The genetics of inflammatory bowel disease

Bijay Baburajan, Miles Parkes

Over the past 10 years major progress has been made in understanding the genetic contribution to inflammatory bowel disease. NOD2 was recently identified as a major susceptibility gene for Crohn's disease. This and a number of other strong genetic leads are discussed.

Inflammatory bowel disease (IBD) describes a state of chronic relapsing intestinal inflammation of unknown aetiology. The two major forms – Crohn's disease (CD) and ulcerative colitis (UC) – have a combined prevalence of 150–250/100 000 in northern Europe (Calkins and Mendeloff, 1995).

The pathogenic mechanisms of IBD remain elusive. Like most complex diseases, the inflammation of IBD appears to result from an environmental trigger in a genetically susceptible host. The key to understanding the pathogenesis lies in identifying the environmental agent and characterizing the genes involved.

Early investigators used candidate gene studies, looking at polymorphisms (variants) within genes of known function to assess their potential contribution to IBD. While some progress was made, a quantum leap was required to advance our understanding of the genetic basis of IBD.

Since 1996, a number of 'hypothesis-free' genome scans have been performed in IBD to identify regions likely to contain disease susceptibility genes. These are not based on prior assumptions about disease pathogenesis, but rely on a systematic screen of polymorphic markers distributed across the human genome in large panels of multiply affected families. Regions of linkage are those where affected sibling pairs show significantly greater allele sharing than expected by chance alone. Where such regions are replicable in independent family panels they merit further investigation: fine mapping by association study or positional candidate gene analysis (*Figure 1*).

The more meiotic steps separate two affected people, the smaller the chromosomal segment they will share. In affected sibling pair methods of linkage analysis the study subjects are closely related (sibs) and distortion of allele sharing is seen over a wide chromosomal segment. This makes linkages relatively easy to detect at a genome-wide level but provides relatively poor resolution for gene localization. In association studies marker allele frequencies are compared in healthy controls and populations of sporadically affected individuals. These latter are at most distantly related and thus separated by many meioses: any distortion of allele frequencies is difficult to detect at the genome level but provides high resolution for gene localization.

In interpreting the results of any genetic study caution is required (Cardon and Bell, 2001). Results should attain stringent levels of statistical significance at a genome-wide level and be replicated in independent panels. Accepted criteria are available (Lander and Kruglyak, 1995; Risch and Merikangas, 1996). Results that fail to attain significance should be interpreted with caution as many will be false positives.

Given the volume of marginally significant results in the IBD genetics literature, this review will not provide an exhaustive list of all studies carried out to date, but will highlight those that illustrate general points and the findings of greatest overall significance.

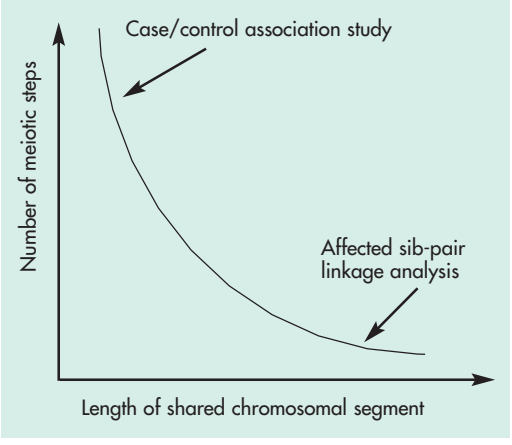


Figure 1. The comparative properties of association studies and linkage analysis for gene localization.

Dr Bijay Baburajan is Specialist Registrar and **Dr Miles Parkes** is Consultant Gastroenterologist in the Department of Gastroenterology, Addenbrooke's Hospital, Cambridge CB2 2QQ and Associate Lecturer, University of Cambridge

Correspondence to: Dr M Parkes

EPIDEMIOLOGICAL EVIDENCE FOR A GENETIC CONTRIBUTION TO IBD

As for the genetic analysis of any disease, the foundation stone for such studies in IBD is the epidemiological evidence. The most convincing data in IBD come from twin and family studies.

Twin studies

Studies of twins point to substantially higher rates of disease concordance in monozygotic compared with dizygotic twins, particularly for CD (35% *vs* 7% respectively) compared to UC (11% *vs* 3% respectively).

Family studies

Estimates of disease risk in multiply affected families indicate that a positive family history outweighs all other known risk factors for the development of IBD. This is formalized as the λs ratio (the risk that the sibling of a patient develops disease compared with the risk that a member of the general population develops disease). For CD this ratio is 20–35 and for UC it is 8–15.

A GENETIC MODEL FOR IBD

A genetic model for IBD must account for a number of factors (*Table 1*). The model that perhaps best fits is one in which CD and UC are multifactorial diseases that share some susceptibility genes but differ at others. Within each disease category, there are likely to be a number of conditions that are non-overlapping or only partially overlapping in their genetic predisposition.

The genetic variants which lead to IBD might be rare mutations, as with monogenic conditions and with NOD2, or common variants, perhaps at a frequency only modestly greater than that in the healthy population, and each contributing only a small fraction to the overall disease phenotype.

TABLE 1.
Factors which must be accounted for in a genetic model of inflammatory bowel disease

Factor	Evidence
More than one gene is involved in CD and UC	Segregation analysis
Shared susceptibility genes between CD and UC	Presence of indeterminate colitis in 10% of cases of inflammatory bowel disease
	Up to 30% of multiply affected families contain cases of both CD and UC
	Shared extra-intestinal manifestations between CD and UC
Heterogeneity within CD and UC	Clinical sub-phenotypes such as site of inflamnmation in CD, breed true in multiply affected families
	Recent NOD2 data (see text)

CD = Crohn's disease; UC = ulcerative colitis

TABLE 2.
The meaning of a 'positive' result in a genetic association study

Implication	Reason
A disease-causing variant	Will clearly be present more frequently in affected individuals than in healthy controls
An allele in linkage disequilibrium with 1	This allele was present on the founder chromosome at a nearby polymorphic locus at the time the disease-causing mutation first occurred, and thus these two variants have been inherited as a single block or 'haplotype'
False positive	Usually a result of poor study design, e.g. inappropriately lax statistical thresholds, failure to replicate or poorly matched controls

MOLECULAR GENETIC STUDIES

Early molecular studies focused on candidate genes implicated by their known function, investigated by association study (*Table 2*). The most successful application of this approach in IBD has been studies of the human leucocyte antigen (HLA) complex on the short (p) arm of chromosome 6. More recently attention has focused on genome scanning studies, but this is beginning to come full circle with acceptance that a powerful strategy for disease gene identification is to combine both approaches – studying positional candidate genes. Such genes are implicated because they map to a region of linkage identified in a genome scan, and also by their expression pattern or function.

Pure candidate gene studies

The HLA region was first implicated in IBD in replicated Japanese studies in which HLA-DR2 was associated with UC (Asakura et al, 1982). However, this allele does not appear to be associated with IBD in Caucasoid populations.

In a Caucasoid panel, Satsangi et al (1996b) identified an association between HLA-DR3, DQ2 and extensive UC particularly in females, and DRB1*0103 and both severe disease requiring surgery and extra-intestinal manifestations. Both these associations have been replicated in at least two further panels.

There is strong evidence that two phenotypically and genotypically distinct forms of arthropathy are associated with IBD. A large joint inflammatory oligo-arthropathy is strongly associated with the DRB1*0103 allele, and a small joint arthropathy is associated with HLA-B44 (Orchard et al, 2000). DRB1*0103 is also strongly associated with erythema nodosum and iritis in both CD and UC, highlighting that some genes might determine phenotype rather than disease susceptibility per se.

Although original association data for CD overall were inconsistent at the HLA, results of linkage studies have re-focused attention. These suggest a role for this region in CD as well as UC, and consistency is appearing in the associa-

tion data both for particular HLA alleles (Ahmad et al, 2002) and tumour necrosis factor promoter polymorphisms. The latter have shown replicable association in Japanese and Caucasian populations. van Heel et al (2002) showed association between IBD and a functionally significant tumour necrosis factor promoter polymorphism which appears to enhance binding of NFκB.

Despite this, it is still unclear which gene or genes on chromosome 6p contribute to IBD, because numerous immunoactive genes map here and there is extensive linkage disequilibrium (LD) in this region – an issue that most studies of genes in this region have failed to address. Strong LD means that few cross-overs tend to occur in this area at meiosis, thus association is seen with 'hitchhiking' polymorphisms on either side of a disease-causing mutation. Differentiating the disease-causing variant from hitchhikers is difficult – in haemochromatosis it took 20 years from identifying HLA A3 association to pinpointing HFE.

Many immunoactive candidate gene polymorphisms have been studied in IBD, and some have shown modest if inconsistent evidence for association. Perhaps the strongest among these is the interleukin (IL)-1 receptor antagonist, but until stronger evidence is forthcoming, judgment as to their likely contribution is best reserved.

Genome scans in IBD provide replicated linkage results

To date, seven full genome scans have been published for IBD which have identified strong candidate regions on chromosomes 1, 3, 5, 6, 12, 14 and 16 (Hugot et al, 1996; Satsangi et al, 1996a; Cho et al, 1998; Hampe et al, 1999; Ma et al, 1999; Duerr et al, 2000; Rioux et al, 2000). These have used panels of multiply affected families and hypothesis-free, non-parametric methods of linkage analysis to detect genomic regions showing significant linkage. Compared with other complex diseases, genome scans in IBD have been highly successful in providing replicable and significant evidence for linkage.

Chromosome 1p: Original evidence for linkage between 1p and CD came from America (Cho et al, 1998). In a second study using an independent panel from a Chaldean population isolate, Cho et al replicated the evidence for linkage and also found association between IBD and an ancestral haplotype at this locus. Further data are awaited.

Chromosome 3p: This was first linked to IBD in the Oxford, UK genome scan (Satsangi et al, 1996a), and then replicated in Germany, Finland, Canada and USA. Linkage appears strongest for CD. This gene-dense region of the genome contains a large number of interesting positional candidate genes. Reports of association at adjacent microsatellite markers await replication.

Chromosome 5q: Initial linkage to CD was detected in the Canadian affected sibpair panel around a cytokine gene cluster on the long (q) arm of chromosome 5 (Rioux et al, 2000). This has been substantiated by a strong association across a broad genomic segment at this locus in this and other panels (Rioux et al, 2001). There is strong LD in this region, which is hindering attempts to pinpoint the disease-causing variant. Candidate gene studies have been negative but are ongoing.

Chromosome 12: This locus was the most strongly linked region in the Oxford IBD genome scan (Satsangi et al, 1996a), with subsequent data suggesting that it contributes more to UC than CD susceptibility (Parkes et al, 2000). The linkage has been replicated in four further datasets worldwide, and held up in the mega-analysis of pooled genotyping data carried out by the international IBD genetics consortium.

Success in fine mapping the chromosome 12 locus has been limited. The region of linkage is large and different groups detect peaks in different positions. This probably reflects the poor resolution of allele sharing methods but might also indicate more than one susceptibility gene here.

Association data have been inconsistent. Both Oxford and Pittsburgh groups identified association at the same microsatellite marker – but with different alleles. Analysis of single nucleotide polymorphisms in this region is underway.

Two positional candidate gene studies, interferon γ and integrin β7, have been negative. Other positional candidate gene studies are in progress.

Chromosome 14q: Replicated linkage to CD has been seen in three independent datasets. Fine mapping data are awaited.

Chromosome 16 – IBD1: The IBD1 locus on chromosome 16 had been extensively replicated using linkage analysis in a number of panels. The optimism that this generated was recently realized with the identification of mutations in the NOD2 gene as playing an important role in the pathogenesis of CD – but not UC.

NOD2 was reported simultaneously by Hugot et al (2001), who used classic LD mapping techniques and identified one frameshift and two missense mutations, and Ogura et al (2001), who used positional candidate gene approach. Average relative risks for CD for genotypes containing zero, one or two of the variants were 1, 3 and 38 respectively in the French study (44 for compound heterozygotes). This has been replicated in many panels, and data suggest that NOD2 mutations are only associated with ileal inflammation and not CD where the colon alone is inflamed (Ahmad et

al, 2002). Double dose mutations (homozygous or compound heterozygous) increase the risk of early onset, fistulating ileal disease.

It is unclear how mutations in NOD2 predispose to CD. The protein product of NOD2 is expressed in monocytes and apparently not the intestinal mucosa. Known or predicted functions include sensing bacterial lipopolysaccharide, activation of NF-κB and regulating apoptosis. The mutations identified appear to predominantly affect the leucine-rich repeat domain (*Figure 2*), which might play a role in sensing bacterial lipopolysaccharide and regulating NF-κB. Whether these or other functions are important pathogenic mechanisms remains to be seen.

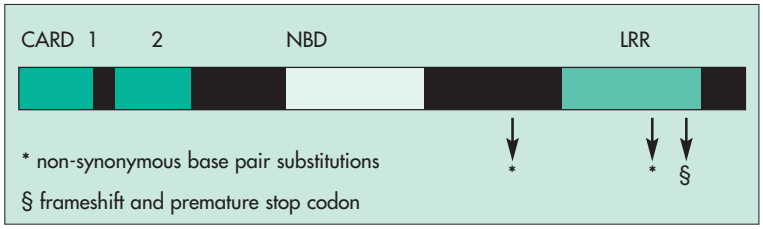
CONCLUSIONS AND CLINICAL IMPLICATIONS

Substantial progress has been made in the past 10 years in understanding the genetic basis of IBD. Much remains to be done, but each advance in complex disease genetics in general and IBD genetics in particular facilitates the next step. The long-term hope is that genetic studies will provide a quantum leap forward in understanding the pathogenesis of IBD, allowing the development of rational new therapies.

Reclassification of CD and UC into genotypically defined sub-groups based on shared allelic variants should clarify much of the unpredictability associated with the management of IBD. Drug trials will be stratified by genotype, and in due course lifestyle advice and therapies will be targeted in a much more specific manner.

A complete understanding of the genetic basis of IBD should provide important insights into the nature of environmental triggers and drive. Identification of NOD2 has provided a clear link between the immune response to enteric bacteria and development of CD. In future, for individuals with an 'at risk' complex of susceptibility alleles at various loci, environmental modification might be appropriate. On present understanding, this would include avoidance of non-steroidal anti-inflammatory drugs and of smoking in those at risk of CD. In future, it might be possible to abrogate the development and progression of IBD entirely. HM

Figure 2. Three functional domains of NOD2: the caspase recruitment domains (CARD) 1 and 2, the nuclear binding domain (NBD) and the leucine rich region (LRR) which may be involved in binding bacterial lipopolysaccharide. The mutations that predispose to Crohn's disease are localized to the LRR.



KEY POINTS

- Genetic susceptibility is important particularly for Crohn's disease but also for ulcerative colitis.
- Firm replication is required before attaching too much credence to a reported genetic association or linkage.
- One or more genes in the human leukocyte antigen region contributes to inflammatory bowel disease susceptibility, and may influence disease phenotype.
- Several replicated regions of linkage in the genome have been identified in inflammatory bowel disease.
- The NOD2 gene is one of the first complex disease genes to have been positionally cloned, and appears specific for ileal Crohn's disease.

Conflict of interest: none.

Ahmad T, Armuzzi A, Bunce M et al (2002) The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* **122**(4): 854–66

Asakura H, Tsuchiya M, Aiso S et al (1982) Association of the human leukocyte DR2 antigen with Japanese ulcerative colitis. *Gastroenterology* **82:** 413–8

Calkins BM, Mendeloff AI (1995) The epidemology of ideopathic inflammatory bowel disease. In: Kirsner JB, Shorter RG, eds. *Inflammatory Bowel Disease.* Williams and Wilkins, Baltimore: 31–68

Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* **2**(2): 91–9

Cho JH, Nicolae DL, Gold L et al (1998) Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q and 4q: evidence for epistasis between 1p and IBD1. *Proc Nat Acad Sci USA* **95:** 7502–7

Duerr RH, Barmada MM, Zhang L et al (2000) High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* **66**(6): 1857–62

Hampe J, Schreiber S, Shaw S et al (1999) A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* **64**(3): 808–16

Hugot JP, Laurent-Puig P, Gower-Rousseau C et al (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* **379:** 821–3

Hugot JP, Chamaillard M, Zouali H et al (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**(6837): 599–603

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11:** 241–7

Ma Y, Ohmen JD, Li Z et al (1999) A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* **5**(4): 271–8

Ogura Y, Bonen DK, Inohara N et al (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**(6837): 603–6

Orchard TR, Thiyagaraja S, Welsh KI et al (2000) Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology* **118**(2): 274–8

Parkes M, Barmada MM, Satsangi J et al (2000) The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *Am J Hum Genet* **67**(6): 1605–10

Rioux JD, Silverberg MS, Daly MJ et al (2000) Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* **66**(6): 1863–70

Rioux JD, Daly MJ, Silverberg MS et al (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* **29**(2): 223–8

Risch N, Merikangas K (1996) The future of genetic studies of human complex diseases. *Science* **273:** 1516–7

Satsangi J, Parkes M, Louis E et al (1996a) Two-stage genome-wide search in inflammatory bowel disease: evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* **14**(2): 199–202

Satsangi J, Welsh KI, Bunce M et al (1996b) Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet* **347:** 1212–17

van Heel DA, Udalova IA, de Silva AP et al (2002) Inflammatory bowel disease is associated with a TNF polymorphism that affects an interaction between the OCT1 and NF(-kappa)B transcription factors. *Hum Mol Genet* **11**(11): 1281–9