Published in: Nature Genetics (2010)
Status: Postprint (Author's version)

Genome-wide meta-analysis increases to 71 the number of confirmed Crohns disease susceptibility loci

Andre Franke[1,70], Dermot P B McGovern[2,3,70], Jeffrey C Barrett[4,70], Kai Wang[5], Graham L Radford-Smith[6], Tariq Ahmad[7], Charlie W Lees[8], Tobias Balschun[9], James Lee[10], Rebecca Roberts[11], Carl A Anderson[4], Joshua C Bis[12], Suzanne Bumpstead[4], David Ellinghaus[1], Eleonora M Festen[13], Michel Georges[14], Todd Green[15], Talin Haritunians[3], Luke Jostins[4], Anna Latiano[16], Christopher G Mathew[17], Grant W Montgomery[18], Natalie J Prescott[17], Soumya Raychaudhuri[15], Jerome I Rotter[3], Philip Schumm[19], Yashoda Sharma[20], Lisa A Simms[6], Kent D Taylor[3], David Whiteman[18], Cisca Wijmenga[13], Robert N Baldassano[21], Murray Barclay[11], Theodore M Bayless[22], Stephan Brand[23], Carsten Buning[24], Albert Cohen[25], Jean-Frederick Colombel[26], Mario Cottone[27], Laura Stronati[28], Ted Denson[29], Martine De Vos[30], Renata D'Inca[31], Maria Dubinsky[32], Cathryn Edwards[33], Tim Florin[34], Denis Franchimont[35], Richard Gearry[11], Jürgen Glas[23,36,37], Andre Van Gossum[35], Stephen L Guthery[38], Jonas Halfvarson[39], Hein W Verspaget[40], Jean-Pierre Hugot[41], Amir Karban[42], Debby Laukens[30], Ian Lawrance[43], Marc Lemann[44], Arie Levine[45], Cecile Libioulle[46], Edouard Louis[46], Craig Mowat[47], William Newman[48], Julián Panés[49], Anne Phillips[47], Deborah D Proctor[20], Miguel Regueiro[50], Richard Russell[51], Paul Rutgeerts[52], Jeremy Sanderson[53], Miquel Sans[49], Frank Seibold[54], A Hillary Steinhart[55], Pieter C F Stokkers[56], Leif Torkvist[57], Gerd Kullak-Ublick[58], David Wilson[59], Thomas Walters[60], Stephan R Targan[2], Steven R Brant[22], John D Rioux[61], Mauro DAmato[62], Rinse K Weersma[63], Subra Kugathasan[64], Anne M Griffiths[60], John C Mansfield[65], Severine Vermeire[52], Richard H Duerr[50,66], Mark S Silverberg[55], Jack Satsangi[8], Stefan Schreiber[1,67], Judy H Cho[20,68], Vito Annese[16,69], Hakon Hakonarson[5,21], Mark J Daly[15,71] & Miles Parkes[10,71]

[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany. [2]Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [3]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [4]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [5]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [6]Inflammatory Bowel Disease Research Group, Queensland Institute of Medical Research, Brisbane, Australia. [7]Peninsula College of Medicine and Dentistry, Exeter, UK. [8]Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Edinburgh, UK. [9]PopGen Biobank, Christian-Albrechts University Kiel, Kiel, Germany. [10]Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. [11]Department of Medicine, University of Otago, Christchurch, New Zealand. [12]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. [13]Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands. [14]Department of Genetics, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium. [15]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [16]Unit of Gastroenterology, IRCCS-CSS Hospital, San Giovanni Rotondo, Italy [17]Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, UK. [18]Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia. [19]Department of Health Studies, University of Chicago, Chicago, Illinois, USA. [20]Section of Digestive Diseases, Department of Medicine, Yale University, New Haven, Connecticut, USA. [21]Department of Pediatrics, Center for Pediatric Inflammatory Bowel Disease, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [22]lnflammatory Bowel Disease Center, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [23]Department of Medicine II, University Hospital Munich Grosshadern, Ludwig-Maximilians-University, Munich, Germany. [24]Department of Gastroenterology, Charité, Campus Mitte, Universitätsmedizin Berlin, Berlin, Germany. [25]Montreal Jewish General Hospital, Montreal, Québec, Canada. [26]Registre EPIMAD, Université de Lille, Lille, France [27]Unit of Gastroenterology, Cervello Hospital, Palermo, Italy. [28]ENEA, Department of Biology of Radiations and Human Health, Rome, Italy. [29]Pediatric Gastroenterology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA. [30]Department of Hepatology and Gastroenterology, Ghent University Hospital, Ghent, Belgium. [31]Division of Gastroenterology, University Hospital Padua, Padua, Italy. [32]Department of Pediatrics, Cedars Sinai Medical Center, Los Angeles, California, USA. [33]Torbay Hospital, Torbay, Devon, UK. [34]Department of Gastroenterology, Mater Health Services, Brisbane, Australia. [35]Department of Gastroenterology, Erasmus Hospital, Free University of Brussels, Brussels, Belgium. [36]Department of Preventive Dentistry and Periodontology, Ludwig-Maximilians-University, Munich, Germany. [37]Department of Human Genetics, RWTH Aachen, Germany. [38]Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah, USA. [39]Department of Medicine, Orebro University Hospital, Orebro, Sweden. [40]Department of Gastroenterology, Leiden University Medical Center, Leiden, The Netherlands. [41]Université Paris Diderot, Paris, France. [42]Department of Gastroenterology, Faculty of Medicine, Technion-lsrael Institute of Technology, Haifa, Israel. [43]School of Medicine and Pharmacology, The University of Western Australia, Fremantle, Australia. [44]GETAID group, Université Paris Diderot, Paris, France [45]Pediatric Gastroenterology Unit, Wolfson Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel. [46]Division of Gastroenterology, CHU, Université de Liège, Liège, Belgium. [47]Department of Medicine, Ninewells Hospital and Medical School, Dundee, UK. [48]Department of Medical Genetics, University of Manchester, Manchester, UK. [49]Department of Gastroenterology, Hospital Clfπic/IDIBAPS, CIBER EHD, Barcelona, Spain [50]Division of Gastroenterology, Hepatology and Nutrition, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [51]Consultant Paediatric Gastroenterologist, Yorkhill Hospital, Glasgow, UK. [52]Division of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium. [53]Department of Gastroenterology, Guy's and St Thomas' National Health Service Foundation Trust, St Thomas' Hospital, London, UK. [54]Division of Gastroenterology, Inselspital, University of Bern, Bern, Switzerland. [55]Mount Sinai Hospital Inflammatory Bowel Disease Centre, University of Toronto, Toronto, Ontario, Canada. [56]Department of Gastroenterology, Academic Medical Center, Amsterdam, The Netherlands. [57]Department of Clinical Science Intervention and Technology, Karolinska Institutet, Stockholm, Sweden. [58]Division of Clinical Pharmacology and Toxicology, University Hospital Zurich, Zurich, Switzerland. [59]Child Life and Health, University of Edinburgh, Edinburgh, UK. [60]The Hospital for Sick Children, University of Toronto, Ontario, Canada. [61]Université de Montréal and the Montreal Heart Institute, Research Center, Montréal, Québec, Canada. [62]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. [63]Department of Gastroenterology, University Medical Center Groningen, Groningen, The Netherlands. [64]Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, USA. [65]lnstitute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK. [66]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [67]Department for General Internal Medicine, Christian-

Published in: Nature Genetics (2010)
Status: Postprint (Author's version)

Albrechts-University, Kiel, Germany. [68]Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA. [69]Unit of Gastroenterology, University Hospital Careggi, Florence, Italy. [70]These authors contributed equally to this work. [71]These authors jointly directed this project.

Abstract

We undertook a meta-analysis of six Crohn's disease genome-wide association studies (GWAS) comprising 6,333 affected individuals (cases) and 15,056 controls and followed up the top association signals in 15,694 cases, 14,026 controls and 414 parent-offspring trios. We identified 30 new susceptibility loci meeting genome-wide significance ($P < 5 \times 10^{-8}$). A series of *in silico* analyses highlighted particular genes within these loci and, together with manual curation, implicated functionally interesting candidate genes including *SMAD3*, *ERAP2*, *IL10*, *IL2RA*, *TYK2*, *FUT2*, *DNMT3A*, *DENND1B*, *BACH2* and *TAGAP*. Combined with previously confirmed loci, these results identify 71 distinct loci with genome-wide significant evidence for association with Crohn's disease.

Crohn's disease (MIM#266600) results from the interaction of environmental factors, including intestinal microbiota, with host immune mechanisms in genetically susceptible individuals. Along with ulcerative colitis, it is one of the main subphenotypes of inflammatory bowel disease (IBD). Genome-wide association studies (GWAS) have highlighted key pathogenic mechanisms in Crohn's disease, including autophagy and Th17 pathways. A meta-analysis of these early scans implicated 32 susceptibility loci but only accounted for 20% of the genetic contribution to disease risk, suggesting that more loci await discovery[1]. Recognizing that an increased sample size would be required to detect these, we expanded the International IBD Genetics Consortium (IIBDGC), approximately doubling the discovery panel size in comparison with the first meta-analysis.

The discovery panel for the current study comprised 6,333 individuals with Crohn's disease (cases) and 15,056 controls, all of European descent, with data derived from six index GWAS studies (for overview, see **Supplementary Table 1**)[2-6]. Imputation using Hap Map3 reference data allowed us to test for association at 953,241 autosomal SNPs. Our discovery panel had 80% power to detect variants conferring odds ratios >1.18 at the genome-wide significance level of $P < 5 \times 10^{-8}$, assuming a minor allele frequency of ≥20% in healthy controls. Underthe same conditions, the sample size of our original meta-analysis had only 11% power[1].

A quantile-quantile plot of the primary meta-statistic using single-SNP z-scores combined across all sample sets, showed a marked excess of significant associations (**Supplementary Fig. 1**). A total of 2,024 SNPs within 107 distinct genomic loci, including all previously defined significant hits from our earlier meta-analysis (**Table 1**), demonstrated association with $P < 10^{-5}$. A Manhattan plot is shown in **Supplementary Figure 2**. We followed up 51 of the regions, representing new loci associated at $P < 5 \times 10^{-6}$, by genotyping the most significant SNPs in an independent panel of 15,694 Crohn's disease cases, 14,026 controls and 414 parent-offspring trios (**Table 2** and **Supplementary Table 2**).

Variants within 30 distinct new loci met a genome-wide significance threshold of $P < 5 \times 10^{-8}$ for association with Crohn's disease in the combined discovery plus replication panel, with at least nominal association in the replication panel (**Table 2**). Two additional loci, encompassing *CARD9* and *IL18RAP*, had previously been reported as associated with Crohn's disease in a candidate gene study[7] and were here both replicated and confirmed at $P < 5 \times 10^{-8}$. Another five loci were identified at genome-wide significance in GWAS published subsequent to our replication experiment being designed. One, the *PUT2* locus, was found in a recent adult Crohn's disease GWAS[6]. Four more (in *ZMIZ1* and in *IL*27 at 16p11, 19q13 and 22q12) were identified in a pediatric IBD population[5]; these identifications were replicated here in our current sample set. Two further loci had produced suggestive evidence of association with replication in our earlier study[1]. Here, these loci clearly exceeded the genome-wide significance threshold in the meta-analysis alone and, given the previous replication evidence, were not followed up further. Thus, cumulatively, 39 loci can now be added to the 32 confirmed Crohn's disease susceptibility loci identified at the time of our original meta-analysis[1]. We did not observe statistically significant heterogeneity of the odds ratios (Breslow Day test $P < 0.05$ after Bonferroni correction; **Supplementary Table 3**) between the panels from our 15 different countries (**Supplementary Tables 1** and **2**) for any of the 71 loci, nor did we observe any evidence of interaction between the associated loci (**Supplementary Fig. 3**).

Published in: Nature Genetics (2010)
Status: Postprint (Author's version)

Table 1 *Association results and in silico analyses for 32 previously confirmed Crohn's disease risk loci*

dbSNP ID	Chr.	Left-right (Mb)	Risk allele	Allele frequency in controls	P_{meta}	OR (95% Cl)	Association reported with other phenotypes	Positional candidate genes of interest
rs11209026	1p31	67.13-67.54	G	0.932	1.00×10^{-64}	2.66 (2.36-3.00)	UC, AS, Ps, PBC, GC, BD	***IL23R***
rs2476601	1p13	113.66-114.42	G	0.907	4.47×10^{-9}	1.26 (1.17-1.37)	TlD[a], RA, SLE, Ps, Vitiligo[a], AITD	***PTPN22***
rs4656940	1q23	158.96-159.20	A	0.801	6.17×10^{-7}	1.15 (1.09-1.21)	SLE, RA	***CD244,*** *ITLN1*
rs7517810	1q24	170.92-171.21	T	0.246	1.51×10^{-15}	1.22 (1.16-1.28)	HepC, SLE, SSc, T2D	***TNFSF18, TNFSF4, FASLG***
rs7554511	1q32	199.11-199.32	C	0.726	1.58×10^{-7}	1.14 (1.08-1.19)	UC, celiac, MS	***C1orf106,*** *KIF21B*
rs3792109	2q37	233.81-234.23	A	0.529	6.76×10^{-41}	1.34 (1.29-1.40)	UC	***ATG16L1***
rs3197999	3p21	48.16-51.73	A	0.297	6.17×10^{-17}	1.22 (1.16-1.27)	UC	***MST1, GPX1, BSN***
rs11742570	5p13	39.88-41.00	C	0.606	7.08×10^{-36}	1.33 (1.27-1.39)	MS	***PTGER4***
rs12521868	5q31	129.41-132.05	T	0.422	1.41×10^{-20}	1.23 (1.18-1.28)	Ps, fibrinogen, asthma, TB, UC	***SLC22A4, SLC22A5, IRF1, IL3***
rs7714584	5q33	150.01-150.38	G	0.088	7.76×10^{-19}	1.37 (1.28-1.47)	TB	***IRGM***
rs6556412	5q33	158.43-158.88	A	0.332	5.37×10^{-14}	1.18 (1.13-1.24)	Ps, SLE, malaria, asthma	***IL12B***
rs6908425	6p22	20.60-21.25	C	0.784	1.41×10^{-8}	1.17 (1.11-1.23)	T2D, Ps, UC	*CDKAL1*
rs1799964	6p21	31.49-32.98	C	0.209	3.98×10^{-11}	1.19 (1.13-1.25)	Multiple including UC	***LTA,*** *HLA-DQA2,* ***TNF, LST1, LTB***
rs6568421	6q21	106.50-106.67	G	0.301	4.37×10^{-8}	1.13 (1.07-1.18)[b]	SLE, RA	***PRDM1***
rs415890	6q27	167.26-167.47	C	0.522	2.51×10^{-12}	1.17 (1.12-1.22)	RA, Graves	***CCR6***
rs1456896	7p12	50.22-50.34	T	0.69	1.20×10^{-8}	1.14 (1.09-1.20)	AD, SLE, MCV, ALL	***IKZF1, ZPBP, FIGNL1***
rs4871611	8q24	126.54-126.65	A	0.609	1.51×10^{-12}	1.17 (1.12-1.23)		
rs10758669	9p24	4.93-5.29	C	0.349	1.00×10^{-13}	1.18 (1.13-1.23)	UC, MPD	*JAK2*
rs3810936	9q32	116.47-116.74	C	0.682	1.00×10^{-15}	1.21 (1.15-1.27)	UC, leprosy, SpA	***TNFSF15, TNFSF8***
rs12242110	10p11	35.22-35.94	G	0.315	1.10×10^{-9}	1.15 (1.10-1.20)	UC	***CREM***
rs10761659	10q21	63.97-64.43	G	0.538	4.37×10^{-22}	1.23 (1.18-1.29)	BC	*ZNF365*
rs4409764	10q24	101.26-101.33	T	0.492	2.29×10^{-20}	1.22 (1.17-1.27)	UC	*NKX2-3*
rs7927997	11q13	75.70-76.04	T	0.389	5.62×10^{-13}	1.17 (1.12-1.22)	Atopy[a]	*C11orf30*
rs11564258	12q12	38.42-39.31	A	0.025	6.17×10^{-21}	1.74 (1.55-1.95)	PD, leprosy	***MUC19,*** *LRRK2*
rs3764147	13q14	43.13-43.54	G	0.245	1.41×10^{-10}	1.17 (1.12-1.23)	Leprosy	***C13orf31***
rs2076756	16q12	49.02-49.41	G	0.26	3.98×10^{-69}	1.53 (1.46-1.60)	Leprosy, atopy, Blau, GvHD	***N0D2***
rs2872507	17q21	34.62-35.51	A	0.458	1.51×10^{-9}	1.14 (1.09-1.19)	Asthma, UC, PBC, T1D, RA, WBC	***GSMDL, ZPBP2, ORMDL3, IKZF3***
rs11871801	17q21	37.57-38.25	A	0.756	2.51×10^{-8}	1.15 (1.10-1.21)	MS[a], obesity, HIES	***MLX,*** *STAT3*
rs1893217	18p11	12.73-12.92	G	0.153	1.29×10^{-14}	1.25 (1.18-1.32)	TlD[a], celiac	*PTPN2*
rs740495	19p13	1.04-1.13	G	0.247	8.13×10^{-12}	1.16 (1.10-1.21)		*GPX4, SBN02*
rs1736020	21q21	15.62-15.77	C	0.579	9.33×10^{-12}	1.16 (1.11-1.21)	UC	
rs2838519	21q22	44.41-44.52	G	0.391	2.09×10^{-14}	1.18 (1.13-1.23)	Celiac, UC	***ICOSLG***

The table reports new data for loci confirmed in the earlier meta-analysis[1]. Left-right association boundaries are given for each index SNP (NCBI's dbSNP build vl30; see Online Methods). Associations with other relevant traits were identified by a literature search using the US National Institutes of Health catalog of genome-wide association studies and the HUGE database (version 1.4)[43,44]. Candidate genes of interest are listed. Those in bold were highlighted by *in silico* analyses (GRAIL connectivity, presence of an eQTL effect with LOD ≥ 5.0 or implicated coding SNP; see main text and **Supplementary Table 6** for more details). Loci tagged by rs4656940 and rs7554511 previously replicated strongly (at 0.00048 and 2.3×10^{-6} replicated previously, respectively[1]) and still pass genome-wide significance on combined analysis. UC, ulcerative colitis; AS, ankylosing spondylitis; Ps, psoriasis; PBC, primary biliary cirrhosis; T1D, type 1 diabetes; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T2D, type 2 diabetes; MS, multiple sclerosis; AD, Alzheimer's disease; MCV, mean corpuscular volume; ALL, acute lymphocytic leukemia; SpA, spondyloarthritis; PD, Parkinson's disease; AITD, autoimmune thyroid disease; BC, breast cancer; BD, Behcet's disease; GC, gastric cancer; HepC, hepatitis C susceptibility; SSc, systemic sclerosis; MPD, myeloproliterative disease; TB, tuberculosis; GvHD, graft versus host disease; WBC, white blood cell count; HIES, hyper immunoglobulin E syndrome.

[a]Association in the opposite direction in different traits.

[b]Loci with more than one independent association

Published in: Nature Genetics (2010)
Status: Postprint (Author's version)

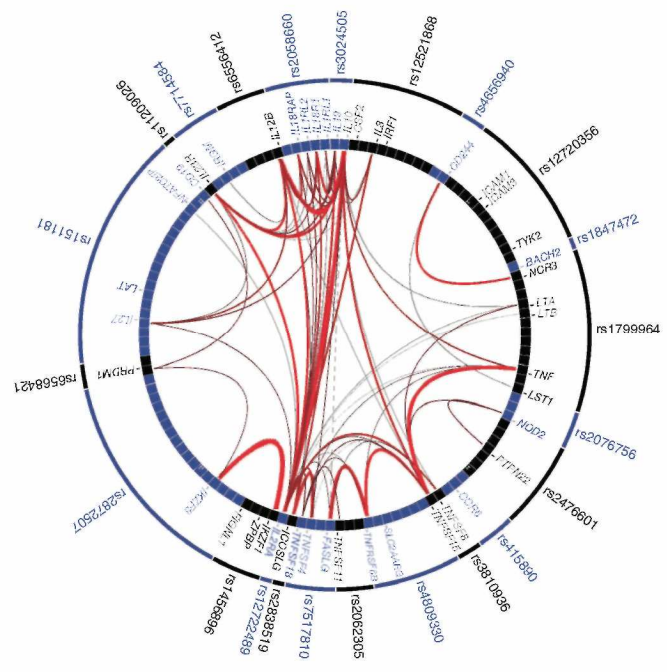
Table 2 *Association results and in silico analyses for 39 newly confirmed Crohn's disease risk loc*

Chr.	Left-right (Mb)	Risk allele	Allele frequency in controls	P_{meta}	$P_{repl.}$	$P_{comb.}$	OR (95% Cl)	Association reported with other phenotypes	Positional candidate genes of interest
1p36	7.66-7.89	A	0.19	2.69×10^{-10}	1.40×10^{-2}	7.10×10^{-9}	1.05 (1.01-1.10)	Celiac	*VAMP3*
1q22	153.24-154.39	A	0.25	1.29×10^{-9}	2.70×10^{-5}	2.30×10^{-13}	1.13 (1.06-1.19)[a]	T2D, asthma, PD	*SCAMP3, MUC1*
1q31	195.58-196.21	G	0.302	4.90×10^{-9}	1.60×10^{-2}	8.70×10^{-9}	1.04 (1.00-1.09)	Asthma	*DENND1B*
lq32	204.87-205.10	T	0.157	8.32×10^{-9}	1.50×10^{-7}	1.60×10^{-14}	1.12 (1.07-1.17)	T1D, UC, SLE, BD, HepC	***ILW, IL19***
2p23	25.30-25.46	G	0.326	1.41×10^{-8}	5.90×10^{-4}	8.50×10^{-10}	1.06 (1.03-1.10)		*DNMT3A*
2p23	27.24-27.71	T	0.418	1.10×10^{-4}	3.30×10^{-8}	4.70×10^{-11}	1.15 (1.10-1.21)	CRP, glucose, TGs	***GCKR***
2p21	43.30-43.80	T	0.129	7.70×10^{-8}	2.90×10^{-8}	1.60×10^{-14}	1.14 (1.09-1.20)[a]	T2D, PC	***THADA***
2p16[b]	60.77-61.74	T	0.42	6.61×10^{-9}	N/A	N/A	1.14 (1.09-1.19)	RA, UC, celiac	***C2orf74,*** *REL*
2q12[c]	102.17-102.67	G	0.231	1.58×10^{-12}	N/A	N/A	1.19 (1.14-1.26)	Celiac, asthma, T1D, HSV	***IL18RAP, IL12RL2, IL18R1, IL1RL1***
2q33	197.85-198.67	A	0.473	1.82×10^{-7}	1.60×10^{-3}	3.50×10^{-9}	1.06 (1.02-1.11)	CAD	***PLCL1***
2q37	230.76-230.94	T	0.187	4.57×10^{-9}	7.40×10^{-6}	3.10×10^{-13}	1.12 (1.07-1.18)	CLL	***SP140***
3p24	18.58-18.86	A	0.322	8.20×10^{-7}	1.00×10^{-3}	6.70×10^{-9}	1.08 (1.03-1.13)		
5q13	72.49-72.62	A	0.600	2.00×10^{-6}	6.40×10^{-7}	5.90×10^{-12}	1.12 (1.07-1.17)		
5q15	96.11-96.45	C	0.409	4.47×10^{-11}	2.00×10^{-3}	1.10×10^{-10}	1.05 (1.02-1.09)	AS, PD, T1D, PET	***ERAP2, LRAP***
5q31	141.39-141.62	C	0.796	1.10×10^{-9}	4.20×10^{-3}	2.00×10^{-9}	1.06 (1.02-1.11)		*NDFIP1*
5q35	173.15-173.47	T	0.571	5.25×10^{-8}	3.30×10^{-6}	2.50×10^{-12}	1.08 (1.04-1.12)		***CPEB4***
6p25	3.35-3.41	T	0.639	6.16×10^{-7}	3.10×10^{-4}	6.70×10^{-9}	1.10 (1.05-1.16)		
6q15	90.86-91.14	G	0.658	3.63×10^{-6}	1.40×10^{-4}	5.10×10^{-9}	1.07 (1.03-1.11)	T1D, celiac	***BACH2***
6q25	159.26-159.46	G	0.393	1.41×10^{-7}	2.40×10^{-5}	2.30×10^{-11}	1.10 (1.05-1.14)	RA, celiac, T1D[d]	*TAGAP*
8q24	129.56-129.67	T	0.865	2.29×10^{-6}	2.40×10^{-13}	3.90×10^{-18}	1.23 (1.17-1.30)		
9q34[c]	138.27-138.54	T	0.411	4.37×10^{-19}	1.50×10^{-19}	1.30×10^{-6}	1.18 (1.13-1.22)	UC, AS	***CARD9, SNAPC4***
10p15	6.07-6.21	C	0.852	8.51×10^{-6}	5.20×10^{-5}	2.90×10^{-9}	1.11 (1.05-1.16)	MS, T1D, vitiligo, RA, AA, asthma, AITD	***IL2RA***
10q21	59.50-59.81	C	0.774	1.41×10^{-7}	1.10×10^{-10}	9.10×10^{-17}	1.19 (1.13-1.25)	AD	*UBE2D1*
10q22[e]	80.67-80.77	G	0.669	2.00×10^{-10}	7.30×10^{-22}	1.10×10^{-30}	1.19 (1.15-1.23)	Celiac, MS, vitiligo, ESC	*ZMIZ1*
11q12	61.28-61.44	C	0.341	7.24×10^{-8}	1.70×10^{-5}	2.30×10^{-11}	1.08 (1.04-1.12)	CAD, dyslipidemia	***FADS1***
11q13	63.58-64.05	A	0.626	3.38×10^{-7}	3.50×10^{-4}	6.00×10^{-10}	1.10 (1.05-1.16)	AA	*PRDX5, ESRRA*
13q14	41.72-42.00	G	0.346	2.00×10^{-6}	5.70×10^{-5}	4.90×10^{-10}	1.10 (1.05-1.15)	BMD, RA	***TNFSF11***
14q24	68.23-68.39	G	0.584	2.00×10^{-7}	4.50×10^{-5}	1.60×10^{-10}	1.07 (1.11-1.04)[a]	Celiac, T1D	*ZFP36L1*
14q35	87.28-87.71	T	0.119	1.29×10^{-8}	5.90×10^{-11}	4.20×10^{-18}	1.23 (1.16-1.31)[a]		***GALC, GPR65***
15q22	65.20-65.27	T	0.233	1.41×10^{-13}	2.00×10^{-8}	2.70×10^{-19}	1.12 (1.07-1.16)	CAD, T2D	*SMAD3*
16p11[e]	28.20-28.94	G	0.386	1.10×10^{-10}	1.20×10^{-3}	1.50×10^{-11}	1.07 (1.03-1.12)	T1D, obesity, asthma, CRC, SLE, RA, IBD	***IL27, SH2B1, EIF3C, LAT, CD19***
17q12[b]	29.51-29.70	A	0.723	1.70×10^{-13}	N/A	N/A	1.20 (1.14-1.26)	HIV resistance	***CCL2, CCL7***
19p13	10.26-10.50	G	0.084	9.20×10^{-10}	1.90×10^{-5}	1.40×10^{-12}	1.12 (1.06-1.19)[a]	T1D, SLE, MS, HIES	***TYK2, ICAM1, ICAM3***
19q13[e]	38.42-38.47	T	0.612	2.69×10^{-7}	2.00×10^{-3}	8.70×10^{-9}	1.06 (1.02-1.11)		
19q13[e]	53.78-53.97	A	0.487	8.60×10^{-10}	5.20×10^{-5}	7.40×10^{-12}	1.07 (1.04-1.11)	B12, norovirus, HP	***FUT2, RASIP1***
20q13	61.65-61.95	G	0.709	2.51×10^{-12}	4.60×10^{-5}	2.70×10^{-15}	1.12 (1.06-1.18)	Glioma	***RTEL1, TNFRSF6B, SLC2A4RG***
22q11	20.14-20.39	T	0.203	6.31×10^{-13}	2.30×10^{-6}	4.80×10^{-16}	1.10 (1.06-1.15)	RA, celiac, SLE, MCV	***YDJC***
22q12[e]	28.23-29.00	C	0.471	5.70×10^{-9}	8.30×10^{-5}	7.30×10^{-12}	1.08 (1.04-1.13)	IBD, T1D	*MTMR3*
22q13	38.00-38.14	C	0.830	1.70×10^{-10}	9.50×10^{-18}	1.10×10^{-26}	1.23 (1.17-1.29)		*MAP3K7IP1*

The table lists Crohn's disease susceptibility loci newly identified since the first meta-analysis[1] with $P < 5 \times 10^{-8}$ in the combined analysis (discovery plus replication sample) and $P < 0.05$ in the replication stage. Seven of these loci have previously been reported (see below). Column headings and abbreviations are as described in **Table 1**. Additional abbreviations are as follows: CRC, colorectal cancer; CRP, C-reactive protein; TGs, triglycerides; PC, prostate cancer; HSV, human simplex virus; CAD, coronary artery disease; CLL, chronic lymphocytic leukemia; BMD, bone mineral density; B12, serum vitamin B12 levels; HP, *Helicobacter pylori;* AA, alopecia areata.
[a]Loci with more than one independent association. [b]Loci that previously showed suggestive association and replication[1] but not at genome-wide significance. [c]*IL18RAP* and *CARD9* associations were reported[7], but not previously at genome-wide significance, [d]Association in the opposite direction in different traits. [e]Loci previously reported at genome-wide significance in GWAS published subsequent to design of the current replication experiment[5,6].

Regional association plots of all 71 susceptibility loci, includingthe underlying genes, are shown in detail in **Supplementary Figure 4**, and complete genotype data, including odds ratios and allele frequencies, are shown in **Supplementary Tables 3** and **4**. Five loci showed evidence for more than one independently associated variant (**Table 2**). Although six of the 30 newly discovered regions contain just a single gene, which is thereby strongly implicated in Crohn's disease pathogenesis (for example, *SMAD3*, *NDFIP1* and *BACH2*), 22 of these regions include more than one gene within the associated interval (**Table 2**). We thus applied additional *in silico* analyses to refine the list of functional candidate genes further. These analyses were (i) interrogation of a publicly available expression quantitative trait loci (eQTL) database[8], which identified genes for which expression correlates with genotype at our most associated SNP (**Supplementary Note**), (ii) use of 1000 Genomes Project Pilot sequence data and HapMap3 data to identify genes containing non-synonymous variants in strong linkage disequilibrium ($r^2 > 0.5$) with the focal SNP within each region (for details on coding SNPs, see **Supplementary Table** 5) and (iii) use of GRAIL[9] to identify non-random and evidence-based connectivity between the genes in the 71 confirmed Crohn's disease loci. Specifically, GRAIL evaluates each gene within a Crohn's disease-associated locus for non-random correlation with genes in the other 70 loci through word-usage in PubMed abstracts related to the gene (**Fig. 1**).

Figure 1 *Gene relationships across implicated loci (GRAIL) pathway analysis. Links between genes at 23 of 71 Crohn's disease-associated oci which scored P < 0.01 using GRAIL.* Specifically, of the 71 Crohn's disease-associated SNPs, 69 are in linkage disequilibrium intervals containing or within 50 kb of at least one gene. In total, there were 355 genes implicated by proximity to these 69 SNPs. Each observed association was scored with GRAIL, which takes each gene mapping within Crohn's disease-associated intervals and evaluates for each whether it is non-randomly linked to the other genes through word usage in PubMed abstracts. The 23 SNPs shown in the outer circle are significant at $P < 0.01$, indicating that the regions which they tag contain genes which are more significantly linked to genes in the other 68 regions than expected by chance at that level. The lines between genes represent individually significant connections that contribute to the positive signal, with the thickness of the lines being inversely proportional to the probability that a literature-based connection would be seen by chance. To accurately assess the statistical significance of this set of connections, we conducted simulations in which we selected 1,000 sets of 69 SNPs implicating in total 355 genes ± 18 genes (5%) (selecting the SNPs randomly and using rejection sampling, only taking lists that implicated the same number of genes). Each of those 1,000 sets was scored with GRAIL. The mean number of $P < 0.01$ hits in a simulated list was 0.91, with a range in the 1,000 sets from 0 to 11, suggesting that the likelihood of observing 23 hits with $P < 0.01$ is far less than 0.1%.



Summary results of these analyses are shown in the rightmost column of **Table 1** and **Table 2** and in **Supplementary Table 6**. Highlighted genes are described briefly in **Box 1**, as are genes that constitute particularly noteworthy candidates from intervals containing one or few genes. Although we believe that these evidence-based approaches are helpful in identifying likely functional candidates, in some instances, the different techniques highlight different genes. This reflects uncertainty as to which genes are causal and emphasizes the need for functional studies.

We identified 30 new signals here beyond those described in the earlier meta-analysis[1] and other subsequent publications. The discovery of these new associations was driven primarily by increased power arising from the expanded sample size rather than improved imputation, as more than two-thirds of the newly discovered loci have good proxies ($r^2 > 0.8$) on both generation arrays used in the earlier studies (Illumina 300K and Affymetrix 500K). Extending this argument beyond the current analysis, it seems likely that many more loci of modest effect size still await discovery.

For many of the loci, associations have been reported previously in other complex diseases, comprising mostly chronic inflammatory disorders (**Tables 1** and **2**). Such diseases can cluster both within families and within individuals, reflecting shared genetic risk factors. For example, IBD and ankylosing spondylitis can co-segregate and both are associated with *IL23R* (refs. 2,10) and *TNFSF15* (refs. 11,12). The *IL10* locus was previously associated with ulcerative colitis[13] and was identified as a new Crohn's disease risk locus here. Thus, *IL10* is a generic IBD locus, which is a functionally intuitive finding of potential therapeutic significance.

For loci previously associated with other inflammatory diseases, the direction of effect in Crohn's disease is usually the same; however, in five instances, the risk allele for one disease appears to be protective in another disease (see arrow symbol in the reported association column in **Tables 1** and **2**). In most such instances, functional annotation suggests modulation of T cells and other immune pathways. Indeed, GRAIL highlights a number of such genes. These inverse associations may reflect overlap in the pathways by which the host regulates effector functions in defense and regulatory functions in self tolerance. This is a delicate balance and, in the face of competing requirements, selection pressures may have conferred an advantage for divergent alleles in a cell- and environment-dependent manner.

The associated SNP rs281379 at 19ql3, also recently identified in another study[6], is highly correlated ($r^2 > 0.80$) with a common nonsense variant (rs601338, also known as c.428G>A or p.Trp142X) at *FUT2*. This is classically referred to as the non-secretor variant, as individuals homozygous for this null allele do not secrete blood group antigens at epithelial surfaces. Recently, non-secretors were identified as having near-complete protection from symptomatic GII.4 norovirus infection[14], and the same null allele is identified here as a risk factor for Crohn's disease. This suggests one potential elusive link between infection and immune-mediated disease.

In contrast to the implication of coding variation in *FUT2*, our previous data showed that most Crohn's disease-associated SNPs were not in linkage disequilibrium with coding polymorphisms[1], suggesting that regulatory effects are likely to be a more common mechanism of disease susceptibility. Providing further direct evidence for this, we here identified a number of new eQTL effects (**Table 2** and **Supplementary Note**), including *CARD9* ($\log_{10}$ odds score (LOD) = 12.4), *ERAP2* (LOD = 47.2) and *TNFSF11* (also known as *RANKL*) (LOD = 5.9). The latter maps adjacent to but outside the associated recombination interval, suggesting another potential long-range *cis*-regulatory effect as previously described for *PTGER4* in Crohn's disease[4]. RANKL has pleiotropic immunological effects and also stimulates osteoclast activity. This finding may be relevant to the osteoporosis clinically associated with Crohn's disease.

Given the importance of regulatory effects, it is notable that variants within the gene encoding a key mediator of epigenetic regulation, *DNMT3A* (the DNA methyltransferase 3A gene), should be associated with Crohn's disease. By inducing transcriptional silencing, DNMT3A is known to play an important role in immunoregulation. For example, DNMT3A methylates IL-4 and IFN-γ promoters following T-cell-receptor stimulation, hence regulating T-cell polarization[15], and induces dynamic regulation of TNF-α transcription following lipopolysaccharide exposure in leukocytes[16]. Genetically determined alterations in DNMT3A activity could thus have far-reaching effects.

The 32 loci described up until 2008 explained approximately 20% of Crohn's disease heritability Adding the 39 loci described since that time increases the proportion of heritability explained to only 23.2%. This pattern of common alleles, explaining a logarithmically decreasing fraction of heritability (**Fig. 2**), is consistent with a recent model of effect size distribution[17], which predicted (based on the previous Crohn's-disease meta-analysis) that our current sample size would likely identify 48 new loci. Furthermore, it is likely that more high-frequency Crohn's disease risk alleles of even smaller effect size remain unidentified: the same model predicts that 140 loci would be identified by a sample size of 50,000, but these loci would explain only a few more percent of Crohn's disease heritability. It is clear, therefore, that larger GWAS alone will not explain all of the missing heritability in Crohn's disease.

Published in: Nature Genetics (2010)
Status: Postprint (Author's version)

Box 1 *Noteworthy genes within loci newly implicated in Crohn's disease pathogenesis*

Although we highlight these as interesting genes, we do not yet have data to confirm causality.

■ ***VAMP3*** (1p36) encodes vesicle-associated membrane protein 3. Following bacterial stimulation of TNF-α production within macrophages, VAMP3 interacts with SNARE proteins first on the trans-Golgi network, where TNF-α is taken up, and then on the cell membrane, where TNF-α is released[19]. VAMP3 also plays a role in cell migration and adhesion by trafficking molecules such as βl integrin to the cell surface, and it has been implicated in autophagy[20].

■ ***MUC1-SCAMP3*** (1q22). *MUC1* encodes a key constituent of mucus, which is the physical barrier that protects the intestinal epithelium from gut bacteria. MUC1 overexpression and hypoglycosylation have been reported in IBD[21], and Mucl knockdown mice exhibit increased small intestinal damage after *C. jejuni* infection[22]. Secretory carrier membrane protein 3 (encoded by *SCAMP3*) regulates EGFR trafficking within endosomal membranes[23]. It is manipulated by intracellular salmonellae to acquire nutrients and influence host immune responses[24].

■ ***DENND1B*** (1q31) has recently been associated with asthma[25] and is expressed in dendritic and natural killer cells. DENN-containing proteins influence MAP-kinase signaling pathways and DENND1B, in particular, has been predicted to interact with TNFR1.

■ ***IL10*** (1q32). Association with Crohn's disease follows its recent implication in ulcerative colitis[13] and the reporting of mutations in the *IL10* receptors in extreme Crohn's disease in infancy[26]. Known to inhibit synthesis of pro-inflammatory cytokines within macrophages and Thl cells, IL-10 also suppresses antigen-presenting cell activity. Knockdown of IL-10 in mice presents one of the best animal models of IBD.

■ ***DNMT3A*** (2p23). DNA methyltransferase 3A is one of three key methyltransferase genes in humans, effecting epigenetic regulation of gene transcription by methylating cytosine residues within CpG islands. Among many other roles, this protein is known to determine dynamic regulation of both adaptive and innate immune mechanisms[15,16].

■ ***GCKR*** (2p23) encodes an inhibitor of glucokinase, with the focal SNPs at this locus also correlating with both fibrinogen and CRP levels[27].

■ ***THADA*** (2p21) is expressed in the small intestine and appears to encode a death receptor-interacting protein, suggesting an apoptotic function[28].

■ ***ERAP2*** (5q15). Regulated by NF-κB, this gene encodes one of two human endoplasmic reticulum aminopeptidases, which work in concert to trim peptides for presentation on MHC class 1 and hence critically affect antigen presentation to T cells[29]. Ankylosing spondylitis is associated with this locus, but with a different pattern of associated variants[30]. Given the close clinical relationship between Crohn's disease and ankylosing spondylitis, and the strong association of HLA-B27 with the latter but not with the former, the divergent association of these closely related molecules is intriguing and will refocus interest on the MHC class 1 associations in Crohn's disease.

■ ***NDFIP1*** (5q31). Nedd4-family interacting protein 1 is a membrane protein involved in maintenance of the Golgi complex[31]. It is important for protein trafficking through exosomes and may play a role in rapid sequestration and removal of proteins during stress[32].

■ ***CPEB4*** (5q35) encodes a regulator of protein translation and cell division and is a transcriptional target of RORγt. Mouse work suggests that the product of *CPEB4* is the effector by which RORγt (a key determinant of Thl7 cell differentiation) inhibits proliferation of thymocytes[33].

■ ***TAGAP*** (6q25). T-cell activation GTPase-activating protein, associated with multiple autoimmune diseases, was originally identified through its involvement in human T-cell activation and co-regulation with IL-2 (ref. 24).

■ ***IL2RA*** (10p15) encodes part of the IL-2 receptor complex, thus mediating IL-2 signaling in host defense and regulating response to autoantigens by Tregs. The associated variants correlate with differential expression of IL2RA (CD25) on $CD4^+$ naive and memory T cells[35] possibly affecting $Foxp3^+$ Treg homeostasis[36].

■ ***FADS2*** (11q12). Fatty acid desaturase 2 is predominantly located in the endoplasmic reticulum. *Fads2* knockdown mice develop duodenal and ileocecal ulceration[37].

■ ***TNFSF11*** (13q14), also called *RANKL* (receptor activator of nuclear factor kappa B) and *ODF* (osteoclast differentiation factor), encodes a member of the TNF cytokine family. RANKL stimulation of dendritic cells leads to proliferation of naive T cells and inducible Tregs[38]. It also regulates osteoclast activity and bone loss. Previous studies have demonstrated increased plasma levels in Crohn's disease[39].

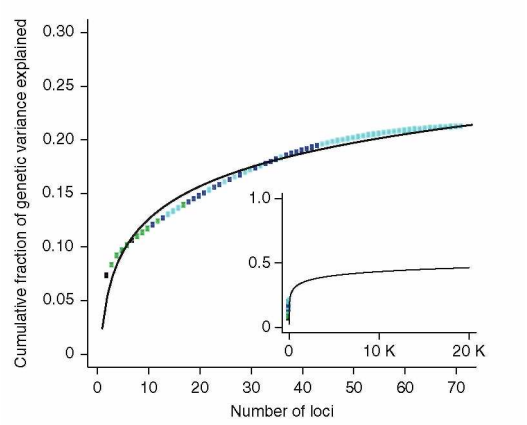
■ ***SMAD3*** (15q22). Phosphorylated following TGF-β signaling through its receptor, the SMAD3 protein complexes with SMAD4 and is then translocated to the nucleus to modulate target gene expression. SMAD3 plays a key role in the TGF-β-mediated induction of $Foxp3^+$ regulatory T cells[40], with SMAD3 deficiency reciprocally enhancing Thl7. Reduced SMAD3 phosphorylation has been observed in IBD and may impair the immunosuppressive effect of TGF-β.

■ ***TYK2*** (19p13) encodes tyrosine kinase 2, a member of the JAK-signal transduction family. It is involved in cytokine signaling by IFN-γ, IL-12 and IL-23 among others, hence affecting Thl and Thl7 lineage development. TYK2 also plays an important role in TLR-mediated responses in dendritic cells, including IL-12 and IL-23 production, and *TYK2* mutations predispose to opportunistic infection[41].

■ ***FUT2*** (19q13) encodes α-(l,2)fucosyltransferase, which regulates expression of the Lewis AB0 (H) histo-blood group of antigens on the surface of epithelial cells and in body fluids. It is strongly associated with Norovirus infection, as well as with with *Helicobacter pylori* infection and serum vitamin B12 levels[14,42].

One key shortcoming of our current model of heritability explained by these loci is a direct consequence of the extent to which GWAS tag SNPs are often imperfect proxies for causal alleles and thus substantially underestimate the true attributable risk. For example, the best tag SNP at the *NOD2* locus in our meta-analysis appears to explain just 0.8% of genetic variance, whereas the three *NOD2* coding mutations themselves account for 5%. If an analogous situation applies to even a small fraction of the other 70 Crohn's disease susceptibility loci, the proportion of overall heritability explained will increase substantially. Indeed, one study of linkage disequilibrium between tag SNPs and causal variants in the heritability of human height[18] suggests that this effect might double the total fraction of heritability explained by GWAS SNPs. Coding variants identified here from the 1000 Genomes Project that are in strong linkage disequilibrium with the focal SNPs in several of our regions (**Supplementary Table 5**) thus now require direct assessment in order to explore this possibility.

Figure 2 *Cumulative fraction of genetic variance explained by 71 Crohn's disease risk loci.* The loci are ordered from largest to smallest individual contribution. Black points were identified pre-GWAS, green points were dentified in the first generation GWAS, blue points were identified in an earlier meta-analysis, and cyan points were identified in this analysis. The inset shows a logarithmic fit to these data extrapolated to an extreme scenario where 20,000 independent common alleles are associated with disease. Even in this situation, less than half of the genetic variance would be explained. This demonstrates that other types of effect (for example, low frequency and rare alleles with higher penetrance) must also exist.



Other factors will also account for the heritability gap, including uncertain epidemiological estimates of disease prevalence and total heritability, as well as our observation that several of the new regions contain more than one independent risk allele. The likelihood is that many more such effects will be identified. Indeed, detailed future analyses will play a key role in helping us to understand the absolute contribution of common causal alleles, as well as in identifying lower frequency variants and rare (even family-specific) mutations. By contrast, our lack of evidence for epistasis among the loci described here suggests that non-additive interactions among common risk alleles do not play an important role in the genetic architecture of Crohn's disease.

The current study has approximately doubled the number of confirmed Crohn's disease susceptibility loci. For many of these loci, we have identified potentially causal genes, though confirmation of their role must await detailed fine mapping, expression and functional studies. Although the alleles detected only modestly affect disease risk, they continue to enhance our understanding of the genetic etiology of Crohn's disease. Looking for evidence of sub-phenotype associations represents an important future goal for the consortium. Thus, we are working toward sharing detailed genotype and clinical data to allow this. In the meantime, extensive resequencing, together with large-scale fine mapping exercises using custom array-based technologies, is already underway and will further elucidate the pathogenic mechanisms of IBD.

URLs. Evoker v1.0, http://www.sanger.ac.uk/resources/software/ evoker/.

ONLINE METHODS

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

Study subjects

All study subjects were of European origin. The meta-analysis was based on data from 6,333 Crohns disease cases and 15,056 controls derived from six index genome-wide scans from Germany, Cedars-Sinai Medical Center (Los Angeles, California, USA)[6], the CHOPSTICKS consortium (of early onset cases)[5], and the NIDDK[45], Wellcome Trust Case Control Consortium (WTCCC)[46] and Belgian-French[4] studies. Details of the numbers of cases and controls genotyped in the respective scans and of the genotyping platforms used are given in **Supplementary Table 1**. The methods of ascertainment of the GWAS cohorts, as well as of quality control procedures applied, were provided in the aforementioned original publications. The GWAS set from Germany has not yet been published but used standard methods of data filtering, removing samples with >1% missing genotypes and using PLINK to identify samples with excess allele sharing (hence removing duplicates and retaining only one individual from pairs or groups of relatives) and to detect population structure, allowing removal of outliers identified by principal components analysis.

Details of the replication panel of 15,694 cases and 14,026 controls plus 414 parent-affected offspring trios are provided in **Supplementary Table 2**. As with the GWAS subjects, the replication cases were ascertained in major IBD centers using standard clinical, endoscopic, radiological and histopathological criteria for diagnosis of Crohns disease. Each center supplying cases also supplied its own panel of controls. The controls for the Cedars-Sinai study were obtained from the Cardiovascular Health Study (CHS), a population-based longitudinal study of risk factors for cardiovascular disease and stroke in adults 65 years of age or older, recruited at four field centers[47]. Five thousand two hundred and one individuals were recruited in 1989-1990 from random samples of Medicare eligibility lists, followed by an additional 687 individuals recruited in 1992-1993 (total n = 5,888).

All participating centers received approval from their local and national institutional review boards, and informed consent was obtained from all participants.

Imputation

GWAS imputation was carried out using BEAGLE[48,49] and the HapMap3 reference samples from the CEU, TSI, MEX and GJT collections, with the exception of the early onset samples, which were imputed using the MACH program and the HapMap2 CEU reference samples. We carried forward a set of 953,242 autosomal markers (HapMap3 X-chromosome data were not available) that were polymorphic in at least one GWAS dataset for association analysis.

Test for association

We used genotype probabilities and empirical variances to summarize the standard 1 degree-of-freedom allele-based test for association as a z-score within each scan and combined the scores across all six datasets (inversely weighted by variance) to produce a single meta-statistic for each SNP. Odds ratios for replicated SNPs were estimated jointly from all case-control replication data available. Genomic positions were retrieved from NCBI's dbSNP build vl30. Linkage disequilibrium regions around focal SNPs were defined by extending the region to the left for 0.1 cM or until another SNP with $P < 10^{-5}$ was reached, in which case the process was repeated from this SNP. Right-hand boundaries were defined in the same way.

Notably, the 'belge' and 'cedar2' GWAS collections showed more pronounced inflation than the other six collections (**Supplementary Fig. 1**). Three lines of evidence suggest that population structure in these collections is not driving the association at our 71 loci: (i) the principal component analysis-corrected association analysis of the belge and cedar2 samples implies that some of the inflation is due to population stratification (the belge inflation decreased from 1.2 to 1.15 and the cedar2 decreased from 1.30 to 1.05 after control for the principal components). Under the null hypothesis that all association signals were driven by this stratification, we would expect a consistent decrease in significance at our 71 loci after correction. By contrast, 28 loci become more significant in the belge subset and 37 loci become more significant in the cedar2 subset after correction, whereas the P values of the remaining loci did not change significantly. This is especially noteworthy because the additional degrees of freedom in the corrected test reduced power on average, and this evidence strongly affirms that the signals at these 71 loci are driven by disease association and not by confounding. For nearly all SNPs, the change before and after corrections was small (**Supplementary Table 3**). (ii) The belge and cedar2 panels only represent 11% of our GWAS sample, making it difficult for them to drive associations in the full meta-analysis. We also tested for heterogeneity of odds ratios across all collections and found no significant differences after correcting for 71 tests (**Supplementary Table 3**). (iii) Some of the overall study inflation is due to true disease signal because the inflation factor decreases from 1.27 to 1.24 after excluding the known loci

(with many more to be identified in the future).

Replication

We selected the most significantly associated SNP from each region that was taken forward to replication and carried out genotyping according to standard protocols associated with platforms described in **Supplementary Table 2**. Samples with > 10% missing data in any experiment were excluded from further analysis.

The following additional quality controls were applied to the data before the association analysis: (i) visual and/or manual inspection of each cluster plot, that is, the genotype assignment and cluster separation were manually checked. For the Sequenom data, we used Evoker v1.0 (see URLs)[50] to examine the cluster plots. For SNPlex, genotype assignments were verified visually using the GeneMapper 4.0 (Applied Biosystems) software, and for TaqMan, we used the SDS v2.3 progam (Applied Biosystems). The aim of examining a cluster plot is twofold: to determine whether a given SNP has been genotyped well (in particular, whether clear distinct clusters can be identified on the plot that would correspond to the three genotypes) and to determine whether the calling algorithm has called the clusters correctly. If both of these requirements were fulfilled, as genotype counts can usually be assumed to be sufficiently accurate. If not, any observed disease association of such a SNP may be due to incorrect genotype counts, (ii) Call rate (genotype success rate) in each case or control panel greater than 95%. (iii) Hardy-Weinberg $P > 10^{-4}$ in each healthy control sample.

Replication and joint *P* values were calculated using the weighted z statistic described above.

Heritability analyses

We estimated the fraction of additive genetic variance explained using the liability threshold model of Risch[51], which assumes an additive effect at each locus and which shifts the mean of a normally distributed distribution of disease liability for each genotype class. Given the lack of evidence for gene-gene interaction among our 71 loci, we consider these contributions to be additive. We assumed a prevalence of Crohns disease of 4 per 1,000 and a total narrow-sense heritability of 50% (ref. 52). Results of this analysis are shown in **Figure 2**.

Interaction analyses

To test for pairwise interactions among the 71 confirmed associated loci, we performed an interaction meta-analysis using the approach described below. Each of the six scans performed an identical pairwise scan of the 71 × 70/2 pairs of SNPs in **Tables 1**,**2** using logistic regression in which three terms (dosage of SNP 1, dosage of SNP 2 and an additive interaction term) were included. The significance of this interaction terms was represented by a directional z-score which was then combined across all six studies using a sample-size-weighted combination of z-scores. The overall combined quantile-quantile plot resulting from this analysis showed no deviation from the null (**Supplementary Fig. 3**), and no results were significant when considering the number of tests performed.

ACKNOWLEDGMENTS

We thank all subjects who contributed samples and the physicians and nursing staff who helped with recruitment globally. This study was supported by the German Ministry of Education and Research through the National Genome Research Network and infrastructure support through the DFG cluster of excellence 'Inflammation at Interfaces' and by the Italian Ministry for Health GR-2008-1144485, with case collections supported by the Italian Group for IBD and the Italian Society for Paediatric Gastroenterology, Hepatology and Nutrition. We acknowledge funding provided by the Royal Brisbane and Women's Hospital Foundation; the University of Queensland (Ferguson Fellowship); the National Health and Medical Research Council, Australia, and by the European Community (5th PCRDT) and by the European Crohn's and Colitis Organization. UK case collections were supported by the National Association for Colitis and Crohn's disease, Action Medical Research, Wellcome Trust, Medical Research Council UK, the University of Edinburgh and the Peninsular College of Medicine and Dentistry Exeter. We also acknowledge the National Institute of Health Research (NIHR) Biomedical Research Centre awards to Guy's & St. Thomas' National Health Service Trust/King's College London and to Addenbrooke's Hospital/University of Cambridge School of Clinical Medicine. The NIDDKIBD Genetics Consortium is funded by the following grants: DK062431 (S.R.B.), DK062422 (J.H.C.), DK062420 (R.H.D.), DK062432 and DK064869 (J.D.R.), DK062423 (M.S.S.), DK062413 (D.P.B.M.), DK76984 (M.J.D.), and

DK084554 (M.J.D. and D.P.B.M.) and DK062429 (J.H.C.). J.H.C. is also funded by the Crohn's and Colitis Foundation of America, and S.L.G. is funded by DK069513 and the Primary Children's Medical Center Foundation. Cedars Sinai was supported by National Center for Research Resources (NCRR) grant M01-RR00425; US National Institutes of Health/NIDDK grant P01-DK046763; DK 063491; and Cedars-Sinai Medical Center Inflammatory Bowel Disease Research Funds. R.K.W is supported by a clinical fellow grant (90700281) from The Netherlands Organization for Scientific Research. E.L., D.F. and S.V. are senior clinical investigators for the Funds for Scientific Research (FWO/FNRS) Belgium. S.B. was supported by the 'Deutsche Forschungsgemeinschaft' (DFG; BR 1912/5-1). J.C. Barrett is supported by Wellcome Trust grant WT089120/Z/09/Z. Replication genotyping was supported by unrestricted grants from Abbott Laboratories Ltd. and Giuliani SpA. We acknowledge the Wellcome Trust Case Control Consortium. We thank the 1958 British Birth Cohort and Banco Nacional de ADN, Salamanca, Spain, who supplied control DNA samples. The CHS research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. Other significant contributors were K. Hanigan, Z.-Z. Zhao, N. Huang, P. Webb, N. Hayward, A. Rutherford, R. Gwilliam, J. Ghori, D. Strachan, W. McCardle, W. Ouwehand, M. Newsky, S. Ehlers, I. Pauselius, K. Holm, C. Sina, L. Baidoo, A. Andriulli and M.C. Renda.

AUTHOR CONTRIBUTIONS

A.F., D.P.B.M., G.L.R.-S., T.A., J.L., R. Roberts, J.C. Bis, T.H., A. Latiano, C.G.M., N.J.P., J.I.R., P.S., Y.S., L.A.S., K.D.T., D. Whiteman, C.W, G.K.-U, J.D.R., M.DA., R.K.W, S.V., R.H.D., J. Satsangi, S.S., V.A., H.H. and M.P. were involved in establishing DNA collections and/or assembling phenotypic data. A.F., D.E., J.C. Barrett, K.W, T.G., S.R., C.A.A., L.J. and M.J.D. performed statistical analyses. D.P.B.M., G.L.R.-S., C.W.L., E.M.F., R.N.B., M.B., T.M.B., S. Brand, C.B., A.C., J.-F.C, M.C., L.S., T.D., M.D.V., R.D.'I., M.D., C.E., T.F., D.F, A.M.G., R.G., J.G., A.V.G., S.L.G., J.H., H.W.V., J.-P.H., A.K., D.L., I.L., M.L., A. Levine, C.L., E.L., C.M., W.N., J.P., A.P., D.D.P., M.R., P.R., R. Russell, J. Satsangi, M.S.S., M.S., F.S., A.H.S., P.C.F.S., S.R.T., L.T., T.W, S.R.B., R.K.W, S.K., A.M.G., J.C.M., S.V., D. Wilson, R.H.D., M.S., J. Sanderson, S.S., J.H.C, V.A. and M.P. recruited patients. A.F., D.P.B.M., T.B., S. Bumpstead, J.I.R., M.G. and G.W.M. supervised laboratory work. A.F., D.P.B.M., J.C. Barrett, K.W, S. Brand, R.H.D., J. Satsangi, S.S., J.H.C, M.J.D. and M.P contributed to writing the manuscript. All authors read and approved the final manuscript before submission.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. 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).
2. Duerr, R.H. *et al.* A genome-wide association study identifies *IL23R* as an nflammatory bowel disease gene. *Science* 314, 1461-1463 (2006).
3. 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).
4. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5pl3.1 and modulates expression of *PTGER4*. *PLoS Genet.* 3, e58 (2007).
5. Imielinski, M. *et al.* Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat. Genet.* 41, 1335-1340 (2009).
6. McGovern, D.P. *et al.* Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum. Mol. Genet.* 19, 3468-3476 (2010).
7. Zhernakova, A. *et al.* Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring *CARD9* and *IL18RAP*. *Am. J. Hum. Genet.* 82, 1202-1210 (2008).
8. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* 39, 1202-1207 (2007).
9. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* 5, el000534 (2009).
10. Rueda, B. *et al.* The IL23R Arg381Gln non-synonymous polymorphism confers susceptibility to ankylosing spondylitis. *Ann. Rheum. Dis.* 67, 1451-1454 (2008).
11. Yamazaki, K. *et al.* Single nucleotide polymorphisms in *TNFSF15* confer susceptibility to Crohn's disease. *Hum. Mol. Genet.* 14, 3499-3506 (2005).
12. Zinovieva, E. *et al.* Comprehensive linkage and association analyses identify haplotype, near to the *TNFSF15* gene, significantly associated with spondyloarthritis. *PLoS Genet.* 5, el000528 (2009).
13. Franke, A. *et al.* Sequence variants in *IL10, ARPC2* and multiple other loc contribute to ulcerative colitis susceptibility. *Nat. Genet.* 40, 1319-1323 (2008).
14. Carlsson, B. *et al.* The G428A nonsense mutation in *FUT2* provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. *PLoS ONF* 4, e5593 (2009).

15. Gamper, C.J., Agoston, A.T., Nelson, W.G. & Powell, J.D. Identification of DNA methyltransferase 3a as a T cell receptor-induced regulator of Thl and Th2 differentiation. *J. Immunol.* 183, 2267-2276 (2009).
16. El Gazzar, M. *et al.* G9a and HP1 couple histone and DNA methylation to TNFα transcription silencing during endotoxin tolerance. *J. Biol. Chem.* 283, 32198-32208 (2008).
17. Park, l.H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* 42, 570-575 (2010).
18. Yang, 1. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565-569 (2010).
19. Murray, R.Z., Kay, I.G., Sangermani, D.G. & Stow, l.L. A role for the phagosome in cytokine secretion. *Science* 310, 1492-1495 (2005).
20. Luftman, K., Hasan, N., Day, P., Hardee, D. & Hu, C. Silencing of VAMP3 inhibits cell migration and integrin-mediated adhesion. *Biochem. Biophys. Res. Commun.* 380, 65-70 (2009).
21. Campbell, B.I., Yu, L.G. & Rhodes, l.M. Altered glycosylation in inflammatory bowel disease: a possible role in cancer development. *Glycoconj. J.* 18, 851-858 (2001).
22. McAuley, l.L. *et al.* MUC1 cell surface mucin is a critical element of the mucosa barrier to infection. *J. Clin. Invest.* 117, 2313-2324 (2007).
23. Aoh, Q.L., Castle, A.M., Hubbard, C.H., Katsumata, 0. & Castle, l.D. SCAMP3 negatively regulates epidermal growth factor receptor degradation and promotes receptor recycling. *Mol. Biol. Cell* 20, 1816-1832 (2009).
24. Mota, L.J., Ramsden, A.E., Liu, M., Castle, l.D. & Holden, D.W. SCAMP3 is a component of the Salmonella-induced tubular network and reveals an interaction between bacterial effectors and post-Golgi trafficking. *Cell. Microbiol.* 11, 1236-1253 (2009).
25. Sleiman, P.M. *et al.* Variants of *DFNND1B* associated with asthma in children. *N. Engl. J. Med.* 362, 36-44 (2010).
26. Glocker, E.O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* 361, 2033-2045 (2009).
27. Danik, l.S. *et al.* Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17,686 women: the Women's Genome Health Study. *Circ. Cardiovasc. Genet.* 2, 134-141 (2009).
28. Rippe, V. *et al.* Identification of a gene rearranged by 2p21 aberrations in thyroid adenomas. *Oncogene* 22, 6111-6114 (2003).
29. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* 6, 689-697 (2005).
30. Burton, P.R. *et al.* Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* 39, 1329-1337 (2007).
31. Harvey, K.F., Shearwin-Whyatt, L.M., Fotia, A., Parton, R.G. & Kumar, S. N4WBP5, a potential target for ubiquitination by the Nedd4 family of proteins, is a novel Golgi-associated protein. *J. Biol. Chem.* 277, 9307-9317 (2002).
32. Putz, U. *et al.* Nedd4 family-interacting protein 1 (Ndfipl) is required for the exosomal secretion of Nedd4 family proteins. *J. Biol. Chem.* 283, 32621-32627 (2008).
33. Xi, H., Schwartz, R., Engel, I., Murre, C. & Kersh, G.l. Interplay between RORγt, Egr3, and E proteins controls proliferation in response to pre-TCR signals. *Immunity* 24, 813-826 (2006).
34. Mao, M. *et al.* T lymphocyte activation gene identification by coregulated expression on DNA microarrays. *Genomics* 83, 989-999 (2004).
35. Dendrou, C.A. *et al.* Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. *Nat. Genet.* 41, 1011-1015 (2009).
36. Burchill, M.A., Yang, J., Vang, K.B. & Farrar, M.A. lnterleukin-2 receptor signaling n regulatory T cell development and homeostasis. *Immunol. Lett.* 114, 1-8 (2007).
37. Stroud, C.K. *et al.* Disruption of *FADS2* gene in mice impairs male reproduction and causes dermal and intestinal ulceration. *J. Lipid Res.* 50, 1870-1880 (2009).
38. Loser, K. *et al.* Epidermal RANKL controls regulatory T-cell numbers via activation of dendritic cells. *Nat. Med.* 12, 1372-1379 (2006).
39. Moschen, A.R. *et al.* The RANKL/OPG system is activated in inflammatory bowe disease and relates to the state of bone loss. *Gut* 54, 479-487 (2005).
40. Lu, L. *et al.* Role of SMAD and non-SMAD signals in the development of Thl7 and regulatory T cells. *J. Immunol.* 184, 4295-4306 (2010).
41. Ghoreschi, K., Laurence, A. & O'Shea, J.J. lanus kinases in immune cell signaling. *Immunol. Rev.* 228, 273-287 (2009).
42. Hazra, A. *et al.* Common variants of *FUT2* are associated with plasma vitamin B12 levels. *Nat. Genet.* 40, 1160-1162 (2008).
43. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106, 9362-9367 (2009).
44. Yu, W., Clyne, M., Khoury, M.l. & Gwinn, M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* 26, 145-146 (2010).
45. 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).
46. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678 (2007).
47. Fried, L.P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann. Epidemiol.* 1, 263-276 (1991).
48. 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, 210-223 (2009).
49. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084-1097 (2007).
50. Morris, J.A., Randall, J.C., Mailer, J.B. & Barrett, J.C. Evoker: a visualization tool for genotype intensity data. *Bioinformatics* 26, 1786-1787 (2010).
51. Risch, N.l. Searching for genetic determinants in the new millennium. *Nature* 405, 847-856 (2000).
52. Ahmad, T., Satsangi, J., McGovern, D., Bunce, M. & lewell, D.P. The genetics of inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 15, 731-748 (2001).