Association between a complex insertion/deletion polymorphism in \textit{NOD1} (\textit{CARD4}) and susceptibility to inflammatory bowel disease

Dermot P.B. McGovern\textsuperscript{1,2,*}, Pirro Hysi\textsuperscript{1}, Tariq Ahmad\textsuperscript{2}, David A. van Heel\textsuperscript{3}, Miriam F. Moffatt\textsuperscript{1}, Alisoun Carey\textsuperscript{4}, William O.C. Cookson\textsuperscript{1} and Derek P. Jewell\textsuperscript{2}

\textsuperscript{1}The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK, \textsuperscript{2}Gastroenterology Unit, University of Oxford, Radcliffe Infirmary, Oxford, \textsuperscript{3}Imperial College, Hammersmith Campus, London, UK and \textsuperscript{4}Oxagen Ltd, Abingdon, Oxfordshire, UK

Received November 19, 2004; Revised March 11, 2005; Accepted March 21, 2005

The identification of the role of genetic variants within \textit{NOD2} (\textit{CARD15}) in Crohn’s disease and ulcerative colitis susceptibility highlight the role of the innate immune system in inflammatory bowel disease (IBD) pathogenesis. \textit{NOD1} (\textit{CARD4}) is located on chromosome 7p14.3, in a region of known linkage to IBD and encodes an intracellular bacterial pathogen-associated molecular pattern receptor that is closely related to \textit{NOD2}. We have identified strong association between haplotypes in the terminal exons of \textit{NOD1} and IBD (multi-allelic $P = 0.0000003$) in a panel of 556 IBD trios. The deletion allele of a complex functional \textit{NOD1} indel polymorphism (\textit{ND1} \textsuperscript{32656} \textsuperscript{*1}) was significantly associated with early-onset IBD ($P = 0.0003$) in unrelated cases and controls. \textit{ND1} \textsuperscript{32656} \textsuperscript{*1} was also associated with extra-intestinal manifestations of IBD ($P = 0.04$). These findings in two independent populations provide strong evidence for a role for \textit{NOD1} variants in IBD susceptibility and reinforce the role of the innate immune system in IBD pathogenesis.

INTRODUCTION

The inflammatory bowel diseases (IBD), ulcerative colitis (UC) and Crohn’s disease (CD) are related polygenic conditions sharing some but not all susceptibility alleles. They have a combined prevalence of 400/100 000 in western Europe (1) and are a significant cause of morbidity among young people. The identification of \textit{NOD2} (\textit{CARD15}) as the first susceptibility gene for CD (2,3) was a breakthrough in understanding IBD pathogenesis and for complex disease genetics in general.

The \textit{NOD2} protein is made up of a leucine rich repeat region (LRR) that recognizes bacterial muramyl dipeptide (4), a nucleotide binding domain and two caspase recruitment domains. Variants within or adjacent to the LRR muramyl dipeptide recognition domain are associated with altered NF\kappa B activation and increased susceptibility to CD (2).

The CD association with \textit{NOD2} has been widely replicated (5) and the association is particularly strong with CD of the small bowel (6,7). More recently, an association between \textit{NOD2} variants and UC has also been described (8). The population attributable risks for \textit{NOD2} variants are $\sim 30$ and 6\% for CD (6) and UC (8), respectively. Genome-wide scans indicate that several other genetic loci for IBD exist (9–14).

A haplotype on 5q31 (\textit{IBD5}) (15) has been widely replicated as an IBD locus and recently it has been suggested that variants in \textit{OCTN1} and \textit{OCTN2}, cation transporter genes, are the \textit{IBD5} susceptibility genes (16,17). Variants in \textit{DLG5}, encoding an epithelial scaffolding protein, on chromosome 10q23 have also been associated with susceptibility to IBD (16,17), although this finding has not yet been replicated. A further susceptibility locus for UC and CD has been identified on chromosome 7p14 in a British genome-wide scan for linkage to IBD (18). Further evidence for linkage to this region has been demonstrated in other genome-wide scans (9,10) and in a recent meta-analysis of genome scans for IBD (19).

The gene encoding \textit{NOD1} (\textit{CARD4}) is located within the chromosome 7p14 IBD locus. \textit{NOD1} is similar in structure to \textit{NOD2}. It contains LRR and NOD domains but differs from \textit{NOD2} in the presence of a single CARD. \textit{NOD1}, like \textit{NOD2}, activates NF\kappa B and enhances apoptosis. \textit{NOD1}
detects a unique tripeptide motif (diaminopimelic acid) found in Gram-negative bacterial peptidoglycan (20). A previous study had typed a single NOD1 polymorphism and failed to detect association with IBD (21). The aim of our study was therefore to assess, more comprehensively, genetic variation in NOD1 for influences on IBD susceptibility.

RESULTS

We genotyped 12 previously identified NOD1 polymorphisms (Table 1) (22) in a panel of 556 IBD trios containing 294 CD trios, 252 UC trios and 10 trios with a diagnosis of indeterminate colitis. We tested for association using the transmission disequilibrium test (TDT). We found that the common deletion allele of a complex polymorphism (ND1 + 32656/C31) was significantly associated with IBD (P = 0.02) and with UC (P = 0.01). Associations were also observed between IBD and ND1 + 233/C31 (P = 0.05), ND1 + 21984/C31 (P = 0.02) and ND1 + 27606/C32 (P = 0.05). The rs5743336 (E266K) NOD1 polymorphism previously examined in IBD (21) had a frequency, 1% in our subjects and was not tested.

Linkage disequilibrium between the markers was assessed in unrelated individuals (Fig. 1). LD was incomplete between all the markers, and in particular LD between the ND1 + 32656/C1 polymorphism and neighbouring markers was weak or absent. These findings are consistent with those previously observed in asthmatic families (22).

We further investigated the locus by examining two-, three- and four-marker haplotypes in a sliding window across the locus. Strong associations were observed between two-marker ND1 + 32656/ND1 + 45343 haplotypes and CD (Multi-allelic TDT P = 0.007), UC (P = 0.000007) and IBD (P = 0.000000003) (Fig. 2). Examination of individual haplotypes indicated the presence of a strong protective effect of the ND1 + 32656/2/ND1 + 45343*1 haplotype (frequency 7%) to IBD (Table 2). Extension of the haplotype to include other markers resulted in increased haplotype diversity, but progressively decreased the evidence for association. This result is consistent with as yet undiscovered effects in the interval between ND1 + 32656 and ND1 + 45343.

Table 1. NOD1 polymorphisms tested in family panels

<table>
<thead>
<tr>
<th>Position ID*</th>
<th>Database ID</th>
<th>Frequency b</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND1 – 664</td>
<td>rs2736726</td>
<td>0.754</td>
</tr>
<tr>
<td>ND1 + 233</td>
<td>rs2075817</td>
<td>0.751</td>
</tr>
<tr>
<td>ND1 + 18915</td>
<td>rs2975632</td>
<td>0.801</td>
</tr>
<tr>
<td>ND1 + 21658</td>
<td>rs3020207</td>
<td>0.709</td>
</tr>
<tr>
<td>ND1 + 21984</td>
<td>rs2075818</td>
<td>0.765</td>
</tr>
<tr>
<td>ND1 + 25816</td>
<td>rs2235099</td>
<td>0.766</td>
</tr>
<tr>
<td>ND1 + 26129</td>
<td>rs3020208</td>
<td>0.924</td>
</tr>
<tr>
<td>ND1 + 27053</td>
<td>rs2075821</td>
<td>0.769</td>
</tr>
<tr>
<td>ND1 + 27606</td>
<td>rs2075822</td>
<td>0.779</td>
</tr>
<tr>
<td>ND1 + 32656c</td>
<td>–</td>
<td>0.782</td>
</tr>
<tr>
<td>ND1 + 45343</td>
<td>rs2907748</td>
<td>0.799</td>
</tr>
<tr>
<td>ND1 + 50150</td>
<td>rs5743368</td>
<td>0.886</td>
</tr>
</tbody>
</table>

aThe position is numbered from the first nucleotide of exon 1. The sequence is obtained from Golden Path (http://genome.ucsc.edu).
bFrequency in unrelated individuals (parents). Allele “/C31 is defined as the more common.
cComplex insertion/deletion polymorphism (indel), partially identified as rs6958571.

Figure 1. Linkage disequilibrium between markers. The marker positions are shown along the diagonal of the figure. Pairwise estimations of D’ are shown from unrelated subjects in the family panel (the parents), on a scale of 1 (complete linkage disequilibrium: red) to 0 (blue). Marker positions are shown as a schematic rather than as actual distances apart.

Figure 2. Family-based association study of NOD1 two-marker haplotypes in a sliding window. P-values were calculated with the extended multi-allelic transmission disequilibrium test (ETDT).
Examination of NOD1, NOD2 and IBD5 in a multiple regression showed that only ND1 + 32656 was significantly associated with the age of diagnosis of CD (P = 0.023) (Table 4), although the trend was also for NOD2 to predict early onset of disease (P = 0.06). IBD5 was not associated with early diagnosis of disease in these subjects (P = 0.47), in contrast to an earlier study (15). Univariate Spearman correlation coefficients were NOD1, R = 0.174, P = 0.002; IBD5, R = 0.01, P = 0.923 and NOD2 R = 0.057, P = 0.31. No genetic interactions between the loci were found to affect the age of diagnosis of disease.

No significant independent association was seen when the IBD samples were stratified by the presence of stenotic or fistulating disease, disease location or smoking status. However, ND1 + 32656 was associated with the presence of IBD extra-intestinal manifestations (EIMs: large and small joint arthritis,ankylosing spondylitis, ocular inflammation, erythema nodosum, pyoderma gangrenosum and primary sclerosing cholangitis) (OR 1.35, 95% CI = 1.02–1.82, P = 0.04 when compared with individuals with IBD and without EIMs).

**DISCUSSION**

We have demonstrated association between genetic variants in NOD1 and susceptibility to IBD in two independent cohorts. ND1 + 32656, which is located on the sliding two-locus haplotype significantly associated with IBD, is located at the beginning of intron IX and affects the binding of an unknown nuclear factor (22). NOD1 is expressed in both large and small bowel as a number of splice variants arising from exon IX (22). It has not yet been established whether ND1 + 32656 may be associated with general susceptibility to IBD rather than CD or UC alone. These findings are consistent with original linkage data (18,19).

We then compared the risk of CD conferred by ND1 + 32656 and the known susceptibility loci NOD2 and IBD5 in the cases and controls by logistic regression. The OR for ND1 + 32656 was 2.0 [95% confidence Interval (CI) 1.2–3.5], which was comparable with the effect of IBD5 (OR = 2.3, 95% CI = 1.7–3.1) but less than that of NOD2 (OR = 6.45; 95% CI = 4.1–10.2). There was no evidence of genetic interaction between the three loci.

Stratification of cases by age showed that early diagnosis (<25 years) of disease was strongly associated with ND1 + 32656 for CD (P = 0.00004) and IBD (P = 0.0003) (Table 3). Wilcoxon life-table analyses demonstrated that the median age of diagnosis of CD was 32.3 years for the ND1 + 32656 *I*1 genotype, 36.2 for *I*2 and 49.6 for *I*2 *I*2 (P = 0.001 for the difference) (Fig. 3). Similar results were seen if all the cases of IBD were examined together.

### Table 2. Transmission of ND1 + 27606/ND1 + 32656 haplotypes in family panels

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
<th>CD T</th>
<th>NT</th>
<th>P-value</th>
<th>CD age diagnosis &lt;25</th>
<th>UC T</th>
<th>NT</th>
<th>P-value</th>
<th>IBD T</th>
<th>NT</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (<em>I</em>1)</td>
<td>0.725</td>
<td>79</td>
<td>67</td>
<td>ns</td>
<td>40</td>
<td>44</td>
<td>ns</td>
<td>65</td>
<td>58</td>
<td>ns</td>
<td>146</td>
</tr>
<tr>
<td>2 (<em>I</em>2)</td>
<td>0.156</td>
<td>73</td>
<td>64</td>
<td>ns</td>
<td>47</td>
<td>35</td>
<td>ns</td>
<td>61</td>
<td>45</td>
<td>ns</td>
<td>136</td>
</tr>
<tr>
<td>3 (<em>I</em>1)</td>
<td>0.069</td>
<td>16</td>
<td>39</td>
<td>0.002</td>
<td>12</td>
<td>26</td>
<td>ns</td>
<td>10</td>
<td>43</td>
<td>6 x 10^-6</td>
<td>26</td>
</tr>
<tr>
<td>4 (<em>I</em>2)</td>
<td>0.050</td>
<td>24</td>
<td>22</td>
<td>ns</td>
<td>19</td>
<td>13</td>
<td>ns</td>
<td>22</td>
<td>12</td>
<td>0.09</td>
<td>47</td>
</tr>
</tbody>
</table>

*aFrequency in unrelated founders.

### Table 3. Genotype frequencies of ND1 + 32656 in healthy controls and patients with IBD

<table>
<thead>
<tr>
<th>Status</th>
<th>ND1 + 32656 Genotype</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>I</em>1</td>
<td><em>I</em>2</td>
</tr>
<tr>
<td>Controls</td>
<td>161</td>
<td>135</td>
</tr>
<tr>
<td>IBD</td>
<td>358</td>
<td>259</td>
</tr>
<tr>
<td>UC</td>
<td>153</td>
<td>128</td>
</tr>
<tr>
<td>CD</td>
<td>290</td>
<td>131</td>
</tr>
<tr>
<td>IBD onset &lt;25</td>
<td>143</td>
<td>71</td>
</tr>
<tr>
<td>CD onset &lt;25</td>
<td>106</td>
<td>44</td>
</tr>
</tbody>
</table>

*aMantel–Haenszel test for linear association compared with controls.

Weaker associations with other two-marker haplotypes were observed in a cluster between ND1 + 21658 and ND1 + 26129 (Fig. 1). These results suggested independent effects within NOD1 on susceptibility to IBD.

To confirm these results we then tested ND1 + 32656 in an independent panel of 664 subjects with IBD (358 CD and 306 UC) and 335 controls. In these subjects the ND1 + 32656 *I* allele was significantly associated with CD (P = 0.003) and IBD (P = 0.017) but not with UC (Table 3). However, although association to UC was not seen in these subjects, the trend was in the same direction as for CD, and the overall data from the family and case-control studies suggest that ND1 + 32656 *I* may be associated with general susceptibility to IBD rather than CD or UC alone. These findings are consistent with original linkage data (18,19).
NOD1 has been examined previously for a role in IBD susceptibility, with the conclusion that it does not influence disease susceptibility (21). However, these investigations were only concerned with rare coding polymorphism. The general recognition now is that complex disease susceptibility is expected often to be mediated through regulatory polymorphisms, as is the case in the present study.

NOD1 is expressed in large and small bowel (22) and plays a role in colonic epithelial defences against intracellular organisms, such as *Shigella flexneri* (23) and *enteroinvasive E. coli* (24). The presence of bacterial flora is essential for IBD to develop in animal models (25), and antibiotics (26) and faecal diversion (27) are effective therapies for CD. The identification of associations between NOD1, NOD2 and IBD suggest that altered recognition of intracellular bacterial pathogen-associated molecular pattern may be a key event in the pathogenesis of the disease. Further work is now required to establish the effect of these variants within the gastrointestinal tract and the role they play in IBD behaviour and response to treatment.

**MATERIALS AND METHODS**

**Study populations**

Five hundred and fifty-six IBD nuclear trios were recruited as previously described (5). An independent panel of 664 subjects with IBD [358 CD and 306 UC, median age of onset of disease 28.2 years (range 1.0–82.2 years), 44.4% male] and 335 controls was studied to confirm potentially positive results. All subjects and controls were Caucasian. The family collection was obtained from probands attending the Oxford IBD clinic and referred from gastroenterologists from around UK. Non-family cases were recruited patients attending the John Radcliffe Hospital, Oxford IBD clinic and were compared with healthy unrelated individuals recruited through the UK blood transfusion service and from healthy individuals attending ‘well-person’ screening clinics at their family doctors in Oxfordshire, UK.

All diseased cases were diagnosed as having IBD according to standard clinical (history of abdominal pain, weight loss, rectal bleeding, diarrhoea, abdominal mass, perianal disease with or without evidence of IBD EMIs), endoscopic (macroscopic evidence of rectal, colonic or terminal ileal mucosal inflammation), radiological (superficial or deep ulceration, presence of fistulae or strictures and distribution of disease) and histological findings (inflammation, inflammatory infiltrates, glandular architecture distortion, goblet cell depletion and the presence of granulomas). Patients were classified with indeterminate colitis in the presence of definitive IBD affecting the colon only but with histology ambiguous to the presence of UC or CD.

All individuals in the study gave written informed consent and ethical approval from the relevant hospital ethical committees had been obtained.

**SNP identification and genotyping**

Polymorphisms in NOD1 were identified and genotyping was performed as previously described (22). Alleles for the three common CD associated NOD2 variants (Arg702Trp, Gly908Arg and the fsinsC1007) and for an IBD5 CD risk haplotype tagging SNP were genotyped as previously described (5,15,28).

**Statistics**

Polymorphisms in the family panel were tested for Mendelian transmission and for Hardy–Weinberg equilibrium before inclusion in the analyses. Association to categorical traits within the families was examined by the TDT. Haplotypes were generated by the MERLIN computer program (29), and D’ between markers estimated by the MERLIN utility HAPLOXT. The patterns of linkage disequilibrium between markers were visualized using GOLD (30). Two-, three- and four-marker haplotypes in the trios were generated across the locus in a sliding window by MERLIN (29), and coded as individual alleles before analysis by the multi-allelic TDT (ETDT) test (31).

In the case–control panel Hardy–Weinberg equilibrium was established before using the Mantel–Haenszel test for linear association to compare genotype frequencies and
life-table analyses with the Wilcoxon (Gehan) statistic to compare age of onset of disease between genotypes (SPSS for OSF, 6.1.4). The effects of NOD1, NOD2 and IBD5 genotypes on disease risk were estimated by logistic regression, and their effects on age of diagnosis were examined by multiple linear regression (SPSS for OSF, 6.1.4). Gene–gene interactions were tested as interaction terms within the regressions. The effects of phenotypic variants were examined by $\chi^2$ estimations of contingency tables selected by phenotypic stratum.

Given the close correlations between various phenotypes and between the markers themselves, simple corrections for multiple comparisons were not applied to the initial test results and replication was used to determine true genetic effects. In order to maintain a conservative approach, two-tailed $P$-values were estimated for the replicate samples.

**ACKNOWLEDGEMENTS**


**REFERENCES**


