# The Contribution of *NOD2* Gene Mutations to the Risk and Site of Disease in Inflammatory Bowel Disease

ANDREW P. CUTHBERT,* SHEILA A. FISHER,* MUDDASSAR M. MIRZA,* KATHY KING,* JOCHEN HAMPE,† PETER J. P. CROUCHER,† SILVIA MASCHERETTI,† JEREMY SANDERSON,§ ALASTAIR FORBES,‖ JOHN MANSFIELD,¶ STEFAN SCHREIBER,† CATHRYN M. LEWIS,* and CHRISTOPHER G. MATHEW*

*Division of Medical and Molecular Genetics and §Department of Medicine, Guy's, King's, and St Thomas' School of Medicine, London, England; †Department of General Internal Medicine, Christian-Albrechts University, Kiel, Germany; ‖St Mark's Hospital, Harrow, England; and ¶Department of Gastroenterology, School of Clinical and Medical Sciences, University of Newcastle, Newcastle upon Tyne, United Kingdom

**See editorial on page 1161.**

***Background & Aims:*** **Mutations in the *NOD2* gene are strongly associated with susceptibility to Crohn's disease (CD). We analyzed a large cohort of European patients with inflammatory bowel disease to determine which mutations confer susceptibility, the degree of risk conferred, their prevalence in familial and sporadic forms of the disease, and whether they are associated with site of disease.** ***Methods:*** **Individuals were genotyped for 4 *NOD2* mutations: P268S, R702W, G908R, and 3020insC. Allelic transmission distortion to 531 CD- and 337 ulcerative colitis–affected offspring was assessed by the transmission disequilibrium test. Association was also tested in an independent cohort of 995 patients with inflammatory bowel disease and 290 controls. Cases were stratified by disease site and compared across *NOD2* genotypes.** ***Results:*** **R702W, G908R, and 3020insC were strongly associated with CD but not with ulcerative colitis. Linkage disequilibrium was observed between P268S and the other mutations, forming 3 independent disease haplotypes. Genotype relative risks were 3.0 for mutation heterozygotes and 23.4 for homozygotes or compound heterozygotes. The frequency of *NOD2* mutations was higher in cases from families affected only with CD and was significantly increased in ileal-specific disease cases compared with colon-specific disease (26.9% vs. 12.7%, $P = 0.0004$).** ***Conclusions:*** **The R702W, G908R, and 3020insC mutations are strong independent risk factors for CD and are associated particularly with ileal disease.**

Crohn's disease (CD [MIM 2666600]) and ulcerative colitis (UC [MIM191390]) are common clinical subtypes of idiopathic inflammatory bowel disease (IBD) characterized by chronic inflammation of the gastrointestinal tract. The combined prevalence of CD and UC in Western countries is 100–200/100,000.[1] The underlying basis of pathogenesis in IBD is not yet clear but may involve persistent bacterial infection, a defective mucosal barrier, or an imbalance in the regulation of the immune response.[2] Epidemiologic studies have identified a significant genetic contribution to the etiology of IBD,[3] but simple mendelian models of IBD inheritance are not supported by segregation analyses.[4–7] Taken together, these observations support a complex immunogenetic model for IBD whereby genetically susceptible individuals harbor an aberrant response to yet-unidentified environmental influences.

Attempts to localize IBD susceptibility genes through genome-wide linkage studies have identified putative loci on many human chromosomes. The original finding of the linkage on chromosome 16[8] in particular has been replicated in several studies and by the IBD International Genetics Consortium.[9] Recently, 3 groups, including our own, showed that sequence variations within the *NOD2* gene (MIM 605956) on chromosome 16q12 were strongly associated with susceptibility to CD but not UC.[10–12] In this report, we examine the contribution of 4 mutations in the *NOD2* gene to IBD susceptibility in 2 large cohorts of both familial and sporadic cases of CD and UC. We propose a genetic model for CD susceptibility that shows a >20-fold increased risk associated with homozygosity for *NOD2* mutations. We also investigate the pathogenic effect of these mutations through stratification by site of disease.

*Abbreviations used in this paper:* SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test.



0016-5085/02/$35.00
doi:10.1053/gast.2002.32415

## Materials and Methods

### Patient Cohorts

Family collections consisted of trios of parents with a single affected offspring or families with at least 2 affected offspring. At least 1 parent was available for genotyping in 82% of families. Patients were recruited from King's College School of Medicine, Guy's Hospital, and St Mark's Hospital (London, England), Charite University Hospital (Berlin, Germany), the Department of General Internal Medicine at the Christian-Albrechts-Universitat (Kiel, Germany), and the Academic Medical Centre (Amsterdam, The Netherlands). A cohort of independent patients with CD or UC was recruited from King's College Hospital, Guy's Hospital, and St. Mark's Hospital (London, England), Royal Victoria Infirmary, Freeman Road Hospital, and North Tyneside Hospital (Newcastle, United Kingdom). Clinical data on the site of disease were available for a subset of British cases. A summary of the family and case-control cohorts is shown in Table 1. In the family cohort, 82% of families had at least 1 parent available for genotyping. One unaffected sibling was typed in 113 families to facilitate inference of missing parental genotypes. The cohort of 559 families included 309 families who were genotyped for the 3020insC mutation in our previous study.[12] None of the 995 sporadic cases were included in the previous study.

Diagnosis of IBD and classification as UC or CD was confirmed by established criteria of clinical, radiologic, and endoscopic analysis, as well as histologic reports.[13,14] These data were reviewed centrally by 1 or more members of a panel composed of several of the authors (J.S., A.F., J.M., and S.S.), as well as J. Lennard-Jones, A. Macpherson, and S. van Deventer. The site of CD was investigated by colonoscopy with supportive histology and/or double-contrast barium enema, with a barium series (usually a barium follow-through) to assess the small bowel. Ileal-specific disease was defined as typical changes seen on small bowel barium study with normal colon histology. A venous blood sample was obtained from affected individuals and for the family collection, where possible, from parents and unaffected siblings. Informed written consent was obtained from all study participants, and recruitment protocols were approved by institutional review committees at each participating center. Normal healthy controls were recruited through the European Collection of Animal Cell Cultures (Wiltshire, England) and from residents in the Newcastle area as part of a regional genetic study.

**Table 1.** Summary of Family and Case-Control Cohorts

| | Phenotype | | | |
|---|---|---|---|---|
| | CD | UC | MX | Total |
| Family cohort | | | | |
| No. of families with 1 affected offspring | 165 | 122 | — | 287 |
| No. of families with ≥2 affected offspring | 143 | 72 | 57 | 272 |
| Total number of families | 308 | 194 | 57 | 559 |
| Total number of affected offspring | 467 | 274 | 127 | 868 |
| Case-control cohort | | | | |
| London cases | 222 | 244 | — | 466 |
| Newcastle cases | 207 | 322 | — | 529 |
| Total cases | 429 | 566 | — | 995 |
| ECACC controls | — | — | — | 188 |
| Newcastle controls | — | — | — | 102 |
| Total controls | — | — | — | 290 |

MX, mixed families with both CD- and UC-affected offspring; ECACC, European Collection of Animal Cell Cultures.

### Mutation Detection and Genotyping

The 12 exons and untranslated regions of *NOD2* were screened for mutations by denaturing high-performance liquid chromatography and genomic DNA resequencing using flanking intronic polymerase chain reaction primers as previously described.[12] Twelve single nucleotide polymorphisms (SNPs) were identified, of which 9 were located in exons. Four of the coding SNPs were nonsynonymous mutations: (1) a frameshift C-insertion mutation 3020insC in exon 11 and (2) 3 missense mutations, 802C → T/P268S (exon 4), 2104C → T/R702W (exon 4), and 2722G → C/G908R (exon 8). Nucleotide positions were determined by alignment of sequencing data with the consensus complementary DNA sequence (GenBank accession no. AF178930). These polymorphisms correspond to SNP5, SNP8, SNP12, and SNP13 described by Hugot et al.[10] The detection methods, sequence context, and allele frequencies for all SNPs we detected in *NOD2* can be viewed at http://www.ncbi.nlm.nih.gov/SNP/snp_search.cgi?searchType=byPub&pub_id=601.

Genotypes of the 4 nonsynonymous polymorphisms P268S, R702W, G908R, and 3020insC were determined using the TaqMan biallelic discrimination system.[15] Genotypes were called manually and exported for mendelian inheritance error checking using PedCheck.[16]

### Statistical Analysis

Families were analyzed for association of the *NOD2* polymorphisms with disease using the transmission disequilibrium test (TDT).[17] The TDT was implemented using TRANSMIT,[18] which infers missing parental genotypes from population allele frequencies assuming Hardy–Weinberg equilibrium and uses additional genotype information from unaffected siblings to reconstruct multilocus haplotypes. In the case-control cohorts, tests for Hardy–Weinberg equilibrium and case-control association analyses were performed using $\chi^2$ statistics. Genotype relative risks were estimated based on control genotype frequencies calculated assuming Hardy–Weinberg equilibrium. The population attributable risk was estimated by $(X - 1)/X$, where $X = (1 - p)^2 + 2p(1 - p)GRR_1 + p^2GRR_2$, given population allele frequency $p$ and genotype risk ratios $GRR_1$ and $GRR_2$ for heterozygous and homozygous genotypes. Separate analyses were performed for CD and UC. Haplotype frequencies were estimated with EHPLUS[19,20] using an independent sample of unrelated individuals. Linkage disequilibrium was assessed using the coeffi-

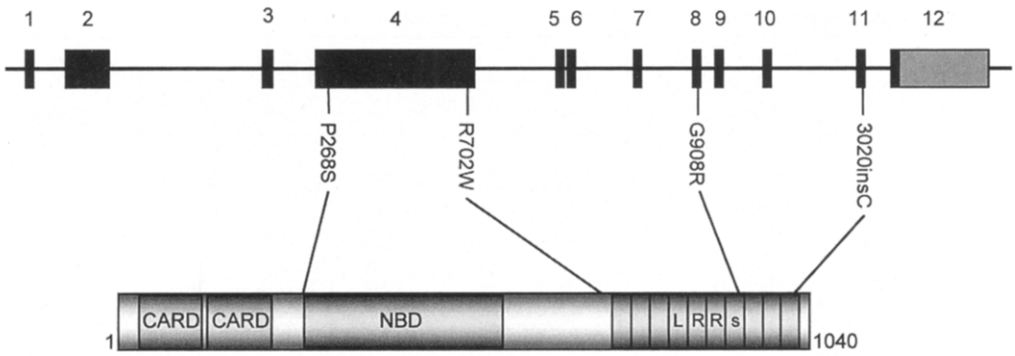


**Figure 1.** Diagram of the genomic structure of the *NOD2* gene, the location of the mutations studied, and the predicted structural domains of the encoded protein. CARD, caspase recruitment domain; NBD, nucleotide binding domain; LRR, leucine-rich repeat region.

cient Δ as a measure of association.[21] Cases were stratified according to site of disease and compared across *NOD2* genotypes using $\chi^2$ statistics.

## Results

### Association of Mutations With CD

Sibling pairs from the British/German family cohort have previously shown evidence of linkage to the *IBD1* locus on chromosome 16.[22] Allelic transmission distortion for 4 SNPs (P268S, R702W, G908R, and 3020insC; see Figure 1) was assessed in this family collection using the TDT. Highly significant evidence of preferential transmission of the rare allele to CD-affected offspring was observed for P268S, R702W, and 3020insC (Table 2). Excess transmission of the rare G908R SNP did not reach statistical significance. Interestingly, transmissions to UC-affected offspring were significantly lower than expected for 3 of the SNPs (R702W, G908R, and 3020insC). A minimal degree of transmission distortion to offspring not affected with CD may be expected in a complex disease,[23] and transmissions to 133 unaffected offspring in the TDT families did not deviate from expected values (data not shown).

The 4 polymorphisms were also genotyped in the case-control cohort consisting of 429 patients with CD, 566 patients with UC, and 290 normal controls. This independent case collection was ascertained as sporadic cases presenting at clinic. Because the patients and controls originated from 2 locations, allele frequencies for each collection were compared. There were no significant differences in allele frequencies between London and Newcastle CD or UC case collections (data not shown). Similarly, allele frequencies in controls originating from the Newcastle area were not significantly different than controls obtained from throughout the United Kingdom, and data from these collections were therefore combined for subsequent analysis. The distribution of genotypes within each phenotypic group was consistent with Hardy–Weinberg equilibrium, with the exception of P268S ($P = 0.016$) and 3020insC ($P = 8.1 \times 10^{-4}$) in patients with CD. All 4 polymorphisms were of higher frequency in patients with CD than normal controls but were not significantly increased in patients with UC (Table 3). To estimate linkage disequilibrium across *NOD2* mutations and assess disease risk, a single affected offspring was selected from each family in the family cohort and combined with patients from the case-control collection, providing an independent set of 1292 patients (688 with CD and 604 with UC) and 250 controls. No evidence of linkage disequilibrium between R702W, G908R, and 3020insC was observed in this independent sample (linkage disequilibrium coefficient, $\Delta < 0.02$ for all pairwise tests). However, some evidence of linkage disequilibrium was observed between P268S and each of the other mutations (R702W, $\Delta = 0.38$; G908R, $\Delta = 0.13$; 3020insC, $\Delta = 0.25$). This was assessed further by analysis of transmission distortion of haplotypes across the 4 SNPs (Table 4). Haplotypes carrying P268S to-

**Table 2.** TDT for *NOD2* Mutations: Transmissions to CD- and UC-Affected Offspring From the Family Cohort Calculated Using TRANSMIT

| | CD | | | | | UC | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SNP | n | Obs | Exp | $\chi^2$ | *P* | n | Obs | Exp | $\chi^2$ | *P* |
| P268S | 478 | 385 | 347.2 | 18.3 | $1.9 \times 10^{-5}$ | 309 | 167 | 178.3 | 2.84 | NS |
| R702W | 499 | 129 | 102.9 | 17.0 | $3.8 \times 10^{-5}$ | 325 | 30 | 38.6 | 5.6 | 0.018 |
| G908R | 486 | 45 | 38.9 | 2.4 | 0.122 | 313 | 12 | 18.8 | 6.2 | 0.013 |
| 3020insC | 512 | 102 | 76.5 | 21.6 | $3.4 \times 10^{-6}$ | 325 | 15 | 19.9 | 3.9 | 0.049 |

n, number of affected offspring; Obs, observed transmissions; Exp, expected transmissions to affected offspring; NS, not significant.

**Table 3.** *NOD2* Genotype and Allele Frequencies for Patients and Controls

| SNP | Cohort | Genotype | | | Allele frequency (%) | $P^a$ |
|---|---|---|---|---|---|---|
| | | −/− | −/+ | +/+ | | |
| P268S | CD | 175 | 153 | 57 | 34.7 | 0.0002 |
| | UC | 271 | 195 | 40 | 27.2 | NS |
| | Controls | 134 | 96 | 11 | 24.5 | |
| R702W | CD | 346 | 64 | 6 | 9.1 | 0.0001 |
| | UC | 512 | 39 | 1 | 3.7 | NS |
| | Controls | 253 | 19 | 0 | 3.5 | |
| G908R | CD | 390 | 28 | 0 | 3.4 | 0.0014 |
| | UC | 490 | 16 | 0 | 1.6 | NS |
| | Controls | 264 | 3 | 0 | 0.6 | |
| 3020insC | CD | 371 | 43 | 6 | 6.6 | 0.0002 |
| | UC | 541 | 17 | 0 | 1.5 | NS |
| | Controls | 271 | 12 | 0 | 2.1 | |

NS, not significant; −/−, homozygous wild-type; −/+, heterozygous; +/+, homozygous mutant.
[a]*P* value for difference in allele frequency between patients and controls.

gether with 1 other mutation were preferentially transmitted to affected offspring; however, the haplotype carrying P268S alone was significantly undertransmitted ($P = 0.0054$), suggesting that P268S does not contribute independently to disease risk. Patients and controls were also stratified according to whether they carried at least 1 copy of R702W, G908R, or 3020insC. The frequency of P268S between patients and controls was not significantly different in individuals who did not carry another mutation (CD, 21.3%; UC, 22.3%; controls, 21.2%), supporting the hypothesis that P268S does not play a functional role in CD susceptibility. In this independent sample of 1542 individuals, the estimated frequency of haplotypes carrying an *NOD2* susceptibility mutation (R702W, G908R, 3020insC) (defined as an '*NOD2* haplotype') accounted for 13% of possible haplotypes. Haplotypes carrying more than 1 *NOD2* mutation were rare, each with an estimated frequency of <0.2%.

### *NOD2* Genotype Relative Risks

*NOD2* genotypes were defined by the 4 possible haplotypes (wild-type, R702W, G908R, and 3020insC) and genotype relative risks calculated under the assumption that 2 of these mutations do not occur on the same haplotype (Table 5). The odds ratio for a heterozygous individual ranged between 2.6 and 6.3 for the 3 mutations, with the highest risk associated with G908R. The risks for homozygotes and compound heterozygotes were similar, with a combined odds ratio of 23.4. In 62 homozygous or compound heterozygous individuals, 55 were affected with CD. A total of 27.8% of patients with CD carried a single copy of an *NOD2* mutation and 8.0% carried 2 mutations, giving a total of 35.8% of patients with CD carrying at least 1 copy of an *NOD2* mutation. The population prevalence of CD is approximately 1 in 1000, so the estimated penetrances of *NOD2* genotypes, assuming the control genotypes to be in Hardy–Weinberg equilibrium, were 0.03% (wild-type −/−), 0.22% (heterozygous −/+), and 1.73% (compound heterozygous, homozygous +/+). Although no controls homozygous for *NOD2* mutations were found, 7 *NOD2* homozygous patients with UC were observed, suggesting incomplete penetrance of this genotype for

**Table 4.** *NOD2* Haplotype Transmissions to 489 CD-Affected Offspring From the Family Cohort Calculated Using TRANSMIT (Only Haplotypes With Frequency >1% Shown)

| Haplotype | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| P268S | R702W | G908R | 3020insC | Frequency (%) | Obs | Exp | $\chi^2$ | *P* |
| − | − | − | − | 65.8 | 577.7 | 609.9 | 13.06 | $3.0 \times 10^{-4}$ |
| + | − | − | − | 17.0 | 139.3 | 157.3 | 7.73 | 0.0054 |
| + | + | − | − | 7.4 | 113.0 | 88.8 | 16.98 | $3.8 \times 10^{-5}$ |
| + | − | + | − | 2.6 | 38.2 | 32.2 | 3.34 | 0.068 |
| + | − | − | + | 5.2 | 92.0 | 69.5 | 19.28 | $1.1 \times 10^{-5}$ |

Obs, observed transmissions; Exp, expected transmissions to affected offspring.

**Table 5.** Disease Risk Estimates From Case-Control Cohort for *NOD2* Mutation Genotypes

| Genotype | CD (n = 688) | UC (n = 604) | Control (n = 250) | Odds ratio for CD | 95% CI |
|---|---|---|---|---|---|
| Homozygous −/− | 442 | 517 | 216 | 1 | |
| Heterozygous −/+ | | | | | |
| R702W | 96 | 46 | 19 | 2.66 | (1.59, 4.46) |
| G908R | 36 | 18 | 3 | 6.33 | (2.15, 18.59) |
| 3020insC | 59 | 16 | 12 | 2.59 | (1.37, 4.91) |
| Compound heterozygous +/+ | 29 | 2 | 0 | 24.85 | (3.82, 161.67) |
| Homozygous +/+ | 26 | 5 | 0 | 22.19 | (3.96, 124.36) |
| *NOD2* Composite genotype | | | | | |
| Heterozygous | 191 | 80 | 34 | 2.96 | (1.99, 4.40) |
| Compound heterozygous/homozygous | 55 | 7 | 0 | 23.38 | (6.50, 84.10) |

−/−, homozygous wild-type; −/+, heterozygous; +/+, homozygous mutant.
CI, confidence interval.

CD. The population attributable risk for *NOD2* was 26%.

The contribution of *NOD2* to the general phenotype of IBD was assessed by comparing the frequency of the *NOD2* haplotype in CD from (1) sporadic cases, (2) families affected only with CD, and (3) families of mixed phenotype with cases of both CD and UC. The frequency of the *NOD2* haplotype was significantly higher in the cases from families with CD (30.9%, n = 173) than cases with CD from mixed families (19.2%, n = 99, $P = 0.004$) or sporadic CD cases (19.3%, n = 405, $P < 0.001$). This provides additional evidence that other IBD susceptibility genes contribute to the risk of CD.

### Stratification of CD Cases by Site of Disease

Clinical data on the site of disease were available for 444 familial or sporadic cases of CD. The possible effect of *NOD2* mutations on specific CD subphenotypes was therefore investigated by stratification of CD cases by site of disease. The frequency of mutant *NOD2* haplotypes was significantly increased in ileum-specific disease cases compared with colon-specific disease cases (Table 6), with an increase in frequency observed for each of the *NOD2* mutations (data not shown). The *NOD2* mutant haplotype frequency in cases with both colonic and ileal disease were also significantly greater than colon-specific disease, suggesting an association between *NOD2* and the presence of ileal disease. The increased mutation frequency in ileal disease was observed in familial and in sporadically ascertained cases, so the difference in frequency between familial and sporadic cases is not a confounding factor for site of disease. The odds ratio of 2.04 (95% confidence interval, 1.23–3.38) indicates a 2-fold increased risk of ileal disease in patients who carry a *NOD2* mutation.

## Discussion

In this study, a cohort of more than 1200 unrelated cases of IBD has been genotyped for 4 mutations in the coding region of the *NOD2* gene to address several questions relating to its contribution to IBD susceptibility. First, we used both TDT analyses and case-control studies to assess which of the mutations are independent risk factors for CD. The P268S, R702W, and 3020insC mutations were strongly associated with CD in the TDT, whereas all 4 mutations were present at a significantly higher frequency in patients with CD than in controls. The nonsignificant excess of transmission of the G908R mutation in the TDT is very likely because of its low population frequency, because only a small number of transmissions will be informative in even a large TDT cohort. A greater sample size is required to obtain power

**Table 6.** Stratification of CD Cases by Site of Disease

| Clinical/epidemiologic factor | n | *NOD2* composite genotype −/− | −/+ | +/+ | Allele frequency (%) | *P* |
|---|---|---|---|---|---|---|
| Site of disease | | | | | | |
| Ileum only | 104 | 60 | 32 | 12 | 26.9 | 0.0004[a] |
| Ileocolitis | 234 | 156 | 64 | 14 | 19.7 | 0.036[a] |
| Colon only | 106 | 83 | 19 | 4 | 12.7 | |

−/−, homozygous wild-type; −/+, heterozygous; +/+, homozygous mutant; n, number of cases.
[a]*P* value for difference in allele frequency between ileal disease and pure colonic disease.

comparable with the case-control design.[24] The analysis of transmission of *NOD2* haplotypes to CD-affected offspring showed that there was overtransmission of P268S only in the presence of one of the other 3 mutations. Thus, it is not an independent risk factor for CD, and its association in the TDT and case-control studies results from it being in linkage disequilibrium with the rarer mutations. We cannot formally exclude the possibility that other as-yet undetected mutations are in tight linkage disequilibrium with R702W, G908R, and 3020insC, but this now seems unlikely given the extent to which we and others have screened the *NOD2* gene.

We found no evidence of excess transmission of any of the 4 mutations to more than 300 patients with UC, and their frequencies in more than 500 sporadic UC cases were very similar to those in controls. Thus, *NOD2* does not seem to be a significant risk factor in UC, which is consistent with weak or absent linkage of the UC phenotype to this region of chromosome 16 and the lack of association observed in initial studies of smaller numbers of patients with UC.[10–12] The reason for the modest undertransmission of mutations to patients with UC is not clear but may simply reflect the small number of transmissions scored in the UC TDT.

Our estimation of disease risk based on 688 unrelated cases of CD shows a 3-fold increased risk in *NOD2* mutation heterozygotes but a >20-fold increased risk in homozygotes or compound heterozygotes. This gene-dosage effect is consistent with the suggested recessive model of inheritance for CD[4,5] and with the work of Hugot et al.[10] An important and unresolved issue is disease penetrance in homozygotes. In our study, the *NOD2* mutations were in Hardy–Weinberg equilibrium in the control group, with no +/+ genotypes observed in 250 individuals. Assuming a Hardy–Weinberg distribution in controls, the penetrance for the +/+ genotype was <2%. There were 7 individuals with this genotype in the UC group. Hugot et al. found no +/+ genotypes in 103 controls. In view of the low frequency of the mutations in the general population, analysis of a large control cohort will be required to obtain a robust estimate of the population frequency of the +/+ genotype and hence of the penetrance of *NOD2* mutations.

An additional important question is the extent of the contribution of *NOD2* to the CD phenotype. In our series of unrelated CD patients, about 36% carried 1 or 2 mutations in *NOD2*, and the overall population risk for CD attributable to *NOD2* mutations was 26%. Hugot et al.[10] detected additional rare missense mutations in *NOD2* with a combined frequency of 17% in patients with CD compared with 5% in controls. It is plausible that rare variants contribute to disease risk in complex disorders,[25] but very large sample sizes would be required to detect the effect of these mutations in association studies. We found a higher frequency of *NOD2* mutations in patients with CD from families affected only with CD than in patients with CD from mixed-phenotype families or in sporadic cases of CD. It may be that the "pure" CD families have a less complex suite of susceptibility genes than the other 2 groups, but this and other case-control studies show that *NOD2* also contributes significantly to the sporadic group.[10–12] However, although *NOD2* clearly makes a substantial contribution to the overall risk for CD, stratification of families by the presence of *NOD2* mutations suggests that this gene does not account for all of the linkage observed in the pericentromeric region of chromosome 16.[10,26]

Finally, we have begun to address the question of whether *NOD2* mutations are associated with a particular subtype of CD. Linkage and association between disease and locus restricted to clinical or epidemiologic subtypes of IBD have been reported.[27–29] Bayless et al.[30] showed concordance for site of disease within familial cases, which might indicate that genetic factors influence location of disease. We investigated this hypothesis in relation to *NOD2* by stratification of CD cases according to site of disease. We found that the frequency of the mutant *NOD2* haplotypes was 2.1-fold higher in ileum-specific disease than that in disease restricted to the colon and was 1.6-fold higher in ileocolitis. Colombel et al.[31] identified a higher rate of small bowel disease in familial cases than in sporadic cases, but in our series the association between *NOD2* and ileal disease was found in both familial and sporadic cases of CD. Family history and site of disease therefore do not seem to be confounding factors. Our findings support the hypothesis that site-specific susceptibility genes exist[30] and suggest that *NOD2* mutations are associated with CD occurring in the ileum. This observation may be related to the fact that certain gut bacteria occur at much higher concentrations in the colon than in the ileum. It is possible that additional mechanisms to combat higher levels of bacteria have evolved in the colon, whereas the ileum may be more dependent on the function of *NOD2*. Testing this hypothesis will require clarification of the role of NOD2 in the immune response.

In summary, we have determined the extent of the contribution of the common *NOD2* mutations to the etiology of IBD in a large cohort of patients and provided a genetic model for disease susceptibility. Through strat-

ification by site of disease, we have shown that *NOD2* is primarily associated with ileal disease. These findings may direct future investigations into the precise function of *NOD2* and its contribution to innate immune responses in the gastrointestinal tract.

## References


1. Probert CS, Jayanthi V, Rampton DS, Mayberry JF. Epidemiology of inflammatory bowel disease in different ethnic and religious groups: limitations and aetiological clues. Int J Colorectal Dis 1996;11:25–28.
2. Sartor JB. Pathogenesis and immune mechanisms of chronic inflammatory bowel diseases. Am J Gastroenterol 1997;92:5S–11S.
3. Orchard TR, Satsangi J, van Heel D, Jewell DP. Genetics of inflammatory bowel disease: a reappraisal. Scand J Immunol 2000;51:10–17.
4. Forabosco P, Collins A, Latiano A, Annese V, Clementi M, Andriulli A, Mortina P, Devoto M, Morton NE, on behalf of the GISC. Combined segregation and linkage analysis of inflammatory bowel disease in the IBD1 region using severity to characterise Crohn's disease and ulcerative colitis. Eur J Hum Genet 2000;8: 846–852.
5. Orholm M, Iselius L, Sorensen TI, Munkholm P, Langholz E, Binder V. Investigation of inheritance of chronic inflammatory bowel diseases by complex segregation analysis. Br Med J 1993; 306:20–24.
6. Monsen U, Iselius L, Johansson C, Hellers G. Evidence for a major additive gene in ulcerative colitis. Clin Genet 1989;36: 411–414.
7. Kuster W, Pascoe L, Purrmann J, Funk S, Majewski F. The genetics of Crohn disease: complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives. Am J Med Genet 1989;32:105–108.
8. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JF, Thomas G. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 1996;379: 821–823.
9. IBD International Genetics Consortium. International collaboration provides convincing linkage replication in complex diseases through analysis of a large pooled data set: Crohn disease and chromosome 16. Am J Hum Genet 2001;68: 1165–1171.
10. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard J-P, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of *NOD2* leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603.
11. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar J-P, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH. A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. Nature 2001;411:603–606.
12. Hampe J, Cuthbert A, Croucher PJP, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPherson AJS, Bridger S, van Deventer S, Forbes A, Nikolaus S, Lennard-Jones JE, Foelsch UR, Krawczak M, Lewis C, Schreiber S, Mathew CG. Association between insertion mutation in *NOD2* gene and Crohn's disease in German and British population. Lancet 2001; 357:1925–1928.
13. Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989;170:2–6.
14. Podolsky DK. Inflammatory bowel disease. N Engl J Med 1991; 325:928–937.
15. Livak KJ, Marmaro J, Todd JA. Towards fully automated genome-wide polymorphism screening. Nat Genet 1995;9:341–342.
16. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 1998;63:259–266.
17. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–516.
18. Clayton D. A generalisation of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 1999;65:1170–1177.
19. Xie X, Ott J. Testing linkage disequilibrium between a disease gene and marker loci. Am J Hum Genet 1993;53:1107.
20. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation test for allelic associations. Hum Hered 2000;50:133–139.
21. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 1995;29:311–322.
22. Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, Macpherson AJS, Cardon LR, Sakul H, Harris TJR, Buckler A, Hall J, Stokkers P, van Deventer SJH, Nürnberg P, Mirza MM, Lee JCW, Lennard-Jones JE, Mathew CG, Curran ME. A genome-wide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. Am J Hum Genet 1999;64:808–816.
23. Lie BA, Ronningen KS, Akselsen HE, Thorsby E, Undlien DE. Application and interpretation of transmission/disequilibrium tests: transmission of HLA-DQ haplotypes to unaffected siblings in 526 families with type I diabetes. Am J Hum Genet 2000;66: 740–743.
24. Lewis CM, Fisher SA. Power of the TDT to detect association: effect of linkage disequilibrium between genetic markers and an unobserved disease mutation. GeneScreen 2000;1:69–71.
25. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124–137.
26. Hampe J, Frenzel H, Mirza MM, Croucher PJ, Cuthbert A, Mascheretti S, Huse K, Platzer M, Bridger S, Meyer B, Nurnberg P, Stokkers P, Krawczak M, Mathew CG, Curran M, Schreiber S. Evidence for a *NOD2*-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. Proc Natl Acad Sci U S A 2002;99:321–326.
27. Brant SR, Panhuysen CIM, Bailey-Wilson JE, Rohal PM, Lee S, Mann J, Ravenhill G, Kirschner BS, Hanauer SB, Cho JH, Bayless RM. Linkage heterogeneity for the *IBD1* locus in Crohn's pedigrees by disease onset and severity. Gastroenterology 2000; 119:1483–1490.
28. de la Concha EG, Fernandez-Arquero M, Lopez-Nava G, Martin E, Allcock RJ, Conejero L, Paredes JG, Diaz-Rubio M. Susceptibility to severe ulcerative colitis is associated with polymorphism in the central MHC gene IKBL. Gastroenterology 2000;119:1491–1495.
29. Parkes M, Barmada MM, Satsangi J, Weeks DE, Jewell DP, Duerr RH. The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. Am J Hum Genet 2000;67: 1605–1610.
30. Bayless TM, Tokayer AZ, Polito JM, Quaskey SA, Mellits ED, Harris ML. Crohn's disease: concordance for site and clinical type in affected family members – potential hereditary influences. Gastroenterology 1996;111:573–579.
31. Colombel JF, Grandbastien B, Gower-Rousseau C, Plegat S,

Evrard JP, Dupas JL, Gendre JP, Modigliani R, Belaiche J, Hostein J, Hugot JP, van Kruiningen H, Cortot A. Clinical characteristics of Crohn's disease in 72 families. Gastroenterology 1996;111: 813–815.

Received October 11, 2001. Accepted January 10, 2002.

Address requests for reprints to: Christopher G. Mathew, Ph.D., Division of Medical and Molecular Genetics, GKT School of Medicine, King's College London, 8th Floor Guy's Tower, Guy's Hospital, London SE1 9RT, England. e-mail: christopher.mathew@kcl.ac.uk; fax: (44) 207-955-4644.

Supported in the United Kingdom by the Wellcome Trust, the Generation Trust, the TR-Golden Charitable Trust, the Kati Jacobs Appeal, Axys Pharmaceuticals Inc., and the Special Trustees of the Royal Victoria Infirmary and by an equipment grant from the Guy's and St. Thomas' Hospitals Charitable Trust. Supported in Germany by the Deutsche Forschungsgemeinschaft (For 423), a Training and Mobility of Research Network grant of the European Union (ERB-4061-PL-97-0389), a competence network "Chronisch-entzündliche Darmerkrankungen," the German Human Genome Project, and the National Genome Research Network (all funded by the German Federal Department for Research and Education).

A.P.C. and S.A.F. contributed equally to this work.

The authors thank Drs. J. Lennard-Jones, A. Macpherson, J. Lee, S. Bridger, S. Hodgson, M. Hudson, and S. van Deventer for patient ascertainment and Gillian Scott, Jo Hirst, and Alison Craggs for help in obtaining samples from IBD patients and families.