

NOD2 (CARD15), the first susceptibility gene for Crohn's disease

"This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."
Sir Winston Churchill, 1942.

Hugot JP, Chamaillard M, Zouali H, *et al.* Association of *NOD2* leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599-603.

Abstract

Crohn's disease and ulcerative colitis, the two main types of chronic inflammatory bowel disease, are multifactorial conditions of unknown aetiology. A susceptibility locus for Crohn's disease has been mapped to chromosome 16. Here we have used a positional-cloning strategy, based on linkage analysis followed by linkage disequilibrium mapping, to identify three independent associations for Crohn's disease: a frameshift variant and two missense variants of *NOD2*, encoding a member of the Apaf-1/Ced-4 superfamily of apoptosis regulators that is expressed in monocytes. These *NOD2* variants alter the structure of either the leucine-rich repeat domain of the protein or the adjacent region. *NOD2* activates nuclear factor NF- κ B; this activating function is regulated by the carboxy-terminal leucine-rich repeat domain, which has an inhibitory role and also acts as an intracellular receptor for components of microbial pathogens. These observations suggest that the *NOD2* gene product confers susceptibility to Crohn's disease by altering the recognition of these components and/or by over-activating NF- κ B in monocytes, thus documenting a molecular model for the pathogenic mechanism of Crohn's disease that can now be further investigated.

Ogura Y, Bonen D, Inohara N, *et al.* A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature* 2001;411:603-6.

Abstract

Crohn's disease is a chronic inflammatory disorder of the gastrointestinal tract which is thought to result from the effect of environmental factors in a genetically predisposed host. A gene location in the pericentromeric region of chromosome 16, *IBD1*, that contributes to susceptibility to Crohn's disease has been established through multiple linkage studies, but the specific gene(s) has not been identified. *NOD2*, a gene that encodes a protein with homology to plant disease resistance gene products is located in the peak region of linkage on chromosome 16. Here we show, by using the transmission disequilibrium test and case-control analysis, that a frameshift mutation caused by a cytosine insertion, 3020insC, which is expected to encode a truncated *NOD2* protein, is associated with Crohn's disease. Wild-type *NOD2* activates nuclear factor NF- κ B, making it responsive to bacterial lipopolysaccharides; however, this induction was deficient in mutant *NOD2*. These results implicate *NOD2* in susceptibility to Crohn's disease, and suggest a link between an innate immune response to bacterial components and development of disease.

Comment

A "working model" of the pathophysiology of Crohn's disease (CD) is of an abnormal immune response to enteric bacteria in genetically susceptible individuals. There is a greater concordance for CD in monozygotic than dizygotic twins (concordance for CD in monozygotic twins 37% compared with 7% for dizygotic twins), suggesting a major genetic contribution to the pathogenesis of CD. The risk to siblings of affected individuals relative to the general population (λ_s) is 15-40 for CD and the calculated heritability for CD is greater than that for schizophrenia or asthma and at least equal to that of type 1 diabetes.^{1,2}

Evidence for the role of enteric bacteria in the pathogenesis of CD comes from data demonstrating that genetically modified mice utilised as models for inflammatory bowel disease/CD fail to manifest the disease in germ free environments and only develop disease when exposed to normal enteric commensal organisms. In addition, Crohn's colitis frequently improves when the faecal stream is diverted following formation of a loop ileostomy and symptoms from CD are transiently improved following administration of antibiotics.

Geneticists have proposed that the genes for complex genetic diseases will be discovered by one or more of the following three methods:

- Positional cloning using microsatellite markers and single nucleotide polymorphisms (SNPs) in multiply affected families.
- Identification of candidate genes through an understanding of the pathophysiology of the underlying defects leading to disease.
- Gene expression studies (for example, microarrays) comparing differential gene expression in tissue from affected and unaffected individuals.

In 1996, Jean-Pierre Hugot *et al* published the first genome wide scan identifying a susceptibility locus for CD adjacent to the centromere on chromosome 16,³ *IBD1*. Replication studies in different populations is the "gold standard" for confirmation of loci, and linkage to *IBD1* has been replicated by a number of centres, most notably by the IBD International Genetics Consortium who demonstrated a LOD score of 5.8 for CD at this locus.⁴

Two independent groups have now identified the gene and subsequent polymorphism(s) at *IBD1* that confer susceptibility to CD.^{5,6} Hugot *et al* continued their previous work by fine mapping *IBD1* using 26 microsatellite markers approximately 1 centiMorgan apart.⁵ Using the transmission disequilibrium test they identified a borderline significant association ($p < 0.05$) in 108 families between CD and an allele of the microsatellite marker

Table 1 Risk of developing Crohn's disease (CD) according to inheritance of a combination of the three polymorphisms based on the data of Hugot and colleagues⁵

| Genotype | Relative risk of CD | Absolute risk of CD |
|-----------------------|---------------------|---------------------|
| No mutations | 1 | 0.0007 |
| Simple heterozygous | 3 | 0.002 |
| Compound heterozygous | 44 | 0.03 |
| Homozygous | 38 | 0.03 |

D16S3136. In order to confirm this association, the authors tested this marker in a further 76 families and found a weak association but with a different allele of this marker. Many investigators may have been dissuaded from pursuing this avenue of investigation by these borderline results but the authors hypothesised that while these may be type 1 errors, their results "may reflect true association in two sets of families drawn from genetically different populations." Thus the authors sequenced a bacterial artificial chromosome clone containing this marker and found significant association with several SNPs and CD. This sequence and a human leucocyte cDNA library were used to identify a novel gene, *NOD2*. The gene was sequenced in 50 CD patients and a total of 13 polymorphisms were identified. Two amino acid changing SNPs (Arg702Trp and Gly908Arg) and one frameshift mutation (3020insC), causing a premature stop codon, were significantly associated with CD ($p=0.001$, $p=0.003$, and $p=0.000006$ respectively). (It is important to note that the nomenclature identifying the polymorphisms used is different in the two papers as *NOD2* has alternative splice sites. In this commentary we have used the nomenclature used by Ogura and colleagues⁶ (see fig 1).) Three haplotypes, each containing one of the three polymorphisms, were also significantly associated with CD (interestingly, another variant, Pro268Ser, occurred on each of these haplotypes). Arg702Trp, Gly908Arg, or the frameshift mutation (3020insC) did not occur on the same haplotype.

The risk of CD conferred from inheriting these mutations is shown in table 1. Inheritance of CD occurring as a result of *NOD2* polymorphisms behaves part way between a gene dosage and an autosomal recessive effect. Heterozygotes have a relatively small increased risk (of the order of that of smokers) for developing CD whereas in contrast, homozygotes and compound heterozygotes have a much greater relative risk. There were no homozygotes (or compound heterozygotes) seen in the healthy control groups (although there is one homozygote with ulcerative colitis (UC)). Hugot *et al* found no overall association between UC and *NOD2* variants.

A group from the University of Michigan Medical School had previously identified NOD1,⁷ an intracellular protein composed of an N terminal caspase recruitment domain (CARD), a centrally located nucleotide binding domain (NBD), and a leucine rich repeat (LRR) domain at its C terminus which could activate nuclear factor κ B (NF κ B) and also promote apoptosis. NOD2 was identified by searching the public databases for genes encoding similar proteins to NOD1 and the gene coding for NOD2 was located on chromosome 16q12. NOD2 differs from NOD1 in that it is the first protein known to contain two CARDS (see fig 1). NOD2 is expressed primarily in monocytes and following stimulation by bacterial lipopolysaccharide (LPS), which occurs at the LRR domain, activates NF κ B.⁸ The same group in collaboration with groups from Chicago, Pittsburgh, Baltimore, and Cleveland identified that a gene involved in the innate immune system and located in a susceptibility region would be an ideal candidate gene for CD. The gene was sequenced in patients with CD known to have linkage to 16q12 and the

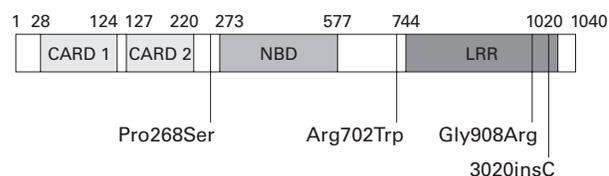


Figure 1 Representation of the *NOD2* domain structure and the position of the polymorphisms associated with Crohn's disease. The numbers represent the amino acid positions. The C insertion at nucleotide 3020 causes a Leu1007Pro amino acid change followed by a stop codon. CARD, caspase recruitment domain; NBD, nucleotide binding domain; LRR, leucine rich repeat.

same three polymorphisms described by Hugot *et al* were identified and found to be significantly associated with CD.⁶

Ogura *et al* also performed functional studies investigating the consequences of the frameshift mutation (3020insC) by cotransfecting embryonic kidney cell lines with 3020insC or wild type plasmids together with an NF κ B reporter construct. The mutant *NOD2* response to LPS from a variety of bacteria was significantly reduced compared with wild type, thereby confirming a functional consequence of the polymorphism.

There has been one subsequent paper reporting a similar association between the *NOD2* frameshift mutation in German and British populations with CD.⁹

Finding a susceptibility gene for CD is exciting but many questions remain unanswered:

• Why does a frameshift mutation in the LRR region of the gene predispose to CD?

In the original paper identifying *NOD2*,⁵ mutational analysis demonstrated that expression of a *NOD2* mutant form lacking the entire LRR region resulted in enhanced NF κ B activity whereas the frameshift mutation causing a truncated protein missing the final 33 amino acids (of the LRR region) resulted in low NF κ B activity. These results are not mutually exclusive but further work is needed to clarify the role of the LRR region in stimulating (or inhibiting) NF κ B activity. Other potential explanations may be: that the frameshift mutation results in innate hyporesponsiveness thereby inducing an abnormal adaptive response to enteric bacteria; that the truncated protein may lead to elevated NF κ B when stimulated by an as yet untested (?unknown) bacterial LPS; that the frameshift mutation may have a differential effect on caspase 9 induced apoptosis; and that wild type *NOD2* may induce anti-inflammatory cytokines such as interleukin 10.

• What are the functional consequences of the other identified *NOD2* polymorphisms?

Recently Hugot's group have demonstrated that polymorphisms within the NBD domain of *NOD2* (*CARD15*) are associated with Blau syndrome (granulomatous arthritis, uveitis, and skin rash with camptodactyly).¹⁰

• Are there other mutations in *NOD2* associated with CD?

Hugot *et al* identified 35, mostly rare amino acid changing *NOD2* polymorphisms, in total, and also commented that the three associated polymorphisms do not account for all of the linkage demonstrated at 16q12. This suggests that other *NOD2* polymorphisms or even other genes in this region may predispose to CD.

• What are the other genes for CD?

The data of Hugot *et al* show that approximately 40% of CD patients carry one of the three *NOD2* variants (although approximately 15% of healthy controls also carry

one of these variants). Other replicated loci do exist, particularly on chromosomes 6 (*IBD3*), 12 (*IBD2*), and 14 (*IBD4*), and it is possible that these areas also contain genes involved in the NOD2/innate immune system pathways. It may be that a NOD2 defect alone may not be sufficient for the development of CD and that interaction with other genes is necessary.

- **What is the prevalence of NOD2 mutations in “sporadic” CD?**

If the prevalence of NOD2 polymorphisms in sporadic cases is as common as in familial cases, this would suggest that all CD may be occurring in genetically susceptible individuals.

- **Why are NOD2 polymorphisms so common in the healthy population?**

The 7% allele frequency for at least one of the three SNPs in healthy controls suggests that these mutations might confer some evolutionary benefit, perhaps against an infection/disease.

- **Why do the three associated polymorphisms all occur on a common background that includes the Pro268Ser polymorphism?**

This could have occurred by chance but is unlikely. It is more probable that if these mutations have persisted due to positive selection, then two NOD2 variants may be needed for any evolutionary benefit to be manifest.

- **Will these discoveries lead to a new molecular classification of disease?**

Is there a particular CD phenotype associated with NOD2 polymorphisms and are other CD-like diseases (for example, Behçets, gastrointestinal tuberculosis, etc.) associated with NOD2 variants? If this is the case, classification based on genotype may allow a better prediction of response to medication, thereby enabling a degree of individualisation of therapy and may also predict the need (or not) for surgery.

- **Will reanalysis of linkage studies and genome wide scans stratified for NOD2 status demonstrate novel loci or evidence of epistasis between the chromosome 16 loci and other loci?**

It is clear that a considerable amount of work is required to answer these and other questions posed by these important discoveries. It is still too early for there to be any clear

benefits for the clinician or patient in clinical practice and it is important to remember that the majority of subjects with NOD2 variants are healthy and the clinical course of NOD2 positive (or negative) CD patients is not yet known. Patients should however, if willing, be encouraged to participate in studies which may help researchers solve some of the outstanding questions and identify the other genes involved in IBD. Identification of this gene may well have insurance implications, not only for NOD2 positive patients, but also for their relatives.

These seminal papers demonstrate the benefits of the human genome project and the publicly available genetic databases and also prove, arguably for the first time, that the genes for complex polygenic diseases can be identified using both the positional cloning and candidate gene approaches. These data also support previous results suggesting that CD is a disease of an abnormal (innate) immune response to enteric bacteria in genetically susceptible individuals.

D P B MCGOVERN
D A VAN HEEL
T AHMAD
D P JEWELL

Wellcome Trust Centre for Human Genetics and Gastroenterology Unit,
University of Oxford, Oxford, OX3 7BN, UK.

Correspondence to: D P B McGovern, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.
Dermot@well.ox.ac.uk

Conflict of interest: Drs McGovern and van Heel, and Professor Jewell are members of the IBD International Genetics Consortium.

- 1 Satsangi J, Rosenberg WMC, Jewell DP. The prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *Eur J Gastroenterol Hepatol* 1994;6:413–16.
- 2 Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994;265:2037–48.
- 3 Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;379:821–3.
- 4 Cavanaugh J, The IBD International Genetics Consortium. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn's disease and chromosome 16. *Am J Hum Genet* 2001;68:1165–71.
- 5 Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599–603.
- 6 Ogura Y, Bonen D, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;411:603–6.
- 7 Inohara N, Koseki T, del Peso L, et al. Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB. *J Biol Chem* 1999;274:14560–8.
- 8 Ogura Y, Inohara N, Benito A, et al. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-B. *J Biol Chem* 2001;276:4812–18.
- 9 Hampe J, Cuthbert A, Croucher PJP, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;357:1925–8.
- 10 Miceli-Richard C, Lesage S, Rybojad M, et al. CARD15 mutations in Blau syndrome. *Nat Genet* 2001;29:19–20.