Background & Aims: Identification of inflammatory bowel disease (IBD) susceptibility genes is key to understanding pathogenic mechanisms. Recently, the North American IBD Genetics Consortium provided compelling evidence for an association between ileal Crohn’s disease (CD) and the IL23R gene using genome-wide association scanning. External replication is a priority, both to confirm this finding in other populations and to validate this new technique. We tested for association between IL23R and IBD in a large independent UK panel to determine the size of the effect and explore subphenotype correlation and interaction with CARD15.

Methods: Eight single nucleotide polymorphism markers in IL23R tested in the North American study were genotyped in 1902 cases of Crohn’s disease (CD), 975 cases of ulcerative colitis (UC), and 1345 controls using MassARRAY. Data were analyzed using χ² statistics, and subgroup association was sought.

Results: A highly significant association with CD was observed, with the strongest signal at coding variant Arg381Gln (allele frequency, 2.5% in CD vs 6.2% in controls [P = 1.1 × 10⁻¹²]; odds ratio, 0.38; 95% confidence interval, 0.29–0.50). A weaker effect was seen in UC (allele frequency, 4.6%; odds ratio, 0.73; 95% confidence interval, 0.55–0.96). Analysis accounting for Arg381Gln suggested that other loci within IL23R also influence IBD susceptibility. Within CD, there were no subphenotype associations or evidence of interaction with CARD15.

Conclusions: This study shows an association between IL23R and all subphenotypes of CD with a smaller effect on UC. This extends the findings of the North American study, providing clear evidence that genome-wide association scanning can successfully identify true complex disease genes.

It is widely recognized that knowledge regarding the genetic basis of inflammatory bowel disease (IBD) and other complex diseases will provide key insights into pathogenic mechanisms. It is this fact that has spurred efforts to identify disease susceptibility genes. Of the many complex diseases investigated using molecular genetic techniques, Crohn’s disease (CD) is exceptional in that specific genetic variants unequivocally associated with disease susceptibility have been successfully identified. Nonetheless, characterization of the unknown number of remaining CD genes is required to complete the picture and remains a priority.

CD is one of the 2 common and related forms of IBD, the other being ulcerative colitis (UC). Within the United Kingdom, they have a combined prevalence of approximately 4/1000. Both are known to have a significant genetic contribution to their etiology, but this is stronger for CD than UC. The epidemiologic evidence also suggests that CD and UC share some susceptibility genes. In 2001, fine mapping of a widely replicated linkage region on chromosome 16 led to the identification of CARD15 as a major CD susceptibility gene, with mutations leading to dysregulation of innate immune pathways.
CARD15 genes have subsequently been shown in meta-analysis to predominantly determine susceptibility to ileal CD. Variants within a number of other genes have been associated with CD, UC, or both, although their exact roles in IBD susceptibility require clarification and, in some cases, replication.

To date, pinpointing of disease genes has depended on detailed evaluation of candidates implicated by their function or patterns of expression or by fine mapping within large regions identified in the course of genome-wide linkage scans. Across the range of common diseases, productivity of such approaches has been limited. Most complex disease genetic studies, including many in IBD, have been beset by poor reproducibility of results and slow progress in identifying disease genes. This has been attributed to a range of factors, some of the most important being the low resolution of sib-pair linkage analysis, use of inappropriate statistical thresholds for significance, and poor matching of controls due to population admixture. One powerful new method for the identification of complex disease genes is genome-wide association scanning, genotyping large panels of affected individuals and appropriately matched population controls for hundreds of thousands of polymorphic markers across the genome and using appropriately stringent statistical thresholds for significance. Within the past year, such studies have become technically and financially possible using sets of markers that capture most of the common variation across the genome using knowledge regarding human haplotype structure available from the International HapMap Project (http://www.hapmap.org). Systematic whole-genome association studies, in comparison with the previous gold standard of linkage analysis, should provide substantially increased power and resolution for detection of complex disease susceptibility genes.

Recently, the results of a 308,332-marker genome scan in a North American panel of 547 non-Jewish case patients with CD and 548 controls were reported. Case patients were selected as having ileal CD to reduce heterogeneity. Three markers showed a highly significant association with CD, of which were in CARD15. The third marker was a rare coding variant rs11209026c (1142G→A; Arg381Gln) found in the interleukin 23 receptor (IL23R) gene on chromosome 1 (P = 5.05 × 10^-9). Nine other markers showed association with P < .0001 either within IL23R or in the intergenic area with the adjacent IL12RB2 gene. Internal replication was achieved in the index study using both a Jewish CD case-control cohort (peak P value, 3.36 × 10^-13) and family-based methodologies, the latter in addition suggesting association with UC in a small non-Jewish cohort. This finding indicates that IL23R may have a general role in the etiology of IBD.

The aims of the current study were to seek replication of the association between IL23R and IBD in a large independent North European cohort representing the full range of CD and UC phenotypes, examine in detail genotype-phenotype relationships, explore evidence for epistasis with the known CD susceptibility gene CARD15, and provide accurate estimates of disease risk for associated variants. Replication of the association in an independent cohort would serve 2 important purposes. First, it is key to confirming the veracity of the original finding and the applicability of these findings in populations outside North America. Further, strong independent replication of the key finding of one of the first published genome-wide association scans would provide proof of principle that this novel methodology can be used to identify risk variants for complex diseases.

### Subjects and Methods

#### Subjects

A total of 2877 individuals with IBD (1902 with CD and 975 with UC) were recruited in 5 centers across England and Scotland. The study was approved by the research ethics committees at each center.

Standard clinical, radiologic, and histologic diagnostic criteria were applied. Phenotypic details were obtained by retrospective case notes review. CD phenotype was classified by age at diagnosis, location, and behavior of disease. Only one member of multiply affected families was included. A total of 1.75% were of Jewish origin, and 2.25% were nonwhite. Demographic and subphenotype data are presented in Table 1.

Control allele frequencies were obtained from 1345 individuals recruited across Britain as part of the 1958 British birth cohort. Cases and controls were categorized into 12 broad geographical regions within Great Britain to minimize confounding due to variation in allele frequencies across the country.

### Table 1. Demographic Details of 2877 Individuals With IBD Used in Case-Control Panel

<table>
<thead>
<tr>
<th></th>
<th>CD (n = 1902)</th>
<th>UC (n = 975)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age at diagnosis (y)</td>
<td>26</td>
<td>38.9</td>
</tr>
<tr>
<td>Gender (F/M)</td>
<td>1153/745</td>
<td>480/495</td>
</tr>
<tr>
<td>Smoking at diagnosis (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>58.4</td>
<td>55.0</td>
</tr>
<tr>
<td>Ex</td>
<td>9.4</td>
<td>30.3</td>
</tr>
<tr>
<td>Current</td>
<td>32.2</td>
<td>14.7</td>
</tr>
<tr>
<td>Jewish ancestry (%)</td>
<td>1.75</td>
<td>1.9</td>
</tr>
<tr>
<td>Nonwhite (%)</td>
<td>2.25</td>
<td>3.25</td>
</tr>
<tr>
<td>Surgery (%)</td>
<td>61.8</td>
<td></td>
</tr>
<tr>
<td>Location/extent (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.7 ileal</td>
<td>16.5 rectum only</td>
<td></td>
</tr>
<tr>
<td>31.8 colonic</td>
<td>35.0 distal to splenic flexure</td>
<td></td>
</tr>
<tr>
<td>35.5 ileocolonic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.1 perianal</td>
<td>48.5 proximal to splenic flexure</td>
<td></td>
</tr>
<tr>
<td>Behavior (%)</td>
<td>36.5 stenosing</td>
<td>17.15 penetrating</td>
</tr>
</tbody>
</table>

#### Table 1. Demographic Details of 2877 Individuals With IBD Used in Case-Control Panel

*CD* (n = 1902) *UC* (n = 975)
Genotyping

Genotyping of cases was undertaken with iPLEX chemistry on a matrix-assisted laser desorption/ionization time-of-flight MassARRAY platform (Sequenom, San Diego, CA). Cases were genotyped for 8 IL23R markers reported in the index study, including the non-synonymous single nucleotide polymorphism (SNP) rs11209026 encoding amino acid change Arg381Gln (primer sequences in Supplementary Table 1; see supplemental material online at www.gastrojournal.org). Two of the North American markers (rs7517847, rs2201841) were omitted due to their location within a sequence of interspersed low-complexity repeats.

Genotyping of controls was undertaken at the Wellcome Trust Sanger Institute using the Illumina 550K chip (Illumina, San Diego, CA). Concordance of genotype calls between the different platforms was confirmed by genotyping 87 control DNAs for all 8 markers using the MassARRAY platform with strong concordance of calls between technologies—98.99% for the 8 markers overall. There was 100% concordance for 3 markers, including the coding variant Arg381Gln (Supplementary Table 2; see supplemental material online at www.gastrojournal.org). The data for 1594 cases of CD genotyped for CARD15 mutations in earlier studies were used to undertake analysis for evidence of interaction between CARD15 and IL23R.18–21

Statistical Methods

Allele frequencies were compared between cases and controls and between phenotypic subgroups using χ² tests of 2 × 2 tables. Odds ratios were calculated for the minor allele at each SNP; confidence intervals (CIs) were calculated using Woolf’s method.22 Pairwise SNP linkage disequilibrium coefficients were estimated using Haploview.23 Conditional association analysis was implemented using COCAPHASE, a module of the UNPHASED program.24 This method tests for equality of odds ratios for haplotypes identical at conditioning loci. The Mantel–Haenszel test for association conditioning on geographical region was implemented using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). Median age at disease diagnosis between groups was compared using the Wilcoxon rank sum test. Age at diagnosis was dichotomized according to the Montreal classification.25 Unless specified otherwise, all analyses were performed using R version 2.2 for Windows (http://www.R-project.org).

Results

All genotypes were in Hardy–Weinberg equilibrium in both cases and controls (P > .05). A highly significant association with CD was observed across the region (Table 2). The strongest association was observed at the non-synonymous SNP Arg381Gln, where the frequency of the A allele was 2.5% in CD compared with 6.2% in controls (P = 0.001). The odds ratio for this protective allele was 0.38 (95% CI, 0.29–0.50). Alternatively, the common wild-type homozygous GG genotype can be considered as the risk genotype with an odds ratio of 2.70. To minimize potential confounding from regional differences in allele frequencies, a Mantel–Haenszel test was performed across 12 regional strata. Mantel–Haenszel odds ratios was very similar to those obtained from pooled data for all SNPs. For example, the Mantel–Haenszel odds ratio was 0.38 (95% CI, 0.29–0.50). The nonsynonymous SNP Arg381Gln was highly associated with UC (Table 2). The strongest signal was observed at this locus.

Several SNPs also showed significant association with UC (Table 2). The strongest signal was observed with common SNPs rs1004819 (P = 0.0071) and rs10889677 (P = 0.0042). The frequency of Arg381Gln was only marginally different between cases and controls (UC, 0.046; controls, 0.062; P = .029), with an odds ratio of 0.73 (95% CI, 0.55–0.96). The non-synonymous SNP Arg381Gln was in tight linkage disequilibrium with one other SNP (rs11465804, r² = 0.85) but weak linkage disequilibrium with all 6 other SNPs (r² = 0.03–0.1). A separate test for CD association was performed for each SNP conditioning on Arg381Gln by conditional regression modeling. This showed a significant association at all SNPs (P < .001) except rs11465804, with the strongest residual association detected at rs10889677 (P = 4.6 × 10⁻³). Hence, the non-synonymous SNP does not account for all the association signal at this locus.

Data were then analyzed for evidence of significant genotype-phenotype correlations based on age at onset of CD, disease location, and disease behavior (Table 3). No
significant subgroup association was observed. In particular, the subgroup of subjects with CD affecting the colon only without small bowel disease (n = 539) appeared to be as strongly associated as those with exclusively ileal/small bowel involvement (n = 668) (minor allele frequencies, 2.3% and 2.0%, respectively). The age at disease onset ranged from 12 to 67 years in patients with CD who carried the A allele of Arg381Gln and from 0 to 80 years in wild-type GG cases. There was no difference in the median age of onset between these 2 groups (AA/AG: median, 28 years [n = 85]; GG: median, 26 years [n = 1650]; P = .26). Stratification of cases by age at diagnosis according to the Montreal classification revealed similar genotype frequencies in all groups (Table 3). For UC, subgroup analysis by disease extent, smoking history, and sex also revealed no significant subgroup association. Age at onset of UC ranged from 14 to 79 years in cases who carried the A allele of Arg381Gln and from 2 to 81 years in wild-type GG cases, with no difference in the median age of onset between the 2 groups (AA/AG: median, 34 years [n = 72]; GG: median, 33 years [n = 708]; P = .14) (Table 4). A total of 1540 subjects with CD were fully genotyped for the 3 CARD15 mutations (G908R, L1007fs, R702W) (Table 3). The frequency of Arg381Gln in 460 cases carrying at least one CARD15 mutation (2.2%) was not significantly different from that in 1081 cases who carried none (2.7%; P = .47). None of the 3 cases who were homozygous for the rare A allele also carried a CARD15 mutation.

### Discussion

This study provides unequivocal confirmation of association between variants in the *IL23R* gene and IBD, suggesting a major effect on overall susceptibility to CD and a more modest effect on UC. Importantly, this study also shows the association at *IL23R* for the first time in a non-American population. The strength of this association at *IL23R* and the fact that it reaches such a magnitude in 2 independent data sets leaves no doubt that it is a true finding. In addition, this is one of the first instances of highly significant, independent replication of data derived from a genome-wide association scan and provides important validation of this technique as a hypothesis-free method for the identification of complex disease genes.
As with the North American genome-wide scan, the strongest evidence for association was seen at the non-synonymous SNP Arg381Gln, where the frequency of the A allele was 2.5% in CD compared with 6.2% in controls ($P = 1.1 \times 10^{-12}$). These allele frequencies are similar to those seen in the North American panel. There was no evidence that IL23R variants associate with any particular subphenotype of CD based on disease behavior or location. Hence, there was no difference in minor allele frequency even between the extremes of pure ileal/small bowel CD and pure colonic CD (2.7% and 2.3%, respectively). Likewise, analysis based on disease behavior did not show any specific subgroup associations (Table 3). This negative result is interesting because it contrasts with the other confirmed CD susceptibility locus CARD15, which seems to have definite associations with ileal disease. These findings are extended by the observation of association with UC overall but not with any known UC subphenotype group, suggesting that IL23R variants may exert a rather generic effect on chronic intestinal inflammation, although the effect size in UC does appear to be smaller than in CD. It is noteworthy that the odds ratio confidence interval at Arg381Gln for UC (0.73 [95% CI, 0.55–0.96]) does not overlap with that for CD (0.38 [95% CI, 0.29–0.50]), suggesting a significantly less marked protective effect of the rare allele for UC compared with CD.

Based on data from our large, independent panel of CD cases, it is possible to provide an accurate estimate of the size of the effect conferred by IL23R variants with regard to the risk of CD. We estimated an odds ratio of 0.38 (95% CI, 0.29–0.49) for Arg381Gln. This is likely to be a more accurate estimate than that provided in the index report from the North American study (odds ratio, 0.26; 95% CI, 0.15–0.43) due to the well-recognized bias of the so-called “winner’s curse,” which leads to overestimation of effect size in discovery panels. Characterizing the exact effect size is important to permit sample size calculation for any further attempts at replication. Where the effect size is overestimated, there is a risk that apparently appropriately powered studies will fail to observe the effect and erroneously conclude that it is a false positive.

In some previous reports of genetic association, effect sizes have been quantitated as a population-attributable risk in addition to the odds ratio. This figure is intended to estimate the proportion of disease incidence attributable to a specific variant. However, it cannot be calculated for a protective minor allele as in the case of Arg381Gln. Further, while it is possible to think of the effect at Arg381Gln as an increased risk conferred by the common G allele, calculations of population-attributable risk based on this assumption lead to an implausibly high figure due to the very high carriage rate of the G allele in the control population.

One important question is whether the nonsynonymous variant accounts for all of the association signal at IL23R. This was tested by conditional regression modeling, looking for evidence of association while controlling for the effect at Arg381Gln. From this analysis, it is clear that there is a strong residual signal, maximal at rs10889677 ($P = 4.6 \times 10^{-9}$), and hence that variation at loci in addition to Arg381Gln, either within or adjacent to IL23R, exert an influence on IBD susceptibility. Whether this reflects a functional impact of the non-coding variants themselves or the fact that they are in linkage disequilibrium with other functionally significant or coding variants is yet to be established. Analysis within Haploview (http://www.broad.mit.edu/mpg/haploview/) of data available from the International HapMap Project shows that this selection of 8 tag SNPs captures only 18 out of 83 informative SNPs, within IL23R and the 3′ intergenic region covered by these markers, genotyped in the CEPH (Utah residents with ancestry from northern and western Europe) panel at $r^2 > 0.8$. Any additional coding variation is likely to be rare because interrogation of Ensembl database release 41 (October 2006) (http://www.ensembl.org/) revealed the

---

**Table 4. Arg381Gln Genotype and Allele Frequencies in UC Cases Stratified by Known Phenotypic Subgroups**

<table>
<thead>
<tr>
<th>Sex</th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
<th>Total</th>
<th>Freq(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1</td>
<td>41</td>
<td>447</td>
<td>489</td>
<td>0.044</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>43</td>
<td>436</td>
<td>481</td>
<td>0.049</td>
</tr>
<tr>
<td>Smoking history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>28</td>
<td>301</td>
<td>329</td>
<td>0.043</td>
</tr>
<tr>
<td>Yes</td>
<td>0</td>
<td>7</td>
<td>81</td>
<td>88</td>
<td>0.040</td>
</tr>
<tr>
<td>Ex</td>
<td>1</td>
<td>20</td>
<td>160</td>
<td>181</td>
<td>0.061</td>
</tr>
<tr>
<td>Disease extent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectum only</td>
<td>1</td>
<td>12</td>
<td>134</td>
<td>147</td>
<td>0.048</td>
</tr>
<tr>
<td>Distal to splenic flexure</td>
<td>1</td>
<td>24</td>
<td>286</td>
<td>311</td>
<td>0.042</td>
</tr>
<tr>
<td>Proximal to splenic flexure</td>
<td>1</td>
<td>43</td>
<td>387</td>
<td>431</td>
<td>0.052</td>
</tr>
<tr>
<td>Age at diagnosis (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 or younger</td>
<td>0</td>
<td>2</td>
<td>39</td>
<td>41</td>
<td>0.025</td>
</tr>
<tr>
<td>17–40</td>
<td>1</td>
<td>41</td>
<td>443</td>
<td>485</td>
<td>0.044</td>
</tr>
<tr>
<td>Older than 40</td>
<td>0</td>
<td>28</td>
<td>228</td>
<td>256</td>
<td>0.055</td>
</tr>
</tbody>
</table>
presence of only 2 additional nonsynonymous coding variants (rs1884444 and rs7530511) in IL23R with minor allele frequency >1% in healthy European populations, both of which were investigated in the North American study with neither showing evidence of association. However, it is known that different splice isoforms of IL23R exist and it is possible that their expression is determined by some of the documented noncoding variation. To clarify these issues, future studies will need to include resequencing of IL23R in a CD panel and fine mapping across the gene using markers identified as a result, as well as studies to assess the potential functional impact of variants identified.

The data with regard to Arg381Gln provide evidence of a very common variant being a disease risk allele, or conversely protection from CD being conferred by the rare allele. The explanations are likely to be complex but for immune-mediated conditions may include the fact that genetic variation at a particular locus confers a spectrum of risk, being protective against some diseases, such as infections, while increasing the risk of others, such as autoimmune or inflammatory conditions. These variations will have been subject to differing selection pressures in diverse populations as a result of different environmental exposures. It is also noteworthy that for some of the markers showing evidence of association, it is the rarer allele that is associated with increased risk of CD. This further supports the argument for more than one variant in IL23R with different effects on gene function.

Recent studies have identified IL-23, the cognate ligand of IL23R, as a key player in both innate and adaptive immune systems. Most IL-23 is secreted by activated dendritic cells, monocytes, and macrophages following their exposure to pathogen-derived molecules that bind at toll-like receptors. IL-23 stimulates a unique CD4+ helper T-cell population characterized by the production of IL-17, tumor necrosis factor, and IL-6 and known as Th17 cells. These cells play a central role in driving autoimmune inflammation in a number of animal models. IL-17 stimulates monocytes and endothelial cells to produce proinflammatory mediators, which in turn promote rapid neutrophil recruitment. The effect of IL-23 has recently been distinguished from that of the related heterodimer IL-12, with which it shares a common p40 subunit. Importantly in this regard, 2 studies in knockout mice lacking the p19 subunit of IL-23 showed marked attenuation of T cell–mediated colitis, while knockout of the p35 subunit of IL-12 produces no such attenuation, suggesting that IL-23 but not IL-12 is essential for the development of colitis. The identification of different roles for IL-12 and IL-23 in control of immune pathways together with the current genetic data suggest that targeting IL-23 (and components of its downstream effector pathway) may be a useful and specific strategy to inhibit IBD while sparing systemic host protective immunity.

As well as focusing attention on the IL-23 pathway in the pathogenesis of IBD, the current study also provides key validation of genome-wide association scanning as a means of identifying complex disease susceptibility genes. The North American study group applied an appropriate, genome-wide significance level, and use of such a stringent threshold has immediately led to replication in our independent panel with a level of significance that makes the association indisputable.

To date, complex disease genetic studies have been beset by poor study design, particularly use of nonconservative thresholds for significance, resulting in publication of many unreplicated false-positive results across the spectrum of common disease, hence the importance of the current study in providing unequivocal early replication in an independent panel of the principal findings from one of the first reported genome-wide association scans. There are recent reports in another complex disease (age-related macular degeneration) that also provide grounds for optimism that this technique produces replicable genetic association data. The clear message is that genome-wide association scanning works and that this study design, which is being so vigorously applied across a number of common diseases, is likely to be highly productive. The hope is that with use of appropriately stringent statistical thresholds and appropriately powered data sets, the success seen here in CD will be generally applicable without the plethora of false-positive results that have vexed the field of complex disease genetics to date.

Appendix

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1053/j.gastro.2007.02.051.

References


Received December 14, 2006. Accepted February 8, 2007.
Address requests for reprints to: Jack Satsangi, FRCP, Gastrointestinal Unit, Division of Medical Sciences, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, England. e-mail: j.satsangi@ed.ac.uk; fax: (44) 0131-537-1085.

The authors acknowledge use of DNA from the 1958 British Birth Cohort collection (R. Jones, S. Ring, W. McArdle, and M. Pembrey), funded by Medical Research Council grant G0000934 and Wellcome Trust grant 068545/Z/02, the National Association for Colitis and Crohn’s Disease and the Wellcome Trust for supporting the case DNA collections, and the Wellcome Trust Case Control Consortium, for which the Crohn’s disease panel was originally assembled.

The authors have no conflicts of interest to declare.

M.T., F.C., and S.A.F. contributed equally to this work.