Online Methods

Study samples

Meta-analysis: All study subjects were of European origin. The meta-analysis was based on data from 6,687 UC cases and 19,718 population controls derived from six index genome-wide scans from Cedars-Sinai[5], Germany[4,33], Sweden[5], the CHOPSTICKS consortium (of early onset cases)[10,34], the NIDDK IBD Genetics Consortium[35] and the WTCCC2[2]. All cases were ascertained using standard clinical, endoscopic and histopathological criteria for diagnosis of UC. Details of the genotyping platform and number of cases and controls in each study are given in **Supplementary Table 1**. Sample ascertainment and quality control procedures are fully described in the index publications. Controls for the Cedars-Sinai study were obtained from the Cardiovascular Health Study (CHS)[36], a population-based longitudinal study of risk factors for cardiovascular disease and stroke in adult 65 years of age or older, recruited at four field centers. 5201 individuals, predominantly of European descent, were recruited in 1989-1990 from random samples of Medicare eligibility lists, followed by an additional 687 African-Americans recruited in 1992-1993 (total n=5888).

Follow-up: All samples were of European origin. Details of the follow-up panel of 9,628 cases and 12,917 controls are provided in **Supplementary Table 2**. Each centre supplying cases also supplied its own panel of population controls. All participating centers received approval from their local and national institutional review boards, and informed consent was obtained from all participants in the study.

Statistical Methods

Imputation: Genome-wide SNP imputation was carried out using BEAGLE[37] and the HapMap 3 reference samples from the CEU, TSI, MEX and GJT cohorts, with the exception of the CHOPSTICKS samples, which were imputed using the MACH

program (http://www.sph.umich.edu/csg/abecasis/MACH/) and the HapMap2 CEU reference samples. In total, 1,428,850 autosomal markers (HapMap3 X chromosome data were not available) were available for association analysis in at least one GWAS dataset.

Meta-analysis: To carry out the meta-analysis we employed methodology used in both our previous CD meta-analyses[14,38]. For each individual and SNP, post-imputation genotype class probabilities were converted to allelic dosages. For example, the allelic dosage for the A allele at a given SNP with post-imputation genotype probabilities of AA:0.8, Aa:0.1 and aa:0.1 in a given individual is 2×0.8 + 1×0.1 + 0×0.1 = 1.7 (and the allelic dosage for allele *a* is 0.3). Per SNP, the dosage at each allele is summated separately across cases and controls within each of the GWAS studies. An empirical variance is calculated on a per SNP per study basis to correctly weight the association analysis according to GWAS sample size. The empirical variance (σ^2) is given by

$$\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

where x_i is the allelic dosage for individual *i* at the reference allele, μ is the mean dosage across all individuals in the study, and *N* is the total number of individuals in the study. Z scores were calculated using the empirical variance (rather than the binomial variance) and summated across studies to yield a meta-analysis Z score and P-value.

Locus-definition: Linkage disequilibrium regions around focal SNPs were calculated by extending in both directions a distance of 0.1cM. If another SNP within this region had a $P<10^{-5}$ the addition of 0.1cM was repeated from this SNP[2,38]. A gene was considered a positional candidate if the index SNP was located within the gene.

Follow-up

Loci previously associated with UC at $P<5\times10^{-8}$ were not taken forward for follow-up. The most associated SNP from remaining loci with $P < 1x10^{-5}$ was selected for follow-up if it was associated at $P < 0.05$ in at least two GWAS (or three studies if only a single SNP in the locus reached $P < 1x10^{-5}$). This reduces the possibility that SNPs are followed up due to bias in any one study. In total, 50 SNPs/loci were taken forward for follow-up. SNPs were selected for follow-up based on the results of a pre-final version of the meta-analysis. As a result, the lead SNP in the final version of the meta-analysis was not selected for follow-up for two loci – 5p15(rs6451493): rs348594, 16q12 (rs11644386):rs1894942 and 17q21(rs12942547): rs744166. Genotyping was carried out according to standard protocols associated with the various genotyping platforms (**Supplementary Table 2**). Individuals with more than 10% missing data were removed in addition to SNPs with more than 10% missing data or a Hardy-Weinberg equilibrium $P < 0.0001$. Duplicate samples across studies were identified (identity by state = 2) and the sample with the lowest call rate removed. Follow-up and joint P-values were calculated using the weighted *Z* score method described above, using binomial variance for genotype data. ORs and corresponding 95% confidence intervals were also calculated. A genome-wide significance threshold of $P < 5\times10^{-8}$ for combined analysis was adopted. Study/cohort specific results for all SNPs listed in **Table 1** and **Table 2** are given in **Supplementary Table 3 and region wide association plots (created using SNAP[39]) are given in Supplementary Figure 2**.

Using genotype data from the follow-up cohort, we tested all newly confirmed loci for association to three disease subphenotypes. Disease location was defined using the Montreal classification system and we compared a) proctitis (E1) to left-sided disease (E2) and pancolitis (E3) and b) proctitis (E1) and left-sided disease (E2) to pancolitis (E3). We also compared cases requiring surgical intervention for acute severe or medically refractory disease to those with no recorded surgical intervention. Individuals receiving colectomy for dysplasia or malignancy were excluded. Cochran-Mantel-Haenszel tests were conducted stratifying by cohort and

the Bonferroni significance threshold (0.05/87:P<5.75×10^{-4}) was adopted.

GRAIL analysis

To obtain insight in the functional relatedness of all UC loci we performed a Gene Relationship Across Implicated Loci (GRAIL) pathway analysis (http://www.broadinstitute.org /mpg/grail). GRAIL is a statistical tool that uses text mining of PubMed abstracts to annotate candidate genes from loci associated with disease risk[40]. We used HG17 and December 2006 PubMed datasets, default settings for SNP rsNumber submission and all 47 UC loci as query and seed. 7/47 SNPs were not in HapMap r21 so the second, third or fourth most strongly associated SNP from the GWAS meta-analysis was used.

eQTL analysis

We used eQTL data from 1469 peripheral blood DNA and RNA (PAXgene) samples from Dutch and UK individuals, described in detail by Dubois et al[11]. All samples had been genotyped using either an Illumina Hap370 or 610-Quad platform. SNPs that had a MAF≥5%, call-rate≥95% and exact HWE P>0.001 were included. Imputation for ungenotyped SNPs was performed using IMPUTE software (https://mathgen.stats.ox.ac.uk/impute/impute.html). We applied a window of 500kb around each SNP (250kb on each side). cis-eQTLs were considered statistically significant with a Spearman P < 0.0055 corresponding to a 5% false discovery rate (FDR). 42/47 UC genome-wide significant UC associated loci were included for analysis (**Supplementary Table 4**).

nsSNP analysis

Data from the 1000 Genomes Project March 2010 release (ftp://ftp.sanger.ac.uk/pub/1000genomes/REL-1005/QCALL/) were used to find non-synonymous, splice or stop-encoding SNPs in high LD ($r^2\geq0.5$) with our most associated SNP within the locus. Results are shown in **Supplementary Table 5**.

Shared association signals between UC and CD

We selected the most associated SNP for each of the 99 reported UC/CD loci from the current study and our recent CD meta-analysis[14]. From those loci, 71 were reported in CD and 47 are reported in UC, including an overlap of 19 loci reported in both studies. When for a given locus these SNPs differed between the two studies, we included both SNPs in downstream analyses and calculated the correlation between them (in terms of r^2 and D'). Loci were defined by physical position, which means that, for some loci, more than one independent signal may be present. The overlapping loci are those with overlapping physical intervals. For all 112 index SNPs meeting our criteria, we extracted association results from both UC and CD scans, and aligned them to the same reference allele. For a signal of association to be considered shared, the index SNP from one disease needed to achieve $P<1x10^{-4}$ in the other and show the same direction of effect. This threshold was selected because of the number of loci tested, similarity of sample size between studies and the known biological and clinical similarities between UC and CD. Due to the inclusion of some shared controls between the CD and UC meta-analyses, we expect some small correlation between their Z-scores. This could explain some excess of shared directionality for non-significant associations, but is not enough to explain a shared association signal achieving $P<1x10^{-4}$ in both scans.

Functional enrichment analysis

To assess the statistical enrichment of functional gene sets from molecular function, biological process categories and pathways for candidate genes, p-values were computed using the hypergeometric test[41] and implemented in the R programming language as described in Hitomi et. al.[42] The gene sets were compiled from multiple sources: molecular function and biological process categories from Gene Ontology (GO)[43] and Panther[44,45]; canonical pathways from MSigDB[46] and Panther. A weighted Jaccard coefficient was used to compute gene overlap[47,48]. Strongly connected components in the network were identified using Tarjan's algorithm[49]. It should be noted that gene length was not taken into account in this enrichment analysis, which can can potentially bias enrichment of pathways containing long genes, however,

the genes that contributed to the strong immune enrichment featured in the network (Supplementary Figure 3 and Supplementary Table 8) are considered relatively short (mean: 32.3kb, median: 18.9kb) as compared to the average length of all GO annotated genes (106.5kb as per Staley SM, Bailey TL, Mattick JS. GENOME: measuring correlations between GO terms and genomic positions. BMC Bioinformatics. 2006; 7:94).

Variance explained

The proportion of variance explained is based on the liability threshold model and formula outlined by Risch and Merikangas[50] and assuming a population prevalence of 0.0024[51], an MZ/DZ concordance of 10%/3%[52], an MZ/DZ correlation of 0.196/0.055 and thus a heritability of 28%. ORs and allele frequency estimates are taken from the follow-up cohort where available and the meta-analysis cohort otherwise.

18. Grenningloh, R., Kang, B.Y. & Ho, I.C. Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. *J Exp Med* **201**, 615-26 (2005).
19. Moisan, J., Grenningloh, R., Bettelli, E., Oukka, M. & Ho, I.C. Ets-1 is a negative regulator of Th17 differentiation. *J Exp Med* **204**, 2825-35 (2007).
20. Sabath, E. et al. Galpha12 regulates protein interactions within the MDCK cell tight junction and inhibits tight-junction assembly. *J Cell Sci* **121**, 814-24 (2008).
21. Bettelli, E., Korn, T., Oukka, M. & Kuchroo, V.K. Induction and effector functions of T(H)17 cells. *Nature* **453**, 1051-7 (2008).
22. Steinberg, M.W. et al. A crucial role for HVEM and BTLA in preventing intestinal inflammation. *J Exp Med* **205**, 1463-76 (2008).
23. Maerten, P. et al. Involvement of 4-1BB (CD137)-4-1BBligand interaction in the modulation of CD4 T cell-mediated inflammatory colitis. *Clin Exp Immunol* **143**, 228-36 (2006).
24. Mahida, Y.R., Wu, K. & Jewell, D.P. Enhanced production of interleukin 1-beta by mononuclear cells isolated from mucosa with active ulcerative colitis of Crohn's disease. *Gut* **30**, 835-8 (1989).
25. Williams, E.J. et al. Distribution of the interleukin-8 receptors, CXCR1 and CXCR2, in inflamed gut tissue. *J Pathol* **192**, 533-9 (2000).
26. Noble, C.L. et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut* **57**, 1398-405 (2008).
27. Schluns, K.S., Kieper, W.C., Jameson, S.C. & Lefrancois, L. Interleukin-7 mediates the homeostasis of naive and memory CD8 T cells in vivo. *Nat Immunol* **1**, 426-32 (2000).
28. Yamazaki, M. et al. Mucosal T cells expressing high levels of IL-7 receptor are potential targets for treatment of chronic colitis. *J Immunol* **171**, 1556-63 (2003).
29. Gregory, S.G. et al. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* **39**, 1083-91 (2007).
30. Fumagalli, M. et al. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *J Exp Med* **206**, 1395-408 (2009).
31. Pandey, A.K. et al. NOD2, RIP2 and IRF5 play a critical role in the type I interferon response to Mycobacterium tuberculosis. *PLoS Pathog* **5**, e1000500 (2009).
32. Liu, L. et al. LSP1 is an endothelial gatekeeper of leukocyte transendothelial migration. *J Exp Med* **201**, 409-18 (2005).

Online Methods

33. Franke, A. et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* **40**, 1319-23 (2008).
34. Imielinski, M. et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet* **41**, 1335-40 (2009).

35. Silverberg, M.S. et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet* **41**, 216-20 (2009).
36. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
37. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* **84**, 210-23 (2009).
38. Barrett, J.C. et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-62 (2008).
39. Johnson, A.D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938-9 (2008).
40. Raychaudhuri, S. et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* **5**, e1000534 (2009).
41. Rivals, I., Personnaz, L., Taing, L. & Potier, M.C. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **23**, 401-7 (2007).
42. Hitomi, J. et al. Identification of a molecular signaling network that regulates a cellular necrotic cell death pathway. *Cell* **135**, 1311-23 (2008).
43. The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res* **38**, D331-5.
44. Thomas, P.D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**, 2129-41 (2003).
45. Mi, H. et al. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* **38**, D204-10.
46. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
47. Dhillon, I.S., Marcotte, E.M. & Roshan, U. Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics* **19**, 1612-9 (2003).
48. Pinto, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368-72.
49. Tarjan, R. Depth-first search and linear graph algorithms. *SIAM Journal of computing* **1**, 146-160 (1972).
50. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-7 (1996).
51. Rubin, G.P., Hungin, A.P., Kelly, P.J. & Ling, J. Inflammatory bowel disease: epidemiology and management in an English general practice population. *Aliment Pharmacol Ther* **14**, 1553-9 (2000).
52. Ahmad, T., Satsangi, J., McGovern, D., Bunce, M. & Jewell, D.P. Review article: the genetics of inflammatory bowel disease. *Aliment Pharmacol Ther* **15**, 731-48 (2001).