(SUPPLEMENTARY TABLE 1) Study Subjects

STUDY POPULATION	CD cases	controls
Primary dataset:		
Samples Received		
Recruited by MAAIS	227	201
Subjects not included		
< 15% African Ancestry	0	1
Final primary dataset Study Population	227	200
Secondary dataset:		
Samples Received		
Univ. of Chicago	66	0
Med Coll of Wisconsin	59	0
Duke University	12	0
IBDGC (22 UC and IC cases also received)	14	7
Balt. Asthma Controls	0	152
Total Samples cohort 2 received	151	159
Subjects not included		
≤ 15% African Ancestry	2	1
Canadian IBDGC study subjects	12	4
Poor genotyping	10	0
Final secondary dataset Study Population	127	154

(SUPPLEMENTARY TABLE 2) Characteristics of 21 SNPs genotyped (combined dataset)

Gene	SNP no.	Nucleotide change	Location	Minor allele frequency (in controls)	Genotype call rate	HWE *P*	HapMap CEU	HapMap YRI
Candidate SNPs								
NOD2	Rs2066847 1007fs (SNP13)	3020insC [-/C]	Exon11	Allele "+": 0.1%	99.2%	0.98	-: 100% +: 0%	-: 100% +: 0%
NOD2	Rs2066845 G908R (SNP12)	C/G	Exon8	Allele "C": 0.3%	98.9%	0.96	C: 1.7% G: 98.3%	C: 0% G: 100%
NOD2	Rs2066844 R702W (SNP8)	C/T	Exon4	Allele "T": 0.6%	95.5%	0.92	C: 89% T: 11%	C: 100% T: 0%
NOD2	Rs2066842 P268S (SNP5)	C/T (ancestral allele C)	Exon4	Allele "T": 4.4%	98.9%	0.38	C: 63.8% T: 36.2%	C: 100% T: 0%
NOD2	rs35285618 R708H	A/G	Exon4	Allele "A": 1.4%	95.5%	0.80	--	--
NOD2	rs5743279 R790Q	A/G (ancestral allele G)	Exon4	Allele "A": 4.0%	95%	0.04	A: 0% G: 100%	A: 2.5% G: 97.5%
NOD2	rs5743278 A725G	C/G (ancestral allele C)	Exon4	Allele "G": 4.4%	95.5%	0.63	C: 100% G: 0%	C: 91.7% G: 8.3%
ATG16L1	Rs2241880 Thr300Ala	A/G (ancestral allele T)	Exon8	Allele "G":29.3%	98.6%	0.72	T: 45.8% C: 54.2%	T: 72.5% C: 27.5%
IBD5_OCTN1 SLC22A4	Rs1050152 Leu503Phe	C/T (ancestral allele C)	Exon9	Allele "T": 7.6%	98.9%	0.44	C: 54.2% T: 45.8%	C: 100% T: 0.0%
IBD5_OCTN2 SLC22A5	Rs2631367	C/G (ancestral allele C)	Promoter -207	Allele "C":35.7%	97.9%	0.07	C: 51.7% G: 48.3%	C: 60.4% G: 39.6%
IBD5_OCTN2 IGR2230	Rs17622208	A/G (ancestral allele G)	Intron2	Allele "A":18.7%	97.6%	0.15	A: 50.8% G: 49.2%	A: 10% G: 90%
IBD5_IGR2198	Rs11739135	C/G (ancestral allele G)	3'-UTR	Allele "G":11.3%	96.6%	0.46	C: 42.5% G: 57.5%	G: 98.3% C: 1.7%
IBD5_IGR2096	Rs12521868	G/T	intergenic	Allele "T":11.7%	97.1%	0.16	G: 55%	G: 99.2%

		(ancestral allele G)					T: 45%	T: 0.8%
IL23R	Rs7517847	G/T (ancestral allele T)	Intron6	Allele "G":21.4%	98.6%	0.20	G: 50% T: 50%	G: 11.7% T: 88.3%
IL23R	Rs2201841	C/T	Intron7	Allele "G": 13.8%	98.3%	0.55	C: 27.5% T: 72.5%	C: 10.8% T: 89.2%
IL23R	Rs11209026 Arg381Gln	A/G (ancestral allele G)	Exon9	Allele "A": 2.56%	98.6%	0.62	A: 6.8% G: 93.2%	A: 1.7% G: 98.3%
IL23R	Rs1495965	A/G (ancestral allele G)	intergenic	Allele "G": 44.5%	98.3%	0.30	A: 58.8% G: 41.2%	A: 53.3% G: 46.7%
IL23R	Rs10889677	A/C (ancestral allele C)	3'-UTR	Allele "A": 14.7%	98.2%	0.81	A: 27.5% C: 72.5%	A: 10% C: 90%
IRGM	Rs4958847	A/G (ancestral allele G)		Allele "A": 48.7%	98.6%	0.40	A: 9.2% G: 90.8%	A: 55% G: 45%
IRGM	Rs13361189	C/T (ancestral allele C)		Allele "C": 45%	98.6%	0.37	C: 2.5% T: 97.5%	C: 53.3% T: 46.7%
IRGM	Rs10065172	C/T (ancestral allele C)		Allele "T": 45.1%	98.7%	0.43	C: 97.5% T: 2.5%	C: 46.5% T: 53.5%

(SUPPLEMENTARY TABLE 3) Heterogeneity test between primary and secondary replicate dataset

	primary dataset						secondary dataset						Q_test
Marker	OR	95%CI	P	MAF	D.ca/ct	d.ca/ct	OR	95%CI	P	MAF	D.ca/ct	d.ca/ct	P
NOD2_Cins	-	-	0.02	0.01	7/0	447/398	1.23	0.08-19.7	1.00	0.00	1/1	247/303	0.07
NOD2.G908R_GC	1.75	0.16-19.3	1.00	0.00	2/1	452/395	2.45	0.22-27.1	0.59	0.01	2/1	246/301	0.85
NOD2.R702W_CT	2.00	0.51-7.79	0.36	0.01	2/3	435/373	4.67	0.52-42.1	0.19	0.01	4/1	244/285	0.51
NOD2.P268S_CT	1.54	0.87-2.72	0.16	0.06	34/20	418/378	2.07	0.96-4.47	0.08	0.05	18/11	230/291	0.54
NOD2.R708H_GA	1.06	0.28-3.99	1.00	0.01	5/4	437/372	1.63	0.51-5.21	0.56	0.02	7/5	241/281	0.63
NOD2.R790Q_GA	0.98	0.46-2.09	1.00	0.03	15/13	427/363	1.07	0.50-2.33	1.00	0.05	13/14	235/272	0.87
NOD2.A725G_CG	1.20	0.61-2.36	0.61	0.04	21/15	421/361	0.56	0.22-1.42	0.27	0.04	7/14	241/272	0.19
ATG16L_AG	1.47	1.10-1.96	0.01	0.32	163/110	291/288	1.34	0.94-1.92	0.12	0.33	89/90	155/210	0.70
IBD5.OCTN1_CT	1.23	0.76-2.00	0.46	0.09	43/31	411/365	1.82	1.02-3.23	0.04	0.10	31/22	217/280	0.31
IBD5.OCTN2_GC	1.38	1.05-1.83	0.02	0.39	191/137	261/259	1.10	0.78-1.56	0.59	0.38	98/108	150/182	0.32
IGR2230_GA	1.14	0.81-1.62	0.48	0.19	88/69	358/321	1.41	0.94-2.10	0.10	0.23	64/60	182/240	0.44
IGR2198_CG	1.29	0.82-2.02	0.31	0.11	52/36	390/348	1.39	0.88-2.21	0.19	0.16	45/41	201/255	0.81
IGR2096_GT	1.14	0.73-1.76	0.58	0.11	52/40	394/344	1.63	1.03-2.57	0.04	0.16	50/40	198/258	0.26
IL23_rs751_GT	1.02	0.73-1.42	0.93	0.20	93/80	361/316	0.85	0.56-1.28	0.47	0.22	49/70	193/234	0.50
IL23_rs220_AG	1.34	0.92-1.96	0.13	0.16	78/54	370/344	1.49	0.95-2.34	0.10	0.16	47/43	193/263	0.73
IL23_rs112_GA	1.10	0.29-4.13	1.00	0.01	5/4	447/395	0.80	0.34-1.88	0.67	0.04	9/14	233/290	0.69
IL23_rs149_TC	1.18	0.89-1.54	0.27	0.45	207/168	239/228	0.94	0.67-1.32	0.73	0.47	112/145	132/161	0.32
IL23_rs108_CA	1.36	0.94-1.99	0.11	0.16	79/64	369/344	1.05	0.67-1.66	0.91	0.17	41/49	201/253	0.39
IRGM_rs495_GA	0.88	0.67-1.16	0.37	0.47	204/192	248/206	0.93	0.66-1.30	0.67	0.50	119/153	125/149	0.82
IRGM_rs133_TC	0.94	0.72-1.23	0.68	0.44	197/178	257/218	1.05	0.75-1.47	0.80	0.46	112/137	130/167	0.61
IRGM_rs100_CT	0.91	0.69-1.20	0.53	0.44	194/179	258/217	1.02	0.73-1.43	0.93	0.45	112/137	134/167	0.64

MAF, minor allele frequency; D.ca/ct, minor allele cases/controls; d.ca/ct, major allele cases/controls

Heterogeneity was assessed by Q statistic. Cochran WG. The combination of estimates from different experiment. Biometrics 1954;10:101-29.

(SUPPLEMENTARY TABLE 4) Association of SNPs among 227 CD and 200 controls in primary dataset

Candidate Gene	SNP no.	0 copy of variant allele		1 copy of variant allele		2 copies of variant allele		Carrier	MAF case/control	Minor Alleleic test
		Case /control	OR (ref)	Case /control	OR (95%CI) p value	Case /control	OR (95%CI) P value	OR (95%CI) P value		OR (95%CI) P value
NOD2 C_insertion	Rs2066847 1007fs (SNP13)	220/199	1.0	7/0	8.63 (1.28-~) 0.02	0/0	--	8.63 (1.28-~) 0.02	1.54% / 0%	--
NOD2 (G->C, variant C)	Rs2066845 G908R (SNP12)	225/197	1.0	2/1	1.75 (0.16-19.5) 1.00	0/0	--	1.75 (0.16-19.5) 1.0	0.44% / 0.25%	1.75 (0.16-19.3) 1.0
NOD2 (C->T, variant T)	Rs2066844 R702W (SNP8)	214/183	1.0	7/3	2.00 (0.51-7.83) 0.32	0/0	--	1.99 (0.51-7.83) 0.36	1.58% / 0.81%	1.98 (0.51-7.71) 0.36
NOD2 (C->T, variant T)	Rs2066842 P268S (SNP5)	196/179	1.0	26/20	1.19 (0.64-2.20) 0.58	4/0	4.79 (0.59-~) 0.15	1.37 (0.75-2.50) 0.37	7.52% / 5.03%	1.54 (0.87-2.72) 0.16
NOD2 (G->A, variant A)	rs35285618 R708H	216/182	1.0	5/4	1.05 (0.28-3.98) 1.00	0/0	--	1.05 (0.28-3.98) 1.0	1.13% / 1.07%	1.05 (0.28-3.95) 1.0
NOD2 (G->A, variant A)	rs5743279 R790Q	206/172	1.0	15/12	1.04 (0.48-2.29) 1.00	0/0	--	1.04 (0.48-2.29) 1.0	3.39% / 3.26%	1.04 (0.48-2.26) 1.0
NOD2 (C->G, variant G)	rs5743278 A725G	201/172	1.0	19/15	1.08 (0.53-2.20) 0.82	1/0	0.85 (0.02-~) 1.00	1.14 (0.57-2.29) 0.73	4.75% / 4.01%	1.19 (0.61-2.35) 0.74
ATG16L1 (A->G, variant G)	Rs2241880 Thr300Ala	91/108	1.0	109/72	1.80 (1.19-2.70) 0.005	27/19	1.69 (0.88-3.23) 0.11	1.77 (1.21-2.61) 0.004	35.9% / 27.6%	1.47 (1.10-1.96) 0.01
IBD5_OCTN1 SLC22A4 (C->T, variant T)	Rs1050152 Leu503Phe	190/167	1.0	31/31	0.88 (0.51-1.51) 0.64	6/0	7.12 (1.02-~) 0.05	1.05 (0.62-1.77) 0.90	9.47% / 7.83%	1.23 (0.76-2.00) 0.46
IBD5_OCTN2 SLC22A5 (G->C, variant C)	Rs2631367	79/82	1.0	103/95	1.13 (0.74-1.71) 0.58	44/21	2.17 (1.19-3.98) 0.01	1.32 (0.89-1.95) 0.19	42.2% / 34.6%	1.38 (1.05-1.82) 0.02
IBD5_OCTN2 SLC22A5 IGR2230 (G->A, variant A)	Rs17622208	145/128	1.0	68/65	0.92 (0.61-1.40) 0.71	10/2	4.40 (0.91-41.9) 0.07	1.03 (0.69-1.54) 0.92	19.7% / 17.7%	1.14 (0.81-1.62) 0.48
IBD5_IGR2198 (C->G, variant G)	Rs11739135	177/156	1.0	36/36	0.88 (0.53-1.47) 0.63	8/0	9.65 (1.47-~) 0.01	1.08 (0.66-1.76) 0.80	11.7% / 9.37%	1.29 (0.82-2.02) 0.31
IBD5_IGR2096	Rs12521868	178/152	1.0	38/40	0.81	7/0	8.13	0.96	11.7% /	1.14

(G->T, variant T)					(0.50-1.33) 0.41		(1.21-~) 0.03	(0.60-1.55) 0.90	10.4%	(0.73-1.76) 0.58
IL23R (T->G, variant G)	Rs7517847	145/123	1.0	71/70	0.86 (0.57-1.29) 0.47	11/5	1.87 (0.63-5.52) 0.26	0.93 (0.62-1.38) 0.76	20.5% / 20.2%	1.02 (0.73-1.42) 0.93
IL23R (A->G, variant G)	Rs2201841	151/149	1.0	68/46	1.46 (0.94-2.26) 0.09	5/4	1.23 (0.33-4.68) 0.76	1.44 (0.94-2.20) 0.11	17.4% / 13.6%	1.34 (0.92-1.96) 0.13
IL23R (G->A, variant A)	Rs11209026 Arg381Gln	221/195	1.0	5/4	1.10 (0.29-4.17) 0.89	0/0	--	1.10 (0.29-4.17) 0.89	1.11% / 1.01%	1.10 (0.29-4.13) 1.0
IL23R (T->C, variant C)	Rs1495965	63/64	1.0	113/100	1.15 (0.74-1.78) 0.54	47/34	1.40 (0.80-2.46) 0.24	1.21 (0.80-1.84) 0.36	46.4% / 42.4%	1.17 (0.89-1.54) 0.26
IL23R (C->A, variant A)	Rs10889677	150/149	1.0	69/46	1.49 (0.96-2.31) 0.07	5/4	1.24 (0.33-4.71) 0.75	1.47 (0.96-2.25) 0.09	17.6% / 13.5%	1.36 (0.94-1.99) 0.11
IRGM (A->G, variant G)	Rs4958847	53/47	1.0	98/98	0.89 (0.55-1.44) 0.62	75/54	1.23 (0.73-2.08) 0.44	1.00 (0.64-1.58) 0.97	45.1% / 48.2%	0.88 (0.67-1.15) 0.37
IRGM (C->T, variant T)	Rs13361189	45/39	1.0	107/100	0.93 (0.56-1.54) 0.77	75/59	1.10 (0.64-1.91) 0.73	0.99 (0.62-1.60) 0.97	43.4% / 44.9%	0.93 (0.71-1.23) 0.67
IRGM (C->T, variant T)	Rs10065172	76/58	1.0	106/101	0.80 (0.52-1.24) 0.32	44/39	0.86 (0.50-1.49) 0.59	0.82 (0.54-1.24) 0.34	42.9% / 45.2%	0.91 (0.69-1.19) 0.53

(SUPPLEMENTARY TABLE 5) Diplotype association test for *IBD5*

diplotype:	freq.case	freq.ctrl	OR	95%CI	P value
CGGCG/CGGCG	0.33	0.35	1.00 (ref)	-	-
CGGCG/CCGCG	0.23	0.26	0.91	0.60-1.37	0.7
CGGCG/TCAGT	0.09	0.11	0.93	0.52-1.64	0.88
CGGCG/CCACG	0.10	0.10	0.98	0.54-1.76	1
CCGCG/CCGCG	0.05	0.03	1.47	0.61-3.65	0.47
CCGCG/CCACG	0.04	0.02	2.42	0.76-9.03	0.16
CGGCG/CGAGT	0.01	0.05	0.25	0.03-1.31	0.12
CGAGT/CCGCG	0.00	0.02	0.42	0-5.46	0.51
CCGCG/TCAGT	0.04	0.03	1.41	0.56-3.70	0.56
CGGCG/TCAGT	0.01	0.00	3.02	0.24-160	0.63
TCAGT/TCAGT	0.02	0.00	8.14	1.16 - ∞	0.03
CCACG/TCAGT	0.02	0.01	3.02	0.53-31	0.3
CCGCG/CCAGT	0.01	0.00	5.27	0.65 - ∞	0.13
CCACG/CCACG	0.00	0.01	0.34	0.006-4.28	0.64
CGAGT/CGAGT	0.00	0.01	0.51	0.008-9.8	1

Diplotype association used PHASE Bayesian approach and partition-ligation algorithm to reconstruct haplotypes and estimate diplotype probabilities for CD and HC, separately. (Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978-89.) In the regression model, (CD served as the dependent variables, and the diplotype probabilities served as continuous variables.

(SUPPLEMENTARY TABLE 6)
Power calculation: Sample size 354 cases and 1:1 control; CD Population risk: 50/100,000; 2-sided p value 0.05;

	Allele freq in present study data and OR estimates from prior and present studies	**Power Estimate**
NOD2 any mutant	Allele freq 1.03% in study AA HCs	
	OR 3.2 (whites)[38], log. additive mode	83%
	OR 3.3 (present data), log. additive mode	86%
SLC22A4	Allele freq 7.6% in study AA HCs	
(variant C)	OR 1.7 (white)[30] dominant mode	78%
	OR 8.4 (present data), recessive mode	95%
SLC22A5	Allele freq 35.7% in study AA HCs	
(variant C)	OR 2.0 (white)[30],dominant mode	99%
	OR 1.9 (present data), recessive mode	87%
ATG16L1	Allele freq 29.3% in study AA HCs	
	OR 2.0 (white)[30], dominant mode	99%
	OR 1.4 (present data), dominant mode	62%
IL23R rs2201841	Allele freq 13.8% in study AA HCs	
	OR 2.2 (white)[30], dominant mode	99%
	OR 1.3 (present data), dominant mode	35%
IRGM rs13361189	Allele freq 45% in study AA HCs	
	OR 1.4 - 1.6 (white)[19], dominant mode*	51 - 78%

*Taken from allelic OR. Data for mode of inheritance not available. Dominant mode assumed over additive mode as more conservative. OR from Ref 19 was 1.6, a slightly larger cohort than the Wellcome Trust discovery cohort (allelic OR of 1.4). *IRGM* power estimates are not given from present study data as OR was 0.98 (Table 2).