The IBD6 Crohn’s disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants

David A. van Heel1,*,†, Bryan M. Dechairo2,†, Gary Dawson2, Dermot P.B. McGovern1, Kenichi Negoro1, Alisoun H. Carey2, Lon R. Cardon1, Ian Mackay2, Derek P. Jewell1 and Nicholas J. Lench2

1Wellcome Trust Centre for Human Genetics and Gastroenterology Unit, University of Oxford, Oxford, UK and 2Oxagen Ltd, Abingdon, UK

Received April 11, 2003; Revised and Accepted August 13, 2003

Genetic studies in inflammatory bowel disease have identified multiple susceptibility loci, whose relevance depends critically on verification in independent cohorts. Genetic variants associated with Crohn’s disease have now been identified on chromosomes 5 (IBD5/5q31 risk haplotype) and 16 (IBD1 locus, CARD15/NOD2 mutations). Stratification of genome-wide linkage analyses by disease associated variants is now possible, offering both increased power for identification of other loci and improved understanding of genetic mechanisms. We performed a genome-wide scan of 137 Crohn’s disease affected relative pairs from 112 families. Multipoint non-parametric linkage analyses were performed, with further stratification of affection status by common CARD15 mutations and the IBD5 haplotype. We verified linkage of Crohn’s disease to regions on chromosome 3 (P = 0.0009) and X (P = 0.001) in our cohort. Linkage to chromosome 16 (IBD1) was observed in Crohn’s disease pairs not possessing common CARD15 mutations (P = 0.0007), ~25 cM q telomeric of CARD15. Evidence for linkage to chromosome 19 (IBD6) was observed in Crohn’s disease pairs not possessing CARD15 mutations (P = 0.0001), and in pairs possessing one or two copies of the IBD5 risk haplotype (P = 0.0005), with significant evidence for genetic heterogeneity and epistasis, respectively. These analyses demonstrate the complex genetic basis to Crohn’s disease, and show that the discovery of disease-causing variants may be used to aid identification of further susceptibility loci in complex disease.

INTRODUCTION

Crohn’s disease (CD; MIM 266600) is a chronic inflammatory disorder of the gastrointestinal tract. A strong genetic component to disease susceptibility is suggested by combined results from twin studies which demonstrate a greater concordance for CD in monozygotic twins (33%) than dizygotic twins (4%) (1–3), and the increased risk to siblings of affected individuals relative to the general population (15–40 times) (4).

Ten genome-wide scans have been performed in an effort to identify multiple susceptibility loci, whose relevance depends critically on verification in independent cohorts. Genetic variants associated with Crohn’s disease have now been identified on chromosomes 5 (IBD5/5q31 risk haplotype) and 16 (IBD1 locus, CARD15/NOD2 mutations). Stratification of genome-wide linkage analyses by disease associated variants is now possible, offering both increased power for identification of other loci and improved understanding of genetic mechanisms. We performed a genome-wide scan of 137 Crohn’s disease affected relative pairs from 112 families. Multipoint non-parametric linkage analyses were performed, with further stratification of affection status by common CARD15 mutations and the IBD5 haplotype. We verified linkage of Crohn’s disease to regions on chromosome 3 (P = 0.0009) and X (P = 0.001) in our cohort. Linkage to chromosome 16 (IBD1) was observed in Crohn’s disease pairs not possessing common CARD15 mutations (P = 0.0007), ~25 cM q telomeric of CARD15. Evidence for linkage to chromosome 19 (IBD6) was observed in Crohn’s disease pairs not possessing CARD15 mutations (P = 0.0001), and in pairs possessing one or two copies of the IBD5 risk haplotype (P = 0.0005), with significant evidence for genetic heterogeneity and epistasis, respectively. These analyses demonstrate the complex genetic basis to Crohn’s disease, and show that the discovery of disease-causing variants may be used to aid identification of further susceptibility loci in complex disease.
respective population attributable risks of \(~20\) and \(~11\%\) indicate that other loci contributing to genetic susceptibility await discovery (18,19).

Previous stratified linkage studies in complex disease, including Crohn’s disease, have conditioned on the weakly indicative familial linkage status at a locus (e.g. family NPL score), whereas true interactions take place among genetic variants. The identification of genetic variants associated with Crohn’s disease therefore offers potential improvements in the power of genome-wide scans to identify further susceptibility loci. Removal of families segregating a known disease-causing variant will improve power to detect other loci in the remaining families, when genetic heterogeneity exists. Conversely, consideration of only families possessing a disease-causing variant allows for the detection of epistatic interactions with other loci. Analysis of susceptibility in the interleukin-10-deficient mouse model of inflammatory bowel disease has revealed multiple loci and epistatic interactions between loci (20), a system likely to be more complex in human inflammatory bowel disease. In other human diseases, epistatic interactions are beginning to be discovered and have so far been identified for Alzheimer’s disease, breast cancer, Hirschprung’s disease and sickle-cell anaemia (21–24).

Here, we report an extended whole-genome scan in a Crohn’s disease cohort. We stratified the genome-wide linkage analyses by disease associated variants, and revealed evidence for both heterogeneity and epistasis between loci.

RESULTS

All families were genotyped for markers spanning the genome at an average resolution of 9.4 cM (largest gap 21.5 cM). Mean information content obtained across the genome was 71.3%, and the combined microsatellite and SNP genotype success rate (after pedigree checks and removal of possible errors) was 95%.

Multipoint analysis of the unstratified CD phenotype is shown in Figure 1. Two regions met the threshold for suggestive linkage. Linkage was observed on chromosome 3q for Crohn’s disease (LOD 2.1, \(P = 0.0009\)) between D3S1279 and D3S3725. Linkage to the p arm of the X chromosome was observed between DXS1226 and DXS1214 (LOD 2.0, \(P = 0.001\)). No increase in evidence for linkage was observed at either locus after stratification by disease-associated variants.

Linkage analyses performed after stratification of Crohn’s disease affection status by possession/non-possession of disease associated variants revealed regions on chromosomes 3, 12, 16 and 19 meeting the threshold for suggestive linkage (genome-wide data not shown). The regions on chromosome 3p (LOD 2.2, \(P = 0.0008\), D3S1266) and 12p (LOD 2.3, \(P = 0.0005\), D12S364) were observed in CD affecteds not possessing the \(IBD5\) haplotype. These findings are supported by \(\chi^2\) tests for the significance associated with the increase in LOD score after stratification (\(P = 0.008\), \(P = 0.001\), respectively), and by empirical significance levels (\(P = 0.004\), \(P = 0.0004\), respectively) obtained in 10 000 linkage analyses of family subsets equivalent in size randomly sampled from the CD data. However, we believe these data should not be over-interpreted due to the relatively small number of ARPs in this subset.

Evidence for linkage to chromosome 16 was obtained with stratification by possession of one or more common \(CARD15\) variants (Fig. 2A). This result was anticipated, and the \(CARD15\) gene maps to the middle of the broad linkage region (peak LOD 2.6, \(P = 0.0003\), D16S3068). Interestingly, evidence for linkage was also obtained in the \(CARD15\) non-possession set between D16S514 and D16S515 (LOD 2.2, \(P = 0.0007\)), \(\sim 25\) cM q telomeric of \(CARD15\). This result is consistent with a second locus on chromosome 16, and suggested genetic heterogeneity between this locus and the three major \(CARD15\) disease associated variants. This finding was confirmed by \(\chi^2\) tests for the significance associated with the increase in LOD score after stratification (\(P = 0.01\)).

Evidence for linkage to chromosome 19 locus was observed in Crohn’s disease affected relative pairs not possessing common \(CARD15\) mutations (LOD 2.9, \(P = 0.0001\),...
DISCUSSION

Remarkable progress has recently been made in understanding the molecular immuno-genetics of Crohn’s disease. Linkage studies have identified susceptibility loci, and recent research has led to the discovery of disease associated variants on chromosomes 5 (IBD5 locus and haplotype) and 16 (IBD1 locus and CARD15/NOD2 gene). These variants, however, only account for a proportion of the population-attributable risk for Crohn’s disease in our British Caucasian population. We hypothesized that stratification of a genome-wide search by the known Crohn’s disease associated variants might enable detection of further susceptibility loci, either by reducing genetic heterogeneity or detecting epistatic interactions.

Two peaks of ‘suggestive’ linkage were observed, in the initial unstratified Crohn’s disease genome-wide analysis, on chromosomes 3 and X. The peak of linkage observed on chromosome 3 (q arm) lies next to a region observed in a North American study (9). Linkage in the latter study was strongest for the combined Crohn’s disease and ulcerative colitis phenotype, although both phenotypes contributed to the evidence. The peak of linkage on the p arm of the X chromosome has been noted in two other studies (7,8). Consistent with a possible X chromosome susceptibility locus, previous studies have reported an association of IBD with Turner’s syndrome (25), and maternal transmission of CD risk (26).

There is increasing recognition that linkage studies in complex disease are relatively underpowered using the number of families that can feasibly be collected by one centre, and that a much larger replication sample may be needed to confirm initial findings (27,28). The lack of linkage to the IBD1 locus observed in our current cohort (despite the proven presence of association with CARD15 variants), and in genome-wide scans from other investigators, further confirms the relatively low statistical power of such studies, which is likely to explain the variability in findings across centres. Whilst one successful approach in inflammatory bowel disease has been to pool data in a large multi-centre analysis (29), the power of linkage studies can also potentially be increased by conditioning on families that harbour a disease associated variant from other investigators, further confirms the relatively low statistical power of such studies, which is likely to explain the variability in findings across centres. Whilst one successful approach in inflammatory bowel disease has been to pool data in a large multi-centre analysis (29), the power of linkage studies can also potentially be increased by conditioning on families that harbour a disease associated variant from another genomic region. In complex disease, such analyses have generally previously been performed using HLA variants, as the identification of non-HLA genes has proved elusive (30). The identification, and replication, of genetic associations with CARD15 variants and the IBD5 haplotype in the Caucasian population provides an opportunity to increase power in CD linkage studies.

After stratification of the chromosome 16 CD data, strong linkage (peak LOD 2.6) was observed in the CARD15 variant possessing set, as might be expected because CARD15 maps to chromosome 16. These data provide a positive control for the stratified analysis, and also suggest the upper bound of significance that might be obtained if stratification of another locus of equal effect size was biologically valid. A similar effect was not, however, observed with chromosome 5 and the IBD5 haplotype. This result is likely to be related to the very modest relative risk conferred by this locus in our British Caucasian population compared with CARD15 variants (18,19), and the common nature of the haplotype making detection by association easier than by linkage (31).
Interestingly, evidence for linkage to chromosome 16 was also observed amongst relative pairs not possessing common CARD15 mutations. This may reflect rarer private mutations segregating in these families (which account for ~20% of all CARD15 disease causing mutations in the Caucasian population) (32). However, peak linkage in the CARD15-negative Crohn's disease set is observed some 25 cM away from the CARD15 gene. Stratified linkage studies by other investigators have also suggested a locus in the same region (33). Thus these findings might alternatively represent a further chromosome 16 susceptibility gene.

We observed strong linkage between Crohn's disease and a locus on chromosome 19, but only after removal of individuals segregating CARD15 mutations, and obtained evidence for genetic heterogeneity between the two loci. A Canadian study had observed significant linkage across for the combined Crohn's disease and ulcerative colitis phenotype, with a strong contribution from both phenotypes when assessed individually (7). The stronger association with the IBD5 haplotype seen in the Canadian versus British population (19), and therefore greater subsequent epistatic interaction between the IBD6 locus, might explain the ease of linkage detection in this study. Peak linkage in these families was observed across a broad 60 cM region of chromosome 19. Consideration of the region of overlapping linkage in the Canadian and the current study might suggest the locus may lie more towards the centromere of 19p, although non-parametric linkage studies do not provide accurate resolution of disease gene position. These data provide evidence for confirmed linkage using the conservative Lander and Kruglyak criteria, and the designation IBD6 for the chromosome 19 locus. Furthermore, a meta-analysis of 10 Crohn's disease genome scans (D.A. van Heel, unpublished data) also confirms this locus. The recent availability of a high-density single-nucleotide polymorphism (SNP)-based linkage disequilibrium map for chromosome 19 should accelerate disease gene identification studies in this region (34).

Genetic heterogeneity exists in animal models, in which it is possible to knock-out, over-express, or mutate distinct genes involved in diverse pathways and produce a final phenotype similar to human IBD. Susceptibility to colitis in the dextran sulfate sodium and interleukin-10-deficient mouse models is also controlled by multiple loci, which exhibit heterogeneity (20,35). Interestingly, more complex, epistatic interactions between loci have also been demonstrated in the interleukin-10-deficient mouse. Epistasis, or compound non-additive interaction between disease genes, occurs widely in other species and monogenic human disease (36). The presence of such interactions in IBD may explain why locus-specific effects identified in linkage studies do not sum to the overall genetic risk observed in epidemiological studies. An interaction has been reported between a locus on chromosome 1p and IBD1 (9), using the surrogate method of weighting families on the basis of linkage scores (37), although this could not be confirmed in the current study or elsewhere (7). Direct stratification by known associated variants is likely to be a more powerful strategy, and we detected evidence for an epistatic interaction between the chromosome 19 locus and the IBD5 CD risk haplotype.

In conclusion, our study confirms the existence of a Crohn's disease susceptibility locus on chromosome 19, which demonstrates genetic heterogeneity with CARD15 and epistasis with IBD5 variants. We provide evidence for the existence of a locus on chromosome 16 located distant from CARD15. The recent discovery of disease-associated variants is likely to accelerate identification of further Crohn's disease susceptibility genes, and the design of such studies should specifically account for potential complex interactions between loci.

**MATERIALS AND METHODS**

**Families**

Multiply-affected families were recruited from the Oxford inflammatory bowel disease clinic, and referred by physicians/surgeons throughout the UK. Clinical, endoscopic, histological and radiological findings were obtained and reviewed by at least one of the study authors for confirmation of diagnosis based on standard criteria (38). All families were Northern European Caucasian, and no families were of Jewish descent. Venous blood was obtained for DNA extraction from lymphocytes. We genotyped 112 families, each comprising at least one Crohn's disease affected relative pair for linkage analysis (Table 1). A first generation whole-genome search,
using single-point parametric methods, had previously been performed on 32 of the current families (6). Research ethics committee approval and written consent from participants were obtained.

Genotyping
Microsatellite genotyping was performed by PCR amplification of genomic DNA with fluorescent tagged primers. Products were pooled and electrophoresed in ABI3700 automated sequencers. The genome scan was performed using the ABI Prism Linkage Mapping Set Version 2.0 (Applied Biosystems, Cheshire, UK). We had previously genotyped regions of chromosomes 3, 7, 12 and 16 using a higher density microsatellite panel (18,39). In these regions we analysed only markers from the LMS panels, such that the marker density in these regions was similar to that of the other chromosomes (to avoid potential bias) (40). Data were obtained from a total of 404 highly polymorphic microsatellite markers (mean heterozygosity 0.77). Microsatellite genotypes were called using GENOTYPER (version 2.0, Applied Biosystems).

CARD15 variants previously shown to be associated with Crohn’s disease (Arg702Trp, Gly908Arg, Leu1007fsinsC) were genotyped as described (18). A marker, IGR2060a_1, which accurately distinguished the 5q31/C21 found in our study) have suggested that a LOD score of 2.0 would be expected to occur once every three or four genome scans (48). This threshold, corresponding to ‘suggestive linkage’ in a real data set, seemed appropriate for our genome wide studies on five non-independent phenotypes (CD, and four stratified CD sub-phenotypes)—and for which a strict Bonferroni correction would be highly conservative.

We performed two tests to determine the significance of apparent effects due to genetic heterogeneity or epistasis observed in the genotype-stratified linkage analyses. First, we used \( \chi^2 \) tests based on the increase in LOD score observed after stratification, as described (37). Second, we chose families at random from the unstratified CD data for linkage analysis (with sample sizes as found in each of the four genotype stratified sets) and repeated the selection 10 000 times to see how often the observed differences in LOD scores occurred.

ACKNOWLEDGEMENTS
We thank all participating families, our research nurses, the National Association for Crohn’s and Colitis, consultants and general practitioners who helped ascertain families. We thank M.J. Daly for advice on the simulation analyses. This work was supported by the UK Medical Research Council (Clinical Training Fellowships to D.A.vH. and D.P.Mc.G., and a LINK Grant), the Wellcome Trust (L.R.C. holds Principal Research Fellowship, DAvH a Clinician Scientist Fellowship), the Hamanako Symposium (K.N.), Oxagen Ltd and NIH grant YE15652 (L.R.C.).

REFERENCES

<p>| Table 1. Summary of families and ARPs analysed in the Crohn’s disease genome-wide scan |
|---------------------------------------------|----------|----------|----------|</p>
<table>
<thead>
<tr>
<th>AFFECTED RELATIVE PAIR TYPE</th>
<th>INFORMATIVE FAMILIES</th>
<th>AFFECTED SIBLING PAIRS</th>
<th>AFFECTED RELATIVE PAIRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Families with affected sibling pair</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families with affected sibling trio</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families with affected sibling pair, and affected aunt/uncle</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families with affected sibling pair, and affected niece/nephew</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families with affected half-sibling pairs</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families with affected avuncular pair</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total families (informative for linkage)</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total affected sibling pairs</td>
<td>128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total affected relative pairs</td>
<td>137</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 2. Crohn’s disease families and ARPs stratified by CARD15 and IBD5 |
|-----------------------------|-------------|-------------|-------------|</p>
<table>
<thead>
<tr>
<th>STRATIFICATION</th>
<th>INFORMATIVE FAMILIES</th>
<th>AFFECTED SIBLING PAIRS</th>
<th>AFFECTED RELATIVE PAIRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD-CARD15-0</td>
<td>56</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td>CD-CARD15-1 + 2</td>
<td>38</td>
<td>42</td>
<td>44</td>
</tr>
<tr>
<td>CD-IBD5-0</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>CD-IBD5-1 + 2</td>
<td>64</td>
<td>78</td>
<td>83</td>
</tr>
</tbody>
</table>

\*CARD15 (or IBD5)-0, wild-type; CARD15 (or IBD5)-1 + 2, possessing one variant (for CARD15 any of Arg702Trp, Gly908Arg, Leu1007fsinsC) or two variants (as homozygote or complex heterozygote).


