

Published in final edited form as:

*Nat Genet.* 2010 December ; 42(12): 1118–1125. doi:10.1038/ng.717.

## Meta-Analysis Increases to 71 the Tally of Confirmed Crohn's Disease Susceptibility Loci

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

<sup>1</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Schittenhelmstr. 12, D-24105 Kiel, Germany <sup>2</sup>Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. <sup>3</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK <sup>4</sup>Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA <sup>5</sup>Inflammatory Bowel Disease Research Group, Queensland Institute of Medical Research, Brisbane, Australia. <sup>6</sup>Peninsula College of Medicine and Dentistry, Barrack Road, Exeter, UK <sup>7</sup>Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, UK <sup>8</sup>popgen Biobank, Christian-Albrechts University Kiel, D-24105 Kiel, Germany <sup>9</sup>Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK

†Corresponding author: Miles Parkes, Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 0QQ, United Kingdom, miles.parkes@addenbrookes.nhs.uk Tel.: +44 (0) 1223-216389, Fax: +44 (0) 1223 596213 .

\*Shared first authorship

†Shared senior authorship

**Contribution of authors** AF, DPBM, GRS, TA, JL, RR, JB, TH, AL, CGM, NP, JIR, PS, YS, LS, KDT, DW, CW, GKU, JDR, MD'A, RW, SV, RHD, JS, SS, VA, HH were involved in establishing DNA collections, and/or assembling phenotypic data; AF, DE, JCB, KW, TG, SR, CAA, LJ, MJD performed statistical analyses; DPBM, GRS, CWL, EMF, RNB, MB, TMB, SB, CB, AC, J-FC, MC, SC, TD, MdV, RD'I, MD, CE, TF, DF, RG, JG, AVG, SLG, JH, DH, J-PH, DL, IL, ML, AL, CL, EL, CM, WN, JP, AP, DDP, MR, PR, JS, MS, FS, AHS, PCFS, SRT, LT, TW, SRB, RW, SK, AMG, JCM, SV, RHD, MSS, JS, SS, JHC, VA recruited patients; AF, DPBM, TB, SB, KT, MG, GM supervised laboratory work; AF, DPBM, JCB, KW, SB, RHD, JS, SS, JHC, MJD, MP contributed to writing the manuscript. All authors read and approved the final manuscript before submission.

All authors declare no financial interest.

<sup>10</sup>Department of Medicine, University of Otago, Christchurch 8140, New Zealand  
<sup>11</sup>Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, USA <sup>12</sup>Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands <sup>13</sup>Department of Genetics, Faculty of Veterinary Medicine, University of Liège B43, 20 Bd de Colonster, 4000 Liège, Belgium <sup>14</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA <sup>15</sup>Unit of Gastroenterology, IRCCS-CSS Hospital, San Giovanni Rotondo, Italy <sup>16</sup>Department of Medical and Molecular Genetics, King's College London School of Medicine, Floor 8 Tower Wing, Guy's Hospital, London, UK <sup>17</sup>Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia 4006 <sup>18</sup>Department of Health Studies, University of Chicago, Chicago, Illinois, USA <sup>19</sup>Section of Digestive Diseases, Department of Medicine, Yale University, New Haven, Connecticut, USA <sup>20</sup>Department of Pediatrics, Center for Pediatric Inflammatory Bowel Disease, The Children's Hospital of Philadelphia, Philadelphia, USA <sup>21</sup>Inflammatory Bowel Disease Center, Dept. of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, U.S.A <sup>22</sup>Department of Medicine II, University Hospital Munich Grosshadern, Ludwig-Maximilians-University, Munich, Germany <sup>23</sup>Department of Gastroenterology, Charité, Campus Mitte, Universitätsmedizin Berlin, Berlin, Germany <sup>24</sup>Montreal Jewish General Hospital, Montréal, Québec, Canada <sup>25</sup>Registre EPIMAD, Université de Lille, Lille, France <sup>26</sup>Unit of Gastroenterology, Cervello Hospital, Palermo, Italy <sup>27</sup>ENEA, Department of Biology of Radiations and Human Health, Rome, Italy <sup>28</sup>Pediatric Gastroenterology, Cincinnati Children's Hospital. Medical Center. 3333 Burnet Ave, Cincinnati, USA <sup>29</sup>Department of Hepatology and Gastroenterology, Ghent University Hospital, Ghent, Belgium, <sup>30</sup>Division of Gastroenterology, University Hospital Padua, Italy <sup>31</sup>Department of Pediatrics, Cedars Sinai Medical Center, Los Angeles, CA, USA <sup>32</sup>Torbay Hospital, Torbay, Devon, UK <sup>33</sup>Department of Gastroenterology, Mater Health Services, Brisbane, Australia 4101 <sup>34</sup>Department of Gastroenterology, Erasmus Hospital, Free University of Brussels, Brussels, Belgium <sup>35</sup>Department of Preventive Dentistry and Periodontology, Ludwig-Maximilians-University, Munich, Germany <sup>36</sup>Department of Human Genetics, RWTH Aachen, Germany <sup>37</sup>Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA <sup>38</sup>Department of Medicine, Örebro University Hospital, Örebro, Sweden. <sup>39</sup>Dept of Gastroenterology, Leiden University Medical Center, Leiden, The Netherlands <sup>40</sup>Université Paris Diderot, Paris, France <sup>41</sup>School of Medicine and Pharmacology, The University of Western Australia, Fremantle, Australia 6160 <sup>42</sup>GETAID group, Université Paris Diderot, Paris, France <sup>43</sup>Pediatric Gastroenterology Unit, Wolfson Medical Center and Sackler School of Medicine, Tel Aviv University, Israel <sup>44</sup>Division of Gastroenterology, CHU, Université de Liège, Liège, Belgium <sup>45</sup>Dept of Medicine, Ninewells Hospital and Medical School, Dundee, UK <sup>46</sup>Department of Medical Genetics, University of Manchester, Manchester, UK <sup>47</sup>Department of Gastroenterology, Hospital Clínic / IDIBAPS. CIBER EHD. Barcelona, Spain <sup>48</sup>Division of Gastroenterology, Hepatology and Nutrition, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA <sup>49</sup>Division of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium <sup>50</sup>Dept Gastroenterology, Guy's & St Thomas' NHS Foundation Trust, St Thomas' Hospital, London, UK <sup>51</sup>Division of Gastroenterology, Inselspital, University of Bern, Bern, Switzerland <sup>52</sup>Mount Sinai Hospital Inflammatory Bowel Disease Centre, University of Toronto, Canada <sup>53</sup>Department of Gastroenterology, Academic Medical Center, Amsterdam, the Netherlands <sup>54</sup>Department of Clinical Science Intervention and Technology, Karolinska Institutet, Stockholm, Sweden <sup>55</sup>Division of Clinical Pharmacology and Toxicology University Hospital Zurich, CH-8091 Zurich, Switzerland <sup>56</sup>Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA <sup>57</sup>The Hospital for Sick Children, University of Toronto, Ontario, Canada <sup>58</sup>Université de Montréal and the Montreal Heart Institute, Research Center, Montréal, Québec, Canada. <sup>59</sup>Department of Biosciences and Nutrition, Karolinska Institute, Stockholm, Sweden <sup>60</sup>Department of Gastroenterology, University Medical Center Groningen, Groningen, The Netherlands <sup>61</sup>Department of Pediatrics, Emory University School of Medicine,

Atlanta, GA, USA <sup>62</sup>Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK <sup>63</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA <sup>64</sup>Department for General Internal Medicine, Christian-Albrechts-University, Schittenhelmstr. 12, D-24105 Kiel, Germany <sup>65</sup>Department of Genetics, Yale School of Medicine, New Haven CT, USA <sup>66</sup>Unit of Gastroenterology, University Hospital Careggi Florence, Italy

We undertook a meta-analysis of six Crohn's disease (CD) genome-wide association studies (GWAS) comprising 6,333 cases and 15,056 controls, and followed up the top association signals in 15,694 cases, 14,026 controls and 414 parent/offspring trios. Thirty new susceptibility loci meeting genome-wide significance ( $P$ -value  $<5 \times 10^{-8}$ ) were identified. 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, the results described here identify a total of 71 distinct loci with genome-wide significant evidence for association with Crohn's disease.

Crohn's disease (OMIM #266600) results from the interaction of environmental factors, including the intestinal microbiota, with host immune mechanisms in genetically susceptible individuals. Along with ulcerative colitis (UC), it is one of the main subphenotypes of inflammatory bowel disease (IBD). GWAS have highlighted key CD pathogenic mechanisms, 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<sup>1</sup>. Recognizing that an increased sample size would be required to detect these, we have expanded the International IBD Genetics Consortium (IBDGC), approximately doubling the discovery panel size in comparison with the first meta-analysis.

The discovery panel for the current study comprised 6,333 CD subjects and 15,056 controls, all of European descent, with data derived from six index GWAS studies (for overview see Supplementary Table 1)<sup>2-6</sup>. Imputation using HapMap3 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 20% in healthy controls. Under the same conditions, the sample size of our original meta-analysis had only 11% power<sup>1</sup>.

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 Figure 1). A total of 2,024 SNPs within 107 distinct genomic loci, including all previously defined significant hits from our earlier meta-analysis, demonstrated association with  $P$ -values  $< 10^{-5}$ . A Manhattan plot is shown in Supplementary Figure 2. 51 of the regions, representing new loci associated at  $P < 5 \times 10^{-6}$ , were followed up by genotyping the most significant SNPs in an independent panel of 15,694 CD cases, 14,026 controls and 414 parent/offspring trios (see Table 1 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 CD in the combined discovery plus replication panel, with at least nominal association in the replication panel (see Table 1). Two additional loci, encompassing the *CARD9* and *IL18RAP* genes, had previously been reported as associated with CD in a candidate gene study<sup>7</sup> and were here both replicated and confirmed at  $P < 5 \times 10^{-8}$ . Five loci were identified at genome-wide significance in GWAS studies

published subsequent to our replication experiment being designed. One, the FUT2 locus, was from a recent adult CD GWAS<sup>6</sup>. Four more (*ZMIZ1*, *IL27* at 16p11, 19q13 and 22q12) were identified in a pediatric IBD population<sup>5</sup>, these replicating here in our current sample set. Two further loci had produced “suggestive” evidence of association with replication in our earlier study<sup>1</sup>. Here, these clearly exceeded the genome-wide significance threshold in the meta-analysis alone and, given the previous replication evidence, were not followed up further (see Table 1). Thus cumulatively, 39 additional loci can now be added to the 32 confirmed CD susceptibility loci identified at the time of the Barrett *et al.* study. We did not observe statistically significant heterogeneity of the odds ratios (Breslow Day test P-value <0.05 after Bonferroni correction; Supplementary Table 4) between the panels from our 15 different countries (Supplementary Tables 1 and 2) for any of the 71 loci. Nor was any evidence of interaction between the associated loci observed (Supplementary Figure 3).

Regional association plots of all 71 susceptibility loci including the 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 had evidence for more than one independently associated variant (Table 1). While 6 of the 30 novel regions contain just a single gene, which is thereby strongly implicated in CD pathogenesis (e.g. *SMAD3*, *NDVIP1* and *BACH2*), 22 include more than one gene within the associated interval (Table 1; two regions without any gene or gene prediction). We thus applied additional *in silico* analyses to refine the list of functional candidate genes further. These were:

1. Interrogation of a publicly available expression quantitative trait loci (eQTL) database<sup>8</sup>. These analyses identified genes for which expression correlates with genotype at our most associated SNP (see Supplementary Results).
2. Use of 1000 Genomes Project Pilot sequence data and HapMap3 to identify genes containing non-synonymous variants in strong LD ( $r^2 > 0.5$ ) with the focal SNP within each region (for details on coding SNP see Supplementary Table 5).
3. Use of GRAIL<sup>9</sup>, to identify non-random and evidence-based connectivity between the genes in the 71 confirmed CD loci. Specifically, GRAIL evaluates each gene in a CD-associated locus for non-random correlation with genes in the other 70 loci via word-usage in PubMed abstracts related to the gene (see Figure 1).

Summary results of these analyses are shown in the rightmost column of Table 1. The highlighted genes are described briefly in Box 1, as are genes that constitute particularly noteworthy candidates from intervals containing one or few genes. While 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 is causal, and highlights the need for functional studies.

30 new signals were identified here beyond those described in the earlier meta-analysis<sup>1</sup> and other subsequent publications. The new associations were driven primarily by increased power arising from the expanded sample size rather than improved imputation, as more than two-thirds of the novel loci identified here have good proxies ( $r^2 > 0.8$ ) on both earlier generation arrays (Illumina 300K and Affymetrix 500k Set). 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 novel loci, associations have been reported previously in other complex diseases, comprising mostly chronic inflammatory disorders (Table 1). Such diseases can cluster both within families and individuals, reflecting shared genetic risk factors. For example, IBD and ankylosing spondylitis can co-segregate and both are associated with

*IL23R*<sup>2,10</sup> and *TNFSF15*<sup>1,12</sup>. The *IL10* locus was previously associated with UC<sup>13</sup> and was identified as a novel CD locus in the present study. 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 CD is usually the same, but in five cases the risk allele for one disease appears to be protective in another disease (see arrow symbol in “Reported association” column in Table 1). In most such instances, functional annotation suggests modulation of T cell 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 advantage for divergent alleles in a cell- and environmentally dependent manner.

The associated SNP rs281379 at 19q13, recently also identified by McGovern *et al.*<sup>6</sup> is highly correlated ( $r^2 > 0.80$ ) with a common nonsense variant (rs601338 also known as G428A or W142X) 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<sup>14</sup> and the same null allele is identified here as a CD risk factor. This suggests one potential elusive link between infection and immune-mediated disease.

In contrast to the implication of coding variation in the *FUT2* gene, our previous data demonstrated that most CD-associated SNPs were not in LD with coding polymorphisms<sup>1</sup>, suggesting that regulatory effects are likely to be a more common mechanism of disease susceptibility. Providing further direct evidence for this, a number of new eQTL effects were identified here (see Table 1 and Supplementary Results Section) – notably including *CARD9* (LOD=12.4), *ERAP2* (LOD=47.2) and *TNFSF11* (*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 CD<sup>4</sup>. *RANKL* has pleiotropic immunological effects and also stimulates osteoclast activity. This finding may be relevant to the osteoporosis clinically associated with CD.

Given the importance of regulatory effects, it is intriguing that variants within the gene encoding a key mediator of epigenetic regulation, *DNA methyltransferase 3a* (*DNMT3A*), should be associated with CD. By inducing transcriptional silencing, DNMT3a is known to play an important role in immunoregulation. For example, it methylates *IL-4* and *IFN- $\gamma$*  promoters following T cell receptor stimulation, hence regulating T cell polarization<sup>15</sup>, and induces dynamic regulation of *TNF- $\alpha$*  transcription following lipopolysaccharide exposure in leukocytes<sup>16</sup>. Genetically determined alterations in DNMT3a activity could thus have far-reaching effects.

The 32 loci described up to 2008 explained approximately 20% of CD heritability. Adding the 39 loci described since increases the proportion of heritability explained to just 23.2%. This pattern of common alleles, explaining a logarithmically decreasing fraction of heritability (Figure 2), is consistent with a recent model of effect size distribution<sup>17</sup>, which predicted (based on the previous CD meta-analysis) that our current sample size would likely identify 48 new loci. Furthermore, it is likely that more high-frequency CD 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 would explain only a few more percent of CD heritability. It is clear, therefore, that larger GWAS alone will not explain all of the missing heritability in CD.

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 CD susceptibility loci, the proportion of overall heritability explained will increase significantly. Indeed, one study of LD between tag SNPs and causal variants in the heritability of human height<sup>18</sup> suggests that this effect might double the total fraction of heritability explained by GWAS SNPs. Coding variants identified here from the 1000 Genomes Project which are in strong LD with the focal SNPs in several of our regions (see Supplementary Table 4) thus now require direct assessment in order to explore this possibility.

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 identifying less common 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 CD.

The current study has approximately doubled the number of confirmed CD susceptibility loci. For many of these loci we have identified potentially causal genes, accepting that confirmation of their role must await detailed fine mapping, expression and functional studies. While the alleles detected only modestly affect disease risk, they continue to enhance our understanding of the genetic etiology of CD. Analysis for evidence of sub-phenotype associations represents an important future goal for the consortium. Thus, we are working towards sharing of 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, are already underway and will further elucidate the pathogenic mechanisms of IBD.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank all subjects who contributed samples, and 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". Also 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 Royal Brisbane and Women's Hospital Foundation; University of Queensland (Ferguson Fellowship); 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, Wellcome Trust, Medical Research Council UK and Peninsular College of Medicine and Dentistry, Exeter. We also acknowledge the NIHR Biomedical Research Centre awards to Guy's & St Thomas' NHS Trust / King's College London and to Addenbrooke's Hospital / University of Cambridge School of Clinical Medicine. The NIDDK IBD Genetics Consortium is funded by the following grants: DK062431 (S.R.B.), DK062422 (J.H.C.), DK062420 (R.H.D.), DK062432 & DK064869 (J.D.R.), DK062423 (M.S.S.), DK062413 (D.P.B.M.), DK76984 (MD), and DK084554 (MD and DPBM), and DK062429 (J.H.C.). J.H.C. is also funded by the Crohn's and Colitis Foundation of America; and SLG by DK069513 and Primary Children's Medical Center Foundation. Cedars Sinai supported by NCCR grant M01-RR00425; NIH/NIDDK grant P01-DK046763; DK 063491; and Cedars-Sinai Medical Center Inflammatory Bowel

Disease Research Funds. RW is supported by a clinical fellow grant (90700281) from the Netherlands Organization for Scientific Research; EL, DF and SV are senior clinical investigators for the Funds for Scientific Research (FWO/FNRS) Belgium. SB was supported by the “Deutsche Forschungsgemeinschaft” (DFG; BR 1912/5-1). JCB 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: 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.

## References

1. Barrett JC, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008; 40:955–62. [PubMed: 18587394]
2. Duerr RH, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science.* 2006; 314:1461–3. [PubMed: 17068223]
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.* 2007; 39:830–2. [PubMed: 17554261]
4. 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.* 2007; 3:e58. [PubMed: 17447842]
5. Imielinski M, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet.* 2009; 41:1335–40. [PubMed: 19915574]
6. McGovern DP, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum Mol Genet.* 19:3468–76. [PubMed: 20570966]
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.* 2008; 82:1202–10. [PubMed: 18439550]
8. Dixon AL, et al. A genome-wide association study of global gene expression. *Nat Genet.* 2007; 39:1202–7. [PubMed: 17873877]
9. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* 2009; 5:e1000534. [PubMed: 19557189]
10. Rueda B, et al. The IL23R Arg381Gln non-synonymous polymorphism confers susceptibility to ankylosing spondylitis. *Ann Rheum Dis.* 2008; 67:1451–4. [PubMed: 18199597]
11. Yamazaki K, et al. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet.* 2005; 14:3499–506. [PubMed: 16221758]
12. Zinovieva E, et al. Comprehensive linkage and association analyses identify haplotype, near to the TNFSF15 gene, significantly associated with spondyloarthritis. *PLoS Genet.* 2009; 5:e1000528. [PubMed: 19543369]
13. Franke A, et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet.* 2008; 40:1319–23. [PubMed: 18836448]
14. Carlsson B, et al. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. *PLoS One.* 2009; 4:e5593. [PubMed: 19440360]
15. Gamper CJ, Agoston AT, Nelson WG, Powell JD. Identification of DNA methyltransferase 3a as a T cell receptor-induced regulator of Th1 and Th2 differentiation. *J Immunol.* 2009; 183:2267–76. [PubMed: 19625655]
16. El Gazzar M, et al. G9a and HP1 couple histone and DNA methylation to TNFalpha transcription silencing during endotoxin tolerance. *J Biol Chem.* 2008; 283:32198–208. [PubMed: 18809684]

17. Park JH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet.* 42:570–5. [PubMed: 20562874]
18. Yang J, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 42:565–9. [PubMed: 20562875]
19. Murray RZ, Kay JG, Sangermani DG, Stow JL. A role for the phagosome in cytokine secretion. *Science.* 2005; 310:1492–5. [PubMed: 16282525]
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.* 2009; 380:65–70. [PubMed: 19159614]
21. Campbell BJ, Yu LG, Rhodes JM. Altered glycosylation in inflammatory bowel disease: a possible role in cancer development. *Glycoconj J.* 2001; 18:851–8. [PubMed: 12820718]
22. McAuley JL, et al. MUC1 cell surface mucin is a critical element of the mucosal barrier to infection. *J Clin Invest.* 2007; 117:2313–24. [PubMed: 17641781]
23. Aoh QL, Castle AM, Hubbard CH, Katsumata O, Castle JD. SCAMP3 negatively regulates epidermal growth factor receptor degradation and promotes receptor recycling. *Mol Biol Cell.* 2009; 20:1816–32. [PubMed: 19158374]
24. Mota LJ, Ramsden AE, Liu M, Castle JD, Holden DW. SCAMP3 is a component of the Salmonella-induced tubular network and reveals an interaction between bacterial effectors and post-Golgi trafficking. *Cell Microbiol.* 2009; 11:1236–53. [PubMed: 19438519]
25. Sleiman PM, et al. Variants of DENND1B associated with asthma in children. *N Engl J Med.* 362:36–44. [PubMed: 20032318]
26. Glocker EO, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med.* 2009; 361:2033–45. [PubMed: 19890111]
27. Danik JS, 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.* 2009; 2:134–41. [PubMed: 20031577]
28. Rippe V, et al. Identification of a gene rearranged by 2p21 aberrations in thyroid adenomas. *Oncogene.* 2003; 22:6111–4. [PubMed: 12955091]
29. Saveanu L, et al. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol.* 2005; 6:689–97. [PubMed: 15908954]
30. Burton PR, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet.* 2007; 39:1329–37. [PubMed: 17952073]
31. Harvey KF, Shearwin-Whyatt LM, Fotia A, Parton RG, Kumar S. N4WBP5, a potential target for ubiquitination by the Nedd4 family of proteins, is a novel Golgi-associated protein. *J Biol Chem.* 2002; 277:9307–17. [PubMed: 11748237]
32. Putz U, et al. Nedd4 family-interacting protein 1 (Ndfip1) is required for the exosomal secretion of Nedd4 family proteins. *J Biol Chem.* 2008; 283:32621–7. [PubMed: 18819914]
33. Xi H, Schwartz R, Engel I, Murre C, Kersh GJ. Interplay between RORgammat, Egr3, and E proteins controls proliferation in response to pre-TCR signals. *Immunity.* 2006; 24:813–26. [PubMed: 16782036]
34. Mao M, et al. T lymphocyte activation gene identification by coregulated expression on DNA microarrays. *Genomics.* 2004; 83:989–99. [PubMed: 15177553]
35. Dendrou CA, et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nat Genet.* 2009; 41:1011–5. [PubMed: 19701192]
36. Burchill MA, Yang J, Vang KB, Farrar MA. Interleukin-2 receptor signaling in regulatory T cell development and homeostasis. *Immunol Lett.* 2007; 114:1–8. [PubMed: 17936914]
37. Stroud CK, et al. Disruption of FADS2 gene in mice impairs male reproduction and causes dermal and intestinal ulceration. *J Lipid Res.* 2009; 50:1870–80. [PubMed: 19351970]
38. Loser K, et al. Epidermal RANKL controls regulatory T-cell numbers via activation of dendritic cells. *Nat Med.* 2006; 12:1372–9. [PubMed: 17143276]
39. Moschen AR, et al. The RANKL/OPG system is activated in inflammatory bowel disease and relates to the state of bone loss. *Gut.* 2005; 54:479–87. [PubMed: 15753532]



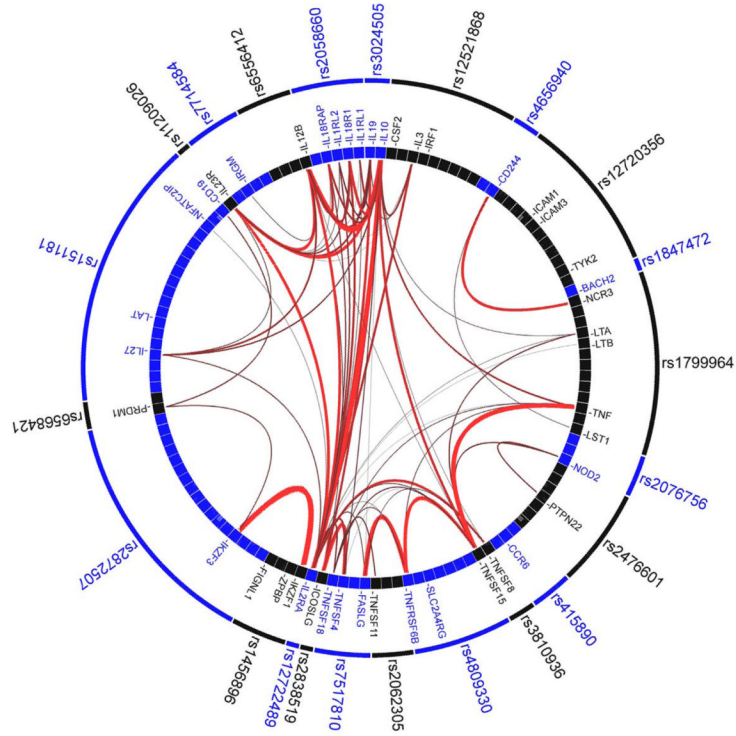
40. Lu L, et al. Role of SMAD and non-SMAD signals in the development of Th17 and regulatory T cells. *J Immunol.* 184:4295–306. [PubMed: 20304828]
41. Ghoreschi K, Laurence A, O’Shea JJ. Janus kinases in immune cell signaling. *Immunol Rev.* 2009; 228:273–87. [PubMed: 19290934]
42. Hazra A, et al. Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nat Genet.* 2008; 40:1160–2. [PubMed: 18776911]
43. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362–7. [PubMed: 19474294]
44. Yu W, Clyne M, Khoury MJ, Gwinn M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics.* 26:145–6. [PubMed: 19864262]

### Box 1

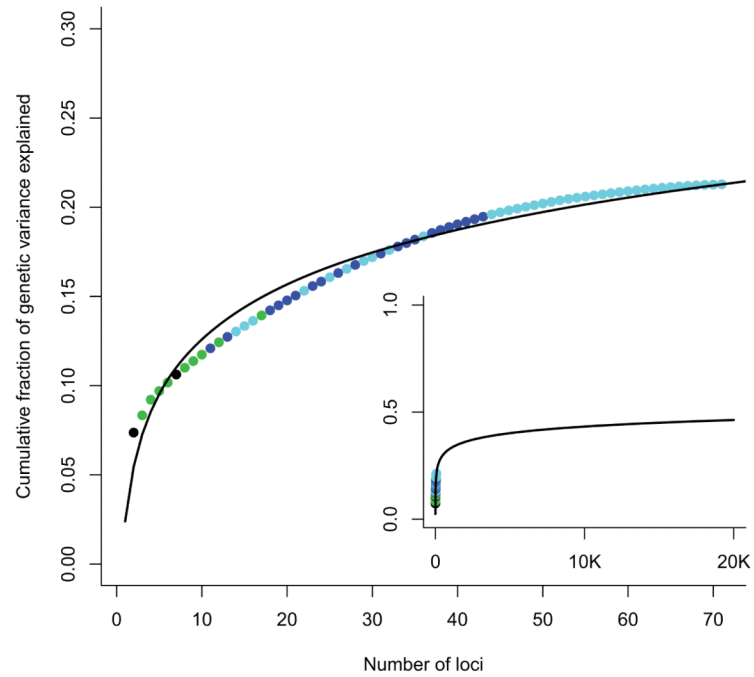
Noteworthy genes within loci newly implicated in Crohn's disease pathogenesis. N.B. Although we highlight these as interesting genes, we do not yet have data to confirm causality

- **VAMP3** (1p36) encodes vesicle-associated membrane protein 3: following bacterial stimulation of TNF- $\alpha$  production within macrophages, VAMP3 interacts with SNARE proteins first on the trans-Golgi network, where TNF- $\alpha$  is taken up, and then on the cell membrane, where TNF- $\alpha$  is released<sup>19</sup>. VAMP3 also plays a role in cell migration and adhesion, by trafficking molecules such as beta1 integrin to the cell surface, and has been implicated in autophagy<sup>20</sup>.
- **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<sup>21</sup> and *Muc1* knockout mice exhibit increased small intestinal damage after *C. jejuni* infection<sup>22</sup>. Secretory carrier membrane protein 3 (*SCAMP3*) regulates EGFR trafficking within endosomal membranes<sup>23</sup>. It is manipulated by intracellular salmonellae to acquire nutrients and influence host immune responses<sup>24</sup>.
- **DENND1B** (1q31) has been recently associated with asthma<sup>25</sup>, and is expressed in dendritic and effector memory T cells. Among other functions it modulates inflammatory signaling pathways, including Th1-Th2 cytokines, through repression of TNFR1 signaling.
- **IL10** (1q32): association with CD follows its recent implication in UC<sup>13</sup> and reporting of mutations in the *IL10* receptors in extreme CD of infancy<sup>26</sup>. Known to inhibit synthesis of pro-inflammatory cytokines within macrophages and Th1 cells, IL10 also suppresses antigen presenting cell activity. Knockdown of IL10 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 is known to determine dynamic regulation of both adaptive and innate immune mechanisms<sup>15,16</sup>.
- **GCKR** (2p23) encodes an inhibitor of glucokinase, with the focal SNPs at this locus also correlating with both fibrinogen and CRP levels<sup>27</sup>.
- **THADA** (2p21) is expressed in small intestine and appears to encode a death receptor-interacting protein suggesting an apoptotic function<sup>28</sup>.
- **ERAP2** (5q15): regulated by NF- $\kappa$ B, this gene encodes one of two human endoplasmic reticulum aminopeptidases, which work in concert to trim peptides for presentation on MHC class I and hence critically affect antigen presentation to T cells<sup>29</sup>. Ankylosing spondylitis is associated with this locus but with a different pattern of associated variants<sup>30</sup>. Given the close clinical relationship between CD and AS, and the strong association of HLA-B27 with the latter but not the former, the divergent association of these closely related molecules is intriguing and will refocus interest on the MHC class I associations in CD.
- **NDFIP1** (5q31): Nedd4 family interacting protein 1 is a membrane protein involved in maintenance of the Golgi complex<sup>31</sup>. It is important for protein trafficking via exosomes and may play a role in rapid sequestration and removal of proteins during stress<sup>32</sup>.

- ***CPEB4*** (5q35) is a regulator of protein translation and cell division, and a transcriptional target of ROR $\gamma$ t. Mouse work suggests that the product of *CPEB4* is the effector by which ROR $\gamma$ t (a key determinant of Th17 cell differentiation) inhibits proliferation of thymocytes<sup>33</sup>.
- ***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<sup>34</sup>.
- ***IL2RA*** (10p15) encodes part of the IL2 receptor complex, thus mediating IL2 signalling in host defence and regulating response to autoantigens by Tregs. The associated variants correlate with differential expression of IL2RA (CD25) on CD4+ naïve and memory T cells<sup>35</sup> possibly affecting Foxp3<sup>+</sup> Treg homeostasis<sup>36</sup>.
- ***FADS 2*** (11q12): fatty acid desaturase 2 is predominantly located in the endoplasmic reticulum. *Fads2* knockout mice develop duodenal and ileocecal ulceration<sup>37</sup>.
- ***TNFSF11*** (13q14) (synonyms RANKL - receptor activator of nuclear factor kappa B; 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<sup>38</sup>; it also regulates osteoclast activity / bone loss. Previous studies have demonstrated increased plasma levels in Crohn's disease<sup>39</sup>.
- ***SMAD3*** (15q22): phosphorylated following TGF- $\beta$  signalling 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- $\beta$ -mediated induction of Foxp3<sup>+</sup> regulatory T cells<sup>40</sup>, with SMAD3 deficiency reciprocally enhancing Th17. Reduced SMAD3 phosphorylation has been observed in IBD, and may impair the immunosuppressive effect of TGF- $\beta$ .
- ***TYK2*** (19p13) encodes tyrosine kinase 2, a member of the JAK-signal transduction family. It is involved in cytokine signaling by IFN- $\gamma$ , IL-12 and IL-23 among others – hence affecting Th1 and Th17 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<sup>41</sup>.
- ***FUT2*** (19q13) encodes alpha-(1,2)fucosyltransferase which regulates expression of the Lewis AB0(H) histo-blood group antigens on the surface of epithelial cells and in body fluids. Strongly associated with Norovirus infection, also with *Helicobacter pylori* infection and serum vitamin B12 levels<sup>14,42</sup>.



**Figure 1. Gene Relationships Across Implicated Loci (GRAIL) pathway analysis**  
 Links between genes at 23 of 71 Crohn’s disease associated loci which scored  $P < 0.01$  using GRAIL. Specifically, of the 71 CD-associated SNPs, 69 are in LD intervals containing or within 50 kb of at least one gene. In total, there are 355 genes implicated by proximity to these 69 SNPs. Each observed CD-association was scored with GRAIL, which takes all genes mapping within CD-associated intervals and evaluates for each whether it is non-randomly linked to the other genes, via word-usage in PubMed abstracts. 23 SNPs shown in the outer circle are  $P < 0.01$  hits - 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 thickness of lines inversely proportional to the probability a literature-based connection would be seen by chance. To accurately assess the statistical significance of this set of connections, we conducted simulations where we selected 1000 sets of 69 SNPs implicating in total 355 genes  $\pm 18$  (5%) (selecting the SNPs randomly and using rejection sampling - only taking lists that implicated the same number of genes). Each of those 1000 sets were scored with GRAIL. The mean number of  $P < 0.01$  hits in a simulated list was 0.91 with a range in the 1000 sets from 0 to 11, suggesting that the likelihood of observing 23 hits with  $P < 0.01$  is far less than 0.1%.



**Figure 2. Cumulative fraction of genetic variance explained by 71 CD loci**

Cumulative fraction of genetic variance explained by the 71 CD loci reported here, ordered from largest to smallest individual contribution. **Black** points were identified pre-GWAS, green in first generation GWAS, blue in an earlier meta-analysis and cyan in this analysis. 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 (e.g. less common and rare alleles with higher penetrance) must also exist.

**Table 1**  
**Association results and *in silico* analyses for all 71 confirmed CD loci**

The upper tier lists new Crohn's disease susceptibility loci (beyond the first international meta-analysis<sup>1</sup>) confirmed in the current study with a genome-wide significant P-value ( $P < 5 \times 10^{-8}$ ) in the combined analysis (discovery + replication sample.) and  $P < 0.05$  on replication. Results for replication are listed for all 39 loci which at the time of study design had not met  $P < 5 \times 10^{-8}$  plus at least nominal evidence of replication in an independent sample set. The 7 loci identified in subsequent studies are identified (see footnotes). The lower tier lists new data for SNPs/loci confirmed in the earlier meta-analysis<sup>1</sup>. Genomic positions were retrieved from NCBI's dbSNP build v130. Linkage disequilibrium (LD) 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. We identified loci previously associated with other relevant traits by a manual literature search and using the NIH catalog of published Genome-wide association studies and the HuGe database (version 1.4) (accessed on May 28<sup>th</sup> 2010) 43,44. UC – ulcerative colitis, AS – ankylosing spondylitis, Ps – psoriasis, PBC – primary biliary cirrhosis, T1D – type 1 diabetes, RA – rheumatoid arthritis, SLE – systemic lupus erythematosus, celiac – celiac disease, T2D – type 2 diabetes, MS – multiple sclerosis, Graves – Graves disease, AD – Alzheimer's disease, CRC – colorectal cancer, CRP – C-reactive protein, ALL – acute lymphocytic leukemia, Lepr. – leprosy, SpA – spondyloarthritis, PD – Parkinson's disease, CAD – coronary artery disease, CLL – chronic lymphocytic leukemia, BD – bone density, B12 – serum vitamin B12 levels, HP – *Helicobacter pylori*, AA – alopecia areata, AITD – autoimmune thyroid disease, BC – breast cancer, BD – Behcet's disease, GC – gastric cancer, Hep.C – hepatitis C susceptibility, SSC – systemic sclerosis, Myelo. – myeloproliferative disease, TB – tuberculosis, GvHD – Graft versus host disease, WBC – white blood cell count, HIES – hyper immunoglobulin E syndrome. Regional association plots for all 71 loci are shown in Supplementary Figure 4 and genotype data is shown in Supplementary Tables 3 and 4.

No.	dbSNP ID	Chr.	Left - right (Mb)	Risk allele - Allele frequency in control population	P-value meta	P-value repl.	P comb.	OR (95% CI); *Loci with evidence of > than 1 independent association	Reported association	Positional candidate genes of interest; some additionally highlighted by ▲ IKG eSNP(s) in LD; GRALL (bold); eQTL (LOD score)
1	rs2797685	1p36	7.66 - 7.89	A - 0.190	$2.69 \times 10^{-10}$	$1.40 \times 10^{-2}$	$7.10 \times 10^{-9}$	1.05 (1.01-1.10)	Celiac	VAMP3
2	rs3180018	1q22	153.24 - 154.39	A - 0.250	$1.29 \times 10^{-9}$	$2.70 \times 10^{-5}$	$2.30 \times 10^{-13}$	1.13 (1.06-1.19)*	T2D, Asthma, PD	SCAMP3, MUC1
3	rs1998598	1q31	195.58 - 196.21	G - 0.302	$4.90 \times 10^{-9}$	$1.60 \times 10^{-2}$	$8.70 \times 10^{-9}$	1.04 (1.00-1.09)	Asthma	DENND1B
4	rs3024505	1q32	204.87 - 205.10	T - 0.157	$8.32 \times 10^{-9}$	$1.50 \times 10^{-7}$	$1.60 \times 10^{-14}$	1.12 (1.07-1.17)	T1D, UC, SLE, BD, Hep. C	IL10, IL19
5	rs13428812	2p23	25.30 - 25.46	G - 0.326	$1.41 \times 10^{-8}$	$5.90 \times 10^{-4}$	$8.50 \times 10^{-10}$	1.06 (1.03-1.10)	CRP, Glucose, TGs	DNMT3A
6	rs780093	2p23	27.24-27.71	T - 0.418	$1.10 \times 10^{-4}$	$3.30 \times 10^{-8}$	$4.70 \times 10^{-11}$	1.15 (1.10-1.21)	T2D, PC	▲ GCKR
7	rs10495903	2p21	43.30 - 43.80	T - 0.129	$7.70 \times 10^{-8}$	$2.90 \times 10^{-8}$	$1.60 \times 10^{-14}$	1.14 (1.09-1.20)*	RA, UC, Celiac	▲ THADA
8 <sup>§</sup>	rs10181042	2p16	60.77 - 61.74	T - 0.420	$6.61 \times 10^{-9}$	N/A	N/A	1.14 (1.09-1.19)	Celiac, Asthma, T1D, HSV	<u>C2orf74 (LOD)</u> , REL
9 <sup>†</sup>	rs2058660	2q12	102.17 - 102.67	G - 0.231	$1.58 \times 10^{-12}$	N/A	N/A	1.19 (1.14-1.26)	Celiac, Asthma, T1D, HSV	IL18RAP, IL12RL2, IL18RI, IL1RL1
10	rs6738825	2q33	197.85 - 198.67	A - 0.473	$1.82 \times 10^{-7}$	$1.60 \times 10^{-3}$	$3.50 \times 10^{-9}$	1.06 (1.02-1.11)	CAD	▲ PLCL1

(a) New Loci meeting Genome-wide significance ( $P$ -value  $< 5.0 \times 10^{-8}$ ) in this study

No.	dbSNP ID	Chr.	Left - right (Mb)	Risk allele - Allele frequency in control population	P-value meta	P-value repl.	P comb.	OR (95% CI); *Loci with evidence of > than 1 independent association	Reported association	Positional candidate genes of interest; some additionally highlighted by ▲ 1KG eSNP(s) in LD; GRAB (bold); eQTL (LOD score)
11	rs7423615	2q37	230.76 - 230.94	T - 0.187	4.57×10 <sup>-9</sup>	7.40×10 <sup>-6</sup>	3.10×10 <sup>-13</sup>	1.12 (1.07-1.18)	CLL	<u>SP140 (8.8)</u>
12	rs13073817	3p24	18.58 - 18.86	A - 0.322	8.20×10 <sup>-7</sup>	1.00×10 <sup>-3</sup>	6.70×10 <sup>-9</sup>	1.08 (1.03-1.13)		<u>TMEM174</u>
13	rs7702331	5q13	72.49 - 72.62	A - 0.600	2.00×10 <sup>-6</sup>	6.40×10 <sup>-7</sup>	5.90×10 <sup>-12</sup>	1.12 (1.07-1.17)		▲ <u>ERAP2</u> , <u>LRAP (47.2)</u>
14	rs2549794	5q15	96.11 - 96.45	C - 0.409	4.47×10 <sup>-11</sup>	2.00×10 <sup>-3</sup>	1.10×10 <sup>-10</sup>	1.05 (1.02-1.09)	AS, PD, T1D	<u>NDPFI</u>
15	rs11167764	5q31	141.39 - 141.62	C - 0.796	1.10×10 <sup>-9</sup>	4.20×10 <sup>-3</sup>	2.00×10 <sup>-9</sup>	1.06 (1.02-1.11)		<u>CFEB4 (6.1)</u>
16	rs359457	5q35	173.15 - 173.47	T - 0.571	5.25×10 <sup>-8</sup>	3.30×10 <sup>-6</sup>	2.50×10 <sup>-12</sup>	1.08 (1.04-1.12)		<u>C6orf85</u>
17	rs17309827	6p25	3.35 - 3.41	T - 0.639	6.16×10 <sup>-7</sup>	3.10×10 <sup>-4</sup>	6.70×10 <sup>-9</sup>	1.10 (1.05-1.16)	T1D, Celiac	<u>BACH2</u>
18	rs1847472	6q15	90.86 - 91.14	G - 0.658	3.63×10 <sup>-6</sup>	1.40×10 <sup>-4</sup>	5.10×10 <sup>-9</sup>	1.07 (1.03-1.11)	RA, Celiac, T1D ↑	<u>TAGAP</u>
19	rs212388	6q25	159.26 - 159.46	G - 0.393	1.41×10 <sup>-7</sup>	2.40×10 <sup>-5</sup>	2.30×10 <sup>-11</sup>	1.10 (1.05-1.14)		
20	rs6651252	8q24	129.56 - 129.67	T - 0.865	2.29×10 <sup>-6</sup>	2.40×10 <sup>-13</sup>	3.90×10 <sup>-18</sup>	1.23 (1.17-1.30)	UC, AS	<u>CARD9 (12.4)</u> , ▲ <u>CARD9</u> , ▲ <u>SNAPC4</u>
21 †	rs4077515	9q34	138.27 - 138.54	T - 0.411	4.37×10 <sup>-19</sup>	1.50×10 <sup>-19</sup>	1.30×10 <sup>-36</sup>	1.18 (1.13-1.22)		
22	rs1272489	10p15	6.07 - 6.21	C - 0.852	8.51×10 <sup>-6</sup>	5.20×10 <sup>-5</sup>	2.90×10 <sup>-9</sup>	1.11 (1.05-1.16)	MS, T1D, Vitiligo, RA, AA, Asthma, AITD	<u>IL2RA</u>
23	rs1819658	10q21	59.50 - 59.81	C - 0.774	1.41×10 <sup>-7</sup>	1.10×10 <sup>-10</sup>	9.10×10 <sup>-17</sup>	1.19 (1.13-1.25)	AD	<u>UBE2D1</u>
24 †	rs1250550	10q22	80.67 - 80.77	G - 0.669	2.00×10 <sup>-10</sup>	7.30×10 <sup>-22</sup>	1.10×10 <sup>-30</sup>	1.19 (1.15-1.23)	Celiac, MS, Vitiligo, BC	<u>ZMIZ1</u>
25	rs102275	11q12	61.28 - 61.44	C - 0.341	7.24×10 <sup>-8</sup>	1.70×10 <sup>-5</sup>	2.30×10 <sup>-11</sup>	1.08 (1.04-1.12)	CAD; Dyslipidemia	<u>FADS1 (5.0)</u>
26	rs694739	11q13	63.58 - 64.05	A - 0.626	3.38×10 <sup>-7</sup>	3.50×10 <sup>-4</sup>	6.00×10 <sup>-10</sup>	1.10 (1.05-1.16)	AA	<u>PRDX3</u> , <u>ESRRA</u>
27	rs2062305	13q14	41.72 - 42.00	G - 0.346	2.00×10 <sup>-6</sup>	5.70×10 <sup>-5</sup>	4.90×10 <sup>-10</sup>	1.10 (1.05-1.15)	BD, RA	<u>TNFSF11</u> , <u>TNFSF11 (5.9)</u>
28	rs4902642	14q24	68.23 - 68.39	G - 0.584	2.00×10 <sup>-7</sup>	4.50×10 <sup>-5</sup>	1.60×10 <sup>-10</sup>	1.07 (1.11-1.04)*	Celiac, T1D	<u>ZFP36L1</u>
29	rs8005161	14q35	87.28 - 87.71	T - 0.119	1.29×10 <sup>-8</sup>	5.90×10 <sup>-11</sup>	4.20×10 <sup>-18</sup>	1.23 (1.16-1.31)*		▲ <u>GALC</u> , ▲ <u>GPR65</u> , <u>GPR65</u>
30	rs17293632	15q22	65.20 - 65.27	T - 0.233	1.41×10 <sup>-13</sup>	2.00×10 <sup>-8</sup>	2.70×10 <sup>-19</sup>	1.12 (1.07-1.16)	CAD, T2D	<u>SMAD3</u>
31 †	rs151181	16p11	28.20 - 28.94	G - 0.386	1.10×10 <sup>-10</sup>	1.20×10 <sup>-3</sup>	1.50×10 <sup>-11</sup>	1.07 (1.03-1.12)	T1D, obesity, Asthma, CRC, SLE, RA	▲ <u>APOB48R</u> , ▲ <u>IL27</u> , ▲ <u>SULT1A2</u> , ▲ <u>SULT1A1</u> , ▲ <u>SH2B1</u> , <u>EHEC (11.3)</u> , <u>IL27</u> , <u>LAT</u> , <u>CD19</u> , <u>NFATC2IP</u>
32 †	rs3091315	17q12	29.51 - 29.70	A - 0.723	1.70×10 <sup>-13</sup>	N/A	N/A	1.20 (1.14-1.26)	HIV resistance	<u>CCL2</u> , <u>CCL7</u>
33	rs12720356	19p13	10.26 - 10.50	G - 0.084	9.20×10 <sup>-10</sup>	1.90×10 <sup>-5</sup>	1.40×10 <sup>-12</sup>	1.12 (1.06-1.19)*	T1D, SLE, MS, HIES	▲ <u>TYK2</u> , <u>TYK2</u> , <u>ICAMI</u> , <u>ICAM3</u>
34 †	rs736289	19q13	38.42 - 38.47	T - 0.612	2.69×10 <sup>-7</sup>	2.00×10 <sup>-3</sup>	8.70×10 <sup>-9</sup>	1.06 (1.02-1.11)		▲ <u>FUT2</u> , ▲ <u>RASIP1</u>
35 †	rs281379	19q13	53.78 - 53.97	A - 0.487	8.60×10 <sup>-10</sup>	5.20×10 <sup>-5</sup>	7.40×10 <sup>-12</sup>	1.07 (1.04-1.11)	B12, Norovirus, HP	▲ <u>RTEL1</u> , <u>TNFRSF6B</u> , <u>SLC2A4RG</u>
36	rs4809330	20q13	61.65 - 61.95	G - 0.709	2.51×10 <sup>-12</sup>	4.60×10 <sup>-5</sup>	2.70×10 <sup>-15</sup>	1.12 (1.06-1.18)	Glioma	

No.	dbSNP ID	Chr.	Left - right (Mb)	Risk allele - Allele frequency in control population	P-value meta	P-value repl.	P comb.	OR (95% CI); *Loci with evidence of > than 1 independent association	Reported association	Positional candidate genes of interest; some additionally highlighted by ▲ 1KG eSNP(s) in LD; GRALL (bold); eQTL (LOD score)
37	rs181359	22q11	20.14 - 20.39	T - 0.203	6.31×10 <sup>-13</sup>	2.30×10 <sup>-6</sup>	4.80×10 <sup>-16</sup>	1.10 (1.06-1.15)	RA, Celiac, SLE, MCV	▲ <b>YDIC</b>
38*	rs713875	22q12	28.23 - 29.00	C - 0.471	5.70×10 <sup>-9</sup>	8.30×10 <sup>-5</sup>	7.30×10 <sup>-12</sup>	1.08 (1.04-1.13)	T1D	<b>MTMR3</b>
39	rs2413583	22q13	38.00 - 38.14	C - 0.830	1.70×10 <sup>-10</sup>	9.50×10 <sup>-18</sup>	1.10×10 <sup>-26</sup>	1.23 (1.17-1.29)		<b>MAP3K7IP1</b>
<b>(b) Loci that met genome-wide significance (P-value &lt; 5.0×10<sup>-8</sup>) in Barrett et al.</b>										
1	rs11209026	1p31	67.13 - 67.54	G - 0.932	1.00×10 <sup>-64</sup>	N/A	N/A	2.66 (2.36-3.00)	UC, AS, Ps, PBC, GC, BD	▲ <b>IL23R, IL23R</b>
2	rs2476601	1p13	113.66 - 114.42	G - 0.907	4.47×10 <sup>-9</sup>	N/A	N/A	1.26 (1.17-1.37)	T1D, RA, SLE, Ps, Vitiligo, ATD	▲ <b>PTPN22, PTPN22</b>
3	rs4656940	1q23	158.96 - 159.20	A - 0.801	6.17×10 <sup>-7</sup>	N/A	N/A	1.15 (1.09-1.21)	SLE, RA	<b>CD244, CD244, ITLN1</b>
4	rs7517810	1q24	170.92 - 171.21	T - 0.246	1.51×10 <sup>-15</sup>	N/A	N/A	1.22 (1.16-1.28)	HepC, SLE, SSC, T2D	<b>TNFSF18, TNFSF4, FASLG</b>
5	rs7554511	1q32	199.11 - 199.32	C - 0.726	1.58×10 <sup>-7</sup>	N/A	N/A	1.14 (1.08-1.19)	UC, celiac, MS	▲ <b>C1orf106, KIF21B</b>
6	rs3792109	2q37	233.81 - 234.23	A - 0.529	6.76×10 <sup>-41</sup>	N/A	N/A	1.34 (1.29-1.40)	UC	▲ <b>ATG16L1</b>
7	rs3197999	3p21	48.16 - 51.73	A - 0.297	6.17×10 <sup>-17</sup>	N/A	N/A	1.22 (1.16-1.27)	UC	▲ <b>MST1, GPXI, BSN</b>
8	rs11742570	5p13	39.88 - 41.00	C - 0.606	7.08×10 <sup>-36</sup>	N/A	N/A	1.33 (1.27-1.39)	MS	<b>PTGER4</b>
9	rs12521868	5q31	129.41 - 132.05	T - 0.422	1.41×10 <sup>-20</sup>	N/A	N/A	1.23 (1.18-1.28)	Ps, Fibinogen, Asthma, TB, UC	▲ <b>SLC22A4, SLC22A5, CSF2, IL3</b>
10	rs7714584	5q33	150.01 - 150.38	G - 0.088	7.76×10 <sup>-19</sup>	N/A	N/A	1.37 (1.28-1.47)	TB	<b>IRGM</b>
11	rs6556412	5q33	158.43 - 158.88	A - 0.332	5.37×10 <sup>-14</sup>	N/A	N/A	1.18 (1.13-1.24)	Ps, SLE, Malaria, Asthma	<b>IL12B</b>
12	rs6908425	6p22	20.60 - 21.25	C - 0.784	1.41×10 <sup>-8</sup>	N/A	N/A	1.17 (1.11-1.23)	T2D, Ps, UC	<b>CDKAL1</b>
13	rs1799964	6p21	31.49 - 32.98	C - 0.209	3.98×10 <sup>-11</sup>	N/A	N/A	1.19 (1.13-1.25)	Multiple including UC	▲ <b>MCCD1, LTA, HL A-DQA2, TNF, LST1, LTB, LTA, NCR3</b>
14	rs6568421	6q21	106.50 - 106.67	G - 0.301	4.37×10 <sup>-8</sup>	N/A	N/A	1.13 (1.07-1.18)*	SLE, RA	<b>PRDMI</b>
15	rs415890	6q27	167.26 - 167.47	C - 0.522	2.51×10 <sup>-12</sup>	N/A	N/A	1.17 (1.12-1.22)	RA, Graves	<b>CCR6</b>
16	rs1456896	7p12	50.22 - 50.34	T - 0.69	1.20×10 <sup>-8</sup>	N/A	N/A	1.14 (1.09-1.20)	AD, SLE, MCV, ALL	<b>IKZF1, ZPBP, FIGLN1</b>
17	rs4871611	8q24	126.54 - 126.65	A - 0.609	1.51×10 <sup>-12</sup>	N/A	N/A	1.17 (1.12-1.23)	UC, Myelo.	<b>JAK2</b>
18	rs10758669	9p24	4.93 - 5.29	C - 0.349	1.00×10 <sup>-13</sup>	N/A	N/A	1.18 (1.13-1.23)	UC, Lepr., SpA	<b>TNFSF15, TNFSF8</b>
19	rs3810936	9q32	116.47 - 116.74	C - 0.682	1.00×10 <sup>-15</sup>	N/A	N/A	1.21 (1.15-1.27)	UC	<b>CREM (6.4)</b>
20	rs12242110	10p11	35.22 - 35.94	G - 0.315	1.10×10 <sup>-9</sup>	N/A	N/A	1.15 (1.10-1.20)	UC	<b>ZNF363</b>
21	rs10761659	10q21	63.97 - 64.43	G - 0.538	4.37×10 <sup>-22</sup>	N/A	N/A	1.23 (1.18-1.29)	BC	<b>NKX2-3</b>
22	rs4409764	10q24	101.26 - 101.33	T - 0.492	2.29×10 <sup>-20</sup>	N/A	N/A	1.22 (1.17-1.27)	UC	<b>C1orf30</b>
23	rs7927997	11q13	75.70 - 76.04	T - 0.389	5.62×10 <sup>-13</sup>	N/A	N/A	1.17 (1.12-1.22)	Atopy	



No.	dbSNP ID	Chr.	Left - right (Mb)	Risk allele - Allele frequency in control population	P-value meta	P-value repl.	P comb.	OR (95% CI); *Loci with evidence of > than 1 independent association	Reported association	Positional candidate genes of interest; some additionally highlighted by ▲ 1KG eSNP(s) in LD; GRAIL (bold) ; eQTL (LOD score)
24	rs11564258	12q12	38.42 - 39.31	A - 0.025	$6.17 \times 10^{-21}$	N/A	N/A	1.74 (1.55-1.95)	PD, Lepr.	▲ <i>MUC19</i> , <i>LR6K2</i>
25	rs3764147	13q14	43.13 - 43.54	G - 0.245	$1.41 \times 10^{-10}$	N/A	N/A	1.17 (1.12-1.23)	Lepr.	▲ <i>CI3orf31</i>
26	rs2076756	16q12	49.02 - 49.41	G - 0.26	$3.98 \times 10^{-69}$	N/A	N/A	1.53 (1.46-1.60)	Lepr., Atopy, Blau, CvHD	<i>NOD2</i>
27	rs2872507	17q21	34.62 - 35.51	A - 0.458	$1.51 \times 10^{-9}$	N/A	N/A	1.14 (1.09-1.19)	Asthma, UC, PBC, T1D, RA, WBC	▲ <i>GSMDL</i> , ▲ <i>ZBP2</i> , <i>ORMDL3</i> (20.3), <i>IKZF3</i>
28	rs11871801	17q21	37.57 - 38.25	A - 0.756	$2.51 \times 10^{-8}$	N/A	N/A	1.15 (1.10-1.21)	MS <sup>↓</sup> , obesity, HIES	▲ <i>MLX</i> , <i>STAT3</i>
29	rs1893217	18p11	12.73 - 12.92	G - 0.153	$1.29 \times 10^{-14}$	N/A	N/A	1.25 (1.18-1.32)	T1D <sup>↓</sup> , celiac	<i>PTPN2</i>
30	rs740495	19p13	1.04 - 1.13	G - 0.247	$8.13 \times 10^{-12}$	N/A	N/A	1.16 (1.10-1.21)	UC	<i>GPX4</i> , <i>SBN2</i>
31	rs1736020	21q21	15.62 - 15.77	C - 0.579	$9.33 \times 10^{-12}$	N/A	N/A	1.16 (1.11-1.21)	Celiac, UC	<i>ICOSLG</i>
32	rs2858519	21q22	44.41 - 44.52	G - 0.391	$2.09 \times 10^{-14}$	N/A	N/A	1.18 (1.13-1.23)		

<sup>†</sup>*IL18RAP* and *CARD9* – association reported by Zhemakova *et al.*<sup>7</sup> but not previously at genome-wide significance

<sup>‡</sup>loci previously reported at genome-wide significance in GWAS studies published subsequent to design of the current replication experiment<sup>5,6</sup>

<sup>§</sup>loci that showed suggestive association and replication in Barrett *et al.*<sup>1</sup> but not previously at genome-wide significance

<sup>↓</sup> association in the opposite direction in different traits

We identified genes of interest based on a variety of *in silico* techniques (see text for more details) – identified as ▲ for coding SNPs identified from 1000 Genomes Project or HapMap in linkage disequilibrium with our most associated SNP (for dbSNP IDs see Supplementary Table 5), bold text for GRAIL connectivity and underlined text for presence of an eQTL effect with LOD 5.0 (for details see Supplementary Results).

Loci tagged by rs4656940 and rs7554511 previously replicated strongly (0.00048 and  $2.3 \times 10^{-6}$  respectively in Barrett *et al.*<sup>1</sup>) and still pass genome-wide significance on combined analysis.