Characterization of Genetic Loci That Affect Susceptibility to Inflammatory Bowel Diseases in African Americans

Chengrui Huang, Johns Hopkins University
Talin Haritunians, Cedars-Sinai Medical Center
David Okou, Emory University
David Cutler, Emory University
Michael Zwick, Emory University
Kent D. Taylor, Harbor-UCLA Medical Center
Lisa W. Datta, Johns Hopkins University
Joseph C. Maranville, University of Chicago
Zhenqiu Liu, Cedars-Sinai Medical Center
Shannon Ellis, Johns Hopkins University

Only first 10 authors above; see publication for full author list.

Journal Title: Gastroenterology
Volume: Volume 149, Number 6
Publisher: Elsevier | 2015-11-01, Pages 1575-1586
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1053/j.gastro.2015.07.065
Permanent URL: https://pid.emory.edu/ark:/25593/rthgw

Copyright information:
© 2015 AGA Institute.
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed September 24, 2017 4:24 PM EDT
Characterization of Genetic Loci That Affect Susceptibility to Inflammatory Bowel Diseases in African Americans

A full list of authors and affiliations appears at the end of the article.

# These authors contributed equally to this work.

Abstract

**Background & Aims**—Inflammatory bowel disease (IBD) has familial aggregation in African Americans (AAs), but little is known about the molecular genetic susceptibility. Mapping studies using the Immunochip genotyping array expand the number of susceptibility loci for IBD in Caucasians to 163, but the contribution of the 163 loci and European admixture to IBD risk in AAs is unclear. We performed a genetic mapping study using the Immunochip to determine whether IBD susceptibility loci in Caucasians also affect risk in AAs and identify new associated loci.

**Methods**—We recruited AAs with IBD and without IBD (controls) from 34 IBD centers in the US; additional controls were collected from 4 other immunochip studies. Association and admixture loci were mapped for 1088 patients with Crohn's disease (CD), 361 with ulcerative colitis (UC), 62 with IBD type-unknown (IBDU), and 1797 controls; 130,241 autosomal single-nucleotide polymorphisms (SNPs) were analyzed.

**Results**—The strongest associations were observed between UC and HLA rs9271366 \((P=7.5\times 10^{-6})\), CD and 5p13.1 rs4286721 \((P=3.0\times 10^{-6})\), and IBD and \(KAT2A\) rs730086 \((P=2.3\times 10^{-6})\). Additional suggestive associations \((P<4.2\times 10^{-5})\) were observed between CD and IBD and African-specific SNPs in \(STAT5A\) and \(STAT3\); between IBD and SNPs in \(IL23R\), \(IL12B\), and C2 open reading frame 43; and between UC and SNPs near \(HDAC11\) and near \(LINC00994\). The latter 3 loci have not been previously associated with IBD, but require replication. Established Caucasian associations were replicated in AAs \((P<3.1\times 10^{-4})\) at \(NOD2\), \(IL23R\), 5p15.3, and \(IKZF3\). Significant admixture \((P=3.9\times 10^{-4})\) was observed for 17q12-17q21.31 (IZKF3 through \(STAT3\)), 10q11.23-10q21.2, 15q22.2–15q23, and 16p12.2–16p12.1. Network analyses showed significant

---

1 Corresponding author Steven R. Brant, M.D., Johns Hopkins University School of Medicine, Meyerhoff Inflammatory Bowel Disease Center, 1501 E. Jefferson St., B136, Baltimore, MD 21231. sbrant@jhmi.edu; Phone: 410-955-9679; Fax: 410-502-9913.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Disclosures:** No authors have conflicts of interest to declare.

**Author Contributions:** Study concept and design: SRB SK DPBM LWHK JHC CH DJC. Analysis and interpretation of data: CH TH DTO DJC MEZ KDT JCM ZL CLS LWHK DPBM SRB SK. Data acquisition and material support: LWD JSA RNB JHC RKC TD TAD RHD IJS IJK SZH KJI KEK HK MK MDK IK RK BSK JFJ AK JHK ML PM DEM BOO AP IDR JIR SS EJS MSS AS SRT JV MHW SLB RPK SSR ADR DPBM SRB SK. Drafting manuscript: CH SRB. Critical revision of the manuscript for important intellectual content: CH TH DTO HK CLS MHW SSR ADR JHC LWHK DPBM SRB SK.

Author names in bold designate shared co-first authors.
enrichment (false discovery rate <1e–5) in genes that encode members of the JAK–STAT, cytokine, and chemokine signaling pathways, as well those involved in pathogenesis of measles.

**Conclusions**—In a genetic analysis of 3308 AA IBD cases and controls, we found that many variants associated with IBD in Caucasians also showed association evidence with these diseases in AAs; we found evidence for loci and variants not previously associated with IBD. The complex genetic factors that determine risk for or protection from IBD in different populations require further study.

**Keywords**
race; ethnicity; genetic variant; intestinal inflammation

---

**Introduction**

Inflammatory bowel disease (IBD) is a complex genetic disorder of immune dysregulation causing chronic idiopathic inflammation of the gastrointestinal tract, estimated to affect about 1.4 million Americans. It comprises two major, genetically related phenotypes: Crohn's disease (CD) and ulcerative colitis (UC).

IBD shares many clinical and immunological characteristics with other complex genetic immune-mediated diseases (IMDs), especially with seronegative auto-inflammatory diseases\(^1\), such as ankylosing spondylitis, psoriasis and primary immune-deficiencies\(^2\), indicating overlapping etiological factors. To facilitate genetic studies on IMDs a custom Illumina array of ~200,000 SNPs, the Immunochip, was designed based on genome wide association (GWA) analyses on Caucasian populations of 12 IMDs including IBD\(^3\). The main purposes were to fine map established associations and to replicate suggestive, but not yet proven, associations\(^3\). The Immunochip also contains ancestry informative markers (AIMs) allowing for genome-wide admixture estimates and adjustment for population stratification.

IBD GWA studies, including those performed using the Immunochip, have expanded the number of IBD susceptibility loci to 163 (including 30 CD- and 23 UC-specific loci)\(^2\), and have enhanced our understanding of IBD immunopathogenesis by identifying key cellular pathways, both known – such as barrier function, the role of T cell subsets and cytokine–cytokine receptor signaling – and unknown – such as autophagy, regulation of interleukin 23 (IL23) signaling, and host defense\(^4\). However, compared to hundreds of IBD genetic studies in Caucasian populations including massive GWA mega-analyses and replication studies like the Caucasian Immunochip Study (CIS)\(^2\), only a handful of IBD-associated gene variations have been evaluated in African Americans (AAs), in relatively small sample sizes of a few hundred cases and controls, and only for CD, not UC\(^5\)-\(^7\).

AAs are a recently admixed population derived from an average of approximately 80% West African and 20% European ancestries\(^8\). IBD prevalence is lower in AAs than Caucasian Americans (CAs) possibly as a result of both genetic and environmental differences\(^9\)-\(^11\). IBD sibling risk in AA IBD patients is relatively high (2.5%), suggesting underlying genetic
risk factors are in-part responsible for IBD in AAs, albeit lower than that observed for Caucasians (4.6%).

An admixed population is one where two or more previously separated populations have interbred. Loci that have different allele frequencies in the founder populations become correlated because of gene flow, a phenomenon known as admixture linkage disequilibrium (ALD). AAs are a typical example of an admixed population, where there has been recent introduction of European genetic lineages into a West-African-derived population. A powerful technique called mapping by ALD (MALD) can be used on admixed populations such as AAs and the underlying assumption is that the difference in disease frequency is due in part to differences in allele frequencies of causal variants between populations.

We therefore undertook an evaluation of the Immunochip in ~4000 AA IBD cases and controls, primarily to determine the importance of the 163 established CIS IBD loci in the understudied AA population and to identify novel IBD loci, including loci identified by MALD.

**Patients and Methods**

**Study Population and Phenotyping**

The study population included unrelated self-identified non-Hispanic AA volunteers recruited from three coordinating centers: (1) Johns Hopkins: Multicenter African American IBD Study (MAAIS) coordinated by Johns Hopkins IBD Genetics Research Center (GRC) of the NIDDK IBD Genetics Consortium (IBDGC) with recruitment from 13 collaborating IBD centers and 4 other IBDGC GRCs, and additional AA control samples from the controls from rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and type 1 diabetes (T1D) Immunochip studies; (2) Emory: GENESIS AA cohort, an ancillary study of the NIDDK IBDGC, coordinated by Emory University with recruitment of IBD cases and matched controls from 12 of their collaborating IBD centers; and (3) Cedars: Cedars-Sinai Medical Center recruited IBD cases and controls with additional controls from the Pharmacogenetics and Risk of Cardiovascular disease Study Group (PARC) from San Francisco General Hospital and University of California, Los Angeles.

All subjects gave informed consent to participate in genetics research studies in protocols approved by each sites institutional review board. Cases were confirmed as CD, UC, or IBD type undetermined (IBDU) in accordance with the NIDDK IBDGC phenotyping manual. Details regarding controls are described in the Supplement. See Acknowledgements for listing of all recruitment centers.

**Genotyping and Quality Control (QC)**

DNA samples were derived from whole blood. All DNA samples were genotyped using the Immunochip and genotype determinations (allele calls) were made using GenomeStudio version 2011.1 and Genotyping Module Version 1.9.4. All MAAIS samples and the RA, SLE, and T1D controls were genotyped at Feinstein Institute for Medical Research. All Emory, Cedars-Sinai and PARC samples were genotyped at Cedars-Sinai Medical Center Genetics Institute.
Several SNP-wise and sample-wise quality filters were applied (Supplementary Figure 1). Samples were excluded if they had <99% data completeness, differed by >3 standard deviations from the mean heterozygosity for the study, had discrepant gender, or were unexpectedly the first-degree relative of any other sample in the study. Matching of cases and controls was done by determining principal components (PC) using the software EIGENSTRAT\textsuperscript{15}, plotting PC1 against PC2, followed by visual inspection and elimination of outlier samples. Three successive rounds were necessary until a satisfactory matching of cases/controls (Supplementary Figure 2) was obtained.

Local and Global Ancestry Estimation

Because AAs are well-modeled as linear combinations of West African (YRI) and European (CEU) ancestries\textsuperscript{16}, to estimate the locus-specific local ancestry, we chose the WINPOP model in LAMP package\textsuperscript{17} for its fast computation with low error rate\textsuperscript{18}. Average West African (global YRI) ancestry was estimated by using ADMIXTURE\textsuperscript{19}.

Study-wide Association and Admixture, and Replication of 163 SNPs Established in Caucasian Immunochip Study (CIS)

For all analyses, we compared three phenotypes, all IBD, CD, and UC, against the same set of AA controls. We performed association and admixture mapping both under generalized linear models. Sex, recruitment coordinating center (Hopkins, Emory or Cedars), GRC (Feinstein or Cedars), global YRI ancestry, and the first 10 PCs were included in multiple regression with IBD affection status, and those significantly associated with IBD were used as covariates in association and admixture mapping.

We estimated the testing burdens empirically by assessing the autocorrelation (of genotypes for association or local ancestry proportions for admixture) of all the SNPs on each autosomal chromosome for each individual, and then summing over the 22 chromosomes and averaging across individuals\textsuperscript{20}.

The testing burdens of independent SNPs or admixture regions, specific to our dataset, were 23,639.4 and 128.4 for association and admixture mapping respectively. Association peaks with \( p < 2.1 \times 10^{-6} \) (5% false positive rate corrected for number of independent SNPs, i.e. \( 0.05/23,639.4 \)) were marked as significant, while those with \( 2.1 \times 10^{-6} \leq p < 4.2 \times 10^{-5} \) were marked as suggestive (i.e. one false positive per study-wide test burden). Admixture peaks with \( p < 3.9 \times 10^{-4} \) (0.05/128.4) were marked as significant.

For chromosomes with multiple significant/suggestive association SNPs, conditional regression was performed to determine number of independent signals. Specifically for \( m \) SNPs, all possible \( m \times (m-1)/2 \) pairwise combinations were tested. For each pair if the target SNP remained significant in the presence of the conditional SNP then they are independent of each other; otherwise, they have arisen from the same signal.

To evaluate the 163 loci with reported genome-wide significance in the CIS, we tested the most statistically significant SNPs at each locus and present on the Immunochip (Maximal-CIS SNPs). Criteria for replication was \( p < 3.1 \times 10^{-4} \) (0.05/163).
Gene Ontology was utilized to functionally annotate the genes within each significant admixture region, using the PANTHER Classification System (http://pantherdb.org). Genes with immunological or gastrointestinal-relevant functions were considered to be suggestive candidate genes within the region.

**eQTL Evaluation**

For significant/suggestive association SNPs we also performed local (*cis*)- and distant (*trans*)- expression quantitative trait locus (eQTL) mapping with R package MatrixEQTL\textsuperscript{21} on peripheral blood mononuclear cells (PBMCs) drawn from a different set of 85 unrelated AA study subjects profiled with Illumina HumanHT-12 v4 Expression BeadChips and genotyped on Illumina HumanOmni1 or HumanOmni2.5 BeadChips\textsuperscript{22}. We defined gene-SNP pairs within 1Mb of another as local, otherwise distant, and those with false discovery rate (FDR) <0.05 were considered significant.

**Network Analysis**

We first selected genes with association p<0.005, and then the identified genes annotated with multiple biologically functional databases including Reactome\textsuperscript{23}, human protein reference databases (www.hprd.org), and NCI/Nature Pathway Interaction Database. The networks were then constructed from the known interactions in those databases. In addition, we identified the top KEGG pathways associated with each network using the enrichment analysis tool in STRING (http://string-db.org/). See Supplement for additional details.

**Results**

After QC measures were performed, 3,308 samples (1,511 IBD cases and 1,797 controls) and 130,241 unique autosomal SNPs remained for association and admixture analyses (Supplementary Figures 1 and 2). Average West African ancestry was higher in controls than cases (81.7±9.7% vs. 80.0±10.3%, p<1e-5), and also differed by coordinating centers (p<1e-10), with YRI proportion lowest for AAs from the American West Coast, similar to other genetic studies\textsuperscript{24} (Supplementary Table 1).

Five variables were significantly associated with IBD (p<0.05) and therefore were included as covariates for all subsequent analyses: sex (p=5e-7), recruitment coordinating center (p=2e-5), GRC (p=4e-4), PC2 (p=0.015), and global ancestry (p=0.03). Quantile-quantile plots for association mappings and genomic inflation factors are shown in Supplementary Figure 3.

**Replication of Maximal-CIS SNPs**

152 of the 163 Maximal-CIS SNPs passed QC. The remaining 11 SNPs did not have any tagging (r\textsuperscript{2}<0.6) SNPs that could be used as alternative markers. Five SNPs met criteria for replication (p<3.1e-4): rs5743289 that tags the three common *NOD2* mutations, rs11209026 that encodes the *IL23R* R381Q protective variant, rs1801274 that encodes the *FCGR2A* H167R risk variant, rs11742570 in the 5p13.1 gene desert near *PTGER4* and rs12946510 on 17q21 just 3’ of *IKZF3* (Table 1).
For most of the 152 SNPs, the AA odds ratios are in the same direction as those from the CIS (correlation=0.78, Figure 1). 81 SNPs had AA ORs within the 95% CIS confidence intervals (CI), including one replicated SNP (rs11742570). Nominal levels of significance (p<0.05 and same risk allele) were observed for 41 of the 152 (27%) variants (Supplementary Table 2).

**Study-wide Association**

Association results for IBD, CD and UC are shown in Figure 2 (left panels). Suggestive association for IBD was observed at the following 7 loci for 15 SNPs (Table 2): 1p at IL23R; 2p near the p-telomere and 4 Mb upstream of an associated locus in the CIS near C2orf43; 5p13.1 gene desert (5 SNPs in tight LD, pairwise r² ranging from 0.52 to 1.00); 5q 70kb telomeric of IL12B (2 SNPs in tight LD, r²=0.87); and 17 in the STAT5A/STAT3 region (6 SNPs in 3 LD blocks). Conditional regression revealed 3 independent signals from the latter region: three STAT5A SNPs (rs7220367, rs7217884 and rs13380828, all monomorphic in CEU, r² ranging from 0.46 to 0.99) in one block; two SNPs 200 kb apart (rs730086 in KAT2A and rs1053004 in STAT3, r²=0.44); and rs7224339. IBD, being the phenotype with the largest sample size had greatest power to detect association evidence. However, 72% of cases had CD, and not surprisingly SNPs at 5p15.3 (PTGER4), IL12B, KAT2A and STAT5A also showed suggestive association for CD, with CD consistently showing greater association evidence at 5p15.3 than all IBD (Table 2).

For UC, the major peak located in the HLA region, with greatest association for rs9271366 (p=7e-6). Conditional analysis revealed that all suggestive HLA SNPs could be accounted for by rs9271366 (Supplementary Figure 4). Two SNPs showed suggestive associations on chromosome 3, at 13Mb and at 64Mb (near HDAC11 and LINC00994), neither within 5Mb of any established IBD loci.

Raw genotyping intensities were visually examined for all replicated/suggestive SNPs and only rs35990859 (at IL12B) had sub-optimal cluster separation (data not shown).

**Admixture Association**

We observed four chromosomal regions of significant admixture association (Figure 2, right panels and Table 3) with all having significant evidence for both IBD and CD except for 16p12 (IBD-only). For all of these regions, CEU (European) ancestry increases risk and YRI (West African) ancestry was protective. Just below significance was an area with increased CD risk from CEU, maximal at rs2111112 (p=4.9e-4) and 3 Mb from NOD2 (with increased CEU ALD extending through NOD2). Conditioning on NOD2 genotype rs5743289 weakened this ALD evidence (p=0.017). A list of annotated genes with immune-related functions that map to each region is included in Supplementary Table 3.

**eQTL Evaluation**

We conducted PBMC eQTL mapping for 26 (5 significant replication and 21 suggestive study-wide association) SNPs with eQTL data. After controlling for multiple comparisons, 0 distant and 6 local gene-SNP pairs showed significant association (Supplementary Table 4): rs1053004 on chromosome 17 was associated with KRT19 (p=1.7e-4) and with TTC25.
expression; 3 SNPs (rs1876141 \(p=5.5e^{-4}\), rs6866402 \(p=7.5e^{-4}\), and rs1505994 \(p=7.5e^{-4}\)) in tight LD \(r^2>0.99\) at 5p13.1 with Complement Component 6 (C6) expression; and rs1801274 with CD84 expression \(p=1.2e^{-3}\). PTGER4 expression was evaluated but was not associated with any SNP evaluated \(p>0.05\).

**Network Analysis**

We found multiple networks demonstrating significant influence for IBD, CD and UC in AAs even after FDR and Bonferroni corrections (Table 4). Cytokine-cytokine receptor interaction was the dominant pathway for CD and IBD overall, but for UC was second in effect to Measles (which also played a significant but less prominent role in IBD and CD). Jak-STAT and Chemokine signaling pathways were included in top 3 networks for IBD and CD (and Jak-STAT alone for UC). In addition, T cell receptor signaling pathway was one of the top pathways in UC and IBD, but not CD.

**Discussion**

In this study of IBD genetics in the AA population, we assembled the largest set of cases and controls (more than 4 times larger than any prior AA IBD study and the first genetic study of AA UC), evaluated for replication the majority of established IBD associations and interrogated the majority of known IBD loci for novel associations, and performed pathway and eQTL analyses to further inform about the nature of AA IBD genetics. Although our genotyping platform (not being GWA) limited association mapping primarily to replicating known associations and interrogating established, immunologically related loci, we also used the more powerful method of MALD to identify novel loci throughout the autosomal genome and to complement the association mapping.

We replicated, using stringent criteria \(p<3.1e^{-4}\), five CIS loci. As our study has more CD than UC cases, most of our replications were for CD. Not surprisingly we replicated loci with greater impact in other populations. Three of the five replications, NOD2, IL23R and 5p13.1, account for the greatest degree of CD variance estimated in CIS. The two other replicated loci also have relatively high OR for Caucasian IBD: FCGR2A and IKZF3 (17th and 47th ranked ORs of the 163 CIS loci).

Two replicated CIS SNPs are more frequent in, and hence IBD risk is more likely to arise from, CEU than YRI genome: rs5743289 that tags the 3 common NOD2 CD-associated mutations rs12946510 at IKZF3 (Table 1). In contrast, the two other replicated SNPs, rs11742570 at 5p13.1 and the rs1801274 at FCGR2A with functional variant H167R, have similar presence in CEU and YRI genomes, whereas the R381Q wildtype albeit risk variant of IL23R at rs11209026 is more common in YRI than CEU genomes.

Given our sample size, we had 80% power to replicate Maximal-CIS SNPs with OR>1.3 and MAF>0.2. Supplementary Figure 5 illustrates our power at different ORs/MAFs at three levels of significance. Although our study was underpowered a-priori for genome-wide association (and obtaining enough AA individuals for genetic studies, as statically powerful as those in Caucasians, may be impractical for years to come), we have identified SNPs in the AA population at a level of association for which to target for independent replication.
Two replicated loci also had additional SNPs with suggestive association: 5p13.1 and IL23R. At 5p13.1 rs4286721 showed greatest association with CD (p=2.97e-6), only marginally below study-wide significance (p<2.1e-6). The suggestive SNPs are in LD, are protective and more frequent in YRI than CEU, and, in our PBMC eQTL analysis, correlated with complement component 6 (C6) instead of PTGER4, the associated gene in the original discovery in lymphoblastoid cell lines. C6 is a reasonable candidate gene, given that C2 and other complement-related genes are associated with numerous IMDs. For rs7515029 at IL23R the protective allele is almost twice as frequent in YRI (9%) than CEU (5%) indicating another African-derived protective variant, in addition to the well-established R381Q replication.

Our strongest UC signal was in the HLA region, similar to that observed in all examined populations. The most significant SNP, rs9271366, known to tag the HLA class II allele DRB1*1502 in other populations, is the same SNP found most statistically significant in both Japanese and Korean UC GWASs (p=1e-70 and p=1e-18 respectively). Hence, our rs9271366 UC association (p=7.47e-6) is justified as replication of an established UC SNP. It also accounted for nearby associations according to conditional analysis (Supplementary Figure 4). Interestingly, it was found to be the most associated HLA SNP in AA SLE, a disease 4-fold more frequent in AAs than CAs.

The strongest IBD association was observed for rs730086, with OR=1.4 (p=2.26e-6, just below study-wide significance), at an intron of KAT2A, a histone acetyltransferase gene recently linked to repression of IFN-beta and innate antiviral immunity through inhibition of TKB2. In LD with rs730086 is rs1053004 (r²=0.44), located in the 3’utr of STAT3, which is eQTL for two genes (FDR=0.038): KRT19, a gene that produces a major type 1 keratin expressed in ileal and colonic epithelium with a broad distribution in simple and stratified epithelia; and TTC25, a gene important for celiogenesis expressed in bone marrow but not in intestine. The risk variants are much more frequent in CEU than YRI.

We observed two other independent signals for IBD and CD on chromosome 17 centered about STAT5A/STAT3 adjacent genes. These two signals are protective and African-specific (monomorphic in CEU). We found no association (p>0.1) for rs12942547, the CIS SNP located in STAT3 intron with nearly identical frequencies in CEU and YRI. Complementing these findings is significant ALD that overlaps this chromosome 17 region (and extends through IKZF3 and STAT3) with the CEU genome producing IBD risk and YRI being protective. Conditioning on the IKZF3 replicated SNP (rs11742570) and the 3 independent STAT5A/STAT3 associations eliminated this ALD (p=0.8). An important implication of our association and admixture results is that the common CEU STAT5A/STAT3 haplotype likely contributes to IBD risk in the Caucasians – a finding made possible by studying an admixed population. Sequencing studies in AAs will be important to identify potential functional variants in LD with the YRI-specific protective alleles.

We found three other regions with significant ALD. Although ALD regions should be unbiased to known loci (as Immunochip contained AIMs throughout the genome) all four regions overlap CIS loci. The most highly associated region is on chromosome 10, includes the IPMK/CISD1/UBE2D1/TFAM locus and extends to within 800 kb of ZNF365 (using a
A region with ALD associated with CD and IBD on chromosome 15 contains SMAD3, a gene associated with need for multiple surgeries in CD30. A region on chromosome 16 contains PRKCB, IL4R and IL21R, and extends centromeric just to SBK1. Each of the ALD regions contains additional potential candidate genes, as noted in Supplementary Table 3.

The top three KEGG pathways in AA IBD (cytokine, Jak-STAT and chemokines) were among the top four KEGG pathways found in CIS. However, not observed in AA IBD was the Leishmania infection pathway – second most significant pathway in the CIS; whereas AA UC, IBD, and CD all showed strong evidence for measles pathway and UC also showed significant evidence for African trypanosomiasis – both pathways not observed in CIS.

The high frequency of genes associated with infectious disease supports the hypothesis that IBD genetic susceptibility may be a by-product of evolutionary adaptation to human pathogens. Hence the differential infectious disease pathways associated with AA and Caucasian IBD may be related to adaptation to geographically distinct infectious diseases. In total, these findings suggest that, while major pathways are similar across populations, some are distinct and may be targets for personalized IBD therapies.

This first, in-depth characterization of AA IBD genetics provides a basis to compare IBD genetics with Caucasian and East Asian populations where multiple GWAS, replication and Immunochip studies have been performed (Figure 3). Five of the AA loci with suggestive/replication evidence are associated in all three populations: IL23R, FCGR2A, IL12B, HLA and STAT5A, with the functional SNPs for IL23R and FCGR2A the same. The higher OR loci tend to be observed in more than one population. The STAT5A/STAT3 locus stands out in AAs: the ORs for the three independent associations are 0.76, 1.40 and 0.55, suggesting that this region likely has a stronger influence in AA than Caucasian IBD (OR=1.1 for rs12942547 in CIS). A Japanese CD GWAS found relatively strong influence for STAT3, maximal at rs9891119, but we found no evidence in AAs for this SNP equally frequent in YRI, CEU and East Asians.

In summary, our study reveals that in AAs HLA SNPs demonstrate the dominant signal for UC; STAT5A/STAT3 shows disproportionate association with IBD and CD including novel African protective haplotypes and common European risk haplotypes that together with an IKZF3 CD replication 3 Mb centromeric likely gives risk to significant chromosome 17 regional ALD. We identified three other chromosomal regions contributing significant IBD risk from European admixture. We replicated Maximal-CIS SNPs for NOD2, IL23R, 5p13.1, and FCGR2A, identified additional risk variants at 5p13.1 and IL23R. We observed a strong correlation in ORs for Maximal-CIS SNPs between our study and the CIS suggesting that additional Caucasian established loci likely play a role in AAs. We also demonstrated other known (IL12B) and novel (C2orf43, HDAC11 and LINC00994) suggestive areas of association interest. The new suggestive associations will need to be validated by independent replication. Finally a network analysis showed that, as in other studies, cytokines, Jak-STAT signaling and chemokine pathways play major roles, but also suggests that measles and African trypanosomiasis pathways may be important for further investigation. This study has yielded vital information not only on the etiopathogenesis of
IBD specifically for AAs, but also about risk variants and ancestral chromosomal regions that may also contribute to IBD pathogenesis in Caucasians, as we continue our exploration of AA IBD genetics.

Supplementary Material
Refer to Web version on PubMed Central for supplementary material.

Authors
Chengrui Huang\textsuperscript{1}, Talin Haritunians\textsuperscript{2}, David T. Okou\textsuperscript{3}, David J. Cutler\textsuperscript{5}, Michael E. Zwick\textsuperscript{5}, Kent D. Taylor\textsuperscript{4}, Lisa W. Datta\textsuperscript{6}, Joseph C. Maranville\textsuperscript{7}, Zhenqiu Liu\textsuperscript{2}, Shannon Ellis\textsuperscript{6}, Pankaj Chopra\textsuperscript{5}, Jonathan S. Alexander\textsuperscript{8}, Robert N. Baldassano\textsuperscript{9}, Raymond K. Cross\textsuperscript{10}, Themistocles Dassopoulos\textsuperscript{11}, Tanvi A. Dhure\textsuperscript{12}, Richard H. Duerr\textsuperscript{13}, John S. Hanson\textsuperscript{14}, Jason K. Hou\textsuperscript{15}, Sunny Z. Hussain\textsuperscript{16}, Kim L. Isaacs\textsuperscript{17}, Kelly E Kachelries\textsuperscript{9}, Howard Kader\textsuperscript{18}, Michael D. Kappelman\textsuperscript{19}, Jeffrey Katz\textsuperscript{20}, Richard Kellermayer\textsuperscript{21}, Barbara S. Kirschner\textsuperscript{22}, John F. Kuemmerle\textsuperscript{23}, Archana Kumar\textsuperscript{3}, John H. Kwon\textsuperscript{24}, Mark Lazarev\textsuperscript{6}, Peter Mannon\textsuperscript{25}, Dedrick E. Moulton\textsuperscript{26}, Bankole O. Osuntokun\textsuperscript{27}, Ashish Patel\textsuperscript{28}, John D. Rioux\textsuperscript{29}, Jerome I. Rotter\textsuperscript{4}, Shehzad Saeed\textsuperscript{30}, Ellen J. Scher\textsuperscript{31}, Mark S. Silverberg\textsuperscript{32}, Ann Silverman\textsuperscript{33}, Stephan R. Targan\textsuperscript{2}, John F. Valentine\textsuperscript{34}, Ming-Hsi Wang\textsuperscript{6}, Claire L. Simpson\textsuperscript{35}, S. Louis Bridges\textsuperscript{36}, Robert P. Kimberly\textsuperscript{36}, Stephen S. Rich\textsuperscript{37}, Judy H. Cho\textsuperscript{38}, Anna Di Rienzo\textsuperscript{7}, Linda W.H. Kao\textsuperscript{1}, Dermot P.B. McGovern\textsuperscript{2}, Steven R. Brant\textsuperscript{1,6,†}, and Subra Kugathasan\textsuperscript{3}

Affiliations
1 Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21231, USA
2 F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90049, USA
3 Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA
4 Institute for Translational Genomics and Population Sciences and Division of Genomic Outcomes, Departments of Pediatrics and Medicine, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA
5 Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA
6 Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA
7 Committee on Clinical Pharmacology and Pharmacogenomics, and the Department of Human Genetics, The University of Chicago, Chicago, IL 60637, USA
8 Department of Molecular and Cellular Physiology, Louisiana State University Health Sciences Center, Shreveport, LA 71130, USA

9 Division of Gastroenterology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

10 Division of Gastroenterology, University of Maryland, Baltimore, MD 21201, USA

11 Department of Medicine, Washington University School of Medicine, St Louis, MO 63110, USA

12 Department of Medicine, Emory University School of Medicine, Atlanta, GA 30322, USA

13 Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, and Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

14 Charlotte Gastroenterology and Hepatology, PLLC, Charlotte, NC 28207, USA

15 Department of Medicine, Baylor College of Medicine; VA HSR&D Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center, Houston, TX 77030, USA

16 Department of Pediatrics, Willis-Knighton Physician Network, Shreveport, LA 71118, USA

17 Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

18 Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD 21201, USA

19 Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

20 Division of Gastroenterology, Case Western Reserve University, Cleveland, OH 44106, USA

21 Section of Pediatric Gastroenterology, Baylor College of Medicine, Houston, TX, 77030

22 Department of Pediatrics, University of Chicago Comer Children's Hospital, Chicago, IL 60637, USA

23 Departments of Medicine and Physiology and Biophysics, VCU Program in Enteric Neuromuscular Sciences, Medical College of Virginia Campus of Virginia Commonwealth University, Richmond VA 23298, USA

24 Section of Gastroenterology, Department of Medicine, University of Chicago, Chicago, IL 60637, USA

25 Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA
Acknowledgements

We are grateful to all of the patients and the controls that volunteered to join this study. This study is dedicated to Dr. Linda Kao who helped plan and direct the study and passed away just prior to the study's completion.

IBDGC sample recruitment was from the Meyerhoff Inflammatory Bowel Disease Center at The Johns Hopkins Hospital and Johns Hopkins University School of Medicine, Departments of Medicine and Surgery (Patricia Ushry, Sharon Dudley-Brown, Theodore M. Bayless, Christina Ha, Jonathon Efron; Susan Gearhart, and Michael Marohn); and the Division of Pediatric Gastroenterology and Nutrition at The Johns Hopkins Children's Center (Maria Oliva-Hemker and Carmen Cuffari); and funded satellite centers at Henry Ford Health System (Martin Zonca, Qiana Samuels and Aref Araya); University of North Carolina Department of Medicine (Dolly Walkup); Baylor College of Medicine; Virginia Commonwealth University (Kasiah Banks; Alisa Maibauer; Amy Newcombe); Washington Hospital Center (Michael S. Gold and Averell Sherker); University of Florida; Howard University (Duane T. Smoot); University of Alabama (Toni Seay; Tajuanna Lucious); Perelman School of
Medicine at the University of Pennsylvania (James D. Lewis); University of Maryland Department of Pediatrics; University of Chicago (Lici Shen); Cornell University; and Columbia University (Arun Swaminath). IBDGC Genetic Research Centers at Yale University, University of Pittsburgh, University of Toronto, and University of Montreal (John D. Rioux holds a Canada Research Chair). The database was developed with assistance from Phil Schumann (University of Chicago). Additional DNA samples on African American IBD cases were provided by Material Transfer Agreements from Washington University Inflammatory Bowel Disease Program with assistance from Rodney Newberry and Ellen Li.

Emory samples were recruited from: Emory University School of Medicine and funded satellite centers at The Children's Hospital of Philadelphia, Cincinnati Children's Hospital Medical Center, Case Western Reserve University, University of Maryland, Vanderbilt Children's Hospital, University of Texas Southwestern Medical Center, University of North Carolina Chapel Hill Department of Pediatrics, University of Chicago Comer Children's Hospital (Thomas Mangatou; Kathleen Van’t Hof, The Barnett and Alscher Families), Louisiana State University Health Sciences Center, Cook Childrens Medical Center, and Willis-Knighton Physician Network.

Cedars samples were recruited from: Cedars-Sinai Medical Center F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute (Stephan Targan, Andrew Ippoliti, Eric Vasiliasiuskas, David Shih, Gil Melmed, Marla Dubinsky), Charlotte Gastroenterology and Hepatology PLLC; and the Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center Institute for Translational Genomics and Population Sciences.

We thank Suna Onengut-Gumuscu for assistance with extraction of control data from RA, SLE and T1D Immunochip studies. We thank Feng Zhou for assistance with Figure 3 and compiling IBD association information in East Asian populations.

**Grant Support:** NIH Grants DK062431 (S.R.B.), DK087694 (S.K.), DK062413 (D.P.B.M and K.T), DK046763-19, AI067068 and U54DE023789-01 (D.P.B.M.), DK062429 and DK062422 (J.H.C.), DK062420 (R.H.D.), DK062432 (J.D.R.), and DK062423 (M.S.S.). Additional support from Harvey M. and Lynn P. Meyerhoff Inflammatory Bowel Disease Center, the Morton Hyatt Family, the Buford and Linda Lewis family (S.R.B.); Endowed professorship from Marcus foundation (SK); The Joshua L and Lisa Z Greer Endowed Chair, HS021747 from the Agency for Healthcare Research and Quality, grant 305479 from the European Union, and The Leona M. and Harry B. Helmsley Charitable Trust (D.P.B.M.); from the Veterans Administration HSR&D Center for Innovations in Quality, Effectiveness and Safety (#CIN 13-413) and the Michael E. DeBakey VA Medical Center (J.K.H.). PARC control samples supported by NIH/NHLBI grant HL06957 (Ronald M. Krauss, PI). RA control samples recruited and supported by the Consortium for the Longitudinal Evaluation of African-Americans with Early Rheumatoid Arthritis (CLEAR), NIH grants N01-AR-02247 and AR-6-2278 (S.L.B.) and the University of Alabama GCRC (M01-RR-00032); SLE control samples recruited and supported by the PROFILE Study group coordinated at the University of Alabama Birmingham and supported by NIH grants P01-AR49084 (R.P.K.) and M01-RR-00032; T1D control samples recruited and supported by the Type 1 Diabetes Genetics Consortium U01 DK062418 (S.S.R.), and sponsored by NIDDK, NIAID, NHGRI, NICHD and the Juvenile Diabetes Research Foundation (JDRF).

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALD</td>
<td>admixture linkage disequilibrium</td>
</tr>
<tr>
<td>AAs</td>
<td>African Americans</td>
</tr>
<tr>
<td>AIMs</td>
<td>ancestry informative markers</td>
</tr>
<tr>
<td>CAs</td>
<td>Caucasian Americans</td>
</tr>
<tr>
<td>CIS</td>
<td>Caucasian Immunochip study</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn's disease</td>
</tr>
<tr>
<td>CEU</td>
<td>European ancestry</td>
</tr>
<tr>
<td>eQTL</td>
<td>expression quantitative trait locus</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>GRC</td>
<td>genetics research center</td>
</tr>
<tr>
<td>GWA</td>
<td>genome wide association</td>
</tr>
</tbody>
</table>
IMD immune-mediated diseases
IBD inflammatory bowel disease
IBDU inflammatory bowel disease type undetermined
OR odds ratio
PBMCs peripheral blood mononuclear cells
QC quality control
RA rheumatoid arthritis
Maximal-CIS SNPs the most statistically significant SNPs at each locus in the Caucasian Immunochip study and present on the Immunochip genotyping array
SLE systemic lupus erythematosus
T1D type 1 diabetes
UC ulcerative colitis
YRI West African ancestry

References


Figure 1.
Comparison of $\log_2$(Odds ratios) estimated in this study with those in Caucasian Immunochip Study (CIS) for 110 IBD-general (correlation=0.78), 26 CD-specific (correlation=0.77) and 23 UC-specific (correlation=0.61) Maximal-CIS SNPs.
Figure 2.
Association mapping (left) and admixture (right) mapping p-values (in $-\log_{10}$) for IBD (top panel), CD (middle panel) and UC (bottom panel).
Figure 3.
Comparison of log_{10}(Odds Ratios) at the risk alleles by chromosomal position (in Mb, Genome Build 37) between Caucasian (black), East Asian (green) and African American (red) loci for IBD (hatched), CD (solid) and UC (stippled). Bars with black outlines represent loci with reported genome-wide significant association evidence in the CIS (black, limited to OR $\geq$1.1) or in East Asian studies (green). Bars without outlines are present for AA or East Asian replications (at p<3.1e-4) of CIS-Maximal SNPs, or for SNPs evaluated in these populations with suggestive evidence for association at established CIS loci (p<4.2e-5) or borderline of genome-wide significance for novel loci (p $\leq$1e-6). Asterisks denote the same associated SNPs in AA and additional populations. Red arrows denote overlaps between significant admixture regions and association loci.
Table 1

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Mb</th>
<th>Nearest Gene</th>
<th>A1</th>
<th>A2</th>
<th>YRI</th>
<th>CEU</th>
<th>CIS</th>
<th>Sig</th>
<th>p</th>
<th>OR</th>
<th>1.95CI</th>
<th>1.95CI</th>
<th>CIS</th>
<th>CIS</th>
<th>CIS</th>
<th>CIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11209026</td>
<td>1p31.3</td>
<td>67.71</td>
<td>IL23R</td>
<td>A</td>
<td>G</td>
<td>0.009</td>
<td>0.041</td>
<td>0.007</td>
<td>0.014</td>
<td>0.067</td>
<td>7.53E-05</td>
<td>0.219</td>
<td>0.108</td>
<td>0.464</td>
<td>IBD</td>
<td>0.497</td>
<td>0.465</td>
</tr>
<tr>
<td>rs1801274</td>
<td>1q23</td>
<td>161.48</td>
<td>FCGR2A</td>
<td>A</td>
<td>G</td>
<td>0.496</td>
<td>0.491</td>
<td>0.506</td>
<td>0.453</td>
<td>0.525</td>
<td>2.60E-04</td>
<td>1.211</td>
<td>1.098</td>
<td>1.343</td>
<td>UC</td>
<td>1.192</td>
<td>1.15</td>
</tr>
<tr>
<td>rs11742570</td>
<td>5p13.1</td>
<td>40.41</td>
<td>PTGER4</td>
<td>A</td>
<td>G</td>
<td>0.385</td>
<td>0.354</td>
<td>0.342</td>
<td>0.390</td>
<td>0.39</td>
<td>1.29E-04</td>
<td>0.812</td>
<td>0.729</td>
<td>0.903</td>
<td>CD</td>
<td>0.773</td>
<td>0.747</td>
</tr>
<tr>
<td>rs5743289</td>
<td>16q21</td>
<td>50.76</td>
<td>NOD2</td>
<td>A</td>
<td>G</td>
<td>0</td>
<td>0.226</td>
<td>0.052</td>
<td>0.032</td>
<td>0.178</td>
<td>6.12E-05</td>
<td>1.363</td>
<td>1.336</td>
<td>2.326</td>
<td>CD</td>
<td>1.557</td>
<td>1.497</td>
</tr>
<tr>
<td>rs1296510</td>
<td>17q21</td>
<td>37.91</td>
<td>IKZF3</td>
<td>A</td>
<td>G</td>
<td>0.119</td>
<td>0.300</td>
<td>0.233</td>
<td>0.201</td>
<td>0.467</td>
<td>9.31E-05</td>
<td>1.319</td>
<td>1.148</td>
<td>1.536</td>
<td>IBD</td>
<td>1.157</td>
<td>1.124</td>
</tr>
</tbody>
</table>

*IBD-general SNP in CIS;*  
*Crohn’s disease-specific SNP in CIS;*  
*SNP position in mega base-pair (Mb, Genome Build 37);*  
*Minor (A1) and major (A2) allele in current dataset respectively;*  
*Frequency corresponding to minor allele (A1) for Hapmap African samples (YRI), Hapmap European samples (CEU), current African American IBD case (Case) and control (Ctrl) samples, and Immunochip case/control samples used in CIS respectively;*  
*Most significant phenotype (Sig), corresponding p-value (p), minor allele odds ratio (OR) with upper (U95CI) and lower (L95CI) 95% confidence interval. *  
*The most significant phenotype (Sig) and the corresponding odds ratio (OR), lower (L95CI) and upper (U95CI) bounds for 95% confidence interval for odds ratio in CIS;*  
*SNP without expression data available in eQTL evaluation.*
<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Mb</th>
<th>A1</th>
<th>A2</th>
<th>YRI</th>
<th>CEU</th>
<th>Case</th>
<th>Ctrl</th>
<th>Nearest Gene</th>
<th>Location</th>
<th>Sig</th>
<th>p</th>
<th>OR</th>
<th>L95CI</th>
<th>U95CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7515020G</td>
<td>1</td>
<td>67.60</td>
<td>G</td>
<td>A</td>
<td>0.093</td>
<td>0.049</td>
<td>0.075</td>
<td>0.101</td>
<td>IL2R</td>
<td>intergenic</td>
<td>IBD</td>
<td>3.17E-05</td>
<td>0.670</td>
<td>0.555</td>
<td>0.809</td>
</tr>
<tr>
<td>rs3072^G</td>
<td>2</td>
<td>20.88</td>
<td>G</td>
<td>A</td>
<td>0.027</td>
<td>0.354</td>
<td>0.100</td>
<td>0.130</td>
<td>C2orf53</td>
<td>intergenic</td>
<td>IBD</td>
<td>2.15E-05</td>
<td>0.697</td>
<td>0.590</td>
<td>0.823</td>
</tr>
<tr>
<td>rs10043340G</td>
<td>5</td>
<td>40.49</td>
<td>C</td>
<td>A</td>
<td>0.466</td>
<td>0.275</td>
<td>0.361</td>
<td>0.434</td>
<td>PTGER4</td>
<td>intergenic</td>
<td>CD^d</td>
<td>3.25E-06</td>
<td>0.754</td>
<td>0.669</td>
<td>0.849</td>
</tr>
<tr>
<td>rs4286721</td>
<td>5</td>
<td>40.50</td>
<td>G</td>
<td>A</td>
<td>0.385</td>
<td>0.252</td>
<td>0.319</td>
<td>0.386</td>
<td>PTGER4</td>
<td>intergenic</td>
<td>CD^d</td>
<td>2.97E-06</td>
<td>0.748</td>
<td>0.662</td>
<td>0.845</td>
</tr>
<tr>
<td>rs6866402</td>
<td>5</td>
<td>40.52</td>
<td>A</td>
<td>G</td>
<td>0.442</td>
<td>0.257</td>
<td>0.368</td>
<td>0.428</td>
<td>PTGER4</td>
<td>intergenic</td>
<td>CD^d</td>
<td>4.05E-05</td>
<td>0.782</td>
<td>0.695</td>
<td>0.879</td>
</tr>
<tr>
<td>rs1876141</td>
<td>5</td>
<td>40.53</td>
<td>G</td>
<td>A</td>
<td>0.442</td>
<td>0.267</td>
<td>0.367</td>
<td>0.428</td>
<td>PTGER4</td>
<td>intergenic</td>
<td>CD^d</td>
<td>4.00E-05</td>
<td>0.782</td>
<td>0.695</td>
<td>0.879</td>
</tr>
<tr>
<td>rs1505994</td>
<td>5</td>
<td>40.53</td>
<td>A</td>
<td>G</td>
<td>0.442</td>
<td>0.267</td>
<td>0.368</td>
<td>0.429</td>
<td>PTGER4</td>
<td>intergenic</td>
<td>CD^d</td>
<td>3.63E-05</td>
<td>0.781</td>
<td>0.694</td>
<td>0.878</td>
</tr>
<tr>
<td>rs35990859G</td>
<td>5</td>
<td>158.80</td>
<td>G</td>
<td>A</td>
<td>-</td>
<td>0.092^e</td>
<td>0.014</td>
<td>0.027</td>
<td>IL12B</td>
<td>intergenic</td>
<td>IBD^d</td>
<td>6.62E-06</td>
<td>0.393</td>
<td>0.261</td>
<td>0.590</td>
</tr>
<tr>
<td>rs36048684G</td>
<td>5</td>
<td>158.82</td>
<td>A</td>
<td>T</td>
<td>-</td>
<td>0.505^e</td>
<td>0.013</td>
<td>0.027</td>
<td>IL12B</td>
<td>intergenic</td>
<td>IBD^d</td>
<td>9.04E-06</td>
<td>0.398</td>
<td>0.265</td>
<td>0.598</td>
</tr>
<tr>
<td>rs730086</td>
<td>17</td>
<td>40.27</td>
<td>G</td>
<td>A</td>
<td>0.097</td>
<td>0.664</td>
<td>0.238</td>
<td>0.177</td>
<td>KAT2A</td>
<td>intron</td>
<td>IBD^d</td>
<td>2.26E-06</td>
<td>1.373</td>
<td>1.204</td>
<td>1.567</td>
</tr>
<tr>
<td>rs7220367</td>
<td>17</td>
<td>40.44</td>
<td>G</td>
<td>C</td>
<td>0.392</td>
<td>0</td>
<td>0.285</td>
<td>0.349</td>
<td>STAT5A</td>
<td>intergenic</td>
<td>CD</td>
<td>2.23E-05</td>
<td>0.758</td>
<td>0.667</td>
<td>0.862</td>
</tr>
<tr>
<td>rs7217884G</td>
<td>17</td>
<td>40.45</td>
<td>G</td>
<td>A</td>
<td>0.270</td>
<td>0</td>
<td>0.211</td>
<td>0.267</td>
<td>STAT5A</td>
<td>intron</td>
<td>IBD^d</td>
<td>2.21E-05</td>
<td>0.765</td>
<td>0.675</td>
<td>0.866</td>
</tr>
<tr>
<td>rs13380828G</td>
<td>17</td>
<td>40.45</td>
<td>A</td>
<td>T</td>
<td>0.297</td>
<td>0</td>
<td>0.210</td>
<td>0.266</td>
<td>STAT5A</td>
<td>intron</td>
<td>IBD</td>
<td>2.57E-05</td>
<td>0.766</td>
<td>0.677</td>
<td>0.867</td>
</tr>
<tr>
<td>rs1053004</td>
<td>17</td>
<td>40.47</td>
<td>A</td>
<td>G</td>
<td>0.110^f</td>
<td>0.611</td>
<td>0.169</td>
<td>0.118</td>
<td>STAT3</td>
<td>utr-3</td>
<td>IBD</td>
<td>1.08E-05</td>
<td>1.409</td>
<td>1.209</td>
<td>1.641</td>
</tr>
<tr>
<td>rs72243309G</td>
<td>17</td>
<td>40.49</td>
<td>A</td>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0.033</td>
<td>0.058</td>
<td>STAT3</td>
<td>intron</td>
<td>IBD</td>
<td>8.76E-06</td>
<td>0.552</td>
<td>0.425</td>
<td>0.717</td>
</tr>
<tr>
<td>rs2665211</td>
<td>3</td>
<td>13.49</td>
<td>G</td>
<td>A</td>
<td>0.179</td>
<td>0.431</td>
<td>0.190</td>
<td>0.265</td>
<td>HDAC11</td>
<td>intergenic</td>
<td>UC</td>
<td>6.97E-06</td>
<td>0.619</td>
<td>0.503</td>
<td>0.763</td>
</tr>
<tr>
<td>rs254855</td>
<td>3</td>
<td>64.07</td>
<td>A</td>
<td>G</td>
<td>0.261</td>
<td>0.301</td>
<td>0.287</td>
<td>0.214</td>
<td>LNCO0994</td>
<td>coding</td>
<td>UC</td>
<td>3.13E-05</td>
<td>1.491</td>
<td>1.235</td>
<td>1.799</td>
</tr>
<tr>
<td>rs2395178G</td>
<td>6</td>
<td>32.41</td>
<td>C</td>
<td>G</td>
<td>0.533</td>
<td>0.602</td>
<td>0.415</td>
<td>0.500</td>
<td>HLA-DRA</td>
<td>intergenic</td>
<td>UC</td>
<td>1.49E-05</td>
<td>1.692</td>
<td>0.586</td>
<td>0.818</td>
</tr>
<tr>
<td>rs9270998G</td>
<td>6</td>
<td>32.57</td>
<td>A</td>
<td>C</td>
<td>0.173</td>
<td>0.204</td>
<td>0.218</td>
<td>0.156</td>
<td>HLA-DRB1</td>
<td>intergenic</td>
<td>UC</td>
<td>1.85E-05</td>
<td>1.559</td>
<td>1.272</td>
<td>1.91</td>
</tr>
<tr>
<td>rs9271366G</td>
<td>6</td>
<td>32.59</td>
<td>G</td>
<td>A</td>
<td>0.168</td>
<td>0.192</td>
<td>0.202</td>
<td>0.141</td>
<td>HLA-DQA1</td>
<td>intergenic</td>
<td>UC</td>
<td>7.47E-06</td>
<td>1.627</td>
<td>1.315</td>
<td>2.013</td>
</tr>
<tr>
<td>rs2097431G</td>
<td>6</td>
<td>32.59</td>
<td>A</td>
<td>G</td>
<td>0.332</td>
<td>0.394</td>
<td>0.445</td>
<td>0.364</td>
<td>HLA-DQA1</td>
<td>intergenic</td>
<td>UC</td>
<td>3.34E-05</td>
<td>1.436</td>
<td>1.210</td>
<td>1.703</td>
</tr>
</tbody>
</table>

^aSNP position in mega base-pair (Mb, Genome Build 37);
b Frequency corresponding to minor allele (A1) for Hapmap African (YRI), Hapmap European (CEU), and current African American case (Case) and control (Ctrl) samples respectively;

c The most significant phenotype (Sig) and corresponding p-values (p), odds ratio (OR) with lower (L95CI) and upper (U95CI) bounds for 95% confidence intervals;

d SNP with suggestive association (p<4.2e-5) for both IBD and CD;

e Minor allele frequency (MAF) for samples from 1000 Genomes pilot 1 CEU low coverage panel;

f MAF for samples from the human variation panel of African Americans;

g SNP without expression data available in eQTL evaluation.
Table 3

Significant (p<3.9e-4) admixture association regions

<table>
<thead>
<tr>
<th>Maximal SNP</th>
<th>Chromosomal^a</th>
<th>Position^a</th>
<th>Phenotype^b</th>
<th>p^b</th>
<th>OR^b</th>
<th>L95CI^b</th>
<th>U95CI^b</th>
<th>Candidate Genes^d</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1684909</td>
<td>10q11.23-10q21.2</td>
<td>49.94-63.25</td>
<td>IBD^c</td>
<td>1.47E-05</td>
<td>0.542</td>
<td>0.472</td>
<td>0.715</td>
<td>ZNF365, REEP3, MBL2</td>
</tr>
<tr>
<td>rs603439</td>
<td>15q22.2-15q23</td>
<td>62.76-70.75</td>
<td>CD^c</td>
<td>1.80E-04</td>
<td>0.549</td>
<td>0.484</td>
<td>0.751</td>
<td>TRIP4, SMAD6, SMAD3, MAP2K1, IGDDC4, IGDDC3</td>
</tr>
<tr>
<td>rs1423086</td>
<td>16p12.2-16p12.1</td>
<td>21.74-28.39</td>
<td>IBD</td>
<td>2.67E-04</td>
<td>0.586</td>
<td>0.465</td>
<td>0.781</td>
<td>IL-4, IL-21R, SLCA11, PRKCB, ITGF1</td>
</tr>
<tr>
<td>rs2227322</td>
<td>17q12-17q21.31</td>
<td>37.39-40.93</td>
<td>IBD^c</td>
<td>1.10E-04</td>
<td>0.580</td>
<td>0.519</td>
<td>0.765</td>
<td>ORM13L3, CSF3, CCR7, Keratin cluster</td>
</tr>
</tbody>
</table>

^a Window containing SNPs with suggestive admixture (p<7.8e-4) by chromosomal banding and physical positions in Mb (Genome Build 37).

^b The most significant phenotype (Phenotype) and the corresponding: p-value (p); odds ratio with each additional African ancestry (YRI) allele (OR); lower (L95CI) and upper (U95CI) bounds for 95% confidence interval for YRI odds ratio.

^c SNP with significant admixture (p<3.9e-4) for both IBD and CD.

^d Candidate genes highlighted based on literature review of genes located in physical window.
### Table 4

Significant KEGG pathways for IBD, CD and UC

<table>
<thead>
<tr>
<th>Top KEGG Pathways</th>
<th>Phenotype</th>
<th>Genes $^b$</th>
<th>p-value $^c$</th>
<th>FDR $^c$</th>
<th>Bonferroni $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytokine-cytokine receptor interaction</td>
<td>IBD</td>
<td>12</td>
<td>1.55E-12</td>
<td>3.66E-10</td>
<td>3.66E-10</td>
</tr>
<tr>
<td>Jak-STAT signaling pathway</td>
<td>IBD</td>
<td>8</td>
<td>4.07E-09</td>
<td>4.82E-07</td>
<td>9.64E-07</td>
</tr>
<tr>
<td>Chemokine signaling pathway</td>
<td>IBD</td>
<td>8</td>
<td>1.59E-08</td>
<td>1.26E-06</td>
<td>3.77E-06</td>
</tr>
<tr>
<td>Measles</td>
<td>IBD</td>
<td>7</td>
<td>3.32E-08</td>
<td>1.97E-06</td>
<td>7.88E-06</td>
</tr>
<tr>
<td>Pathways in cancer</td>
<td>IBD</td>
<td>9</td>
<td>8.65E-08</td>
<td>4.10E-06</td>
<td>2.05E-05</td>
</tr>
<tr>
<td>T cell receptor signaling pathway</td>
<td>IBD</td>
<td>5</td>
<td>6.33E-06</td>
<td>2.50E-04</td>
<td>1.50E-03</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>IBD</td>
<td>6</td>
<td>9.10E-06</td>
<td>3.08E-04</td>
<td>2.16E-03</td>
</tr>
<tr>
<td>Gap junction</td>
<td>IBD</td>
<td>4</td>
<td>5.97E-05</td>
<td>1.77E-03</td>
<td>1.42E-02</td>
</tr>
<tr>
<td>Fc gamma R-mediated phagocytosis</td>
<td>IBD</td>
<td>4</td>
<td>7.47E-05</td>
<td>1.97E-03</td>
<td>1.77E-02</td>
</tr>
<tr>
<td>Vascular smooth muscle contraction</td>
<td>IBD</td>
<td>4</td>
<td>1.74E-04</td>
<td>4.13E-03</td>
<td>4.13E-02</td>
</tr>
<tr>
<td>Cytokine-cytokine receptor interaction</td>
<td>CD</td>
<td>18</td>
<td>3.11E-21</td>
<td>7.37E-19</td>
<td>7.37E-19</td>
</tr>
<tr>
<td>Chemokine signaling pathway</td>
<td>CD</td>
<td>10</td>
<td>4.13E-11</td>
<td>4.89E-09</td>
<td>9.79E-09</td>
</tr>
<tr>
<td>Jak-STAT signaling pathway</td>
<td>CD</td>
<td>9</td>
<td>2.29E-10</td>
<td>1.81E-08</td>
<td>5.42E-08</td>
</tr>
<tr>
<td>Measles</td>
<td>CD</td>
<td>6</td>
<td>1.15E-06</td>
<td>6.84E-05</td>
<td>2.74E-04</td>
</tr>
<tr>
<td>Pathways in cancer</td>
<td>CD</td>
<td>8</td>
<td>1.76E-06</td>
<td>8.32E-05</td>
<td>4.16E-04</td>
</tr>
<tr>
<td>Measles</td>
<td>UC</td>
<td>7</td>
<td>1.97E-09</td>
<td>4.67E-07</td>
<td>4.67E-07</td>
</tr>
<tr>
<td>Cytokine-cytokine receptor interaction</td>
<td>UC</td>
<td>8</td>
<td>1.03E-08</td>
<td>1.22E-06</td>
<td>2.43E-06</td>
</tr>
<tr>
<td>Jak-STAT signaling pathway</td>
<td>UC</td>
<td>6</td>
<td>2.05E-07</td>
<td>1.62E-05</td>
<td>4.87E-05</td>
</tr>
<tr>
<td>T cell receptor signaling pathway</td>
<td>UC</td>
<td>5</td>
<td>8.88E-07</td>
<td>5.26E-05</td>
<td>2.11E-04</td>
</tr>
<tr>
<td>African trypanosomiasis</td>
<td>UC</td>
<td>3</td>
<td>2.33E-05</td>
<td>1.11E-03</td>
<td>5.53E-03</td>
</tr>
<tr>
<td>NOD-like receptor signaling pathway</td>
<td>UC</td>
<td>3</td>
<td>1.15E-04</td>
<td>4.56E-03</td>
<td>2.74E-02</td>
</tr>
<tr>
<td>Adipocytokine signaling pathway</td>
<td>UC</td>
<td>3</td>
<td>1.88E-04</td>
<td>6.38E-03</td>
<td>4.47E-02</td>
</tr>
</tbody>
</table>

$^a$ Associated phenotype;

$^b$ Number of genes in the pathway;

$^c$ Nominal p-value, false discovery rate (FDR), and Bonferroni-corrected p-value respectively.