Supplementary Methods

Research Subjects

1. IBD discovery and replication cohorts:

Subject ascertainment and diagnostic classification.

Affected individuals with pediatric onset IBD (both CD and UC) were ascertained through the Children's Hospital of Wisconsin and Medical College of Wisconsin, Children's Hospital of Philadelphia, and Cincinnati Children's Hopsital Medical Center. Additional UC cases were recruited from Primary Children's Medical Center and from the University of Utah and the Clinica Pediatrica, Università 'La Sapienza', Roma, Italy. In addition, colonic mucosal biopsies from affected IBD patients were obtained fronm Cincinnati Children's Medical center and from Children's Hospital of Wisconsin during the diagnostic endoscopic procedures. Only subjects of European ancestry were used in the final analysis which consisted of 1,011 individuals with IBD (including 647 CD and 317 UC, with the remainder being indeterminate colitis) in the discovery cohort and 173 cases in the replication cohort where the age of onset for IBD was before their 19th birthday. All subjects had genotypes with call rates above 95%. Informed consent was obtained from all participants, and protocols were approved by the local institutional review board in all participating institutions. The diagnosis of IBD was made after fulfilling standard criteria across the participating centers that requires (i) one or more of the following symptoms: diarrhea, rectal bleeding, abdominal pain, fever or complicated perianal disease; (ii) occurrence of symptoms on two or more occasions separated by at least 8 weeks or ongoing symptoms of at least 6 weeks' duration and (iii) objective evidence of inflammation from radiologic, endoscopic, video capsule endoscopy. Histological evidence of IBD[1] was considered mandatory for the diagnosis of CD or UC and inclusion in the study.

Phenotypic classification (Supplementary Table 3) was based on the Montreal classification[2]. For CD we defined disease location based on each subjects all available endoscopic and radiographic evaluation. Based on macroscopic evidence of disease location, we classified each subject by the following: Ileum only: disease of the small bowel proximal to the cecum and distal 4th portion of duodenum; Colon only: any colonic location between cecum and rectum with no small bowel disease; Ileocolonic: disease of the small bowel and any location between cecum and rectum. In addition, any of the above categories may *have upper GI tract involvement:* disease involving esophagus, stomach, duodenum and perianal disease including: perianal fistulae, perianal and anal lesions including more than single skin tags and anal ulcers. For example, subjects with ileal only, colonic only or

ileocolonic disease may also have concomitant upper tract and/or perianal disease.

2. Control Subjects from Philadelphia:
The discovery control group included 4250 children and the replication group consisted of an independent cohort of 3481 children with self reported European descent, mean age 9.5 years; 53.0 % male and 47.0 % female, who did not have IBD (CD or UC). These individual were recruited by CHOP's clinicians and nursing staff within the CHOP's Health Care Network, including four primary care clinics and several group practices and outpatient practices that included well child visits. The Research Ethics Board of CHOP approved the study, and written informed consent was obtained from all subjects.

Genotyping
Illumina Infinium™ assay: We performed high throughput genome-wide SNP genotyping, using the Illumina Infinium™ II HumanHap550 BeadChip technology[3,4] (Illumina, San Diego), at the Center for Applied Genomics at CHOP. We used 750ng of genomic DNA to genotype each sample, according to the manufacturer's guidelines. On day one, genomic DNA was amplified 1000-1500-fold. Day two, amplified DNA was fragmented ~300-600bp, then precipitated and resuspended followed by hybridization on to a BeadChip. Single base extension utilizes a single probe sequence ~50bp long designed to hybridize immediately adjacent to the SNP query site. Following targeted hybridization to the bead array, the arrayed SNP locus-specific primers (attached to beads) were extended with a single hapten-labeled dideoxynucleotide in the SBE reaction. The haptens were subsequently detected by a multi-layer immunohistochemical sandwich assay, as recently described. The Illumina BeadArray Reader scanned each BeadChip at two wavelengths and created an image file. As BeadChip images were collected, intensity values were determined for all instances of each bead type, and data files were created that summarized intensity values for each bead type. These files consisted of intensity data that was loaded directly into Illumina's genotype analysis software, BeadStudio. A bead pool manifest created from the LIMS database containing all the BeadChip data was loaded into BeadStudio along with the intensity data for the samples. BeadStudio used a normalization algorithm to minimize BeadChip to BeadChip variability. Once the normalization was complete, the clustering algorithm was run to evaluate cluster positions for each locus and assign individual genotypes. Each locus was given an overall score based on the quality of the clustering and each individual genotype call was given a GenCall score. GenCall scores provided a

quality metric that ranges from 0 to 1 assigned to every genotype called. GenCall scores were then calculated using information from the clustering of the samples. The location of each genotype relative to its assigned cluster determined its GenCall score.

Gene Array Analysis. The global pattern of gene expression in colon was determined in the Microarray Core of the CCHMC Digestive Health Center[5]. Following informed consent, colonic biopsies were obtained from pediatric patients with CD and UC and healthy controls. All of the biopsies for IBD patients and healthy controls were obtained from the ascending colon, with the exception of one subject with UC whose biopsy was obtained from the rectum.. Colon biopsies were immediately placed in RNAlater stabilization reagent (Qiagen, Germany) at 4 C. Total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen) and stored at -80 C. Samples where then submitted to the CCHMC Digestive Health Center Microarray Core where the quality and concentration of RNA was measured by the Agilent Bioanalyser 2100 (Hewlett Packard) using the RNA 6000 Nano Assay to confirm a 28S/18S ratio of 1.6-2.0. 100 ng of total RNA was amplified using Target 1-round Aminoallyl-aRNA Amplification Kit 101 (Epicentre, WI). The biotinylated cRNA was hybridized to Affymetrix GeneChip Human Genome HG-U133 Plus 2.0 arrays, containing probes for approximately 22,634 genes. The images were captured using Affymetrix Genechip Scanner 3000. The complete dataset is available at the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo); the accession number is GSE10616. GeneSpring♠ software was used in the CCHMC Digestive Health Center Bioinformatics core to analyze fold changes in gene expression between patient groups and healthy controls. Data were normalized to an internal control reference sample, and then to the median of the healthy control values, to allow for array to array comparisons, and differences between groups were detected in GeneSpring™ with significance at the 0.05 level relative to healthy control samples. In order to allow for comparison between the IBD sub-groups, mucosal inflammation was quantified in colon biopsies using the Crohn's Disease Histological Index of Severity.

DCR3 ELISA. The DuoSet DCR3 ELISA Kit was obtained from R&D Systems (Minneapolis, MN), catalog #DY142. Ninety-six-well Nunc Maxisorb plates were used. PBS was obtained from Sigma-Aldrich (St. Louis, MO). All other materials were obtained from R&D Systems. Serum samples were diluted 1:100 in PBS/0.05% Tween 20/1% BSA, and the assay was performed as per manufacturer's instruction. The sensitivity of the assay was 47.42 pg/ml.

Supplementary references

1. Bousvaros, A. et al. Differentiating ulcerative colitis from Crohn disease in children and young adults: report of a working group of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the Crohn's and Colitis Foundation of America. *J Pediatr Gastroenterol Nutr* **44**, 653-74 (2007).

2. Silverberg, M.S. et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* **19 Suppl A**, 5-36 (2005).

3. Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. & Chee, M.S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**, 549-54 (2005).

4. Steemers, F.J. et al. Whole-genome genotyping with the single-base extension assay. *Nat Methods* **3**, 31-3 (2006).

5. Carey, R. et al. Activation of an IL-6:STAT3-dependent transcriptome in pediatric-onset inflammatory bowel disease. *Inflamm Bowel Dis* (2007).

Supplementary Table 1: CD case-control association study results for GWA significant markers. Minor allele frequencies (MAF), *P*-values and odds ratios (OR) are shown. The ORs shown are for the minor alleles (as observed in the controls). Combined *P*-values are also shown, together with the most relevant gene in which the markers reside or which they are nearest to. *P*-values are two-sided and uncorrected in each instance. Aff. allele freq., allele frequency in affected individuals; Chr., chromosome; CI, confidence interval; Ctrl allele freq., allele frequency in unaffected individuals.

CHR	SNP	Position (B36)	Minor Allele	MAF Aff	MAF Ctrl	*P*-value	Bonferonni *P*	OR	95% CI	Relevant Gene
16	rs5743289	49314275	T	0.257	0.172	1.21×10^{-13}	6.22×10^{-8}	1.671	1.46-1.92	*CARD15*
1	rs11209026	67478546	A	0.018	0.061	3.35×10^{-10}	0.00017	0.281	0.18-0.43	*IL23R*
2	rs2241880	233848107	T	0.396	0.488	7.63×10^{-10}	0.00039	0.687	0.61-0.77	*ATG16L1*
2	rs2289472	233846979	A	0.398	0.489	1.10×10^{-9}	0.00056	0.691	0.61-0.78	*ATG16L1*
2	rs13391356	233835108	T	0.399	0.489	1.31×10^{-9}	0.00067	0.693	0.61-0.78	*ATG16L1*
16	rs2076756	49314382	G	0.338	0.258	1.88×10^{-9}	0.00097	1.465	1.30-1.66	*CARD15*
2	rs3792109	233849156	T	0.399	0.488	3.41×10^{-9}	0.0018	0.699	0.62-0.79	*ATG16L1*
16	rs2066843	49302700	T	0.351	0.272	3.61×10^{-9}	0.0019	1.449	1.28-1.64	*CARD15*
1	rs11465804	67475114	G	0.024	0.065	7.64×10^{-9}	0.0039	0.355	0.25-0.51	*IL23R*

Supplementary Table 2: UC case-control association study results for GWA significant markers. Minor allele frequencies (MAF), *P*-values and odds ratios (OR) are shown. The ORs shown are for the minor alleles (as observed in the controls). Combined *P*-values are also shown, together with the most relevant gene in which the markers reside or which they are nearest to. *P*-values are two-sided and uncorrected in each instance. Aff. allele freq., allele frequency in affected individuals; Chr., chromosome; CI, confidence interval; Ctrl allele freq., allele frequency in unaffected individuals.

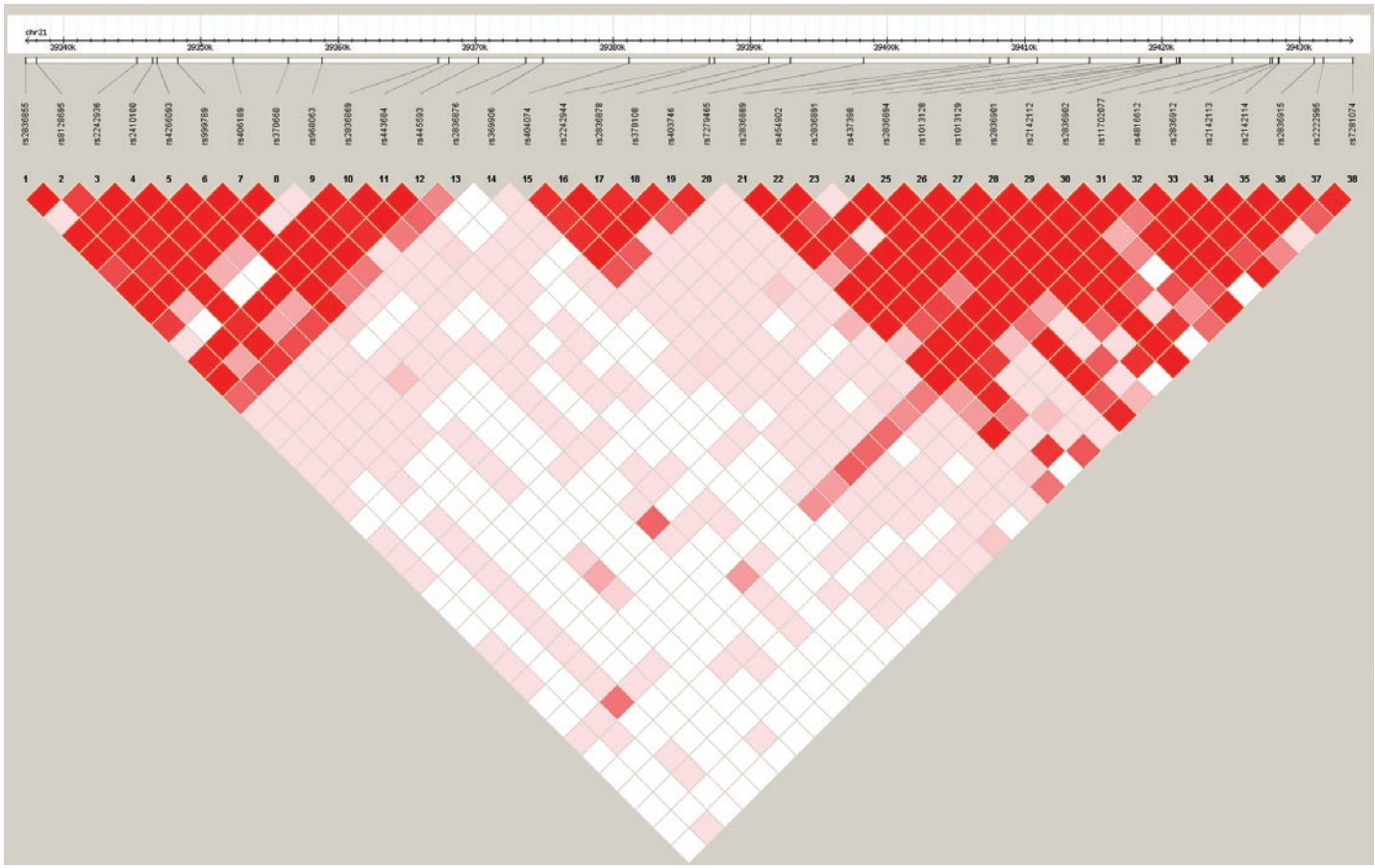
CHR	SNP	Position (B36)	Minor Allele	MAF Aff	MAF Ctrl	*P*-value	Bonferonni *P*	OR	95% CI	Relevant Gene	Eigenstra *P*
6	rs9271568	32698441	A	0.148	0.301	8.22×10^{-16}	4.22×10^{-10}	0.402	0.32-0.51	*MHC*	5.21×10^{-10}
6	rs2516049	32678378	G	0.167	0.313	1.17×10^{-14}	6.02×10^{-9}	0.440	0.36-0.54	*MHC*	4.20×10^{-10}
6	rs477515	32677669	T	0.167	0.313	1.24×10^{-14}	6.36×10^{-9}	0.440	0.36-0.54	*MHC*	4.45×10^{-10}
6	rs2395185	32541145	T	0.177	0.325	1.97×10^{-14}	1.01×10^{-8}	0.447	0.36-0.55	*MHC*	1.06×10^{-9}
6	rs3104404	32790152	A	0.353	0.230	3.10×10^{-12}	1.59×10^{-6}	1.823	1.54-2.16	*MHC*	
6	rs3129882	32517508	G	0.579	0.452	5.76×10^{-10}	0.00030	1.670	1.42-1.97	*MHC*	
6	rs6903608	32536263	C	0.445	0.328	1.71×10^{-9}	0.00088	1.644	1.40-1.94	*MHC*	
6	rs3129763	32698903	A	0.374	0.264	1.80×10^{-9}	0.00093	1.667	1.41-1.97	*MHC*	
6	rs602875	32681607	G	0.377	0.268	3.75×10^{-9}	0.0019	1.650	1.40-1.95	*MHC*	
6	rs382259	32317005	G	0.429	0.317	6.93×10^{-9}	0.0036	1.617	1.37-1.90	*MHC*	
3	rs2245556	102098240	T	0.063	0.145	8.34×10^{-9}	0.0043	0.396	0.29-0.55	*ABI3BP*	
6	rs660895	32685358	G	0.101	0.188	4.39×10^{-8}	0.023	0.485	0.37-0.63	*MHC*	
3	rs2595893	102160532	C	0.066	0.144	4.44×10^{-8}	0.023	0.421	0.31-0.58	*ABI3BP*	
6	rs1035798	32259200	T	0.375	0.274	4.57×10^{-8}	0.023	1.591	1.35-1.88	*MHC*	
3	rs2245473	102098826	G	0.064	0.142	4.64×10^{-8}	0.024	0.414	0.30-0.57	*ABI3BP*	
4	rs7663239	38462245	G	0.125	0.068	7.50×10^{-8}	0.039	1.965	1.53-2.52	*TLR1*	
6	rs3135363	32497626	C	0.391	0.290	8.32×10^{-8}	0.043	1.571	1.33-1.86	*MHC*	

Supplementary Table 3. Patient characteristics

Patient's characteristics	Crohn's disease (n=647)	Ulcerative colitis (n=317)
Gender (male)	56%	53%
Mean Age of onset (range)	11.1 (1-18)	11.1 (1-18)
Disease location (n)%	Colon only (29%) Ileocolon (54%) Ileum only (17%)	Pancolitis (89%) Left sided colitis (11%)

Supplemental Figures.

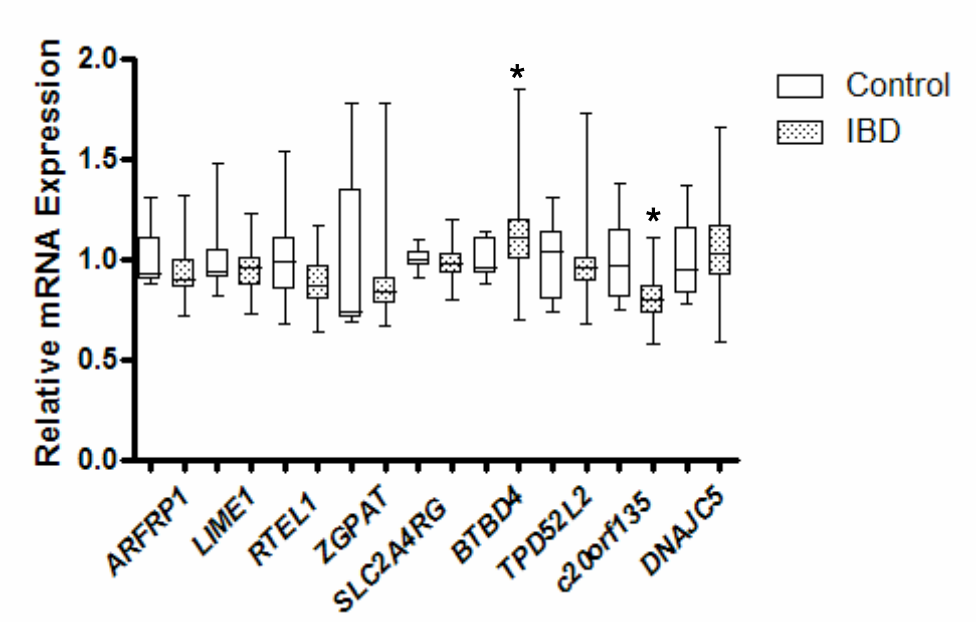
Supplemental Figure 1. Linkage disequilibrium (D') between SNPs at the 21q22 locus in the control cohort together. The association signal resides in a region of LD that harbors no genes; however, *PSMG1* represents the nearest gene geographically.


chr21
39240k
39250k
39260k
39270k
39280k
39290k
39300k
39310k
39320k
39330k
39340k
rs2836855
rs8128695
rs2242936
rs2410160
rs4266093
rs999789
rs4065189
rs370660
rs968063
rs2836869
rs443364
rs445593
rs2836876
rs369006
rs404074
rs2242944
rs2836878
rs378108
rs403746
rs7279465
rs2836889
rs454902
rs2836891
rs437398
rs2836894
rs1013128
rs1013129
rs2836901
rs2142112
rs2836902
rs11702877
rs4816612
rs2836912
rs2142113
rs2142114
rs2836915
rs2229985
rs7281074
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38

Supplemental Figure 2. Colonic expression of genes at the 20q13 locus and *PSMG1*. Colon biopsies were obtained from healthy controls (n=11), and affected segments for CD patients with ileo-colonic (n=18) or colon-only (n=14) location and UC patients (n=10). RNA was prepared and the global pattern of gene expression was determined using the Affymetrix GeneChip Human Genome HG-U133 Plus 2.0 array. Data were normalized to an internal control reference sample, and then to the median value of the healthy control samples, to allow for comparison of mRNA expression in the IBD patient colon samples relative to the healthy control samples. Results for A) genes at the 20q13 locus and B) *PSMG1* are shown. *$p<0.05$ vs. control.

Supplemental Figure 2

A)



B)