이학석사 학위논문

크론병-궤양성대장염 질병 간 연관 분석을 통한 유전자좌 발굴

Case-case genome-wide association analysis identifying genetic loci with divergent effects on Crohn′s disease and ulcerative colitis

울 산 대 학 교 대 학 원

의 과 학 과

김 용 재

크론병-궤양성대장염 질병 간
연관 분석을 통한 유전자좌 발굴

지 도 교 수 이 호 수

이 논문을 이학석사학위 논문으로 제출함

2023 년 8 월

울 산 대 학 교 대 학 원
의 과 학 과
김 용 재

김용재의 이학석사학위 논문을 인준함

심사위원 박 상 형 (인)

심사위원 예 병 덕 (인)

심사위원 이 호 수 (인)

울 산 대 학 교 대 학 원

2023 년 8 월

ABSTRACT

Crohn's disease (CD) and ulcerative colitis (UC), two major subtypes of inflammatory bowel disease, show substantial differences in their clinical course and treatment response. To identify the genetic factors underlying the distinct characteristics of these two diseases, we performed a genome-wide association study (GWAS) between CD (n = 2,359) and UC (n = 2,175) in a Korean population, followed by replication in an independent sample of 772 CD and 619 UC cases. Two novel loci were identified with divergent effects on CD and UC: rs9842650 in *CD200* and rs885026 in *NCOR2*. In addition, the 7 established susceptibility loci (MHC, *TNFSF15, OTUD3, USP12, IL23R, FCHSD2,* and *RIPK2*) reached genome-wide significance. Of the 9 loci, 6 (MHC, *TNFSF15, OTUD3, USP12, IL23R,* and *CD200*) were replicated in the case-case GWAS (CC-GWAS) of European populations. The proportion of variance explained in CD-UC status by polygenic risk score analysis was up to 22.6%. The area under the receiver-operating characteristic curve value was 0.74, suggesting acceptable discrimination between CD and UC. This CD-UC GWAS provides new insights into genetic differences between the two diseases with similar symptoms and might be useful in improving their diagnosis and treatment.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AUC	Area under curve
CD	Crohn’s disease
eQTL	Expression quantitative trait loci
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
IBD	Inflammatory bowel disease
LD	Linkage disequilibrium
LDSC	Linkage disequilibrium score regression
MHC	Major histocompatibility complex
OR	Odds ratio
PP	Posterior probability
PRS	Polygenic risk score
ROC	Receiver operating characteristic
SNP	Single nucleotide polymorphism
UC	Ulcerative colitis

1. INTRODUCTION

Inflammatory bowel disease (IBD), a chronic inflammatory disorder of the gastrointestinal tract, is thought to develop due to dysregulated mucosal immune responses to the gut flora in genetically susceptible individuals (1). Crohn's disease (CD) and ulcerative colitis (UC) are the two main forms of IBD. Although these two forms of IBD share similar clinical and pathological features, there are differences in the disease localization, histopathology, and endoscopic features, suggesting differences in the underlying pathogenic mechanisms of each disease.

Previous large-scale studies on populations of European ancestry have greatly advanced our understanding of IBD-related genetics. A meta-analysis by the International IBD Genetics Consortium (IIBDGC), which combined genome-wide association studies (GWASs) 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 (ORs) between European and Asian cohorts (2). The latest genome-wide meta-analysis performed on populations of European ancestry reported 241 susceptibility loci for IBD, revealing substantial overlapping of the genetic risk for CD and UC (3).

Understanding the genetic factors that contribute to the disease-specific characteristics is crucial for improving diagnosis and treatment. Previously, approximately 35% of loci were designated as CD- or UC-associated loci based on their relative strength of association (2, 3). The largest genotype-phenotype study also reported that specific genetic loci contribute to phenotype differences between CD and UC (4). While the shared genetic component is substantial (2, 3), previous Asian studies have also implicated differential genetic architecture between these two disorders. We and a Japanese group found the association of the same single nucleotide polymorphism (SNP) in the major histocompatibility complex (MHC) region with CD and UC, but the association was in the opposite direction (5-7). We therefore hypothesized that analysis of the genetic differences between CD and UC might lead to a better understanding of these two diseases. To this end, we performed a GWAS of CD and UC, estimated the proportion of variance explained in CD-UC status, and compared the findings with the results of a newly developed method case-case GWAS (CC-GWAS) using publicly available summary statistics.

2. MATERIALS AND METHODS

2.1 Study subjects

The datasets used in this study and the strategy for identifying loci with divergent effects on CD and UC are presented in Figure 1. Diagnostic criteria for CD and UC were described in our previous clinical studies (28-30). As we removed overlapping samples across genotyping platforms, all the samples were independent. For discovery, we combined 3 previously published datasets consisting of 2 GWASs and an Immunochip dataset without overlapping samples. The 3 cohorts were collected using the same protocol but in different years and genotyped using different platforms: cohort I included 725 CD and 1,001 UC cases genotyped most recently using the Infinium Asian Screening Array-24 v1.0 (Illumina) (8), cohort II included 896 CD and 573 UC cases genotyped firstly using the OmniExpress and Omni1-Quad (Illumina) (9), and cohort III included 738 CD and 601 UC cases genotyped using the Infinium ImmunoArray-24 v2 (Illumina) (7). The novel candidate loci were genotyped using the TaqMan genotyping assay in the independent cohort IV of 772 CD and 619 UC cases. In total, 5,925 samples including 3,131 CD and 2,794 UC cases were used for the meta-analysis. The clinical characteristics of the patients are summarized in Table 1. All patients were recruited from the IBD Clinic of Asan Medical Center.

2.2 Quality control and imputation

Standard quality control (QC) procedures were performed for each cohort dataset using PLINK v1.9 (https://www.cog-genomics.org/plink2) and R 3.6.1 as described previously for the Korean IBD GWAS (9). We excluded SNPs according to the following QC criteria: SNPs on the X, Y, and mitochondrial chromosomes; a minor allele frequency (MAF) < 0.01; a call rate < 98%; and $P < 1.0 \times 10^{-5}$ for healthy controls or $P < 5.0 \times 10^{-8}$ for patients with IBD cases in the Hardy-Weinberg equilibrium (HWE) test. Then, we removed samples with a proportion of missing genotypes > 4% or close genetic relatedness (PI_HAT > 0.2, IBS > 0.8) to any other samples. Population outliers were excluded following principal component analysis (PCA) using 194 reference samples including European (CEU), Asian (CHB + JPT), and African (YRI) populations from the International HapMap Project. After the standard QC analyses, 457,272 SNPs in cohort I (725 CD cases, 1,001 UC cases, and 378 controls), 522,285 SNPs in cohort II (896 CD cases, 573 UC cases, and 4,041 controls), and 168,049 SNPs in cohort III (738 CD cases, 601 UC cases, and 488 controls) were remained.

Imputation was performed based on the genotyped SNPs that passed the QC criteria in each cohort dataset based on the multi-ethnic 1000 Genomes Project reference panel release v5 (https://www.international genome.org/). The software IMPUTE v.2.3.2 (31) was used for

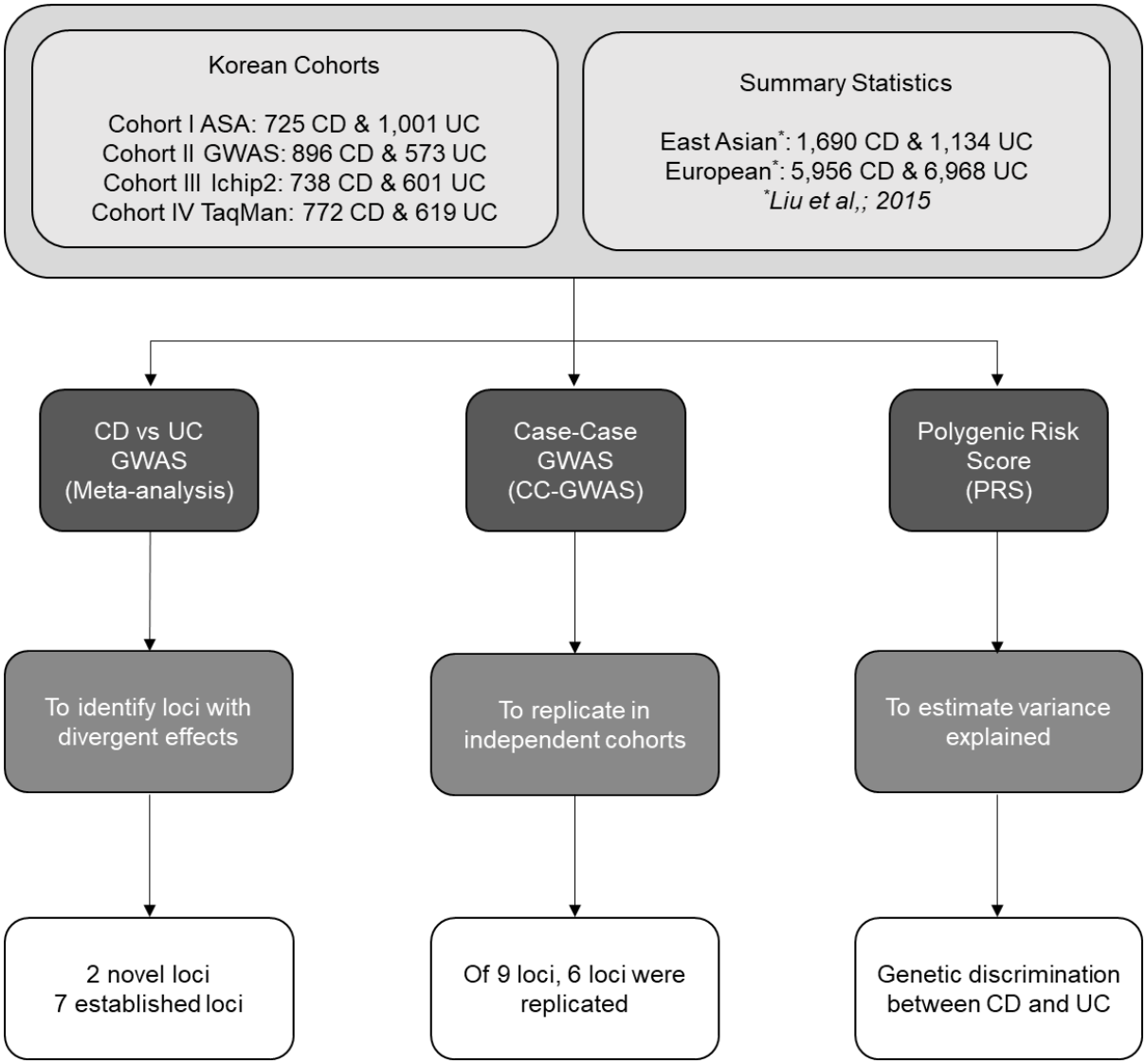


Figure 1. Study design to identify genetic loci with divergent effects on Crohn's disease (CD) and ulcerative colitis (UC).

Table 1. Clinical characteristics of Korean patients with CD and UC.

	Cohort I		Cohort II		Cohort III		Cohort IV		Total	
	CD	UC	CD	UC	CD	UC	CD	UC	CD	UC
No. of samples	725	1,001	896	573	738	601	772	619	3,131	2,794
Male (%)	561 (77.4)	605 (60.4)	633 (70.6)	320 (55.9)	551 (74.7)	361 (60.1)	555 (71.9)	345 (55.7)	2,300 (73.5)	1,631 (58.4)
Mean age at diagnosis (yr)	24.2 ± 8.8	37.1 ± 14.5	22.3 ± 8.2	36.0 ± 13.9	24.8 ± 8.8	36.3 ± 13.8	24.4 ± 9.1	36.5 ± 13.4		
Age group at diagnosis (%)										
≤ 16	104 (14.6)	37 (3.8)	237 (26.5)	51 (8.9)	73 (9.9)	21 (3.5)	70 (9.1)	13 (2.1)	484 (15.5)	122 (4.4)
17~40	566 (79.5)	557 (56.5)	621 (69.3)	280 (49.0)	612 (82.9)	352 (58.6)	652 (84.5)	373 (60.3)	2,451 (78.3)	1,562 (55.9)
≥ 40	42 (5.9)	392 (39.8)	38 (4.2)	241 (42.1)	53 (7.2)	228 (37.9)	50 (6.5)	233 (37.6)	183 (5.8)	1,094 (39.1)
NA	13	15		1					13	16
Location, no. (%)										
Ileum	106 (20.4)		158 (18.0)		189 (25.7)		193 (25.0)		646 (22.2)	
Colon	28 (5.4)		48 (5.5)		22 (3.0)		24 (3.1)		122 (4.2)	
Ileocolon	385 (74.2)		674 (76.6)		523 (71.3)		555 (71.9)		2,137 (73.6)	
NA	206		16		4				226	
Extent, no. (%)										
Proctitis		235 (34.3)		155 (27.2)		204 (34.3)		174 (28.1)		768 (27.5)
Left-sided colitis		184 (26.8)		179 (31.5)		184 (31.0)		195 (31.5)		742 (26.6)
Extensive colitis		267 (38.9)		235 (41.3)		206 (34.7)		250 (40.4)		958 (34.3)
NA		315		4		7				326
Behavior, no. (%)										
Inflammatory	267 (49.1)		343 (39.1)		345 (47.0)		400 (51.8)		1,355 (43.3)	
Stricturing	98 (18.0)		173 (19.7)		122 (16.6)		122 (15.8)		515 (16.4)	
Penetrating	179 (32.9)		362 (41.2)		267 (36.4)		242 (31.3)		1,050 (33.5)	
NA	181		18		4		8		211	
Perianal fistula, no. (%)										
No	264 (38.1)		325 (38.5)		392 (53.2)		331 (42.9)		1,312 (41.9)	
Yes	429 (61.9)		519 (61.5)		345 (46.8)		441 (57.1)		1,734 (55.4)	
NA	32		52		1				85	

CD, Crohn's disease; UC, ulcerative colitis; NA, not applicable.

imputing the genotype data of untyped SNPs following pre-phasing using SHAPEIT v.2 (32). For the QC of imputed SNPs, we removed all imputed SNPs with an info score < 0.8, a missing rate > 10%, an MAF < 1%, an HWE test $P < 1.0 \times 10^{-5}$ for controls and 5.0×10^{-8} for cases, or a posterior probability score < 0.8. After imputation and QC procedures, there were 6,139,980 imputed SNPs in cohort I, 6,088,678 imputed SNPs in cohort II, and 2,701,234 SNPs in cohort III.

2.3 Association analysis comparing CD with UC

We performed an association test between CD and UC in each cohort dataset using the additive model of frequentist association test of SNPTEST v2.5.2 (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) (10). Then, to increase power, we performed a fixed-effects meta-analysis using the summary statistics from 3 cohorts based on the inverse-variance method of meta v1.7. Of 2,765,594 shared SNPs among datasets from the 3 cohorts, all SNPs with a heterogeneity of $P < 0.05$ in the meta-analysis were removed due to possible heterogeneity across studies. Analysis was also performed including a continuous model for age at diagnosis and a binary model for sex as covariates. A quantile-quantile plot was generated using R 3.6.1 (http://www.r-project.org/) 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 (11), which provides 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 (3.6.1). Conditional regression analysis was performed to identify possible independent associations at genome-wide significant loci. To identify novel genetic loci with divergent effects on CD and UC, we selected SNPs with $P_{meta} < 1 \times 10^{-7}$ in the meta-analysis for replication using independent cohort IV, which consisted of 772 CD and 619 UC cases. Genotyping of the cohort IV was performed using a TaqMan genotyping assay with the Applied Biosystems 7900HT Fast Real-Time PCR System according to the manufacturer's instructions. After a combined analysis using the total datasets from the 4 cohorts datasets, we considered the selected SNPs with $P_{combined} < 5 \times 10^{-8}$ and a heterogeneity of $P > 0.05$ as statistically significant signal. We previously used ~4900 shared controls for a GWAS of CD or UC (6-9); however, to compare direction of the allelic effects between CD and UC, we performed fixed-effects meta-analyses using case-control summary statistics from cohorts I, II, and III and non-overlapping controls comprising a total of 2,359 CD cases vs. 2,454 controls or 2,175 UC

cases vs. 2,453 controls.

2.4 CC-GWAS analysis using summary statistics of Europeans or East Asians

To identify loci with divergent effects on CD and UC using the publicly available summary statistics of Europeans or East Asians (downloaded from www.ibdgenetics.org), we applied a CC-GWAS (case–case genome-wide association study) method, comparing the allele frequency between cases of the two disorders based on the respective case–control GWAS summary statistics (18). The CC-GWAS weighted the effect sizes from the respective case–control GWAS using two methods of CC-GWAS ordinary least squares (CC-$GWAS_{OLS}$) and CC-GWAS exact (CC-$GWAS_{exact}$) to control type I error. SNPs with a CC-$GWAS_{OLS}$ $P < 5 \times 10^{-8}$ and a CC-$GWAS_{exact}$ $P < 1 \times 10^{-4}$ were considered statistically significant. Based on these CC-GWAS methods, we compared CD and UC cases using case-control summary statistics of Europeans (5,956 CD cases and 14,927 controls/ 6,968 UC cases and 20,464 controls) or East Asians (1,690 CD cases and 3,719 controls/ 1,134 UC cases and 3,719 controls) (2).

2.5 Expression quantitative trait loci (eQTL) analysis

To gain insight into the potential functional roles of the loci with divergent effects on CD and UC, we performed cis-eQTL analyses by extensively searching publicly available data from the GTEx (v.8) (33) and whole blood cis-eQTL databases from (13), ImmuNexUT (14), Koreans (34), and Japanese (35). In the GTEx database, we selected disease-relevant samples including whole blood, small intestine, transverse colon, and sigmoid colon.

2.6 Gene annotation

To determine a set of causal SNPs, fine-mapping analysis was performed using 'FM-summary' (https://github.com/hailianghuang/FM-summary/blob/master/getCredible.r) based on the summary statistics from the meta-analysis of cohorts I and II 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 a posterior probability (PP) > 95% in the fine-mapping analysis. In addition, using Multi-marker Analysis of GenoMic Annotation (MAGMA) v.1.07b (12), we performed gene analysis to prioritize causal genes. We used summary statistics from the fixed-effects meta-analysis using the datasets from the 3 cohorts of the association test between CD and UC cases, the LD reference of East Asians (JPT + CHB) in the 1000 Genomes Project reference panel, the location file of 14,182 reference genes. In the gene analysis, we annotated genes applying Bonferroni corrected $P < 3.53 \times 10^{-6}$ (0.05/14,182).

2.7 Polygenic risk scores

We performed polygenic risk score (PRS) analysis using PRSice-2 (36) to estimate each individual's genetic score based on the genotype profiles of the independent cohort and effect size information from the summary statistics of the fixed-effects meta-analysis data. To avoid overfitting, we used the meta-analysis of cohorts I and II as the base data for estimating the effect sizes and cohort III consisting of 738 CD and 601 UC samples as the target data for evaluating the PRS. We treated the MHC region (chromosome 6: 25−34 Mb) with additional caution to minimize overfitting due to a tight LD by selecting only the most significant variant in this region. After extracting SNPs with an MAF > 0.05 and performing LD clumping (--clump-kb 250, --clump-p 1.00, and --clump-r2 0.10) using the 1000 Genomes East Asian data (CHB+JPT) as a reference panel, a total of 34,808 SNPs remained. We selected 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) 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. We also calculated the area under the receiver operating characteristic (ROC) curve (AUC) using PRSs to estimate the sensitivity/specificity. The AUC estimates the probability that a randomly selected case has predicted value more extreme than that of a randomly chosen control.

3. RESULTS

3.1 Differentiating genetic contribution to CD and UC

To identify genetic loci with divergent effects on CD and UC, we performed an association test between CD and UC in each dataset of cohort I (725 CD and 1,001 UC cases) with 6,597,252 SNPs (8), cohort II (896 CD and 573 UC cases) with 6,610,963 SNPs (9), and cohort III (738 CD and 601 UC cases) with 2,869,283 SNPs (7) using the additive model of frequentist association test of SNPTEST v2.5.2 (10). Using shared 2,765,594 SNPs among the summary statistics of the 3 cohorts, we performed a fixed-effects meta-analysis based on the inverse-variance method of meta v1.7. We excluded 126,084 SNPs with heterogeneity $P < 0.05$ in the meta-analysis due to possible heterogeneity across studies. The resulting meta-analysis of the 3 GWAS datasets demonstrated moderate inflation of test statistics ($\lambda_{GC} = 1.15$, Figure 2), though the LDSC intercept (1.00) indicated that the inflation is driven by trait polygenicity rather than confounding bias (11). Applying a threshold of $P_{meta} < 5 \times 10^{-8}$, 8 loci including 7 established loci (MHC, *TNFSF15*, *OTUD3*, *USP12*, *IL23R*, *FCHSD2*, and *RIPK2*) and a novel locus of *NCOR2* were identified as genetic loci with divergent effects on CD and UC (Figure 3 and Table 2). By applying a threshold of $P_{meta} < 1 \times 10^{-7}$, we selected 2 additional novel candidate loci, *ZBTB10* and *CD200*, for the replication study (Table 2). We genotyped the 3 SNPs from 3 novel candidate loci (*NCOR2*, *ZBTB10* and *CD200*) in an independent cohort IV consisted of 772 CD and 619 UC cases using TaqMan technology. By combining association results from the 4 cohorts, 2 novel loci were identified: rs885026 in *NCOR2* at 12q24 ($P_{combined} = 7.83 \times 10^{-10}$, OR = 0.78) and rs9842650 in *CD200* at 3q13 ($P_{combined} = 8.26 \times 10^{-10}$, OR = 1.29) (Table 3 and 4). These 2 loci did not show additional independent signals following conditional analyses on the top SNPs (Figure 4A and B). In total, we identified 9 genetic loci with divergent effects on CD and UC (Table 2). Consistent associations were observed even after adjusting for age and sex (Table 5). We also compared the direction of allelic effects of the 9 loci (MHC, *TNFSF15*, *OTUD3*, *USP12*, *IL23R*, *FCHSD2*, *NCOR2*, *RIPK2*, and *CD200*) with divergent effects on CD and UC using case versus non-overlapping healthy control data. *TNFSF15* and *OTUD3* showed the same direction of effects with significantly different effect sizes, while the others showed the opposite direction of effects in a total of 2,359 CD cases versus 2,454 controls or 2,175 UC cases versus 2,453 controls (Table 6 and Figure 5). To identify credible sets of causal SNPs from the identified GWA variants, we performed fine-mapping analysis using FM-summary. A total of 2 loci (MHC and *TNFSF15*) showed SNPs with posterior probability (PP) >50%, providing evidence of a causal SNP associated (Table 7). At the MHC locus, the lead SNP rs9270965 was the only variant with PP = 99.9% within the 95% credible set and rs722126 at the *TNFSF15* locus was the only variant with

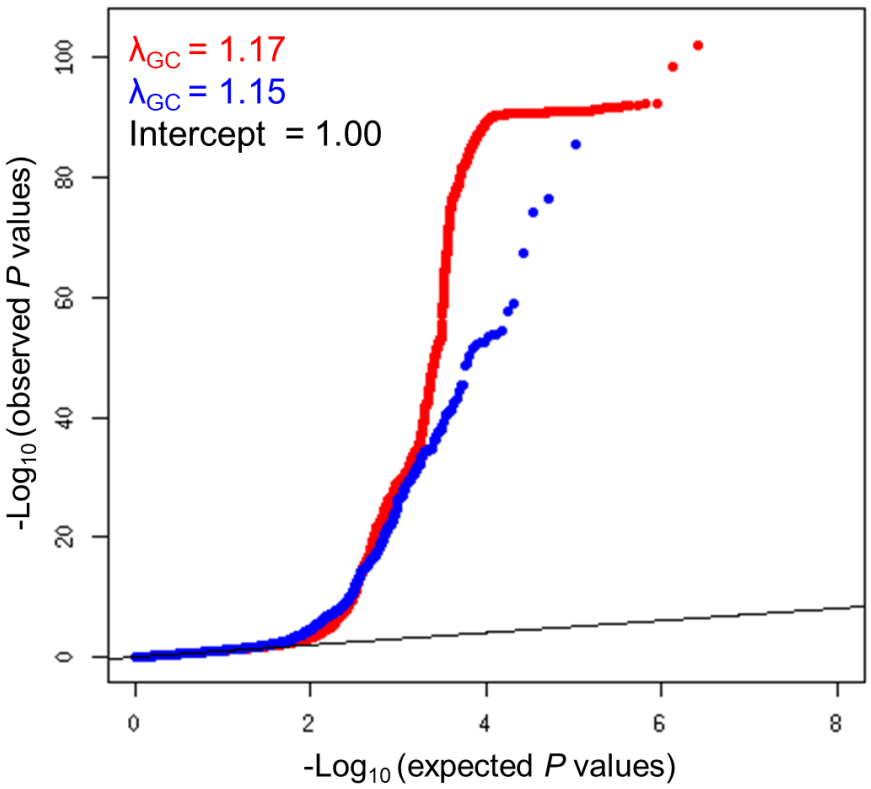


Figure 2. Quantile-quantile plot for meta-analysis of the 3 GWAS datasets. The -$\log_{10}P$ values of 2,639,510 SNPs (red dots) and 103,800 LD-pruned ($r^2 < 0.2$) SNPs (blue dots) were plotted against the expected null distributions. The intercept was determined using linkage disequilibrium score regression (LDSC).

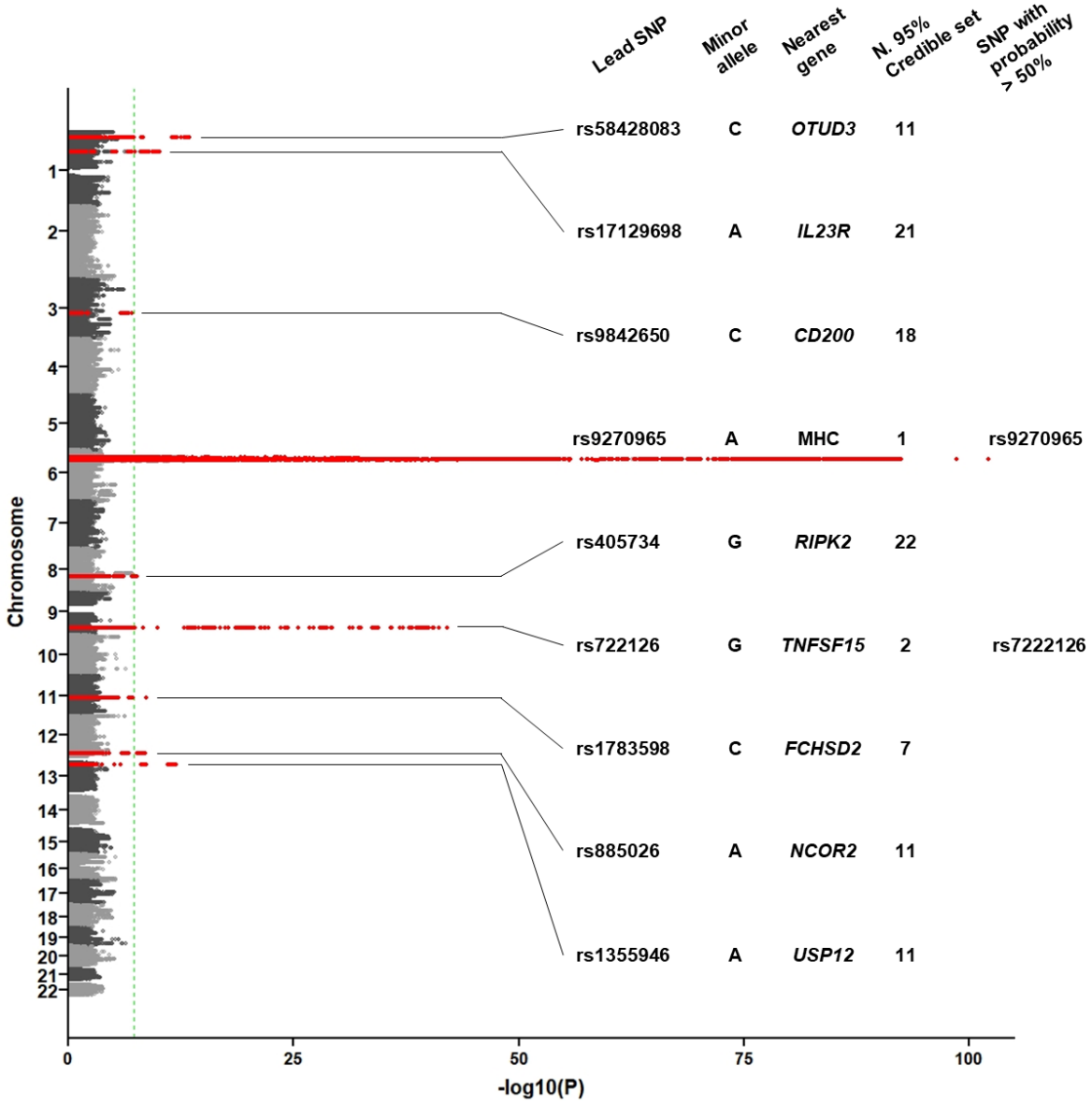


Figure 3. Manhattan plot for the 9 identified loci with divergent effects in CD and UC. From left to right; Manhattan plot of the GWAS meta-analysis results between CD versus UC (genome-wide significance level – $P < 5 \times 10^{-8}$ – indicated with green line); single nucleotide polymorphisms (SNPs) located in GWAS significant loci are colored red; Minor allele; Nearest gene – the closest gene to the lead SNP; Number of SNPs in the 95% credible set – the minimum set of variants from Bayesian fine-mapping analysis with > 95% likely to contain the causal variant; SNP with probability > 50% - single variant (if detected) with > 50% probability of being causal.

Table 2. Lead SNPs in the 10 loci with $P_{meta} < 1.00 \times 10^{-7}$ from fixed-effects meta-analysis.

					Meta-analysis (2,359 CD / 2,175 UC)			Cohort I (725 CD / 1,001 UC)				Cohort II (896 CD / 573 UC)				Cohort III (738 CD / 601 UC)			
								MAF				MAF				MAF			
CHR	SNP	Position (hg19)	Candidate gene	Minor allele	P^*	OR (95% CI)	$P_{het}^\dagger$	CD	UC	$P^\ddagger$	OR (95% CI)	CD	UC	$P^\ddagger$	OR (95% CI)	CD	UC	$P^\ddagger$	OR (95% CI)
6	rs9270965	32,573,471	MHC	A	7.54E-103	0.36 (0.26-0.45)	5.07E-01	0.16	0.36	3.64E-38	0.33 (0.28-0.39)	0.14	0.34	5.74E-35	0.31 (0.26-0.37)	0.16	0.36	1.28E-33	0.33 (0.27-0.39)
9	rs722126	117,592,778	*TNFSF15*	G	8.76E-43	0.53 (0.44-0.62)	9.59E-01	0.23	0.36	5.65E-17	0.52 (0.45-0.61)	0.23	0.36	2.24E-14	0.53 (0.45-0.62)	0.22	0.36	1.12E-14	0.51 (0.43-0.60)
1	rs58428083	20,190,931	*OTUD3*	C	3.08E-14	1.50 (1.39-1.60)	9.33E-01	0.24	0.18	7.68E-06	1.44 (1.22-1.70)	0.23	0.16	5.01E-06	1.56 (1.29-1.89)	0.25	0.19	3.80E-05	1.47 (1.22-1.77)
13	rs1359946	27,536,972	*USP12*	A	1.15E-12	0.68 (0.59-0.78)	4.97E-01	0.16	0.23	1.46E-07	0.63 (0.53-0.75)	0.17	0.23	9.32E-05	0.70 (0.58-0.84)	0.16	0.21	2.61E-03	0.74 (0.61-0.90)
1	rs17129698	67,654,072	*IL23R*	A	6.92E-11	1.96 (1.76-2.16)	6.89E-01	0.05	0.03	2.75E-04	1.91 (1.35-2.71)	0.06	0.03	4.62E-04	2.00 (1.36-2.93)	0.06	0.03	2.46E-05	2.37 (1.56-3.59)
11	rs1783598	72,851,463	*FCHSD2*	C	1.88E-09	1.30 (1.22-1.40)	9.14E-01	0.45	0.38	4.47E-05	1.33 (1.16-1.52)	0.47	0.41	1.00E-03	1.28 (1.11-1.49)	0.43	0.38	3.02E-03	1.25 (1.07-1.46)
12	rs885026	125,032,789	*NCOR2*	A	2.67E-09	0.76 (0.67-0.85)	9.75E-01	0.29	0.35	2.54E-04	0.76 (0.66-0.88)	0.29	0.35	3.95E-04	0.75 (0.64-0.88)	0.29	0.35	2.03E-03	0.78 (0.66-0.92)
8	rs405734	90,768,439	*RIPK2*	G	2.12E-08	0.78 (0.70-0.87)	9.45E-01	0.36	0.41	1.20E-03	0.79 (0.69-0.91)	0.35	0.42	6.52E-04	0.77 (0.66-0.89)	0.36	0.41	2.20E-03	0.79 (0.67-0.92)
8	rs12543811	81,278,885	*ZBTB10*	A	7.66E-08	0.79 (0.70-0.87)	4.53E-01	0.33	0.40	3.02E-05	0.74 (0.64-0.85)	0.36	0.40	4.07E-02	0.86 (0.73-1.00)	0.36	0.41	2.89E-03	0.80 (0.68-0.93)
3	rs9842650	112,069,392	*CD200*	C	8.00E-08	1.29 (1.20-1.38)	9.33E-01	0.32	0.27	2.54E-03	1.25 (1.08-1.45)	0.30	0.24	9.45E-04	1.33 (1.13-1.58)	0.30	0.25	2.85E-03	1.29 (1.09-1.53)

CD, Crohn's disease; CI, confidence interval; CHR, Chromsome; hg19, human genome version 19; MAF, minor allele frequency; OR, odds ratio; Position, chromosome position; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

*Fixed-effects meta-analysis *P* value.

†*P* value for heterogeneity.

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

Table 3. Combined analysis of 2 novel loci with divergent effects on CD and UC in Koreans.

Locus	SNP	Position (hg19)	Candidate gene	Minor allele	Study	Number of samples		MAF		OR (95% CI)	*P*	P_{het} *
						CD	UC	CD	UC			
12q24	rs885026	125,032,789	*NCOR2*	A (0.32)#	Cohort I	725	1,001	0.29	0.35	0.76 (0.66-0.88)	2.54×10^{-4}†	
					Cohort II	896	573	0.29	0.35	0.75 (0.64-0.88)	3.95×10^{-4}†	
					Cohort III	738	601	0.29	0.35	0.78 (0.66-0.92)	2.03×10^{-3}†	
					Cohort IV	772	619	0.30	0.33	0.85 (0.72-1.00)	5.21×10^{-2}†	
					Combined	3,131	2,794	0.29	0.34	0.78 (0.78-0.93)	7.83×10^{-10}‡	6.98×10^{-1}
3q13	rs9842650	112,069,392	*CD200*	C (0.27)#	Cohort I	725	1,001	0.32	0.27	1.25 (1.07-1.46)	2.54×10^{-3}†	
					Cohort II	896	573	0.30	0.24	1.33 (1.13-1.57)	9.45×10^{-4}†	
					Cohort III	738	601	0.30	0.25	1.29 (1.09-1.53)	2.85×10^{-3}†	
					Cohort IV	772	619	0.30	0.25	1.29 (1.09-1.53)	2.86×10^{-3}†	
					Combined	3,131	2,794	0.31	0.26	1.29 (1.19-1.40)	8.26×10^{-10}‡	9.87×10^{-1}

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

*P value for heterogeneity.

#Minor allele frequency of 4,907 healthy controls from cohort I, II, and III dataset.

†Association P value of SNPTEST v2.5.2.

‡Combined P value using a fixed-effect meta-analysis model.

Table 4. Combined analysis of the 8q21 locus using cohort IV.

Locus	SNP	Position (hg19)	Candidate gene	Minor allele	Study	Number of samples		MAF		OR (95% CI)	*P*	P_{het}*
						CD	UC	CD	UC			
8q21	rs12543811	81,278,885	*ZBTB10*	A (0.37)#	Cohort I	725	1,001	0.33	0.40	0.74 (0.64-0.85)	3.02×10^{-5}†	
					Cohort II	896	573	0.36	0.40	0.86 (0.73-1.00)	4.07×10^{-2}†	
					Cohort III	738	601	0.36	0.41	0.80 (0.68-0.93)	2.89×10^{-3}†	
					Cohort IV	772	619	0.36	0.38	0.92 (0.79-1.08)	3.10×10^{-1}†	
					Combined	3,131	2,794	0.36	0.40	0.82 (0.76-0.88)	2.38×10^{-7}‡	1.87×10^{-1}

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

*P value for heterogeneity.

#Minor allele frequency of 4,041 healthy controls from GWAS dataset.

†Association P value of SNPTEST v2.5.2.

‡Combined P value using a fixed-effect meta-analysis model.

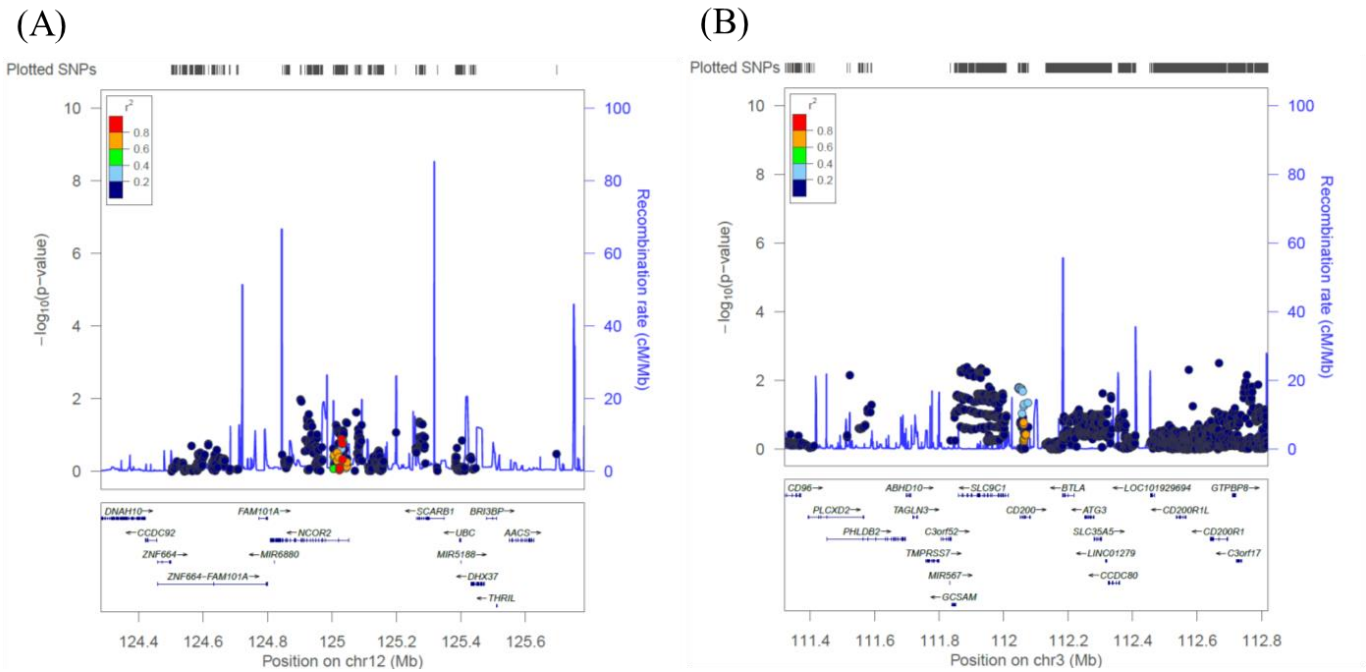


Figure 4. Conditional association signals for 2 novel loci. (A) Conditioned on rs885026 at 12q24; (B) Conditioned on rs9842650 at 3q13. SNPs were plotted according to their chromosomal positions [NCBI Build 37] with -$\log_{10}P$ values from the meta-analysis in the region flanking 750 kb on either side of the marker SNP. 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 [November 2014] are plotted to reflect the local LD structure. Plots were generated using LocusZoom.

Table 5. Associations of the 9 loci after adjustment for age and sex.

CHR	SNP	Position (hg19)	Candidate gene	Minor allele	Crude		Age		Sex		Sex + Age	
					*P**	OR (95% CI)	*P**	OR (95% CI)	*P**	OR (95% CI)	*P**	OR (95% CI)
1	rs58428083	20,190,931	*OTUD3*	C	3.08E-14	1.50 (1.39-1.60)	5.12.E-10	1.47 (1.30-1.66)	2.99E-13	1.49 (1.34-1.66)	1.10.E-09	1.46 (1.30-1.66)
1	rs17129698	67,654,072	*IL23R*	A	6.92E-11	1.96 (1.76-2.16)	1.59.E-07	1.96 (1.52-2.52)	1.53E-09	1.96 (1.58-2.46)	7.80.E-07	1.90 (0.47-2.46)
3	rs9842650	112,069,392	*CD200*	C	8.00E-08	1.29 (1.20-1.38)	1.41.E-04	1.23 (1.11-1.37)	7.38E-07	1.27 (1.16-1.40)	3.78.E-04	1.22 (1.09-1.36)
6	rs9270965	32,573,471	MHC	A	7.54E-103	0.36 (0.26-0.45)	1.39.E-73	0.33 (0.29-0.37)	2.16E-91	0.33 (0.29-0.37)	2.51.E-72	0.32 (0.29-0.37)
8	rs405734	90,768,439	*RIPK2*	G	2.12E-08	0.78 (0.70-0.87)	1.15.E-06	0.78 (0.71-0.86)	3.24E-09	0.77 (0.70-0.84)	3.43.E-07	0.77 (0.69-0.85)
9	rs722126	117,592,778	*TNFSF15*	G	8.76E-43	0.53 (0.44-0.62)	1.52.E-33	0.51 (0.46-0.57)	9.87E-41	0.52 (0.47-0.57)	3.23.E-33	0.51 (0.46-0.57)
11	rs1783598	72,851,463	*FCHSD2*	C	1.88E-09	1.30 (1.22-1.40)	6.09.E-07	1.29 (1.17-1.43)	3.86E-10	1.33 (1.22-1.45)	2.97.E-07	1.31 (1.18-1.45)
12	rs885026	125,032,789	*NCOR2*	A	2.67E-09	0.76 (0.67-0.85)	7.92.E-08	0.75 (0.67-0.83)	5.63E-08	0.77 (0.70-0.85)	7.87.E-08	0.75 (0.67-0.83)
13	rs1359946	27,536,972	*USP12*	A	1.15E-12	0.68 (0.59-0.78)	1.29.E-11	0.65 (0.58-0.74)	3.18E-12	0.68 (0.61-0.76)	1.40.E-11	0.65 (0.57-0.74)

CHR, chromosome; CI, confidence interval; hg19, human genome version 19; OR, odds ratio; *P*, *P* value; Position, chromosome position; SNP, single nucleotide polymorphism

*Fixed-effects meta-analysis *P* value.

Table 6. Nine loci with divergent effects on CD and UC in Koreans.

CHR	SNP	Candidate gene	Position (hg19)	Minor allele	CD vs UC (2,359 CD / 2,175 UC)			CD vs controls (2,359 cases / 2,454 controls)			UC vs controls (2,175 cases / 2,453 controls)			Minor allele frequency		
					P^{*}	OR (95% CI)	$P_{het}^{\#}$	P^{*}	OR (95% CI)	$P_{het}^{\#}$	P^{*}	OR (95% CI)	$P_{het}^{\#}$	CD	Control†	UC
1	rs58428083	*OTUD3*	20,190,931	C	3.08×10^{-14}	1.50 (1.39-1.60)	9.33×10^{-1}	5.30×10^{-1}	0.97 (0.86-1.07)	3.37×10^{-1}	5.68×10^{-14}	0.62 (0.50-0.75)	9.01×10^{-1}	0.241	0.243	0.178
1	rs17129698	*IL23R*	67,654,072	A	6.92×10^{-11}	1.96 (1.76-2.16)	6.89×10^{-1}	3.43×10^{-4}	1.46 (1.25-1.66)	5.80×10^{-1}	1.81×10^{-3}	0.66 (0.40-0.92)	4.43×10^{-1}	0.058	0.043	0.029
3	rs9842650	*CD200*	112,069,392	C	8.00×10^{-8}	1.29 (1.20-1.38)	9.33×10^{-1}	4.62×10^{-4}	1.19 (1.09-1.29)	0.92×10^{-1}	2.78×10^{-2}	0.88 (0.77-0.99)	6.78×10^{-1}	0.307	0.265	0.258
6	rs9270965	MHC	32,573,471	A	7.54×10^{-103}	0.36 (0.26-0.45)	5.07×10^{-1}	3.97×10^{-53}	0.44 (0.33-0.54)	4.33×10^{-1}	4.64×10^{-12}	1.46 (1.35-1.57)	1.00×10^{-1}	0.150	0.282	0.357
8	rs405734	*RIPK2*	90,768,439	G	2.12×10^{-8}	0.78 (0.70-0.87)	9.45×10^{-1}	2.45×10^{-3}	0.87 (0.77-0.96)	0.44×10^{-1}	5.73×10^{-3}	1.15 (1.05-1.26)	5.41×10^{-1}	0.354	0.386	0.412
9	rs722126	*TNFSF15*	117,592,778	G	8.76×10^{-43}	0.53 (0.44-0.62)	9.59×10^{-1}	6.07×10^{-48}	0.49 (0.39-0.58)	8.02×10^{-1}	5.10×10^{-2}	0.90 (0.80-1.01)	9.11×10^{-1}	0.225	0.382	0.359
11	rs1783598	*FCHSD2*	72,851,463	C	1.88×10^{-9}	1.30 (1.22-1.40)	9.14×10^{-1}	1.14×10^{-3}	1.16 (1.07-1.25)	9.88×10^{-1}	4.41×10^{-1}	0.90 (0.80-1.00)	8.44×10^{-1}	0.451	0.429	0.387
12	rs885026	*NCOR2*	125,032,789	A	2.67×10^{-9}	0.76 (0.67-0.85)	9.75×10^{-1}	6.06×10^{-3}	0.87 (0.77-0.97)	9.10×10^{-1}	1.10×10^{-1}	1.09 (0.99-1.20)	3.67×10^{-1}	0.289	0.323	0.347
13	rs1359946	*USP12*	27,536,972	A	1.15×10^{-12}	0.68 (0.57-0.78)	4.97×10^{-1}	3.60×10^{-1}	0.94 (0.82-1.07)	4.92×10^{-1}	1.23×10^{-9}	1.48 (1.36-1.61)	0.48×10^{-1}	0.164	0.171	0.222

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

*Combined *P* value using a fixed-effect meta-analysis model.

#*P* value for heterogeneity.

†4,907 healthy controls from cohort I, II, and III dataset.

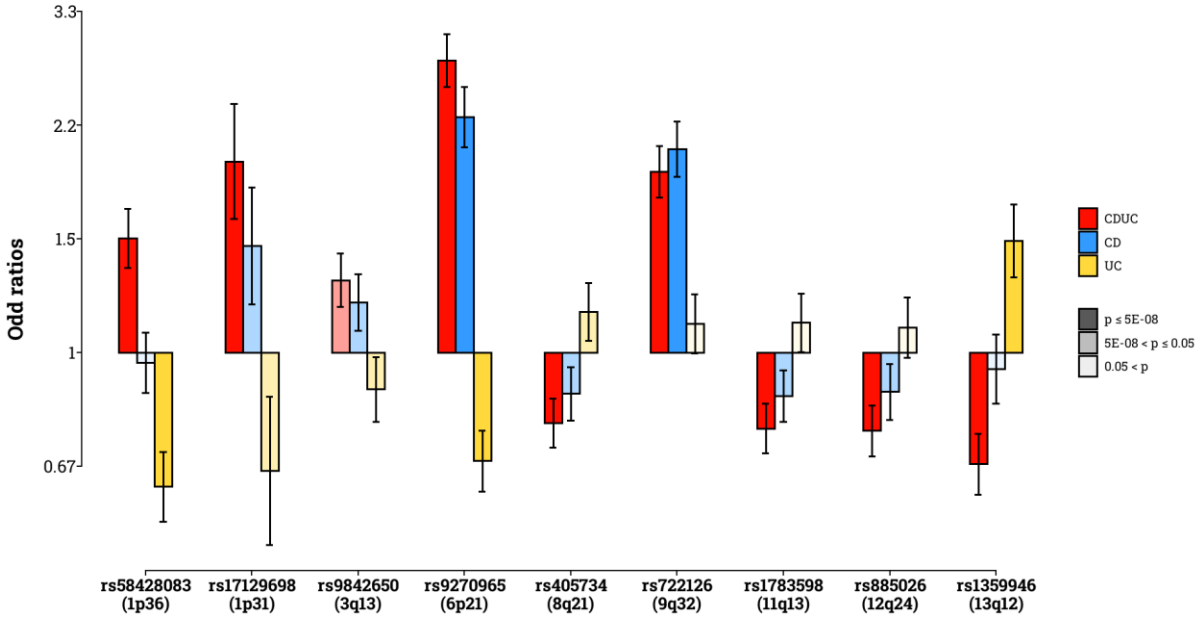


Figure 5. Comparison of OR values in CD and UC for the 9 identified loci with divergent effects. OR values from the association analyses in CD versus UC, CD versus controls, and UC versus controls were presented with 95% confidence intervals. OR values for the association tests between CD versus UC, CD versus controls, and UC versus controls were in red, blue, and yellow, respectively; darker colors indicated a GWAS significant effect ($P \leq 5 \times 10^{-8}$), lighter colors indicated a nominal significance level ($5 \times 10^{-8} < P \leq 0.05$) and white indicated non-significance ($0.05 < P$).

Table 7. List of SNPs in the 95% credible set of 9 identified loci.

Locus	Lead SNP	Gene	Number of SNPs in 95% credible set	SNP list of 95% credible set(posterior probability)*
1q36	rs58428083	*OTUD3*	11	rs58428083(0.168), rs4654903(0.129), rs6670226(0.123), rs6667256(0.123), rs58359414(0.116), rs7553638(0.086), rs6674040(0.066), rs2066130(0.055), rs6697482(0.039), rs10799591(0.039), rs10916670(0.02)
1p31	rs17129698	*IL23R*	21	rs17129698(0.155), rs17129700(0.155), rs61780309(0.115), rs61780308(0.114), rs12565567(0.112), rs17129680#(0.063), rs61780311(0.039), rs61780312(0.034), rs12562213(0.03), rs61780310(0.027), rs28464018(0.027), rs78377598#(0.011), rs12566159#(0.011), rs117633859#(0.009), rs2024825(0.009), rs117282985(0.009), rs12564219(0.008), rs61780314(0.007), rs6693659#(0.006), rs61780315(0.006), rs12069782(0.004)
3q13	rs9842650	*CD200*	18	rs9842650(0.205), rs59981538(0.081), rs9881834(0.056), rs2399418(0.052), rs60377655(0.052), rs1050572(0.052), rs111268897(0.051), rs57404826(0.051), rs2399417(0.049), rs11921546(0.048), rs4582023(0.046), rs60497880(0.036), rs59863561(0.031), rs60673403(0.031), rs3817425(0.03), rs16859484(0.029), rs4575866(0.028), rs12106675(0.028)
6p21	rs9270965	MHC	1	**rs9270965(0.999)**
8q21	rs405734	*RIPK2*	22	rs405734(0.064), rs402886(0.051), rs40452(0.05), rs39503(0.05), rs39504(0.05), rs39500(0.05), rs39761(0.05), rs43225(0.05), rs447618(0.049), rs43134(0.049), rs2735882(0.047), rs40545(0.046), rs39505(0.046), rs39506(0.046), rs465(0.044), rs416324(0.041), rs372981(0.04), rs39509(0.029), rs40247(0.028), rs400411(0.028), rs40453(0.028), rs411279(0.021)
9q32	rs722126	*TNFSF15*	2	**rs7222126(0.843)**, rs7040029(0.109)
11q13	rs1783598	*FCHSD2*	7	rs1783598(0.45), rs573529(0.441), rs7126070(0.014), rs6592510(0.014), rs3862794(0.013), rs6592500(0.013), rs12294037(0.012)
12q24	rs885026	*NCOR2*	11	rs885026(0.212), rs4765577(0.124), rs11057658(0.097), rs12307174(0.094), rs7136910(0.093), rs10846683(0.085), rs4765578(0.077), rs116928246(0.069), rs10773091(0.041), rs12228332(0.039), rs11057665(0.029)
13q12	rs1359946	*USP12*	11	rs1359946#(0.215), rs9512464(0.158), rs6491170(0.104), rs7983353(0.076), rs1556039(0.072), rs1556040(0.071), rs17085007#(0.069), rs9551344(0.062), rs73154069#(0.056), rs9579054(0.05), rs9553939(0.04)

Lead SNP, lead single nucleotide polymorphism.

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

#Overlapped SNPs in credible sets of European fine-mapping study. (PMID=28658209)

Bold: SNPs with posterior probability > 0.5.

PP = 84.3% within the 95% credible set including 3 SNPs (Table 7).

We then examined the previously established European IBD-associated loci (231 independent SNPs in 200 loci) in our GWAS of CD and UC (2). Data from 182 SNPs in 164 loci were available. Of these, a total of 5 SNPs from 4 loci (*OTUD3*, MHC, *TNFSF15*, and *USP12*) were identified as genetic loci with divergent effects on CD and UC at a GWAS significant threshold. These 5 SNPs had been classified "UC, IBD", "UC" or "NA" based on the maximum likelihood modeling, which were consistent with our findings.

3.2 Two novel loci with divergent effects on CD and UC

NCOR2 locus at 12q24

Lead SNP rs885026 ($P_{combined} = 7.83 \times 10^{-10}$, OR = 0.78) was located in intron 1 of *NCOR2* (nuclear receptor corepressor 2), the only gene within the LD region of 47.3 kb at 12q24 (Table 3 and Figure 6A). This locus did not show additional independent signals following conditional analysis on the top SNP rs885026 (Figure 6A). The 95% credible set at the 12q24 locus consisted of 11 SNPs including rs885026 (PP = 21.2%) in the fine-mapping analysis (Table 7). Gene analysis using MAGMA v.1.07b (12) identified a significant association with *NCOR2* ($P = 2.38 \times 10^{-7}$). rs885026 had eQTL effects for *NCOR2* in the whole blood ($P = 6.33 \times 10^{-20}$) (The eQTLGen database) (13) and in 25 immune cell-types (the ImmuNexUT, Immune Cell Gene Expression Atlas from the University of Tokyo) (14) (Table 8). In both databases, minor allele A, which is significantly more frequent in UC than in CD, was associated with a higher expression level of *NCOR2* than major allele C, suggesting higher expression of *NCOR2* in UC than in CD. *NCOR2*, which encodes a transcriptional co-repressor to inhibit the expression of target genes by modifying chromatin structure, was identified as a key regulator for IL-4-induced monocyte differentiation (15).

CD200 locus at 3q13

At 3q13, lead SNP rs9842650 ($P_{combined} = 8.26 \times 10^{-10}$, OR = 1.29) was located in intron 6 of *CD200* (CD200 molecule) and was the only gene within the LD region of 77.8 kb (Table 3 and Figure 6B). Conditioning on the top SNP showed no independent signals at the 3q13 locus (Figure 6B). The 95% credible set at the 3q13 locus consisted of 18 SNPs including rs9842650 (PP = 20.5%) in the fine-mapping analysis (Table 7). Gene analysis using MAGMA v.1.07b (12) identified a significant association with *CD200* ($P = 5.33 \times 10^{-8}$). The Japanese eQTL database showed that minor C allele (risk allele in CD, and protective allele in UC) was associated with a lower expression level of *CD200* in CD4 T cells than major T allele, suggesting higher expression

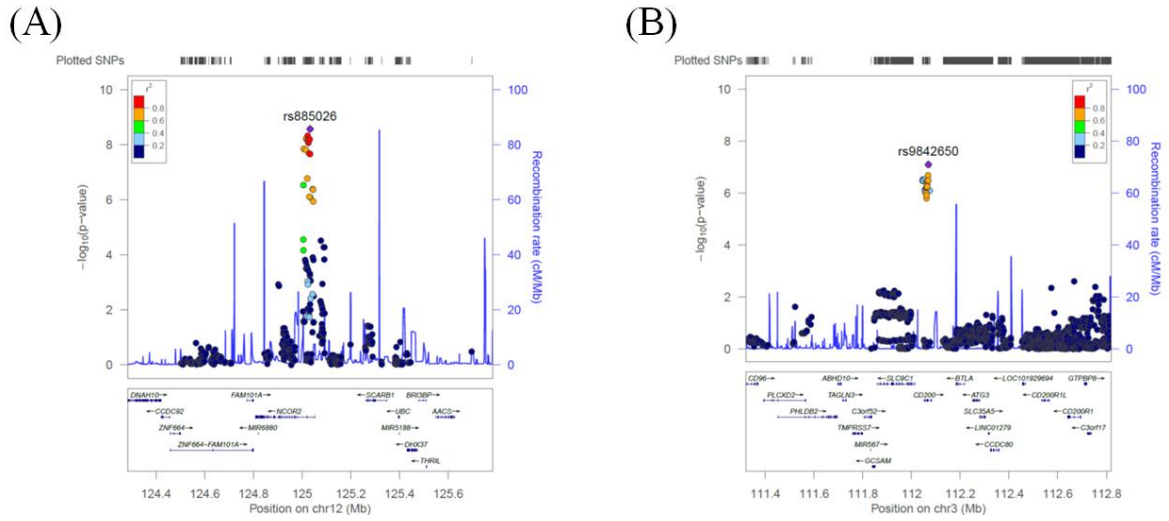


Figure 6. Regional association plots for two novel loci: (A) rs885026 at 12q24 and (B) rs9842650 at 3q13. SNPs are plotted according to their chromosomal positions [NCBI Build 37] with $-\log_{10}P$ values from the meta-analysis in the region flanking 750 kb on either side of the marker SNP. The most 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 [November 2014] are plotted to reflect the local LD structure. Plots were generated using LocusZoom.

Table 8. eQTLs in the 9 loci with divergent effects on CD and UC in Koreans.

Locus (Gene)	SNP	Ref	Alt	Gene Symbol	*P*	Effect	Tissue	Database
3q13 *(CD200)*	rs9842650	T	C	*CD200*	7.41E-04	-0.59	CD4+ T Cell	Japanese‡
				GCSAM	8.87E-07	4.92	Whole Blood	eQTLGen#
6p21 (MHC)	rs9270965	A	G	*HLA-DRB5*	9.10E-37	-0.46	Whole Blood	GTEx*
				HLA-DQB2	6.60E-32	0.55	Whole Blood	GTEx*
				HLA-DQB1	4.20E-29	-0.41	Whole Blood	GTEx*
				HLA-DRB9	2.70E-25	0.47	Whole Blood	GTEx*
				HLA-DRB6	1.40E-22	0.45	Whole Blood	GTEx*
				HLA-DQA2	1.20E-17	0.46	Whole Blood	GTEx*
				HLA-DQA1	7.40E-10	-0.15	Whole Blood	GTEx*
				HLA-DQB1-AS1	1.30E-06	-0.16	Whole Blood	GTEx*
				PBS2	7.30E-05	-0.07	Whole Blood	GTEx*
				PSMB9	1.30E-04	0.09	Whole Blood	GTEx*
				HLA-DRB5	7.60E-22	-0.60	Colon - Transverse	GTEx*
				HLA-DRB5	7.80E-18	-0.66	Colon - Sigmoid	GTEx*
				HLA-DQB2	7.30E-18	0.56	Colon - Transverse	GTEx*
				HLA-DQB2	9.70E-17	0.62	Colon - Sigmoid	GTEx*
				HLA-DRB6	6.50E-14	0.49	Colon - Transverse	GTEx*
				HLA-DQB1	1.00E-11	-0.47	Colon - Sigmoid	GTEx*
				HLA-DQA2	1.80E-11	0.47	Colon - Transverse	GTEx*
				HLA-DRB9	4.80E-09	0.40	Colon - Transverse	GTEx*
				HLA-DQB1	7.10E-09	-0.33	Colon - Transverse	GTEx*
				HLA-DRB6	1.50E-07	0.41	Colon - Sigmoid	GTEx*
				HLA-DRB1	1.80E-06	-0.25	Colon - Sigmoid	GTEx*
				HLA-DQA2	3.10E-06	0.38	Colon - Sigmoid	GTEx*
				HLA-DRB1	3.20E-06	-0.17	Colon - Transverse	GTEx*
				HLA-DQA1	1.40E-05	-0.25	Colon - Sigmoid	GTEx*
				NOTCH4	4.10E-05	-0.18	Colon - Sigmoid	GTEx*
				HLA-DRB9	4.70E-11	0.64	Small intestine - Terminal ileum	GTEx*
				HLA-DQB2	1.70E-09	0.48	Small intestine - Terminal ileum	GTEx*
				HLA-DRB5	1.80E-08	-0.54	Small intestine - Terminal ileum	GTEx*
				HLA-DRB6	6.50E-08	0.52	Small intestine - Terminal ileum	GTEx*
				HLA-DQB1	3.30E-07	-0.38	Small intestine - Terminal ileum	GTEx*
				HLA-DQA2	9.60E-07	0.51	Small intestine - Terminal ileum	GTEx*
				HLA-DRB5	3.27E-310	-75.24	Whole Blood	eQTLGen#
				HLA-DQB1	3.04E-250	-33.86	Whole Blood	eQTLGen#
				HLA-DRB6	6.14E-26	10.53	Whole Blood	eQTLGen#
				HLA-DQA2	3.13E-16	80.17	Whole Blood	eQTLGen#
				C4B	4.50E-14	7.55	Whole Blood	eQTLGen#
				SKIV2L	1.37E-13	7.40	Whole Blood	eQTLGen#
				C4A	1.04E-12	-7.13	Whole Blood	eQTLGen#
				HLA-DQA1	1.44E-11	-6.75	Whole Blood	eQTLGen#
				PSMB9	7.88E-09	5.77	Whole Blood	eQTLGen#
				HLA-DQB1-AS1	9.98E-09	-5.73	Whole Blood	eQTLGen#
				HSPA1B	4.13E-08	-5.49	Whole Blood	eQTLGen#
				AGER	1.23E-08	5.48	Whole Blood	eQTLGen#
				AGPAT1	5.65E-07	-5.00	Whole Blood	eQTLGen#
				DDAH2	1.76E-06	-4.78	Whole Blood	eQTLGen#
				HLA-DQB2	1.37E-05	4.35	Whole Blood	eQTLGen#
				HLA-DRB1	1.13E-12	-1.03	Peripheral blood	Japanese‡
				HLA-DRB5	6.78E-10	-0.92	Peripheral blood	Japanese‡
				HLA-DQA1	5.18E-07	-0.72	B cell	Japanese‡
				HLA-DRB5	6.34E-11	-0.89	B cell	Japanese‡
				HLA-DRB1	4.10E-08	-0.79	B cell	Japanese‡
				HLA-DRB5	7.66E-13	-0.98	Natural killer cell	Japanese‡

Table 8. Cont'd

Locus (Gene)	SNP	Ref	Alt	Gene Symbol	*P*	Effect	Tissue	Database
8p12 (*RIPK2*)	rs405734	A	G	*RP11-37B2.1*	4.50E-16	-0.22	Whole Blood	GTEx*
				RP11-37B2.1	2.40E-22	-0.41	Colon - Transverse	GTEx*
				RP11-37B2.1	3.30E-12	-0.38	Colon - Sigmoid	GTEx*
				RP11-37B2.1	9.70E-18	-0.48	Small intestine - Terminal ileum	GTEx*
				RP11-37B2.1	3.27E-310	-53.93	Whole Blood	eQTLGen#
				RIPK2	8.74E-195	29.77	Whole Blood	eQTLGen#
				DECR1	3.28E-14	-7.59	Whole Blood	eQTLGen#
				RP11-37B2.1	1.02E-07	-0.69	B cell	Japanese‡
				RIPK2	7.44E-12	0.85	CD8+ T Cell	Japanese‡
				RIPK2	6.62E-49	0.57	Naive CD8 T cell	ImmuNexUT†
				AF117829.1	2.37E-17	-0.46	Switched memory B cell	ImmuNexUT†
9q32 (*TNFSF15*)	rs722126	G	T	*TNFSF15*	4.20E-11	-0.23	Whole Blood	GTEx*
				TNFSF8	5.02E-233	-32.59	Whole Blood	eQTLGen#
				TNFSF15	2.37E-08	-0.81	Peripheral blood	Japanese‡
				TNFSF8	1.86E-08	0.79	Natural killer cell	Japanese‡
				TNFSF15	8.06E-11	-0.89	Monocyte	Japanese‡
				TNFSF15	1.34E-39	-0.73	Plasmacytoid Dendritic Cell	ImmuNexUT†
				TNC	2.80E-13	0.38	Memory CD8 T cell	ImmuNexUT†
				TNFSF8	1.19E-07	0.25	Natural killer cell	ImmuNexUT†
11q13 (*FCHSD2*)	rs1783598	C	T	*STARD10*	2.35E-46	14.30	Whole Blood	eQTLGen#
				ARAP1	1.88E-07	-5.21	Whole Blood	eQTLGen#
				STARD10	3.05E-22	-0.46	Neutrophil	ImmuNexUT†
12q24 (*NCOR2*)	rs885026	C	A	*NCOR2*	6.33E-20	9.14	Whole Blood	eQTLGen#
				NCOR2	5.56E-12	0.22	Naive CD4 T cell	ImmuNexUT†

Effect, effect size; *P* , *P* value; SNP, single nucleotide polymorphism.

*Tissues from GTEx database: 670 samples of whole blood, 174 samples of small Intestine - terminal Ileum, 368 samples of colon - transverse, and 318 samples of colon - sigmoid (ref. 36).

#31,684 whole blood samples from 37 eQTLGen Consortium cohorts (ref. 14).

†9,852 samples consisted of 28 immune cell types from 416 donors from the ImmuNexUT (ref. 15).

‡105 whole blood samples from healthy controls in Japanese (ref.38).

No eQTLs were identified for the remaining loci (*OTUD3, IL23R,* and *USP12*).

of *CD200* in UC than CD (Table 8). *CD200*, a membrane glycoprotein containing an immunoglobulin domain, is involved in suppression of T-cell proliferation and interferon (IFN)-γ production through interaction with the CD200 receptor (CD200R) (16, 17).

3.3 Replications of loci with divergent effects on CD and UC

To investigate whether the associations identified in this study were replicable in independent datasets, we applied a recently developed method (case-case GWAS; CC-GWAS) (18) to test for differences in allele frequency between CD and UC using the summary statistics from the respective case–control GWASs of European or East Asian origin. In Europeans, *OTUD3* and MHC showed a statistically significant CC-GWAS$_{OLS}$ $P < 5 \times 10^{-8}$ and CC-GWAS$_{exact}$ $P < 1 \times 10^{-4}$, and 3 additional loci including *IL23R*, *CD200* and *USP12* showed a nominal significance level with CC-GWAS$_{OLS}$ and CC-GWAS$_{exact}$ $P < 0.05$ (Table 9 and Figure 7). Of the 4 loci available in the summary statistics of East Asians, only *TNFSF15* locus was statistically significant, and 2 additional loci including *IL23R* and *USP12* showed a nominal significance level with CC-GWAS$_{OLS}$ and CC-GWAS$_{exact}$ $P < 0.05$ (Table 9). The most significant association of the CC-GWAS between CD versus UC in Europeans was rs2076756, which is located in an intron of *NOD2*, and that in East Asians was rs9268831, located 15kb away from the 3′-end of HLA-DRA (Table 10 and Figure 7).

3.4 Polygenic risk score analysis

We calculated the polygenic risk score (PRS) using the fixed-effects meta-analysis data from cohorts I and II as a base file and genotype profiles of independent cohort III (738 CD and 601 UC cases) as the target file. The best-fit PRSs explained the highest phenotypic variance of 22.6% at a threshold of $P < 2.33 \times 10^{-5}$ (based on 24 SNPs) and 19.1% at $P < 5 \times 10^{-8}$ (Figure 8A and B). Although the PRS observed at the best *P* value threshold (*P* value of the Shapiro-Wilk test = 2.11×10^{-4}) showed a skewed distribution, it showed a normal distribution when the samples were stratified based on the genotype of the most significantly associated SNP in the HLA region (rs9270965); this finding is likely due to a mixture of normal distributions (Figure 9). The mean values of best-fit PRSs showed statistically significant differences (*P* value in T test < 2.2×10^{-16}) between 738 patients with CD and 601 patients with UC; however, there were no significant differences between patients with ileal, ileocolonic, and colonic CD (Figure 8C). We also calculated the AUC using the best-fit PRSs to predict accuracy in correctly classifying patients with CD and UC. The AUC value was 0.74 (95% CI = [0.71 − 0.77]), suggesting that PRSs provide acceptable discrimination between patients with CD and UC (Figure 10). Following removal of the largest-effect locus MHC, the PRS explained 11.0% variance at the threshold of $P < 2.33 \times 10^{-5}$ and was based on 23 SNPs.

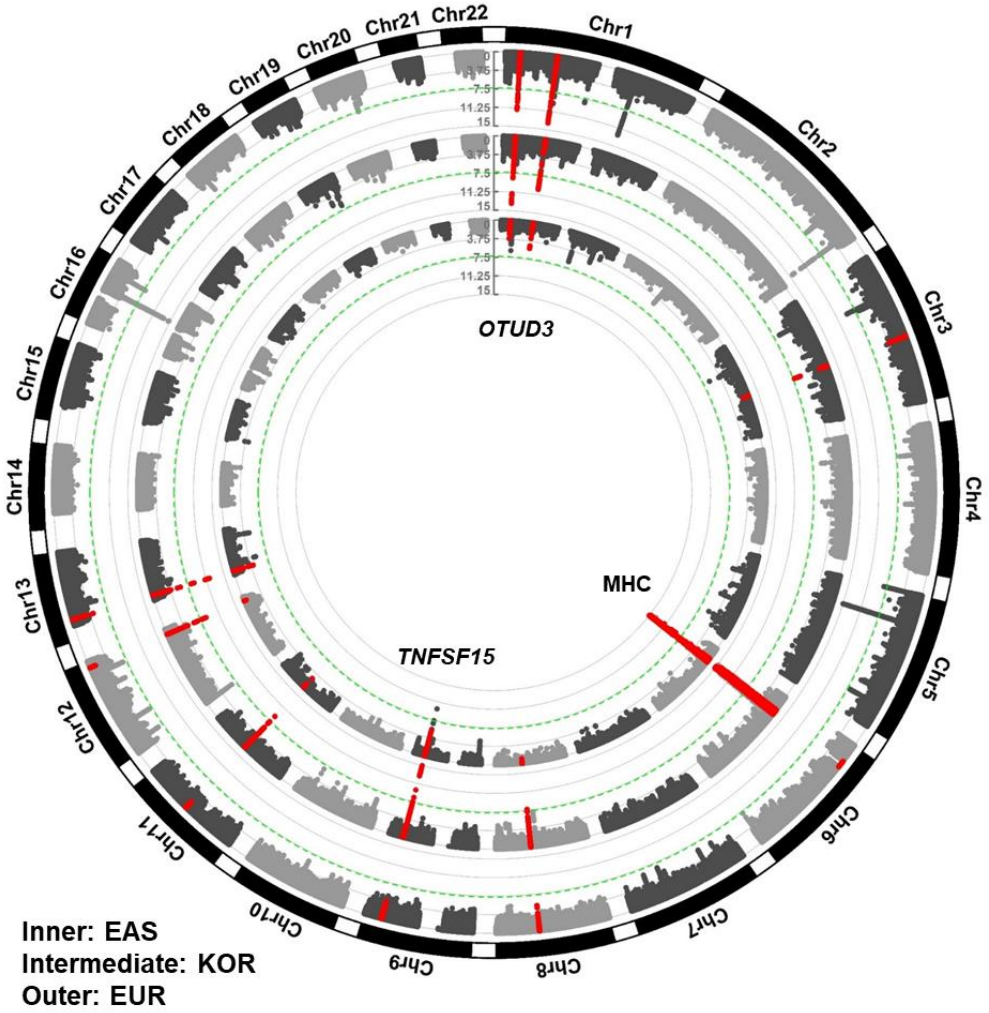


Figure 7. Replication of the loci with divergent effects on CD and UC in other populations using a case-case GWAS (CC-GWAS). In the circular Manhattan plot for CD versus UC, SNPs located in the 9 identified loci are colored red. The green line is the genome-wide significance threshold (CC-GWAS$_{OLS}$ $P < 5 \times 10^{-8}$). The inner-plot represents CD versus UC among East Asians (1,690 CD/ 1,134 UC cases); The intermediate-plot represents CD versus UC among Koreans (2,359 CD/ 2,175 UC cases); The outer-plot represents CD versus UC among Europeans (5,956 CD/ 6,968 UC cases).

Table 9. Associations of the 9 loci from the CC-GWAS analysis in East Asians and Europeans.

					East Asians*							Europeans*						
					CD		UC		CC-GWAS#			CD		UC		CC-GWAS#		
CHR	SNP	Candidate gene	Position (hg19)	Minor allele†	*P*	OR (95% CI)	*P*	OR (95% CI)	P_{OLS}	P_{exact}	Freq‡	*P*	OR (95% CI)	*P*	OR (95% CI)	P_{OLS}	P_{exact}	Freq‡
1	rs58428083	*OTUD3*	20,190,931	C	.	.	.	.	.	.	0.26	8.19E-01	1.01 (0.96-1.05)	9.12E-22	0.82 (0.78-0.86)	4.18E-12	9.30E-13	0.49
1	rs17129698	*IL23R*	67,654,072	A	1.42E-03	1.37 (1.18-1.57)	3.69E-01	0.89 (0.64-1.14)	6.46E-04	1.22E-03	0.04	2.97E-01	1.06 (0.95-1.17)	7.67E-02	0.91 (0.80-1.01)	2.82E-02	2.69E-02	0.03
3	rs9842650	*CD200*	112,069,392	C	3.16E-02	1.11 (1.01-1.20)	9.44E-01	1.00 (0.89-1.11)	8.06E-02	1.14E-01	0.33	1.45E-02	1.06 (1.01-1.11)	7.01E-01	0.99 (0.95-1.03)	1.89E-02	2.08E-02	0.35
6	rs9270965	MHC	32,573,471	A	.	.	.	.	.	.	0.29	5.41E-01	0.98 (0.93-1.04)	2.30E-11	1.19 (1.14-1.24)	4.20E-08	1.97E-08	0.35
8	rs405734	*RIPK2*	90,768,439	G	.	.	.	.	.	.	0.57	6.46E-01	1.01(0.97-1.06)	1.04E-01	1.04 (0.99-1.08)	4.15E-01	3.91E-01	0.62
9	rs722126	*TNFSF15*	117,592,778	G	7.97E-32	0.57 (0.47-0.66)	4.93E-01	0.97 (0.87-1.06)	1.53E-20	5.89E-17	0.43	9.65E-07	0.88 (0.83-0.93)	9.00E-07	0.89 (0.84-0.94)	6.21E-01	7.64E-01	0.30
11	rs1783598	*FCHSD2*	72,851,463	C	.	.	.	.	.	.	0.41	9.88E-01	1.00 (0.94-1.06)	9.62E-01	1.00 (0.95-1.05)	9.60E-01	9.59E-01	0.19
12	rs885026	*NCOR2*	125,032,789	A	.	.	.	.	.	.	0.39	6.02E-01	1.01 (0.97-1.06)	3.81E-01	0.98 (0.94-1.02)	2.76E-01	2.72E-01	0.44
13	rs1359946	*USP12*	27,536,972	A	8.54E-01	0.99 (0.88-1.10)	4.18E-05	1.27 (1.38-1.15)	4.78E-04	1.73E-04	0.24	7.74E-01	1.01 (0.95-1.07)	3.84E-09	1.17 (1.12-1.22)	3.77E-05	2.12E-05	0.19

CC-GWAS, case-case genome-wide association study; CD, Crohn's disease; CI, confidence interval; hg19, human genome version 19; OR, odds ratio; *P* , *P* value; Position, chromosome position; MAF, minor allele frequency; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

*Summary statistics of East Asians (1,690 CD cases and 3,719 controls/ 1,134 UC cases and 3,719 controls) or Europeans (5,956 CD cases and 14,927 controls/ 6,968 UC cases and 20,464 controls) (PMID: 26192919).

#Statistical method to compare allele frequency between cases of two disorders based on the respective case–control GWAS summary statistics (PMID: 33686288). The CC-GWAS calculated *P* values of two methods of ordinary least squares ($CC\text{-}GWAS_{OLS}$) and $CC\text{-}GWAS_{exact}$ to control type I error.

†Minor allele frequency of Koreans (2,359 CD and 2,175 UC).

‡Allele frequency of 1,008 East Asians or 1,006 Europeans from the 1000 Genomes Project database.

Table 10. Significant loci from the CC-GWAS analysis in East Asians and Europeans (CC-GWAS$_{OLS}$ $P < 5 \times 10^{-8}$ and CC-GWAS$_{exact}$ $P < 10^{-4}$).

(A) East Asians

CHR	Position (hg19)	SNP	Gene	CCGWAS$_{EAS}$*	
				P_{OLS}	P_{exact}
6	32,427,748	rs9268831	MHC	1.30E-53	3.38E-54
9	117,592,638	rs2006996	*TNFSF15*	2.90E-24	2.46E-20

(B) Europeans

CHR	Position (hg19)	SNP	Gene	CCGWAS$_{EUR}$*	
				P_{OLS}	P_{exact}
16	50,756,881	rs2076756	*NOD2*	3.37E-41	2.27E-39
2	234,161,769	rs6431654	*ATG16L1*	2.46E-21	3.21E-20
1	20,171,860	rs6426833	*OTUD3*	2.65E-17	3.01E-18
5	40,446,549	rs4957294	*TTC33*	1.06E-15	7.60E-15
1	67,667,936	rs1977160	*IL23R*	2.57E-14	2.72E-13
1	172,857,050	rs6704109	*TNFSF18*	3.03E-13	7.63E-13
1	70,991,829	rs1995301	*CTH*	2.48E-10	2.24E-10
3	53,037,695	rs2564917	*SFMBT1*	8.33E-10	6.51E-10
1	114,377,568	rs2476601	*PTPN22*	6.16E-09	8.38E-09
5	537,890	rs56108664	*MIR4456*	1.16E-08	9.07E-09
16	73,155,804	rs11648199	*ZFHX3*	2.61E-08	2.97E-08
6	32,573,471	rs9270965	MHC	4.20E-08	1.97E-08
20	43,068,239	rs6031606	*HNF4A*	4.52E-08	3.41E-08

CC-GWAS, case-case genome-wide association study; CHR, Chromosome; hg19, human genome version 19; P_{OLS}, ordinary least squares P value; P_{exact}, exact P value ;Position, chromosome position; SNP, single nucleotide polymorphism.

*Summary statistics of East Asians (1,690 CD and 1,134 UC) or Europeans (5,956 CD and 6,968 UC) (PMID: 26192919).

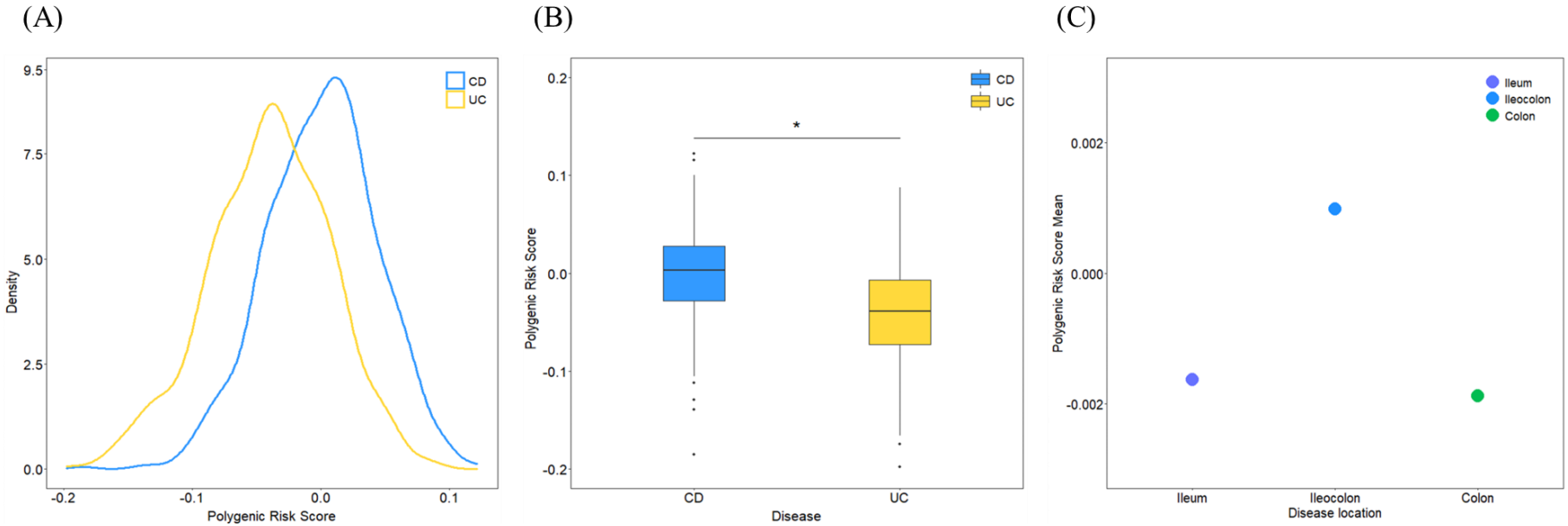
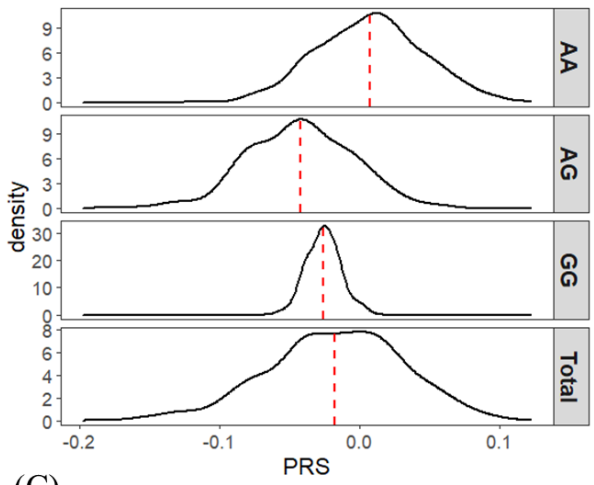
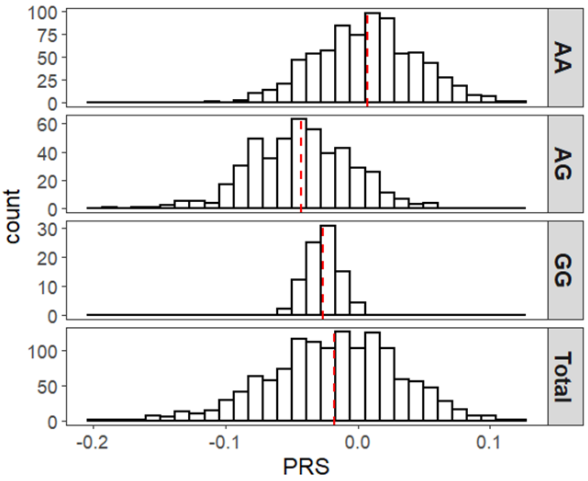


Figure 8. PRS analysis of patients with CD and UC. (A) Distributions and (B) Box plot among patients with CD versus those with UC in cohort III (738 CD / 601 UC; *P*-value by two tailed t-test = 2.22×10^{-16}); (C) Distribution of patients with CD with disease location information in cohort III (colon: 22, ileocolon: 523, ileum: 189).

(A)



(B)



(C)

	Shapiro-Wilk test			
	rs9270965			Total
	AA	AG	GG	
Shapiro-Wilk *P*-value	0.69	0.23	0.89	2.11×10^{-4}

Figure 9. Normal distributions of the PRS for cohort III based on cohorts I and II. (A) Density plot, (B) histogram, and (C) Shapiro-Wilk normality tests of whether variables are normally distributed, based on the PRS separated by the most significant SNP (rs9270965) in the major histocompatibility region (chromosome 6: 25−34 Mb). The red line indicates the mean PRS.

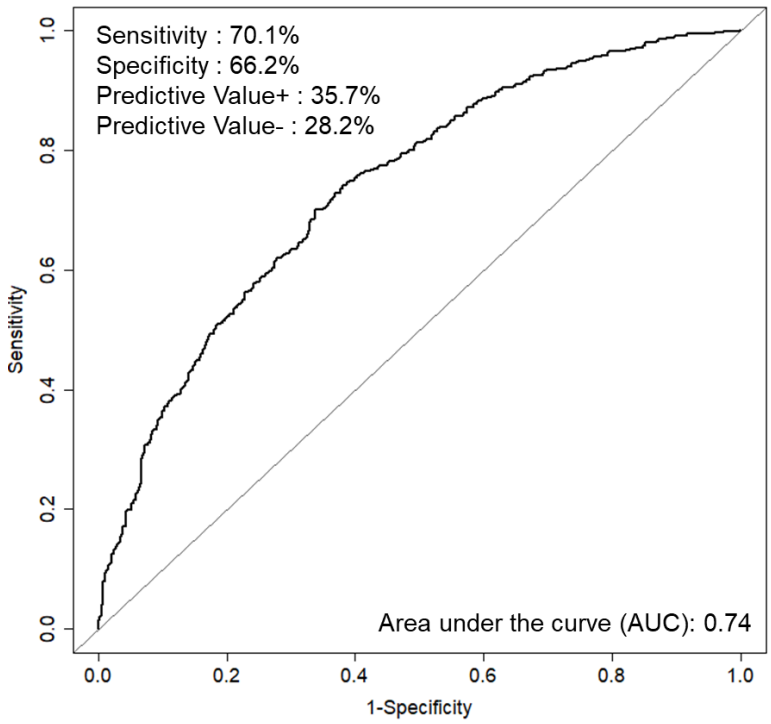


Figure 10. R**eceiver operating characteristic (ROC) curve to predict accuracy in correctly classifying patients with CD and UC (738 CD/ 601 UC cases).** The AUC was 0.74, suggesting acceptable discrimination between CD and UC.

4. DISCUSSION

By comparing individual-level data of 3,131 CD cases versus 2,794 UC cases in Korean population, we identified 2 novel and 7 established susceptibility loci contributing to the phenotypic differences between CD and UC. Despite substantial overlap between CD and UC, replication of 6 loci (MHC, *TNFSF15*, *OTUD3*, *USP12*, *IL23R*, and *CD200*) in the independent cohorts of Europeans and/or East Asians support small but widespread differences in genetic architecture between CD and UC.

We explored the application of our case–case direct comparison findings to estimate case-case polygenic risk prediction. Recent studies have tried to link PRS analysis to the clinical decision-making including disease screening, therapeutic intervention, and life planning (19-21). In the PRS analysis using an independent cohort of 738 CD and 601 UC cases as target data, the PRSs explained up to 22.6% of the phenotypic variance and showed significantly different mean values between patients with CD and those with UC (Table 11). Furthermore, the AUC value based on the PRSs was 0.74 (95% CI = [0.71 − 0.77]), supporting that the PRS analysis using the CD-UC GWAS has potential to support clinical diagnosis of CD and UC.

Finally, we aimed to quantify the genetic overlap between CD and UC, as this has not previously been systematically examined using genome-wide data. LDSC analysis showed that there was a lower genetic correlation between CD and UC in Koreans (r_{gKOR} [SE] = 0.2 [0.13], P = 0.13) than in Europeans (r_{gEUR} [SE] = - 0.6 [0.13], $P = 6.22 \times 10^{-58}$). This finding might be due to trait-specific genetic contributions in East Asians.

Notably, there was a distinct contribution of HLA to CD and UD in Asians. With the development of GWASs, there have been attempts at genome-wide molecular phenotyping in IBD (22). The most important achievement is the distinction between ileal and colonic CD (4, 23-25). Interestingly, in Asian patients with CD, unlike in European patients, the proportions of colonic CD are low (26), and the reason for this difference remains unknown. The largest European study to evaluate genotype-phenotype relations reported that a colonic disease location for CD was better predicted by the HLA susceptibility alleles of UC than those of CD (4). The most consistent genetic risk with UC is the rare HLA class II allele HLA-DRB*0103, which has a frequency of < 2% in European populations. This allele is strongly associated with colonic CD, particularly with isolated colonic disease with of an OR from 5.1 to 18.5 compared with CD at other sites (27). This allele is absent in Koreans/Japanese (7); however, the major risk haplotype spanning HLA-Cw*1202-B*5201-DRB1*1502 in Asian patients with UC reduces the risk of CD (5, 7). After removing HLA, the genetic correlation between CD and UC increased (r_{gKOR} [SE] = 0.47 [0.11],

Table 11. Variance explained in CD-UC status by PRS.

Phenotype	Target data	Variance explained	*P*	Number of SNPs used
CD versus UC	Cohort 3	22.60%	2.33E-05	24
	Cohort 3 excluding the MHC region	19.10%	5.00E-08	4
	Cohort 3 excluding the MHC region	11.00%	2.33E-05	23

CD, Crohn's disease; *P*, *P* value for variance explained; PRS, polygenic risk score; SNP, single nucleotide polymorphism; UC, ulcerative colitis.
For the MHC region (chromosome 6:25−34Mb, hg19), only the most significant SNP was selected from Korean GWAS to minimize over-fitting

$P = 1.43 \times 10^{-5}$) (Table 12), for which was a significant increment (Fisher's Z-transformation method, Z-score = 1.69; $P = 0.045$); furthermore, the variance explained was decreased, suggesting that the rather low genetic correlation between CD and UC in the Korean population might be driven by HLA. This was consistent with the finding of a previous Japanese report that the most significant HLA haplotype associated with UC reduced the risk of CD (5) and with the findings of our previous report that in HLA, the effects for CD were more population-specific than those for UC (7).

We recognize several limitations of our study. First, the statistical power to detect SNPs with a rare MAF or low effect size was limited due to our small sample size. Particularly, due to the small sample size of patients with colonic CD, we could not replicate the findings by Cleynen et al. (4) that ileal CD and colonic CD are quite distinct in terms of genetics. Further studies with a larger sample size are required to identify genetic differences between ileal vs. colonic CD in the Korean population. Second, as Immunochip data were used to increase the sample size, the SNP coverage was almost cut in half after performing the meta-analysis with the GWAS chip datasets. The residual phenotypic variance between CD and UC needs to be explained using additional GWASs with increased sample sizes and SNP coverage in the future. Third, we also acknowledge the diagnostic uncertainties and difficulties related to IBD, despite the careful and repeated evaluations performed with systematic follow-up of the present cohort.

Our GWAS between Korean patients with CD and UC provide new insights into the genetic differences between these two diseases with similar symptoms, which might be useful in improving their diagnosis and treatment. Future studies with large-scale data are needed to verify our observation in diverse populations.

Table 12. Genetic correlation between Koreans and Europeans using LDSC.

(A) Original data

	KOR_CD / KOR_UC	KOR_CD / EUR_CD	KOR_CD / EUR_UC	KOR_UC / EUR_CD	KOR_UC / EUR_UC	EUR_CD / EUR_UC
Sample	6,040 / 5,993	6,040 / 20,883	6,040 / 27,432	5,993 / 20,883	5,993 / 27,432	20,883 / 27,432
overlapping SNPs	895,361	921,871	922,267	923,056	923,438	1,202,265
r_g (se)	0.20 (0.14)	0.47 (0.10)	0.41 (0.10)	0.26 (0.10)	0.54 (0.13)	0.67 (0.06)
P	0.13	4.73×10^{-06}	2.74×10^{-05}	9.20×10^{-03}	2.21×10^{-05}	2.52×10^{-33}
z-score	1.45	4.58	4.19	2.6	4.24	12.03

(B) Excluding the HLA region

	KOR_CD / KOR_UC	KOR_CD / EUR_CD	KOR_CD / EUR_UC	KOR_UC / EUR_CD	KOR_UC / EUR_UC	EUR_CD / EUR_UC
Sample	6,040 / 5,993	6,040 / 20,883	6,040 / 27,432	5,993 / 20,883	5,993 / 27,432	20,883 / 27,432
overlapping SNPs	887,373	918,137	918,531	919,535	919,914	1,199,223
r_g (se)	0.47 (0.11)	0.54 (0.10)	0.47 (0.10)	0.29 (0.11)	0.61 (0.14)	0.69 (0.05)
P	1.43×10^{-05}	1.57×10^{-08}	7.73×10^{-07}	8.40×10^{-03}	6.83×10^{-06}	2.46×10^{-38}
z-score	4.39	5.65	4.94	2.64	4.50	12.95

CD, Crohn's disease; *P*, *P* value; r_g, genetic correlation; se, standard error; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

Summary statistics of Korean (1,621 CD cases and 4,419 controls/ 1,574 UC cases and 4,419 controls) (PMID: 33853113) and Europeans (5,956 CD cases and 14,927 controls/ 6,968 UC cases and 20,464 controls) (PMID: 26192919).

Web resources

The URLs for data presented herein are as follows:

METAL, http://www.sph.umich.edu/csg/abecasis/metal/

The 1000 Genomes Project, http://www.1000genomes.org/

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

IIBDGC, www.ibdgentics.org

RegulomeDB v2, http://www.broadinstitute.org/mammals/haploreg/haploreg.php

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

eQTL Blood Browser, http://www.genenetwork.nl/bloodeqtlbrowser/

Geuvadis/1000 Genomes resources, http://www.ebi.ac.uk/Tools/geuvadis-das/

References

1 Khor, B., Gardet, A. and Xavier, R.J. (2011) Genetics and pathogenesis of inflammatory bowel disease. *Nature*, **474**, 307-317.

2 Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T. *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.*, **47**, 979-986.

3 de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G. *et al.* (2017) Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics*, **49**, 256-261.

4 Cleynen, I., Boucher, G., Jostins, L., Schumm, L.P., Zeissig, S., Ahmad, T., Andersen, V., Andrews, J.M., Annese, V., Brand, S. *et al.* (2016) Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*, **387**, 156-167.

5 Okada, Y., Yamazaki, K., Umeno, J., Takahashi, A., Kumasaka, N., Ashikawa, K., Aoi, T., Takazoe, M., Matsui, T., Hirano, A. *et al.* (2011) HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. *Gastroenterology*, **141**, 864-871.e865.

6 Yang, S.K., Hong, M., Zhao, W., Jung, Y., Tayebi, N., Ye, B.D., Kim, K.J., Park, S.H., Lee, I., Shin, H.D. *et al.* (2013) Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic susceptibility with Caucasians. *Inflamm. Bowel Dis.*, **19**, 954-966.

7 Han, B., Akiyama, M., Kim, K.K., Oh, H., Choi, H., Lee, C.H., Jung, S., Lee, H.S., Kim, E.E., Cook, S. *et al.* (2018) Amino acid position 37 of HLA-DRβ1 affects susceptibility to Crohn's disease in Asians. *Hum. Mol. Genet.*, **27**, 3901-3910.

8 Jung, S., Ye, B.D., Lee, H.S., Baek, J., Kim, G., Park, D., Park, S.H., Yang, S.K., Han, B., Liu, J. *et al.* (2021) Identification of three novel susceptibility loci for inflammatory bowel disease in Koreans in an extended genome-wide association study. *J. Crohns Colitis*, **15**, 1898-1907.

9 Yang, S.K., Hong, M., Oh, H., Low, H.Q., Jung, S., Ahn, S., Kim, Y., Baek, J., Lee, C.H., Kim, E. *et al.* (2016) Identification of loci at 1q21 and 16q23 that affect susceptibility to inflammatory bowel disease in Koreans. *Gastroenterology*, **151**, 1096-1099.e1094.

10 Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906-913.

11 Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium, Duncan, L. *et al.* (2015) An atlas of genetic correlations across human diseases and traits. *Nat. Genet.*, **47**, 1236-1241.

12 de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.

13 Võsa, U., Claringbould, A., Westra, H.J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S. *et al.* (2021) Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.*, **53**, 1300-1310.

14 Ota, M., Nagafuchi, Y., Hatano, H., Ishigaki, K., Terao, C., Takeshima, Y., Yanaoka, H., Kobayashi, S., Okubo, M., Shirai, H. *et al.* (2021) Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*, **184**, 3006-3021.e3017.

15 Sander, J., Schmidt, S.V., Cirovic, B., McGovern, N., Papantonopoulou, O., Hardt, A.L., Aschenbrenner, A.C., Kreer, C., Quast, T., Xu, A.M. *et al.* (2017) Cellular differentiation of human monocytes is regulated by time-dependent interleukin-4 signaling and the transcriptional regulator NCOR2. *Immunity*, **47**, 1051-1066.e1012.

16 Jenmalm, M.C., Cherwinski, H., Bowman, E.P., Phillips, J.H. and Sedgwick, J.D. (2006) Regulation of myeloid cell function through the CD200 receptor. *J. Immunol.*, **176**, 191-199.

17 Copland, D.A., Calder, C.J., Raveney, B.J., Nicholson, L.B., Phillips, J., Cherwinski, H., Jenmalm, M., Sedgwick, J.D. and Dick, A.D. (2007) Monoclonal antibody-mediated CD200 receptor signaling suppresses macrophage activation and tissue damage in experimental autoimmune uveoretinitis. *Am. J. Pathol.*, **171**, 580-588.

18 Peyrot, W.J. and Price, A.L. (2021) Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. *Nat. Genet.*, **53**, 445-454.

19 Kuchenbaecker, K.B., McGuffog, L., Barrowdale, D., Lee, A., Soucy, P., Dennis, J., Domchek, S.M., Robson, M., Spurdle, A.B., Ramus, S.J. *et al.* (2017) Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J. Natl. Cancer Inst.*, **109**, djw302.

20 Natarajan, P., Young, R., Stitziel, N.O., Padmanabhan, S., Baber, U., Mehran, R., Sartori, S., Fuster, V., Reilly, D.F., Butterworth, A. *et al.* (2017) Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation*, **135**, 2091-2101.

21 Maas, P., Barrdahl, M., Joshi, A.D., Auer, P.L., Gaudet, M.M., Milne, R.L., Schumacher, F.R., Anderson, W.F., Check, D., Chattopadhyay, S. *et al.* (2016) Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.*, **2**, 1295-1302.

22 Furey, T.S., Sethupathy, P. and Sheikh, S.Z. (2019) Redefining the IBDs using genome-scale molecular phenotyping. *Nat. Rev. Gastroenterol. Hepatol.*, **16**, 296-311.

23 Atreya, R. and Siegmund, B. (2021) Location is important: differentiation between ileal and colonic Crohn's disease. *Nat. Rev. Gastroenterol. Hepatol.*, **18**, 544-558.

24 Subramanian, S., Ekbom, A. and Rhodes, J.M. (2017) Recent advances in clinical practice: a systematic review of isolated colonic Crohn's disease: the third IBD? *Gut*, **66**, 362-381.

25 Dulai, P.S., Singh, S., Vande Casteele, N., Boland, B.S., Rivera-Nieves, J., Ernst, P.B., Eckmann, L., Barrett, K.E., Chang, J.T. and Sandborn, W.J. (2019) Should we divide Crohn's disease into ileum-dominant and isolated colonic diseases? *Clin. Gastroenterol. Hepatol.*, **17**, 2634-2643.

26 Park, S.H., Kim, Y.J., Rhee, K.H., Kim, Y.H., Hong, S.N., Kim, K.H., Seo, S.I., Cha, J.M., Park, S.Y., Jeong, S.K. *et al.* (2019) A 30-year trend analysis in the epidemiology of inflammatory bowel disease in the Songpa-Kangdong district of Seoul, Korea in 1986-2015. *J. Crohns Colitis*, **13**, 1410-1417.

27 Ahmad, T., Marshall, S.E. and Jewell, D. (2006) Genetics of inflammatory bowel disease: the role of the HLA complex. *World J. Gastroenterol.*, **12**, 3628-3635.

28 Park, S.H., Yang, S.K., Park, S.K., Kim, J.W., Yang, D.H., Jung, K.W., Kim, K.J., Ye, B.D., Byeon, J.S., Myung, S.J. *et al.* (2014) Long-term prognosis of crohn's disease and its temporal change between 1981 and 2012: a hospital-based cohort study from Korea. *Inflamm. Bowel Dis.*, **20**, 488-494.

29 Lee, H.S., Park, S.H., Yang, S.K., Lee, J., Soh, J.S., Lee, S., Bae, J.H., Lee, H.J., Yang, D.H., Kim, K.J. *et al.* (2015) Long-term prognosis of ulcerative colitis and its temporal change between 1977 and 2013: a hospital-based cohort study from Korea. *J. Crohns Colitis*, **9**, 147-155.

30 Park, S.H., Kim, Y.J., Rhee, K.H., Kim, Y.H., Hong, S.N., Kim, K.H., Seo, S.I., Cha, J.M., Park, S.Y., Jeong, S.K. *et al.* (2019) A 30-year Trend Analysis in the Epidemiology of Inflammatory Bowel Disease in the Songpa-Kangdong District of Seoul, Korea in 1986-2015. *J. Crohns Colitis*, **13**, 1410-1417.

31 Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.

32 Delaneau, O., Marchini, J. and Zagury, J.F. (2011) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179-181.

33 Consortium, G. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318-1330.

34 Jung, S., Liu, W., Baek, J., Moon, J.W., Ye, B.D., Lee, H.S., Park, S.H., Yang, S.K., Han, B., Liu, J. *et al.* (2020) Expression quantitative trait loci (eQTL) mapping in Korean patients with Crohn's disease and identification of potential causal genes through integration with disease associations. *Front. Genet.*, **11**, 486.

35 Ishigaki, K., Kochi, Y., Suzuki, A., Tsuchida, Y., Tsuchiya, H., Sumitomo, S., Yamaguchi, K., Nagafuchi, Y., Nakachi, S., Kato, R. *et al.* (2017) Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat. Genet.*, **49**, 1120-1125.

36 Choi, S.W. and O'Reilly, P.F. (2019) PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience*, **8**, giz082.

국문요약

크론병과 궤양성대장염은 두 가지 주요 염증성 장질환으로, 임상 경과와 치료 반응에서 차이를 보인다. 두 질병의 유전적 차이를 이해하는 것은 질병의 진단과 치료를 개선하는 데 중요하다. 한국인을 대상으로 크론병 환자 2,359명과 궤양성대장염 환자 2,175명의 전장유전체 연관분석을 실시하였고, 독립적인 크론병 환자 772명과 궤양성대장염 환자 619명을 대상으로 검증하였다. 크론병과 궤양성대장염 간 차이를 보이는 2개의 새로운 유전자 좌를 발굴하였다 (*CD200* at 3q13, *NCOR2* at 12q24). 또한, 이전에 보고된 7개의 감수성 유전자 좌 (MHC, *TNFSF15, OTUD3, USP12, IL23R, FCHSD2,* and *RIPK2*) 에서 두 질병 간 유의한 차이를 확인했다. 크론병과 궤양성대장염을 구분하는 9개의 유전자 부위 중 6개의 유전자 좌 (MHC, *TNFSF15, OTUD3, USP12, IL23R,* and *CD200*) 가 독립적인 유럽인과 동북아시아인 코호트에서도 확인되었다. 다유전자 위험 점수 분석에서 크론병과 궤양성대장염의 차이를 설명하는 분산이 22.6%였으며, 수신자 조작 특성 면적 값 0.73으로 두 질병 간의 유전적 특성을 기반으로 구분할 수 있다는 것을 확인하였다. 이러한 결과는 유사한 증상을 가지는 두 질병 간의 유전적 차이에 대한 통찰력을 제공하며, 이를 통해 질병 진단과 치료를 향상시키는 데 도움이 될 것이다.