nature genetics

Ulcerative colitis–risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study

Mark S Silverberg[1,22], Judy H Cho[2,3,22], John D Rioux[4], Dermot P B McGovern[5], Jing Wu[6], Vito Annese[7], Jean-Paul Achkar[8], Philippe Goyette[4], Regan Scott[9], Wei Xu[10], M Michael Barmada[11], Lambertus Klei[12], Mark J Daly[13], Clara Abraham[2], Theodore M Bayless[14], Fabrizio Bossa[7], Anne M Griffiths[15], Andrew F Ippoliti[5], Raymond G Lahaie[16], Anna Latiano[7], Pierre Paré[17], Deborah D Proctor[2], Miguel D Regueiro[9], A Hillary Steinhart[1], Stephan R Targan[5], L Philip Schumm[18], Emily O Kistner[18], Annette T Lee[19], Peter K Gregersen[19], Jerome I Rotter[5,20], Steven R Brant[14,21], Kent D Taylor[5,20], Kathryn Roeder[6] & Richard H Duerr[9,11]

© 2009 Nature America, Inc. All rights reserved.

Ulcerative colitis is a chronic inflammatory disease of the colon that presents as diarrhea and gastrointestinal bleeding. We performed a genome-wide association study using DNA samples from 1,052 individuals with ulcerative colitis and preexisting data from 2,571 controls, all of European ancestry. In an analysis that controlled for gender and population structure, ulcerative colitis loci attaining genome-wide significance and subsequent replication in two independent populations were identified on chromosomes 1p36 (rs6426833, combined $P = 5.1 \times 10^{-13}$, combined odds ratio OR = 0.73) and 12q15 (rs1558744, combined $P = 2.5 \times 10^{-12}$, combined OR = 1.35). In addition, combined genome-wide significant evidence for association was found in a region spanning *BTNL2* to *HLA-DQB1* on chromosome 6p21 (rs2395185, combined $P = 1.0 \times 10^{-16}$, combined OR = 0.66) and at the *IL23R* locus on chromosome 1p31 (rs11209026, combined $P = 1.3 \times 10^{-8}$, combined OR = 0.56; rs10889677, combined $P = 1.3 \times 10^{-8}$, combined OR = 1.29).

Ulcerative colitis (UC) and Crohn's disease (CD) are the two main forms of inflammatory bowel disease (IBD). Extensive evidence suggests that IBD occurs when ubiquitous commensal enteric bacteria initiate and perpetuate a dysregulated mucosal immune response in genetically susceptible individuals[1]. CD and UC share many characteristics but unique clinical features also distinguish them, suggesting that they share some genetic susceptibility loci but differ at others.

Genome-wide association studies (GWAS) in CD and a CD GWAS meta-analysis have provided genome-wide significant evidence for 32 CD-associated loci[2–7]. The most significant CD-associated loci include *IL23R*, *NOD2*, *ATG16L1* and a gene desert on chromosome 5p13, implicating both the innate and acquired immune responses. A nonsynonymous SNP scan found association between UC and *ECM1* on chromosome 1q21 (ref. 8). The same study and another also reported association between UC and some CD loci[8,9]. A GWAS in a pediatric-onset IBD sample identified general IBD loci on chromosomes 20q13 and 21q22 (ref. 10). A UC GWAS in European samples found combined, genome-wide significant evidence for

[1]Mount Sinai Hospital Inflammatory Bowel Disease Group, University of Toronto, 600 University Avenue, Toronto, ON M5G1X5, Canada. [2]Section of Digestive Diseases, Department of Medicine and [3]Department of Genetics, Yale University, 333 Cedar Street, LMP1080, New Haven, Connecticut 06520, USA. [4]Université de Montréal and the Montreal Heart Institute, Research Center, 5000 rue Belanger, Montréal, QC H1T1C8, Canada. [5]Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048, USA. [6]Department of Statistics, Carnegie Mellon University, Baker/Porter Hall A60K, Pittsburgh, Pennsylvania 15213, USA. [7]IRCCS-CSS Hospital, Viale Cappuccini, 1, S. Giovanni Rotondo, Fg 71013, Italy. [8]Digestive Disease Institute, Cleveland Clinic, 9500 Euclid Avenue, Desk A31, Cleveland, Ohio 44195, USA. [9]Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, UPMC-PUH, Mezzanine Level, C-Wing, 200 Lothrop Street, Pittsburgh, Pennsylvania 15213, USA. [10]Princess Margaret Hospital, Department of Public Health Sciences, University of Toronto, 610 University Avenue, Toronto, ON M5G2M9, Canada. [11]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, 130 Desoto Street, Pittsburgh, Pennsylvania 15261, USA. [12]University of Pittsburgh Medical Center, Department of Psychiatry, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213, USA. [13]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [14]Johns Hopkins University School of Medicine, Department of Medicine, Harvey M. and Lyn P. Meyerhoff Infammatory Bowel Disease Center, 600 N. Wolfe Street, Baltimore, Maryland 21287, USA. [15]The Hospital for Sick Children, Department of Pediatrics, 555 University Avenue, Toronto, ON M5G1X8, Canada. [16]Department of Medicine, Université de Montréal and CHUM–Hôpital Saint-Luc, 1058, rue Saint-Denis, Montréal, QC H2X3J4, Canada. [17]Division of Gastroenterology, Department of Medicine, Laval University, Quebec City, CHAUQ–Hôpital du St-Sacrement, 1050, Chemin Ste–Foy, QC G1S4L8, Canada. [18]Department of Health Studies, University of Chicago, 5841 S. Maryland Avenue, Chicago, Illinois 60637, USA. [19]The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, New York 11030, USA. [20]Medical Genetics Institute, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048, USA. [21]Johns Hopkins University Bloomberg School of Public Health, Department of Epidemiology, 1501 E. Jefferson Street, B136, Baltimore, Maryland 21231, USA. [22]These authors contributed equally to this work. Correspondence should be addressed to R.D. (duerr@pitt.edu).

Received 30 May 2008; accepted 14 October 2008; published online 4 January 2009; corrected after print 28 April 2009; doi:10.1038/ng.275

© 2009 Nature America, Inc. All rights reserved.

Table 1 UC GWAS, replication and combined association evidence for four regions that had combined genome-wide significant association with UC

Locus					UC GWAS sample				Replication samples						Combined	
Chr	SNP	Position	Local loci	Minor allele	MAF cases	MAF controls	UC GWAS GEM *P*	UC GWAS GEM OR	North American cases MAF	North American controls MAF	Italian cases MAF	Italian controls MAF	Replication CMH *P*	Replication CMH OR	*P*	OR
1	rs3806308	20015453		A	0.30	0.37	4.7×10^{-8}	0.70	0.33	0.38	0.26	0.29	6.2×10^{-3}	0.85	6.7×10^{-9}	0.78
1	rs10753575	20036455		G	0.31	0.39	2.0×10^{-9}	0.68	0.34	0.39	0.31	0.34	1.7×10^{-3}	0.83	9.4×10^{-11}	0.76
1	rs6426833	20044447		G	0.37	0.46	6.8×10^{-10}	0.68	0.39	0.46	0.33	0.38	2.3×10^{-5}	0.77	5.1×10^{-13}	0.73
1	rs1004819	67442801	*IL23R*	A	0.35	0.30	6.1×10^{-5}	1.30	0.34	0.30	0.39	0.35	1.4×10^{-3}	1.22	1.5×10^{-6}	1.25
1	rs11465804	67475114	*IL23R*	C	0.039	0.071	3.9×10^{-4}	0.60	0.038	0.066	0.045	0.062	1.4×10^{-4}	0.61	9.4×10^{-7}	0.61
1	rs11209026	67478546	*IL23R*	A	0.037	0.070	3.1×10^{-5}	0.55	0.036	0.067	0.040	0.060	1.9×10^{-5}	0.57	1.3×10^{-8}	0.56
1	rs10889677	67497708	*IL23R*	A	0.36	0.30	5.7×10^{-7}	1.38	0.36	0.31	0.38	0.35	1.0×10^{-3}	1.22	1.3×10^{-8}	1.29
6	rs2395185	32541145		A	0.24	0.33	1.4×10^{-6}	0.72	0.21	0.32	0.17	0.23	3.6×10^{-13}	0.61	1.0×10^{-16}	0.66
12	rs7134599[a]	66786342		A	0.44	0.36	5.4×10^{-7}	1.55	0.42	0.38	0.41	0.35	4.8×10^{-4}	1.23	6.0×10^{-9}	1.32
12	rs1558744	66790859		A	0.48	0.38	5.5×10^{-10}	1.47	0.45	0.41	0.45	0.37	1.5×10^{-4}	1.26	2.5×10^{-12}	1.35
12	rs2870946[a]	66882928	*IL26*	G	0.10	0.055	1.0×10^{-5}	2.01	0.079	0.062	0.12	0.085	2.6×10^{-3}	1.37	4.8×10^{-7}	1.54

[a]SNP from HumanHap550 only SNPs. Chr, chromosome. MAF, minor allele frequency.

association with *IL10* and other suggestive associations[11]. There are many reports of UC association with HLA class II alleles[12] and major histocompatibility complex (MHC) SNPs[8,10,11].

We conducted a GWAS that included 1,052 individuals with UC extending beyond the rectum and 2,571 controls, all of white, non-Hispanic, European ancestry. Principal components of ancestry analysis demonstrated population structure (**Supplementary Fig. 1** online). We used genetic matching (GEM) within gender to control for structure and potential effects of gender, followed by allelic association tests by conditional logistic regression on gender-ancestry strata[13] (**Supplementary Table 1** online). GEM reduces the false-positive rate attributable to ancestry mismatches between cases and controls[13]. The genomic control inflation factor[14] in the combined GEM results was 1.04, compared to 1.17 for pre-GEM traditional chi-squared tests for allelic association. Quantile-quantile plots showed deviation from expectation at a substantially higher chi-squared value in the GEM dataset (**Supplementary Fig. 2b** online) compared to the pre-GEM dataset (**Supplementary Fig. 2a**). A second stage of our study aimed to replicate association in two regions that showed genome-wide significance and to extend the association evidence at other independent loci with GWAS *P*-values less than 1×10^{-4} (**Supplementary Table 2** online). We chose $P < 1 \times 10^{-4}$ as the threshold for our follow-up studies because the corresponding chi-squared value (15.136) is just above the point where the quantile-quantile plot of GEM results began to deviate from expectation (**Supplementary Fig. 2b**). Two independent replication samples included 768 UC cases and 721 controls from North America, all of white, non-Jewish, non-Hispanic, European ancestry, and 619 UC cases and 394 controls from southern Italy after quality control (**Supplementary Table 1**).

Table 1 and **Figure 1** show results for four regions that had combined genome-wide significant evidence for association with UC in the UC GWAS and replication samples. We observed genome-wide significant association in the GWAS, with subsequent replication, on chromosome 1p36 (**Table 1** and **Fig. 1a**) at the correlated ($r^2 = 0.76$) markers rs6426833 (GWAS $P = 6.8 \times 10^{-10}$, GWAS and replication combined $P = 5.1 \times 10^{-13}$) and rs10753575 (GWAS $P = 2.0 \times 10^{-9}$, GWAS and replication combined $P = 9.4 \times 10^{-11}$). We observed a residual association signal for rs6426833 but not rs10753575 when these SNPs were paired in a conditional analysis of their GWAS data (**Supplementary Table 3** online). A third SNP in this region, rs3806308, also showed genome-wide significant association in the GWAS with subsequent replication (GWAS $P = 4.7 \times 10^{-8}$, GWAS and replication combined $P = 6.7 \times 10^{-9}$, **Table 1** and **Fig. 1a**); rs3806308 was not correlated with either rs6426833 or rs10753575 ($r^2 < 0.01$), and we observed residual association signals when rs3806308 and each of the other two SNPs were paired in conditional analyses of their GWAS data (**Supplementary Table 3**). Therefore, the chromosome 1p36 locus had at least two independent association signals (rs6426833 and rs3806308) separated by recombination hotspots (**Fig. 1a**). There was also significant transmission/disequilibrium test (TDT)[15] evidence for undertransmission of the rs6426833 G allele (transmitted:untransmitted = 197:239, $P = 4.4 \times 10^{-2}$) in 436 trios of UC GWAS or replication cases and their parents.

On chromosome 12q15 (**Table 1** and **Fig. 1d**), we observed genome-wide significant evidence for association and subsequent replication at rs1558744 (GWAS $P = 5.5 \times 10^{-10}$, GWAS and replication combined $P = 2.5 \times 10^{-12}$). rs7134599, which was correlated with rs1558744 ($r^2 = 0.92$), showed association evidence equivalent to the evidence for rs1558744 in a conditional analysis of the GWAS data (**Supplementary Table 3**) and genome-wide significant combined evidence for association in the GWAS and replication

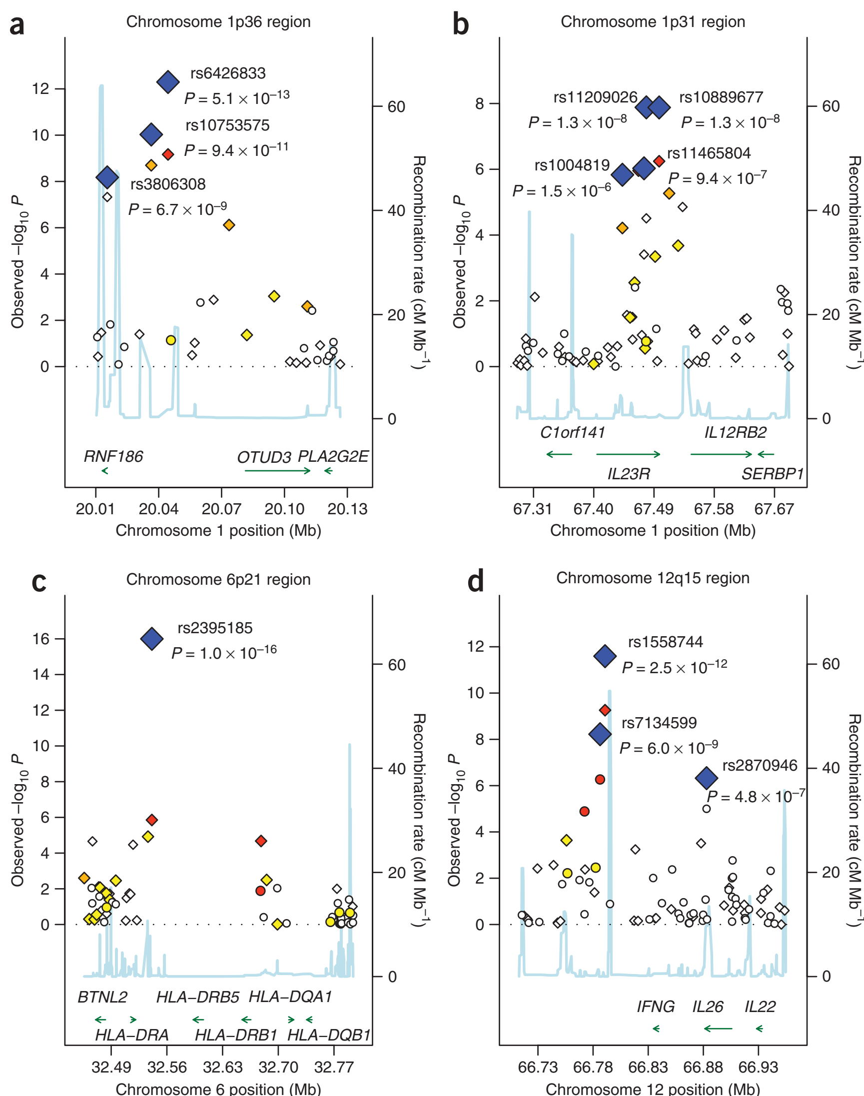

Figure 1 Association signals and recombination rates for loci showing genome-wide significant association with UC. (**a**–**d**) Associations at chromosomes 1p36 (**a**), 1p31 (**b**), 6p21 (**c**) and 12q15 (**d**). The vertical axis scale on the left side of each plot represents the $-\log_{10}$ of the *P*-values. The vertical axis scale on the right side of each plot represents the recombination rate in cM Mb^{-1}. The UC GWAS $-\log_{10}$ GEM *P*-values are plotted as small diamonds (all arrays shared SNPs GEM data set; **Supplementary Table 1**) and small circles (HumanHap550 only SNPs GEM data set; **Supplementary Table 1**). Linkage disequilibrium (r^2) in the GWAS control data to the single most significantly associated regional SNP is color-coded (red, $r^2 > 0.8$; orange, r^2 0.5–0.8; yellow, r^2 0.2–0.5; white, $r^2 < 0.2$). Large blue diamonds with corresponding rs numbers and *P*-value labels represent combined UC GWAS and replication case–control significance estimates. Light blue lines represent the recombination rates. Conditional analyses (**Supplementary Table 3**) of replicated markers (large blue diamonds) suggest the presence of independent association signals on chromosome 1p36 and on chromosome 12q15 separated by recombination hotspots.

samples (GWAS $P = 5.4 \times 10^{-7}$, GWAS and replication combined $P = 6.0 \times 10^{-9}$). rs2870946, which was only weakly correlated with rs1558744 and rs7134599 ($r^2 = 0.031$ and 0.029, respectively), showed suggestive combined evidence for association in the GWAS and replication samples that was close to the threshold for genome-wide significance (GWAS $P = 1.0 \times 10^{-5}$, GWAS and replication combined $P = 4.8 \times 10^{-7}$), and we observed residual association signals when rs2870946 and each of the other two SNPs were paired in conditional analyses of the GWAS data (**Supplementary Table 3**). Therefore, the chromosome 12q15 locus seemed to have two independent association signals (rs1558744 and rs2870946) separated by recombination hotspots (**Fig. 1d**). The rs1558744 A allele also showed TDT[15] evidence for association with UC (transmitted:untransmitted = 238:180, $P = 4.6 \times 10^{-3}$).

The GWAS and replication samples showed combined genome-wide significant evidence for association at two additional loci that have been implicated in UC previously, namely *IL23R* (interleukin-23 receptor) on chromosome 1p31 (ref. 2; **Table 1** and **Fig. 1b**) and the MHC on chromosome 6p21 (refs. 8,10–12; **Supplementary Table 2**, **Table 1** and **Fig. 1c**). Conditional analyses of the UC GWAS data suggested that there are at least two independent association signals in the *IL23R* region, consistent with our earlier study that identified the CD- and UC-associated *IL23R* locus[2], and multiple independent association signals in the MHC (**Supplementary Table 3**). The most significantly associated SNPs at the *IL23R* and MHC loci also showed TDT[15] evidence for association with UC (rs11209026 A allele, transmitted:untransmitted = 25:63, $P = 5.1 \times 10^{-5}$; rs10889677 A allele, transmitted:untransmitted = 223:150, $P = 1.6 \times 10^{-4}$; rs2395185 A allele, transmitted:untransmitted = 121:224, $P = 2.9 \times 10^{-8}$). See **Supplementary Note** online for further discussion about the association evidence at rs2395185.

OTUD3 (OTU domain containing 3) and part of *PLA2G2E* (phospholipase A2, group IIE) were located within an approximately 100 kb region containing the most significant chromosome 1p36 association signal (rs6426833, **Fig. 1a**). *OTUD3* is expressed broadly and has homology to an OTU-like cysteine protease[16]. PLA2G2E is a member of the secretory phospholipase A2 family of proteins that release arachidonic acid from membrane phospholipids, which leads to the production of proinflammatory lipid mediators, such as prostaglandins and leukotrienes[17]. Furthermore, *PLA2G2E* expression in the lung and small intestine is induced with lipopolysaccharide stimulation, suggesting a role in bacterially associated inflammation[17,18]. The second independent association signal on chromosome 1p36 (rs3806308, **Fig. 1a**) was located within 1.1 kb of *RNF186* (ring finger protein 186). Ring finger proteins are involved in ubiquitination of proteins and in diverse cellular processes[19].

The most significant chromosome 12q15 association signal (rs1558744, **Fig. 1d**) was located in a region devoid of established coding genes, but the *IFNG* (interferon-γ; IFN-γ), *IL26* (interleukin-26; IL-26) and *IL22* (interleukin-22) genes are located 44 kb, 91 kb and 137 kb, respectively, telomeric to rs1558744. A second, highly suggestive association signal at rs2870946 was independent of rs1558744 and located in *IL26* (**Fig. 1d**). IFN-γ is critical in the immune response to

© 2009 Nature America, Inc. All rights reserved.

pathogens, in part through regulation of macrophage function; it regulates many levels of immune homeostasis, including T-cell subsets, NK cells and NK T cells[20]. IL-22 and IL-26 are secreted by T_H-17 cells, which mediate host defense against infections as well as tissue inflammation in many chronic, immune-mediated diseases, including IBD[21]. Besides a role in mediating host defense to bacterial pathogens[21], IL-22 can mediate protection during acute inflammation, including intestinal inflammation[22].

The chromosome 6p21 association signal mapped to a broad region (**Supplementary Table 2**) that spans many genes in the MHC and shows extensive linkage disequilibrium. However, the maximal association signal at rs2395185 was located in an approximately 300-kb interval spanning the *BTNL2* (butyrophilin-like 2), *HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* genes (**Fig. 1c**). This region was also implicated in other studies[8,10–12]. rs9268877 showed the most significant evidence for association with UC in a previous study[11] and is correlated ($r^2 = 0.61$) in HapMap data from the CEPH collection of samples from Utah residents with ancestry from northern and western Europe (CEU data)[23] with rs6903608, which was the second-most-significantly associated MHC SNP in our UC GWAS (**Supplementary Table 2**). rs2516049, which was correlated with rs2395185 ($r^2 = 0.80$) and showed association evidence equivalent to that for rs2395185 in our UC GWAS data (**Supplementary Table 3**), showed association that was nearly as significant as the evidence for the maximally associated MHC SNP in another study[10]. The rs2395185 risk allele was associated with increased average expression of *HLA-DRB1* and *HLA-DQA1* in a lymphoblastoid cell line eQTL analysis[24].

Of the remaining loci tested in our replication samples (**Supplementary Table 2**), a cluster of SNPs on chromosome 7q31 flanked on one side by *SLC26A3* (solute carrier family 26, member 3) or located within *DLD* (dihydrolipoamide dehydrogenase) or *LAMB1* (laminin-β1) on the other side showed promising extension of the evidence for association (rs4730276, combined $P = 9.0 \times 10^{-6}$; rs4598195, combined $P = 9.6 \times 10^{-7}$; rs2158836, combined $P = 6.7 \times 10^{-6}$). *LAMB1* is a plausible functional candidate because laminins play a role in intestinal health and disease[25].

Finally, we determined whether CD loci implicated in a recent meta-analysis and CD or UC loci reported in other studies showed evidence for association in our UC GWAS (**Supplementary Table 4** online). When we assessed the best UC GWAS proxy for each previously implicated SNP, we observed only nominally ($P < 0.05$) significant evidence (**Supplementary Table 4a**) or no significant evidence (**Supplementary Table 4b**) for association with these loci in our UC GWAS and no significant evidence for association for some of these loci that we tested in the replication samples (**Supplementary Table 4a**,**b**).

In this study, we identified UC loci on chromosomes 1p36 and 12q15 where genes involved in inflammation and immunity include *PLA2G2E*, *IFNG*, *IL26* and *IL22*. We also confirmed association between UC and the MHC region, where associated SNPs are correlated with altered HLA class II expression, and between UC and *IL23R*. This report extends the model suggesting that there exist both unique and shared genetic factors for UC and CD. It is highly likely that more UC loci remain to be found because our study had limited power to detect small effect sizes, especially for lower risk-allele frequencies (**Supplementary Table 5** online)[26]. The experience in CD and other complex traits is that individual studies with sample sizes similar to the present study sample are sufficiently powered to detect risk loci with larger effects but find only a fraction of loci with smaller effects; loci with smaller effects can be detected through meta-analysis of the individual studies[7]. The loci identified through GWAS may eventually establish new therapeutic targets for the treatment of IBD.

METHODS

UC GWAS study subjects, genotyping and analysis. We genotyped 1,052 white, non-Hispanic, European ancestry patients with UC extending beyond the rectum and 2,571 white, European ancestry controls using HumanHap300v1, HumanHap300v2, HumanHap550v1 or HumanHap550v3 Genotyping BeadChips (Illumina) at the Feinstein Institute for Medical Research. Control data were obtained from our CD GWAS[2,4] or from studies 64 and 65 deposited in the Illumina iControlDB (see URLs below). Informed consent was obtained using protocols approved by each local institutional review board. More details about study subjects and the Illumina Genotyping BeadChips used to genotype groups of study subjects are available in the **Supplementary Methods** and **Supplementary Table 1** online.

Data management and quality control filtering are described in the **Supplementary Methods** online. We matched cases to controls by gender and ancestry and then performed association analyses by conditional logistic regression on gender-ancestry strata in the matched data sets (**Supplementary Table 1**) using GEM[13]. The GEM analysis is described in more detail in the **Supplementary Methods**. We also computed the genomic control inflation factor[14]. SNPs with nominal P-values $< 5 \times 10^{-8}$ were considered to have genome-wide significant evidence for association[27].

Replication study subjects, genotyping and analysis. We included two independent replication samples in the second stage of our study (**Supplementary Table 1**). A North American replication sample consisted of 769 patients with UC and 727 controls, all of white, non-Jewish, non-Hispanic, European ancestry. The second replication sample consisted of 633 subjects with UC and 415 controls from southern Italy.

We included SNPs with UC GWAS P-values $< 1 \times 10^{-4}$ on a list of replication genotyping candidates, assigning priority scores for inclusion in the replication genotyping according to significance of the association evidence. Among correlated ($r^2 > 0.5$) SNPs with UC GWAS P-values $<1 \times 10^{-4}$, we chose only the SNP with the most significant P-value as the 'best region' SNP and placed it on the replication candidate list. Other replication genotyping candidates included (i) SNPs at previously implicated CD or UC loci or their best UC GWAS proxies ($r^2 > 0.5$ in a 1-Mb window of HapMap CEU data[23] centered on the implicated SNP) that showed nominal evidence for association ($P < 0.05$) in our UC GWAS and (ii) previously implicated SNPs that did not have a proxy in our UC GWAS. Finally, we added to the replication candidate list more SNPs from loci that showed genome-wide significant evidence for association. rs6426833, rs2395185 and rs1558744 were genotyped by Pyrosequencing (Biotage) at the University of Pittsburgh. SNPs on the replication candidate list that had designable Sequenom iPlex (Sequenom) assays and could be multiplexed in three oligonucleotide pools were genotyped on the Sequenom MassArray platform at the University of Pittsburgh (North American samples and some of the Italian sample genotyping) and the Montreal Heart Institute (Italian samples). Samples with high genotyping failure rates were attempted a second time. Data management and quality control filtering are described in the **Supplementary Methods**. Association evidence in the two replication samples was assessed using the Cochran-Mantel-Haenszel (CMH) test, and the Fisher method was used to combine UC GWAS GEM P-values and replication CMH P-values. SNPs with combined P-values $< 5 \times 10^{-8}$ were considered to have genome-wide significant evidence for association[27]. SNPs with genome-wide significant evidence for association in the UC GWAS and CMH P-values < 0.05 in the replication samples were considered to have confirmed association with UC.

We also genotyped parents of UC GWAS and replication sample cases when DNA samples were available and analyzed affected offspring trios using the transmission/disequilibrium test[15] (**Supplementary Table 1**).

Conditional analysis. We were interested in whether one or more causal SNPs might explain the observed association in regions showing genome-wide significant association with UC. To explore this question, we fit a conditional logistic model, with pairs of SNPs in the model, conditional on ancestry and gender. P-values for the SNPs are interpretable as the residual variation

© 2009 Nature America, Inc. All rights reserved.
npg

explained by the SNP, conditional on the inclusion of the other SNP. Significant residual association signal was defined as $P < 0.05$ in the conditional analysis. We also computed pairwise r^2 values in our GWAS control data for the same pairs of SNPs.

Demarcation of regions containing genome-wide significant association signals. We demarcated regions containing genome-wide significant association signals by first defining the set of SNPs with $r^2 > 0.5$ to all independent SNPs showing genome-wide significant association using a 1-Mb window of HapMap CEU data[23] centered on the index SNP. SNPs were considered to be independent if they had a $r^2 < 0.3$ in our GWAS control data. Recombination hotspots[28] flanking these sets of SNPs were set as the boundaries of regions containing genome-wide significant association signals.

Power analysis. We computed a *post hoc* power analysis for a two-stage study design with parameters similar to our study design using the CaTS software[26] (**Supplementary Table 5**).

URLs. Illumina iControlDB, http://www.illumina.com/pages.ilmn?ID=231.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
Thanks to L. Wu Datta, J. Fultz and J. Stempak for coordinating study subject recruitment, to A. Andriulli and O. Palmieri for providing clinical data and to J. Lian, A. Liew and H. Khalili for technical support. The NIDDK IBD Genetics Consortium is funded by the following grants: DK062431 (S.R.B.), DK062422 (J.H.C.), DK062420 (R.H.D.), DK062432 (J.D.R.), DK062423 (M.S.S.), DK062413 (K.D.T.) and DK062429 (J.H.C.). The authors would also like to acknowledge additional support from the Atran Foundation (S.R.B.), Board of Governor's Chair in Medical Genetics at Cedars-Sinai Medical Center (J.I.R.), Bohmfalk Funds for Medical Research (J.H.C.), Burroughs Wellcome Medical Foundation (J.H.C.), Crohn's and Colitis Foundation of America (C.A., T.M.B., S.R.B., J.H.C., R.H.D., J.D.R.), Crohn's and Colitis Foundation of Canada (M.S.S.), Feintech Chair in Immunobiology (S.R.T.), Gale and Graham Wright Research Chair in Digestive Diseases (M.S.S.), Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center (T.M.B., S.R.B.), US National Institutes of Health grants RR00052 (S.R.B.), DK077905 (C.A.), DK068112 (J.-P.A.), DK072373 (J.H.C.), RR024139 (J.H.C.), DK076025 (R.H.D.), DK064869 (J.D.R.), MH057881 (K.R.), DK046763 (J.I.R., S.R.T., K.D.T.) and RR00425 (K.D.T.), the Rainin IBD Genetics Research Fund (J.-P.A.) and the W. Buford Lewis family (S.R.B.).

AUTHOR CONTRIBUTIONS
J.H.C., E.O.K. and L.P.S. developed and maintained the US National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) IBD Genetics Consortium Data Coordinating Center infrastructure. C.A., J.-P.A., V.A., T.M.B., F.B., S.R.B., J.H.C., R.H.D., A.M.G., A.F.I., R.G.L., A.L., D.P.B.M., P.P., D.D.P., M.D.R., J.D.R., J.I.R., R.S., M.S.S., A.H.S., S.R.T. and K.D.T. provided clinical samples and information. S.R.B., J.H.C., M.J.D., R.H.D., P.K.G., A.T.L., J.D.R., J.I.R., M.S.S. and K.D.T. designed the GWAS. M.M.B. and R.H.D. performed quality control and preliminary association analyses of the GWAS data. J.W. performed the GEM, quantile-quantile and conditional analyses of the GWAS data under the supervision of K.R., with contributions from L.K. M.J.D. defined the 'best region' SNPs among SNPs with GWAS $P < 0.0001$. W.X. identified the best GWAS proxies for CD and UC loci. R.H.D., P.G., J.D.R. and R.S. designed and performed the replication study. R.H.D. and P.G. analyzed the replication data. The manuscript was written by J.H.C., R.H.D., K.R. and M.S.S. with contributions from C.A., M.M.B., S.R.B., P.K.G., D.P.B.M., J.D.R., J.I.R., R.S. and J.W. R.H.D. coordinated the genotyping, analysis and manuscript writing efforts of this multicenter study.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Xavier, R.J. & Podolsky, D.K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434 (2007).
2. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
3. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3**, e58 (2007).
4. Rioux, J.D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* **39**, 596–604 (2007).
5. The Wellcome Trust Case Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
6. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* **39**, 830–832 (2007).
7. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
8. Fisher, S.A. *et al.* Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat. Genet.* **40**, 710–712 (2008).
9. Franke, A. *et al.* Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat. Genet.* **40**, 713–715 (2008).
10. Kugathasan, S. *et al.* Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat. Genet.* **40**, 1211–1215 (2008).
11. Franke, A. *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat. Genet.* **40**, 1319–1323 (2008).
12. Stokkers, P.C., Reitsma, P.H., Tytgat, G.N. & van Deventer, S.J. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. *Gut* **45**, 395–401 (1999).
13. Luca, D. *et al.* On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am. J. Hum. Genet.* **82**, 453–463 (2008).
14. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
15. Ewens, W.J. & Spielman, R.S. The transmission/disequilibrium test: history, subdivision, and admixture. *Am. J. Hum. Genet.* **57**, 455–464 (1995).
16. Makarova, K.S., Aravind, L. & Koonin, E.V. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and Chlamydia pneumoniae. *Trends Biochem. Sci.* **25**, 50–52 (2000).
17. Murakami, M. *et al.* Arachidonate release and eicosanoid generation by group IIE phospholipase A(2). *Biochem. Biophys. Res. Commun.* **292**, 689–696 (2002).
18. Suzuki, N. *et al.* Structures, enzymatic properties, and expression of novel human and mouse secretory phospholipase A(2)s. *J. Biol. Chem.* **275**, 5785–5793 (2000).
19. Joazeiro, C.A. & Weissman, A.M. RING finger proteins: mediators of ubiquitin ligase activity. *Cell* **102**, 549–552 (2000).
20. Schoenborn, J.R. & Wilson, C.B. Regulation of interferon-gamma during innate and adaptive immune responses. *Adv. Immunol.* **96**, 41–101 (2007).
21. Aujla, S.J., Dubin, P.J. & Kolls, J.K. Th17 cells and mucosal host defense. *Semin. Immunol.* **19**, 377–382 (2007).
22. Sugimoto, K. *et al.* IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *J. Clin. Invest.* **118**, 534–544 (2008).
23. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
24. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
25. Teller, I.C. & Beaulieu, J.F. Interactions between laminin and epithelial cells in intestinal health and disease. *Expert Rev. Mol. Med.* **3**, 1–18 (2001).
26. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
27. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
28. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).

© 2009 Nature America, Inc. All rights reserved.
npg

Corrigendum: Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study

Mark S Silverberg, Judy H Cho, John D Rioux, Dermot P B McGovern, Jing Wu, Vito Annese, Jean-Paul Achkar, Philippe Goyette, Regan Scott, Wei Xu, M Michael Barmada, Lambertus Klei, Mark J Daly, Clara Abraham, Theodore M Bayless, Fabrizio Bossa, Anne M Griffiths, Andrew F Ippoliti, Raymond G Lahaie, Anna Latiano, Pierre Pare, Deborah D Proctor, Miguel D Regueiro, A Hillary Steinhart, Stephan R Targan, L Philip Schumm, Emily O Kistner, Annette T Lee, Peter K Gregersen, Jerome I Rotter, Steven R Brant, Kent D Taylor, Kathryn Roeder & Richard H Duerr

***Nat. Genet.* 41, 216–220 (2009), published online 4 January 2009; corrected after print 28 April 2009**

In the first paragraph of the second column on the third page, rs11209026 A allele was incorrectly listed as rs111209026 A allele. The error has been corrected in the HTML and PDF versions of the article.

© 2009 Nature America, Inc. All rights reserved.

npg