



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility

Citation for published version:

Wellcome Trust Case Control Consor, Parkes, M, Barrett, JC, Prescott, NJ, Tremelling, M, Anderson, CA, Fisher, SA, Roberts, RG, Nimmo, ER, Cummings, FR, Soars, D, Drummond, H, Lees, CW, Khawaja, SA, Bagnall, R, Burke, DA, Todhunter, CE, Ahmad, T, Onnie, CM, McArdle, W, Strachan, D, Bethel, G, Bryan, C, Lewis, CM, Deloukas, P, Forbes, A, Sanderson, J, Jewell, DP, Satsangi, J, Mansfield, JC, Cardon, L & Mathew, CG 2007, 'Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility', *Nature Genetics*, vol. 39, no. 7, pp. 830-832.
<https://doi.org/10.1038/ng2061>

Digital Object Identifier (DOI):

[10.1038/ng2061](https://doi.org/10.1038/ng2061)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature Genetics

Publisher Rights Statement:

Published in final edited form as:
Nat Genet. 2007 July ; 39(7): 830–832. doi:10.1038/ng2061.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Published in final edited form as:

Nat Genet. 2007 July ; 39(7): 830–832. doi:10.1038/ng2061.

Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn disease susceptibility

Miles Parkes^{*1}, Jeffrey C Barrett^{*2}, Natalie Prescott^{*3}, Mark Tremelling¹, Carl A Anderson², Sheila A Fisher³, Roland G Roberts³, Elaine R Nimmo⁴, Fraser R Cummings⁵, Dianne Soars³, Hazel Drummond⁴, Charlie W Lees⁴, Saud A Khawaja³, Richard Bagnall³, Denis A Burke⁶, Catherine E Todhunter⁶, Tariq Ahmad⁵, Clive M Onnie³, Wendy McArdle⁷, David Strachan⁸, Graeme Bethel⁹, Claire Bryan⁹, Panos Deloukas⁹, Alastair Forbes¹⁰, Jeremy Sanderson¹¹, Derek P Jewell⁵, Jack Satsangi⁴, John C Mansfield⁶, the Wellcome Trust Case Control Consortium¹², Lon Cardon², and Christopher G Mathew³

¹IBD research group, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 2QQ, UK

²Bioinformatics and Statistical Genetics, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

³Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Tower, Guy's Hospital, London, SE1 9RT, UK

⁴Gastrointestinal Unit, Division of Medical Sciences, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU UK

⁵Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford, OX2 6HE, UK

⁶Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK

⁷ALSPAC, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK

⁸Division of Community Health Services, St George's University of London, Cranmer Terrace, London SW17 0RE, UK

⁹The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

¹⁰Institute for Digestive Diseases, University College London Hospitals Trust, London, NW1 2BU, UK

¹¹Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, SE1 7EH, UK

¹²Wellcome Trust Case Control Consortium – authors as listed [Supplementary material]

Abstract

A genome-wide association scan in Crohn disease by the Wellcome Trust Case Control Consortium1 detected strong association at 6 novel loci. We tested 37 SNPs from these and other loci for association in an independent case control sample. Replication was obtained for the *IRGM* gene on chromosome 5q33.1 which induces autophagy (replication $P = 6.6 \times 10^{-4}$, combined $P = 2.1 \times 10^{-10}$), and for 9 other loci including *NKX2-3* and gene deserts on chromosomes 1q and 5p13.

*Joint first authors

Crohn disease (CD) is a common form of chronic inflammatory bowel disease. Established CD susceptibility genes *NOD2* (*CARD15*), *IL23R* and *ATG16L1* 2-5 showed strong evidence of association in the Wellcome Trust Case Control Consortium (WTCCC) genome-wide scan of 1748 CD cases and 2938 controls genotyped using the Affymetrix 500K chip. Six other loci also showed highly significant association. Although satisfying stringent statistical thresholds for significance ($P < 5 \times 10^{-7}$), replication in independent panels represents a key validation step.

We followed up 37 SNPs from 31 distinct loci associated at $P < 10^{-5}$ on initial analysis of the WTCCC dataset. Support for some of these markers diminished in the final WTCCC analysis after extensive data filtering¹. Two markers were selected for each locus where low LD between associated SNPs in areas of unbroken LD suggested distinct causal variants. SNPs were genotyped in a new panel of 1182 Caucasian CD cases using TaqMan assays (Supplementary Methods and Supplementary Table 1 online). Concordance with Affymetrix data was 99.7%, based on genotyping 96 WTCCC cases on both platforms. To target SNPs for replication testing and limit unnecessary genotyping, preliminary comparison was made between allele frequencies in the new CD panel and 5746 non-autoimmune WTCCC cases (bipolar disorder, coronary artery disease and hypertension). The latter represent a powerful resource of population allele frequencies¹ assuming no overlap with CD aetiology. Twelve SNPs showed a difference in allele frequency between these two groups (Table 1), while for 25 other markers the allele frequencies converged (Supplementary Table 2). To exclude possible bias introduced by this approach, the 12 markers implicated in this preliminary comparison were then genotyped in 2024 independent population controls from the 1958 British Birth Cohort to test formally for association. Results are presented in Table 1, with odds ratios estimated from replication samples only.

Among novel CD loci achieving genome-wide significance in the WTCCC scan¹ the strongest replication adjacent to a known gene was for SNPs rs13361189 and rs4958847 ($P_{\text{rep}}=6.6 \times 10^{-4}$ and 3.1×10^{-4} respectively) immediately flanking the *IRGM* gene [MIM 608212] on chromosome 5q33.1. Association in the combined panels of 2930 CD cases and 4962 controls was highly significant ($P_{\text{comb}}= 2.1 \times 10^{-10}$ and 3.8×10^{-9} respectively) (Figure 1). The long coding exon of the human *IRGM* gene encodes a 20 kD protein of 181 amino acid residues.^{6,7} Since none of the associated SNPs are known to be functional, we re-sequenced this coding exon of *IRGM* and the 4 small putative downstream exons in 48 CD cases homozygous or heterozygous for risk alleles (Supplementary Methods). We detected two novel non-synonymous sequence variants, c.51G>C (p.Glu17Asp) and c.281C>A (p.Thr94Lys) and an exonic synonymous SNP, c.313T>C (rs10065172, p.L105). These SNPs were then genotyped in 769 unselected CD cases from our study and 705 controls. Only the silent 313T>C variant was associated with CD ($P=0.008$), and was in near perfect LD with SNP rs13361189 ($r^2=0.91$). Sequencing of the *IRGM* coding region in a further 100 unselected CD cases and 100 controls did not detect additional variants. These results strongly suggest that the causal variant(s) do not change the amino acid sequence of the *IRGM* protein. They may lie in regulatory sequences, in LD with the associated SNPs, and affect *IRGM* expression; alternatively, the exonic (c.313T>C) SNP itself might affect transcript splicing or the rate of translation of the protein. Analysis of *IRGM* expression by PCR screening against uncloned cDNA showed transcription in several tissues, including colon, small bowel, peripheral blood leukocytes and the U937 monocytic cell line (data not shown).

IRGM belongs to the p47 immunity-related GTPase family. LRG-47 (*irgm1*), its mouse homologue, critically controls intracellular pathogens by autophagy, and *LRG-47*^{-/-} mice showed markedly increased susceptibility to *Toxoplasma gondii* and *Listeria*

monocytogenes.⁸ Consistent with this, *IRGM* induces autophagy and thereby control of intracellular *M. tuberculosis* in human macrophages.⁷

Three other associations with genome-wide significance in the WTCCC scan produced strong evidence of replication ($P = 0.01$), two of which are novel. The strongest was SNP rs9292777 ($P_{\text{rep}}=2.9 \times 10^{-7}$; $P_{\text{comb}}=3.2 \times 10^{-18}$) which maps to a 1.2 Mb gene desert on chromosome 5p13.1 recently associated with CD.⁹ The most significant novel association was SNP rs10883365 ($P_{\text{rep}}=0.0037$, $P_{\text{comb}}=3.7 \times 10^{-10}$), which maps within the *NKX2-3* gene (NK2 transcription factor related, locus 3) on chromosome 10q24.2. *Nkx2.3*-deficient mice develop splenic and gut-associated lymphoid tissue abnormalities with disordered segregation of T- and B- cells.¹⁰ The second novel locus at rs9858542 ($P_{\text{rep}}=0.010$, $P_{\text{comb}}=4.9 \times 10^{-8}$) on chromosome 3p21 is a 1 Mb region of high LD that contains over 20 genes, including *MST1* (macrophage stimulating 1), encoding a protein which induces phagocytosis by resident peritoneal macrophages.

The modest evidence of replication for SNP rs2542151 ($P = 0.048$) at the *PTPN2* locus (protein tyrosine phosphatase, non-receptor type 2) on chromosome 18p11 ($P_{\text{comb}} = 3.2 \times 10^{-8}$) is of interest since *PTPN2* encodes a T cell protein tyrosine phosphatase, a key negative regulator of inflammation and is also associated with Type 1 Diabetes.¹¹ Allele frequencies for rs10761659 on chromosome 10q21, strongly associated in the WTCCC scan, were similar in replication CD cases and population controls (SOM table 2), but association in this intergenic region was recently detected in a North American whole-genome CD scan.¹²

Allele frequencies for most of the markers from the 25 other loci studied that did not achieve genome-wide significance in the WTCCC scan but had an initial $P < 10^{-5}$ converged with control frequencies in the CD replication panel (Supplementary Table 2). Five of these loci, however, provided evidence of replication, thus indicating that a systematic follow-up of all CD hits in the 10^{-5} - 10^{-7} range is likely to yield a number of additional CD susceptibility loci. Four of these loci are novel putative CD risk loci; intriguingly they include two gene deserts, both on chromosome 1q. SNP rs10801047 ($P_{\text{rep}}=4.8 \times 10^{-5}$, $P_{\text{comb}}=2.8 \times 10^{-8}$) is located in a gene desert of 1.7Mb on chromosome 1q31.2, and rs12035082 ($P_{\text{rep}}=0.017$, $P_{\text{comb}}=2.1 \times 10^{-7}$) maps to chromosome 1q24 (the association signal does not extend to nearest gene, *TNFSF18*). Although previous studies did not implicate *IL12B* (interleukin 12B),^{4,13} modest WTCCC association at rs6887695 from this locus replicated here ($P_{\text{rep}}=8.4 \times 10^{-5}$, $P_{\text{comb}}=9.2 \times 10^{-6}$) and warrants follow up, particularly given its association with psoriasis¹⁴ and the common association of IL23R with psoriasis and CD.^{4,14} SNP rs10077785 ($P_{\text{rep}}=0.008$) tags the known *IBD5* locus¹⁵ on chromosome 5q31 in which, despite extensive resequencing, disease-causing coding variants have yet to be confirmed. Finally, rs2836754 ($P_{\text{rep}}=0.012$, $P_{\text{comb}}=5.5 \times 10^{-7}$) maps to an intron in *FLJ45139*, a gene of unknown function on chromosome 21 that is expressed in bone marrow and colon.

We performed 'within-cases' interaction analysis between known CD susceptibility variants and replicating SNPs. For *CARD15*, rs2066844, rs2066845 and rs2066847 were tested independently or in combination classifying individuals as homozygous wild type, heterozygous or compound heterozygote/rare-allele homozygote. No significant interactions were seen. Testing CD associated variants rs11805303 in *IL23R* and rs10210302 in *ATG16L1* also showed no interactions. Thus risk loci in CD appear to act independently.

We examined six sub-phenotypes of CD and two modifiers for differential effects within cases at markers studied. Cases were classified as affected or unaffected for disease site ('pure ileal', 'pure colorectal', 'perianal involvement') and behaviour ('inflammatory', 'stenosing' and 'penetrating'); and by age of onset and smoking status (Supplementary

Table 3). Inflammatory behaviour appeared less strongly associated with rs10210302 (in *ATG16L1*, $P=0.0003$, Bonferroni $P=0.010$) than stenosing or penetrating behaviour. No further significant differential associations were seen, although ‘pure ileal’ tends to associate more strongly with most replicating SNPs than ‘pure colorectal’.

This study confirms all four previously unpublished CD loci which achieved genome-wide significance in the WTCCC scan1, and the recently reported 5p13 gene desert9, as well as implicating four additional novel risk loci. These data, combined with the association in the WTCCC scan and the North American study of an intergenic region on chromosome 10q211,12, underline the value of the whole-genome scan approach. Taken together, the genetic evidence regarding *IRGM*, *ATG16L1* (converging on autophagy pathways), *CARD15* and *IL23R* strongly implicates defects in the early immune response, particularly innate immune pathways and the handling of intracellular bacteria, in the pathogenesis of CD. Whether the latter is restricted to a single pathogen or, more likely, a class effect of sub-pathogenic bacteria relying on defective host innate immunity for survival awaits further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge use of DNA from the 1958 British Birth Cohort collection (R. Jones, S. Ring, W. McArdle and M. Pembrey), funded by the Medical Research Council (grant G0000934) and The Wellcome Trust (grant 068545/Z/02). We also acknowledge the National Association for Colitis and Crohn's disease and the Wellcome Trust for supporting the case DNA collections, and UCB Pharma who supported this study with an unrestricted educational grant. We thank Dr Daniel Kelberman, all subjects who contributed samples, and consultants and nursing staff across the UK who helped with recruitment of study subjects.

References

1. Wellcome Trust Case Control Consortium. *Nature*. 2007 in press.
2. Hugot JP, et al. *Nature*. 2001; 411:299–603. [PubMed: 11385576]
3. Ogura Y, et al. *Nature*. 2001; 411:603–606. [PubMed: 11385577]
4. Duerr RH, et al. *Science*. 2006; 314:1461–1463. [PubMed: 17068223]
5. Hampe J, et al. *Nat. Genet.* 2007; 39:207–211. [PubMed: 17200669]
6. Bekpen C, et al. *Genome Biol.* 2005; 6:R92.1–92.18. [PubMed: 16277747]
7. Singh SB, et al. *Science*. 2006; 313:1438–1441. [PubMed: 16888103]
8. Collazo CM, et al. *J. Exp. Med.* 2001; 194:181–187. [PubMed: 11457893]
9. Libioulle C, et al. *PLOS Genet.* 2007; 3:e58. [Epub ahead of print]. [PubMed: 17447842]
10. Pabst O, et al. *EMBO J.* 2000; 19:2015–2023. [PubMed: 10790368]
11. Todd J, et al. *Nat. Genet.* 2007 this issue.
12. Rioux JD, et al. *Nat. Genet.* 2007; 39:596–604. [PubMed: 17435756]
13. Zwiars A, et al. *Genes Immun.* 2004; 5:675–677. [PubMed: 15483662]
14. Cargill M, et al. *Am. J. Hum. Genet.* 2007; 80:273–290. [PubMed: 17236132]
15. Rioux JD, et al. *Nat. Genet.* 2001; 29:223–228. [PubMed: 11586304]

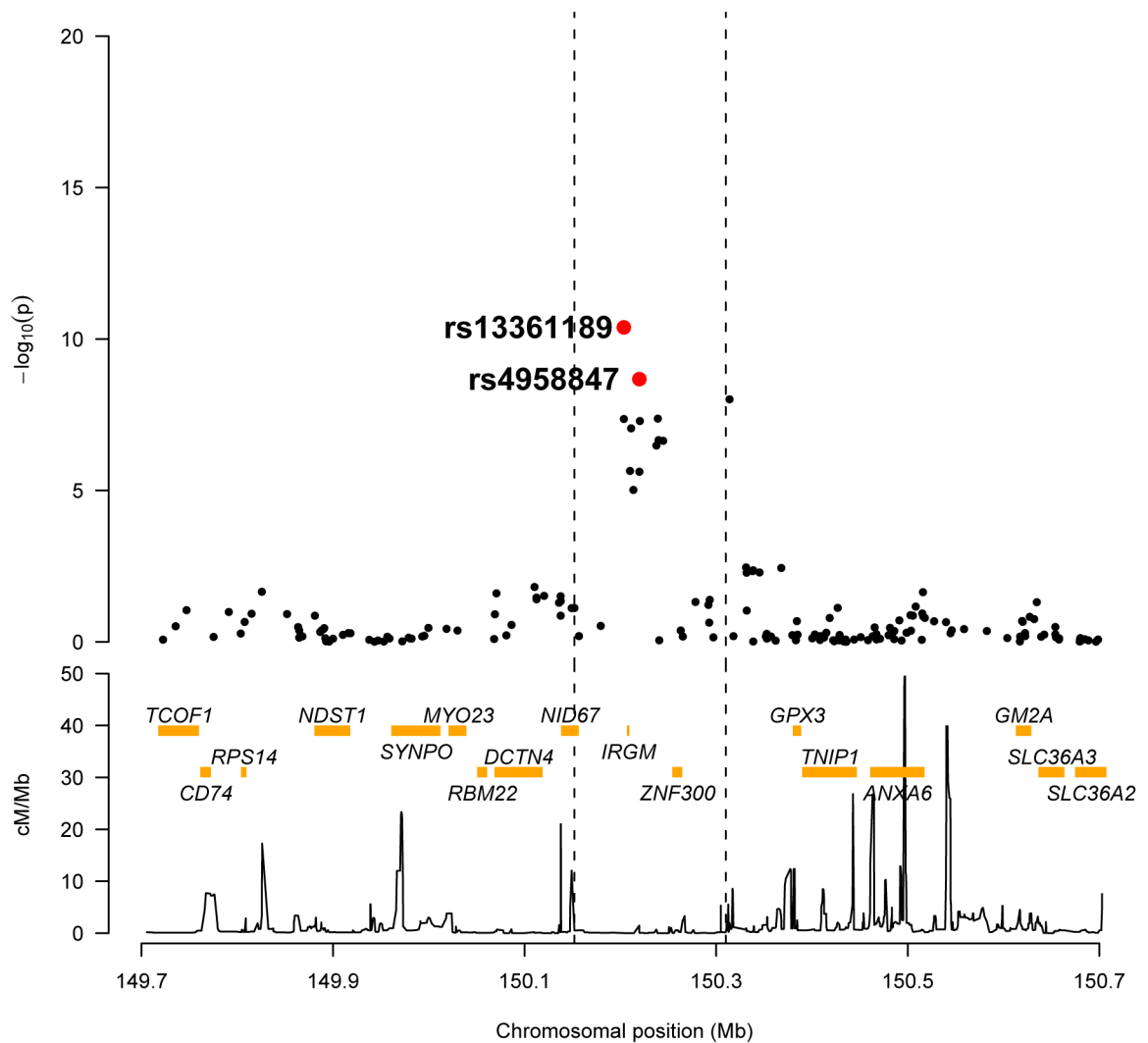


Figure 1.

Associated region at the gene *IRGM*. Strength of association of Affymetrix 500K SNPs in the region, 1 with combined p-values shown in red for replicated SNPs. Recombination map showing how the signals are clearly demarcated by recombination hotspots. Nearby genes are shown in yellow.

Table 1
12 SNPs genotyped in 1182 Crohn disease cases and 2024 controls.

Allele frequencies and association results for 12 SNPs genotyped in 1182 cases from the replication Crohn's disease panel (RCD) compared to frequencies for these markers in the original WTCCC Crohn Disease samples (WTCD), WTCCC Unaffected control samples (WTU1), WTCCC Unaffected case samples (WTU2) (from coronary artery disease, hypertension and bipolar disorder cohorts)¹ and 2024 replication unaffected samples (RU) from the 1958 British Birth Cohort. Quoted p-values are for the final, clean WTCCC dataset. 1 Some SNPs with $p > 10^{-5}$ in this table showed stronger evidence in the preliminary data from which we fast-tracked them for follow-up. Odds ratios are estimated from the replication samples. Regions are shown in alternating rows of white and grey (adjacent rows of the same colour are from the same region).

SNP	Chr	Region (Mb)	Gene	WTCD	RCD	WTU1	WTU2	RU	WTCCC p	Rep p	Combined p	OR (95%CI)
rs12035082	1q24	169.53 - 169.67		0.438	0.429	0.389	0.396	0.398	2.15×10^{-6}	0.01692	2.07×10^{-7}	1.14 (1.02 - 1.27)
rs10801047	1q31	188.17 - 188.32		0.09	0.108	0.067	0.072	0.076	8.15×10^{-5}	4.78×10^{-5}	2.83×10^{-8}	1.47 (1.22 - 1.76)
rs9858542	3p21	49.3 - 49.87	Many	0.331	0.324	0.282	0.289	0.291	7.71×10^{-7}	0.01058	4.91×10^{-8}	1.17 (1.04 - 1.31)
rs17234657	5p13	40.32 - 40.66		0.181	0.151	0.125	0.132	0.132	2.13×10^{-13}	0.05305	3.80×10^{-12}	1.16 (1.00 - 1.35)
rs9292777	5p13	40.32 - 40.66		0.32	0.332	0.394	0.391	0.4	1.89×10^{-12}	2.88×10^{-7}	3.20×10^{-18}	1.34 (1.20 - 1.50)
rs10077785	5q23	131.40 - 131.90	<i>IBD5</i>	0.192	0.202	0.234	0.241	0.232	1.81×10^{-6}	0.008086	7.53×10^{-8}	1.19 (1.05 - 1.36)
rs13361189	5q33	150.15 - 150.31	<i>IRGM</i>	0.098	0.108	0.067	0.079	0.081	4.38×10^{-8}	0.000664	2.09×10^{-10}	1.38 (1.15 - 1.66)
rs4958847	5q33	150.15 - 150.31	<i>IRGM</i>	0.147	0.158	0.113	0.123	0.125	2.41×10^{-6}	0.00031	3.77×10^{-9}	1.36 (1.17 - 1.59)
rs6887695	5q33	158.61 - 158.76	<i>IL12B</i>	0.343	0.369	0.318	0.314	0.318	0.009871	8.42×10^{-5}	9.21×10^{-6}	1.26 (1.12 - 1.41)
rs10883365	10q24	101.26 - 101.32	<i>NKX2-3</i>	0.537	0.523	0.477	0.489	0.482	1.41×10^{-8}	0.00373	3.71×10^{-10}	1.18 (1.05 - 1.32)
rs2542151	18p11	12.76 - 12.91	<i>PTPN2</i>	0.208	0.196	0.163	0.168	0.175	4.56×10^{-8}	0.0479	3.16×10^{-8}	1.15 (1.00 - 1.32)
rs2836754	21q22	39.21 - 39.24		0.399	0.387	0.353	0.35	0.354	1.07×10^{-5}	0.01224	5.48×10^{-7}	1.15 (1.03 - 1.28)