SUPPLEMENTARY METHODS ONLINE

Participants

1182 CD patients of European descent distinct from the WTCCC panel gave written consent and blood for DNA extraction following Research Ethics Committee approval (Cambridge, Oxford, Newcastle, London and Edinburgh local research ethics committees). Standard endoscopic, radiological and histological diagnostic criteria were applied and phenotypic details obtained by case notes review.[1] The Montreal classification[2] (age of diagnosis, disease location and behaviour) was used for CD sub-phenotype (Table 1). Only one member of multiply affected families was included. The 2024 ethnically matched controls came from the 1958 British Birth Cohort which includes all subjects born in one week of March 1958 in England, Scotland and Wales.[3] There was no overlap with the 1500 samples genotyped as controls in the WTCCC.

SNP Genotyping and Sequencing

37 SNPs were genotyped using validated Taqman assays (rs1931047 was custom designed) (ABI). The human *IRGM* gene contains one large exon encoding all 181 amino acids of the major isoform; 4 small exons extending 50 kb 3' to this have been proposed to encode additional isoforms with short extensions of the C-terminal end of the protein.[4] The large coding exon 1 was screened by direct re-sequencing of two overlapping amplimers (1.1 and 1.2) with ABI BigDye v3.1 dye- terminator chemistry and analysed on an ABI 3730XL Sequencer as were all other downstream exons 2-5 (Supplementary Table 5). Identified *IRGM* variants were genotyped by direct re-sequencing as above.

Statistical analysis and quality control

Using data from a preliminary analysis of the WTCCC GWA scan,[5] 37 SNPs showing allelic association at $p < 10^{-5}$ using the Cochran-Armitage trend test were selected for follow-up. Supplementary Table 4 shows missing data rates and HWE p values for the 37 SNPs. Twelve SNPs showed consistent

case/control allele frequencies in additional genotyped cases and three non-autoimmune samples from the WTCCC. Using PLINK[6] (http://pngu.mgh.harvard.edu/~purcell/plink/) these 12 SNPs were evaluated using the same test, both alone and combined with the WTCCC dataset.

For a multiplicative effect, with a perfect proxy to the locus the power to detect the effect with the 'replication' sample set of 1182 cases and 2024 controls at $p < 0.01$ was as follows: GRR = 1.2, MAF = 0.1: 38%; GRR = 1.2, MAF = 0.2: 66%; GRR = 1.3, MAF = 0.1: 78%; GRR = 1.3, MAF = 0.2: 96%.

Case-only interaction tests were performed for replicating SNPs against *CARD15, IL23R* and *ATG16L1* using PLINK. We tested for association at these loci within-cases in the original WTCCC sample for six sub-phenotypes and two modifying effects. Significance would indicate the variant preferentially affects one subphenotype more than the others.

Human *IRGM*: Refseq XM_293893

References

1. Lennard-Jones, J.E.. *Scan. J. Gastroenterol.* Suppl. **170**, 2-6 (1989).
2. Silverberg, M.S. *et al. Can. J. Gastroenterol.* **19** Suppl A,5-36 (2005).
3. Power, C. and Elliot, J. *Int. J. Epidemiol.* **35**, 34-41 (2006).
4. Bekpen, C. *et al. Genome Biol.* **6**, R92.1-92.18 (2005).
5. Wellcome Trust Case Control Consortium, *Nature*, in press (2007).
6. Purcell, S. *et al. Am. J. Hum. Genet.,* in press (2007).