Supplementary Information

Description of African American control samples

All controls were self-identified as African American or black race, required to be able to sign informed consent and able to provide a blood sample.

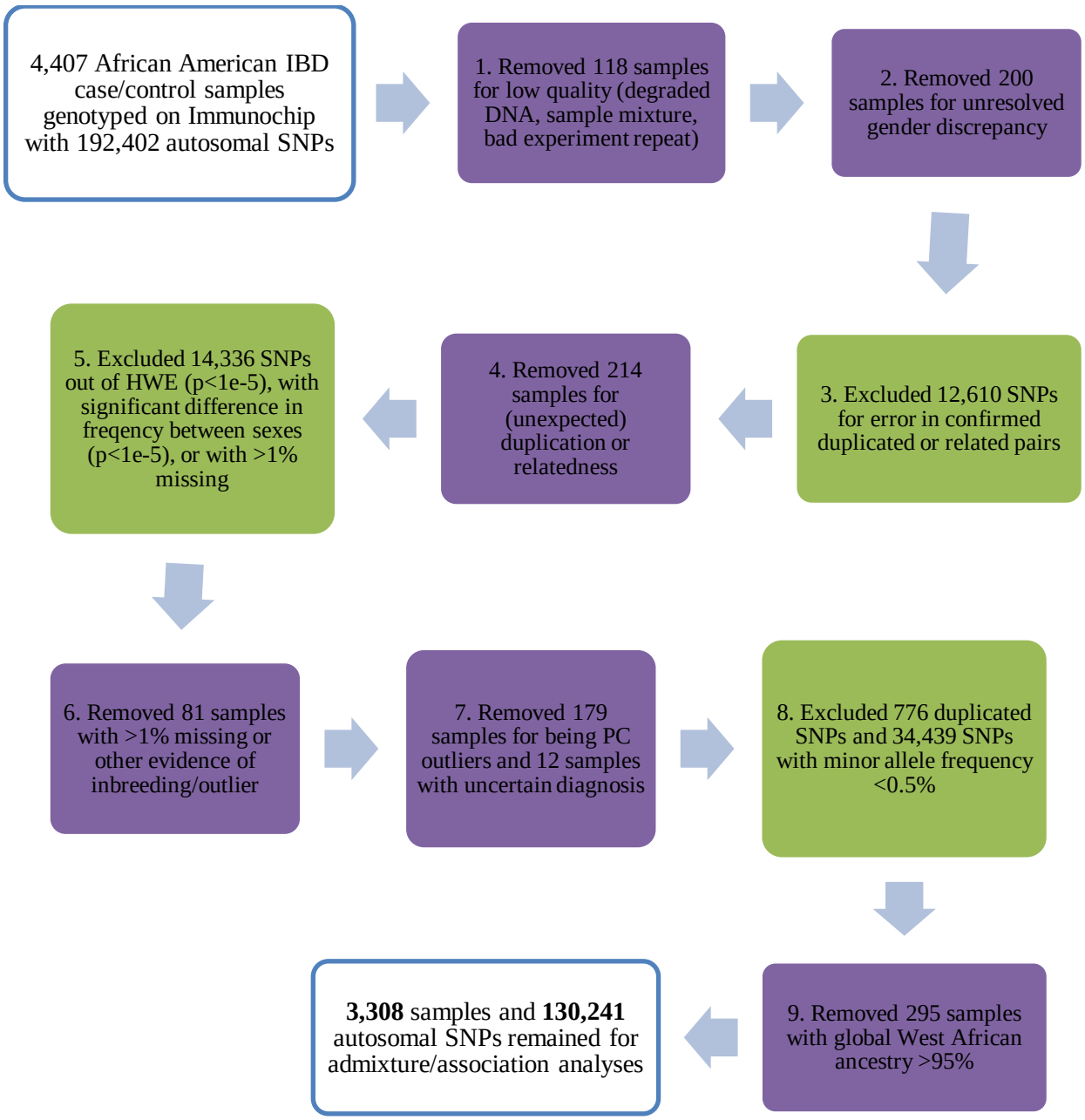
A. Controls from Multicenter African American IBD Study (MAAIS) coordinated by Johns Hopkins University IBD Genetics Research Center (GRC) of the NIDDK IBD Genetics Consortium (IBDGC) were excluded for personal and family history of IBD;
B. Controls from the rheumatoid arthritis (RA) and the systemic lupus erythematosus (SLE) Immunochip studies were screened for connective tissue disease using the Connective tissue disease Screening Questionnaire (CSQ)[1] which focuses on diseases such as Sjogren's, scleroderma, lupus, polymyositis, and gout;
C. Controls from the type 1 diabetes (T1D) Immunochip study were collected as part of the New York Cancer project[2,3];
D. Controls from the GENESIS African American cohort coordinated by Emory University were selected for: (1) no known diagnosis of IBD and (2) no known diagnosis of Autoimmune Disease;
E. Controls from the Pharmacogenetics and Risk of Cardiovascular disease Study Group (PARC) were excluded for: (1) the use of lipid-lowering medication; (2) a recent or planned change in dietary intake or a weight change of > 4.5 kg; (3) the use of corticosteroids, immunosuppressive drugs, or drugs affecting the CYP3A4 system; (4) known liver disease or elevated transaminase levels more than twice the upper limit of normal; (5) elevated creatine phosphokinase levels > 10 times the upper limits of normal; (6) uncontrolled hypertriglyceridemia, blood pressure, or diabetes mellitus; (7) abnormal renal or thyroid function; (8) current alcohol or drug abuse; (9) recent major illness in the preceding 3 months; (10) current pregnancy; and (11) known intolerance to statins.

Given the prevalence of IBD in the US of 348 per 100,000[4], the number of controls who may be affected with IBD would be small, and this is unlikely to have any significant impact on power.

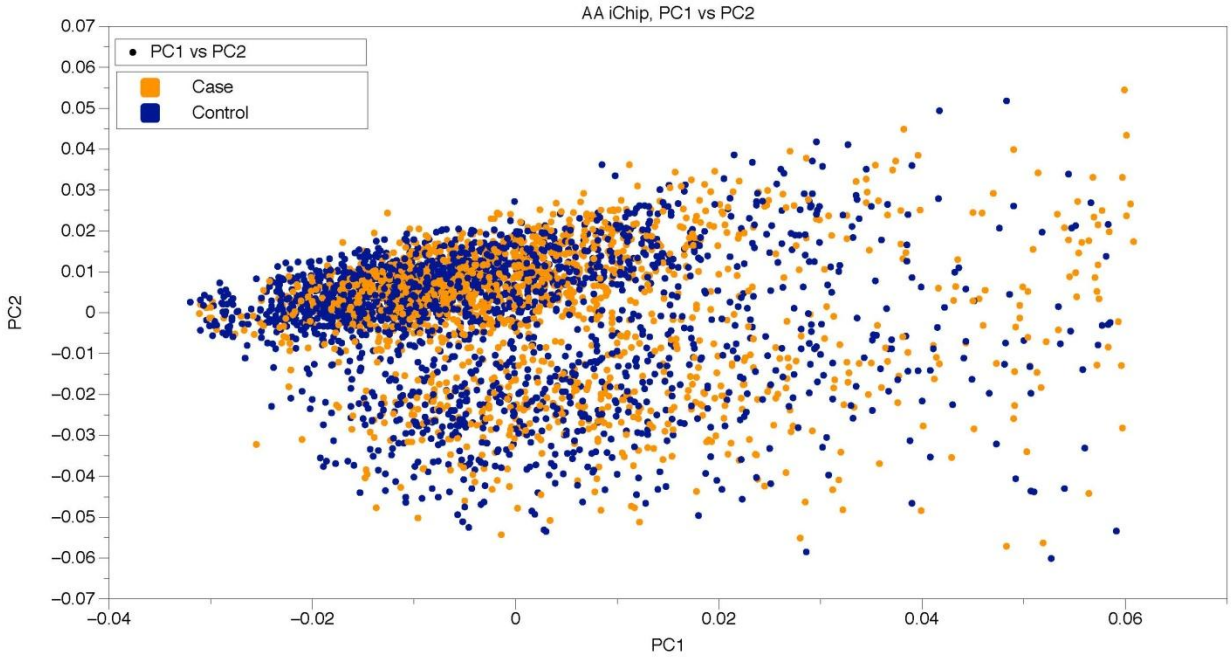
Network Analysis

After we have the lists of CD and UC associated genes, we performed gene enrichment analysis with multiple national biologically functional databases including Reactome, PPI database, and NCI/Nature Pathway Interaction Database. The networks for UC and CD associated genes were constructed from the known interactions in one of these databases. To make sure the interactions are not random, a random background network that preserves the degree distribution of the associated genes in a given list was constructed. This technique is entitled Graph with Given Degree Sequence (RGGDS) similar to Franceschini et al. (2013)[5]. The distribution of the number of interactions (edges) on our network was then compared with that on the random network for interaction enrichment analysis. A small probability ($P < 1e -4$) with Poisson distribution was achieved for both UC and CD networks corresponding to a strong interaction enrichment.

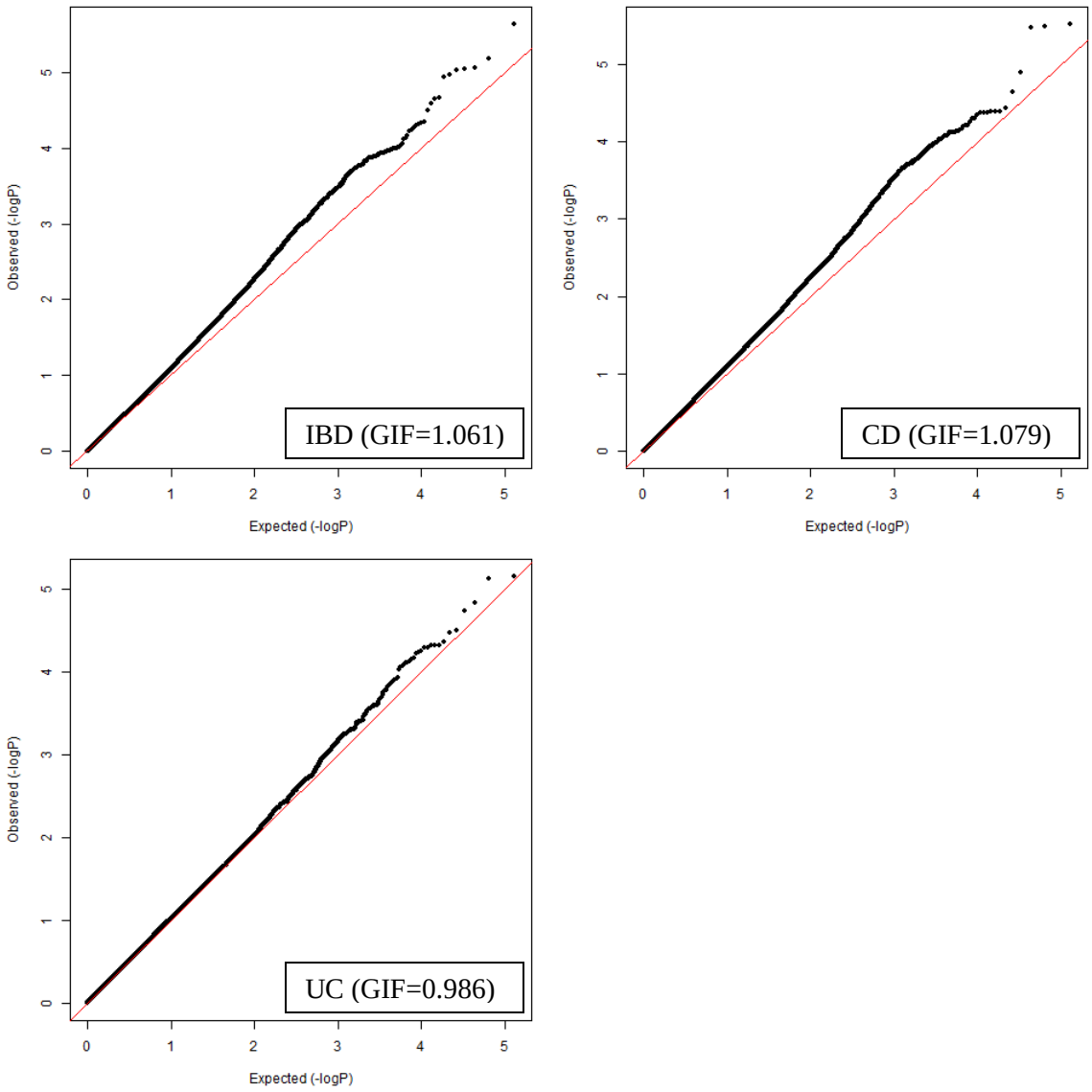
Pathway enrichment analysis for both UC and CD associated genes was performed with STRING (http://string-db.org/). Hypergeometric distribution was used to test if there is an association between a gene list and a pathway. The null hypothesis is that a gene on a pathway and a gene list are independent, or equivalently the association between a gene list and a pathway is likely due to random chance alone. A small P-value indicates there is a statistically significant association between a gene list and a pathway. Since multiple pathways were involved for a given gene list, multiple hypothesis corrections were used in STRING to reduce Type I error. Both Bonferroni and False Discovery Rate were implemented. Bonferroni is the most conservative one and adjusts the significant P-value for the test by the number of pathways investigated, while the False Discovery Rate (FDR) adjusts the significant P-value based on the rank of the predicted level of significance. The resulted P-values for pathway enrichment analysis were reported in the main manuscript.



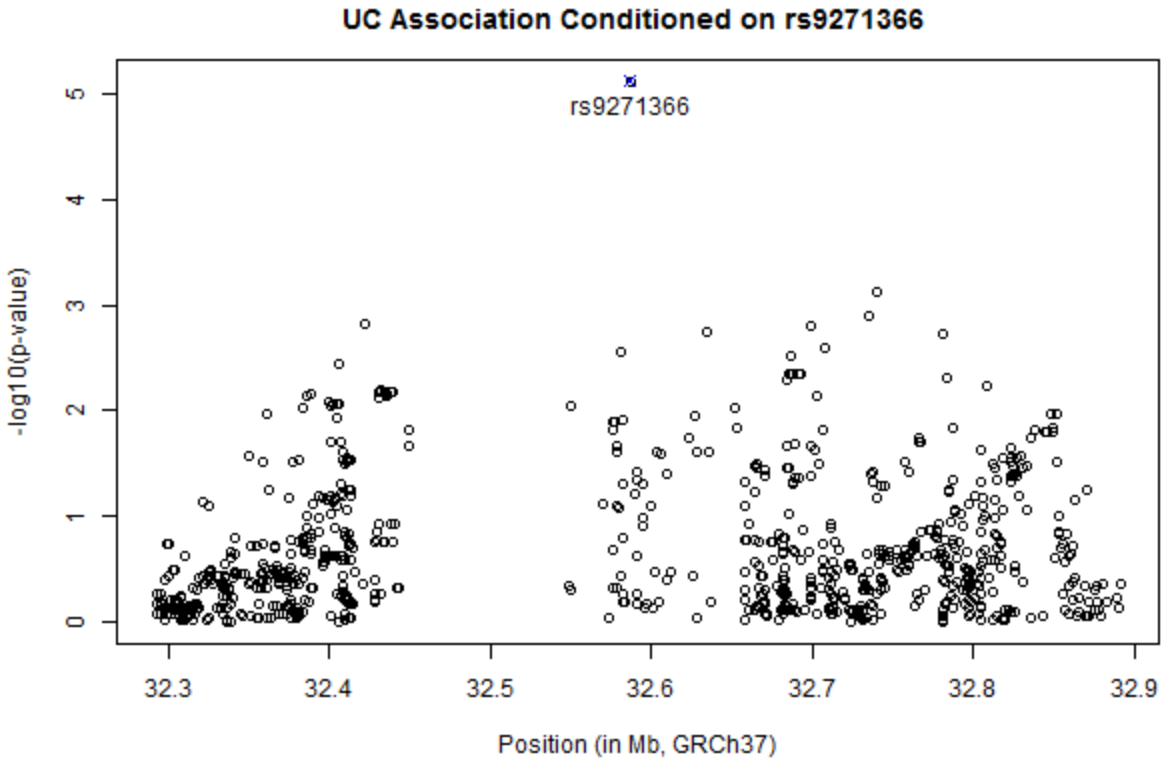
Supplementary Figure 1 Flow chart of SNP-wise and sample-wise quality control.



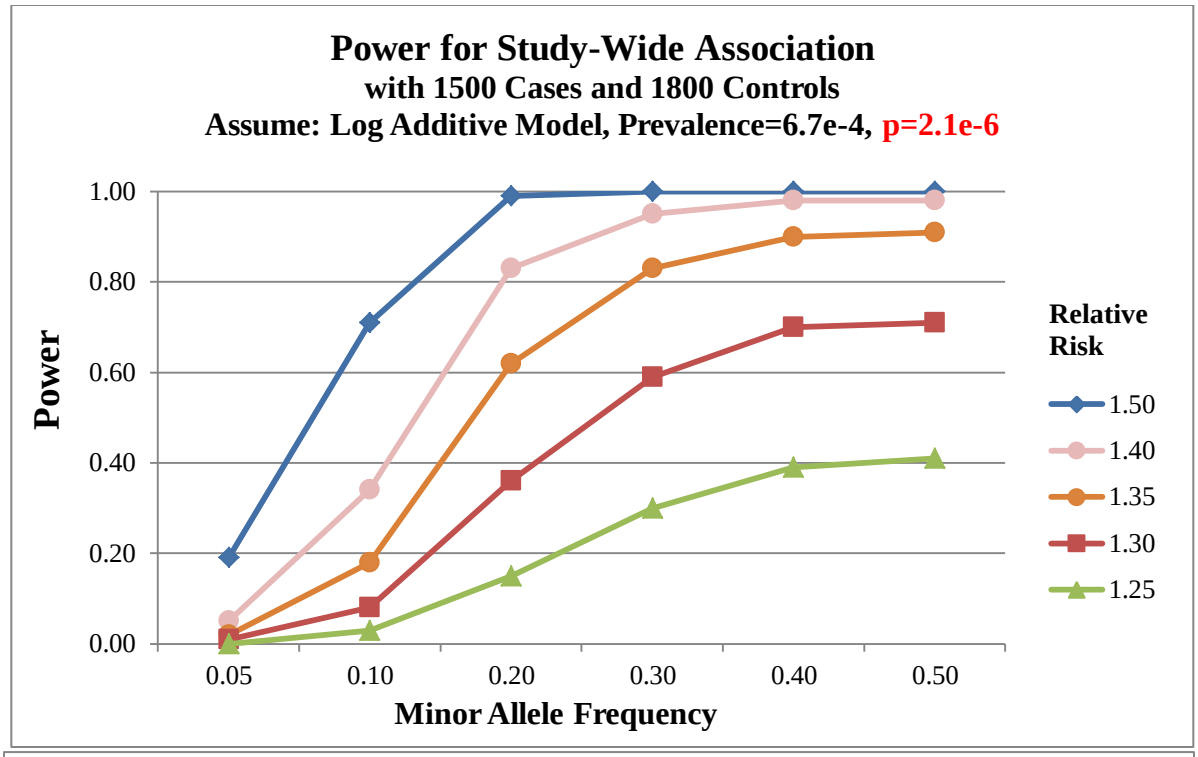
Supplementary Figure 2 Plot of the first two principle components (PCs).

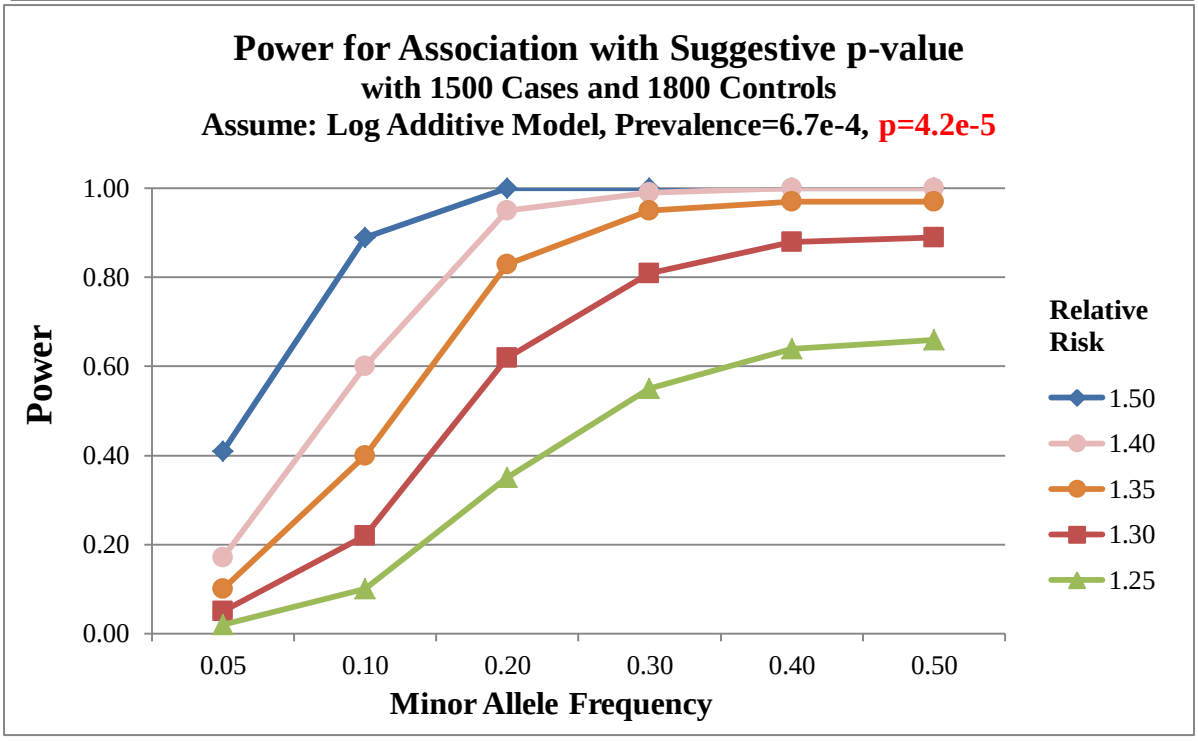


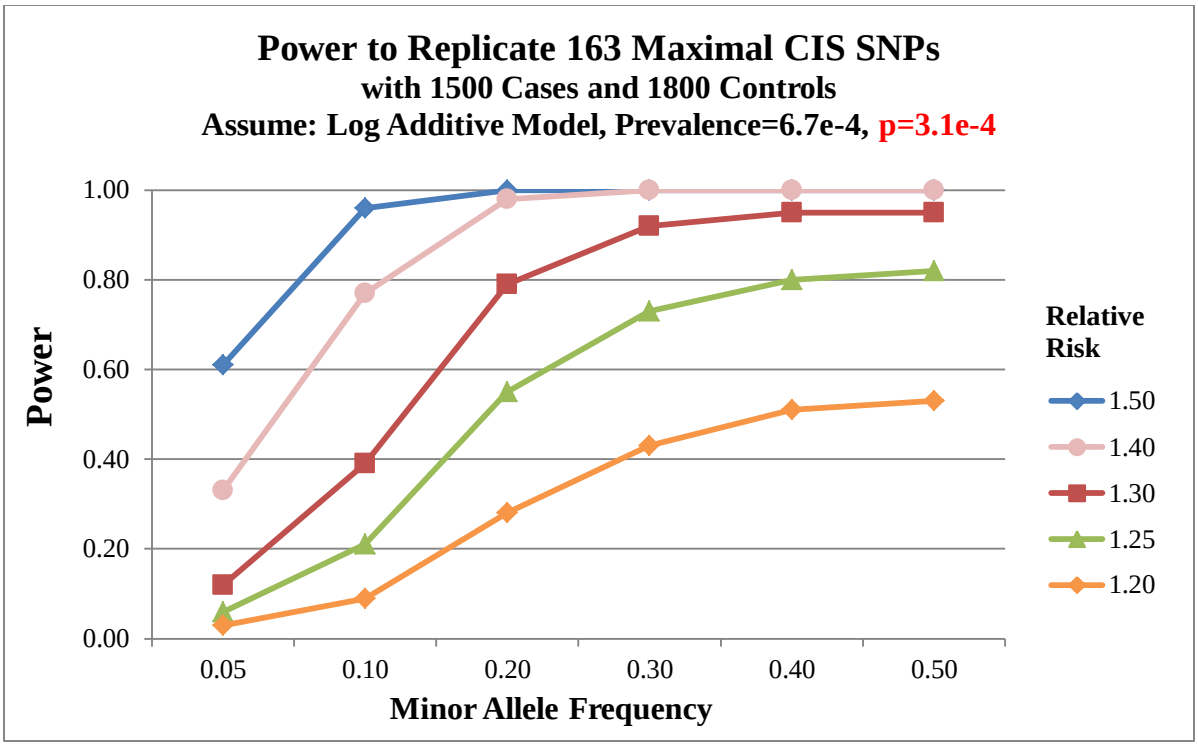
Supplementary Figure 3 Quantile-Quantile (QQ) plots for IBD (genomic inflation factor or GIF=1.061), CD (GIF=1.079) and UC (GIF=0.986) associations after adjusting for sex, recruitment coordinating center, genetic research center and global West African (YRI) ancestry.



Supplementary Figure 4 Association using 361 ulcerative colitis (UC) cases and 1797 controls for SNPs in *HLA* area conditioned on rs9271366.


Power for Study-Wide Association
with 1500 Cases and 1800 Controls
Assume: Log Additive Model, Prevalence=6.7e-4, p=2.1e-6
Power
1.00
0.80
0.60
0.40
0.20
0.00
0.05
0.10
0.20
0.30
0.40
0.50
Minor Allele Frequency
Relative Risk
1.50
1.40
1.35
1.30
1.25


Power for Association with Suggestive p-value
with 1500 Cases and 1800 Controls
Assume: Log Additive Model, Prevalence=6.7e-4, p=4.2e-5
Power
1.00
0.80
0.60
0.40
0.20
0.00
0.05
0.10
0.20
0.30
0.40
0.50
Minor Allele Frequency
Relative Risk
1.50
1.40
1.35
1.30
1.25



Supplementary Figure 5 Power of study-wide (top and middle panels) replication (bottom) association analyses with1500 cases and 1800 controls, assuming: additive model, disease prevalence=6.7e-4 and p=2.1e-6, 4.2e-5, and 3.1e-4 respectively. Note: when disease prevalence (P_0) is extremely low, relative risk (RR) is equivalent to odds ratio (OR), i.e. $RR = OR / (1 - P_0 + P_0 * OR) \approx OR$.

Supplementary Table 1 Descriptive characteristics of African American samples used in admixture/association analysis (count and percentage, or means ± standard deviation)

	Cases & Controls	Healthy Controls	All IBD Cases[a]	CD Cases	UC Cases
Total	3308	1797 (54.3)	1511 (45.7)	1088 (32.9)	361 (10.9)
Hopkins[b]	2108	1309 (62.1)	799 (37.9)	533 (25.3)	229 (10.9)
Emory[c]	712	166 (23.3)	546 (76.7)	434 (61.0)	90 (12.6)
Cedars[d]	488	322 (66.0)	166 (34.0)	121 (24.8)	42 (8.6)
Female	2084 (63.0)	1223 (68.1)	861 (57.0)	613 (56.3)	209 (57.9)
Hopkins	1459 (69.2)	948 (72.4)	511 (64.0)	343 (64.4)	143 (62.4)
Emory	375 (52.7)	108 (65.1)	267 (48.9)	208 (47.9)	46 (51.1)
Cedars	250 (51.2)	167 (51.9)	83 (50.0)	62 (51.2)	20 (47.6)
Global YRI[e] (%)	80.9±10.0	81.7±9.7	80.0±10.3	80.2±10.1	80.0±10.6
Hopkins	81.6±9.4	82.4±9.1	80.4±9.7	80.6±9.3	80.1±10.7
Emory	81.0±10.2	82.0±10.0	80.7±10.2	81.1±9.8	80.2±10.9
Cedars	77.8±11.6	78.7±11.1	76.1±12.3	74.9±13.0	78.6±9.8

a. Inflammatory bowel disease (IBD) cases include Crohn's disease (CD), ulcerative colitis (UC), and IBD type undetermined (IBDU) cases;

b. Samples recruited from or coordinated by Johns Hopkins University, including case and control samples from the National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK) IBD Genetics Consortium (IBDGC) and control-only samples from a rheumatoid arthritis (RA) study at the University of Alabama, a systemic lupus erythematosus (SLE) study at the University of Alabama and a type 1 diabetes (T1D) study at University of Virginia;

c. Samples recruited from or coordinated by Emory University;

d. Samples recruited from or coordinated by Cedars-Sinai Medical Center, including case and control samples from Cedars-Sinai and control-only samples from the Pharmacogenetics and Risk of Cardiovascular disease Study Group (PARC);

e. Mean and standard deviation (in parentheses) of global West African (specifically Yoruba in Ibadan, Nigeria or YRI) ancestry proportion for each individual defined as local YRI ancestry estimates average across the autosomal genome.

Supplementary Table 2 Replication of SNPs significant in Caucasian Immunochip Study (CIS)

[See Excel file for details and the following table for header abbreviations]

Header	Meaning
IC_SNP	Tested Immunochip SNP name
CHR	Chromosome
POS	Position for tested SNP
GWA_SNP	Meta-analysis GWA SNP name (see Jostin et al 2012 for details)
R2_GWA_IC	Correlation between Immunochip and GWA SNPs
Type	Association general/specific to phenotype in Caucasian Immunochip Study
A1	Reference allele; Minor allele in current study
A2	Major allele in current study
Frq_CEU	Referance allele frequency for CEU samples in HapMap project
Frq_YRI	Referance allele frequency for YRI samples in HapMap project
MAF	Referance allele frequency for all samples in current study
Frq_Aff	Referance allele frequency for IBD case samples in current study
Frq_Unaff	Referance allele frequency for control samples in current study
IBD_pval	IBD association p-value
IBD_OR	IBD association odds ratio
IBD_L95	Lower bound for 95% confidence interval for IBD association odds ratio
IBD_U95	Upper bound for 95% confidence interval for IBD association odds ratio
CD_pval	CD association p-value
CD_OR	CD association odds ratio
CD_L95	Lower bound for 95% confidence interval for CD association odds ratio
CD_U95	Upper bound for 95% confidence interval for CD association odds ratio
UC_pval	UC association p-value
UC_OR	UC association odds ratio
UC_L95	Lower bound for 95% confidence interval for UC association odds ratio
UC_U95	Upper bound for 95% confidence interval for UC association odds ratio

Supplementary Table 3 Detailed candidate genes in the region of significant admixture

[See Excel file]

Supplementary Table 4 Significant local (*cis*-) eQTL (FDR<0.05)

Chr[a]	SNP	Gene	Beta[b]	Statistic[b]	p-value[c]	FDR[c]
17q21.2	rs1053004	*KRT19*	-0.752	-3.932	1.74E-04	0.038
17q21.2	rs1053004	*TTC25*	0.677	3.479	8.04E-04	0.038
5p13	rs1876141	*C6*	0.536	3.596	5.47E-04	0.038
5p13	rs6866402	*C6*	0.519	3.500	7.52E-04	0.038
5p13	rs1505994	*C6*	0.519	3.500	7.52E-04	0.038
1q24	rs1801274	*CD84*	0.524	3.350	1.22E-03	0.048

a. Chromosomal location;

b. Regression beta and test statistic respectively;

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

References

1. Karlson EW, Sanchez-Guerrero J, Wright EA, et al. A connective tissue disease screening questionnaire for population studies. Ann Epidemiol 1995;5:297-302.

2. Mitchell MK, Gregersen PK, Johnson S, Parsons R, Vlahov D, New York Cancer Project. The New York Cancer Project: rationale, organization, design, and baseline characteristics. J Urban Health 2004;81:301-310.

3. Hom G, Graham RR, Modrek B, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 2008;358:900-909.

4. Betteridge JD, Armbruster SP, Maydonovitch C, Veerappan GR. Inflammatory bowel disease prevalence by age, gender, race, and geographic location in the U.S. military health care population. Inflamm Bowel Dis 2013;19:1421-1427.

5. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013;41:D808-15.