5				204		
Perianal fistula, no. (%)																
No		264 (38.1)				325 (38.5)				246 (44.3)				835 (39.9)		
Yes		429 (61.9)				519 (61.5)				309 (55.7)				1,257 (60.1)		
NA		32				52				27				111		

IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis.

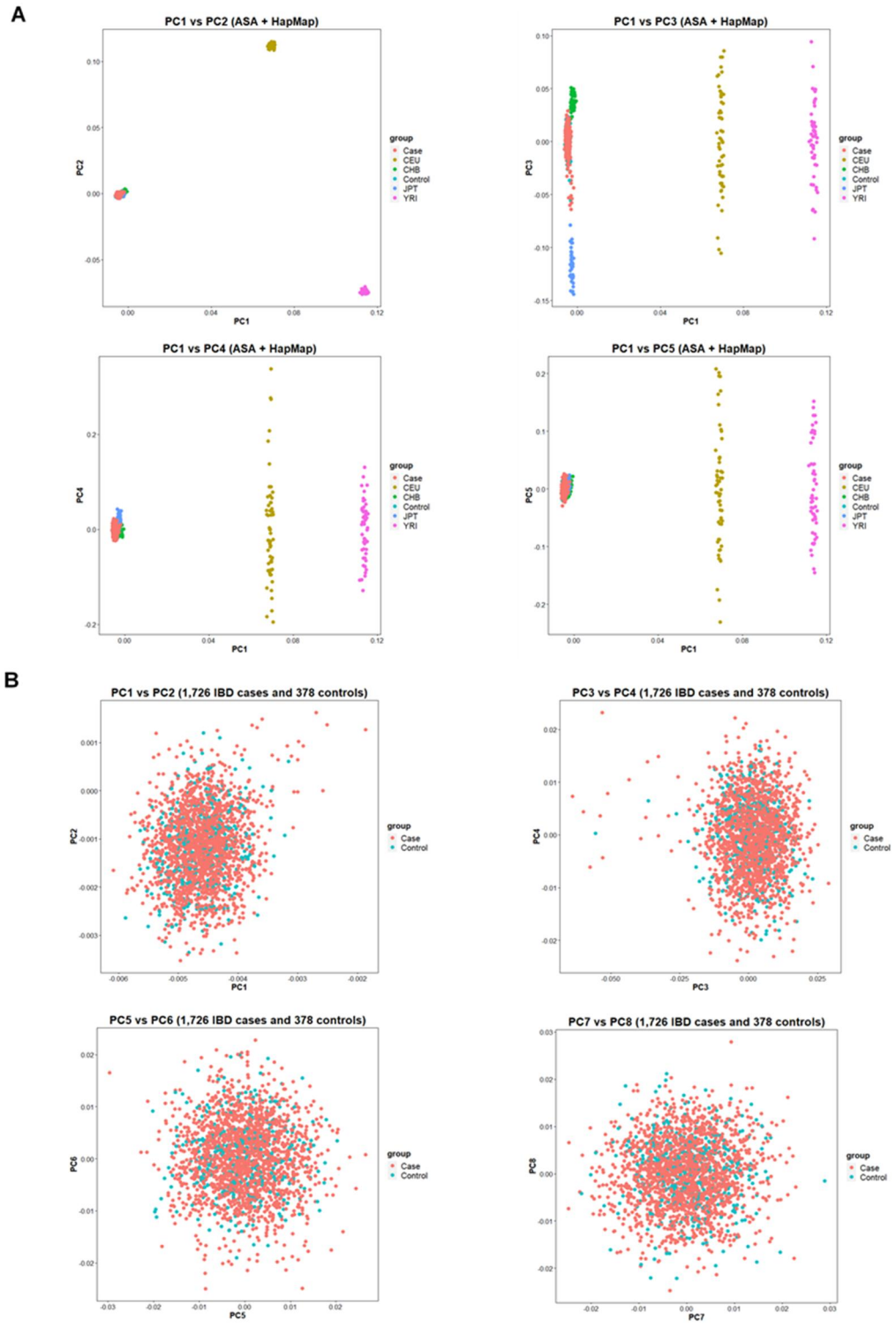

Figure 1. Principal component analysis (PCA) of cohort I. (A) PCA of 2,104 samples and 194 reference DNA samples from HapMap. (B) Plots of the first 8 components from the PCA using 2,104 samples (1,726 cases, 378 controls).

Table 3. Quality control measures of cohort I

		Samples		SNPs
		Cases (CD / UC)	Controls	
Initial counts		1,746 (734 / 1,012)	384	659,183
Pre-QC:	Gender mis-matched samples	8 (3 / 5)	3	
Successfully genotyped		1,738 (731 / 1,007)	381	659,183
SNPs exclusion criteria:	Non-autosomal SNPs			33,446
	In/del SNPs			8,500
	Duplicated SNPs			2,737
	SNP call rate < 98%			19,180
	MAF < 0.01			137,518
	P< 1E-05 for controls, *P*< 5E-08 for cases in HWE			445
Remaining SNPs				457,358
Samples exclusion criteria:	Sample call rate < 96%	2 (1 / 1)	3	
	IBD PI-HAT > 0.2, IBS > 0.8	9 (4 / 5)		
QCed individual data		1,727 (726 / 1,001)	378	
	P< 1E-5 in differential missingness			85
	PCA	1 (1 / 0)		
Final QCed data		1,726 (725 / 1,001)	378	457,272

CD, Crohn's disease; HWE, Hardy-Weinberg equilibrium; IBD, identity-by-descent; MAF, minor allele frequency; PCA, Principal component analysis; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

Table 4. Quality control measures of cohort II

	IBD case1 529 CD/398 UC		IBD case2 391 CD/198 UC		Control 1		Control 2	
	Samples	**SNPs**	**Samples**	**SNPs**	**Samples**	**SNPs**	**Samples**	**SNPs**
Platform	OmniExpress		OmniExpressExome		Omni1-Quad		Omni1-Quad	
Successfully genotyped	927	730,508	589	949,480	800	1,010,300	3351	1,009,702
SNPs exclusion criteria:								
Non-autosomal SNPs		20,415		23,527		26,305		26,029
In/dels		0		136		352		442
SNP call rate < 98%		9,730		0		5,075		4,326
MAF < 0.01		118,701		321,230		251,067		250,744
P< 1E-05 for controls, *P*< 5E-08 for cases in HWE		477		95		222		1,739
Remaining SNPs		581,185		604,492		727,279		726,422
Samples exclusion criteria:								
Sample genotype call rate < 96 %	0		0		0		0	
IBD PI_HAT > 0.2, IBS > 0.8	14		1		51		40	
Duplicated samples	2							
QCed individual data	911	581,185	588	604,492	749	727,279	3311	726,422
Merged data	1,499 cases / 4,060 controls, 531,045 SNPs							
	Samples					SNPs		
SNP call rate < 98%						0		
MAF < 0.01						0		
P< 1E-05 for controls, *P*< 5E-08 for cases in HWE						61		
Sample genotype call rate < 96 %	0							
IBD PI_HAT > 0.2, IBS > 0.8	7 cases, 16 controls							
P< 1E-5 in differential missingness	8,542							
Build-up fail (hg18 to hg19)	64							
PCA	1 case, 3 controls							
Final QCed data	**1,491 (909 CD/ 582 UC) cases/4,041controls-522,378 SNPs**							

CD, Crohn's disease; HWE, Hardy-Weinberg equilibrium; IBD, identity-by-descent; MAF, minor allele frequency; PCA, Principal component analysis; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

9 UC cases) between the two discovery cohorts were removed from cohort II.

2.3. Imputation

We performed pre-phasing using SHAPEIT v2[32] for estimation of haplotypes from genotype data in each cohort. And then, we imputed missing genotypes based on the observed genotypes, estimated haplotypes from pre-phasing, and the multi-ethnic 1000 Genomes Project reference panel release v5 (https://www.international genome.org/) using IMPUTE version 2.3.2.[33] For the QC of imputed SNPs, we removed all imputed SNPs with info score < 0.8, posterior probability score < 0.8, missing rate > 10%, MAF < 1% or HWE test $P < 1 \times 10^{-5}$ for controls and 5×10^{-8} for cases. A total of 6,139,980 imputed SNPs passed the QC criteria and were combined with 457,272 genotyped SNPs for association analysis in cohort I. For cohort II, a total of 6,088,678 imputed SNPs passed the QC criteria and were combined with 522,285 genotyped SNPs for association analysis. After imputation of each dataset, a total of 6,193,769 SNPs are shared between the two cohorts.

2.4. Statistical analysis

We performed association tests for IBD, CD, and UC in each cohort using SNPTEST v2.5.2 (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html)[34] based on the additive model of frequentist association test. A quantile-quantile plot was generated using R 3.5.0 to evaluate the overall significance of the genome-wide associations and the potential impact of population stratification. The impact of population stratification was also evaluated by calculating the genomic control inflation factor (λ_{GC}). As the polygenic architecture and linkage disequilibrium (LD) with true causal variants can influence λ_{GC}, we also evaluated λ_{GC} after stringent LD pruning ($r^2 < 0.1$). In addition, we used the recently developed LD score regression (LDSC) approach,[35] which gives an equivalent correction factor to λ_{GC} after accounting for the polygenic architecture. A Manhattan plot was generated with $-\log_{10}P$ values using R 3.5.0. Conditional analysis was performed to assess whether candidate novel signals were due to long-range LD with variants in previously reported loci. For all variants in candidate loci that were less than 3 Mb away from a known locus, conditional analysis was performed on each of the three datasets separately followed by a meta-analysis or on the combined dataset. Secondary SNPs with

conditional $P < 5 \times 10^{-8}$ were assumed to be independent from the reported lead SNP in the region.

2.5. Fixed-effects meta-analysis

To identify novel susceptibility loci for IBD, CD, and UC, the association results of Asian Screening Array and previously published GWAS were combined using the inverse-variance method based on a fixed-effects model as implemented in meta v1.7.[36] SNPs with $P_{meta} < 1 \times 10^{-6}$ were selected for replication in an independent cohort consisting of 1,088 individuals with IBD (582 CD, 506 UC) and 845 healthy controls. Genotyping of the replication cohort was performed using TaqMan genotyping assay with the Applied Biosystems 7900HT Fast Real-Time PCR System according to the manufacturer's instructions. Between-study heterogeneity was quantified using the I^2 heterogeneity score, and the statistical significance was assessed using the Q test statistic. All SNPs with heterogeneity $P < 0.05$ were excluded to consider possible heterogeneity across studies. For the fixed-effects model, the significance was defined as $P_{combined} < 5 \times 10^{-8}$. We classified the association signals into 4 categories by using the same approach applied by Jostins et al.[14] Four multinomial logistic regression models with parameters βCD and βUC were fitted with the following constraints: (1) CD-specific model: βUC = 0, βCD fitted by maximum likelihood; (2) UC-specific model: βCD = 0, βUC fitted by maximum likelihood; (3) IBD unsaturated (same effect size) model: βCD = βUC = βIBD, βIBD fitted by maximum likelihood; and (4) IBD saturated (different effect size) model: βCD and βUC both fitted by maximum likelihood. The log likelihoods were calculated for each model, and 3 likelihood-ratio tests were conducted by comparing models 1–3 against the IBD saturated model. If all 3 tests showed $P < 0.05$, then the single-nucleotide polymorphism was classified as associated with both CD and UC, but with evidence of different effect sizes. Otherwise, of the 3 constrained models, the SNP was classified based on the model with the largest likelihood. If IBD unsaturated was the best-fitting model, the locus could be interpreted as being associated with both CD and UC without evidence of different effect sizes.

2.6. Fine-mapping

To determine a set of causal SNPs, fine-mapping analysis was carried out using 'FM-summary' (https://github.com/hailianghuang/FM-summary/blob/master/getCredible.r) based on

summary statistics from the meta-analysis of cohort and and the LD reference of East Asians (JPT + CHB) in the 1000 Genomes Project reference panel. The 95% credible set in each locus was defined as the minimum list of SNPs with posterior probability (PP) > 95% in the fine-mapping analysis.

2.7. RNA sequencing using whole blood of 101 CD patients

Total RNA was isolated from the peripheral blood of 101 CD patients in cohort II using PAXgene Blood RNA system (PreAnalytiX, QIAGEN, Germany). The clinical characteristics of the 101 CD patients are shown in Table 5. Whole blood was taken and immediately store in a PAXgene Blood RNA tube at room temperature for > 4h. The total RNA was extracted using the PAXgene Blood RNA kit, following the manufacturer's instructions. RNA quality and quantity were checked using a 2100 Bioanalyzer (Agilent Technologies, CA, USA) and the samples with an RNA integrity number ≥ 7 were deep-sequenced. Sequencing libraries were prepared with the Illumina TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero^{TM} Globin (Illumina, CA, USA) and paired-end RNA sequencing of 101 bp reads was performed using Illumina HiSeq 2500 platform.

We evaluated the number and quality of the total reads, GC percent and adapters in the raw fastq files using FastQC v0.11.7 (https://www.bioinformatics.babraham.ac.uk/ projects/fastqc/). Every one of the 101 samples had > 93 million raw reads (117 million reads on average) and passed the read quality check (average Phred quality score > 36), and GC percent check (46 − 56%). The adapter sequences and reads with low quality were excluded using cutadapt[37] applying the quality Phred score cut-off < 33 and read length cut-off < 20 bp. We performed alignment for the trimmed reads using STAR[38] and GRCh37 reference genome in GENCODE release 19 (https://www.gencodegenes.org/human/release_19.html).[39] For confirmation of unique mapping rate and ribosomal RNA rate, we used RNA-SeQC.[40] As 15 CD patients showed high ribosomal RNA ratio (> 40%) in sample QC, the RNA sequencing of these 15 samples were repeated and aligned to the reference genome. We confirmed high unique mapping rate (94−99%) in all the 101 samples. After alignment, we used RNA-SeQC to estimate the transcript abundance, expected read counts, and transcripts per million reads (TPM) for each gene by selecting the uniquely mapped reads with a mapping quality > 255, and ≤6 mismatched bases to the reference genome.

Table 5. Clinical characteristics of 101 CD patients

	No. of samples
CD patients	101
Male (%)	62 (61.4)
Mean age at diagnosis (yr)	24.2 ± 7.6
Age group at diagnosis (%)	
17~40	56 (55.4)
> 40	45 (44.6)
Location, no. (%)	
Ileum	18 (17.8)
Colon	7 (6.9)
Ileocolon	76(75.2)
Behavior, no. (%)	
Inflammatory	21 (20.8)
Stricturing	24 (23.8)
Penetrating	56 (55.4)
Perianal fistula, no. (%)	
No	41 (40.6)
Yes	60 (59.4)

CD, Crohn's disease

2.8. Korean CD-specific eQTL analysis

After removing low-expressed genes from the mRNA expression data estimated by RNA-SeQC, 21,718 genes with TPM > 0.1 and the number of reads > 6 in ≥ 20% of the 101 CD samples were included. We generated a multi-dimensional scaling (MDS) plot using an R package: edgeR (http://bioconductor.org/packages/release/bioc/html/edgeR.html)[41] using the read count data of the 21,718 genes from the 101 samples to confirm absence of batch effect in the RNA sequencing data from 15 samples that had been re-sequenced due to high rRNA contamination (Figure 2). We used a trimmed mean of M-values (TMM) for normalization of gene expression values considering total mRNA read counts of each sample using edgeR.[41] The genomic input data of the 101 CD patients included a total of 6,451,113 SNPs from GWAS data of cohort II. The cis window was defined as the 1-Mb region up- and downstream of the transcription start site (TSS). Dosage was used for the association analysis for imputed SNPs. Nominal *P* values were calculated for each SNP-gene pair with FastQTL[42] using the linear regression model with 27 covariates including 15 PEER[43] factors, 3 PCs calculated using GWAS dataset of 101 CD samples, repeat or not, gender, age, age of diagnosis, follow-up year, family history, smoking or not, Montreal classification, and disease behaviors. Significance of the top associated variant per gene was estimated by adaptive permutation with the setting '--permute 1000 10000' in FastQTL. The beta distribution-adjusted empirical *P* values were used to calculate the q-values and false discovery rate (FDR) thresholds of each gene using R package: qvalue (https://github.com/StoreyLab/qvalue). The FDR threshold of < 0.05 was applied to identify all the significant cis-eQTL in the whole blood tissue of 101 CD patients.

2.9. Enrichment analysis on eGenes from Korean CD-specific eQTL analysis

To annotate the biological mechanisms related to the eGenes in the eQTL analysis of the 101 CD patients, we performed the Gene Ontology (GO)[44,45] enrichment analysis in the web application, AmiGO2 (http://amigo.geneontology.org/amigo)[46] using 3,816 eGenes with FDR < 0.05 in the cis-eQTL analysis. By the default setting (GO aspect: biological process, Species: Homo sapiens), the result page showed the over- or underrepresented GO terms with significant *P* values.

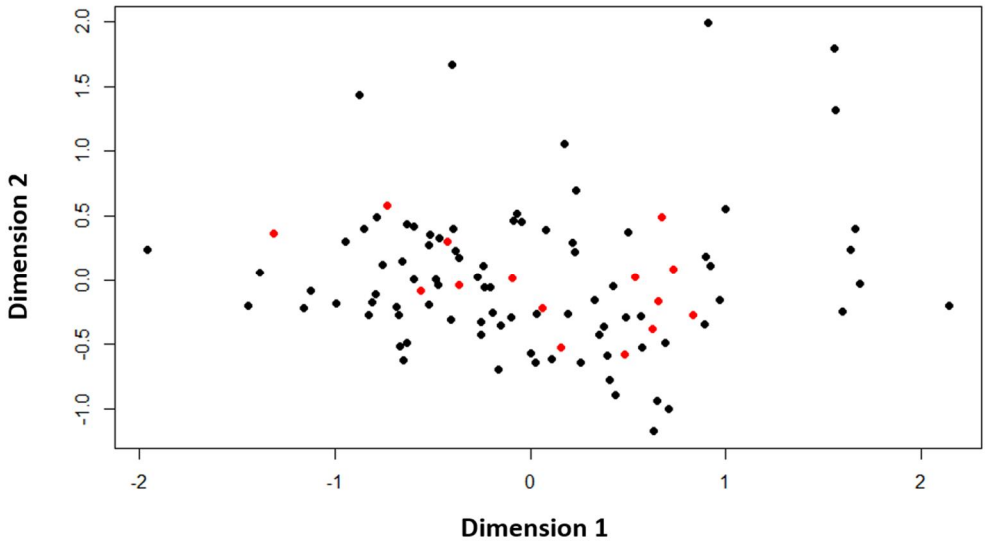

Figure 2. A multi-dimensional scaling (MDS) plot of the read count data of 21,718 high-expressed genes in 101 samples. Red circles in the MDS plot represent 15 repeated samples. Distances on the plot correspond to root mean square average of the largest $\log_2$(fold change) between each pair of samples.

2.10. Comparisons of direction of allelic effects between cis-eQTL databases

We compared the allelic directions of SNP-gene associations shared among the Korean CD cis-eQTL, the existing whole blood cis-eQTL databases of Japanese (105 healthy individuals) (https://humandbs.biosciencedbc.jp/en/hum0099-v1#hum0099.v1.eqtl.v1),[47] and GTEx V7 (369 individuals) (https://gtexportal.org/home/datasets).[48] The number of cis-eQTL in the Korean CD, Japanese, and GTEx datasets was 135,164, 335,813 and 1,052,542, respectively. Using only significant cis-eQTLs with q value $\leq$ 0.05 in each dataset, we compared the slope of the overlapping SNP-gene associations between three pairs of Korean CD-Japanese, Korean CD-GTEx, and GTEx-Japanese datasets.

2.11. Colocalization analysis

To estimate the probability of colocalization between the lead SNP of the GWAS meta-analysis of two cohorts and whole blood eQTL data of 101 CD patients in Koreans, we applied eCAVIAR[49] to estimate the probability of eQTL and GWAS sharing the same causal variants. The eCAVIAR calculated co-localization posterior probability (CLPP) score, indicating the level of colocalization, using each Z score of eQTL and GWAS data, as well as linkage disequilibrium (LD) information. We used the LD reference of East Asians (JPT + CHB) in the 1000 genomes (https://www.international genome.org/). We also tried Japanese eQTL and GTEx eQTL datasets for colocalization analysis. For GTEx, LD reference of European (CEU + FIN + GBR + IBS + TSI) in the 1000 genomes was used since eCAVIAR allows different LD structures for eQTL and GWAS datasets. We selected 100 SNPs upstream and downstream of the lead SNPs in the susceptibility loci (excluding the major histocompatibility complex region, 25 ~ 34 Mb) to calculate the CLPP score. We used the default of two causal variants for locus and eCAVIAR method's recommended significant cut-off, co-localization posterior probability (CLPP) > 0.01, and 0.95 for total credible set posterior probability.

2.12. eQTL and bioinformatics analysis

To gain insight into the potential functional roles of the novel loci, we performed *cis*-eQTL analysis extensively by searching publicly available data from the eQTL Blood Browser,[50] Genotype-Tissue Expression (GTEx) database,[48] and whole blood cis-eQTL databases for

Japanese.[47] Whole blood, small intestine, transverse colon, and sigmoid colon data were selected in the GTEx browser for the analysis. To explore the epigenetic profiles of susceptibility loci, ENCODE[51] histone modification data, HaploReg v4.1,[52] and Regulome DB[53] were used to examine whether any of the SNPs or their proxies ($r^2 \geq 0.8$ in the 1000 genomes of JPT+CHB reference panel) were annotated as regulatory variants. Identified loci were examined for previous implications in other autoimmune or immune-related phenotypes using the Ensembl,[54] UCSC Genome Browser,[55] GeneCards (https://www.genecards.org/), and GWAS Catalog[56] databases. When the SNP was not directly typed, a proxy SNP was used ($r^2 \geq 0.8$).

2.13. Gene annotation

We performed gene analysis using Multi-marker Analysis of GenoMic Annotation (MAGMA) v.1.07b (http://ctg.cncr.nl/software/magma)[57] to prioritize causal genes at susceptibility loci for IBD, CD and UC. By using the summary statistics from the meta-analysis of cohort and , and LD information of East Asian population as input, all SNPs located between the transcription start and end sites were aggregate to that gene to calculate the gene *P* value based on a multiple regression model. Of 19,257 reference genes, 17,371 genes for IBD, 17,361 genes for CD, and 17,396 genes for UC were included in the gene analysis. By applying the threshold of Bonferroni correction, we annotated 29 genes with $P < 2.88 \times 10^{-6}$ (0.05/17,371) for IBD, 58 genes with $P < 2.88 \times 10^{-6}$ (0.05/17,361) for CD, and 39 genes with $P < 2.87 \times 10^{-6}$ (0.05/17,396) for UC.

2.14. Pathway analysis

To identify biological pathways associated with annotated genes for IBD, CD, and UC, we performed gene-set analysis using MAGMA v.1.07b.[57] The analysis results were used as input data. We used the gene sets of 9,976 Gene Ontology pathways from MSigDB v.7.0[58] to calculate the *P* value of each pathway. We set the statistical significance at Bonferroni corrected $P < 5.01 \times 10^{-6}$ (0.05/9,976). We also performed pathway analysis using the previously published summary statistics for a European IBD dataset.[16] The European dataset was comprised of 12,194 cases and 28,072 controls for CD and 12,366 cases and 33,609 controls for UC.

2.15. Polygenic risk scores

We performed polygenic risk score (PRS) analysis using PRSice-2.[59] A PRS estimates an individual's genetic liability to disease based on genotype profile and relevant GWAS data. PRSs are calculated by summing risk alleles, which are weighted by effect sizes derived from GWAS results. To avoid overfitting, we used one of the two cohorts as the base data for estimating effect sizes and the other as the target data for evaluating PRS. Specifically, we calculated the PRSs in the cohort I newly genotyped by ASA based on the effect sizes estimated from the Korean GWAS (PRS_{KOR})[25] of cohort II or the European ancestry IBD GWAS (PRS_{EUR}).[16] We used a total of 5,601,568 shared SNPs between cohort and the Korean GWAS to calculate the PRS_{KOR}, and a total of 4,391,300 shared SNPs between cohort and European ancestry IBD GWAS to calculate the PRS_{EUR}. For the MHC region (chromosome 6: 25~34 Mb), only the most significant SNP was selected from the Korean or European GWAS. After LD clumping (--clump-kb 250, --clump-p 1.00, and --clump-r2 0.10) using the East Asian (CHB+JPT) or European (CEU + FIN + GBR + IBS + TSI) 1000 Genomes data as a reference panel, 151,164 SNPs in Korean GWAS and 131,117 SNPs in European GWAS remained. We selected LD-clumped SNPs based on thresholds of *P* values (5×10^{-8}, 5×10^{-6}, 5×10^{-4}, 5×10^{-3}, 5×10^{-2}, 0.1, 0.2, 0.5, and 1) from the Korean or European GWAS for the PRS analysis. We then compared the full model (including the PRS) with the null model (with the PRS variable excluded) and estimated the variance explained using Nagelkerke's pseudo-R^2.

3. RESULTS

3.1. Fixed-effects meta-analyses using two GWAS datasets in Koreans

Following the QC and imputation of the cohort I including 1,726 IBD cases (725 CD and 1001 UC) and 378 unrelated healthy controls and cohort II including 1,469 IBD patients (896 CD and 573 UC) and 4,041 controls separately, we performed association tests on each of the IBD, CD, and UC using the additive model of frequentist association test of SNPTEST v2.5.2.[34] In cohort I, the quantile-quantile plots for IBD, CD, and UC appeared normal and the genomic inflation factor (λ_{GC}) of IBD, CD, and UC was decreased to less than 1.04 following LD pruning,

as shown in Figure 3A-C. These results suggest that a slight inflation in λ_{GC} might reflect the polygenic architecture of the disease, rather than population stratification. As shown in the Manhattan plot for IBD, CD, and UC (Figure 4A-C), the association test of cohort I data identified 2 loci (*TNFSF15-TNFSF8*, *MHC*) previously established for CD and 1 locus (*MHC*) previously established for UC with genome-wide significance ($P < 5 \times 10^{-8}$). To maximize the statistical power for identification of novel susceptibility loci for IBD, CD, and UC in the Korean population, we performed fixed-effects meta-analyses of cohort I and cohort II consisted of 7,614 individuals including 3,195 IBD patients (1,621 CD and 1,574 UC) and 4,419 healthy controls using meta v1.7.[36] The meta-analysis showed that the λ_{GC} of IBD (1.036), CD (1.033), and UC (1.044) was decreased to 1.029, 1.026, and 1.029, respectively, after LD score regression (Table 6). A total of 10 previously established loci were confirmed with genome-wide significance ($P_{meta} < 5 \times 10^{-8}$) including *TNFSF15-TNFSF8*, *MHC*, *TNFRSF6B*, *TBC1D1-KLF3*, *GPR35*, *PYGO2-SHC1*, *STAT3-STAT5B-STAT5A*, *NCF4-CSF2RB*, *DUSP5-SMNDC1*, and *ZNF365* in the meta-analysis for IBD (Figure 5A). Following the meta-analyses for CD and UC separately, 1 novel locus (*LOC731275*) and 12 established loci for CD (Figure 5B), and 1 novel locus (*LCOR-SLIT1*) and 4 established loci for UC (Figure 5C) exceeded the genome-wide significance level. To identify novel susceptibility loci for IBD, CD, or UC, we selected 8 additional novel candidate loci (2 loci for IBD, 5 loci for CD, and 1 locus for UC) for the replication study applying a threshold of $P_{meta} < 1 \times 10^{-6}$ (Table 7). We genotyped these 10 lead SNPs from 2 novel and 8 suggestive loci in an independent replication sample consisting of 1,088 individuals with IBD (582 CD and 506 UC) and 845 healthy controls using TaqMan genotyping technology. By combining association results from the meta-analysis and replication study, 3 novel susceptibility loci were identified including 1 UC-specific locus and 2 CD- specific loci: rs76227733 in the *LCOR–SLIT1* region at 10q24 ($P_{combined} = 6.56 \times 10^{-9}$, OR = 1.32) for UC, rs2240751 in the *MFSD12-C19orf71-FZR1-DOHH* region at 19p13 ($P_{combined} = 3.03 \times 10^{-8}$, OR = 1.25) for CD, and rs6936629 in the *RFX6-GPRC6A-FAM162B* region at 6q22 ($P_{combined} = 3.63 \times 10^{-8}$, OR = 1.25) for CD (Table 8 and Figure 6A-C). These 3 SNPs showed consistent association across the three independent samples without any indication of genetic heterogeneity ($P > 0.05$). The 3 loci did not show additional independent genome-wide significant signals following conditional analyses (Figure 7A-C).

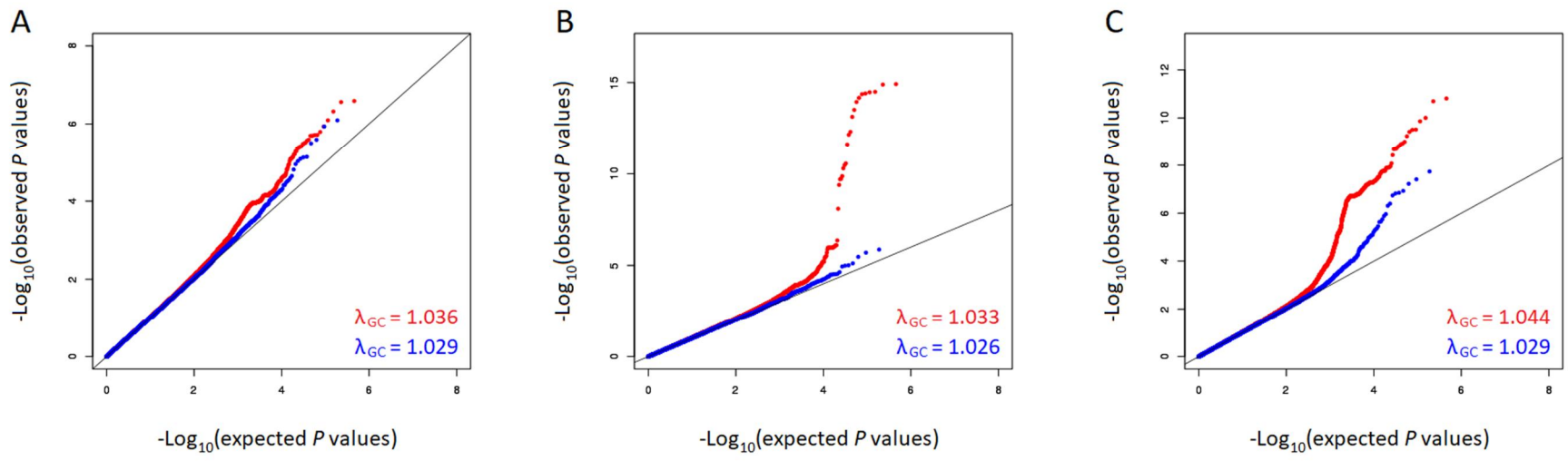

Figure 3. Quantile-quantile plots in cohort I. The $-\log_{10} P$ values of 457,272 genotyped SNPs (red dots) and 188,146 LD-pruned ($r^2 < 0.2$) SNPs (blue dots) were plotted against the expected null distribution. (A) IBD (1,726 cases, 378 controls). (B) CD (725 cases, 378 controls). (C) UC (1,001 cases, 378 controls).

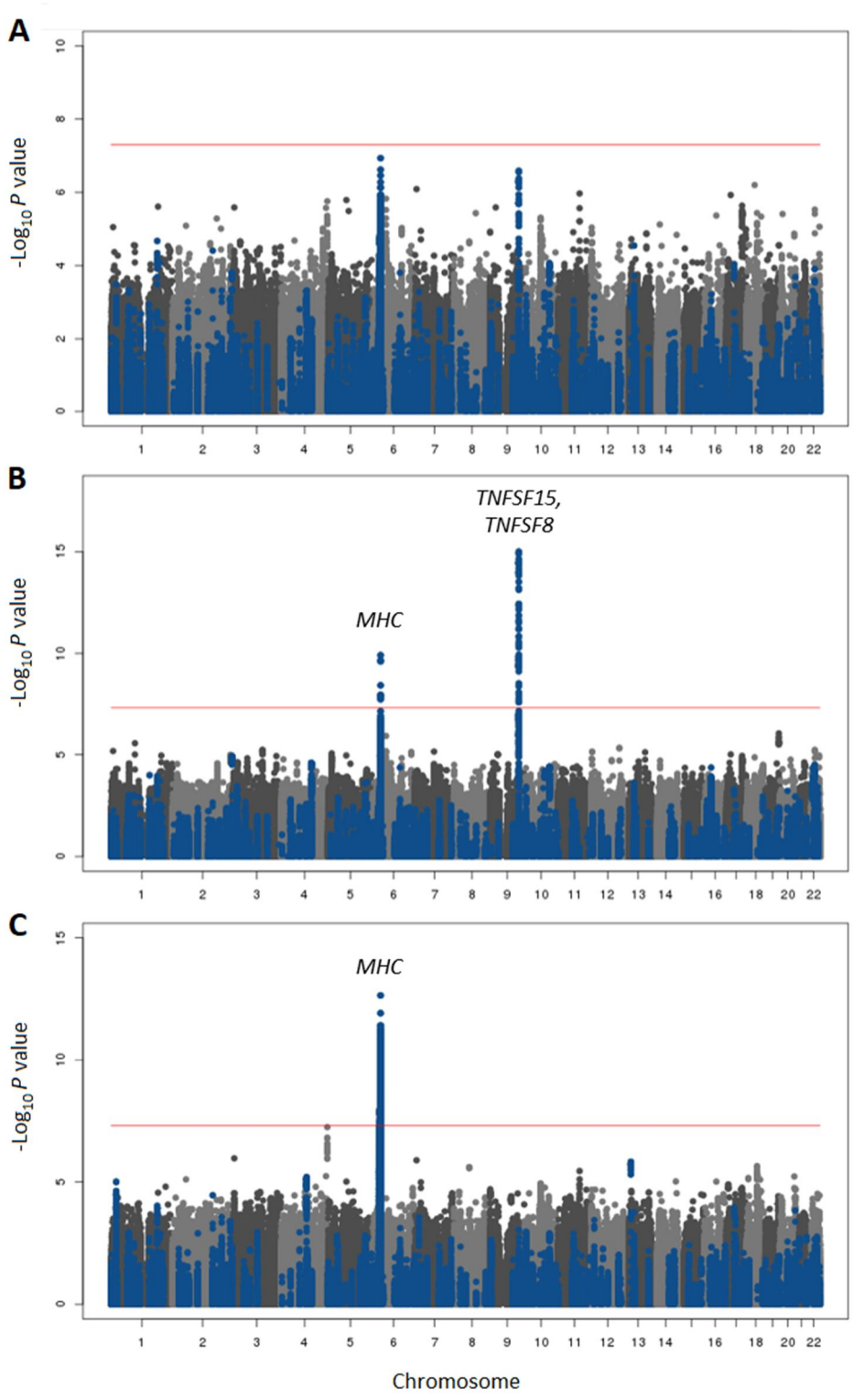

Figure 4. Manhattan plots of (A) IBD, (B) CD, and (C) UC for cohort □. The red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Blue dots indicate SNPs in previously established loci.

Table 6. LD score regression test using the LDSC method

Cohort	SNPs analyzed in IBD/CD/UC	λ_{GC} (λ_{GC} using LDSC)*		
		IBD	CD	UC
I	6,597,252	1.033 (1.000)	1.028 (1.009)	1.035 (0.986)
II#	6,610,963	1.112 (1.029)	1.091 (1.015)	1.055 (1.006)
Combined	5,881,497/5,885,673/5,890,884	1.112 (1.035)	1.106 (1.017)	1.078 (1.010)

CD, Crohn's disease; IBD, inflammatory bowel disease; UC, ulcerative colitis.

*The genomic control inflation factor λ_{GC} was caucluated using R v3.5.0 and LD score regression of LDSC v1.0.0.

#Cohort II: Previous our Korean GWAS data from Yang et al (ref. 25).

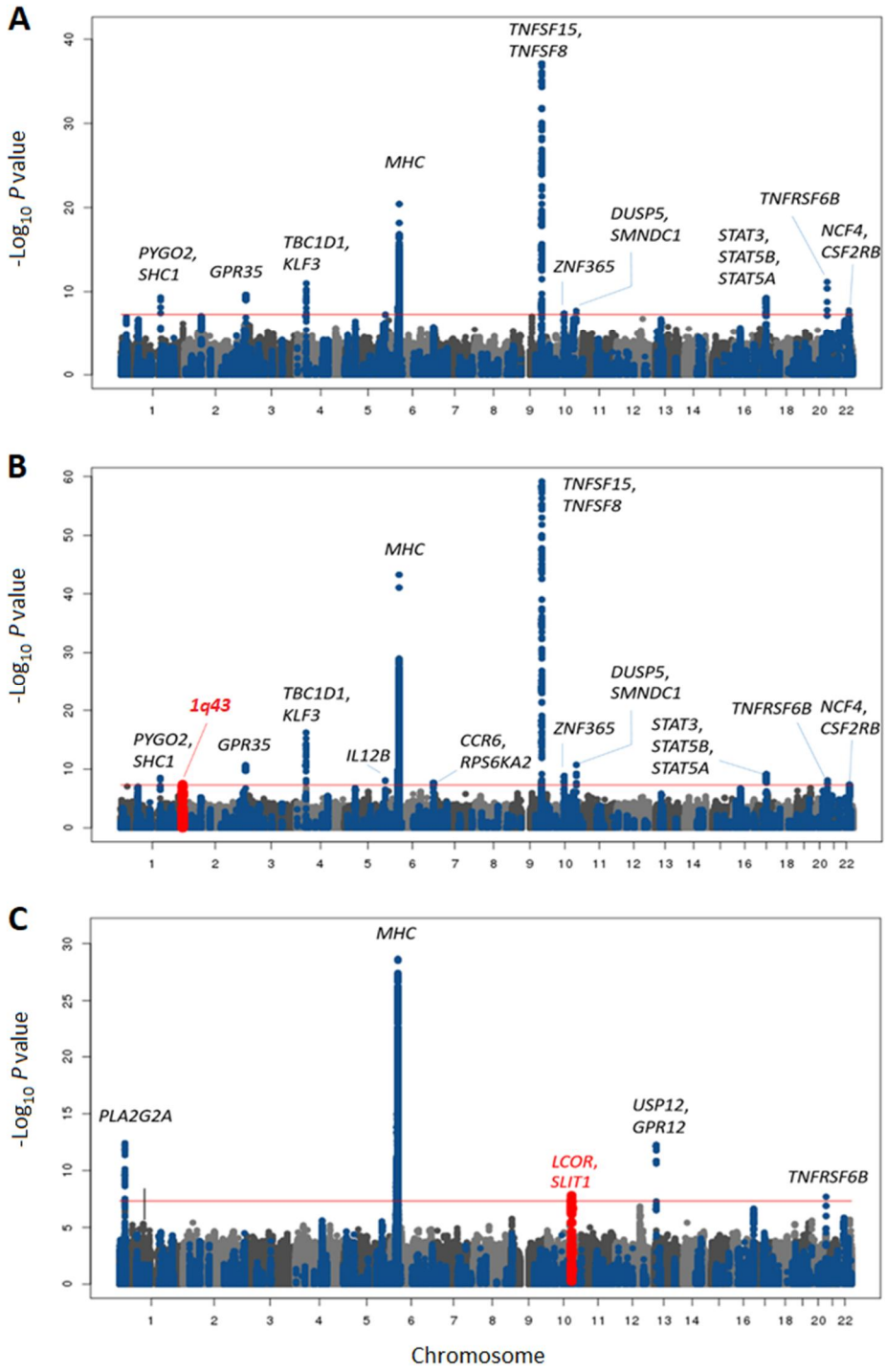

Figure 5. Manhattan plots of (A) IBD, (B) CD, and (C) UC for meta-analysis using cohort □ and □. SNPs located in novel loci are colored red, and those in previously known regions are colored blue. The red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$).

Table 7. Lead SNPs in 10 novel candidate loci with P_{meta} value < 1.00×10^{-6} from fixed-effects meta-analysis of cohort I and II

Phenotype	Locus	SNP	Position (hg19)	Candidate gene(s)	Risk allele	Meta-analysis				Cohort I		Cohort II	
						RAF	OR (95% CI)	P_{meta}*	P_{het}†	OR (95% CI)	P‡	OR (95% CI)	P‡
IBD	9q21	rs2351466	78,922,100	*PCSK5*	G	0.038	1.75 (1.43-2.16)	9.03×10^{-8}	5.28×10^{-1}	1.55 (1.00-2.40)	4.92×10^{-2}	1.82 (1.44-2.28)	5.44×10^{-7}
	12q23	rs1357766	103,164,639	.	A	0.873	1.35 (1.20-1.51)	1.74×10^{-7}	1.59×10^{-1}	1.15 (0.90-1.47)	2.71×10^{-1}	1.40 (1.24-1.58)	1.17×10^{-7}
CD	1q43	rs10754788	243,043,895	.	C	0.225	1.36 (1.22-1.51)	3.92×10^{-8}	6.32×10^{-1}	1.30 (1.04-1.61)	1.95×10^{-2}	1.38 (1.22-1.56)	5.85×10^{-7}
	1p36	rs11249215	25,297,184	*RUNX3*	G	0.458	1.28 (1.17-1.40)	9.07×10^{-8}	1.18×10^{-1}	1.44 (1.21-1.72)	5.12×10^{-5}	1.22 (1.10-1.35)	1.33×10^{-4}
	19q13	rs255773	54,723,546	*RPS9*, *LILRA6*	T	0.500	1.30 (1.18-1.43)	1.64×10^{-7}	1.46×10^{-1}	1.47 (1.21-1.77)	9.02×10^{-5}	1.24 (1.11-1.39)	1.64×10^{-4}
	19p13	rs2240751	3,548,231	*DOHH*, *FZR1*, *C19orf71*, *MFSD12*	G	0.347	1.27 (1.15-1.39)	4.73×10^{-7}	4.96×10^{-1}	1.34 (1.11-1.60)	1.78×10^{-3}	1.24 (1.12-1.38)	6.09×10^{-5}
	6q22	rs6936629	117,239,141	*FAM162B*, *GPRC6A*, *RFX6*	C	0.364	1.26 (1.15-1.38)	6.50×10^{-7}	9.48×10^{-1}	1.27 (1.06-1.52)	9.94×10^{-3}	1.26 (1.13-1.40)	2.08×10^{-5}
	5q14	rs6872414	91,799,986	*LOC105379080*	A	0.075	1.54 (1.30-1.82)	7.08×10^{-7}	2.38×10^{-1}	1.28 (0.91-1.81)	1.62×10^{-1}	1.63 (1.34-1.97)	9.49×10^{-7}
UC	10q24	rs76227733	98,556,649	*LCOR*, *SLIT1*	C	0.307	1.39 (1.24-1.56)	1.62×10^{-8}	4.71×10^{-1}	1.47 (1.22-1.76)	3.91×10^{-5}	1.35 (1.16-1.55)	8.22×10^{-5}
	12q23	rs970332	97,278,178	*NEDD1*	G	0.468	1.31 (1.18-1.45)	1.46×10^{-7}	3.51×10^{-1}	1.23 (1.04-1.45)	1.81×10^{-2}	1.36 (1.20-1.53)	1.69×10^{-6}

CI, confidence interval; hg19, human genome version 19; RAF, risk allele frequency; OR, odds ratio; Position, chromosome position; SNP, single nucleotide polymorphism.

*Fixed-effects meta-analysis *P*.

†*P* value for heterogeneity.

‡Association *P* value of SNPTEST v2.5.2.

Table 8. Three novel susceptiblilty loci for ulcerative colitis or Crohn's diease in Korean population

Phenotype	Locus	SNP	Position (hg19)	Candidate gene(s)	Risk allele	Study	Number of samples		RAF		OR (95% CI)	*P*	P_{het}*
							Case	Control	Case	Control			
UC	10q24	rs76227733	98,556,649	*LCOR, SLIT1*	C	ASA	1,001	378	0.344	0.264	1.47 (1.22-1.76)	3.91×10^{-5}†	
						GWAS	573	4,041	0.348	0.296	1.35 (1.16-1.56)	8.22×10^{-5}†	
						Replication	491	837	0.349	0.311	1.19 (1.01-1.41)	4.35×10^{-2}†	
						Combined	2,065	5,256	0.346	0.296	1.32 (1.20-1.46)	6.56×10^{-9}‡	2.42×10^{-1}
CD	19p13	rs2240751	3,548,231	*MFSD12, C19orf71, FZR1, DOHH*	G	ASA	723	378	0.387	0.319	1.34 (1.11-1.60)	1.78×10^{-3}†	
						GWAS	896	4,041	0.384	0.334	1.24 (1.12-1.38)	6.09×10^{-5}†	
						Replication	461	826	0.377	0.331	1.22 (1.03-1.44)	1.93×10^{-2}†	
						Combined	2,080	5,245	0.384	0.332	1.25 (1.16-1.36)	3.03×10^{-8}‡	7.32×10^{-1}
	6q22	rs6936629	117,239,141	*RFX6, GPRC6A, FAM162B*	C	ASA	723	375	0.399	0.343	1.27 (1.06-1.52)	9.94×10^{-3}†	
						GWAS	896	4,041	0.404	0.351	1.26 (1.13-1.40)	2.08×10^{-5}†	
						Replication	486	826	0.387	0.338	1.21 (1.04-1.42)	1.64×10^{-2}†	
						Combined	2,105	5,242	0.398	0.348	1.25 (1.15-1.35)	3.63×10^{-8}‡	9.14×10^{-1}

CD, Crohn's disease; CI, confidence interval; hg19, human genome version 19; OR, odds ratio; *P*, *P* value; Position, chromosome position; RAF, risk allele frequency; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

*P value for heterogeneity.

†Association P value of SNPTEST v2.5.2.

‡P value of fixed-effects meta-analysis.

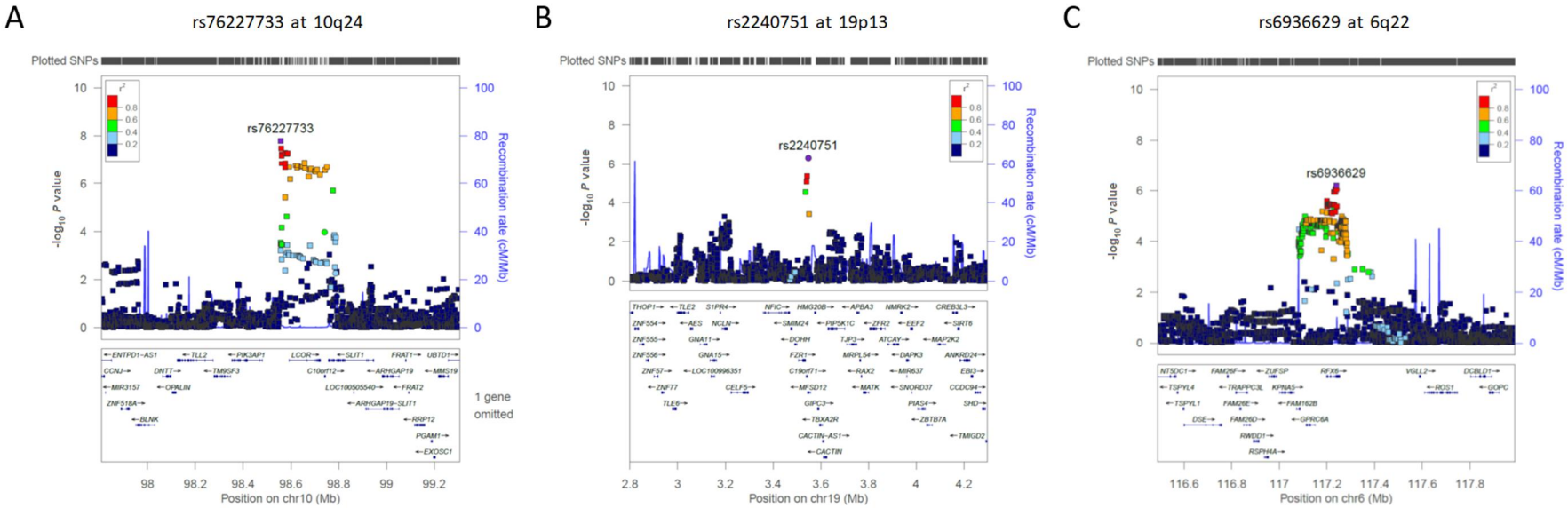

Figure 6. Regional association plots for 3 novel IBD loci. (A) rs76227733 at 10q24. (B) rs2240751 at 19p13. (C) rs6936629 at 6q22. SNPs are plotted according to their chromosomal positions (NCBI Build 37) with $-\log_{10}P$ values from the meta-analysis in the region flanking 750 kb on either side of the marker SNP. Circles indicate genotyped SNPs and squares indicate imputed SNPs. The most strongly associated SNP in the discovery stage is shown as a small purple circle. Linkage disequilibrium (LD; r^2 values) between the lead SNP and other SNPs is indicated using colors. The relative location of the annotated genes and the direction of transcription are shown in the lower portion of the figure. The estimated recombination rates of Asian samples from the 1000 Genomes Project (Nov 2014) are plotted to reflect the local LD structure. Plots are generated using LocusZoom.

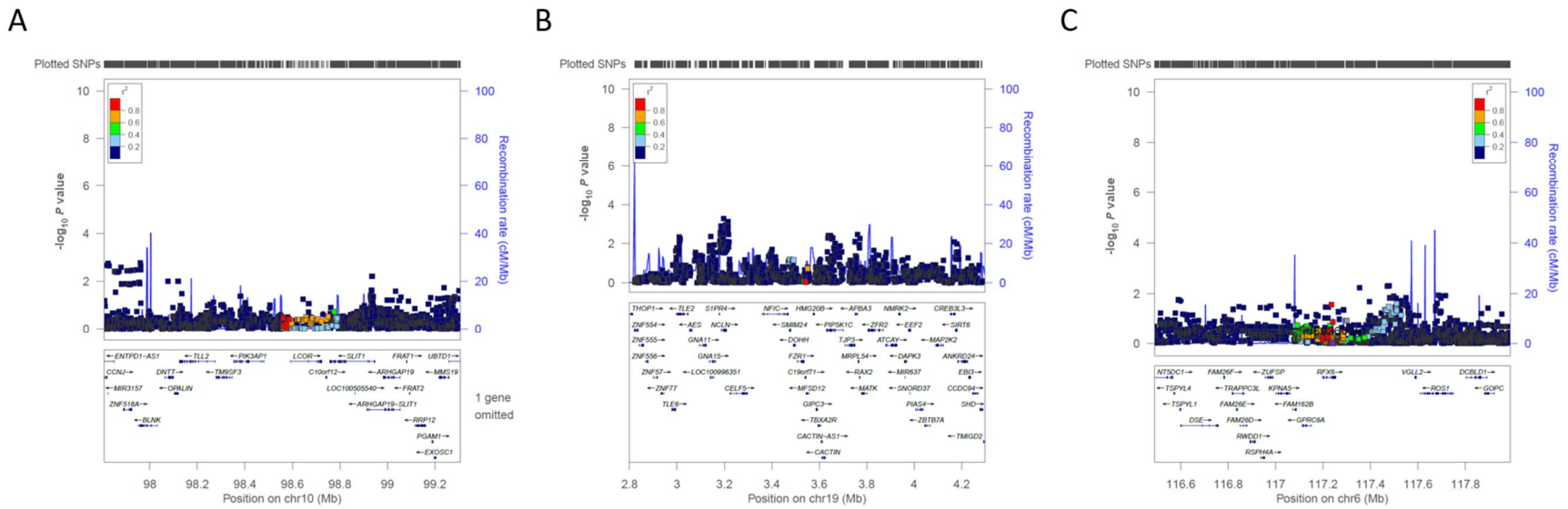

Figure 7. Conditional association signals for 3 novel IBD loci. (A) Conditioned on rs76227733 at 10q24, (B) Conditioned on rs2240751 at 19p13, and (C) Conditioned on rs6936629 at 6q22. SNPs are plotted according to their chromosomal positions (NCBI Build 37) with −log10 *P* values from the meta-analysis after conditioning analyses in the region flanking 750 kb on either side of the lead SNP. Circles indicate genotyped SNPs and squares indicate imputed SNPs. Linkage disequilibrium (LD; r^2 values) between the lead SNP and other SNPs is indicated using colors. The relative location of the annotated genes and the direction of transcription are shown in the lower portion of the figure. The estimated recombination rates of Asian samples from the 1000 Genomes Project (Nov 2014) are plotted to reflect the local LD structure. Plots are generated using LocusZoom.

3.2. Gene prioritization of novel associations

Locus 10q24 for UC

The most significant association at rs76227733 on 10q24 ($P_{combined} = 6.56 \times 10^{-9}$, OR = 1.32) was located 35.4 kb away from 5'-end of *LCOR* (ligand dependent nuclear receptor corepressor), and 201.1 kb away from the 3'-end of *SLIT1* (slit guidance ligand 1) (Table 8 and Figure 6A). rs76227733 was within a LD region of 250.6 kb, which included *LCOR* and *SLIT1*. The 95% credible set at the 10q24 locus consisted of 28 SNPs including rs76227733 (PP = 18.7%) in the fine-mapping analysis (Table 9). Gene analysis using MAGMA v.1.07b[57] identified a significant association with *LCOR* ($P = 9.58 \times 10^{-7}$) (Table 10). *LCOR* is a transcriptional corepressor that can attenuate agonist-activated nuclear receptor signaling through multiple mechanisms.[60] *SLIT1* belongs to the Slit family of proteins implicated in guiding the migration of neurons and leukocytes.[61] rs76227733 did not show cis-eQTL effects in the eQTL Blood Browser,[50] GTEx,[48] Japanese eQTL,[47] and Korean CD-specific eQTL. However, rs57807455, which was in high LD with rs76227733 ($r^2 = 0.99$), showed enhancer histone marks and DNAse I hypersensitivity sites in gastrointestinal tissues, altered TATA box motifs in Haploreg v4.1,[52] and a high regulomeDB[53] score of 2b (Table 11). Analysis of the ENCODE project[51] data showed that rs57807455 was located at the candidate cis-regulatory elements (cCREs) of EH38E1491359 (chromosome 10:96,798,560-96,798,898 bp), a binding site of POLR2A in the transverse colon and Peyer's patch cells (Table 12).

Locus 19p13 for CD

At the 19p13 locus identified in CD meta-analysis, the lead SNP was rs2240751 ($P_{combined} = 3.03 \times 10^{-8}$, OR = 1.25) within a LD region of 66.5 kb, which included *MFSD12* (major facilitator superfamily domain containing 12), *C19orf71* (chromosome 19 open reading frame 71), *FZR1* (fizzy and cell division cycle 20 Related 1), and *DOHH* (deoxyhypusine hydroxylase)(Table 8 and Figure 6B). In the fine-mapping analysis, rs2240751 was the only single variant with greater than 50% certainty (PP=83.4%) within the 95% credible set including 3 SNPs (Table 9). Rs2240751 did not show any cis-eQTL effects in disease-relevant tissues. Rs2240751 was in exon 3 of *MFSD12*, resulting in an amino acid substitution from a polar uncharged tyrosine to a basic

Table 9. List of SNPs in 95% credible set of three novel IBD susceptibility loci

Phenotype	Locus	Lead SNP	No. of SNPs	SNP list in 95% credible set (posterior probability)*
UC	10q24	rs76227733	28	rs76227733 (0.187), rs57807455 (0.096), rs12414968 (0.062), rs12416214 (0.062), rs11188935 (0.059), rs12257954 (0.058), rs10882843 (0.055), rs11188927 (0.048), rs12265203 (0.024), rs12416231 (0.024), rs77379630 (0.023), rs10882852 (0.019), rs7924085 (0.019), rs11188950 (0.019), rs11188964 (0.017), rs11188962 (0.017), rs7088648 (0.017), rs77517784 (0.016), rs10882861 (0.016), rs142885347 (0.016), rs11188952 (0.015), rs10882855 (0.014), rs10786316 (0.014), rs10219031 (0.013), rs3814163 (0.012), rs12251684 (0.012), rs10882856 (0.012), rs7098255 (0.012).
CD	19p13	rs2240751	3	**rs2240751 (0.834)**, rs12608592 (0.096), rs12984831 (0.053).
CD	6q22	rs6936629	277	rs6936629 (0.046), rs1321371 (0.034), rs1406982 (0.034), rs1321372 (0.033), rs9489066 (0.033), rs1321366 (0.032), rs12201912 (0.026), rs35974852 (0.026), rs6927262 (0.026), rs630045 (0.012), rs339326 (0.009), rs339327 (0.009), rs339328 (0.009), rs339331 (0.009), rs654971 (0.009), rs339344 (0.009), rs339350 (0.009), rs610424 (0.009), rs339297 (0.009), rs339334 (0.009), rs339340 (0.009), rs339341 (0.009), rs339301 (0.009), rs339302 (0.009), rs434499 (0.009), rs1358793 (0.008), rs2145173 (0.008), rs6568967 (0.008), rs339351 (0.008), rs339353 (0.007), rs12202378 (0.007), rs12201923 (0.006), rs97457 (0.005), rs339343 (0.004), rs339347 (0.004), rs13199826 (0.004), rs339299 (0.004), rs339300 (0.004), rs645426 (0.003), rs1761875 (0.003), rs2274911 (0.003), rs5024233 (0.002), rs1761877 (0.002), rs1761878 (0.002), rs1512655 (0.002), rs1512657 (0.002), rs339312 (0.002), rs6901971 (0.002), rs88520 (0.002), rs9400968 (0.002), rs339323 (0.002), rs3907920 (0.002), rs9481676 (0.002), rs1512658 (0.002), rs339358 (0.002), rs339359 (0.002), rs339305 (0.002), rs339306 (0.002), rs339365 (0.002), rs2203192 (0.002), rs339309 (0.002), rs339310 (0.002), rs339311 (0.002), rs339314 (0.002), rs339315 (0.002), rs339318 (0.002), rs339319 (0.002), rs7756165 (0.002), rs9400959 (0.002), rs339316 (0.002), rs143357 (0.002), rs339320 (0.002), rs339321 (0.002), rs1631116 (0.002), rs1741682 (0.002), rs1741683 (0.002), rs1761842 (0.002), rs339356 (0.002), rs1406981 (0.002), rs1741688 (0.002), rs5024230 (0.002), rs5024231 (0.002), rs5024232 (0.002), rs7755357 (0.002), rs1741663 (0.002), rs1360755 (0.002), rs1334653 (0.002), rs1334654 (0.002), rs1334656 (0.002), rs1334657 (0.002), rs1334658 (0.002), rs1334661 (0.002), rs1334662 (0.002), rs1334663 (0.002), rs1334665 (0.002), rs1334666 (0.002), rs1406983 (0.002), rs1413731 (0.002), rs1413732 (0.002), rs1413734 (0.002), rs1413735 (0.002), rs1413736 (0.002), rs1413738 (0.002), rs1413739 (0.002), rs1413740 (0.002), rs1413742 (0.002), rs1618412 (0.002), rs1618533 (0.002), rs16849 (0.002), rs17175 (0.002), rs1741652 (0.002), rs1741671 (0.002), rs1741676 (0.002), rs1741677 (0.002), rs1741679 (0.002), rs1741680 (0.002), rs1741681 (0.002), rs1741687 (0.002), rs1761843 (0.002), rs1761845 (0.002), rs1761859 (0.002), rs1761860 (0.002), rs1761865 (0.002), rs1761867 (0.002), rs1970175 (0.002), rs2210714 (0.002), rs768581 (0.002), rs768582 (0.002), rs7755339 (0.002), rs1334667 (0.002), rs1334669 (0.002), rs1741655 (0.002), rs1741656 (0.002), rs1741658 (0.002), rs1741660 (0.002), rs1741661 (0.002), rs1741662 (0.002), rs1741684 (0.002), rs1761861 (0.002), rs1761874 (0.002), rs1761876 (0.002), rs56627688 (0.002), rs1413741 (0.002), rs1761844 (0.002), rs1406984 (0.002), rs1406985 (0.002), rs1761841 (0.002), rs1741659 (0.002), rs1334659 (0.002), rs1761879 (0.002), rs339304 (0.002), rs1741674 (0.002), rs1761863 (0.002), rs1761864 (0.002), rs615199 (0.002), rs6929458 (0.002), rs9372473 (0.002), rs1741675 (0.002), rs1761880 (0.002), rs9400969 (0.002), rs1631199 (0.002), rs682726 (0.002), rs9400970 (0.002), rs1632019 (0.002), rs2353358 (0.002), rs2750416 (0.002), rs6901250 (0.002), rs6924002 (0.002), rs149641179 (0.002), rs168127 (0.002), rs2782298 (0.002), rs9374627 (0.002), rs993394 (0.002), rs600928 (0.002), rs6907088 (0.002), rs7761566 (0.002), rs7761872 (0.002), rs9400975 (0.002), rs1761881 (0.002), rs662657 (0.002), rs7760125 (0.002), rs9400964 (0.002), rs678230 (0.002), rs1606366 (0.002), rs12211764 (0.002), rs55915982 (0.002), rs9400962 (0.002), rs142100674 (0.002), rs201775380 (0.002), rs665401 (0.002), rs200816531 (0.002), rs675495 (0.002), rs9374624 (0.002), rs6938235 (0.002), rs7740481 (0.002), rs63749614 (0.002), rs674621 (0.002), rs585957 (0.002), rs2750417 (0.002), rs2750418 (0.002), rs1084813 (0.002), rs2175622 (0.002), rs584917 (0.001), rs168128 (0.001), rs587174 (0.001), rs596616 (0.001), rs607372 (0.001), rs339360 (0.001), rs339361 (0.001), rs633898 (0.001), rs339362 (0.001), rs561114 (0.001), rs632159 (0.001), rs594785 (0.001), rs595698 (0.001), rs596732 (0.001), rs610979 (0.001), rs611349 (0.001), rs616347 (0.001), rs617426 (0.001), rs619307 (0.001), rs625821 (0.001), rs627551 (0.001), rs630434 (0.001), rs630458 (0.001), rs630695 (0.001), rs631089 (0.001), rs639170 (0.001), rs639646 (0.001), rs643550 (0.001), rs643943 (0.001), rs645745 (0.001), rs657963 (0.001), rs661894 (0.001), rs673055 (0.001), rs673065 (0.001), rs675233 (0.001), rs675266 (0.001), rs675811 (0.001), rs676152 (0.001), rs688949 (0.001), rs75414267 (0.001), rs673906 (0.001), rs597688 (0.001), rs143356 (0.001), rs339368 (0.001), rs339303 (0.001), rs339307 (0.001), rs631642 (0.001), rs339313 (0.001), rs339317 (0.001), rs35565998 (0.001), rs7764347 (0.001), rs636252 (0.001), rs9320588 (0.001), rs4946205 (0.001), rs9489067 (0.001), rs4946204 (0.001), rs7770158 (0.001), rs7774506 (0.001), rs9320585 (0.001), rs9320586 (0.001), rs984258 (0.001), rs966900 (0.001), rs339322 (0.001), rs9384991 (0.001), rs9387439 (0.001), rs339324 (0.001), rs587637 (0.001), rs4946206 (0.001), rs664846 (0.001), rs9400976 (0.001), rs614922 (0.001), rs614924 (0.001), rs1606365 (0.001), rs615850 (0.001), rs339357 (0.001).

CD, Crohn's disease; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

*Posterior probability was estimated using FM-summary (https://github.com/hailianghuang/FM-summary/blob/master/getCredible.r).

Bold: SNPs with posterior probability > 0.5.

Table 10. Annotated genes from gene analysis using MAGMA

IBD			CD			UC		
Gene	Chr:position (hg19)	*P*	Gene	Chr:position (hg19)	*P*	Gene	Chr:position (hg19)	*P*
TNFSF15	9:117,546,915 -117,568,408	1.16E-35	*TNFSF15*	9:117,546,915 -117,568,408	6.13E-54	*HLA-DRB1*	6:32,546,546 -32,557,613	2.46E-17
HLA-DPA1	6:33,032,346 -33,048,555	5.60E-13	*TNFSF8*	9:117,655,623 -117,692,875	2.40E-15	*HLA-DQB1*	6:32,627,241 -32,634,466	3.10E-16
ITPR3	6:33,587,951 -33,664,351	7.33E-12	*HLA-DRB1*	6:32,546,546 -32,557,613	2.42E-15	*AGER*	6:32,148,745 -32,152,099	3.69E-14
TNFSF8	9:117,655,623 -117,692,875	1.76E-11	*POU5F1*	6:31,132,114 -31,138,451	1.38E-12	*NOTCH4*	6:32,162,620 -32,191,844	3.11E-13
HLA-DRB1	6:32,546,546 -32,557,613	4.12E-11	*TCF19*	6:31,126,303 -31,134,183	1.50E-12	*MICB*	6:31,462,054 -31,478,901	1.73E-12
HLA-DQB1	6:32,627,241 -32,634,466	7.14E-11	*HLA-DPA1*	6:33,032,346 -33,048,555	1.80E-12	*PPT2*	6:32,121,229 -32,131,458	6.09E-12
BTNL2	6:32,362,513 -32,374,900	1.52E-09	*ITPR3*	6:33,587,951 -33,664,351	1.61E-11	*OTUD3*	1:20,208,356 -20,239,438	5.80E-11
HLA-DQA2	6:32,709,156 -32,714,664	1.76E-09	*CCHCR1*	6:31,110,216 -31,126,015	3.00E-11	*OR11A1*	6:29,393,281 -29,424,848	5.83E-11
MCCD1	6:31,496,739 -31,498,008	3.41E-09	*HLA-DPB1*	6:33,043,703 -33,057,473	3.40E-11	*NFKBIL1*	6:31,514,628 -31,526,606	1.71E-10
HLA-DPB1	6:33,043,703 -33,057,473	6.44E-09	*HLA-DQA2*	6:32,709,156 -32,714,664	3.74E-11	*DDX39B*	6:31,497,996 -31,510,252	2.15E-10
AGER	6:32,148,745 -32,152,099	1.22E-08	*DDX39B*	6:31,497,996 -31,510,252	9.25E-11	*OR12D3*	6:29,341,200 -29,343,068	1.03E-09
FKBPL	6:32,096,484 -32,098,067	2.61E-08	*ATP6V1G2*	6:31,512,228 -31,514,625	1.26E-10	*HLA-DRA*	6:32,407,619 -32,412,823	1.42E-09
HLA-G	6:29,794,756 -29,798,899	2.70E-08	*HLA-DQB2*	6:32,723,837 -32,731,330	1.89E-10	*OR10C1*	6:29,407,716 -29,408,754	2.28E-09
NOTCH4	6:32,162,620 -32,191,844	4.82E-08	*TAP2*	6:32,789,610 -32,806,547	1.98E-10	*HLA-DQA2*	6:32,709,156 -32,714,664	3.01E-09
STAT3	17:40,465,342 -40,540,586	5.67E-08	*NOTCH4*	6:32,162,620 -32,191,844	2.43E-10	*ITPR3*	6:33,587,951 -33,664,351	3.87E-09
LOC101929163	6:32,371,234 -32,373,967	5.90E-08	*EHMT2*	6:31,847,536 -31,865,464	1.13E-09	*SCGN*	6:25,652,429 -25,702,011	5.02E-09
PSORS1C1	6:31,082,608 -31,107,869	7.50E-08	*SLC44A4*	6:31,830,969 -31,846,823	1.45E-09	*PLA2G2E*	1:20,246,800 -20,250,110	8.28E-09
SLC39A7	6:33,168,603 -33,172,214	1.53E-07	*GPR35*	2:241,544,825 -241,570,676	2.03E-09	*HLA-DPA1*	6:33,032,346 -33,048,555	1.01E-08
NCR3	6:31,556,660 -31,560,762	1.59E-07	*MCCD1*	6:31,496,739 -31,498,008	2.37E-09	*COL11A2*	6:33,130,469 -33,160,245	3.24E-08
GPR35	2:241,544,825 -241,570,676	1.96E-07	*FKBPL*	6:32,096,484 -32,098,067	2.38E-09	*LST1*	6:31,553,956 -31,556,686	4.85E-08
HLA-DRA	6:32,407,619 -32,412,823	4.78E-07	*HLA-DQA1*	6:32,605,169 -32,612,152	2.96E-09	*BTNL2*	6:32,362,513 -32,374,900	6.13E-08
GPSM3	6:32,158,543 -32,163,300	5.74E-07	*HLA-DRA*	6:32,407,619 -32,412,823	2.97E-09	*TAP2*	6:32,789,610 -32,806,547	6.46E-08
PTRF	17:40,554,467 -40,575,506	6.34E-07	*BRD2*	6:32,936,437 -32,949,282	3.33E-09	*AGPAT1*	6:32,135,983 -32,145,888	9.07E-08
OTUD3	1:20,208,356 -20,239,438	8.46E-07	*TNXB*	6:32,008,932 -32,077,151	5.75E-09	*MCCD1*	6:31,496,739 -31,498,008	1.28E-07
CSF2RB	22:37,309,639 -37,336,491	1.05E-06	*LOC101929163*	6:32,371,234 -32,373,967	6.34E-09	*PBX2*	6:32,152,510 -32,157,963	2.78E-07
NKX2-3	10:101,292,690 -101,296,281	1.54E-06	*LTA*	6:31,539,876 -31,542,101	1.23E-08	*SLC17A1*	6:25,783,125 -25,832,287	2.82E-07
HSPA1L	6:31,777,396 -31,790,093	1.58E-06	*HLA-G*	6:29,794,756 -29,798,899	1.36E-08	*HLA-DOB*	6:32,780,540 -32,784,825	3.58E-07
DDR1	6:30,850,694 -30,867,933	1.76E-06	*SLC39A7*	6:33,168,603 -33,172,214	2.78E-08	*C6orf15*	6:31,079,000 -31,080,332	5.67E-07
SULT1A1	16:28,616,908 -28,634,907	2.24E-06	*MUC22*	6:30,973,729 -31,003,179	3.30E-08	*PSORS1C1*	6:31,082,608 -31,107,869	5.94E-07
-	-	-	*PSORS1C1*	6:31,082,608 -31,107,869	3.79E-08	*OR2H1*	6:29,424,947 -29,432,099	6.02E-07
-	-	-	*HLA-F*	6:29,691,117 -29,695,073	4.07E-08	*HSPA1L*	6:31,777,396 -31,790,093	6.11E-07
-	-	-	*STAT3*	17:40,465,342 -40,540,586	5.46E-08	*ATP6V1G2*	6:31,512,228 -31,514,625	6.57E-07
-	-	-	*MLN*	6:33,762,449 -33,771,793	7.91E-08	*MAS1L*	6:29,454,543 -29,455,679	6.92E-07
-	-	-	*LILRB3*	19:54,720,147 -54,726,997	8.92E-08	*LCOR*	10:98,741,041 -98,745,585	9.58E-07
-	-	-	*GPSM3*	6:32,158,543 -32,163,300	8.93E-08	*TCF19*	6:31,126,303 -31,134,183	1.24E-06
-	-	-	*GPANK1*	6:31,629,006 -31,634,060	1.03E-07	*TNXB*	6:32,008,932 -32,077,151	1.29E-06
-	-	-	*HSPA1L*	6:31,777,396 -31,790,093	1.11E-07	*CFAP54*	12:96,966,648 -97,269,333	1.31E-06
-	-	-	*AGPAT1*	6:32,135,983 -32,145,888	1.40E-07	*TGFBR3*	1:92,145,900 -92,371,559	1.56E-06
-	-	-	*LOC554223*	6:29,759,683 -29,765,584	1.58E-07	*POU5F1*	6:31,132,114 -31,138,451	1.69E-06
-	-	-	*C9orf91*	9:117,373,706 -117,408,703	1.58E-07	-	-	-
-	-	-	*BTNL2*	6:32,362,513 -32,374,900	1.80E-07	-	-	-
-	-	-	*DDR1*	6:30,850,694 -30,867,933	2.51E-07	-	-	-
-	-	-	*HLA-DMA*	6:32,916,391 -32,920,900	3.27E-07	-	-	-
-	-	-	*BTN2A1*	6:26,458,132 -26,476,849	3.73E-07	-	-	-
-	-	-	*NCR3*	6:31,556,660 -31,560,762	5.13E-07	-	-	-
-	-	-	*BAG6*	6:31,606,805 -31,620,953	5.55E-07	-	-	-
-	-	-	*ATG16L1*	2:234,160,217 -234,204,320	6.35E-07	-	-	-
-	-	-	*CSF2RB*	22:37,309,639 -37,336,491	7.34E-07	-	-	-
-	-	-	*SULT1A1*	16:28,616,908 -28,634,907	1.01E-06	-	-	-
-	-	-	*NFKBIL1*	6:31,514,628 -31,526,606	1.14E-06	-	-	-
-	-	-	*NCF4*	22:37,257,030 -37,274,059	1.16E-06	-	-	-
-	-	-	*RFX6*	6:117,198,376 -117,253,326	1.26E-06	-	-	-
-	-	-	*CD40*	20:44,746,899 -44,758,384	1.52E-06	-	-	-
-	-	-	*ABHD16A*	6:31,654,726 -31,671,137	1.74E-06	-	-	-
-	-	-	*PTRF*	17:40,554,467 -40,575,506	2.07E-06	-	-	-
-	-	-	*AGER*	6:32,148,745 -32,152,099	2.52E-06	-	-	-
-	-	-	*LSM2*	6:31,765,169 -31,774,761	2.62E-06	-	-	-
-	-	-	*ATF6B*	6:32,083,045 -32,096,017	2.76E-06	-	-	-

CD, Crohn's disease; Chr, chromosome; hg19, human genome version 19; IBD, inflammatory bowel disease; *P*, *P* value; Position, chromosome position; UC, ulcerative colitis.

29 genes with $P < 2.88 \times 10^{-6}$ (0.05/17,371) in IBD, 58 genes with $P < 2.88 \times 10^{-6}$ (0.05/17,361) in CD, and 39 genes with $P < 2.87 \times 10^{-6}$ (0.05/17,396) in UC.

Table 11. Summary of genomic annotation by Haploreg v4.1 and RegulomeDB for 3 novel associations

Chr	SNP	Histone marks		DNAse	Proteins bound	Motifs changed	RegulomeDB score
		Promoter	Enhancer				
6	rs630045	PANC				Irf,Rhox11	5
6	rs339351	PANC				HMG-IY,Maf,Nanog,Sox (4 altered motifs)	6
6	rs339353	PANC			RAD21,CTCF		3a
6	rs339326		ESC, GI, PANC			Crx,FXR,Pou2f2,Sox (4 altered motifs)	5
6	rs339327						No Data
6	rs339328		GI			PU.1	5
6	rs339331		GI, PANC, LNG		KAP1	Hoxa13,Hoxb13,Hoxc10,Hoxd10 (4 altered motifs)	2b
6	rs610424					CTCF	No Data
6	rs339334					AP-1,BATF,Evi-1,GR,Irx,Pax-4,Pou2f2 (7 altered motifs)	No Data
6	rs434499					Mef2,Pbx3,SP2,SRF (4 altered motifs)	6
6	rs339340		PANC			Mef2	No Data
6	rs339341		PANC				No Data
6	rs36057271		PANC				No Data
6	rs339343	ESDR	LIV, GI, PANC		FOXA1		5
6	rs339344		PANC			ATF3,SRF	No Data
6	rs339347					AP-1,CTCFL,ERalpha-a,VDR (4 altered motifs)	No Data
6	rs654971						No Data
6	rs339350					Pou3f3	No Data
6	rs339297					BDP1,LUN-1	6
6	rs339299					GR,HNF4	No Data
6	rs339300					CHX10,Cdx,Dbx1,Eomes,Hoxd8, Ncx,Nkx6-2,Pou6f1,RXRA (9 altered motifs)	6
6	rs10706356					CTCF,Cdx2,HNF1,Hoxa9,Hoxc10,Hoxc9 (6 altered motifs)	No Data
6	rs339301					CDP,HP1-site-factor,TATA	6
6	rs339302					Hoxd10,Nkx6-1,OTX,Pou1f1,Pou2f2,Pou3f2, Sox,TATA,TEF,p300 (10 altered motifs)	No Data
6	rs1358793					Bsx,CEBPD,Dlx2,E4BP4,Hoxa5,Myc,Osr (7 altered motifs)	6
6	rs12201912					Dbx1,Dbx2,HNF1,Hoxa7,Hoxc6,Isl2,Lhx3,Lhx4,Msx-1,Nkx2,Nkx6-1, Nkx6-2,OTX,Pax-4,Pax7,Pou2f2,Pou3f2,Prrx2,Sox (19 altered motifs)	6
6	rs35974852					CEBPB,E4BP4,Evi-1,HLF,HNF1, Hdx,Pou1f1,TATA (8 altered motifs)	6
6	rs2145173					Barhl1,CTCF,Cdc5,HNF1,Myb,Pdx1 (6 altered motifs)	6
6	rs6568967					CDP,Esx1,Evi-1,Fox,HNF1,Pax-1,Pbx-1 (7 altered motifs)	No Data
6	rs6927262					ELF1,Egr-1,NRSF,SETDB1,YY1, Zfp161,Znf143,p300 (8 altered motifs)	No Data
6	rs9489065					GR	No Data
6	rs1321371		PANC			AhR	No Data
6	rs1406982					Dbx1,Foxp1,HMG-IY,HNF1,Hoxd10,Lhx3, Mef2,PLZF,Pou2f2,TATA (10 altered motifs)	6
6	rs1321372					Evi-1,Foxc1,Mef2,Pou2f2 (4 altered motifs)	No Data
6	rs6936629†		PANC				No Data
6	rs9489066		PANC			Bcl6b,Foxd1,NR4A,RAR,SF1 (5 altered motifs)	No Data
6	rs1321366		PANC			BCL,Irf,Pax-5	6
6	rs12202378					Foxo,Homez,Sox	6
10	rs76227733†		GI			AIRE,Ik-2,MZF1::1-4,PRDM1,RXRA (5 altered motifs)	No Data
10	rs57807455		LIV, GI	GI,GI,GI,LIV (4 tissues)	CEBPB,FOXA1, HNF4A,HNF4G, P300,SP1 (6 bound proteins)	Barx2,Dbx1,Dbx2,HNF1,Hlx1,Hoxa10,Hoxb13,Hoxd10,Hoxd8,Isl2,Lhx3, Lhx4,Mef2,Ncx,Nkx6-1,PLZF,Pax-6,Pou2f2,Sox,TATA (20 altered motifs)	2b
10	rs77379630					Foxc1,Pax-5,Pou2f2	No Data
10	rs10882843						5
10	rs11188927	SKIN	BLD, LIV	ESC,BRST,SKIN, SKIN,BLD,CRVX, BRST (7 tissues)	CTCF,NFKB,TCF4, ZZZ3,ELF1,YY1 (6 bound proteins)	Evi-1,Foxp1,HDAC2,Irf,Irx,Pou2f2,Pou3f2, TATA,Zfp105,p300 (10 altered motifs)	3a
19	rs12608592		BLD, BRN	SKIN,PLCNT, CRVX		BCL,GR,Mef2, Pax-5,RXRA (5 altered motifs)	2b
19	rs2240751†		ESDR, IPSC, BLD, SKIN, BRN, PANC, CRVX (7 tissues)	BLD,BLD,SKIN, HRT,BRST, BLD (6 tissues)			4

BLD, blood; BRN, brain; BRST, breast; Chr, chromosome; CRVX, cervix; ESC, embryonic stem cell; ESDR, embryonic stem cell derived; GI, gastrointestine; IPSC, induced pluripotent stem cell; LIV, liver; LNG, lung; PANC, pancreas; PLCNT, placenta.

†Lead SNPs. SNPs with $r^2 \geq 0.8$ were also selected.

Table 12. List of histone marks and binding site of transcription factors at candidate cis-regulatory element (EH38E1491359) in the ENCODE project database

Transcriptional regulation	Name	Cell type(s)
Histone marks (6)	H3K27ac	colonic mucosa, colonic mucosa, hepatocyte, large intestine, mucosa of rectum, mucosa of rectum, Peyer's patch, sigmoid colon, sigmoid colon, small intestine, small intestine, small intestine, small intestine, transverse colon (14 cell types)
	H3K4me1	body of pancreas, colonic mucosa, duodenal mucosa, duodenal mucosa, hepatocyte, HepG2, large intestine, liver, liver, mucosa of rectum, small intestine (11 cell types)
	H3K4me3	HepG2, HepG2, mucosa of rectum
	H3K9ac	colonic mucosa
	H3K4me2	HepG2
	H3K79me2	B cell
Transcription factors (19)	POLR2A	HepG2, Peyer's patch, transverse colon, transverse colon (4 cell types)
	FOXA1	HepG2, HepG2
	FOXA2	HepG2, HepG2
	HNF4A	HepG2, liver
	SP1	HepG2
	CHD4	HepG2
	CREM	HepG2
	JUND	HepG2
	TBX3	HepG2
	ASH2L	HepG2
	EP300	HepG2
	HDAC2	HepG2
	HNF1A	HepG2
	HNF4G	HepG2
	KDM1A	HepG2
	NCOR1	HepG2
	RAD21	HepG2
	POLR2G	HepG2
	HNRNPUL1	HepG2

histidine at position 182 of the transmembrane helical domain. In silico evaluation of rs2240751 based on sequence homology and physico-chemical similarity predicted the substitution to be deleterious with a SIFT[62] score of 0 and probably damaging with a PolyPhen-2[63] score of 1. *MFSD12* belongs to the major facilitator superfamily (MFS) of membrane proteins, the largest family of secondary transporters. MFS proteins catalyze the transport of a wide range of substrates in both directions across the membrane.[64] *MFSD12* mRNA levels are low in the depigmented skin of vitiligo patients, probably due to the autoimmune-related destruction of melanocytes.[65] *MFSD12* has also been found to be associated with skin pigmentation in Africans.[66] The minor allele G of rs2240751 showed active regulatory features in 11 immune cell types including CD4 T cells, macrophages, and granulocytes in the Ensembl Genome Browser[54](Table 13).

Locus 6q22 for CD

rs6936629 at the 6q22 locus identified in CD meta-analysis ($P_{combined} = 3.63 \times 10^{-8}$, OR = 1.25) was in a LD region of 446.5 kb, which included *RFX6* (regulatory factor X6), *GPRC6A* (G protein-coupled receptor class C group 6 member A), and *FAM162B* (family with sequence similarity 162 member B)(Table 8 and Figure 6C). The 95% credible set consisted of 277 SNPs including rs6936629 with a PP of 4.6% (Table 9). Gene analysis using MAGMA[57] identified *RFX6* as an annotated gene with a significant *P* value of 1.26×10^{-6} (Table 10). rs6936629, located in intron 9 of *RFX6*, did not have eQTL effects on the genes at this locus. *RFX6* belongs to the regulatory factor X (RFX) family of transcription factors, which can bind to X-box motifs highly conserved in the promoter regions of various MHC class II genes.[67] Among 37 SNPs in high LD ($r^2 \geq 0.8$) with rs6936629 in Haploreg v4.1,[52] rs339331 ($r^2 = 0.97$) with enhancer histone marks in gastrointestinal tissues had the highest RegulomeDB[53] score (2b) (Table 11). Based on previous reports indicating that the prostate cancer risk-associated SNP rs339331 lies within a functional *HOXB13*-binding site and that the T risk allele increases the transcription of *RFX6* by promoting the binding of *HOXB13* to a transcriptional enhancer,[68,69] the C risk allele for CD appeared to be associated with the decreased expression of *RFX6* ($P_{meta} = 3.53 \times 10^{-6}$). Recently, *RFX6* was reported to be an essential transcriptional regulator of enteroendocrine cell specification in mice, which sheds light on the molecular mechanisms of intestinal failures in human *RFX6*-deficiencies such as Mitchell-Riley syndrome.[70,71]

Table 13. Summary of regulatory effects of rs2240751 from the Ensembl database

SNP	Allele	Regulatory feature	Active cell lines	Binding transcription factors	Binding affinity
rs2240751	G	ENSR00000106069	Neutrophil (VB), eosinophil (VB), neutrophil myelocyte (BM), M0 macrophage (VB), Monocytes-CD14+, neutrophil (CB), CD14+CD16- monocyte (VB), CD14+CD16- monocyte (CB), naive_thymus_derived_CD4_positive__alpha_beta_T_cell, CD14_positive_monocyte, neutrophil (11 cell lines)	MYBL2, MYBL1, RFX3, SRF	Decrease

CB, cord blood; BM, bone marrow; SNP, single nucleotide polymorphism; VB, venous blood.

We further examined why these 3 novel loci were not identified as significant in the European population (Table 14), even though previous European studies had much larger sample sizes. rs76227733 at 10q24 identified in UC meta-analysis did not show a significant association in the European population ($P = 9.82 \times 10^{-2}$). It showed a relatively different allele frequency between two populations (MAF = 3.8% in the European population vs. 28.6% in the East Asian population) (Table 14). However, rs57807455, ~1.9 kb away from rs76227733 with low LD ($r^2 < 0.2$ in Europeans), showed suggestive association with UC in Europeans ($P = 3.78 \times 10^{-3}$, OR = 1.08). rs2240751 at 19p13 identified in CD meta-analysis showed effects in the same direction between the Korean and European data; however, the *P* value (4.22×10^{-1}) was not significant (Table 14). The risk allele frequency of rs2240751 was around 1% in the European population, whereas it was 26.9% in the East Asian population. Among the 3 novel loci, rs6936629 at the 6q22 locus including *RFX6-GPRC6A-FAM162B* showed significant association in European population ($P = 1.88 \times 10^{-4}$, OR = 1.07) (Table 14).

3.3. Previously reported loci

With our Korean data, we first examined 245 IBD-associated loci (276 independent SNPs) that were previously established in European studies (Table 15).[15,16] A total of 39 independent SNPs in 36 loci were monomorphic in the East Asian population, and 13 additional independent SNPs in 10 loci did not have proxy SNPs ($r^2 > 0.8$). Of the remaining 224 SNPs in 199 available loci, 29 independent SNPs in 27 loci were Bonferroni significant (0.05/276, $P<1.81 \times 10^{-4}$). Although the top SNPs were different with low LD ($r^2<0.2$) between the Korean and European data, 3 additional loci including *GPR35*, *TNFRSF6B*, and *NCF4* showed a genome-wide significant association in the discovery phase (Figure 5A and B). Of the 7 loci first identified in Asians, 3 loci including the *PYGO2-SHC1* locus at 1q21,[25] *CDYL2* at 16q23,[25] and *CDKN2A-AS1-CDKN2A-CDKN2B-AS1-CDKN2B* at 9p21[24] were not present in the list of 245 IBD loci. These 3 Asian-specific loci had Bonferroni significant $P_{meta} < 1.67 \times 10^{-2}$ (0.05/3) in the discovery phase; thus we consider them as replicated as well (Table 16). In total, 35 SNPs from 33 loci, including the 27, 3, and 3 loci described above, were replicated in the Korean population. An additional 77 independent SNPs in 73 loci did not reach the Bonferroni threshold but showed nominal $P < 0.05$ in Koreans. Of those, 5 loci had an opposite direction of effects with Europeans (Table 15).

Table 14. European data* of the 3 novel loci identified in the Korean population

Phenotype	Locus	SNP	Position (hg19)	Candidate gene(s)	Risk allele	RAF†		OR (95% CI)	*P*
						EUR	EAS		
UC	10q24	rs76227733	98,556,649	*LCOR*, *SLIT1*	C	0.038	0.286	1.07 (0.99-1.16)	9.82×10^{-2}
CD	19p13	rs2240751	3,548,231	*MFSD12*, *C19orf71*, *FZR1*, *DOHH*	G	0.010	0.269	1.06 (0.92-1.23)	4.22×10^{-1}
	6q22	rs6936629	117,239,141	*RFX6*, *FPRC6A*, *FAM162B*	C	0.319	0.424	1.07 (1.03-1.10)	1.88×10^{-4}

CD, Crohn's disease; Chr, chromosome; CI, confidence interval; EAS, East Asian population; EUR, European population; hg19, human genome version 19; OR, odds ratio; *P*, *P* value; Position, chromosome position; RAF, risk allele frequency; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

*European data was from summary statistics of de Lange KM et al (ref. 16).

†European and East Asian frequency from dbSNP (https://www.ncbi.nlm.nih.gov/snp/).

Table 15. Associations of the previously established Caucasian inflammatory bowel disease 276 independent SNPs in Koreans

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
1	rs12103	1,247,494	*TNFRSF18, TNFRSF4*	T	C	0.20	0.99	2.09E-07	5.79E-01	0.00	2.80E-04	4.61E-01	0.00	3.09E-06	1.65E-01	0.00	
1	rs6667605	2,502,780	*TNFRSF14*	T	C	0.49	0.48	5.58E-06	**3.57E-02**	0.02	5.86E-01	9.06E-01	0.00	2.65E-09	**9.35E-04**	0.07	
1	rs3766606	8,022,197	*TNFRSF9, ERRFI1, PARK7*	T	G	0.16	0.07	1.35E-12	3.45E-01	0.03	2.78E-06	6.84E-01	0.01	3.94E-09	2.50E-01	0.01	
1	rs10799838	20,135,822	*PLA2G2A*	T	C	0.24	0.23	3.48E-06	5.34E-01	0.02	8.48E-01	4.59E-01	0.00	8.48E-14	9.81E-01	0.19	
1	rs3806308	20,142,866	*PLA2G2A*	T	C	0.37	0.48	1.26E-12	8.57E-02	0.17	4.88E-01	6.19E-01	0.00	7.04E-24	**4.36E-03**	0.59	
1	rs6426833‡	20,171,860	*PLA2G2A*	G	A	0.46	0.21	1.01E-19	**1.63E-05**	0.17	8.60E-01	6.06E-01	0.00	3.04E-42	**7.94E-11**	0.65	UC
1	rs12568930‡	22,702,231		C	T	0.18	0.18	7.53E-15	**5.38E-03**	0.22	1.31E-04	5.25E-01	0.01	1.60E-15	**1.48E-04**	0.25	IBD_S
1	rs1748195	63,049,593		G	C	0.33	0.23	–	3.67E-01	NA	–	9.60E-01	NA	–	9.23E-02	NA	
1	rs6588248	67,652,984	*IL23R*	T	G	0.47	0.36	7.41E-10	6.50E-01	0.06	1.54E-20	9.43E-01	0.38	4.63E-01	3.95E-01	0.00	
1	rs7517847‡	67,681,669	*IL23R*	G	T	0.43	0.40	7.44E-80	**1.82E-05**	1.00	5.84E-97	**9.88E-06**	1.00	3.96E-18	**2.48E-02**	0.32	IBD_U
1	rs80174646	67,708,155	*IL23R*	T	G	0.07	0.00	8.71E-107	–	NA	8.20E-90	–	NA	4.49E-40	–	NA	
1	rs2651244	70,995,562		A	G	0.40	0.11	7.91E-01	9.82E-02	0.00	1.02E-05	7.43E-02	0.00	1.05E-06	3.63E-01	0.01	
1	rs17391694	78,623,626		T	C	0.11	0.00	6.83E-04	–	NA	9.90E-06	–	NA	5.21E-01	–	NA	
1	rs34856868	92,554,283		A	G	0.02	0.00	8.86E-03	–	NA	4.15E-03	–	NA	2.67E-01	–	NA	
1	rs11583043	101,466,054	*EDG1*	T	C	0.27	0.13	8.91E-04	2.82E-01	0.00	3.91E-01	1.39E-01	0.00	6.67E-05	7.66E-01	0.00	
1	rs6679677	114,303,808	*PTPN22*	A	C	0.10	0.00	9.91E-05	–	NA	1.77E-15	–	NA	2.33E-01	–	NA	
1	rs2641348†	120,437,884		G	A	0.10	0.03	9.87E-04	5.56E-02	0.00	1.27E-04	4.91E-01	0.00	1.01E-01	**3.02E-03**	0.00	
1	rs4845604	151,801,680	*RORC*	A	G	0.14	0.04	7.09E-14	–	NA	6.06E-07	–	NA	1.57E-11	–	NA	
1	rs670523	155,878,732		A	G	0.33	0.89	1.28E-04	3.36E-01	0.00	3.22E-05	2.45E-01	0.00	1.12E-02	5.85E-01	0.00	
1	rs34687326	159,799,910	*SLAMF8*	A	G	0.10	0.00	2.58E-05	–	NA	1.06E-08	–	NA	1.25E-01	–	NA	
1	rs4656958	160,856,964	*SLAMF1, CD48*	A	G	0.31	0.28	1.68E-08	3.82E-01	0.05	6.39E-07	4.38E-01	0.03	4.69E-05	5.52E-01	0.01	
1	rs1801274‡	161,479,745	*FCGR2A, FCGR2B, FCGR3B, FCGR3A*	G	A	0.49	0.23	9.34E-14	**5.79E-03**	0.09	1.59E-02	6.77E-01	0.00	1.52E-18	**5.72E-05**	0.17	UC
1	rs6025	169,519,049	*SELP, SELE, SELL*	T	C	0.03	0.00	8.92E-05	–	NA	2.81E-03	–	NA	1.33E-02	–	NA	
1	rs7517810	172,853,460	*TNFSF18, FASLG*	T	C	0.24	0.92	8.03E-09	6.41E-01	0.01	1.55E-21	2.08E-01	0.06	7.64E-01	3.95E-01	0.00	
1	rs10798069	186,875,459	*PTGS2, PLA2G4A*	T	G	0.48	0.44	2.29E-02	1.07E-01	0.00	5.27E-05	1.14E-01	0.01	9.76E-01	**4.55E-02**	0.00	
1	rs10801047	191,559,356	.	A	T	0.08	0.12	6.66E-01	1.60E-01	0.00	1.75E-01	4.13E-01	0.00	4.90E-01	1.66E-01	0.00	
1	rs2488389	197,631,141		A	G	0.22	0.19	5.58E-11	2.84E-01	0.08	1.56E-11	3.17E-01	0.11	4.28E-04	5.32E-01	0.01	
1	rs7555082	198,598,663	*PTPRC*	A	G	0.12	0.00	3.82E-03	–	NA	8.77E-06	–	NA	7.21E-01	–	NA	
1	rs2816958	200,101,920		A	G	0.11	0.00	8.71E-08	–	NA	1.31E-01	–	NA	2.00E-13	–	NA	
1	rs7554511	200,877,562		A	C	0.27	0.00	1.00E-21	–	NA	1.49E-10	–	NA	4.27E-16	–	NA	
1	rs3024505	206,939,904	*IL10, IL19, IL20, FAIM3, IL24, MAPKAPK2, PIGR*	A	G	0.16	0.03	6.04E-31	**2.15E-02**	0.05	2.88E-14	2.81E-01	0.01	1.53E-23	**7.32E-03**	0.03	
1	rs59043219	209,970,610	*DIEXF,IRF6*	A	G	0.36	0.57	1.09E-08	3.94E-01	0.06	7.12E-07	6.60E-01	0.04	1.17E-04	2.27E-01	0.01	
2	rs11894081	5,664,008	.	G	T	0.25	0.39	6.79E-01	9.25E-01	0.00	2.27E-01	7.64E-01	0.00	5.64E-01	8.28E-01	0.00	
2	rs13407913	25,097,644		G	A	0.44	0.46	3.39E-07	1.92E-01	0.03	9.04E-08	6.49E-01	0.05	3.55E-03	1.21E-01	0.00	
2	rs1260326	27,730,940	*UCN*	T	C	0.42	0.55	9.61E-08	8.27E-01	0.04	6.32E-11	4.96E-01	0.11	2.61E-02	5.17E-01	0.00	
2	rs925255	28,614,794	*FOSL2, BRE*	T	C	0.44	0.19	–	2.08E-01	NA	–	2.72E-01	NA	–	2.59E-01	NA	
2	rs10495903	43,806,918		T	C	0.13	0.01	5.53E-09	–	NA	4.41E-11	–	NA	9.21E-03	–	NA	
2	rs7608910	61,204,856	*REL*	G	A	0.39	0.04	2.72E-28	1.79E-01	0.02	5.88E-14	5.23E-02	0.01	4.35E-23	8.41E-01	0.03	
2	rs10865331	62,551,472		A	G	0.39	0.35	4.26E-01	1.09E-01	0.00	4.93E-05	7.02E-02	0.01	1.18E-02	2.43E-01	0.00	
2	rs6740462‡	65,667,272	*SPRED2*	C	A	0.26	0.16	6.05E-04	**1.06E-07**	0.00	2.91E-03	**1.04E-05**	0.00	3.60E-02	**1.65E-04**	0.00	IBD_U

Table 15. Cont'd (1)

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
2	rs10185424	102,662,888	*IL1R1, IL1R2, IL1RL2*	T	G	0.46	0.27	5.03E-09	**3.45E-02**	0.04	2.69E-05	8.60E-01	0.01	2.57E-07	**3.44E-04**	0.02	
2	rs6708413	103,063,369	*IL18RAP, IL18R1, IL1RL1, IL1RL2*	G	A	0.23	0.52	1.18E-10	**2.75E-02**	0.18	2.67E-13	**4.08E-02**	0.32	2.25E-03	2.49E-01	0.01	
2	rs11681525	145,492,382		G	C	0.09	0.00	2.87E-07	–	NA	5.19E-07	–	NA	1.68E-03	–	NA	
2	rs4664304	160,794,008	*LY75*	A	G	0.47	0.71	8.38E-02	8.66E-01	0.00	2.14E-01	4.45E-01	0.00	2.84E-01	4.01E-01	0.00	
2	rs2111485	163,110,536	*IFIH1, DPP4*	A	G	0.41	0.84	1.16E-05	5.13E-01	0.01	3.79E-03	5.49E-01	0.00	1.75E-05	7.89E-01	0.01	
2	rs6740847	182,308,352	*ITGA4*	A	G	0.45	0.70	1.22E-13	1.06E-01	0.12	9.72E-11	7.43E-02	0.07	1.43E-06	8.89E-01	0.02	
2	rs144344067	187,576,378	*ITGAV,FAM171B*	I2	D	0.12	0.00	1.29E-08	–	NA	8.27E-07	–	NA	1.37E-04	–	NA	
2	rs1517352	191,931,464	*STAT1, STAT4*	A	C	0.40	0.48	1.66E-04	4.13E-01	0.01	1.05E-03	9.74E-01	0.01	5.07E-03	3.20E-01	0.00	
2	rs1440088	198,871,417		G	T	0.20	0.26	9.42E-02	5.21E-01	0.00	3.65E-01	1.73E-01	0.00	1.26E-03	5.76E-01	0.01	
2	rs17229285	199,523,122		T	C	0.49	0.24	1.23E-03	5.80E-01	0.00	9.82E-01	5.03E-01	0.00	6.93E-08	**2.96E-02**	0.02	
2	rs1405108	199,745,018		C	A	0.32	0.20	1.59E-06	5.30E-01	0.01	7.54E-01	7.83E-01	0.00	1.66E-11	1.13E-01	0.07	
2	rs3116494	204,592,021	*ICOS, CD28, CTLA4*	G	A	0.26	0.07	9.98E-05	2.50E-01	0.00	2.16E-02	5.47E-01	0.00	6.27E-04	4.08E-01	0.00	
2	rs2382817‡	219,151,218	*SLC11A1, IL8RA, IL8RB*	A	C	0.41	0.45	1.90E-07	**9.28E-04**	0.04	7.08E-04	**7.76E-05**	0.01	2.02E-05	1.99E-01	0.02	CD
2	rs111781203	228,660,112	*CCL20*	C	T	0.33	0.16	5.09E-05	**8.80E-03**	0.01	4.29E-02	**5.53E-03**	0.00	7.75E-05	2.03E-01	0.01	
2	rs1811711	228,670,476	.	G	C	0.18	0.07	1.96E-07	–	NA	3.25E-02	–	NA	6.09E-09	–	NA	
2	rs6716753	231,097,129		C	T	0.19	0.00	5.71E-06	–	NA	1.46E-13	–	NA	7.79E-01	–	NA	
2	rs12994997‡	234,173,503		G	A	0.48	0.68	1.60E-28	**1.13E-04**	0.59	2.17E-61	**1.33E-06**	0.99	1.12E-01	2.89E-01	0.00	CD
2	rs3749171	241,569,692	*GPR35*	T	C	0.17	0.09	2.20E-14	**2.13E-03**	0.08	1.20E-06	**2.40E-04**	0.01	4.58E-12	2.36E-01	0.05	
2	rs6724516	241,586,810	*GPR35*	G	A	0.28	0.01	1.39E-09	8.45E-01	0.00	1.17E-02	6.71E-01	0.00	5.21E-13	8.20E-01	0.00	
2	rs76527535	242,484,701	.	T	C	0.23	0.02	2.87E-08	7.21E-02	0.00	5.32E-06	4.94E-01	0.00	7.89E-04	5.79E-02	0.00	
2	rs35320439	242,737,341	*PDCD1, ATG4B*	C	T	0.33	0.34	2.38E-01	3.74E-01	0.00	–	**3.65E-02**	NA	2.72E-01	4.07E-01	0.00	
3	rs4256159	18,767,404		T	C	0.14	0.13	4.47E-04	**6.03E-03**	0.01	2.90E-05	**4.05E-03**	0.01	9.98E-02	1.32E-01	0.00	
3	rs113010081	46,457,412	*FLJ78302, LTF, CCR1, CCR3, CCR5*	C	T	0.11	0.00	1.08E-03	–	NA	9.37E-01	–	NA	1.73E-06	–	NA	
3	rs9868809	48,681,053		T	C	0.10	0.09	2.68E-13	1.18E-01	0.15	1.78E-06	1.95E-01	0.03	1.80E-11	5.71E-01	0.11	
3	rs3172494	48,731,487	*IHPK2, UCN2, PFKFB4*	T	G	0.12	0.00	1.80E-06	–	NA	5.05E-05	–	NA	1.07E-03	–	NA	
3	rs4541435	49,228,501	*USP4*	T	C	0.01	0.00	4.14E-02	–	NA	9.91E-04	–	NA	9.69E-01	–	NA	
3	rs3197999	49,721,532	*MST1R*	A	G	0.29	0.07	1.43E-33	**5.25E-03**	0.13	3.30E-23	**9.28E-04**	0.06	8.40E-20	5.31E-01	0.04	
3	rs9847710	53,062,661	*PRKCD*	C	T	0.42	0.57	1.63E-01	1.50E-01	0.00	3.24E-02	7.79E-02	0.00	6.57E-05	4.42E-01	0.01	
3	rs2581828	53,133,149	*RFT1,RP11-894J14.5*	C	G	0.41	0.24	1.99E-03	1.47E-01	0.00	6.46E-09	**1.15E-03**	0.04	5.30E-01	2.91E-01	0.00	
3	rs2593855	71,175,495	*FOXP1*	T	C	0.29	0.57	2.54E-09	8.44E-01	0.11	6.50E-06	8.82E-01	0.04	1.25E-05	8.17E-01	0.03	
3	rs503734	101,023,748	*IMPG2,SENP7*	G	A	0.48	0.31	2.67E-08	3.66E-01	0.03	4.37E-07	3.75E-01	0.03	5.15E-04	4.32E-01	0.01	
3	rs616597	101,569,726	*NFKBIZ*	A	C	0.22	0.35	4.83E-04	**4.37E-03**	0.01	9.44E-04	9.40E-02	0.01	4.50E-02	**1.60E-03**	0.00	
3	rs724016	141,105,570		G	A	0.43	0.36	2.60E-03	**1.85E-03**	0.00	7.60E-08	**3.11E-04**	0.04	7.47E-01	1.57E-01	0.00	
3	rs56116661	188,401,160	*LPP*	T	C	0.20	0.06	9.27E-10	–	NA	5.67E-10	–	NA	9.21E-05	–	NA	
4	rs2073505	3,444,503		A	G	0.08	0.12	2.16E-04	**8.97E-04**	0.02	1.29E-03	**7.66E-04**	0.01	1.77E-02	6.84E-02	0.00	
4	rs4692386	26,132,361		T	C	0.40	0.17	1.40E-04	2.38E-01	0.00	6.10E-05	6.14E-01	0.01	6.35E-02	1.95E-01	0.00	
4	rs6856616‡	38,325,036		C	T	0.07	0.25	1.42E-04	**3.39E-10**	0.10	4.30E-04	**2.74E-13**	0.08	1.24E-02	**4.57E-02**	0.02	IBD_S
4	rs11734570	38,588,453	.	A	G	0.41	0.69	4.80E-08	2.19E-01	0.04	3.53E-07	4.99E-01	0.03	4.86E-04	1.85E-01	0.01	
4	rs7438704	48,363,245	*TXK*	A	G	0.35	0.63	2.19E-02	2.51E-01	0.00	1.26E-05	4.95E-01	0.02	7.42E-01	1.97E-01	0.00	
4	rs2457996	74,856,535	*IL8, CXCL3, CXCL2, CXCL6, CXCL1, CXCL5, PF4*	C	T	0.11	0.02	1.31E-03	3.32E-01	0.00	8.24E-01	4.96E-01	0.00	5.84E-07	2.10E-01	0.00	

Table 15. Cont'd (2)

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
4	rs13126505	102,865,304		A	G	0.07	0.00	3.88E-06	–	NA	3.79E-08	–	NA	1.11E-02	–	NA	
4	rs3774937	103,434,253	*NFKB1*	C	T	0.33	0.35	8.14E-01	1.73E-01	0.00	9.45E-04	7.42E-01	0.01	6.15E-03	**5.75E-04**	0.00	
4	rs2189234‡	106,075,498		T	G	0.37	0.37	9.38E-04	**1.13E-04**	0.01	2.15E-01	5.64E-02	0.00	1.47E-05	**1.43E-05**	0.02	UC
4	rs7657746	123,161,619	*IL2*	G	A	0.25	0.04	1.96E-07	3.63E-01	0.00	4.98E-07	3.12E-01	0.00	1.44E-03	9.72E-01	0.00	
5	rs11739663	594,083		C	T	0.24	0.02	7.58E-04	9.24E-01	0.00	2.56E-01	7.99E-01	0.00	3.74E-08	8.83E-01	0.00	
5	rs2930047‡	10,695,526	*DAP*	C	T	0.38	0.75	5.08E-06	**7.54E-05**	0.01	4.24E-07	**1.30E-02**	0.02	1.07E-02	**2.01E-04**	0.00	IBD_U
5	rs3194051	35,876,274	*IL7R,CAPSL*	G	A	0.27	0.04	6.37E-03	6.27E-01	0.00	3.54E-01	7.80E-01	0.00	1.78E-03	5.77E-01	0.00	
5	rs395157	38,867,732	*OSMR, FYB*	T	C	0.50	0.26	4.63E-10	**9.73E-03**	0.05	1.56E-07	9.31E-02	0.03	4.36E-06	**1.12E-02**	0.01	
5	rs1842076	40,237,018		C	T	0.29	0.03	8.27E-13	4.85E-01	0.00	1.79E-12	4.90E-01	0.00	9.29E-05	5.40E-01	0.00	
5	rs11742570‡	40,410,584		T	C	0.40	0.82	3.64E-40	**2.71E-05**	0.63	1.11E-55	**2.52E-04**	0.91	2.76E-07	**2.49E-03**	0.01	IBD_U
5	rs1505992‡	40,498,577	*PTGER4*	A	T	0.32	0.63	1.55E-21	**3.47E-05**	0.53	2.31E-38	**1.07E-05**	0.95	2.10E-02	**3.68E-02**	0.00	CD
5	rs10065637	55,438,851	*IL6ST, IL31RA*	T	C	0.21	0.03	1.09E-05	–	NA	5.04E-06	–	NA	1.53E-02	–	NA	
5	rs4703855	71,693,899		T	C	0.30	0.56	1.41E-05	1.20E-01	0.03	1.50E-04	1.11E-01	0.02	2.40E-03	3.10E-01	0.01	
5	rs10061469	72,518,148		C	T	0.32	0.20	2.93E-03	1.79E-01	0.00	1.17E-04	3.99E-01	0.01	8.47E-01	**4.83E-02**	0.00	
5	rs1363907	96,252,803		A	G	0.41	0.28	1.10E-10	5.44E-01	0.07	1.42E-14	7.27E-02	0.17	1.22E-02	3.01E-01	0.00	
5	rs7705924	101,946,798	*SLCO4C1,SLCO6A1*	G	A	0.04	0.01	8.93E-02	8.50E-01	0.00	5.96E-01	4.27E-01	0.00	1.18E-01	4.39E-01	0.00	
5	rs10051722	130,104,076		C	A	0.30	0.39	–	7.47E-01	NA	–	6.36E-01	NA	–	8.54E-01	NA	
5	rs11743851	130,613,600		C	T	0.38	0.00	2.58E-12	–	NA	3.86E-19	–	NA	4.42E-02	–	NA	
5	rs17622378	131,778,452	*IRF1, IL4, IL13, IL5*	G	A	0.42	0.00	2.40E-26	–	NA	8.82E-35	–	NA	2.63E-06	–	NA	
5	rs254560	134,443,606		A	G	0.39	0.19	1.63E-06	**9.37E-04**	0.01	3.29E-01	**2.69E-02**	0.00	2.62E-08	**2.26E-03**	0.02	
5	rs6863411	141,513,204		A	T	0.37	0.32	5.02E-10	6.76E-02	0.07	7.24E-12	1.14E-01	0.12	2.78E-04	1.02E-01	0.01	
5	rs17656349	149,605,994	*SLC6A7,CAMK2A*	C	T	0.44	0.37	5.17E-09	1.65E-01	0.05	3.14E-03	**4.15E-02**	0.00	1.54E-08	9.51E-01	0.05	
5	rs11741861‡	150,277,909	*TNIP1*	G	A	0.08	0.38	3.28E-15	**2.31E-06**	0.94	2.03E-19	**3.97E-04**	0.99	1.10E-05	**9.65E-06**	0.23	IBD_U
5	rs6556412	158,787,385	*IL12B*	A	G	0.33	0.44	5.51E-22	**1.62E-03**	0.53	1.73E-19	**1.72E-03**	0.47	2.92E-10	2.09E-01	0.11	
5	rs9313808	158,820,844	*IL12B*	A	G	0.17	0.00	1.38E-16	–	NA	3.42E-11	–	NA	5.39E-09	–	NA	
5	rs56167332‡	158,827,769	*IL12B*	A	C	0.34	0.34	2.52E-38	**5.34E-08**	0.91	1.19E-27	**8.97E-09**	0.75	1.14E-23	**3.23E-03**	0.57	IBD_U
5	rs564349	172,324,978	*DUSP1*	G	A	0.32	0.34	2.35E-06	**3.55E-02**	0.03	2.24E-04	6.18E-02	0.01	6.67E-05	2.55E-01	0.01	
5	rs72810983	173,318,254		G	A	0.30	0.07	4.15E-03	2.43E-01	0.00	3.64E-04	**4.11E-02**	0.00	2.47E-01	8.37E-01	0.00	
5	rs4976646	176,788,570	*DOK3*	C	T	0.34	0.33	3.99E-09	**1.88E-02**	0.07	3.05E-04	**2.86E-02**	0.01	4.48E-08	1.11E-01	0.05	
6	rs7773324	382,559	*IRF4, DUSP22*	G	A	0.40	0.76	1.69E-01	6.72E-01	0.00	7.24E-03	6.10E-01	0.00	8.51E-01	8.04E-01	0.00	
6	rs13204048	3,420,406		C	T	0.39	0.50	4.70E-02	4.30E-01	0.00	6.74E-04	**3.91E-02**	0.01	9.75E-01	3.72E-01	0.00	
6	rs17119	14,719,496		G	A	0.21	0.09	9.35E-11	4.60E-01	0.02	1.50E-07	9.04E-01	0.01	6.54E-07	3.17E-01	0.01	
6	rs113986290	19,781,009	.	T	C	0.03	0.00	6.43E-05	–	NA	3.43E-01	–	NA	7.59E-09	–	NA	
6	rs6908425	20,728,731		T	C	0.22	0.18	4.46E-11	7.12E-02	0.07	1.12E-10	2.47E-01	0.08	6.44E-05	2.25E-01	0.01	
6	rs9358372	20,812,588		G	A	0.36	0.57	2.35E-07	3.41E-01	0.04	1.25E-08	4.11E-01	0.07	8.28E-03	6.19E-01	0.00	
6	rs71559680	21,430,728		T	C	0.47	0.68	3.72E-07	8.48E-02	0.03	2.32E-08	5.05E-02	0.05	3.17E-02	1.81E-01	0.00	
6	rs116392568	31,274,380		C	T	0.37	0.48	8.18E-10	5.92E-01	0.13	5.31E-18	9.73E-01	0.50	6.25E-02	8.15E-01	0.00	
6	rs9273363‡	32,626,272		A	C	0.29	0.32	9.92E-34	**1.05E-02**	0.93	3.61E-07	**3.03E-02**	0.07	2.41E-45	**2.28E-13**	0.99	UC
6	rs67289879	42,007,403	*CCND3*	T	C	0.20	0.00	3.04E-08	–	NA	2.03E-07	–	NA	2.64E-04	–	NA	
6	rs943072	43,795,968	.	G	T	0.10	0.13	6.41E-05	5.48E-01	0.02	1.35E-01	9.76E-01	0.00	1.81E-04	2.88E-01	0.02	

Table 15. Cont'd (3)

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
6	rs1847472	90,973,159		A	C	0.35	0.05	1.79E-09	2.68E-01	0.00	5.93E-09	2.22E-01	0.00	5.81E-04	8.71E-01	0.00	
6	rs7746082	106,435,269		C	G	0.29	0.00	9.03E-13	–	NA	2.04E-12	–	NA	3.32E-05	–	NA	
6	rs3851228	111,848,191	*TRAF3IP2, FYN*	T	A	0.07	0.00	3.91E-10	–	NA	1.51E-03	–	NA	5.93E-10	–	NA	
6	rs2858829	116,768,917	*FAM26F,TRAPPC3L*	G	A	0.41	0.47	1.51E-03	5.35E-01	0.01	7.64E-01	**3.96E-02**	0.00	8.69E-06	5.15E-01	0.02	
6	rs2503322	127,457,260		A	G	0.47	0.40	1.82E-03	**4.66E-02**	0.00	3.02E-07	**2.31E-02**	0.04	9.94E-01	3.51E-01	0.00	
6	rs13204742	128,245,765		T	G	0.13	0.00	1.40E-03	–	NA	4.74E-06	–	NA	2.89E-01	–	NA	
6	rs6920220	138,006,504	*TNFAIP3*	A	G	0.21	0.00	1.00E-08	–	NA	3.19E-01	–	NA	2.89E-15	–	NA	
6	rs12199775	143,898,894		G	A	0.07	0.06	2.91E-05	**1.46E-02**	0.01	1.04E-04	5.08E-02	0.01	6.55E-02	2.00E-01	0.00	
6	rs7758080	149,577,079	*MAP3K7IP2*	G	A	0.28	0.48	1.09E-04	1.17E-01	0.02	6.09E-05	**1.26E-02**	0.02	2.24E-01	9.71E-01	0.00	
6	rs212388	159,490,436	*TAGAP*	C	T	0.41	0.66	1.62E-05	7.36E-01	0.02	9.52E-11	4.47E-01	0.09	2.41E-01	6.93E-01	0.00	
6	rs1819333‡	167,373,547	*CCR6, RPS6KA2*	G	T	0.48	0.59	8.81E-15	**3.94E-06**	0.20	1.68E-20	**3.83E-08**	0.43	1.12E-03	2.44E-01	0.01	CD
7	rs1182188	2,869,985	*CARD11*	C	T	0.30	0.16	4.86E-06	7.02E-01	0.01	6.11E-01	8.74E-01	0.00	8.59E-10	4.87E-01	0.03	
7	rs11768365	6,545,188	*FLJ20306,DAGLB,KDELR2,GRID2IP*	G	A	0.22	0.23	3.88E-08	5.85E-02	0.05	6.80E-05	**4.83E-02**	0.02	1.91E-06	3.33E-01	0.03	
7	rs1077773	17,442,679	*AHR*	G	A	0.46	0.37	1.61E-03	1.60E-01	0.00	8.74E-01	3.33E-01	0.00	6.25E-06	**4.35E-02**	0.02	
7	rs149169037	20,577,298	.	A	G	0.08	0.00	3.26E-08	–	NA	3.01E-05	–	NA	2.87E-05	–	NA	
7	rs10486483	26,892,440	*SKAP2*	A	G	0.24	0.10	8.58E-02	**1.90E-02**	0.00	4.32E-04	**2.22E-03**	0.00	9.60E-01	9.74E-01	0.00	
7	rs4722672	27,231,762		C	T	0.18	0.43	3.84E-07	8.66E-01	0.11	1.15E-02	8.15E-01	0.01	1.26E-07	2.97E-01	0.13	
7	rs864745	28,180,556		T	C	0.50	0.73	4.68E-03	9.96E-01	0.00	1.61E-05	1.43E-01	0.01	9.86E-01	1.52E-01	0.00	
7	rs12718244	50,175,654	*IKZF1*	A	G	0.41	0.19	7.88E-05	1.52E-01	0.01	2.02E-03	6.21E-02	0.00	4.40E-03	9.44E-01	0.00	
7	rs1456896	50,304,461	*IKZF1*	C	T	0.31	0.56	4.50E-11	**3.82E-02**	0.14	9.38E-12	**1.47E-02**	0.19	1.88E-03	2.50E-01	0.01	
7	rs9297145	98,759,117	*SMURF1*	C	A	0.26	0.09	8.37E-09	4.47E-01	0.01	1.24E-05	3.25E-01	0.00	6.76E-07	7.77E-01	0.01	
7	rs314313	100,423,365	*EPO*	C	T	0.30	0.02	5.47E-06	3.50E-01	0.00	9.47E-05	6.09E-01	0.00	2.00E-03	6.56E-01	0.00	
7	rs7805114	107,450,033		G	T	0.43	0.31	–	5.28E-01	NA	–	**3.45E-02**	NA	–	1.40E-01	NA	
7	rs4380874	107,480,315		T	C	0.41	0.12	8.22E-16	**3.34E-02**	0.04	2.08E-02	1.40E-01	0.00	9.07E-21	**2.64E-02**	0.09	
7	rs38911	116,895,163		A	G	0.47	0.31	6.24E-06	6.84E-01	0.01	2.60E-03	5.47E-01	0.00	1.33E-05	1.63E-01	0.01	
7	rs4728142	128,573,967	*IRF5*	A	G	0.44	0.13	9.12E-05	1.28E-01	0.00	6.97E-01	5.35E-01	0.00	3.23E-10	**7.09E-04**	0.02	
7	rs2538470	148,220,448		A	G	0.37	0.20	3.77E-05	**1.84E-02**	0.01	6.10E-05	**5.48E-03**	0.01	8.90E-03	1.31E-01	0.00	
7	rs243505	148,435,339	*CUL1,EZH2*	G	A	0.38	0.29	3.04E-10	1.99E-01	0.06	5.52E-07	1.14E-01	0.03	3.46E-05	4.29E-01	0.01	
8	rs17057051	27,227,554	*PTK2B*	G	A	0.31	0.21	9.90E-04	**4.23E-02**	0.00	5.43E-04	1.63E-01	0.01	2.65E-01	**4.27E-02**	0.00	
8	rs7011507	49,129,242		A	G	0.13	0.22	5.84E-06	**1.72E-02**	0.05	4.21E-03	3.79E-01	0.01	3.64E-05	**5.02E-03**	0.04	
8	rs12677663	74,007,347	*SBSPON*	G	T	0.42	0.17	2.21E-01	9.64E-01	0.00	6.38E-02	3.02E-01	0.00	7.58E-01	1.94E-01	0.00	
8	rs7015630	90,875,918	*RIPK2, NBN*	C	T	0.26	0.17	9.98E-04	**4.64E-02**	0.00	3.34E-05	**2.23E-03**	0.01	5.26E-01	8.00E-01	0.00	
8	rs921720	126,534,671	*TRIB1*	A	G	0.39	0.59	2.73E-12	7.74E-01	0.14	2.61E-15	5.35E-01	0.27	2.95E-03	9.00E-01	0.00	
8	rs6651252	129,567,181		C	T	0.13	0.04	5.63E-07	6.04E-01	0.00	8.68E-11	4.76E-01	0.01	2.50E-01	9.03E-01	0.00	
8	rs13277237	130,604,563		G	A	0.44	0.53	4.64E-06	**1.06E-02**	0.02	1.37E-02	**9.19E-03**	0.00	1.16E-05	2.22E-01	0.02	
9	rs75900472	4,981,602	*JAK2*	C	A	0.35	0.34	–	**6.66E-03**	NA	–	7.63E-02	NA	–	**1.34E-02**	NA	
9	rs9408254	34,736,158	*CCL21,FAM205A*	A	G	0.15	0.02	8.06E-04	5.10E-01	0.00	1.47E-04	6.64E-02	0.00	2.59E-01	5.08E-01	0.00	
9	rs4743820	93,928,416	*NFIL3*	C	T	0.30	0.37	3.63E-06	5.65E-01	0.03	3.04E-03	6.67E-01	0.01	8.33E-05	6.24E-01	0.02	
9	rs4246905‡	117,553,249	*TNFSF15, TNFSF8*	T	C	0.28	0.30	4.62E-27	**3.76E-21**	0.68	1.13E-20	**4.93E-33**	0.51	5.86E-15	**3.25E-02**	0.27	CD
9	rs11554257	117,605,070	*TNFSF15, TNFSF8*	C	T	0.13	0.17	7.41E-13	8.81E-01	0.24	5.67E-10	8.79E-01	0.15	4.29E-06	6.22E-01	0.04	

Table 15. Cont'd (4)

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
9	rs13300483‡	117,643,362	*TNFSF15, TNFSF8*	T	C	0.25	0.37	9.00E-09	**1.53E-30**	0.09	2.00E-11	**1.03E-53**	0.20	8.07E-03	7.44E-02	0.00	CD
9	rs4986790	120,475,302	*TLR4*	G	A	0.06	0.00	4.66E-03	–	NA	2.77E-05	–	NA	5.73E-01	–	NA	
9	rs10781499	139,266,405	*CARD9*	A	G	0.41	0.29	5.06E-36	**4.14E-02**	0.75	6.40E-30	6.51E-01	0.65	2.07E-16	**7.33E-03**	0.20	
9	rs13300218	139,399,641	*CARD9*	A	G	0.11	0.00	2.13E-13	–	NA	1.01E-08	–	NA	1.12E-09	–	NA	
10	rs12722515	6,081,230	*IL2RA, IL15RA*	A	C	0.15	0.10	1.30E-06	5.91E-01	0.01	4.63E-07	6.76E-01	0.02	4.65E-03	6.83E-01	0.00	
10	rs7911117	27,179,596	.	G	T	0.14	0.13	1.42E-04	4.74E-01	0.01	9.83E-01	2.73E-01	0.00	1.84E-08	8.14E-01	0.05	
10	rs1042058	30,728,101	*MAP3K8*	T	C	0.41	0.57	2.81E-12	4.23E-01	0.14	4.63E-10	7.17E-01	0.10	5.26E-06	4.38E-01	0.02	
10	rs11010067	35,295,431	*CREM*	G	C	0.35	0.27	1.54E-12	4.13E-01	0.10	1.32E-14	7.32E-01	0.18	1.87E-04	5.74E-01	0.01	
10	rs1199103	59,947,231		G	A	0.22	0.40	5.43E-06	1.01E-01	0.05	3.22E-06	3.66E-01	0.06	9.72E-03	1.38E-01	0.01	
10	rs10995235	64,369,749		A	G	0.17	0.21	1.21E-05	2.79E-01	0.02	6.72E-02	9.85E-02	0.00	6.34E-06	7.60E-01	0.03	
10	rs10761659	64,445,564		A	G	0.46	0.23	2.30E-36	8.88E-02	0.58	7.65E-29	**3.73E-04**	0.44	1.33E-15	2.29E-01	0.12	
10	rs224090‡	64,541,319		T	C	0.41	0.58	3.91E-14	**4.80E-05**	0.18	2.60E-16	**1.30E-06**	0.28	5.18E-04	3.16E-01	0.01	IBD_S
10	rs2227551	75,669,190	*PLAU*	G	T	0.26	0.52	–	2.60E-01	NA	–	1.95E-01	NA	–	7.91E-01	NA	
10	rs1250546	81,032,532		G	A	0.40	0.49	3.10E-11	1.30E-01	0.12	6.38E-14	**1.51E-02**	0.23	3.56E-03	9.02E-01	0.00	
10	rs7097656	82,250,831		T	C	0.20	0.02	2.50E-10	3.45E-01	0.00	2.69E-10	4.32E-01	0.00	6.20E-04	4.01E-01	0.00	
10	rs12778642	94,464,307		T	G	0.43	0.74	1.04E-06	9.83E-01	0.02	2.41E-04	4.06E-01	0.01	8.86E-05	6.18E-01	0.01	
10	rs4409764‡	101,284,237		T	G	0.49	0.49	1.90E-34	**1.74E-06**	0.80	1.59E-24	**1.12E-05**	0.59	2.49E-21	**4.01E-04**	0.44	IBD_U
10	rs3740415	104,232,716	*NFKB2*	G	A	0.45	0.75	8.47E-04	5.39E-01	0.00	6.79E-02	5.70E-01	0.00	8.85E-04	7.97E-01	0.00	
10	rs11195128‡	112,186,148	.	T	C	0.33	0.17	2.74E-09	**1.59E-08**	0.03	5.41E-11	**1.97E-11**	0.05	2.07E-03	1.38E-01	0.00	CD
10	rs111456533	126,439,381	*METTL10,FAM175B,RP11-12J10.3,FAM53B*	A	G	0.16	0.20	1.18E-09	3.61E-01	0.10	3.45E-06	2.86E-01	0.04	1.33E-06	8.87E-01	0.04	
10	rs10734105	133,172,119	.	G	A	0.31	0.18	4.46E-01	5.75E-01	0.00	3.45E-01	8.42E-01	0.00	9.18E-01	2.23E-01	0.00	
11	rs907611	1,874,072		A	G	0.31	0.21	1.06E-06	**2.47E-02**	0.02	3.71E-02	3.32E-01	0.00	1.36E-07	**9.71E-03**	0.02	
11	rs11229030	57,203,009	*RP11-872D17.8,SLC43A3*	C	T	0.39	0.18	5.51E-01	5.25E-01	0.00	2.98E-01	9.24E-01	0.00	7.37E-01	5.58E-01	0.00	
11	rs11229555	58,408,687	*CNTF*	T	G	0.25	0.21	2.59E-05	6.11E-02	0.01	1.12E-01	2.25E-01	0.00	9.13E-06	7.77E-02	0.02	
11	rs11230563	60,776,209	*CD5, GPR44, CD6*	T	C	0.35	0.18	1.95E-06	3.35E-01	0.01	5.77E-04	6.08E-01	0.00	2.40E-04	1.77E-01	0.00	
11	rs174537	61,552,680		T	G	0.33	0.33	5.32E-05	**4.58E-02**	0.01	3.65E-07	**3.29E-03**	0.04	1.03E-01	9.42E-01	0.00	
11	rs559928	64,150,370	*RPS6KA4*	T	C	0.18	0.13	1.75E-05	1.44E-01	0.01	4.52E-06	1.48E-01	0.01	4.69E-02	5.09E-01	0.00	
11	rs568617	65,653,242	*RELA, FOSL1, SIPA1*	T	C	0.19	0.44	1.75E-03	5.61E-01	0.01	1.13E-04	9.13E-01	0.03	7.30E-02	8.74E-02	0.00	
11	rs11235667	72,863,697	.	G	A	0.00	0.11	NA	6.38E-02	NA	NA	**6.97E-04**	NA	NA	2.44E-01	NA	
11	rs2155219	76,299,194		G	T	0.49	0.47	1.81E-28	2.11E-01	0.67	5.52E-23	7.26E-01	0.55	5.33E-14	1.40E-01	0.19	
11	rs6592362†	87,125,438	.	A	G	0.26	0.67	1.44E-03	**4.37E-02**	0.01	2.91E-02	4.18E-01	0.00	1.10E-03	**1.18E-02**	0.01	
11	rs483905	96,023,427		A	G	0.29	0.27	1.17E-04	7.12E-01	0.01	6.39E-01	8.28E-01	0.00	6.81E-07	2.78E-01	0.03	
11	rs561722	114,386,830		T	C	0.34	0.64	3.69E-06	8.25E-02	0.03	1.92E-01	9.34E-02	0.00	3.66E-10	2.05E-01	0.10	
11	rs566416	118,759,610	*CXCR5*	G	T	0.23	0.07	3.89E-01	**3.01E-02**	0.00	2.03E-02	1.56E-01	0.00	6.06E-01	**1.32E-02**	0.00	
11	rs11221332	128,380,974	*ETS1*	T	C	0.22	0.04	5.64E-08	3.01E-01	0.00	9.80E-05	3.07E-01	0.00	7.29E-06	3.13E-01	0.00	
12	rs7954567	6,491,125	*CD27, TNFRSF1A, LTBR*	A	G	0.31	0.05	2.69E-06	–	NA	4.69E-09	–	NA	1.64E-01	–	NA	
12	rs11054935†	12,648,843	*DUSP16*	G	A	0.27	0.09	2.47E-03	1.88E-01	0.00	1.81E-01	8.63E-01	0.00	6.46E-04	**1.30E-02**	0.00	
12	rs12422544	40,528,432		C	T	0.02	0.04	4.58E-14	2.50E-01	0.77	1.97E-15	**3.17E-02**	0.83	3.43E-04	9.09E-01	0.07	
12	rs4768236	40,756,472		C	A	0.34	0.58	2.83E-04	3.58E-01	0.01	1.87E-06	6.48E-01	0.03	6.75E-01	2.34E-01	0.00	
12	rs11168249	48,208,368	*RAPGEF3, SENP1*	C	T	0.46	0.07	7.19E-06	1.81E-01	0.00	1.88E-01	3.15E-01	0.00	3.18E-07	5.79E-01	0.00	

Table 15. Cont'd (5)

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
12	rs7134472	68,499,986	*IL22, IFNG, IL26*	A	G	0.38	0.00	2.12E-25	–	NA	3.49E-06	–	NA	1.70E-31	–	NA	
12	rs653178	112,007,756	*SH2B3*	T	C	0.49	0.00	2.14E-09	–	NA	6.95E-07	–	NA	1.46E-05	–	NA	
12	rs11064881	120,146,925		A	G	0.07	0.00	5.34E-05	–	NA	5.94E-05	–	NA	2.90E-03	–	NA	
13	rs17085007‡	27,531,267		C	T	0.18	0.19	6.87E-10	**7.12E-06**	0.09	3.60E-01	5.46E-01	0.00	1.21E-14	**7.88E-13**	0.24	UC
13	rs915286	40,695,992		G	A	0.44	0.76	9.77E-04	**2.22E-02**	0.00	4.65E-04	**1.11E-02**	0.01	2.24E-01	5.08E-01	0.00	
13	rs17061048	40,833,012		A	T	0.05	0.01	2.64E-07	3.77E-01	0.00	6.20E-04	3.82E-01	0.00	1.51E-06	7.74E-01	0.00	
13	rs941823	41,013,977		T	C	0.24	0.11	3.59E-07	**1.77E-03**	0.01	7.58E-03	**3.25E-02**	0.00	4.04E-07	**5.70E-03**	0.01	
13	rs7329174‡	41,558,110	*KBTBD6,KBTBD7,WBP4,ELF1*	G	A	0.01	0.24	4.21E-01	**5.65E-07**	0.32	–	**2.62E-06**	NA	4.46E-01	**1.02E-03**	0.13	IBD_U
13	rs80244186	42,917,861	*AKAP11*	C	T	0.14	0.14	6.46E-06	9.42E-01	0.02	3.66E-08	2.43E-01	0.05	2.23E-01	1.67E-01	0.00	
13	rs9525625	43,018,030	*TNFSF11*	T	C	0.48	0.82	1.60E-03	5.84E-01	0.00	3.72E-07	9.22E-01	0.01	8.25E-01	1.86E-01	0.00	
13	rs3764147	44,457,925		G	A	0.24	0.36	2.74E-08	1.91E-01	0.09	1.38E-13	**2.79E-03**	0.32	7.37E-02	1.78E-01	0.00	
13	rs2026029	49,595,331	*MLNR,FNDC3A*	A	G	0.33	0.45	3.02E-04	**1.29E-02**	0.01	2.58E-03	**2.79E-02**	0.01	9.75E-02	5.92E-02	0.00	
13	rs3742130	99,907,341	*EBI2*	A	G	0.22	0.05	1.34E-06	**4.82E-02**	0.00	1.03E-05	3.50E-01	0.00	3.05E-03	6.04E-02	0.00	
14	rs194749	69,273,905		C	T	0.22	0.30	–	6.22E-01	NA	–	1.86E-01	NA	–	6.60E-01	NA	
14	rs1569328	75,741,751	*FOS*	T	C	0.16	0.26	1.98E-05	**2.83E-04**	0.03	1.36E-07	**1.88E-03**	0.09	7.69E-02	**2.64E-02**	0.00	
14	rs8005161	88,472,595	*GPR65, GALC*	T	C	0.09	0.14	2.71E-11	**1.98E-04**	0.25	4.72E-12	**1.05E-02**	0.31	2.67E-05	**2.10E-04**	0.04	
15	rs16967103	38,899,190	*RASGRP1, SPRED1*	C	T	0.20	0.04	6.35E-03	2.22E-01	0.00	1.40E-06	1.45E-01	0.00	6.73E-01	4.19E-01	0.00	
15	rs28374715	41,563,950		G	A	0.26	0.00	1.31E-03	–	NA	5.99E-01	–	NA	3.27E-07	–	NA	
15	rs17293632	67,442,596	*SMAD3*	T	C	0.23	0.02	3.01E-21	7.64E-01	0.01	2.40E-19	8.89E-01	0.01	2.11E-08	3.83E-01	0.00	
15	rs7165170	91,181,489	*CRTC3*	C	A	0.19	0.18	3.32E-05	2.29E-01	0.01	5.82E-04	9.15E-01	0.01	3.63E-04	**1.76E-02**	0.01	
16	rs423674	11,373,405	*SOCS1*	T	G	0.20	0.04	9.05E-07	8.17E-01	0.00	1.04E-07	1.02E-01	0.00	6.48E-03	7.65E-02	0.00	
16	rs11641184‡	11,704,651	*LITAF*	A	C	0.48	0.41	2.79E-07	**5.78E-04**	0.04	4.29E-06	**3.04E-05**	0.03	1.77E-04	1.51E-01	0.01	CD
16	rs7404095	23,864,590		T	C	0.43	0.39	1.63E-07	**2.34E-02**	0.04	1.17E-03	3.28E-01	0.01	1.74E-07	**3.46E-03**	0.04	
16	rs26528	28,517,709	*IL27*	C	T	0.45	0.35	1.94E-14	**2.90E-03**	0.18	1.62E-11	**8.16E-03**	0.12	1.06E-06	**1.30E-02**	0.03	
16	rs11150589†	30,482,494	*ITGAL*	T	C	0.47	0.03	1.60E-05	3.26E-01	0.00	7.45E-02	**3.65E-02**	0.00	1.21E-06	5.83E-01	0.00	
16	rs78534766	50,335,074	*ADCY7*	A	C	0.01	0.00	9.67E-13	–	NA	1.20E-05	–	NA	1.35E-13	–	NA	
16	rs2066844	50,745,926	*NOD2*	T	C	0.06	0.00	1.42E-38	–	NA	6.26E-99	–	NA	9.45E-02	–	NA	
16	rs2066845	50,756,540	*NOD2*	C	G	0.02	0.00	3.39E-24	–	NA	3.93E-57	–	NA	7.15E-01	–	NA	
16	rs5743293	50,763,781	*NOD2*	D	I	–	–	–	–	NA	–	–	NA	–	–	NA	
16	rs1728785	68,591,230		A	C	0.23	0.19	2.67E-05	2.69E-01	0.01	4.17E-01	6.62E-01	0.00	3.76E-08	1.03E-01	0.04	
16	rs11548656	81,916,912	*PLCG2*	G	A	0.04	0.00	5.18E-11	–	NA	2.96E-05	–	NA	7.92E-08	–	NA	
16	rs10492862	82,867,456	*CDH13*	A	C	0.28	0.05	1.24E-05	–	NA	1.26E-09	–	NA	6.85E-01	–	NA	
16	rs2361755	86,009,686	*IRF8*	C	G	0.08	0.01	1.33E-06	–	NA	6.37E-05	–	NA	5.68E-03	–	NA	
17	rs2945412	25,843,643	*NOS2A, LGALS9*	G	A	0.41	0.69	2.68E-02	**5.55E-03**	0.00	1.50E-09	**4.23E-03**	0.07	3.67E-02	2.63E-01	0.00	
17	rs3091315	32,593,665	*CCL13, CCL2, CCL11, CCL1, CCL7*	G	A	0.28	0.62	6.74E-13	**7.78E-03**	0.22	3.76E-18	1.35E-01	0.50	2.08E-02	**2.39E-03**	0.00	
17	rs12946510‡	37,912,377	*IKZF3*	T	C	0.46	0.34	1.69E-26	**3.82E-04**	0.55	1.35E-16	7.25E-02	0.25	1.52E-16	**9.29E-05**	0.23	UC
17	rs12942547‡	40,527,544	*STAT3, STAT5B, STAT5A*	G	A	0.40	0.34	1.90E-17	**1.29E-09**	0.26	1.54E-11	**6.05E-09**	0.11	1.20E-10	**7.25E-04**	0.09	IBD_S
17	rs3853824	54,880,993		T	C	0.35	0.18	9.79E-06	**3.21E-02**	0.01	4.01E-06	1.16E-01	0.01	1.32E-02	**4.94E-02**	0.00	
17	rs1292053	57,963,537		G	A	0.44	0.65	2.04E-05	**1.30E-02**	0.01	1.12E-06	**1.08E-02**	0.03	1.13E-01	3.68E-01	0.00	
17	rs17780256	70,642,923		C	A	0.20	0.15	3.74E-11	6.42E-02	0.07	1.68E-05	3.64E-01	0.01	3.56E-10	**5.00E-02**	0.06	

Table 15. Cont'd (6)

Chr	SNP	Position (hg19)	GRAIL gene	Allele		A1 allele frequency		Inflammatory bowel disease			Crohn's disease			Ulcerative colitis			LR phenotype**
								European	Korean		European	Korean		European	Korean		
				A1	A2	European	Korean	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	$P_{combined}$ Δ	P_{meta} §	Power*	
17	rs17736589	76,737,118		G	A	0.20	0.04	8.51E-05	–	NA	4.17E-02	–	NA	1.28E-05	–	NA	
18	rs1893217	12,809,340	*PTPN2*	G	A	0.16	0.15	1.37E-15	**2.06E-02**	0.20	3.59E-16	**1.05E-02**	0.25	3.98E-06	3.21E-01	0.02	
18	rs7240004	46,395,022	*SMAD7*	G	A	0.38	0.42	4.30E-08	–	NA	3.10E-06	–	NA	1.19E-05	–	NA	
18	rs9319943	56,879,827		C	T	0.19	0.20	2.10E-02	1.85E-01	0.00	6.60E-05	7.41E-02	0.01	9.78E-01	9.01E-01	0.00	
18	rs727088	67,530,439	*CD226, DOK6*	G	A	0.48	0.29	2.21E-03	**3.84E-02**	0.00	5.43E-03	2.32E-01	0.00	8.13E-02	**4.06E-02**	0.00	
18	rs7236492	77,220,616	*NFATC1*	T	C	0.16	0.00	9.99E-04	–	NA	1.39E-02	–	NA	2.13E-02	–	NA	
19	rs2024092	1,124,031		A	G	0.21	0.20	5.00E-17	**2.99E-03**	0.28	1.04E-20	8.93E-02	0.43	1.48E-04	**8.16E-04**	0.01	
19	rs12720356	10,469,975	*TYK2, ICAM1, ICAM3*	C	A	0.09	0.00	1.44E-13	–	NA	2.09E-09	–	NA	2.46E-08	–	NA	
19	rs11879191	10,512,911	*TYK2, ICAM1, ICAM3*	A	G	0.21	0.40	1.53E-11	1.36E-01	0.26	2.83E-13	2.85E-01	0.40	3.89E-03	1.44E-01	0.01	
19	rs17694108	33,731,551	*CEBPG*	A	G	0.28	0.29	1.36E-12	**2.35E-02**	0.19	1.04E-09	8.05E-02	0.12	1.07E-08	**4.25E-02**	0.08	
19	rs587259	34,656,406	*LSM14A*	T	C	0.37	0.01	1.81E-01	–	NA	1.41E-03	–	NA	4.03E-01	–	NA	
19	rs4802307	46,849,806		T	G	0.30	0.04	1.90E-05	6.56E-01	0.00	3.59E-07	9.43E-01	0.00	9.92E-02	2.65E-01	0.00	
19	rs11083840	47,119,910	*PTGIR*	G	T	0.41	0.34	1.54E-05	4.00E-01	0.02	1.30E-01	9.48E-01	0.00	8.83E-07	1.44E-01	0.03	
19	rs516246	49,206,172	*SPHK2, DBP, IZUMO1*	T	C	0.47	0.01	5.01E-06	–	NA	3.58E-11	–	NA	3.56E-01	–	NA	
19	rs17771967	55,380,214	*NLRP2, NLRP7*	G	A	0.44	0.39	1.66E-03	3.19E-01	0.01	7.81E-01	8.60E-01	0.00	1.29E-06	1.01E-01	0.03	
20	rs4256018	6,093,889	*FERMT1*	G	T	0.28	0.59	1.23E-08	2.98E-01	0.08	2.48E-07	2.50E-01	0.06	1.58E-04	5.89E-01	0.02	
20	rs4243971	30,849,517	*HCK*	T	G	0.45	0.00	9.81E-06	–	NA	2.46E-04	–	NA	5.19E-03	–	NA	
20	rs6087990	31,349,908		C	T	0.40	0.89	1.20E-02	8.22E-01	0.00	1.15E-01	4.41E-01	0.00	2.13E-02	3.04E-01	0.00	
20	rs6088765	33,799,280	*PROCR*	G	T	0.44	0.32	8.47E-03	9.03E-01	0.00	6.83E-01	5.23E-01	0.00	6.85E-05	5.11E-01	0.01	
20	rs4812833	43,068,996	*HNF4A*	G	A	0.48	0.04	1.07E-06	8.16E-01	0.00	9.86E-01	7.35E-01	0.00	2.05E-13	2.63E-01	0.00	
20	rs6074022‡	44,740,196	*CD40, MMP9*	C	T	0.26	0.36	1.40E-06	**3.80E-04**	0.04	1.46E-07	**3.61E-05**	0.07	4.71E-02	1.87E-01	0.00	IBD_U
20	rs913678	48,955,424	*CEBPB, PTPN1, TMEM189-UBE2V1*	C	T	0.34	0.67	4.21E-07	2.18E-01	0.04	5.31E-03	9.30E-01	0.00	1.00E-05	8.70E-02	0.02	
20	rs259964	57,824,309		A	G	0.46	0.09	3.37E-09	1.17E-01	0.01	4.15E-07	1.66E-01	0.01	2.88E-04	2.66E-01	0.00	
20	rs6062504	62,348,907	*TNFRSF6B*	A	G	0.31	0.67	–	2.48E-01	NA	–	1.69E-01	NA	–	7.93E-01	NA	
21	rs2823286	16,817,938		A	G	0.29	0.13	4.43E-29	**8.03E-04**	0.28	5.99E-26	**1.01E-03**	0.25	1.59E-13	**1.93E-02**	0.05	
21	rs2284553	34,776,695	*IL10RB, IFNAR1, IFNGR2, IFNAR2*	A	G	0.40	0.31	7.40E-09	4.96E-01	0.05	1.14E-14	5.57E-01	0.18	1.21E-01	7.68E-01	0.00	
21	rs2836878	40,465,534		A	G	0.27	0.20	2.30E-29	9.05E-02	0.56	3.90E-07	6.10E-01	0.03	1.83E-32	**1.76E-02**	0.64	
21	rs7282490‡	45,615,741	*ICOSLG, AIRE*	G	A	0.39	0.59	1.85E-23	**6.85E-06**	0.50	1.90E-18	**1.42E-05**	0.36	5.33E-13	**4.26E-03**	0.16	IBD_U
22	rs2256609‡	21,925,017	*MAPK1*	G	A	0.19	0.35	3.95E-15	**5.83E-07**	0.44	1.53E-15	**1.51E-05**	0.51	9.90E-06	**4.99E-05**	0.04	IBD_U
22	rs5763767	30,493,882	*OSM, LIF*	A	G	0.46	0.29	4.17E-08	7.30E-02	0.03	7.91E-07	1.21E-01	0.02	3.21E-04	**4.44E-02**	0.01	
22	rs138788	35,729,721	*TOM1*	G	A	0.42	0.74	7.64E-06	5.10E-01	0.01	2.34E-01	6.53E-01	0.00	2.95E-08	4.96E-01	0.03	
22	rs4821544	37,258,503	*NCF4*	C	T	0.33	0.14	2.03E-03	5.59E-01	0.00	1.76E-08	8.66E-01	0.02	6.10E-01	2.04E-01	0.00	
22	rs2413583	39,659,773	*MAP3K7IP1*	T	C	0.17	0.04	4.60E-24	**2.16E-03**	0.05	7.69E-21	**1.92E-02**	0.04	3.15E-10	**1.12E-02**	0.01	
22	rs12627970	39,721,745	*MAP3K7IP1, ATF4*	G	A	0.20	0.82	1.70E-17	**6.69E-04**	0.21	1.37E-12	**1.75E-02**	0.12	4.54E-09	**6.46E-04**	0.05	
22	rs727563	41,867,377		C	T	0.22	0.42	6.52E-03	3.52E-01	0.01	8.20E-04	2.59E-01	0.01	1.45E-01	9.17E-01	0.00	
22	rs5771069	50,435,480	*PIM3,IL17REL,TTLL8*	A	G	0.49	0.50	1.08E-05	**4.23E-02**	0.02	5.95E-01	1.29E-01	0.00	2.47E-10	1.66E-01	0.09	

Chr, chromosome; hg19, human genome version 19; SNP, single nucleotide polymorphism.

Δ European data was from summary statistics of de Lange KM et al (ref. 16).

§ Fixed-effects meta-analysis *P* value using two discovery cohorts in Korean population.

* Power values were calculated using Quanto v1.2 (http://hydra.usc.edu/gxe).

** A total of 29 SNPs in 27 loci assigned to phenotype using likelihood ratio modeling as described previously Jostins et al (ref. 14).

‡ These 29 SNPs in 27 loci passed the *P* value threshold (0.05/276 independent SNPs following Bonferroni correction). *P* values of these 29 SNPs were marked in red, and SNPs with $P<0.05$ were marked in blue.

† SNPs with $P < 0.05$ and opposite direction of effect with European data.

Table 16. Associations of 7 SNPs previously reported in Asian IBD GWAS in Korean and European data

Phenotype	Locus	SNP	Position (hg19)	Candidate gene(s)	Risk allele	RAF		KOR*		EUR†	
						EAS	EUR	OR (95% CI)	*P*	OR (95% CI)	*P*
IBD	1q21	rs3766920	154,934,963	*PYGO2, SHC1*	A	0.050	0	1.50 (1.32-1.71)	4.94×10^{-10}	NA	NA
	16q23	rs8182227‡	80,790,593	*CDYL2*	C	0.114	0.092	1.20 (1.09-1.33)	2.07×10^{-4}	1.01 (0.96-1.05)	7.68×10^{-1}
	9p21	rs3731257	21,966,221	*CDKN2A-AS1, CDKN2A, CDKN2B-AS1, CDKN2B*	G	0.449	0.746	1.17 (1.08-1.26)	4.27×10^{-5}	1.02 (0.99-1.05)	2.56×10^{-1}
CD	4p14	rs6856616	38,325,036	*TBC1D1, KLF3*	C	0.232	0.060	1.46 (1.32-1.62)	2.74×10^{-13}	1.12 (1.05-1.19)	4.30×10^{-4}
	10q25	rs11195128	112,186,148	*SMNDC1, DUSP5*	T	0.148	0.320	1.50 (1.33-1.69)	1.97×10^{-11}	1.12 (1.08-1.16)	5.41×10^{-11}
	13q14	rs7329174	41,558,110	*SLC25A15, ELF1, WBP4*	G	0.247	0.018	1.28 (1.16-1.42)	2.62×10^{-6}	1.01 (0.96-1.05)	7.86×10^{-1}
	11q13	rs11235667	72,863,697	*ATG16L2, FCHSD2*	G	0.107	0	1.27 (1.11-1.46)	6.97×10^{-4}	NA	NA

CD, Crohn's disease; CI, confidence interval; EUR, European; hg19, human genome version 19; IBD, inflammatory bowel disease; KOR, Korean; RAF, risk allele frequency; OR, odds ratio; *P*, *P* value; Position, chromosome position; SNP, single nucleotide polymorphism;

*Meta-analysis result using fixed-effect model using two discovery cohorts in Korean population.

†Summary statistics of de Lange et al (ref.16).

‡A proxy SNP in high LD ($r^2 = 0.96$) with rs16953946.

The replication of only a small fraction of the established loci in our discovery samples may be attributed to the limited power of our study. An estimation of the statistical power of our GWAS samples for detecting the 276 European susceptibility SNPs based on the reported OR and allele frequency in the Korean population (Table 15) showed that our samples had limited power: 9 SNPs in 8 loci and 48 SNPs in 39 loci had >80% power at $P<1.81 \times 10^{-4}$ and $P<0.05$, respectively.

We also compared the effect sizes of 224 SNPs from the established loci between Asians and Europeans using Pearson's correlation coefficient. The comparison showed positive correlations in the direction of effects for CD and UC between the European and Korean populations with significant *P* values (Figure 8A and B) ($r = 0.60$ and $P = 7.30 \times 10^{-22}$ for CD; $r = 0.62$ and $P = 8.25 \times 10^{-24}$ for UC), which are consistent with the findings of previous studies.[15,25]

3.4. Cis-eQTL analysis using whole blood RNA-seq of Korean CD patients

To identify the cis-eQTL variants within 1 Mb on either side of the transcription start site of each gene, cis-eQTL analysis was performed using the GWAS and RNA sequencing data from the peripheral blood of 101 Korean CD patients. Applying the threshold of FDR < 0.05, we found 135,164 cis-eQTL, 104,900 eSNPs, and 3,816 eGenes which had at least one cis-eQTL (Table 17). Of the total 104,900 eSNPs, the number of the target genes was one for 83,848 eSNPs (79.9%), two for 15,508 eSNPs (14.8%), and over three for 5,544 eSNPs (5.3 %). The proportion of the eSNPs to the total SNPs in each chromosome was 0.6 − 3.5% (total = 1.6%), and the ratio of the eGenes to the total genes ranged from 14.4% to 23.0% (total = 17.6%). The distance from an eSNP to the transcription start site (TSS) of the target gene was ≤ 500 kb in 95.7% (129,333 cis-eQTL) and ≤ 250 kb in 86.9% (117,433 cis-eQTL) of the total 135,164 cis-eQTL, and locations of eSNPs were more likely to be near the TSS of their target genes (Figure 9). The gene biotypes of 3,816 eGenes were composed of 2,700 protein coding genes (70.8%), 418 pseudogenes (11.0%), 272 antisense RNAs (7.1%), 270 long intergenic non-coding RNAs (7.1%), 44 sense intronic non-coding RNAs (1.2%), and 43 processed transcripts (1.1%), and 69 other gene biotypes (1.8%) (Figure 10). To annotate the biological processes significantly related to the 3,816 eGenes, we performed GO enrichment analysis using web-based AmiGO2 (http://amigo.geneontology.org/amigo).[46] Of those, 1,051 eGenes were excluded from the analysis mainly due to being non-coding genes or being absent in the reference genes of the GO dataset. A total of 2,765 eGenes

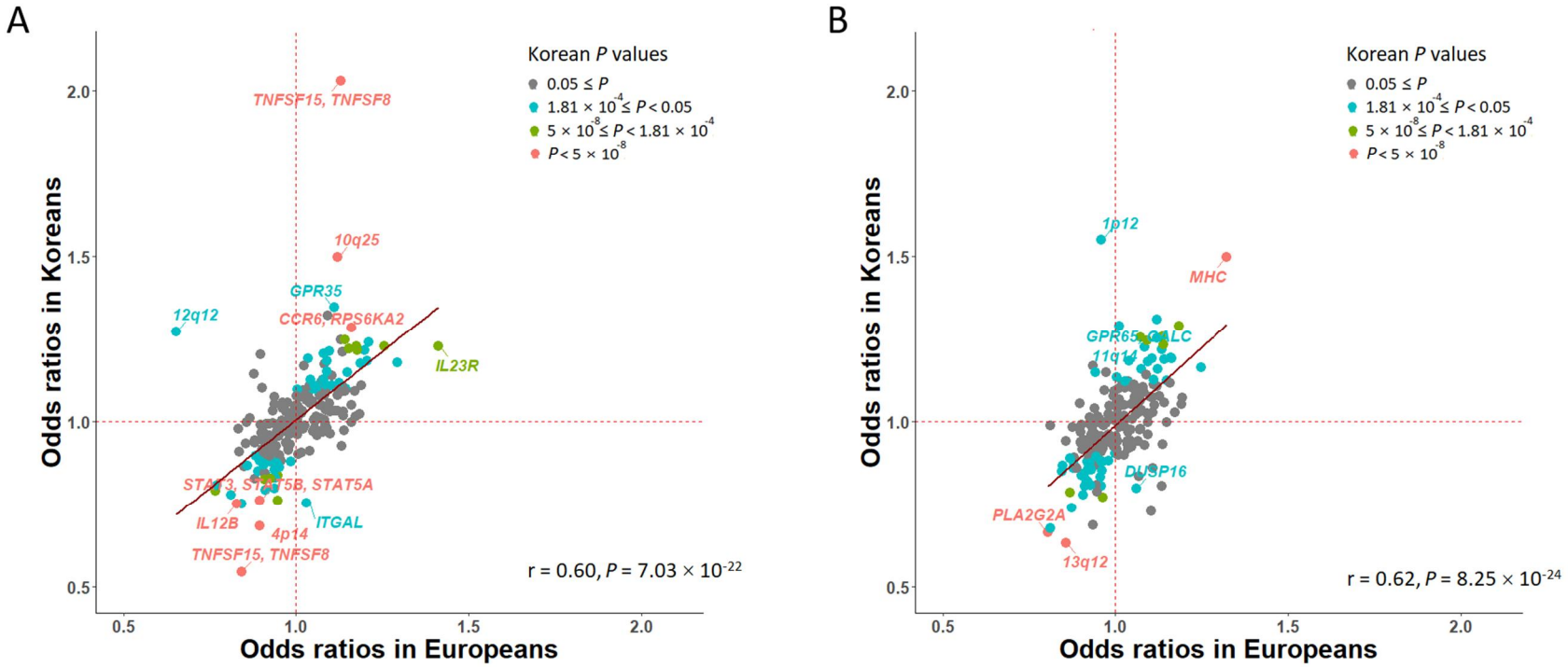

Figure 8. Comparison of the odds ratios (ORs) from CD or UC GWAS between Korean and European CD and UC. Each dot represents the OR of each SNP (total of 224 SNPs in 199 established loci**)** in (A) CD and (B) UC. Four colors denote the range of *P* values from the fixed-effects meta-analysis of cohort I and II. The Pearson correlation coefficient (r) with *P* value and regression line (red solid line) indicate the strength of the correlation between the ORs of two populations. Scatter plots and correlation coefficients with *P* values at the bottom-right corner are generated using R.

Table 17. The number of cis-eQTLs, eSNPs and eGenes in each chromosome

CHR	Cis-eQTLs	GWAS SNPs			eSNPs					eGenes		
		Genotyped	Imputed	Total	1 eQTL	2 eQTLs	>3 eQTLs	Total	Proportion (%)*	Genes	eGenes	Proportion (%)#
1	11,420	43,125	460,009	503,134	8,178	926	378	9,482	1.9	2,280	360	15.8
2	12,314	42,445	492,868	535,313	7,395	947	1,006	9,348	1.7	1,580	287	18.2
3	7,487	35,633	434,092	469,725	4,931	1,161	78	6,170	1.3	1,296	210	16.2
4	6,768	30,082	448,317	478,399	4,218	341	456	5,015	1.0	854	174	20.4
5	6,341	31,540	383,596	415,136	3,777	811	310	4,898	1.2	1,043	192	18.4
6	15,859	35,863	427,405	463,268	9,345	2,644	403	12,392	2.7	1,066	224	21.0
7	7,180	28,274	355,223	383,497	4,331	1,145	156	5,632	1.5	1,083	216	19.9
8	3,781	27,460	327,106	354,566	2,968	345	40	3,353	0.9	806	158	19.6
9	3,376	24,614	259,305	283,919	2,290	366	98	2,754	1.0	916	141	15.4
10	8,172	28,916	308,710	337,626	5,568	1,126	86	6,780	2.0	842	157	18.6
11	6,204	27,091	294,143	321,234	4,382	803	72	5,257	1.6	1,220	196	16.1
12	11,754	26,471	287,778	314,249	4,257	977	1,184	6,418	2.0	1,194	255	21.4
13	1,633	20,487	217,333	237,820	1,182	218	5	1,405	0.6	392	66	16.8
14	3,997	17,482	194,929	212,411	2,756	481	93	3,330	1.6	817	133	16.3
15	4,507	16,715	165,539	182,254	2,557	641	222	3,420	1.9	769	136	17.7
16	4,254	16,922	171,964	188,886	2,262	578	184	3,024	1.6	1,054	168	15.9
17	6,265	14,941	143,621	158,562	3,929	724	245	4,898	3.1	1,329	211	15.9
18	1,846	16,131	164,466	180,597	1,274	112	116	1,502	0.8	350	68	19.4
19	5,366	10,993	119,710	130,703	4,050	266	219	4,535	3.5	1,438	222	15.4
20	2,183	13,804	122,336	136,140	1,881	151	0	2,032	1.5	556	80	14.4
21	1,381	7,919	78,373	86,292	754	180	81	1,015	1.2	256	59	23.0
22	3,076	7,727	69,655	77,382	1,563	565	112	2,240	2.9	577	103	17.9
Total	135,164	524,635	5,926,478	6,451,113	83,848	15,508	5,544	104,900	1.6	21,718	3,816	17.6

CHR, chromosome; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; SNP, single nucleotide polymorphism;

*Proprotion of the eSNPs per chromosome GWA-SNPs.

#Proportion of eGenes in total number of genes used for eQTL analysis per chromosome.

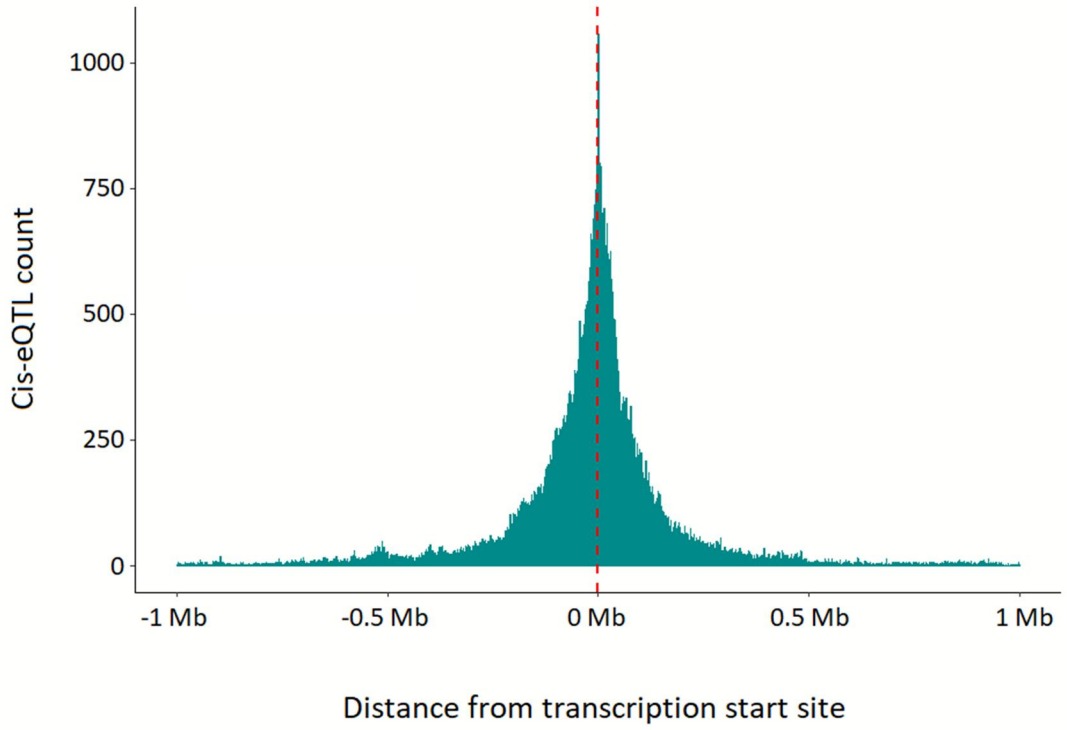

Figure 9. Location of eSNPs relative to the transcription start site of eGene. The distance in histogram was determined per 1 kb bins using 104,900 eSNPs and 3,816 eGenes.

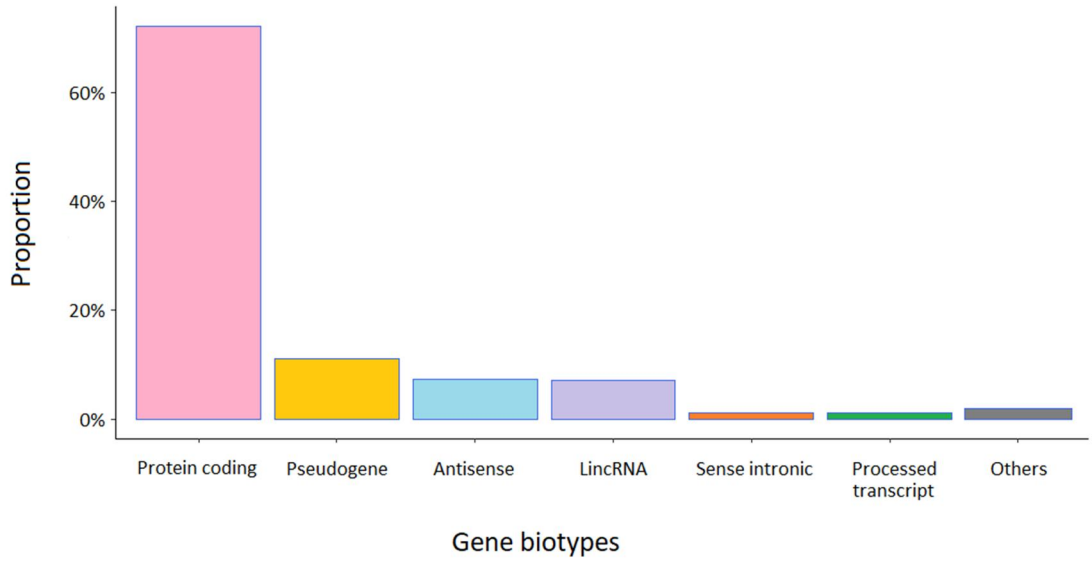

Figure 10. Histogram of the gene biotypes of 3,816 eGenes. The 3,816 eGenes included 2,700 protein coding genes (70.8%), 418 pseudogenes (11.0%), 272 antisense RNA (7.1%), 270 long intergenic non-coding RNA (7.1%), 44 sense intronic non-coding RNA (1.2%), 43 processed transcript (1.1%) and 69 other gene biotypes (1.8%).

were used as the input gene list for the GO enrichment analysis. Of the significantly shared GO terms with the Bonferroni-corrected *P* value < 0.05, granulocyte activation (GO:0036230) and neutrophil activation (GO:0042119) showed the top two highest fold enrichment values, respectively (Table 18).

3.5. Comparisons of three cis-eQTL databases

Using the Korean CD eQTL data and two existing cis-eQTL datasets derived from whole blood samples, including the Japanese eQTL[47] and GTEx,[48] we compared the direction of allelic effects of all the common SNP-gene pairs. Of the 135,164 eQTL in the Korean CD, 335,813 in the Japanese, and 1,052,542 in the GTEx datasets, the number of shared significant cis-eQTL (a threshold of q value ≤ 0.05) in each pair was 50,848 between the Korean CD and Japanese eQTL, 58,197 between the Korean CD and GTEx, and 120,158 between the GTEx and Japanese eQTL datasets (Figure 11). In total, 96.5-98.7 % of shared eGenes in each pair of the three cis-eQTL datasets showed the same direction of allelic effects. The proportion of shared eGenes with the opposite direction of allelic effects was 16 of 1,201 (1.3 %) shared eGenes in the pair of Korean CD-Japanese, 44 of 1,873 (2.3 %) shared eGenes in the pair of GTEx-Japanese cis-eQTL datasets and 56 of 1,581 (3.5 %) shared eGenes in the pair of Korean CD-GTEx (Table 19). Between the 16 and 56 eGenes with the opposite direction of allelic effects in the pair of Korean CD-Japanese or -GTEx, 9 eGenes overlapped.

3.6. Identification of target genes in the IBD susceptibility loci

We also performed colocalization analyses of the current meta-analysis of IBD, CD, and UC GWAS and whole blood eQTL data of the Korean CD, Japanese,[47] and GTEx[48] datasets using eCAVIAR.[49] eCAVIAR computes the CLPP score using Z statistics of GWAS and cis-eQTL data of 100 SNPs upstream and downstream of the reported lead SNP in the 54 susceptibility loci for IBD in Koreans. We identified 228 eGenes within 1 Mb window from the lead SNP in the 54 loci as target genes for cis-eQTL data of the Korean CD dataset. Applying thresholds of CLPP > 0.01 and total credible set posterior probability > 0.95, two loci including *TNFSF15* (TNF Superfamily Member 15) at 9q32 and *GPR35* (G-protein coupled receptor 35) at 2q37 were identified in colocalization analysis between the GWAS and cis-eQTL data of the Korean CD (Table 20).

Table 18. Gene Ontology enrichment analysis using 3,816 eGenes

GO biological process complete	Fold enrichment	*P* value
Granulocyte activation (GO:0036230)	1.92	1.99×10^{-6}
Neutrophil activation (GO:0042119)	1.91	3.64×10^{-6}
Neutrophil mediated immunity (GO:0002446)	1.91	5.04×10^{-6}
Neutrophil degranulation (GO:0043312)	1.89	1.58×10^{-5}
Neutrophil activation involved in immune response (GO:0002283)	1.89	1.27×10^{-5}
Myeloid leukocyte mediated immunity (GO:0002444)	1.86	1.52×10^{-5}
Leukocyte degranulation (GO:0043299)	1.85	3.28×10^{-5}
Myeloid cell activation involved in immune response (GO:0002275)	1.83	3.50×10^{-5}
Myeloid leukocyte activation (GO:0002274)	1.78	4.25×10^{-5}
Leukocyte activation involved in immune response (GO:0002366)	1.74	6.47×10^{-5}
Cell activation involved in immune response (GO:0002263)	1.73	9.79×10^{-5}
Regulated exocytosis (GO:0045055)	1.60	2.97×10^{-3}
Small molecule biosynthetic process (GO:0044283)	1.59	2.81×10^{-2}
Leukocyte activation (GO:0045321)	1.57	2.39×10^{-4}
Exocytosis (GO:0006887)	1.56	3.28×10^{-3}
Leukocyte mediated immunity (GO:0002443)	1.56	2.15×10^{-3}
Cell activation (GO:0001775)	1.53	2.01×10^{-4}
Organonitrogen compound biosynthetic process (GO:1901566)	1.52	5.12×10^{-7}
Cellular amide metabolic process (GO:0043603)	1.52	1.66×10^{-2}
Oxoacid metabolic process (GO:0043436)	1.48	5.90×10^{-3}
Organic acid metabolic process (GO:0006082)	1.47	7.49×10^{-3}
Carboxylic acid metabolic process (GO:0019752)	1.46	4.13×10^{-2}
Secretion (GO:0046903)	1.42	2.38×10^{-2}
Small molecule metabolic process (GO:0044281)	1.42	1.16×10^{-5}
Vesicle-mediated transport (GO:0016192)	1.38	1.13×10^{-4}
Cellular biosynthetic process (GO:0044249)	1.36	1.43×10^{-7}
Organic substance biosynthetic process (GO:1901576)	1.35	2.32×10^{-7}
Organic substance catabolic process (GO:1901575)	1.34	1.05×10^{-2}
Biosynthetic process (GO:0009058)	1.34	7.08×10^{-7}
Cellular catabolic process (GO:0044248)	1.33	1.16×10^{-2}
Establishment of localization in cell (GO:0051649)	1.33	1.77×10^{-2}
Phosphorus metabolic process (GO:0006793)	1.31	2.94×10^{-3}
Phosphate-containing compound metabolic process (GO:0006796)	1.29	1.11×10^{-2}
Cellular localization (GO:0051641)	1.28	1.60×10^{-2}
Cellular nitrogen compound metabolic process (GO:0034641)	1.27	3.02×10^{-5}
Transport (GO:0006810)	1.26	2.82×10^{-7}
Establishment of localization (GO:0051234)	1.26	2.23×10^{-7}
Heterocycle metabolic process (GO:0046483)	1.24	1.08×10^{-2}
Cellular aromatic compound metabolic process (GO:0006725)	1.23	1.62×10^{-2}
Organic cyclic compound metabolic process (GO:1901360)	1.23	1.04×10^{-2}
Organonitrogen compound metabolic process (GO:1901564)	1.21	6.35×10^{-6}
Localization (GO:0051179)	1.20	1.86×10^{-5}
Cellular metabolic process (GO:0044237)	1.20	8.16×10^{-10}
Organic substance metabolic process (GO:0071704)	1.19	1.31×10^{-9}
Nitrogen compound metabolic process (GO:0006807)	1.19	1.10×10^{-7}
Cellular macromolecule metabolic process (GO:0044260)	1.19	2.20×10^{-3}
Metabolic process (GO:0008152)	1.18	8.51×10^{-10}
Primary metabolic process (GO:0044238)	1.17	1.45×10^{-6}
Cellular process (GO:0009987)	1.07	3.00×10^{-4}
Biological_process (GO:0008150)	1.06	2.47×10^{-9}
Detection of chemical stimulus (GO:0009593)	0.34	1.00×10^{-5}
Sensory perception of chemical stimulus (GO:0007606)	0.33	2.99×10^{-6}
Detection of stimulus involved in sensory perception (GO:0050906)	0.33	2.22×10^{-6}
Detection of chemical stimulus involved in sensory perception (GO:0050907)	0.26	9.70×10^{-8}
Sensory perception of smell (GO:0007608)	0.25	2.44×10^{-7}
Detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	0.20	1.41×10^{-8}

GO, Gene Ontology;

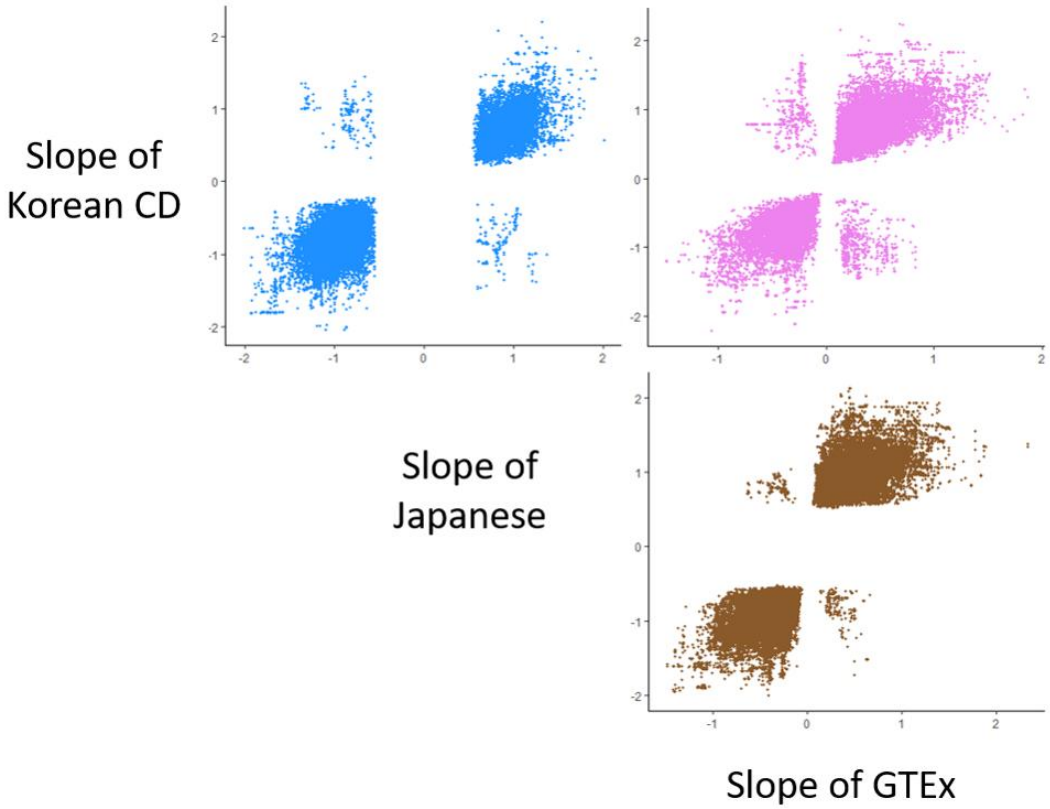

Figure 11. Scatter plots for comparison of the direction of allelic effects among the whole blood cis-eQTL data from the Korean CD, Japanese samples, and GTEx project. Each point on the scatter plots represents the allelic effect of a SNP to a gene expression. The scatter plots included 50,848 cis-eQTLs of 1,201 eGenes between the Korean CD and Japanese samples, 58,197 cis-eQTLs of 1,581 eGenes between the Korean CD and GTEx, and 120,158 cis-eQTLs of 1,873 eGenes between the GTEx and Japanese samples.

Table 19. Shared eGenes with the opposite direction of the allelic effect in each pair combination of the three cis-eQTL databases

Pair of cis-eQTL databases	Gene symbol	Description
Korean CD*-Japanese# (16 genes)	*CEACAM21*	Carcinoembryonic Antigen Related Cell Adhesion Molecule 21
	ZNF749	Zinc Finger Protein 749
	MED22	Mediator Complex Subunit 22
	RP11-705C15.2	.
	NOTCH2NL	Notch 2 N-Terminal Like
	FLCN	Folliculin
	PRR4	Proline Rich 4
	SURF6	Surfeit 6
	CYP2D7P1	Cytochrome P450 Family 2 Subfamily D Member 7 (Gene/Pseudogene)
	RP11-453E17.3	.
	TAP2	Transporter 2, ATP Binding Cassette Subfamily B Member
	CD300C	CD300c Molecule
	LRRC37A2	Leucine Rich Repeat Containing 37 Member A2
	UBA6-AS1	UBA6 Antisense RNA 1 (Head To Head)
	TYW1B	TRNA-YW Synthesizing Protein 1 Homolog B
	NUDT19	Nudix Hydrolase 19
Korean CD*-GTEx$^{\Delta}$ (56 genes)	*PRR4*	Proline Rich 4
	SURF6	Surfeit 6
	TAP2	Transporter 2, ATP Binding Cassette Subfamily B Member
	CEACAM21	Carcinoembryonic Antigen Related Cell Adhesion Molecule 21
	TYW1B	TRNA-YW Synthesizing Protein 1 Homolog B
	MED22	Mediator Complex Subunit 22
	ZNF749	Zinc Finger Protein 749
	NOTCH2NL	Notch 2 N-Terminal Like
	CD300C	CD300c Molecule
	HLA-A	Major Histocompatibility Complex, Class I, A
	HLA-C	Major Histocompatibility Complex, Class I, C
	RP11-347C12.2	.
	JRK	Jrk Helix-Turn-Helix Protein
	ZNF718	Zinc Finger Protein 718
	CCDC125	Coiled-Coil Domain Containing 125
	CEACAM3	Carcinoembryonic Antigen Related Cell Adhesion Molecule 3
	CHD1L	Chromodomain Helicase DNA Binding Protein 1 Like
	CCHCR1	Coiled-Coil Alpha-Helical Rod Protein 1
	RP11-497H16.6	.
	TCF19	Transcription Factor 19
	BARD1	BRCA1 Associated RING Domain 1
	TTC4	Tetratricopeptide Repeat Domain 4
	C22orf34	Chromosome 22 Open Reading Frame 34
	DDX11L10	DEAD/H-Box Helicase 11 Like 10
	MICB	MHC Class I Polypeptide-Related Sequence B
	AC016747.3	.
	RRP7A	Ribosomal RNA Processing 7 Homolog A
	RPA2	Replication Protein A2
	POLR1A	RNA Polymerase I Subunit A
	LIN28	Lin-28 Homolog A
	FANCA	FA Complementation Group A
	AFAP1	Actin Filament Associated Protein 1
	IQCG	IQ Motif Containing G
	GUF1	GUF1 Homolog, GTPase
	TOM1L2	Target Of Myb1 Like 2 Membrane Trafficking Protein
	KIAA1715	Lunapark, ER Junction Formation Factor
	TMOD3	Tropomodulin 3
	DZIP3	DAZ Interacting Zinc Finger Protein 3
	SFI1	SFI1 Centrin Binding Protein
	RAB7L1	RAB29, Member RAS Oncogene Family
	PHACTR4	Phosphatase And Actin Regulator 4
	CWF19L2	CWF19 Like Cell Cycle Control Factor 2

Table 19. Cont'd

	LINC00667	Long Intergenic Non-Protein Coding RNA 667
	MTHFS	Methenyltetrahydrofolate Synthetase
	MRPL10	Mitochondrial Ribosomal Protein L10
	DIP2A	Disco Interacting Protein 2 Homolog A
	RNF185	Ring Finger Protein 185
	LINGO2	Leucine Rich Repeat And Ig Domain Containing 2
Korean CD[*]-GTEx[Δ]	*COG4*	Component Of Oligomeric Golgi Complex 4
(56 genes)	*ACSF3*	Acyl-CoA Synthetase Family Member 3
	AKR1E2	Aldo-Keto Reductase Family 1 Member E2
	FAM21C	WASH Complex Subunit 2C
	CCL5	C-C Motif Chemokine Ligand 5
	CCZ1	CCZ1 Homolog, Vacuolar Protein Trafficking And Biogenesis Associated
	DHFR	Dihydrofolate Reductase
	MAN1B1	Mannosidase Alpha Class 1B Member 1
	TAGLN	Transgelin
	TWISTNB	TWIST Neighbor
	SERINC2	Serine Incorporator 2
	XXbac-BPG248L24.12	.
	CHI3L2	Chitinase 3 Like 2
	DZIP3	DAZ Interacting Zinc Finger Protein 3
	CTD-3214H19.6	Purkinje Cell Protein 2
	LRRC37A2	Leucine Rich Repeat Containing 37 Member A2
	FTCDNL1	Formiminotransferase Cyclodeaminase N-Terminal Like
	RP11-75L1.2	.
	RP11-705C15.2	.
	DDX11L10	DEAD/H-Box Helicase 11 Like 10
	AC004967.7	.
	RP4-717I23.3	.
	CDK10	Cyclin Dependent Kinase 10
	RP11-457M11.5	.
	NAPSB	Napsin B Aspartic Peptidase, Pseudogene
	GRM2	Glutamate Metabotropic Receptor 2
	DECR2	2,4-Dienoyl-CoA Reductase 2
	RP1-257A7.4	.
	EPHA1-AS1	EPHA1 Antisense RNA 1
GTEx[Δ]-Japanese[#]	*ZFAND2A*	Zinc Finger AN1-Type Containing 2A
(44 genes)	*RFWD3*	Ring Finger And WD Repeat Domain 3
	RRP7A	Ribosomal RNA Processing 7 Homolog A
	AK5	Adenylate Kinase 5
	DHFR	Dihydrofolate Reductase
	STYXL1	Serine/Threonine/Tyrosine Interacting Like 1
	NPIPA5	Nuclear Pore Complex Interacting Protein Family Member A5
	AC079325.6	.
	IGLC7	Immunoglobulin Lambda Constant 7
	CTD-2228K2.7	.
	KRT17P2	Keratin 17 Pseudogene 2
	HLA-DRB6	Major Histocompatibility Complex, Class II, DR Beta 6 (Pseudogene)
	RP11-419C5.2	.
	EFCAB12	EF-Hand Calcium Binding Domain 12
	SNX19	Sorting Nexin 19
	ASB1	Ankyrin Repeat And SOCS Box Containing 1
	NSG1	Neuronal Vesicle Trafficking Associated 1
	NBPF3	NBPF Member 3
	PCP2	Purkinje Cell Protein 2
	TUBB2A	Tubulin Beta 2A Class IIa
	CTC-457L16.2	.
	DDX1	DEAD-Box Helicase 1
	PRMT2	Protein Arginine Methyltransferase 2

eQTL, expression quantitative trait loci;

[*]Cis-eQTLs identified in whole blood tissues of 101 Korean CD patients (http://asan.crohneqtl.com/).

[#]Cis-eQTLs identified in whole blood tissues of 105 Japanese healthy volunteers (https://humandbs.biosciencedbc.jp/en/hum0099-v1#hum0099.v1.eqtl.v1).

[Δ]Cis-eQTLs identified in whole blood tissues of 369 postmortem samples (https://gtexportal.org/home/datasets).

Table 20. Colocalization analysis between current meta-analysis of GWAS and whole blood cis-eQTL databases using eCAVIAR

Phenotype	eQTL database	Locus	Lead SNP*	Target gene	SNP	LD (r^2)**	Position (hg19)	Allele		Credible set posterior probability#	CLPP$^{\Delta}$	GWAS§		Cis-eQTL	
								Risk	Non-risk			*P*	OR	*P*	Slope‡
IBD	Korean CD	2q37	rs3749172	*GPR35*	rs2953153	1.00	241,566,012	G	A	0.50	0.16	2.46×10^{-10}	1.29	4.40×10^{-6}	-0.61
					rs3749172	Exact	241,570,249	A	C	0.47	0.15	2.37×10^{-10}	1.29	4.40×10^{-6}	-0.61
		9q32	rs6478109	*TNFSF15*	rs6478109	Exact	117,568,766	G	A	0.50	0.14	7.83×10^{-38}	1.63	3.59×10^{-10}	-0.75
					rs7848647	1.00	117,569,046	C	T	0.50	0.14	8.09×10^{-38}	1.63	3.59×10^{-10}	-0.75
	GTEx	9q32	rs6478109	*TNFSF15*	rs7848647	1.00	117,569,046	C	T	0.59	0.09	8.09×10^{-38}	1.63	1.59×10^{-7}	-0.25
					rs6478109	Exact	117,568,766	G	A	0.41	0.06	7.83×10^{-38}	1.63	2.60×10^{-7}	-0.25
CD	Korean CD	2q37	rs3749172	*GPR35*	rs3749172	Exact	241,570,249	A	C	0.59	0.16	2.18×10^{-11}	1.38	4.40×10^{-6}	-0.61
					rs2953153	1.00	241,566,012	G	A	0.38	0.10	4.35×10^{-11}	1.37	4.40×10^{-6}	-0.61
		9q32	rs56211063	*TNFSF15*	rs6478109	0.51	117,568,766	G	A	0.48	0.02	3.82×10^{-59}	2.08	3.59×10^{-10}	-0.75
					rs7848647	0.51	117,569,046	C	T	0.45	0.02	4.17×10^{-59}	2.08	3.59×10^{-10}	-0.75
	GTEx	9q32	rs56211063	*TNFSF15*	rs7848647	0.51	117,569,046	C	T	0.59	0.07	4.17×10^{-59}	2.08	1.59×10^{-7}	-0.25
					rs6478109	0.51	117,568,766	G	A	0.41	0.05	3.82×10^{-59}	2.08	2.60×10^{-7}	-0.25

BP, base pair; CD; Crohn's disease; CLPP, co-localization posterior probability; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; LD, linkage disequilibrium; OR, odds ratio; *P*. *P* value; RAF, risk allele frequency; SNP, single nucleotide polymorphism;

Target genes with total credible set > 0.95 and significant cis-eQTL *P* value (FDR < 0.05).

*Lead SNP in the fixed-effects meta-analysis using cohort I and II.

**Linkage disequilibrium (LD) between the GWAS lead SNP and SNP identified using eCAVIAR in East Asians (http://www.1000genomes.org).

#Posterior probability of each causal variant within the credible set.

$^{\Delta}$Colocalization posterior probability indicates the level of colocalization (applied threshold > 0.01).

§Fixed-effects meta-analysis *P* and OR using cohort I and II.

‡Effect size of gene expression level of risk versus non-risk allele.

In the 9q32 locus, rs6478109 at 360 bp upstream of *TNFSF15* (Figure 12A) showed a significant *P* value in both GWAS (P=7.83 × 10^{-38} for IBD, and 3.82 × 10^{-59} for CD) and cis-eQTL (P=3.59 × 10^{-10}) with a CLPP score of 0.14 for IBD, and 0.02 for CD. The other causal SNP, rs7848647 at 640 bp upstream of *TNFSF15* is in complete LD (r^2 = 1) with rs6478109 in East Asians. Both risk allele G of rs6478109 and risk allele C of rs7848647 in CD GWAS were related to lower expression of *TNFSF15* than each protective allele A or T in the whole blood tissue (Figure 12B). A lead SNP in the 2q37 locus, rs3749172 is located in exon 6 (p.Ser294Arg) with a SIFT[62] score of 0.6 (tolerated) and a PolyPhen-2[63] score of 0 (benign) (Figure 12C). The CLPP score of rs3749172 was 0.15 for IBD and 0.16 for CD, suggesting that rs3749172 was a shared causal variant in the GWAS (P = 2.37 × 10^{-10} for IBD, and 2.18 × 10^{-11} for CD) and whole blood cis-eQTL (P = 4.40 × 10^{-6}) (Table 20). The other causal SNP, rs2953153 (CLPP = 0.16 for IBD, and 0.10 for CD) in intron 5 and 21,164 bp downstream of the TSS of *GPR35*, is in high LD (r^2 = 0.98) with rs3749172. Expression of *GPR35* in whole blood was down-regulated at risk allele A of rs3749172 and risk allele G of rs2953153, but up-regulated at protective allele C and A, respectively (Figure 12D).

The colocalization analysis of the 486 target genes between the current meta-analysis of GWAS in Koreans and GTEx only identified rs6478109 and rs7848647 at the 9q32 locus as the most likely causal SNP (Table 20), distinct from the European lead SNP of rs10114470 (r^2 = 0.77 with rs6478109) for IBD and CD in the largest GWAS.[16] These two causal SNPs showed significant CLPP scores and a 100% credible set posterior probability in IBD and CD GWAS (Table 20). Both rs2953153 and rs3749172 in the 2q37 locus had non-significant CLPP scores < 0.01 and cis-eQTL *P* values with FDR > 0.05 for *GPR35* expression. The lead SNP in the 2q37 locus from the largest GWAS of IBD and CD in Europeans was rs34236350, in moderate LD (r^2 = 0.34) in Europeans and complete LD (r^2=1) in East Asians with the Korean lead SNP rs3749172. The colocalization analysis of the 2q37 locus between the GWAS of European ancestry[16] and GTEx failed to identify shared causal SNPs.

In the colocalization analysis of the 164 target genes between the current meta-analysis of GWAS in Koreans and Japanese cis-eQTL data, there was no significant causal variants shared between GWAS and eQTL datasets. In 9q32 locus including *TNFSF15*, two causal SNPs of rs6478109 and rs7848647 showed a total of credible set posterior probabilities > 0.95; however,

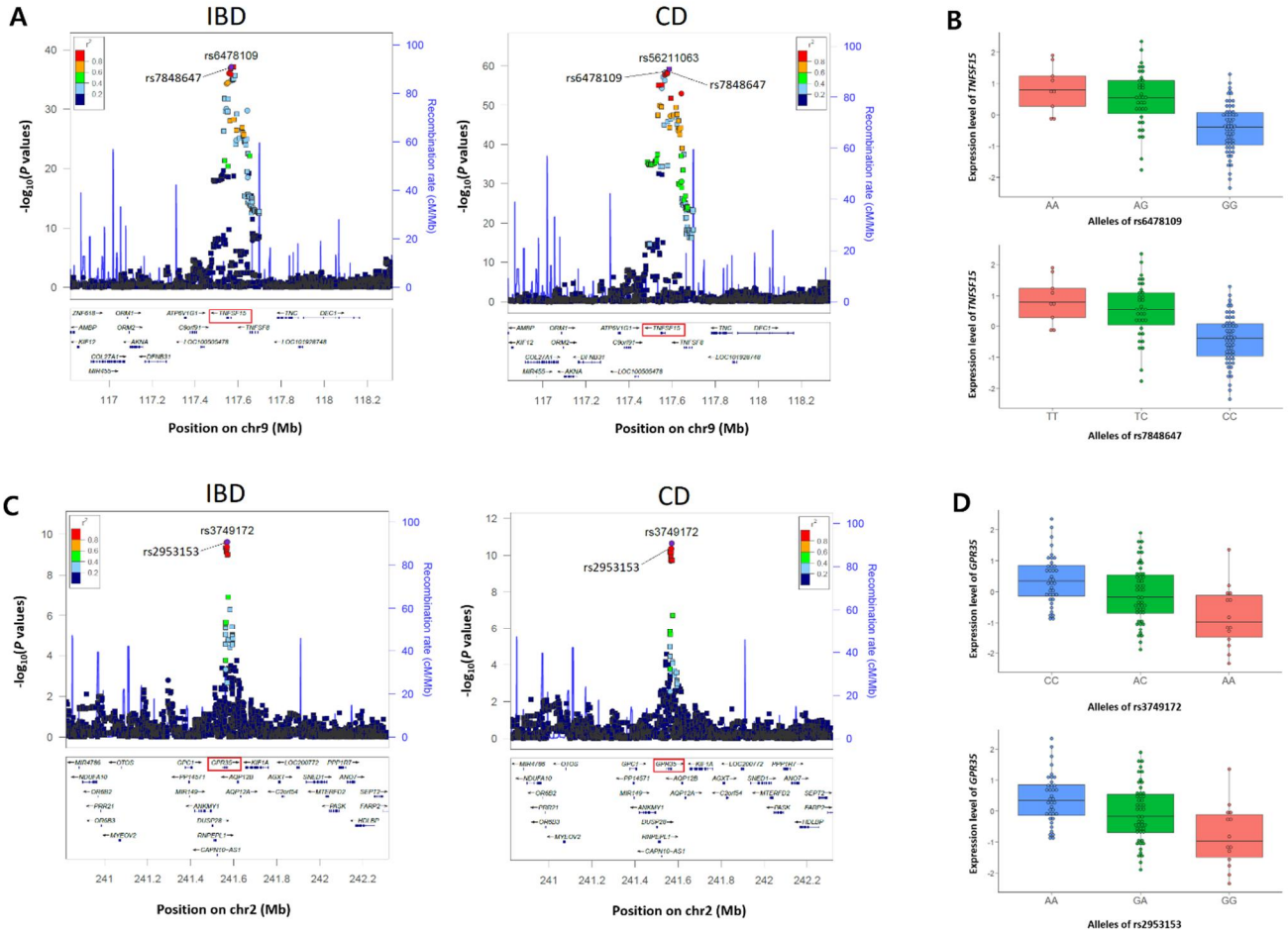

Figure 12. Two loci including *TNFSF15* and *GPR35* identified with colocalization analysis between GWAS and Korean CD eQTL data. (A and C) Regional association plots of the (A) *TNFSF15* locus at 9q32 using 4,791 SNPs for IBD (left), and 4,785 SNPs for CD (right), and (C) *GPR35* locus at 2q37 using 6,002 SNPs for IBD (left), and 6,088 SNPs for CD (right) are plotted according to their chromosomal positions (hg19) with -log10 *P* values from the current GWAS meta-analysis. All SNPs in the regional association plots are in ± 750 kb from the lead SNP, shown as purple circles in each plot. LD (r^2) indicated with colors was calculated using East Asian population data (JPT + CHB) for the 1000 Genomes. Regional association plots were generated using a web browser, LocusZoom (http://locuszoom.org/genform.php?type=yourdata). (B and D) Box plots of the (B) *TNFSF15* expression level according to alleles of rs6478109 and rs7848647, and (D) *GRP35* expression level according to alleles of rs3749172 and rs2953153. Small circles in the box plot indicate the normalized expression level using trimmed mean of M-values (TMM).

CLPP scores in the colocalization analysis of IBD, CD, and UC were below the threshold of 0.01. In the 2q37 locus including *GPR35*, both CLPP scores and credible set posterior probabilities of two causal SNPs (rs2953153 and rs3749172) were not significant in the colocalization analysis of IBD, CD, and UC

3.7. Pathway analysis based on GWAS

To identify biological processes associated with candidate genes for CD and UC, we conducted pathway analyses using the summary statistics obtained through the meta-analyses of cohort and as input. Pathway analysis using MAGMA v.1.07b.[57] for 9,976 Gene Ontology gene sets from MSigDB v.7.0[58] identified 30 and 5 pathways for CD and UC, respectively, with Bonferroni significance (0.05/9,976, $P < 5.01 \times 10^{-6}$) (Table 21). MHC class protein complex was the most significant pathway for both phenotypes ($P = 1.56 \times 10^{-9}$ for CD and 5.65×10^{-10} for UC). Pathways including T cell differentiation ($P = 2.02 \times 10^{-7}$ for CD and 1.48×10^{-1} for UC) and T helper 17 type immune response ($P = 3.29 \times 10^{-6}$ for CD and 2.58×10^{-2} for UC) were specifically significant for CD only.

We also performed additional pathway analyses using the summary statistics of the largest meta-analysis in the European population[16] and identified 157 and 29 pathways for CD and UC, respectively (Table 21). Then, we compared the list of the top 10 pathways for CD and UC between the Korean and European data (Figure 13A-D). In the case of CD, T cell differentiation-related pathways were significant in both populations. MHC class protein complex, antigen binding, and response to antigenic stimulus-related pathways were significant in the Korean population, whereas cytokine and transcription factor-related pathways were significant in the European population (Figure 13A and B). In the case of UC, MHC and antigen binding-related pathways identified in the Korean population were also significant in the European population (Figure 13C and D).

The Gene Ontology pathways for prioritized genes in 3 novel loci at 10q24, 19p13, and 6q22 failed to show Bonferroni significant *P* values. However, transcription factor binding ($P = 6.68 \times 10^{-4}$) including *LCOR* at the 10q24 locus, whole membrane ($P = 3.16 \times 10^{-2}$) including *MFSD12* at the 19p13 locus, and regulation of peptide secretion ($P = 9.47 \times 10^{-3}$) including *RFX6* at the 6q22 locus showed $P < 0.05$.

Table 21. Gene Ontology pathways identified in Korean and European populations

CD				UC			
Korean		European		Korean		European	
Pathway	*P*	Pathway	*P*	Pathway	*P*	Pathway	*P*
MHC class II protein complex	1.56E-09	**cytokine mediated signaling pathway**	**2.59E-19**	**MHC class II protein complex**	**5.65E-10**	**MHC class II protein complex**	**4.50E-13**
MHC protein complex	2.32E-09	**response to cytokine**	**6.49E-17**	**MHC class II receptor activity**	**6.73E-09**	**MHC protein complex**	**3.50E-10**
MHC class II receptor activity	**2.18E-08**	**T cell differentiation**	**7.16E-15**	**MHC protein complex**	**2.54E-08**	regulation of immune response	7.34E-10
antigen binding	1.42E-07	T cell activation	1.53E-14	**peptide antigen binding**	**5.71E-07**	positive regulation of immune system process	1.58E-09
lumenal side of membrane	1.61E-07	inflammatory response	1.55E-14	roof of mouth development	9.27E-07	interleukin 23 mediated signaling pathway	1.23E-08
T cell differentiation	**2.02E-07**	**defense response**	**8.45E-14**	-	-	lymphocyte differentiation	1.40E-08
regulation of immune system process	**2.36E-07**	lymphocyte differentiation	1.11E-13	-	-	regulation of immune system process	2.12E-08
cytokine mediated signaling pathway	**2.56E-07**	interleukin 23 mediated signaling pathway	3.74E-13	-	-	lymphocyte activation	2.44E-08
positive regulation of inflammatory response	3.49E-07	**regulation of immune system process**	**4.77E-13**	-	-	response to cytokine	3.68E-08
antigen receptor mediated signaling pathway	3.86E-07	cytokine production	5.37E-13	-	-	leukocyte differentiation	1.69E-07
alpha beta T cell differentiation	**8.70E-07**	leukocyte cell cell adhesion	6.18E-13	-	-	leukocyte cell cell adhesion	2.16E-07
T cell selection	**1.01E-06**	response to interferon gamma	6.46E-13	-	-	antigen receptor mediated signaling pathway	2.18E-07
regulation of cell activation	**1.21E-06**	regulation of dna binding transcription factor activity	6.89E-13	-	-	positive regulation of immune response	3.43E-07
positive regulation of inflammatory response to antigenic stimulus	1.24E-06	lymphocyte activation	9.93E-13	-	-	cytokine mediated signaling pathway	4.42E-07
regulation of chronic inflammatory response	1.65E-06	positive regulation of cell activation	1.59E-12	-	-	immune response regulating signaling pathway	4.43E-07
positive regulation of leukocyte mediated immunity	**1.74E-06**	**positive regulation of immune system process**	**1.82E-12**	-	-	**peptide antigen binding**	**6.91E-07**
regulation of leukocyte differentiation	**2.39E-06**	positive regulation of cytokine production	1.98E-12	-	-	immune response regulating cell surface receptor signaling pathway	7.10E-07
leukocyte differentiation	**3.02E-06**	positive regulation of dna binding transcription factor activity	3.62E-12	-	-	T cell receptor signaling pathway	8.88E-07
regulation of leukocyte mediated immunity	**3.03E-06**	**regulation of immune response**	**5.70E-12**	-	-	T cell activation	8.94E-07
regulation of inflammatory response to antigenic stimulus	3.09E-06	**leukocyte differentiation**	**7.00E-12**	-	-	**MHC class II receptor activity**	**1.30E-06**
T helper 17 type immune response	**3.29E-06**	regulation of adaptive immune response	7.92E-12	-	-	cell activation	1.48E-06
response to cytokine	**3.43E-06**	**alpha beta T cell differentiation**	**8.51E-12**	-	-	regulation of branching involved in lung morphogenesis	1.72E-06
defense response	**3.58E-06**	regulation of lymphocyte activation	1.21E-11	-	-	activation of immune response	2.14E-06
peptide antigen binding	3.61E-06	regulation of inflammatory response	1.42E-11	-	-	positive regulation of T cell proliferation	2.23E-06
T cell receptor signaling pathway	4.00E-06	alpha beta T cell activation	2.07E-11	-	-	defense response	3.08E-06
regulation of lymphocyte mediated immunity	**4.03E-06**	positive regulation of lymphocyte activation	2.13E-11	-	-	positive regulation of leukocyte cell cell adhesion	3.10E-06
positive regulation of immune system process	**4.25E-06**	positive regulation of molecular function	2.90E-11	-	-	alpha beta T cell activation	4.02E-06
regulation of defense response	**4.77E-06**	**regulation of defense response**	**2.95E-11**	-	-	T cell selection	4.95E-06
T helper 17 cell lineage commitment	**4.91E-06**	positive regulation of leukocyte cell cell adhesion	3.38E-11	-	-	regulation of defense response	4.96E-06
regulation of immune response	**4.98E-06**	regulation of cell cell adhesion	3.66E-11	-	-	-	-

Table 21. Cont'd (1)

CD				UC			
Korean		European		Korean		European	
Pathway	*P*	Pathway	*P*	Pathway	*P*	Pathway	*P*
-	-	lymphocyte activation involved in immune response	4.09E-11	-	-	-	-
-	-	adaptive immune response	4.26E-11	-	-	-	-
-	-	cd4 positive alpha beta T cell activation	4.33E-11	-	-	-	-
-	-	T cell differentiation involved in immune response	5.57E-11	-	-	-	-
-	-	positive regulation of cell cell adhesion	1.11E-10	-	-	-	-
-	-	interleukin 12 receptor binding	1.26E-10	-	-	-	-
-	-	regulation of T cell activation	1.27E-10	-	-	-	-
-	-	regulation of lymphocyte differentiation	1.92E-10	-	-	-	-
-	-	**regulation of cell activation**	**2.08E-10**	-	-	-	-
-	-	positive regulation of cell adhesion	2.59E-10	-	-	-	-
-	-	cytokine metabolic process	2.86E-10	-	-	-	-
-	-	T cell activation involved in immune response	3.48E-10	-	-	-	-
-	-	positive regulation of memory T cell differentiation	3.65E-10	-	-	-	-
-	-	response to molecule of bacterial origin	5.86E-10	-	-	-	-
-	-	positive regulation of intracellular signal transduction	1.40E-09	-	-	-	-
-	-	immune system development	1.54E-09	-	-	-	-
-	-	**regulation of lymphocyte mediated immunity**	**1.55E-09**	-	-	-	-
-	-	**T helper 17 type immune response**	**1.66E-09**	-	-	-	-
-	-	regulation of cell adhesion	1.93E-09	-	-	-	-
-	-	positive regulation of protein metabolic process	2.02E-09	-	-	-	-
-	-	positive regulation of immune response	2.15E-09	-	-	-	-
-	-	regulation of T cell differentiation	2.62E-09	-	-	-	-
-	-	cd4 positive or cd8 positive alpha beta T cell lineage commitment	3.46E-09	-	-	-	-
-	-	response to biotic stimulus	3.73E-09	-	-	-	-
-	-	positive regulation of lymphocyte differentiation	4.45E-09	-	-	-	-
-	-	**T helper 17 cell lineage commitment**	**4.77E-09**	-	-	-	-
-	-	regulation of hemopoiesis	6.75E-09	-	-	-	-
-	-	cell activation	6.81E-09	-	-	-	-
-	-	positive regulation of signaling	7.00E-09	-	-	-	-
-	-	positive regulation of lymphocyte mediated immunity	8.42E-09	-	-	-	-
-	-	cd4 positive alpha beta T cell lineage commitment	9.12E-09	-	-	-	-
-	-	positive T cell selection	9.68E-09	-	-	-	-
-	-	**regulation of leukocyte differentiation**	**9.78E-09**	-	-	-	-
-	-	alpha beta T cell lineage commitment	1.03E-08	-	-	-	-
-	-	response to bacterium	1.47E-08	-	-	-	-
-	-	positive regulation of interferon gamma production	1.58E-08	-	-	-	-
-	-	positive regulation of catalytic activity	1.85E-08	-	-	-	-
-	-	positive regulation of phosphorus metabolic process	2.56E-08	-	-	-	-
-	-	regulation of alpha beta T cell activation	2.78E-08	-	-	-	-
-	-	interferon gamma mediated signaling pathway	3.16E-08	-	-	-	-
-	-	production of molecular mediator of immune response	3.35E-08	-	-	-	-
-	-	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	3.44E-08	-	-	-	-
-	-	regulation of phosphorus metabolic process	3.91E-08	-	-	-	-
-	-	innate immune response	4.06E-08	-	-	-	-

Table 21. Cont'd (2)

CD				UC			
Korean		European		Korean		European	
Pathway	*P*	Pathway	*P*	Pathway	*P*	Pathway	*P*
-	-	immunological memory process	4.73E-08	-	-	-	-
-	-	regulation of response to stress	5.05E-08	-	-	-	-
-	-	interleukin 12 production	5.14E-08	-	-	-	-
-	-	positive regulation of adaptive immune response	5.81E-08	-	-	-	-
-	-	regulation of protein modification process	6.14E-08	-	-	-	-
-	-	positive regulation of biosynthetic process	7.08E-08	-	-	-	-
-	-	response to defenses of other organism involved in symbiotic interaction	7.17E-08	-	-	-	-
-	-	**T cell selection**	**7.42E-08**	-	-	-	-
-	-	positive regulation of interleukin 12 production	7.66E-08	-	-	-	-
-	-	myeloid cell differentiation	7.80E-08	-	-	-	-
-	-	receptor signaling pathway via stat	8.01E-08	-	-	-	-
-	-	response to lipid	8.38E-08	-	-	-	-
-	-	positive regulation of transcription by rna polymerase II	8.66E-08	-	-	-	-
-	-	positive regulation of hemopoiesis	8.87E-08	-	-	-	-
-	-	regulation of B cell activation	1.04E-07	-	-	-	-
-	-	cytokine production involved in immune response	1.08E-07	-	-	-	-
-	-	positive regulation of T helper 17 cell differentiation	1.15E-07	-	-	-	-
-	-	regulation of response to cytokine stimulus	1.28E-07	-	-	-	-
-	-	positive regulation of protein modification process	1.33E-07	-	-	-	-
-	-	positive regulation of interleukin 17 production	1.35E-07	-	-	-	-
-	-	protein phosphorylation	1.39E-07	-	-	-	-
-	-	positive regulation of T helper cell differentiation	1.53E-07	-	-	-	-
-	-	regulation of leukocyte proliferation	2.03E-07	-	-	-	-
-	-	positive regulation of nf kappab transcription factor activity	2.12E-07	-	-	-	-
-	-	positive regulation of leukocyte differentiation	2.19E-07	-	-	-	-
-	-	T cell lineage commitment	2.20E-07	-	-	-	-
-	-	tyrosine phosphorylation of stat protein	2.35E-07	-	-	-	-
-	-	regulation of production of molecular mediator of immune response	2.38E-07	-	-	-	-
-	-	positive regulation of immune effector process	2.46E-07	-	-	-	-
-	-	tumor necrosis factor superfamily cytokine production	2.68E-07	-	-	-	-
-	-	lymphocyte mediated immunity	2.87E-07	-	-	-	-
-	-	**positive regulation of leukocyte mediated immunity**	**2.91E-07**	-	-	-	-
-	-	regulation of receptor signaling pathway via stat	2.93E-07	-	-	-	-
-	-	positive regulation of rna biosynthetic process	3.00E-07	-	-	-	-
-	-	regulation of cell population proliferation	3.38E-07	-	-	-	-
-	-	**regulation of leukocyte mediated immunity**	**3.77E-07**	-	-	-	-
-	-	T cell mediated immunity	4.76E-07	-	-	-	-
-	-	regulation of response to external stimulus	4.86E-07	-	-	-	-
-	-	B cell activation	4.97E-07	-	-	-	-
-	-	immunological memory formation process	5.65E-07	-	-	-	-
-	-	activation of protein kinase activity	5.74E-07	-	-	-	-
-	-	cellular response to biotic stimulus	5.92E-07	-	-	-	-
-	-	regulation of transferase activity	6.79E-07	-	-	-	-
-	-	positive regulation of transferase activity	7.30E-07	-	-	-	-

Table 21. Cont'd (3)

CD				UC			
Korean		European		Korean		European	
Pathway	*P*	Pathway	*P*	Pathway	*P*	Pathway	*P*
-	-	positive regulation of cytokine biosynthetic process	8.04E-07	-	-	-	-
-	-	positive regulation of T helper 17 type immune response	8.09E-07	-	-	-	-
-	-	regulation of T helper cell differentiation	8.42E-07	-	-	-	-
-	-	positive regulation of kinase activity	8.50E-07	-	-	-	-
-	-	negative regulation of multicellular organismal process	8.56E-07	-	-	-	-
-	-	negative regulation of inflammatory response	8.57E-07	-	-	-	-
-	-	myeloid leukocyte differentiation	9.68E-07	-	-	-	-
-	-	lymphocyte migration	1.01E-06	-	-	-	-
-	-	negative regulation of regulatory T cell differentiation	1.11E-06	-	-	-	-
-	-	regulation of cd4 positive alpha beta T cell activation	1.19E-06	-	-	-	-
-	-	**MHC class II receptor activity**	**1.26E-06**	-	-	-	-
-	-	cell cell adhesion	1.37E-06	-	-	-	-
-	-	B cell activation involved in immune response	1.44E-06	-	-	-	-
-	-	positive regulation of gene expression	1.44E-06	-	-	-	-
-	-	activation of janus kinase activity	1.46E-06	-	-	-	-
-	-	immune response regulating signaling pathway	1.52E-06	-	-	-	-
-	-	i kappab kinase nf kappab signaling	1.52E-06	-	-	-	-
-	-	epithelial cell apoptotic process	1.56E-06	-	-	-	-
-	-	negative regulation of immune system process	1.62E-06	-	-	-	-
-	-	regulation of kinase activity	1.63E-06	-	-	-	-
-	-	positive regulation of receptor signaling pathway via stat	1.93E-06	-	-	-	-
-	-	T cell proliferation	2.01E-06	-	-	-	-
-	-	negative regulation of immune response	2.09E-06	-	-	-	-
-	-	regulation of intracellular signal transduction	2.33E-06	-	-	-	-
-	-	leukocyte proliferation	2.36E-06	-	-	-	-
-	-	regulation of T helper 17 cell lineage commitment	2.38E-06	-	-	-	-
-	-	regulation of alpha beta T cell differentiation	2.67E-06	-	-	-	-
-	-	regulation of T cell mediated immunity	2.78E-06	-	-	-	-
-	-	regulation of cd4 positive alpha beta T cell differentiation	2.79E-06	-	-	-	-
-	-	positive regulation of endoplasmic reticulum stress induced intrinsic apoptotic signaling pathway	3.30E-06	-	-	-	-
-	-	regulation of response to interferon gamma	3.36E-06	-	-	-	-
-	-	positive regulation of activated T cell proliferation	3.50E-06	-	-	-	-
-	-	positive regulation of cell fate commitment	3.54E-06	-	-	-	-
-	-	B cell differentiation	3.68E-06	-	-	-	-
-	-	positive regulation of cd4 positive alpha beta T cell activation	3.75E-06	-	-	-	-
-	-	negative regulation of cytokine production	3.97E-06	-	-	-	-
-	-	regulation of cell death	4.80E-06	-	-	-	-
-	-	positive regulation of alpha beta T cell activation	4.84E-06	-	-	-	-
-	-	immune effector process	4.95E-06	-	-	-	-

CD, Crohn's disease; *P*, *P* value; UC, ulcerative colitis.

Significant pathways with *P* value < 5.01×10^{-6} (0.05/9,976): 30 and 5 pathways for CD and UC respectively in Korean, and 157 and 29 pathways for CD and UC respectively in European population.

Bold: shared pathways between Korean and European population (19 pathways in CD, and 4 pathways in UC).

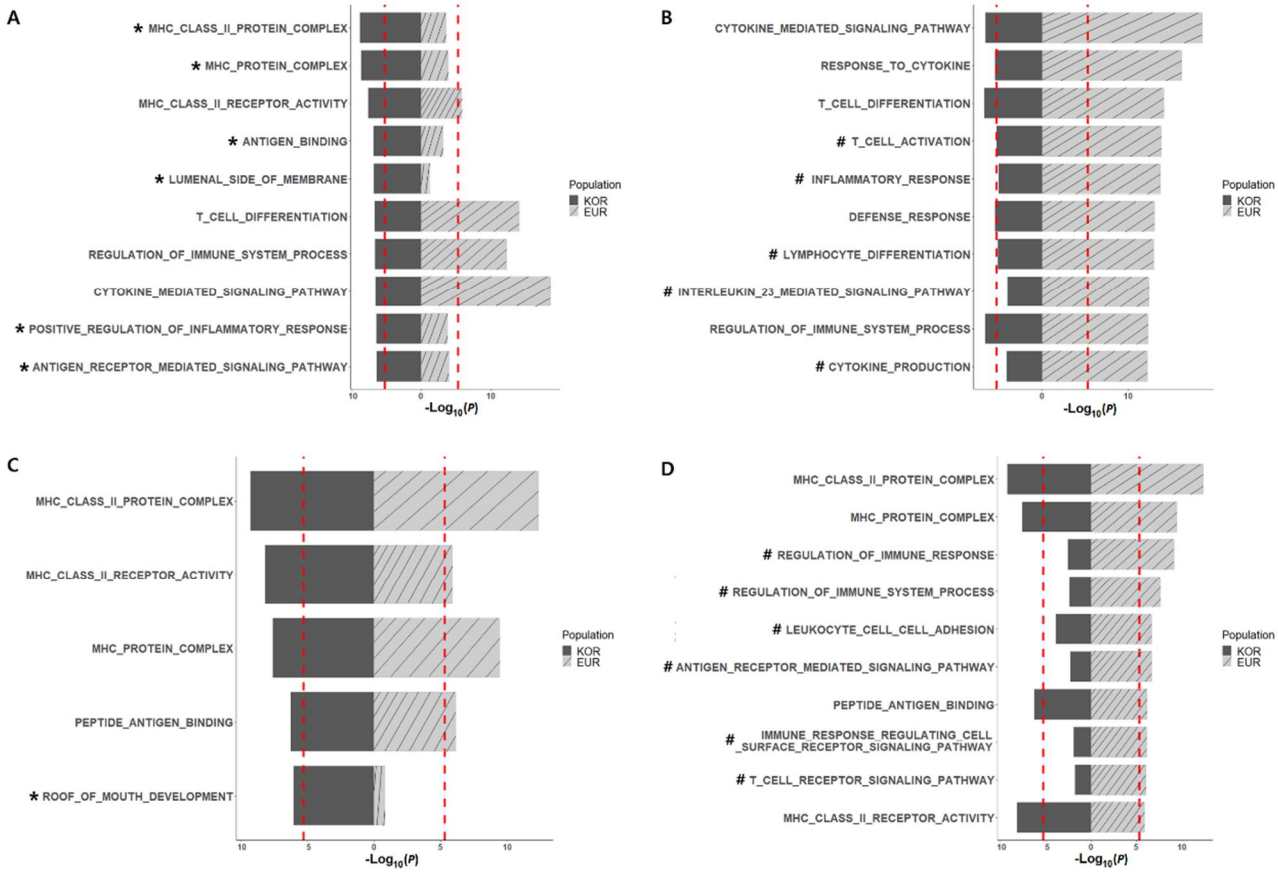

Figure 13. Comparison of biological pathways associated with Crohn's disease and ulcerative colitis between the Korean and European data. Top 10 pathways with Bonferroni significant $P < 5.01 \times 10^{-6}$ (0.05/9,976) (red dashed line) for CD in the (A) Korean and (B) European populations and for UC in the (C) Korean and (D) European populations. Each bar denotes the significance of each biological process (P value on a $-\log_{10}$ scale) in pathway analysis using MAGMA v.1.07b for 9,976 Gene Ontology (GO) sets. *Bonferroni significant pathways in Koreans only, #in Europeans only.

3.8. Estimation of variance explained by polygenic risk scores

We also calculated variance explained by the polygenic risk scores (PRS) for genome-wide significant variants using Korean versus European effect sizes. Despite the fact that the European studies had much larger samples size, PRS derived from Korean data explained up to 14.3 % of phenotype variance of CD whereas those derived from European data explained 9.9% (Table 22). However, for UC, the variance explained by PRS_{EUR} was far better than those explained by PRS_{KOR} (11.8% vs. 7.3%). To confirm this phenomenon, we used the summary statistics from the published Immunochip data in East Asians and Europeans[15] as base file to estimate variance explained by PRS for the target file of cohort I. The variance of CD explained by PRS based on East Asian data (PRS_{EAS}) explained up to 12.9%, whereas those based on European data explained 8.0% (Table 23). The variance of UC explained by PRS based on European data was higher than the variance explained by PRS_{EAS} (11.2% vs. 2.5%).

In order to further examine a larger variance being explained by PRS_{KOR} than by PRS_{EUR} for CD, we re-estimate PRS in the presence or absence of SNPs with larger effect sizes in Koreans. After removing all SNPs in the *TNFSF15* (chromosome 9: 117.4~118.7 Mb, hg19) or *HLA* (chromosome 6: 25~34 Mb, hg19) region with the largest effect size in CD GWAS (Figure 5B), the variance explained by PRS_{KOR} became similar to the variance explained by PRS_{EUR} (Table 24). When we removed both of them, the variance explained by PRS_{KOR} became smaller than the variance explained by PRS_{EUR}.

4. Discussion

In this extended GWAS of IBD in the Korean population, we successfully identified 1 novel locus for UC and 2 novel loci for CD and replicated 35 SNPs from 33 previously reported loci in the Korean population, indicating distinct as well as common pathways associated with IBD in Europeans and Asians. Of the 3 novel loci, 1 novel CD locus (rs2240751 at 19p13) was not replicated in the European data, suggesting the presence of population specific IBD susceptibility loci. The above 36 loci (3 novel + 33 confirmed) identified in the current study did not include previously reported 12 loci for CD and 16 loci for UC with 10 shared loci. Thus, a total of 54 IBD

Table 22. Variance explained by polygenic risk scores (PRS) with five different thresholds for including SNPs

Phenotype	Population	Threshold	Variance explained*	*P*	Number of SNPs used
CD	PRS_{KOR}	**5.00E-08**	**14.32%**	**2.22E-24**	**11**
		5.00E-07	15.31%	1.05E-25	18
		5.00E-06	12.85%	5.12E-22	37
		5.00E-05	9.19%	1.49E-16	116
		5.00E-04	7.38%	9.78E-14	644
	PRS_{EUR}	**5.00E-08**	**9.94%**	**9.07E-18**	**141**
		5.00E-07	10.50%	1.14E-18	210
		5.00E-06	8.45%	2.08E-15	308
		5.00E-05	5.88%	2.15E-11	572
		5.00E-04	1.99%	8.69E-05	1498
UC	PRS_{KOR}	**5.00E-08**	**7.27%**	**7.20E-16**	**3**
		5.00E-07	7.58%	5.36E-16	4
		5.00E-06	5.64%	1.26E-12	19
		5.00E-05	3.92%	3.13E-09	79
		5.00E-04	3.08%	9.28E-08	508
	PRS_{EUR}	**5.00E-08**	**11.81%**	**2.68E-24**	**76**
		5.00E-07	11.41%	1.41E-23	111
		5.00E-06	11.03%	7.18E-23	195
		5.00E-05	9.18%	1.44E-19	395
		5.00E-04	4.67%	5.73E-11	1179

CD, Crohn's disease; EUR, European; KOR, Korean; *P*, *P* value for variance explained; PRS, polygenic risk scores; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

PRS_{KOR} or PRS_{EUR} were calculated using previous Korean GWAS (ref.25) or the largest European ancestry IBD GWAS (ref.16), respectively.

*Variance explained was calculated by SNPs captured by PRS_{KOR} and PRS_{EUR}, respectively.

Bold: variance explained by PRS calculated by SNPs with *P* < 5E-08.

Table 23. Variance explained by polygenic risk scores (PRS) derived from East Asian or European Immunochip data

Phenotype	Population	Variance explained*	*P*	Number of SNPs used	Number of samples	
					Case	Control
CD	PRS_{EAS}	12.86%	4.79E-22	12	1,690	3,719
	PRS_{EUR}	7.99%	8.46E-15	220	14,594	26,715
UC	PRS_{EAS}	2.46%	3.76E-06	5	1,134	3,719
	PRS_{EUR}	11.18%	4.71E-23	126	10,679	26,715

CD, Crohn's disease; EAS, East Asian; EUR, European; *P*, *P* value for variance explained; PRS, polygenic risk scores; SNP, single nucleotide polymorphism; UC, ulcerative colitis. PRS_{EAS} or PRS_{EUR} were calculated using genome-wide significant variants identified in the Liu et al (ref.15).

*Variance explained was calculated by SNPs captured by PRS_{EAS} and PRS_{EUR}, respectively.

Table 24. Variance explained by polygenic risk scores (PRS) excluding *TNFSF15* or *MHC* region

Phenotype	Target data	Population	Variance explained*	*P*	Number of SNPs used
CD	Cohort I	PRS_{KOR}	14.32%	2.22E-24	11
		PRS_{EUR}	9.94%	9.07E-18	141
	Cohort I excluding *TNFSF15* region	PRS_{KOR}	7.73%	1.58E-14	7
		PRS_{EUR}	7.37%	8.37E-14	138
	Cohort I excluding *MHC* region	PRS_{KOR}	10.91%	3.66E-19	10
		PRS_{EUR}	9.69%	2.19E-17	140
	Cohort I excluding both *TNFSF15* and *MHC* region	PRS_{KOR}	3.38%	2.97E-07	6
		PRS_{EUR}	7.13%	2.01E-13	137

CD, Crohn's disease; EUR, Europeans; KOR, Koreans; *P*, *P* value for variance explained; PRS, polygenic risk scores; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

TNFSF15 region: chromosome 9, 117.4~118.7 Mb (hg19).

For the *MHC* region (chromosome 6: 25~34 Mb, hg19), only the most significant SNP was selected from Korean or European GWAS to minimize over-fitting.

PRS_{KOR} or PRS_{EUR} were calculated using genome-wide significant SNPs identified in the previous Korean GWAS (ref.25) or the largest European ancestry IBD GWAS (ref.16), respectively.

*Variance explained was calculated by SNPs captured by PRS_{KOR} and PRS_{EUR}, respectively.

susceptibility loci including 41 loci for CD and 44 loci for UC with 31 loci overlapping were identified in Koreans. Of these, 6 loci (*ATG16L2*, *MFSD12*, *SHC1*, *CDKN2A*, *ELF1*, *CDYL2*) were Asian-specific (Figure 14). Odds ratios of 224 SNPs from the 241 previously established loci for CD and UC showed significantly positive correlation between Korean and European population (Figure 8A and B), consistent with previous report that the genetic architectures of CD and UC were broadly similar across these populations.[15,25] Non-significant observations at many established European IBD loci might be mainly due to the sample size. The sample size of the current study is much smaller than those of European dataset (3,195 Korean cases and 4,419 controls vs. 25,042 European cases and 34,915 controls). Our study had limited power for detecting previously reported associations: 9 SNPs in 8 loci and 48 SNPs in 39 loci had >80% power at $P<1.81 \times 10^{-4}$ and $P<0.05$, respectively. We expect that future Asian studies with larger sample sizes would allow more loci to be replicated in this population.

We built the eQTL database (http://asan.crohneqtl.com/) including 135,164 cis-eQTLs and 3,816 eGenes based on a whole blood RNA-seq dataset from Korean patients with CD. We found that the most significantly enriched GO terms of the 3,816 eGenes was granulocyte activation, especially neutrophil degranulation. The role of neutrophils in the pathogenesis of CD has been much better described in a theory that the common predisposition to CD is a failure of the inflammatory response to tissue damage and innate immunity.[72] Failure of neutrophil migration to the inflammatory site is one of the mechanisms involved in granulomatous inflammation, characteristic of CD, which leads to an intense adaptive immune response and the tissues being infiltrated with large number of T cells. These cells as well as macrophages will react by producing cytokines that cause local inflammation and systemic symptoms. Following the colocalization analysis with Korean CD eQTL, Japanese or GTEx whole blood eQTL data, only the Korean CD eQTL identified functionally related target genes to CD at two previously established susceptibility loci: *TNFSF15* at 9q32 and *GPR35* at 2q37. Colocalization analysis with GTEx whole blood identified *TNFSF15* only. Despite the eQTL signals being consistent among the three eQTL data, we were able to colocalize the *GPR35* locus using only the Korean CD eQTL. The top signals of Korean and European CD GWAS at the *GPR35* locus were rs3749172 and rs34236350, respectively, in moderate LD ($r^2 = 0.34$) in Europeans while in high LD in Asians. Rs34236350 of European CD GWAS signal was eQTL for *GPR35* in GTEx sigmoid colon tissue only with

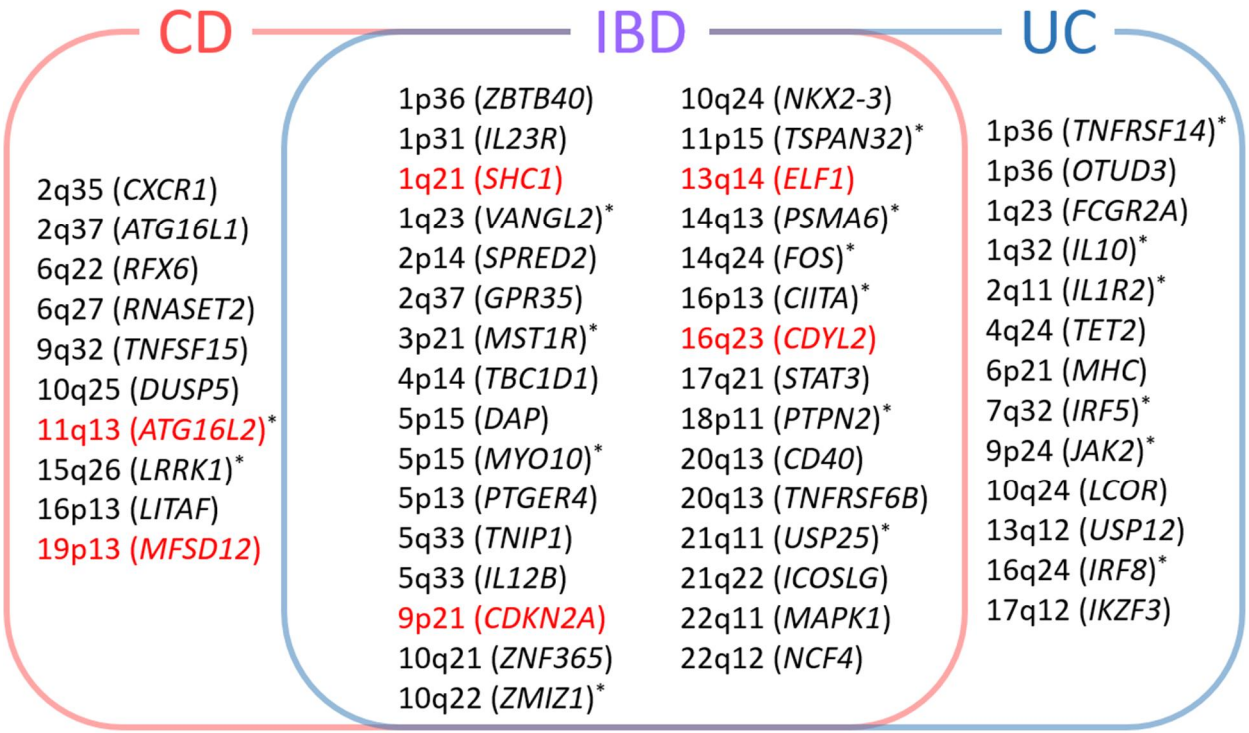

Figure 14. IBD susceptibility loci in Korean population. Including previous studies in Koreans (ref.19-25), IBD GWAS in Koreans identified 6 Asian-specific loci (red), and replicated 48 established loci. A total of 54 IBD susceptibility loci included 41 loci for CD, 44 loci for UC, and 31 shared loci between CD and UC. *18 loci previously reported in Koreans but not present in the current study included 12 loci for CD, 16 loci for UC, and 10 shared loci (ref.19-25).

up-regulation at risk allele, which is in the opposite direction relative to the whole blood eQTL of Koreans. Due to non-significant expression of *GPR35* in GTEx whole blood (FDR > 0.05), colocalization using GTEx did not identify *GPR35*, suggesting population-specific eQTL effects. Therefore, our data highlight the utility of building a population-specific data set, even of modest size.

TNFSF15, a TNF-like ligand for death receptor 3 (DR3) or decoy receptor 3 (DcR3), can induce nuclear factor kappa B or caspase activity, which allows it to play a role in both pro- and anti-inflammation.[73,74] Our data showed that the CD risk allele was associated with lower expression than protective allele, consistent with a recent finding involving peripheral blood monocytes derived from 90 Europeans.[75] Mining of the GTEx database in whole blood also showed the same direction of effect that we observed. However, reports on the effects of *TNFSF15* risk alleles for CD have been inconsistent. Earlier studies reported that the *TNFSF15* risk allele was associated with an increase in *TNFSF15* expression,[76-78] and *TNFSF15* expression was upregulated in intestinal tissues of IBD patients.[79-82] However, protective effect of the *TNFSF15-DR3* signaling on intestinal inflammation via maintenance of regulatory T cells has been also reported.[83,84] These studies highlighted that *TNFSF15* may be more pleiotropic than originally thought, costimulating lymphocytes that control both pro- and anti-inflammatory activities. *GPR35*, an orphan G protein-coupled receptor interacting with the sodium-potassium pump,[85] is highly expressed in the gastrointestinal tract from the stomach to rectum.[86] *GPR35* signaling in macrophages had a protective role during intestinal inflammation in mice.[87] Stimulation of *GPR35* promoted wound repair in the colon via enhancement of colonic epithelial cell migration and ameliorated DSS-induced colitis in mice.[88] Our data showed that the CD risk allele was associated with decreased expression of *GPR35*. We failed to identify target genes in the three novel or established susceptibility loci because of several limitations. First, because of our small sample size of GWAS and eQTL analysis, we had limited statistical power to detect colocalized signals with rare allele frequency or low effect size. Second, we performed eQTL analysis using the peripheral blood containing heterogeneous cell populations. The peripheral blood consists of multiple distinct cell types with specific gene regulatory profiles as shown by eQTL of isolated different blood cell types,[89] contributing to the low yield of causal genes identified for CD using the GWAS-eQTL integration approach.

Comparisons of the top 10 pathways for CD between the Korean and European data (Figure 13A and B) showed that T cell differentiation-related pathways were significant for CD in both populations. However, MHC class protein complex, antigen binding, and response to antigenic stimulus-related pathways were more significant in the Korean population, whereas cytokine and transcription factor-related pathways were more significant in the European population. In the case of UC, MHC and antigen binding-related pathways identified in the Korean population were also significant in the European population (Figure 13C and D). These findings from the pathway analysis were in line with our previous report that in HLA, the effects for CD were more population-specific than for UC.[90]

Recently PRS attracted increasing interest from clinical community for their predictive value for multiple common diseases.[29,91,92] One of the major challenges associated with clinical utility of PRS is that their accuracy is highly dependent on the study population represented in the training GWAS. Recent studies showed that PRS developed using data from Europeans can be less predictive in non-Europeans.[29-31] Indeed, even with our modest sample sizes, the variance of CD explained by PRS_{KOR} was higher than those by PRS_{EUR} (Table 22). Of note is that this large variance explained by PRS_{KOR} was not driven by overfitting, because we ensured that the training data (cohort II) and target data (cohort I) for PRS were independent (Methods). To further validate this phenomenon, we used the summary statistics from the Immunochip IBD GWAS.[15] This study published summary statistics for both East Asians and Europeans. Although the East Asians included Koreans, these individuals were not included in our current study. Thus, we can use these summary statistics to calculate the PRS of our target data (cohort I). When we calculated the PRS, we obtained a similar observation that PRS_{EAS} explained more of the variance than PRS_{EUR} (12.9% versus 8.0%) for CD and the variance of UC explained by PRS_{EUR} was higher than the variance explained by PRS_{EAS} (11.2% versus 2.5%) (Table 23). We wanted to further examine this phenomenon of a larger variance being explained by PRS_{KOR} than by PRS_{EUR} for CD. If the genetic structures had been similar between the two populations, the variance explained by PRS_{EUR} would have been larger than the variance explained by PRS_{KOR}, as shown in our UC data, because PRS_{EUR} was calculated based on a larger base data. Our observation in CD, however, suggested that the effect sizes might be population-specific in CD. We focused on the observation that this tendency was consistent regardless of the *P* value threshold for the PRS calculation (Table 22). This

suggested a possibility that the SNPs with large effects (small *P* values), which contributed to PRS regardless of the *P* value thresholds, might have population-specific effects. To test this hypothesis, we re-calculated the variance explained by PRS_{KOR} and PRS_{EUR} after removing the largest-effect loci, *TNFSF15* and *MHC*. When we removed each of the two, the variance explained by PRS_{KOR} became similar to the variance explained by PRS_{EUR} (Table 24). Furthermore, when we removed both of them, the variance explained by PRS_{KOR} became smaller than the variance explained by PRS_{EUR}. This showed that the large variance explained by PRS_{KOR} might have been driven by these two loci with population-specific effects. In the future, in order to achieve an equitable benefit of PRS for IBD patients, we will need to perform large genetic analyses in diverse populations and create tools for population genetic admixture.

Web Resources

Blood eQTL browser, https://genenetwork.nl/bloodeqtlbrowser/

edgeR, http://bioconductor.org/packages/release/bioc/html/edgeR.html

Ensembl Genome Browser, https://asia.ensembl.org/index.html

FastQC v0.11.7, https://www.bioinformatics.babraham.ac.uk/ projects/fastqc/

FM-summary, https://github.com/hailianghuang/FM-summary/blob/master/getCredible.r

GENCODE, https://www.gencodegenes.org/

GeneCards, https://www.genecards.org/

Genotype-Tissue Expression (GTEx) project, http://www.gtexportal.org/home/

IIBDGC, www.ibdgentics.org/

LocusZoom, http://locuszoom.sph.umich.edu//genform.php?type=yourdata

MAGMA, http://ctg.cncr.nl/software/magma

PLINK v1.9, https://www.cog-genomics.org/plink2

qvalue, https://github.com/StoreyLab/qvalue

R, http://www.r-project.org/

RegulomeDB v2, https://www.regulomedb.org/regulome-search/

The 1000 Genomes Project, https://www.internationalgenome.org/

UCSC Genome Browser, http://genome.ucsu.edu/

References

1. Podolsky DK. Inflammatory bowel disease. N Engl J Med 2002;347:417-29.

2. Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. Lancet 2007;369:1641-57.

3. Gasche C, Scholmerich J, Brynskov J, D'Haens G, Hanauer SB, Irvine EJ et al. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 2000;6:8-15.

4. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19 Suppl A:5A-36A.

5. Thia KT, Loftus EV, Sandborn WJ, Yang SK. An update on the epidemiology of inflammatory bowel disease in Asia. Am J Gastroenterol. 2008;103:3167-82.

6. Yang SK, Yun S, Kim JH, Park JY, Kim HY, Kim YH, et al. Epidemiology of inflammatory bowel disease in the Songpa-Kangdong district, Seoul, Korea, 1986-2005: a KASID study. Inflamm Bowel Dis 2008;14:542-9.

7. Ng SC. Epidemiology of inflammatory bowel disease: focus on Asia. Best Pract Res Clin Gastroenterol. 2014;28:363-72.

8. Park SJ, Kim WH, Cheon JH. Clinical characteristics and treatment of inflammatory bowel disease: A comparison of Eastern and Western perspectives. World J Gastroenterol 2014;20:11525-37.

9. Burisch J, Pedersen N, Čuković-Čavka S, Brinar M, Kaimakliotis I, Duricova D, et al. East-west gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. Gut 2014;63: 588-97.

10. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel disease. Gastroenterology 2011;140:1785-94.

11. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature 2011;474:307-17.

12. Yang H, McElree C, Roth MP, Shanahan F, Targan SR, Rotter JI. Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. Gut 1993;34:517-24.

13. Park JB, Yang SK, Byeon JS, Park ER, Moon G, Myung SJ, et al. Familial occurrence of inflammatory bowel disease in Korea. Inflamm Bowel Dis 2006;12:1146-51.

14. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119-24.

15. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979-86.

16. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in infalmmatory bowel disease. Nat Genet 2017;49:256-61.

17. Asano K, Matsushita T, Umeno J, Hosono N, Takahashi A, Kawaguchi T, et al. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. Nat Genet 2009;41: 1325-9.

18. Yamazaki K, Umeno J, Takahashi A, Hirano A, Johnson TA, Kumasaka N, et al. A genome-wide association study identifies 2 susceptibility loci for Crohn's disease in a Japanese population. Gastroenterology 2013;144:781-8.

19. Yang SK, Hong M, Zhao W, Jung Y, Tayebi N, Ye BD, et al. Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic suceptibility with Caucasians. Inflamm Bowel Dis 2013;19:954-66.

20. Yang SK, Hong M, Zhao W, Jung Y, Baek J, Tayebi N, et al. Genome-wide association study of Crohn's disease in Koreans reveal three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. Gut 2014;63:80-7.

21. Yang SK, Hong M, Choi H, Zhao W, Jung Y, Haritunians T, et al. Immunochip analysis identification of 6 additional susceptibility loci for Crohn's Disease in Koreans. Inflamm Bowel Dis 2015;21:1-7.

22. Ye BD, Choi H, Hong M, Yun WJ, Low HQ, Haritunians T, et al. Identification of ten additional susceptibility loci for ulcerative colitis through immunochip analysis in Koreans. Inflamm Bowel Dis 2016;22: 13-9.

23. Hong M, Ye BD, Yang SK, Jung S, Lee HS, Kim BM, 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;12:730-41.

24. Lee HS, Lee SB, Kim BM, Hong M, Jung S, Hong J, et al. Association of CDKN2A/CDKN2B with inflammatory bowel disease in Koreans. J Gastroenterol Hepatol 2018;33:887-93.

25. Yang SK, Hong M, Oh H, Low HQ, Jung S, Ahn S, et al. Identification of loci at 1q21 and 16q23 that affect susceptibility to inflammatory bowel disease in Koreans. Gastroenterology 2016;151:1096-9.e4.

26. Hu S, Uniken Venema WT, Westra HJ, Vich Vila A, Barbieri R, Voskuil MD et al. Inflammation status modulates the effect of host genetic variation on intestinal gene expression in inflammatory bowel disease. Nat Commun 2021;12:1122.

27. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17:392-406.

28. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219-24.

29. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51:584-91.

30. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry 2017;81:470-7.

31. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun 2019;10:3328.

32. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2011;9:179-81.

33. Howie B, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.

34. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007;39:906-13.

35. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47: 291-5.

36. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 2010;42:436-40.

37. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011;17:10-2.

38. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.

39. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 2019;47:766-73.

40. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 2012;28: 1530-2.

41. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.

42. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 2016;32:1479-85.

43. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 2012;7:500-7.

44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9.

45. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res 2017;45:D331-8.

46. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics 2009;25:288-9.

47. Ishigaki K, Kochi Y, Suzuki A, Tsuchida Y, Tsuchiya H, Sumitomo S, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. Nat Genet 2017;49:1120-5.

48. GTEx consortium. Genetic effects on gene expression across human tissues. Nature 2017;550:204-13.

49. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. Am J Hum Genet 2016;99:1245-60.

50. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans-eQTLs as putative drivers of known disease associations. Nat Genet 2013;45:1238–43.

51. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.

52. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:930-4.

53. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790-7.

54. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res 2020;48:682-8.

55. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res 2021;49:1046-57.

56. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47:1005-12.

57. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11:e1004219.

58. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545-50.

59. Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. Gigascience 2019;8:1-6.

60. Fernandes I, Bastien Y, Wai T, Nygard K, Lin R, Cormier O, et al. Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. Mol Cell 2003;11:139-50.

61. Wong K, Park HT, Wu JY, Rao Y. Slit proteins: molecular guidance cues for cells ranging from neurons to leukocytes. Curr Opin Genet Dev 2002;12:583-91.

62. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012;40(Web Server issue):W452-7.

63. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.

64. Madej MG, Dang S, Yan N, Kaback HR. Evolutionary mix-and-match with MFS transporters. Proc Natl Acad Sci USA 2013;110:5870-4.

65. Yu R, Broady R, Huang Y, Wang Y, Yu J, Gao M, et al. Transcriptome analysis reveals markers of aberrantly activated innate immunity in vitiligo lesional and non-lesional skin. PLoS One 2012;7:e51040.

66. Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, Bowman SL, et al. Loci associated with skin pigmentation identified in African populations. Science 2017;358(6365):eaan 8433.

67. Aftab S, Semenec L, Chu JS, Chen N. Identification and characterization of novel human tissue-specific RFX transcription factors. BMC Evol Biol 2008;8:226.

68. Huang Q, Whitington T, Gao P, Lindberg JF, Yang Y, Sun J, et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. Nat Genet 2014;46:126-35.

69. Spisák S, Lawrenson K, Fu Y, Csabai I, Cottman RT, Seo JH, et al. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncodong GWAS variants. Nat Med 2015;21:1357-63.

70. Gehart H, van Es JH, Hamer K, Beumer J, Kretzschmar K, Dekkers JF, et al. Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. Cell 2019;176:1158-73.

71. Piccand J, Vagne C, Blot F, Meunier A, Beucher A, Strasser P, et al. Rfx6 promotes the differentiation of peptide-secreting enterendocrine cells while repressing genetic programs controlling serotonin production. Mol Metab 2019;29:24-39.

72. Segal AW. Studies on patients establish Crohn's disease as a manifestation of impaired innate immunity. J Intern Med 2019;286:373-88.

73. Zhai Y, Ni J, Jiang GW, Lu J, Xing L, Lincoln C et al. VEGI, a novel cytokine of the tumor necrosis factor family, is an angiogenesis inhibitor that suppresses the growth of colon carcinomas in vivo. FASEB J 1999;13:181-9.

74. Migone TS, Zhang J, Luo X, Zhuang L, Chen C, Hu B et al. TL1A is a TNF-like ligand for DR3 and TR6/DcR3 and functions as a T cell costimulator. Immunity 2002;16:479-92.

75. Richard AC, Peters JE, Savinykh N, Lee JC, Hawley ET, Meylan F, et al. Reduced monocyte and macrophage TNFSF15/TL1A expression is associated with susceptibility to inflammatory bowel disease. PLoS Genet 2018;14:e1007458.

76. Sun R, Hedl M, Abraham C. TNFSF15 Promotes Antimicrobial Pathways in Human Macrophages and These Are Modulated by TNFSF15 Disease-Risk Variants. Cell Mol Gastroenterol Hepatol 2021;11:249-72.

77. Zucchelli M, Camilleri M, Andreasson AN, Bresso F, Dlugosz A, Halfvarson J et al. Association of TNFSF15 polymorphism with irritable bowel syndrome. Gut 2011;60:1671-77.

78. Kakuta Y, Ueki N, Kinouchi Y, Negoro K, Endo K, Nomura E et al. TNFSF15 transcripts from risk haplotype for Crohn's disease are overexpressed in stimulated T cells. Hum Mol Genet 2009;18:1089-98.

79. Jin S, Chin J, Seeber S, Niewoehner J, Weiser B, Beaucamp N et al. TL1A/TNFSF15 directly induces proinflammatory cytokines, including TNFα, from CD3+CD161+ T cells to exacerbate gut inflammation. Mucosal Immunol 2013;6:886-99.

80. Bamias G, Kaltsa G, Siakavellas SI, Gizis M, Margantinis G, Zampeli E et al. Differential expression of the TL1A/DcR3 system of TNF/TNFR-like proteins in large vs. small intestinal Crohn's disease. Dig Liver Dis 2012;44:30-6.

81. Bamias G, Martin C 3rd, Marini M, Hoang S, Mishina M, Ross WG et al. Expression, localization, and functional activity of TL1A, a novel Th1-polarizing cytokine in inflammatory bowel disease. J Immunol 2003;171:4868-74.

82. Prehn JL, Mehdizadeh S, Landers CJ, Luo X, Cha SC, Wei P et al. Potential role for TL1A, the new TNF-family member and potent costimulator of IFN-gamma, in mucosal inflammation. Clin Immunol 2004;112:66-77.

83. Taraban VY, Slebioda TJ, Willoughby JE, Buchan SL, James S, Sheth B et al. Sustained TL1A expression modulates effector and regulatory T-cell responses and drives intestinal goblet cell hyperplasia. Mucosal Immunol 2011;4:186-96.

84. Jia LG, Bamias G, Arseneau KO, Burkly LC, Wang EC, Gruszka D et al. A Novel Role for TL1A/DR3 in Protection against Intestinal Injury and Infection. J Immunol 2016;197:377-86.

85. Schneditz G, Elias JE, Pagano E, Zaeem Cader M, Saveljeva S, Long K et al. GPR35 promotes glycolysis, proliferation, and oncogenic signaling by engaging with the sodium potassium pump. Sci Signal 2019;12:eaau9048.

86. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A et al. Tissue-based map of the human proteome. Science 2015;347:1260419

87. Kaya B, Doñas C, Wuggenig P, Diaz OE, Morales RA, Melhem H et al. Lysophosphatidic Acid-Mediated GPR35 Signaling in CX3CR1+ Macrophages Regulates Intestinal Homeostasis. Cell Rep 2020;32:107979.

88. Tsukahara T, Hamouda N, Utsumi D, Matsumoto K, Amagase K, Kato S. G protein-coupled receptor 35 contributes to mucosal repair in mice via migration of colonic epithelial cells. Pharmacol Res 2017;123:27-39.

89. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 2012;44:502-10.

90. Han B, Akiyama M, Kim KK, Oh H, Choi H, Lee CH, et al. Amino acid position 37 of HLA-DRβ1 affects susceptibility to Crohn's disease in Asians. Hum Mol Genet 2018;27:3901-10.

91. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet 2019;28:R133-42.

92. Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell 2020;182:1198-213.

국문요약

염증성장질환은 위장관에 만성 염증을 일으키는 질환으로 크론병과 궤양성 대장염을 포함한다. 크론병은 위장관 전체에 전층 염증을 유발하며, 궤양성 대장염은 대장 조직에 국한된 점막염을 일으킨다. 염증성장질환은 유전적으로 취약한 개체에서 장내 세균총에 대한 점막 면역 반응이 잘 조절 되지 않아 발생하는 것으로 생각된다.

현재까지 서양인에서 수행된 염증성장질환에 대한 전장유전체연관분석 데이터의 메타분석으로 240 개가 넘는 감수성 유전좌위를 발굴하여 염증성장질환 관련 유전학에 대한 이해가 향상되었다. 그러나 발굴된 변이들로 설명할 수 있는 염증성장질환의 유전력은 일부에 지나지 않는다. 더욱이, 염증성장질환의 임상적 특성이 인종에 따라 차이가 있음에도 불구하고 서양인 외의 인종에서 수행된 유전학 연구는 극히 제한적이었다.

동양인에서 새로운 염증성장질환 감수성 유전좌위를 발굴하기 위해 한국인 염증성장질환 환자 1,726 명과 대조군 378 명 시료의 유전형을 Infinium Asian Screening Array-24 v1.0 (Illumina)로 분석하여 전장유전체연관분석 연구를 수행하였고, 한국인 염증성장질환 환자 1,469 명과 대조군 4,041 명으로 구성된 기존 전장유전체연관분석 데이터와 역분산 고정효과 모델을 이용한 메타분석을 수행하였다. $P_{meta} < 1 \times 10^{-6}$ 기준을 적용하여 새로운 감수성 유전좌위 후보 10 개를 선정하여, 추가 염증성장질환 환자 1,088 명과 대조군 845 명 시료에서 재현 연구를 수행하였다. 전장 유전체 연관 분석 데이터 두개를 메타 분석하여 궤양성 대장염 감수성 유전좌위 10q24 에 있는 rs76227733 ($P_{combined} = 6.56 \times 10^{-9}$), 크론병 감수성 유전좌위 19p13 에 있는 rs2240751 ($P_{combined} = 3.03 \times 10^{-8}$)과 6q22 에 있는 rs6936629 ($P_{combined} = 3.63 \times 10^{-8}$)를 새롭게 발굴하였다. 또한 서양인에서 발굴된 염증성장질환 감수성 유전좌위 245 개를 한국인 메타분석 결과에서 찾아 33 개 유전좌위에 대한 유의한 연관성을 확인하였다.

발굴한 유전좌위의 기능 규명을 돕기 위하여, 한국인 크론병 환자 101 명의 혈액에서 추출한 RNA 의 염기서열분석을 수행하였고, 발현정량적형질유전자좌 (eQTL) 연구 결과 데이터 베이스 (http://asan.crohneqtl.com/)를 구축하였다. 발현정량적형질유전자좌 연구에서 오류 발견률 0.05 미만을 만족하는 단일염기다형성에 따른 유전자 발현량 변화 정보 135,164 개와 발현 차이를 보이는 3,816 개 유전자를

발굴하였다. 발현정량적형질유전자좌 연구와 위의 전장유전체연관분석 결과를 통합 분석한 결과 이전에 보고된 크론병 감수성 유전좌위 9q32 와 2q37 에서 *TNFSF15* 과 *GPR35* 를 타겟 유전자로 발굴 하였으며, 전장유전체연관분석 연구에서 발굴한 단일염기다형성의 크론병 위험도를 높이는 대립유전자는 두 유전자의 발현 감소와 유의하게 연관되어있음을 밝혔다.

한국인과 서양인에서 크론병 혹은 궤양성대장염 관련 생물학적 경로를 비교하기 위해 한국인 전장유전체연관분석 데이터의 메타분석 결과와 서양인 전장 유전체연관분석 결과의 요약통계 데이터를 사용하여 경로분석을 수행하였다. 크론병의 경우 한국인에서 MHC 와 항원자극 관련 경로가 유의한 연관성을 보였으며, 서양인에서는 사이토카인과 전사인자 관련 경로가 유의하였다. 궤양성 대장염의 경우 한국인과 서양인에서 모두 MHC 와 항원결합 관련 경로가 유의하였다. 또한 크론병 또는 궤양성대장염의 유전자위험점수를 이용하여 표현형 분산을 계산하였는데, 크론병의 유전자위험점수로 계산한 표현형 분산은 한국인 데이터 사용시 14%, 서양인 데이터 사용시 10%를 설명하였다. 궤양성대장염에서는 서양인 데이터 사용시 표현형 분산 값이 12%로 한국인 데이터를 이용한 계산 값인 7% 보다 높았다.

한국인에서 염증성장질환 감수성 유전좌위 3 개를 발굴하고 이미 보고된 33 개 유전좌위의 연관성을 확인했는데, 이는 아시아인과 유럽인 염증성장질환에서 공통된 것과 서로 다른 경로가 연관되어 있음을 시사한다. 본 연구로 한국인 염증성장질환 감수성 유전좌위는 54 개로 증가했다. 경로분석으로 크론병 관련 생물학적 경로가 동양인과 서양인에서 다름을 보여주었다. 또한 유전자위험점수 분석에서 서양인 데이터를 사용한 크론병 유전자위험점수는 한국인 데이터를 사용했을 때 보다 예측력이 더 낮아지는 결과를 보였다. 이런 연구 결과들은 이전에 우리가 보고한 크론병이 궤양성대장염 보다 인종 특이적이라는 결과와 동일하며, 유전학 연구에서의 다양성을 강조한다.

Key words: inflammatory bowel disease; GWAS; eQTL; pathway analysis; polygenic risk scores