Unravelling the complex genetics of inflammatory bowel disease

R K Russell, D C Wilson, J Satsangi

The rapid pace of progress in molecular genetics over the past 15 years—since the seminal description of the polymerase chain reaction—has led to the identification of the genes involved in many single gene disorders. These successes in the laboratory have already led directly to clinical applications in diagnosis, pharmacogenetics, and the development of new therapies. Progress in unravelling the genetics of complex diseases has been less straightforward. However, real excitement has followed the identification of the NOD2/CARD15 gene as an important determinant of susceptibility to Crohn’s disease. Not only has this finding provided a proof of principle for the technique of genome-wide scanning in complex disorders, but the discovery also has given real insight into the primary pathophysiology involved in chronic inflammatory bowel disease. The background to this discovery and its implications form the basis for the present article.

The chronic inflammatory bowel diseases, Crohn’s disease and ulcerative colitis are now common causes of gastrointestinal morbidity in children and young adults in the United Kingdom. Moreover, whereas incidence rates in adults appeared to have stabilised in recent years, compelling data both from our own unit, and elsewhere in the United Kingdom, suggest that the incidence of early onset cases in children continues to rise. These studies have shown a rise in the incidence of early onset inflammatory bowel disease in Scottish children from 2.6 cases per 100,000 population in 1968 to 6.5 cases per 100,000 population in 1999. Crohn’s disease is characterised by patchy transmural inflammation affecting any part of the gastrointestinal tract, whereas ulcerative colitis characteristically is limited to the colon, producing continuous mucosal inflammation always involving the rectum. Extra-intestinal manifestations affecting the skin, joints, and the eyes occur in both Crohn’s disease and ulcerative colitis. Colorectal cancer is a recognised complication of long standing colonic involvement. A recent meta-analysis has suggested the risk of colorectal cancer in children with ulcerative colitis is double that of adult onset disease. In a proportion of patients suffering from inflammatory bowel disease affecting the colon, it proves impossible to distinguish between Crohn’s disease and ulcerative colitis, even after colectomy, and a diagnosis of indeterminate colitis is made.

In the 21st century, inflammatory bowel disease has a great impact in both adult as well as paediatric gastroenterology, related to the morbidity of the illnesses and the therapies now available. Growth impairment is a particular feature in children but has effects lasting into adult life, with up to 35% of subjects showing evidence of permanent growth failure. In addition to growth problems, pubertal development, education, employment potential, and quality of life all suffer as a consequence of these chronic disabling illnesses.

GENE-ENVIRONMENTAL INTERACTIONS UNDERLIE THE PATHOGENESIS OF INFLAMMATORY BOWEL DISEASE

There are strong epidemiological data to suggest that both environmental and genetic influences are involved in the pathogenesis of chronic inflammatory bowel diseases, Crohn’s disease and ulcerative colitis. The temporal trends in incidence—in particular the increase in the incidence of inflammatory bowel disease over the past three decades in young people—cannot be explained by the influence of genetic factors alone. The environmental factors involved in disease pathogenesis remain under investigation.

Smoking habit clearly influences susceptibility to both Crohn’s disease and ulcerative colitis. However smoking habit has contrasting effects: in Crohn’s disease smoking is associated with an increased susceptibility to disease, rapid disease progression, and further need for surgery and immune suppression. In contrast, smoking is protective against the development of ulcerative colitis; dose-response relations have been identified. As most children do not smoke at the time of inflammatory bowel disease diagnosis the role of passive smoking is more relevant.

Lashner et al identified smoking at birth rather than at the time of diagnosis of inflammatory bowel disease as a risk factor both for the development Crohn’s disease and ulcerative colitis in childhood.

Attempts to identify infective agents involved in disease pathogenesis have attracted great attention. The role of Mycobacterium paratuberculosis in the pathogenesis of Crohn’s disease remains a subject of impassioned debate. The results of a number of studies are awaited—notably an MRC funded study in the United Kingdom and a placebo controlled double blinded study of anti-mycobacterial therapy in Australia.

See end of article for authors’ affiliations

Correspondence to:
Dr R Russell,
Department of Medical Sciences, Edinburgh
EH4 2U, UK;
richardkrussell71@hotmail.com

Accepted
15 November 2003
measles virus and the MMR vaccine have been the subject of an editorial in this journal, and have been evaluated in great detail elsewhere. The most compelling evidence for the involvement of microbial agents implicates the gut flora. Animal models of colitis—genetically engineered, chemically induced, or spontaneous—all require the presence of gut flora in order for disease to become manifest. These data complement clinical studies—the effect of faecal diversion in Crohn’s disease, and increasing evidence that antibiotic and probiotic therapy may attenuate disease.

Diet, childhood deprivation, breast feeding, and passive smoking all require to be evaluated in large well designed studies of early onset inflammatory bowel disease.

**EVIDENCE FOR GENETIC SUSCEPTIBILITY**

Strong epidemiological data provided the basis for the recent molecular genetic studies in inflammatory bowel disease. In particular, ethnic differences in disease susceptibility, and concordance rates in twin pairs and multiply affected families all provided the catalyst for detailed evaluation of the molecular genetics of Crohn’s disease and ulcerative colitis. Ethnic differences in the susceptibility of inflammatory bowel disease are well documented. The Ashkenazi Jewish population, living in Western Europe and Northern America have the highest reported prevalence rates, both of sporadic and familial disease. In contrast, disease prevalence rates are reported lower in Afro-Americans than in any other ethnic group studied. It is likely that a combination of genetic and environmental factors underlies the differences in prevalence amongst ethnic groups; this complexity is best illustrated by considering prevalence rates among Asian migrants in the United Kingdom. Data from Leicester suggest that the Asian population in the United Kingdom have an increased susceptibility to ulcerative colitis, compared with the indigenous population.

Concordance rates in monozygotic and dizygotic twin pairs provide further evidence for the importance of both genetic and environmental facts. Three studies in Europe have reported on 326 twin pairs, with recent follow up data available from the first study. The overall concordance rates for Crohn’s disease were 36%, and 4% (monozygotic and dizygotic pairs respectively), with the corresponding results for ulcerative colitis reported as 16% and 4%. This suggests a greater genetic role in Crohn’s disease compared with ulcerative colitis. The derived coefficient of heritability in Crohn’s disease is equivalent to that reported for other chronic childhood diseases including insulin dependent diabetes and asthma.

The prevalence of familial inflammatory bowel disease has been studied widely. It is apparent that a positive family history of inflammatory bowel disease is the best established risk factor for development of disease. Although precise estimates vary, consistent findings are present in the studies that have been performed. Between 6% and 32% of patients with inflammatory bowel disease have an affected relative. Siblings are at greatest risk with lower relative risks reported for parents, offspring, and second degree relatives. In Crohn’s disease, the relative risk to a sibling, compared with the population prevalence, has been estimated as between 13 and 36. The equivalent figure in ulcerative colitis has been estimated between 7 and 17. In the unusual situation of a child being born to parents who both have inflammatory bowel disease, they have a 33% chance of developing inflammatory bowel disease by 28 years of age. Again these data are consistent with a strong genetic component in disease susceptibility.

A number of other points are notable from the studies of familial disease. It is apparent that the relatives of patients with Crohn’s disease are at an increased risk, not only of Crohn’s disease, but also of ulcerative colitis, supporting common predisposing factors to these phenotypes of inflammatory bowel disease. Moreover, not only susceptibility, but also disease behaviour, appears to have a strong familial basis. Finally, it is apparent that early onset disease has a stronger familial and therefore perhaps genetic contribution. These data are best illustrated by reviewing the Cleveland Clinic study involving a large number of patients with early onset disease in whom 35% had a positive family history.

**GENETIC MODEL**

Although the epidemiological data suggest genetic factors interact with the environment in disease pathogenesis, the model whereby these interactions occur remains under debate. Complex segregation analyses have suggested that a simple recessive model of inheritance may be pertinent to a small proportion of patients with Crohn’s disease, and a simple dominant model pertinent to a proportion of patients with ulcerative colitis. However, the model which has been most widely accepted by the investigators at the present time is that Crohn’s disease and ulcerative colitis are both polygenic diseases, sharing some but not all susceptibility genes. In addition, the variability of clinical presentation of disease—disease phenotype—is likely to represent the effects of allelic variations of these genes, and the interaction between these allelic variations and environmental factors.

**PROGRESS TOWARDS GENE IDENTIFICATION**

Two complementary methods have been employed by investigators searching for susceptibility genes in complex diseases—candidate gene analysis and genome-wide scanning. The analysis of candidate genes relies on an understanding of disease pathophysiology. In inflammatory bowel disease for example, the immunopathology of the disease led to examination of genes involved in the regulation of the immune system, genes involved in the maintenance of mucosal integrity, and genes involved in cell-cell interactions. The frequency of allelic variants of these genes in patients with inflammatory bowel disease is compared with allelic frequencies in a well matched control population. A significant distortion of frequencies in the groups under comparison would provoke further investigation of the gene of interest. Many candidate genes have been subject to analysis in inflammatory bowel disease—notably the genes of the HLA system, genes involved in the regulation of cytokine production, mucin synthesis, and other aspects of epithelial barrier function.

It is, however the complementary technique of genome-wide scanning in which success has become most apparent. The development of a linkage map of the human genome, involving informative microsatellite markers provided a framework for the systematic analysis of the human genome in both single gene disorders, and complex diseases. Studies in complex diseases have required access to large numbers of multiply affected families (typically sibling pairs), semi-automated technology for genotyping, and particularly the evolution of techniques for analysis. The technique has been applied by investigators in many common disorders—encompassing metabolic, respiratory, cardiovascular, endocrine, and neuropsychiatric disease. In each disorder, a number of linkages with regions throughout the genome have been described. However, proceeding from the initial observation of linkage through replication to gene identification has defeated many investigators. Inflammatory bowel disease has reached an enviable position. Four regions of the genome have been replicated, with sufficient strength to satisfy stringent criteria laid down by the statistical geneticist. Moreover, progress has been most evident in pursing
the IBD 1 locus on chromosome 16. Widespread replication has been followed by gene identification.

OTHER IBD LOCI
IBD 2 is located on chromosome 12 and is the region most strongly implicated from the only reported UK genome-wide scan. A combination of a number of international studies has failed to show strong support for inflammatory bowel disease susceptibility linkage at this locus when Crohn’s disease and ulcerative colitis are considered together. However, this region does show strongest linkage within pure ulcerative colitis families, suggesting it is mainly an ulcerative colitis susceptibility locus. No studies of the candidate genes within this region, however, have shown positive linkage with ulcerative colitis.

IBD 3 is located on chromosome 6 surrounding the region of the major histocompatibility complex. This not only represents a susceptibility locus for ulcerative colitis and Crohn’s disease, but also is implicated in determining disease phenotype. This is shown in ulcerative colitis where possession of the HLA-DRB1*0103 allele has association with severe colitis and the presence of extra intestinal manifestations. There has also been considerable interest in IBD 3 because within this region lies the gene encoding tumour necrosis factor α. Associations within the promoter region of this gene have been linked to susceptibility to inflammatory bowel disease.

The IBD 4 locus is located on chromosome 14 and has been implicated in susceptibility to Crohn’s disease in North America and Europe. IBD 5 is a Crohn’s disease susceptibility locus located at 5q31–33. By using a linkage disequilibrium approach Rioux identified a common haplotype in this region spanning 250 kb that contains a cytokine gene cluster. The risk for heterozygotes in possession of the risk alleles was a twofold risk of Crohn’s disease and for homozygotes, sixfold. Earlier age of onset of Crohn’s disease was identified in those carrying the risk alleles.

A number of other sub-chromosomal regions have been implicated by genome-wide scanning, including the X chromosome linkage described in independent European populations. The strength of linkage evidence for each of these other putative loci is relatively weaker than for IBD 1–5, but these may each contain determinants of susceptibility or disease behaviour.

IBD 1: FROM LINKAGE TO GENE
Jean Pierre Hugot, a paediatric gastroenterologist in Paris, initially described the IBD 1 locus on chromosome 16, in a landmark publication in 1996. The investigators described linkage with Crohn’s disease to a region spanning the centromere on chromosome 16. In spite of worries regarding the relatively weak evidence of linkage, this linkage has been reproduced widely, in Europe, North America, and most strongly in the Australian population. An international collaborative group pooled data from 12 centres, and confirmed the strength of the linkage of chromosome 16.

IBD 1 has now been confirmed as a Crohn’s disease locus which has been linked to early onset severe disease. Between 1996 and 2001, investigators were attempting to narrow the region of linkage and identify the gene lying therein. Three parallel publications in May/June 2001 confirm the identity of the IBD 1 gene as nucleotide oligomerisation domain (NOD) 2. NOD 2 was subsequently renamed caspase activating recruitment domain (CARD) 15. Hugot and colleagues applied the classical strategy of positional cloning to narrowing the region of linkage, and were then able to construct a physical map of the region. A study of markers within a physical map involving bacterial artificial chromosomes led to the identification of the NOD 2/CARD 15 gene. In Hugot’s initial publication, allelic variants of the NOD 2/CARD 15 gene were present in 43% of patients with Crohn’s disease. Three polymorphisms were identified to be associated with Crohn’s disease: two missense mutations Arg702Trp and Gly908Arg, and a frameshift mutation Leu1007fsincC. A number of other rarer mutations have been identified following more detailed analysis. The frameshift mutation has been studied in greatest depth, involving the insertion of a cytosine repeat in exon 10 of the gene, which gives rise to a premature stop codon and a truncated form of the NOD 2/CARD 15 protein. It is of interest that Hugot’s data suggest that the NOD 2 gene may act in a recessive mode of inheritance. The relative risk for simple heterozygote is estimated at 3, whereas the risk for simple homozygote is 38 and the risk for compound heterozygote is quoted at 44. Patients possessing NOD 2 mutations are not at increased risk of ulcerative colitis.

Hugot’s data were mirrored by those from two other sets of investigators. The investigators in the United States led by Judy Cho and Gabriel Nunez were able to publish in the same volume of Nature as Hugo and the European team. These authors also concentrated on the insertion of a cytosine repeat in exon 10, and were able to show that the allelic frequencies of this insertion were significantly increased in both Jewish and non-Jewish patients with Crohn’s disease compared with healthy controls. Once again an increased risk in homozygotes was very clear in this publication. Further confirmatory data from Europe, including the first study in the United Kingdom, followed from Schreiber’s group, based in Kiel. Again these investigators concentrated on the frameshift mutation and were able to show the mutations in almost 20% of patients studied.

A flurry of confirmatory studies, involving more detailed analysis of the gene and genotype-phenotype relations have emerged (see table 1).

The data for Caucasian western populations contrast starkly with that of Asian populations where NOD 2/CARD 15 mutations have not been found in Crohn’s disease patients, ulcerative colitis patients, or healthy controls.

The wealth of data promises to lead to a molecular reclassification of Crohn’s disease and inflammatory bowel disease. It is apparent from very painstaking studies carried out in France and the United Kingdom (both London and Oxford) that NOD 2/CARD 15 mutations are associated with susceptibility to Crohn’s disease, and that these mutations may protect against the development of ulcerative colitis.

Moreover, within Crohn’s disease, it is apparent that NOD 2/CARD 15 mutations predisposed towards ileal disease, but not colonic disease. Patients with NOD 2/CARD 15 mutations have also been linked to strictureing disease behaviour, but this may be a secondary phenomenon as strictureing is seen most commonly as a complication of ileal disease location. Several studies have linked NOD 2 mutations with earlier disease onset. The data for specific paediatric populations, however, are only starting to emerge.

The most striking effect of early onset disease has been for homozygotes/compound heterozygotes in which homozygotes represent 34% of patients with an age of onset <10 years, compared with 3% with an onset of 40 years or more.

NOD 2 FUNCTION AND DYSFUNCTION
Function of the wild type and mutant NOD 2/CARD 15 is clearly an important focus for investigation. Initially it appeared the NOD 2/CARD 15 gene was only expressed in monocytes, but it is clear now that it is not only expressed in all cells of the monocyte lineage but also in primary epithelial cells and intestinal epithelial cells. Within the intestinal epithelial cells the greatest concentration of NOD 2 mRNA is found in Paneth cells both in healthy controls and in patients
with Crohn’s disease. Patients with Crohn’s disease however have greater concentrations of NOD 2 mRNA in the Paneth cells. Paneth cells are found in greatest concentration within the ileum and may help explain the association of NOD 2/CARD 15 mutations with terminal ileal disease.

Tumour necrosis factor α, an important pro-inflammatory cytokine in Crohn’s disease has a key role in the regulation of expression of NOD 2/CARD 15 within the intestinal epithelial cells.

The NOD 2/CARD 15 gene has structural homology with plant disease resistant genes, and the toll-like receptor family of genes which are involved in regulation of the innate immune response. The NOD 2/CARD 15 gene contains three regions: 2N-terminal CARDs which are involved in protein-protein interactions, a centrally located nucleotide binding domain which mediates self oligomerisation that is needed for self activation, and at the C-terminal the leucine rich repeat (LRR) domain that is important for binding of bacteria. Wild type NOD 2/CARD 15 is involved in the activation of nuclear factor-κB (NF-κB) which is triggered by bacteria binding to the LRR. Specifically, it is muramyl dipeptide, the minimal essential structure of bacterial peptidoglycan that is recognised by the NOD 2/CARD 15 pathway within the cell. All three common mutations of the NOD 2 gene result in failure of NF-κB activation after muramyl dipeptide binding giving a uniform loss of function as their mechanism of action. Figure 1 shows the interaction of NOD 2/CARD 15 within the cell.

### NOD 2/CARD 15 GENE-GENE INTERACTION

The interaction of NOD 2/CARD 15 mutations with other genes is only starting to be explored. The gene at the IBD 5 locus has not been identified but two risk alleles on the IBD 5 haplotype have. Mirza et al were able to replicate the findings of Rioux’s Canadian study in a British/German population, but the increase Crohn’s disease risk in this population for those in possession of the risk alleles was much smaller. Mirza et al also showed an earlier age of onset in those possessing the risk alleles, an effect that was increased if NOD 2/CARD 15 mutations were also present. The study showed homozygotes for the risk alleles plus one CARD mutation had a 58% chance of Crohn’s disease by age 21 compared with 27% who carried neither of the genotypes. This figure had risen to 93% by age 36. This suggests cooperation between IBD 5 and NOD 2/CARD 15 in disease causation, giving evidence of gene-gene interaction, especially in an early onset population. Another British study has, however, been unable to replicate these findings.

### NOD 2/CARD 15 MUTATIONS: ROLE IN OTHER DISEASES

The identification of NOD 2/CARD 15 gene has lead to a search for its role in other inflammatory conditions. Blau syndrome is a rare disorder sharing some features in common with Crohn’s disease characterised by skin rashes, uveitis, arthritis, and granuloma formation. Mutations with the NBD domain of the NOD 2/CARD 15 gene have been associated with Blau syndrome, contrasting with Crohn’s disease where the mutations are located in the LRR region. Negative studies have been published for multiple sclerosis, ankylosing spondylitis, systemic lupus erythematosus, rheumatoid arthritis, Wegener’s granulomatosis, and psoriasis. Interestingly a recent study of German schoolchildren suggested NOD 2/CARD 15 mutations were more common

---

**Table 1** Allele frequencies of common NOD 2 mutations in different populations

<table>
<thead>
<tr>
<th>Population</th>
<th>Arg702Trp</th>
<th>Gly908Arg</th>
<th>Leu1007 finsC</th>
<th>Phenotype correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>ND</td>
</tr>
<tr>
<td>USA</td>
<td>ND</td>
<td>ND</td>
<td>8.2</td>
<td>ND</td>
</tr>
<tr>
<td>UK/German</td>
<td>ND</td>
<td>ND</td>
<td>16</td>
<td>ND</td>
</tr>
<tr>
<td>Canadian</td>
<td>9.1</td>
<td>3.4</td>
<td>6.6</td>
<td>ileal/familial</td>
</tr>
<tr>
<td>British</td>
<td>12.9</td>
<td>5.2</td>
<td>10.3</td>
<td>ileal</td>
</tr>
<tr>
<td>European</td>
<td>11</td>
<td>6</td>
<td>11</td>
<td>ileal/earlier onset</td>
</tr>
<tr>
<td>Scottish</td>
<td>7.1</td>
<td>1.8</td>
<td>4.7</td>
<td>No association found</td>
</tr>
<tr>
<td>Finnish</td>
<td>3.3</td>
<td>0.6</td>
<td>4.8</td>
<td>ileal/stricturing</td>
</tr>
<tr>
<td>Dutch</td>
<td>ND</td>
<td>ND</td>
<td>14</td>
<td>ND</td>
</tr>
<tr>
<td>Greek</td>
<td>ND</td>
<td>ND</td>
<td>2.7</td>
<td>ND</td>
</tr>
<tr>
<td>Irish</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>ileal/earlier onset</td>
</tr>
</tbody>
</table>

Figures in bold signify a significant difference in Crohn’s disease patients compared to healthy controls. ND, not done.

---

**Figure 1** The interaction of NOD 2/CARD 15 within the cell. (Reprinted from Bonen DK et al. The genetics of inflammatory bowel disease. Gastroenterology 2003;124(2):530, with permission from the American Gastroenterological Association)
in atopic dermatitis and allergic rhinitis but not asthma. This finding clearly needs to be explored further in other atopic populations.

THE FUTURE

What are the prospects of translating this scientific progress to clinical application? Although many unanswered questions exist, there is really a feeling of optimism among clinicians and scientists following this recent progress. Clearly further studies of gene function—genotype-phenotype relations, and gene-gene and gene-environmental interactions need to be carried out. However, the increased understanding of disease pathophysiology is likely to impact on clinical practice.

Patient counselling may well be improved, when more genetic data are available. The choice of drug therapies for an individual patient may be rationalised, on the basis of genotype. Already, some examples of this exist in inflammatory bowel disease—the use of thiopurine methyltransferase genotyping or phenotyping in patients receiving azathioprine.


Bonon DK, Cho JH. The genetics of inflammatory bowel disease. Gastroenterology 2003;124:521-34.


The genetics of inflammatory bowel disease

603


