# Ulcerative colitis and Crohn's disease: molecular genetics and clinical implications

**Miles Parkes and Derek Jewell**

**The genetics of inflammatory bowel disease (IBD) is an area of enormous topical interest. Crohn's disease is one of the first polygenic diseases in which a susceptibility gene (*NOD2*) has been positionally cloned, providing proof of principle for methodologies that to date had been mired in uncertainty. IBD has been highly unusual among polygenic diseases in providing replicated candidate susceptibility regions using genome-scanning methodologies and linkage analysis. Such studies have identified loci with replicated evidence for linkage on chromosomes 1, 3, 6, 12, 14 and 16 – the latter recently yielding *NOD2*. The next stage is fine mapping each of the remaining candidate loci and identifying the precise genes involved. Although progress is slow, the identification of *NOD2* together with technological advances and the availability of increasingly large panels of patients and multiply affected families provide the basis for great optimism. The ultimate goal is a better understanding of disease pathogenesis – both genetic and environmental – with a rational redefinition of Crohn's disease and ulcerative colitis, and a rational, specific therapeutic approach to each genotypically defined subclass of IBD.**

Inflammatory bowel disease (IBD) describes a state of chronic relapsing intestinal inflammation of unknown aetiology. It causes significant morbidity in populations of European origin, with the two major forms – Crohn's disease and ulcerative colitis – having a combined prevalence of 150–250/100 000 (Ref. 1). They are both more common in Ashkenazi Jews, and less common in Afro-Caribbeans. Both forms of IBD have a peak incidence in early adult life (but can develop at any age), and they affect the sexes to an approximately equal extent.

Environmental risk factors for the development of IBD are poorly defined. The best characterised is smoking, which clearly increases the risk of developing Crohn's disease but protects

Miles Parkes (corresponding author)
Consultant Gastroenterologist, Gastroenterology Unit, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK. Tel: +44 (0) 1223 245151; Fax: +44 (0) 1223 216548; E-mail: miles.parkes@addenbrookes.nhs.uk

Derek Jewell
Professor of Gastroenterology, Gastroenterology Unit, Radcliffe Infirmary, Oxford, OX2 6HE, UK. Tel: +44 (0) 1865 224829; Fax: +44 (0) 1865 790792; E-mail: derek.jewell@clinical-medicine.oxford.ac.uk



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

against ulcerative colitis, for reasons that are not clear. Increased intestinal permeability to gut luminal antigen might play a role in triggering and perpetuating intestinal inflammation, and the increased risk associated with use of non-steroidal anti-inflammatory drugs is thought to be secondary to this. Several specific microbes have also been suggested to play a role in IBD pathogenesis, most recently the measles virus and *Mycobacteria paratuberculosis* with respect to Crohn's disease, but the evidence is at best equivocal.

Ulcerative colitis typically causes symptoms of bloody diarrhoea and faecal urgency. Inflammation involves the rectum, and can extend in a continuous manner to part or all of the proximal colon. Histological examination reveals an infiltrate of chronic inflammatory cells, which are restricted to the superficial layers of the colonic mucosa. Granulomata are not a feature.

Crohn's disease is rather more pleomorphic. It is characterised pathologically by discontinuous segments of transmural inflammation, which can affect any part of the gastrointestinal tract from mouth to anus, but most commonly involve the ileo–caecal region. Fistulae (abnormal tracts between epithelial surfaces – usually between bowel and adjacent bowel, bladder, vagina or skin) and strictures (narrowing of the bowel lumen caused by inflammation and oedema, or by fibrotic scarring as the inflammation heals) are often seen, and granulomata are a histological hallmark. Clinical features are rather variable depending on the site of bowel involvement. If the colon is inflamed the symptoms can exactly mimic those of ulcerative colitis. With the rather more typical pattern of ileal inflammation, obstructive symptoms of abdominal pain and distension after meals can be more prominent than the symptoms of diarrhoea.

With both Crohn's disease and ulcerative colitis active inflammation tends to be associated with non-specific symptoms of malaise and fatigue. Some 30% of patients also have a number of typical extra-intestinal manifestations, which include arthritis, mouth ulceration and ocular inflammation.

Treatment of the active phase of IBD broadly relies on steroid therapy and immunosuppression. 5-Aminosalicylic acid preparations have a role, particularly in the maintenance therapy of ulcerative colitis, and a relatively recent innovation has been the use of anti-tumour necrosis factor (TNF) antibody for Crohn's disease refractory to conventional therapies. It seems likely that it will also be efficacious in ulcerative colitis, but data from controlled trials are awaited. Several other 'biological' therapies are currently being tested. A better understanding of the pathogenesis of IBD, and a rational subdivision of Crohn's disease and ulcerative colitis into genetically distinct subgroups, might allow a more specific targeting of these new therapies, as well as the development of novel therapeutic strategies.

## An introduction to IBD genetics

The explosion of interest in molecular genetics that followed publication of the seminal genome scans in type 1 diabetes in 1994 has affected the study of IBD no less than any other common, complex disease (Refs 2, 3). Prior to that, horizons had largely been limited to studying polymorphisms within genes of known function, to assess their potential contribution to disease susceptibility. While some progress was made, a quantum leap was required to advance our understanding of the molecular genetics and hence pathogenesis of IBD.

Since 1994, much effort has been expended pursuing hypothesis-free methods for identifying susceptibility genes for IBD. These are based on no prior assumptions about disease pathogenesis. Instead they rely on a systematic screen of polymorphic markers distributed across the human genome in large panels of multiply affected families, looking for regions that show linkage to disease susceptibility (Ref. 4). The linkage analysis typically used is the affected sibling or relative pair method, which, being non-parametric, does not require specification of a model of inheritance. Regions of linkage are those where the affected sibling or relative pairs show significantly greater allele sharing than expected by chance alone. Guidelines for the interpretation of linkage results derived from genome scans in polygenic diseases are shown in Table 1. These have become widely accepted, albeit not without criticism for being too stringent. Perhaps the best test of a candidate region's validity pending gene identification is whether it can also be identified in an independent panel of families. Where such candidate regions are replicable they are likely to contain disease susceptibility genes, and merit further investigation using the various techniques



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

**Table 1. Guidelines for the interpretation of linkage results derived from genome scans in complex traits as proposed by Lander and Kruglyak (tab001mpo)**

| Degree of significance[a] | Genome-wide *P* value[b] | Lod score[c] |
|---|---|---|
| 'Suggestive' | $P < 7.4 \times 10^{-4}$ | 2.2 |
| 'Significant' | $P < 2.2 \times 10^{-5}$ | 3.6 |
| 'Highly significant' | $P < 3 \times 10^{-7}$ | 5.4 |
| 'Confirmed' | Replication of an initial significant result with $P \leq 0.01$[d] | – |

[a] Four levels of significance have been proposed by Lander and Kruglyak to interpret data derived from genome scans; these are based on the *P* value and the derived Lod score (Ref. 24).
[b] The *P* value is derived from one of several possible statistics, all of which measure the extent to which allele sharing ratios at a given locus in a panel of affected sib pairs (or affected relative pairs) deviate from that expected by chance. At a locus adjacent to a susceptibility gene, a panel of affected sib pairs are likely to show greater than 25% sharing of two alleles, and correspondingly reduced sharing of zero alleles below the 25% that would be expected at an unlinked (i.e. random) locus.
[c] The Lod score is the log of the odds ratio as defined by: probability that the locus is linked to disease given the observed data divided by the probability that the locus is not linked to disease given the observed data. It thus provides a measure of the strength of linkage data.
[d] The level of significance required in the replication dataset is considerably lower as a specific hypothesis is now being tested.

of fine mapping and positional candidate gene analysis (Ref. 5).

What progress has been made, and are current methodologies likely to be successful? Will the imminent availability of the full sequence of 3 billion base pairs of the human genome (Refs 6, 7), the publication of maps containing over 100 000 single nucleotide polymorphisms (SNPs) (Ref. 8), and the ever-increasing capabilities of high-throughput genotyping and bioinformatics permit a comprehensive understanding of the genetic basis of common diseases? And to what extent will an understanding of the genetic basis of Crohn's disease and ulcerative colitis impact on our ability to manage patients in the clinic of the future? This article attempts to answer these questions. It does not provide an exhaustive list of all the candidate-gene studies carried out in IBD to date, but highlights those that illustrate general points, and focuses primarily on hypothesis-free methodologies.

## 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 that genes play an important role in determining susceptibility to Crohn's disease and ulcerative colitis. This has come from a variety of sources, but the most convincing data come from twin studies (Refs 9, 10, 11) and estimates of disease risk within multiply affected families (Refs 12, 13, 14). Studies of twins point to substantially higher rates of disease concordance in monozygotic compared with dizygotic twins, particularly for Crohn's disease (35% versus 7%, respectively) relative to ulcerative colitis (11% versus 3%, respectively) (Table 2). Estimates of disease risk in multiply affected families indicate that the presence of a positive family history outweighs all other known risk factors for the development of IBD. This is formalised as the $\lambda s$ 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 Crohn's disease this ratio is 20–35, and for ulcerative colitis it is 8–15. Although the high $\lambda s$ ratio undoubtedly reflects some contribution from the shared family environment, the consistent message from epidemiological studies has been that the genetic contribution to the pathogenesis of Crohn's disease is greater than that for ulcerative colitis and, indeed, greater than that for most other common diseases, such as type 1 diabetes ($\lambda s = 10–15$), hypertension ($\lambda s = 2–3$) or schizophrenia ($\lambda s = 5–10$).

Early modelling studies suggested that the $\lambda s$ ratio should be proportional to the ease of



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

**Table 2. Concordance for Crohn's disease and ulcerative colitis in monozygotic and dizygotic twins in which at least one of the pair has inflammatory bowel disease (tab002mpo)**

| | Disease | Concordance in monozygotic pairs[a] | Concordance in dizygotic pairs[a] |
|---|---|---|---|
| Sweden | Crohn's disease | 8/18 | 1/26 |
| | Ulcerative colitis | 1/16 | 0/20 |
| Denmark | Crohn's disease | 3/11 | 0/25 |
| | Ulcerative colitis | 1/15 | 1/42 |
| UK | Crohn's disease | 16/48 | 8/81 |
| | Ulcerative colitis | 7/52 | 3/68 |
| Combined data | Crohn's disease | 27/77 (35%) | 9/132 (7%) |
| | Ulcerative colitis | 9/83 (11%) | 4/130 (3%) |

[a] Data are provided for the three most complete studies reported to date, from Sweden, Denmark and the UK (Refs 9, 10, 11).

mapping of susceptibility loci (i.e. a high λs ratio should indicate loci that can be easily mapped) (Ref. 15). This resulted in great optimism that Crohn's disease susceptibility genes might – of all the common diseases – prove relatively easy to localise. This optimism was supported by: the knowledge that the phenotype can be objectively assessed, by clinical, radiological and pathological means; the recognition that disease subtypes breed true (Refs 16, 17, 18, 19); and the hope that subclinical markers such as anti-*Saccharomyces cerevisiae* antibody (ASCA) and anti-neutrophil cytoplasmic antibody (ANCA) might allow further subdivision within ulcerative colitis and Crohn's disease. These all suggested that it should be possible to minimise at least some of the heterogeneity within the study population, and hence further improve the power of studies to isolate susceptibility genes for these two diseases. The identification of *NOD2* has proved this optimism to be well founded.

The other area of interest relating to heterogeneity is the nature of the pathogenic relationship between Crohn's disease and ulcerative colitis, and the extent to which they share genetic susceptibility loci. Although the epidemiological and clinical distinctions are clearly recognised, up to 10% of cases of IBD are classified as indeterminate and up to 30% of multiply affected families contain cases of both Crohn's disease and ulcerative colitis (Refs 14, 20). Indeed, it is possible to calculate a λs ratio for ulcerative colitis given a sibling with Crohn's disease (λs = 4), although the risk of developing Crohn's disease for a first-degree relative of an ulcerative colitis patient is little higher than that of the general population (Ref. 21). There have even been two or three pairs of monozygotic twins identified in which one has Crohn's disease and the other has ulcerative colitis. Clearly, molecular genetic studies have the potential to illuminate the relationship between Crohn's disease and ulcerative colitis.

## A genetic model for IBD

The genetic model that perhaps best fits the epidemiological and clinical data is one in which Crohn's disease and ulcerative colitis are multifactorial diseases that share some susceptibility genes but differ at others. As for most common diseases, it might be that, within each disease category, there are several conditions that are non-overlapping or only partially overlapping in their genetic predisposition. This heterogeneity is just one of the factors that will confound attempts to isolate complex-disease genes (Box 1).



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

## Box 1. Outstanding research questions

To what extent do Crohn's disease and ulcerative colitis share susceptibility genes?

How many genotypically distinct diseases are encompassed by the terms Crohn's disease and ulcerative colitis? Are these reflected in the current phenotypic subdivisions of, for example, distal versus extensive ulcerative colitis, or stricturing versus fistulising versus inflammatory Crohn's disease?

How many genes are involved in inflammatory bowel disease (IBD)? How many of the disease-predisposing genetic variants are true mutations and how many are polymorphisms present at high frequency (albeit perhaps not together, or without the environmental trigger) in the general population? If genetic variants are polymorphisms, what size of sample will be required to detect them?

Is linkage disequilibrium sufficiently strong and detectable in outbred populations to allow widespread localisation of polygenes using positional cloning strategies, or will resolution of candidate susceptibility regions require large-scale analysis of positional candidate genes?

Will identification of IBD susceptibility genes allow elucidation of the environmental contribution to disease pathogenesis? Is this environmental agent a specific pathogen, or is it a ubiquitous agent that acts only in the genetically predisposed?

Will the general preconceptions regarding the likely nature of candidate genes – for example, immunoregulatory or involved in intestinal barrier function – be confirmed, or will novel mechanisms be revealed?

Animal models have been relatively under-used in IBD as compared with other complex diseases, and could prove very useful. However, there are major concerns as to whether currently available models have any relationship to human IBD. Given the likely epistatic interactions and environmental influences, will it ever be possible to generate an animal model of IBD that truly reflects the complexity of Crohn's disease and ulcerative colitis? And could large-scale analysis of currently available models provide novel insights into the genetic predisposition to IBD?

Other confounding factors include phenocopy (i.e diseases that share the same phenotype but are unrelated in terms of pathogenesis), variable penetrance (the probability that a person with a given genotype will manifest the trait) and variable expression (differing manifestations of the same phenotype for a given genotype). These will depend on both the genetic and the environmental context of the particular susceptibility allele. Furthermore, unlike monogenic diseases, which tend to be caused by rare mutations, the susceptibility alleles that underlie common diseases such as IBD might be common variants. This is called the 'common disease – common variant' hypothesis (Refs 22, 23). Each common variant is likely to contribute a small fraction to the overall disease phenotype and be difficult to recognise by classic association studies, as its frequency in a heterogeneous 'disease' population might not be substantially greater than that in a matched control group. Although this situation might pertain, it should be noted that it was not found to be the case for *NOD2*, where the susceptibility alleles are all rare variants, as discussed below. A further potential confounding factor is ethnic variation in allele frequency, which further underlines the need for caution in association studies in complex diseases (Ref. 22).

In summary, the terms Crohn's disease and ulcerative colitis describe related diseases (or groups of diseases) caused by the interaction of a number of genes, some of which are common to both, together with an undefined environmental



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

stimulus. The latter probably takes the form of a specific trigger, and subsequently a permissive pro-inflammatory effect of the gut luminal microflora.

## Genome scans in IBD provide replicated linkage results

To date, at least seven full genome scans have been published for IBD, focusing on Crohn's disease alone or combining Crohn's disease and ulcerative colitis. These studies have identified strong candidate regions on chromosomes 1, 3, 6, 12, 14 and 16. All studies reported to date have used panels of multiply affected families and hypothesis-free, non-parametric methods of linkage analysis to detect genomic regions showing significant linkage, as discussed above. Compared with genome scans in other complex diseases, such studies in IBD have proven highly successful in providing replicable and significant evidence for linkage at a number of loci.

Given the propensity for this technique to produce false positive results, only linkages that have been replicated in independent datasets, and have achieved the criteria for 'suggestive' linkage put forward by Lander and Kruglyak (Table 1) (Ref. 24), or which are otherwise of interest, are discussed further here. Unsurprisingly, given the number of studies that have now been conducted, nominal levels of significance (e.g. Lod score >1) have been observed on all but a handful of chromosomes. Although these results might indicate true but modest genetic effects, such loci should probably be ignored, at least until further evidence is put forward to implicate them. Given the number of significant and replicated regions of linkage in IBD, it is rational that attention should be focused on these.

### Loci identified by genome scans

In the first genome scan published, Hugot et al. studied the equivalent of 110 sib pairs with Crohn's disease and identified a region of linkage spanning the centromere on chromosome 16 with a peak multipoint Lod score of 3.17 (Ref. 25). Although conferring a locus-specific $\lambda s$ ratio of just 1.3, this linkage was subsequently replicated by several other groups (Refs 26, 27, 28, 29, 30, 31, 32, 33) (Table 3), and *NOD2* has recently been identified as the causative susceptibility gene (Refs 34, 35, 36). There are many lessons to be learned from this finding, as discussed below. Of interest, a meta-analysis of worldwide genotyping data relating to this locus carried out shortly before *NOD2* was identified had confirmed its contribution to Crohn's disease (Lod = 5.2 from 581 affected sib pairs) (Ref. 37), and suggested that this locus is specific to Crohn's disease – making little, if any, contribution to susceptibility to ulcerative colitis. The latter observation has been confirmed with the identification of *NOD2*.

The Oxford (UK) IBD genome scan of 186 sib pairs identified regions that appeared to contribute to both Crohn's disease and ulcerative colitis on chromosomes 12 (Lod = 5.47), 7 (Lod = 3.1) and 3 (Lod = 2.6), and a region close to the human leukocyte antigens (HLAs) on chromosome 6p appeared linked to ulcerative colitis sib pairs (Ref. 38). The regions on chromosomes 12 (Refs 31, 32, 39, 40, 41) (Table 3) and 3 (Refs 42, 43, 44) have subsequently been replicated in independent datasets, with recent work in a combined Oxford and Pittsburgh (USA) study suggesting that a locus on chromosome 12 makes a substantially stronger contribution to ulcerative colitis than Crohn's disease (Ref. 45). This might in part explain the failure of some Crohn's-disease-dominated datasets to reveal linkage in this region (Refs 29, 46, 47, 48). If its contribution to Crohn's disease is only minor then very large datasets would be required to provide sufficient power to replicate the linkage (Refs 49, 50).

More recent genome scans have identified several additional linkages. Cho et al. studied a panel of 297 affected relative pairs from Chicago and Johns Hopkins, Baltimore, USA and identified peak linkage on chromosome 1, with the interesting observation of an epistatic (gene–gene) interaction between this locus and the chromosome 16 locus (Ref. 28). Currently, the power of available linkage programs to detect such multiplicative interactions and conditional gene effects is rather low. It will expand, however, and such expansion will bring with it the potential to reveal novel linkages and a better understanding of gene–gene and gene–environment interactions.

In conducting a genome scan using the largest sib pair panel reported to date (353 affected sib pairs), Hampe et al. replicated the linkages on chromosomes 12 and 16. They also found evidence for linkage of both Crohn's disease and ulcerative colitis around the HLA region on the short arm of chromosome 6



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

**Table 3. Worldwide linkage data for the inflammatory bowel disease candidate regions on chromosomes 12 and 16 (tab003mpo)**

| Centre | Chromosome 12 linkage (MLS)[a] | Chromosome 16 linkage (MLS)[a] | Refs |
|---|---|---|---|
| Oxford | 4.11 | 2.25 | 27, 38 |
| Paris | NS | 3.2 | 48, 85 |
| Los Angeles | 2.7 | 2.5 | 26, 41, 86 |
| Pittsburgh | 2.75 | NS | 39 |
| Chicago / Baltimore | NS | 2.5 | 28, 29 |
| Toronto | 0.3 | 0.2 | 46 |
| Canberra | 0.6 | 6.3 | 30 |
| New York | 2.08 | 0.9 | 40 |
| Leuven | NS | NS | 47 |
| Rome | NS | 2.1 | 87 |
| Helsinki | NS | 1.15 | 43 |
| Axys Pharmaceuticals, La Jola | 1.8 | 1.7 | 31, 32 |

[a]MLS is the maximum Lod score attained in the results of multipoint linkage analysis – i.e. the maximum evidence for linkage observed at the specified locus.
Abbreviation: NS, not significant.

(combined Lod = 2.0). This area of linkage had previously been detected with varying levels of significance (Refs 28, 38, 51), and was subsequently replicated by Rioux et al. using a Crohn's-disease-dominated panel (Lod = 2.3) (Ref. 44). Another interesting finding from Hampe et al. was a Lod score of 1.76 on the X chromosome; this is consistent with the clinical association between Turner's syndrome (which results from the deletion of an X chromosome, i.e. XO) and IBD. Linkage to the X chromosome has also been reported by Vermeire et al. (Ref. 52), and break-point mapping to identify which genes are disrupted in cases of Turner's syndrome and IBD might prove fruitful in fine mapping this locus. Peak linkage in the Hampe study was found on chromosome 10 (Lod = 2.3), but this has not been detected in other studies.

Peak linkage in a Canadian study was found on chromosome 19p13 (Lod = 4.6) (Ref. 44). Of the six other genome scans carried out by other groups, none has found evidence for linkage here. This study made much of the linkage to chromosome 5q31 (Lod = 3.9), primarily because of the large number of immunoactive genes known to map to this region; however, the contribution of such genes to IBD susceptibility has yet to be established. Although this Lod score is clearly respectable, other centres have found only modest evidence for linkage on chromosome 5, and then only in disease subgroups (Refs 28, 41). Perhaps of greater interest is the fact that a genome scan in a mouse model of IBD produced evidence for linkage at a murine region syntenic to human chromosome 5 (Ref. 53).

Both Ma et al. (Ref. 41) and Duerr et al. (Ref. 54) found significant evidence for linkage to chromosome 14q11 (Lod = 2.8 and 3.6, respectively) in Crohn's-disease-dominated panels, thereby together satisfying the criteria for a 'confirmed' locus. It has subsequently been replicated in a Belgian panel of families (S. Vermeire, University of Leuven, Belgium, pers. commun.). This is clearly a region of interest.



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

## Fine mapping the candidate regions and positional candidate gene studies

There are therefore five regions in which linkage to IBD appears to be both strong and replicable – on chromosomes 1, 3, 6, 12, 14 and 16. The next – and undoubtedly greater – challenge is to fine map each of these loci and identify the gene or genes responsible (Box 1). Fine mapping strategies are discussed below, and an algorithm is presented in Figure 1 that outlines possible approaches, but, in a rapidly developing field such as the genetics of complex diseases, the strategies are in constant evolution. Of note, the two groups who simultaneously reported the role for *NOD2* in Crohn's disease used differing approaches: Hugot et al. followed a positional cloning strategy and Ogura et al. used a positional candidate gene approach (Refs 34, 35). Serendipity and luck are likely to play an important role in the successful identification of other IBD susceptibility genes. The chance finding of a family in which a cytogenetic abnormality co-segregates with susceptibility to IBD might localise the gene more effectively than any amount of painstaking family collection and methodical genotyping, or any number of candidate gene studies. This emphasises the need for clinicians treating IBD patients to be genetically literate and aware of the opportunity that such cases might present.

## Methodological issues and uncertainties

### *Fine mapping using pure positional cloning approaches*

It is now well recognised that affected sib pair and affected relative pair methods of linkage analysis are poorly suited to fine mapping, with models suggesting that prohibitively large panels of 700–2000 pairs would be required to narrow candidate intervals to even the modest extent of, for example, 1 cM (Refs 23, 55, 56). It is also increasingly recognised that the saturation of a candidate interval with ever more markers contributes very little to its narrowing by linkage. For reasons that are incompletely understood, but that probably relate to accumulation of genotyping errors and mapping errors, such saturation often results in a diminution of Lod scores and an increasingly spiky linkage curve (Ref. 57). However, the saturation strategy might still be useful for the purposes of detecting evidence of linkage disequilibrium (LD), and also the ever-increasing availability of accurate sequence data and increasingly comprehensive transcript and expressed sequence tag (EST) maps will reduce much of the physical mapping effort required in gene localisation.

With regard to fine mapping the candidate regions identified in genome scans, increasing attention is now turning to the techniques of association (or LD) mapping (Ref. 58). This includes family-based methods such as the transmission disequilibrium test (TDT) (Ref. 59) and its increasingly powerful derivatives (Ref. 60), as well as classic association studies using panels of patients and healthy controls (Box 1). Such studies are based on the premise that a founder effect is present and detectable. Where a mutation arises de novo, it is in LD with alleles at all its neighbouring polymorphic loci, but with succeeding generations this LD gradually decays as a result of meiotic recombination. In a panel of individuals who share a disease caused by the mutation, allele frequencies at distant loci will achieve equilibrium with population allele frequencies early, whereas allele frequencies at those loci very close to the mutation will continue to show distortion for many generations.

Given the large number of meiotic steps that would be expected to separate 'sporadically' affected individuals (i.e. given their distant relationship), distortion in allele frequencies would be expected to occur over a much narrower chromosomal segment in association studies as compared with, for example, distortion in allele sharing in a panel of affected relative pairs (Ref. 5). The family-based methods, which generally use the non-inherited alleles to provide control frequencies, have the advantage that they avoid potentially spurious associations arising as a result of poor matching in the control group (so-called population admixture).

By studying large panels of patients who are geographically or culturally isolated (as described in Ref. 61), or who share phenotypes that are narrow (e.g. ileal Crohn's disease only, or fistulating Crohn's disease only) or extreme (e.g. paediatric Crohn's disease, or Crohn's disease occurring in multiply affected families), it should be possible to diminish genetic heterogeneity sufficiently to allow the currently large genetic intervals to be narrowed. Two examples demonstrating the effects of heterogeneity on Lod scores in the context of linkage studies in IBD have recently been published (Refs 45, 62).



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

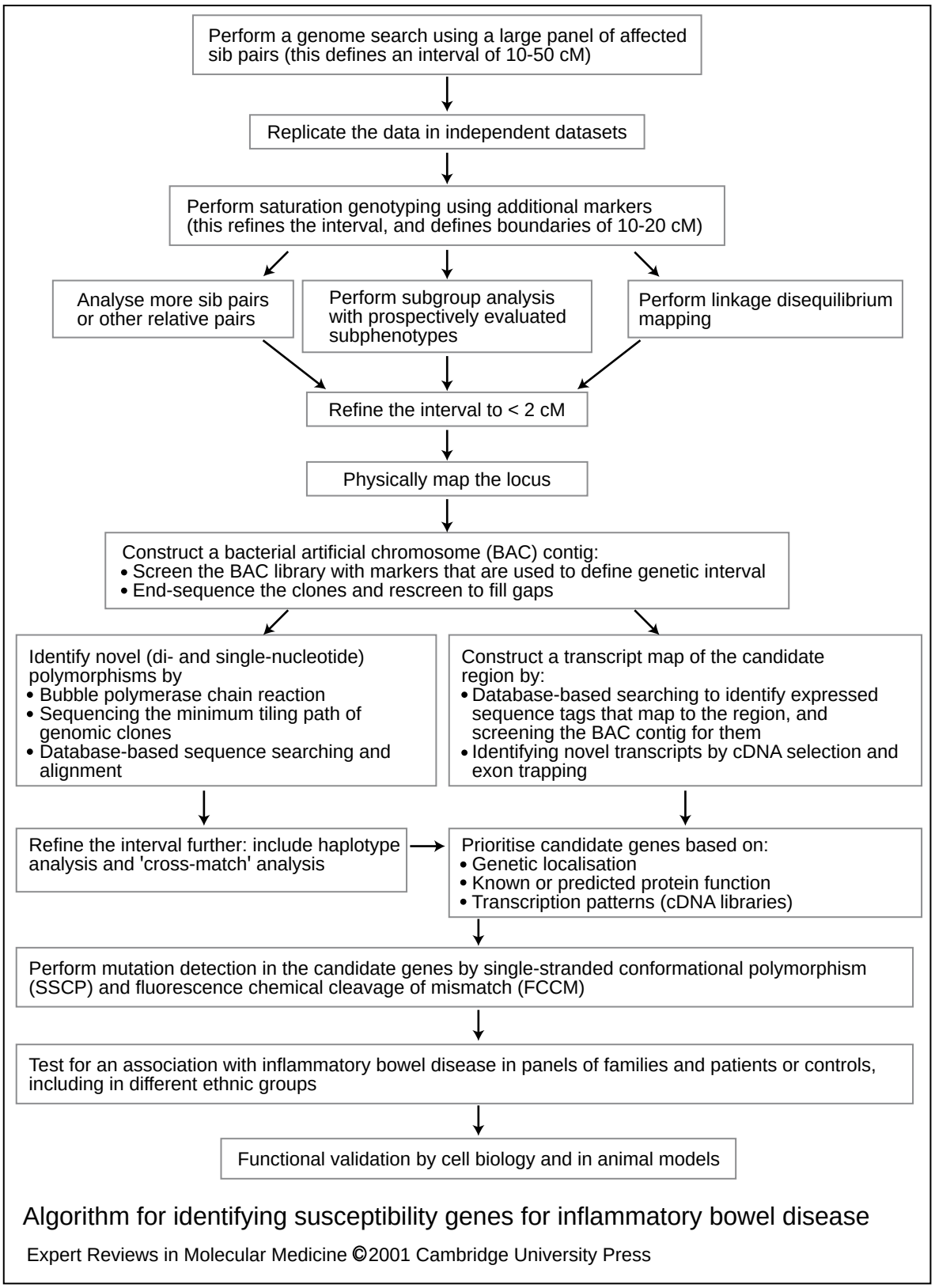



**Figure 1. Algorithm for identifying susceptibility genes for inflammatory bowel disease** (*see next page for legend*) **(fig001mpo)**.



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

**Figure 1. Algorithm for identifying susceptibility genes for inflammatory bowel disease.** This schema gives one approach, and outlines the types of methods that are currently being employed. With the advances in technologies, and increase in availability of polymorphisms and sequence information in databases, the steps suggested in this schema are subject to continuous change, and some might not be required at all **(fig001mpo)**.

### *Positional candidate gene approaches*

Positional candidate gene studies provide one means for reducing the labour involved in positional cloning, but are not without problems. As illustrated by the study of Olavesen et al., the comprehensive assessment of the contribution of a gene is not straightforward (Ref. 63). All polymorphisms must be detected, including potential regulatory sequences present in 5' non-coding regions and introns. Allele frequencies at these loci must then be compared in panels of patients and controls. One of the striking features of their study was the lack of LD even between polymorphic loci that mapped to the same gene. This emphasises the need to identify and study all polymorphic loci that occur at significant frequency before dismissing the contribution of a particular gene. Even where association is demonstrated it might not indicate a contribution of that gene, but might rather reflect LD with polymorphisms in a neighbouring gene.

To an extent, the study of Olavesen et al. also underlines the hazards of candidate gene approaches overall, in their innate assumptions regarding pathogenic disease mechanisms. One of the lessons of positional cloning in monogenic diseases is the extent to which genes that on first appearances (predicted function, expression patterns and so on) are unlikely candidates are subsequently proven to harbour the pathogenic mutation. Who would have predicted that the HFE gene, which, when defective, causes haemochromatosis, would be a close relative of class I HLA molecules (Ref. 64), or that Darier's disease (which causes acantholysis and gross derangement of skin keratinization) would be caused by a defect in a calcium-ATPase transport protein (Ref. 65)? In each of these positional cloning efforts, identification of the correct gene was delayed as other, seemingly more likely candidate genes were pursued. Pre-supposition regarding the nature of the genetic defects in IBD is at least as likely to hinder progress.

To date, progress in fine mapping the candidate regions has been variably successful. The following discussion of results from fine mapping studies focuses on chromosomes 1p, 6p, 12 and 16. This is not a comprehensive review, but rather illustrates some of the more important leads, and the methodologies that are being applied.

### Chromosome 1p

Cho et al. used a founder population of Chaldeans who migrated to North America from Iraq in the 16th century to fine map the chromosome 1 region originally identified in their genome scan of outbred affected relative pairs (Ref. 61). They identified a consensus haplotype of SNPs and microsatellite alleles that appears to co-segregate with disease susceptibility, and refines the 1p locus to a region of approximately 1 cM. Although linkage to this region has not been strong in other panels, this fine mapping is clearly exciting. With a region so refined, and given the availability of ever more sequence data, SNPs and methods for gene identification, there is genuine optimism that an IBD susceptibility gene will be identified from this candidate region in the near future.

### Chromosome 6p – IBD3

Fine mapping the chromosome 6 locus is likely to be confounded by the extreme LD that exists across this region, with the extended haplotypes reflecting the relative paucity of recombination events around the HLA region in ancestral meioses. Indeed, this LD contributed to the difficulties and delays in fine mapping the haemochromatosis gene that lies nearby – a gene for a monogenic disease whose position was known to within 5 cM for some 20 years before it was cloned (Ref. 64).

The HLA region was first implicated in IBD in Japanese studies using serological techniques for genotyping in which HLA-DR2 was identified as being associated with ulcerative colitis (Ref. 66). This finding was subsequently confirmed in the Japanese population using molecular methods of allele-specific polymerase chain reaction (PCR) for genotyping, in which the DRB1*1502 allele of DR2 was implicated (Refs 67, 68).

In an association study using 348 Caucasian IBD patients and 472 ethnically matched controls, Satsangi et al. identified association between the HLA-DR3, DQ2 haplotype and extensive ulcerative colitis particularly in females, and the



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

rare DRB1*0103 allele, which showed association with both severe disease requiring surgery and also the presence of extra-intestinal manifestations (Ref. 51). The latter association was confirmed by Roussomoustakaki in a panel of 107 ulcerative colitis patients requiring colectomy: DRB1*0103 was present in 14.1% of patients versus 3.2% of controls ($P < 1 \times 10^{-6}$), and also appeared to be associated with the presence of extra-intestinal manifestations of IBD (Ref. 69).

The role of DRB1*0103 in the extra-intestinal manifestations of IBD has been further explored by Orchard et al. (Ref. 70). This work has provided strong evidence to suggest that there are two phenotypically and genotypically distinct forms of arthropathy associated with IBD. A large-joint inflammatory oligo-arthropathy (type 1) is strongly associated with DRB1*0103, whereas a small-joint arthropathy, which is less closely associated with flares of intestinal inflammation, is not associated with DRB1*0103 but is associated with HLA-B44 (Ref. 70). Furthermore, DRB1*0103 has been shown to be strongly associated with other extra-intestinal manifestations such as erythema nodosum and iritis, highlighting the fact that some genes might play an important role in determining the phenotype of IBD, possibly independently of a contribution to disease susceptibility per se.

A detailed analysis of the gene-rich HLA region including the nearby MICA gene polymorphisms is currently under way in our laboratory in Oxford. Recent interest has focused on association between Crohn's disease and a functional polymorphism within the promoter region of the TNF-α gene (Ref. 71). Several SNPs have been detected within the promoter region of this gene, with data suggesting that some of these determine levels of TNF production. Such a role has been attributed to the relatively recently described –1031 polymorphism, where the substitution of thymidine for cytosine results in increased production of this pro-inflammatory cytokine. Association between this SNP and Crohn's disease was first observed in a Japanese population (Ref. 72), where the frequency of the C allele in 107 Crohn's disease patients was 24% and in healthy controls was 16%. This finding has subsequently been replicated in association studies on two independent panels of UK Caucasian Crohn's disease patients. In a case–control study the allele frequency in 121 patients with Crohn's disease was 27.6% compared with 18.4% in controls ($P = 0.005$), and in a panel that included a number of multiply affected families a significant distortion of alleleic transmission was observed in a TDT ($P = 0.01$) (Ref. 71).

Several factors testify to the validity of this result. The association is with a functionally significant polymorphism within a gene known to encode a molecule of great significance to the pathogenesis of Crohn's disease. Association is seen not only in more than one dataset, but also in panels of different ethnicity – increasing the likelihood that the result is true, and reducing the chance that it is due to LD. And finally, the association is found at a locus at which strong and replicated linkage has been observed – both in sib pair linkage studies and using the TDT, which has properties of both linkage and association. Despite these factors, however, interpretation of these data will have to take account of the numerous immunoactive genes that map to the short arm of chromosome 6 and also the extensive LD in this area – an issue that most studies of genes in this region to date have failed to address.

Although a positive association with an intragenic polymorphism might implicate the gene, a negative result does not necessarily exclude it. Particularly in situations where LD is less strong or variable, a rigorous assessment of the contribution of any given gene requires detection of all common variants, and either determination of patterns of LD or study of each polymorphism in the panel of IBD patients and controls. Only once all polymorphic loci in a given gene have been studied can a contribution of that gene to disease susceptibility be excluded.

## Chromosome 16 – IBD1

The IBD1 locus on chromosome 16 had been extensively replicated using techniques of linkage analysis in a number of panels of multiply affected families around the world, and the optimism that this generated was recently realised with the identification of mutations in the *NOD2* gene as playing an important role in the pathogenesis of Crohn's disease – but not ulcerative colitis.

In identifying *NOD2* Hugot et al. applied classic techniques of LD mapping (Ref. 34). A marginally positive TDT result was identified at a particular microsatellite marker (D16S3136) from a dense panel of 26 markers that had been genotyped across the candidate interval in a panel of 108 families. The finding at this marker was replicated in a second panel – interestingly, with



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

a different allele implicated. A bacterial artificial chromosome (BAC) clone containing this marker was sequenced, and SNPs were identified and typed in the combined family panel. Alleles at several of the markers showed strong association with Crohn's disease, but only three (a frameshift mutation and two point mutations) were present on independent haplotypes and were non-synonymous. A database search demonstrated an EST match, and the gene was identified as *NOD2*. Average relative risks for Crohn's disease computed for genotypes containing zero, one or two of the variants were 1, 3 and 38, respectively (44 for compound heterozygotes) – suggesting a recessive nature of Crohn's disease susceptibility, as had been indicated in previous segregation analyses (Refs 73, 74).

Ogura et al. studied *NOD2* as a positional candidate gene (Ref. 35), based on its known map position within the IBD1 locus and known role in recognising bacterial components. These investigators sequenced exons and flanking introns in 12 Crohn's patients and four controls, and identified the same frameshift mutation as observed by Hugot et al. Typing this mutation in a large panel of patients and their parents demonstrated marked preferential transmission of the 3020insC mutation from heterozygous parents to affected individuals. The gene is known to regulate activity of the nuclear factor NF-κB. Functional data from this group produced the somewhat puzzling finding that the mutant form of NOD2 is significantly less active than the wild-type protein in conferring responsiveness (stimulation) to bacterial lipopolysaccharide.

Hampe et al. simultaneously reported association with the same 3020insC mutation in a European panel of individuals with Crohn's disease (Ref. 36), again with a gene dosage effect and a particularly high relative risk of 42 in individuals homozygous for the *NOD2* insertion.

As is frequently the case in genetic studies it is currently unclear how mutations in the *NOD2* gene predispose to Crohn's disease. The protein product of *NOD2* is expressed in monocytes and apparently not the intestinal mucosa, and known functions include activation of NF-κB (thereby determining responsiveness to bacterial lipopolysaccharide) and a role in regulating apoptosis. The mutations identified appear to predominantly affect the leucine-rich repeat domain, which plays a major role in regulating NF-κB. Whether these or other functions are important pathogenic mechanisms remains to be seen.

Perhaps the major lesson from *NOD2* at this stage is as a proof of principle, demonstrating that polygenes can indeed be identified using the strategies that had been postulated. A cautionary note for investigators in other diseases might be sounded, as the a priori expectation was that Crohn's disease held many advantages, and could be viewed as a model complex disease for genetic analysis. Nonetheless, the finding will lend hope across the field of the genetics of complex diseases, and by conditioning datasets for the presence or absence of *NOD2* (i.e. reducing heterogeneity) should also facilitate identification of other genes as well as environmental factors contributing to the pathogenesis of Crohn's disease.

## Chromosome 12 – IBD2

Success in fine mapping the chromosome 12 locus has been limited to date. Although it appears to make a stronger contribution to ulcerative colitis than Crohn's disease, the distortion in allele sharing is seen over a wide chromosomal segment (Ref. 45). Different groups have detected peaks in different positions across a 40 cM segment (Refs 31, 32, 38, 39, 40, 41), which probably reflects the poor resolution of allele methods of linkage analysis (Ref. 75) but might also point to the presence of more than one susceptibility gene in this region. A precedent for the latter comes from fine mapping in mouse models of type 1 diabetes (Ref. 76).

Preliminary data from our group in Oxford has demonstrated positive evidence for association (particularly with ulcerative colitis) at three adjacent microsatellite markers within a 150 kb segment (Ref. 77). The alleles implicated also show strong LD, suggesting that the distortion observed is likely to be true. Additional evidence for this is currently being sought by our group by typing a panel of SNPs that map to the region. Given that association studies provide substantially higher levels of resolution for gene localisation than allele sharing methods of linkage analysis, this work might be the first clue to the gene localisation within this locus. Two positional candidate gene studies – of interferon γ (IFN-γ) and integrin β7 – have been negative (Refs 78, 79), although the study of IFN-γ focused on a single (dinucleotide repeat) polymorphism and therefore cannot completely exclude a role for this gene. Other positional candidate gene studies are in progress.



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

## Technology to the rescue?

Ultimately, the fine mapping of any complex-disease locus is likely to require application of the range of technological advances that promise to revolutionise the field of molecular genetics. This includes non-PCR-based methods of high-throughput genotyping (Ref. 8), with SNPs possibly providing a more robust and easily automated substrate than microsatellite markers (Ref. 23). Chip technology based on microarray platforms is likely to play a prominent role, particularly with the increased flexibility provided by the combination of universal array technology and robust allele-discriminating techniques such as minisequencing and oligonucleotide ligation (Ref. 8).

Large panels of clinically well-characterised families and patients/controls will have to be genotyped to provide adequate power for the detection of genes that make only a modest contribution to IBD susceptibility. The burgeoning field of bioinformatics will aid the data handling and quality-control issues presented by the huge amounts of genotyping data that will be generated. Developments in statistical software packages also promise more powerful, valid and robust data analysis with the capability to tease out what might be the weakest of signals among the noise. These will include analysis for haplotype association, which is likely to be substantially more powerful than association at individual polymorphic loci (Ref. 80). That the signal is likely to be weak for the majority of polygenes is suggested by the 'common disease – common variant' hypothesis as discussed above (Refs 22, 23). The 'mutations' underlying common diseases might be present at a relatively high frequency in the normal population, and it is the combination of such variants in a number of genes, together with exposure to one or more currently unidentified environmental triggers, that leads to disease. Although this hypothesis has gained widespread acceptance, it is clear from *NOD2* that rare mutations might be equally important in predisposing to complex disease.

## Potential clinical implications of IBD gene mapping

The potential advantage that identifying susceptibility genes will give in terms of understanding disease pathogenesis and hence allowing rational new therapies to be developed is emphasised by the large investment being made in this field by the pharmaceutical industry. Even before new drugs are developed, the genetics of complex diseases will have an impact in allowing stratification of patients for drug trials. It might be, for example, that the responsiveness or otherwise of a given Crohn's disease patient to azathioprine or methotrexate has a genetic basis, or that side effects and toxicity are likewise genetically determined. By stratifying trial patients by patterns of allele sharing at particular loci, it might in future be possible to predict which patients are likely to respond to a given therapy, and which patients will develop side effects. Subdividing the trial cohort using rational objective genotyping data might thus increase the power of drug trials to detect and aid understanding of both positive beneficial effects and negative toxicity events. In so doing, it might expedite the passage of drugs through the trial phase, hence markedly reducing the cost to industry, and possibly reducing the risk of potentially useful drugs being rejected if their toxicity and therapeutic effects can be better predicted.

Such pharmacogenetic strategies have already been attempted with regard to TNF polymorphisms and the response to the anti-TNF antibody infliximab in Crohn's disease. Indeed, it has recently been reported that, although TNF microsatellite haplotypes did not appear to predict response to therapy, response to anti-TNF treatment was associated with a promoter polymorphism at position –308 (Refs 81, 82). Many similar studies are to be expected in the future – particularly given the recently characterised functional polymorphism at –1031 in the promoter region of the TNF gene.

Genetically determined variation in thiopurine methyltransferase (TPMT) activity has also been shown to have a marked influence on the risk of developing leukopaenia in response to azathioprine therapy – as used in treatment of refractory IBD. This occurs particularly in individuals homozygous (1 in 300) or heterozygous (1 in 8) for a TPMT polymorphism, which results in reduced enzyme activity and hence reduced drug metabolism. At present, the full range of predisposing risk alleles are incompletely understood (Ref. 83), and patients are not routinely genotyped prior to starting azathioprine, but there might come a time when this is deemed necessary.



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

A reclassification of Crohn's disease and ulcerative colitis into genotypically defined subgroups based on the patterns of shared allelic variants should clarify much of the unpredictability associated with the management of IBD based on the current phenotypic principles. With knowledge of the genetic basis of any individual's disease, it should be substantially easier to predict future patterns of disease behaviour, likely environmental triggers and likely therapeutic responses. Lifestyle advice and therapies will thus be targeted in a much more specific manner, with less prescribing on a 'try it and see' basis and a reduction in drug-associated toxicity.

A complete understanding of the genetic basis of IBD should provide important insights into the nature of environmental triggers and drives (Ref. 84). Already the identification of *NOD2* has provided a clear link between the immune response to enteric bacteria and the development of Crohn's disease. In the future, for individuals deemed to be at risk by virtue of possessing an 'at risk' complex of susceptibility alleles at various loci, it might thus be appropriate to seek environmental modification. On the basis of present understanding, this would include avoidance of non-steroidal anti-inflammatory drugs and of smoking in those who are at risk of developing Crohn's disease. In the future, it might include vaccination against whatever environmental agents are found to play a causative role – or possibly induction of immune tolerance by intestinal exposure to particular antigens at particular stages of development. By such strategies, it might be possible to abrogate the development and progression of IBD entirely, avoiding the need for ongoing therapies such as infliximab, and the next generation of increasingly expensive molecular therapies. These are the ultimate goals of the IBD molecular geneticists, and with the application of new technologies they should be ultimately attainable.

## Acknowledgements and funding

We thank Professor Jack Satsangi (Department of Gastroenterology, Western General Hospital, Edinburgh, UK) and Dr Severine Vermeire (Department of Gastroenterology, University of Leuven, Belgium) for their critical review of the manuscript and for their helpful comments. Miles Parkes was a Medical Research Council Clinical Research Fellow.

## References

1 Calkins, B.M. and Mendeloff, A.I. (1995) The epidemiology of idiopathic inflammatory bowel disease. In Inflammatory Bowel Disease (4th edn) (Kirsner, J.B. and Shorter, R.G., eds), pp. 31-68, Williams and Wilkins, Baltimore
2 Davies, J.L. et al. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371, 130-136, PubMed ID: 94352408
3 Hashimoto, L. et al. (1994) Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 371, 161-164, PubMed ID: 94352412
4 Lander, E.S. and Schork, N.J. (1994) Genetic dissection of complex traits. Science 265, 2037-2048, PubMed ID: 94378000
5 Strachan, T. and Read, A. (1996) Human Molecular Genetics, BIOS Scientific Publishers Limited
6 Nature 2001 Vol. 409 (6882), http://www.nature.com/cgi-taf/dynapage.taf?file=/nature/journal/v409/n6822/index.html
7 Science 2001 Vol. 291 (5507), http://www.sciencemag.org/content/vol291/issue5507/index.shtml
8 Weaver, T. (2000) High-throughput SNP discovery and typing for genome-wide genetic analysis. Trends in Genetics: New Technologies for Life Sciences: a Trends Guide, 36-42
9 Tysk, C. et al. (1988) Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut 29, 990-996, PubMed ID: 88284515
10 Orholm, M. et al. (1996) Inflammatory bowel disease in a Danish twin register. Gut 39 (Suppl 3), A187
11 Subhani, J. et al. (1998) Concordance rates of twins and siblings in IBD. Gastroenterology 114, A3113
12 Farmer, R.G., Michener, W.M. and Mortimer, E.A. (1980) Studies of family history among patients with inflammatory bowel disease. Clin Gastroenterol 9, 271-277, PubMed ID: 80223442
13 McConnell, R.B. (1990) Genetics of inflammatory bowel disease. In Inflammatory Bowel Disease (Allan, R.N. et al., eds), pp. 11-23, Churchill Livingstone
14 Satsangi, J., Rosenberg, W.M.C. and Jewell, D.P. (1994) The prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. Euro J Gastro Hepatol 6, 413-416
15 Risch, N. (1990) Linkage strategies for genetically



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

complex traits. II. The power of affected relative pairs. Am J Hum Genet 46, 229-241, PubMed ID: 90144732
16 Satsangi, J. et al. (1996) Clinical patterns of familial inflammatory bowel disease. Gut 38, 738-741, PubMed ID: 96334966
17 Colombel, J.F. et al. (1996) Clinical characteristics of Crohn's disease in 72 families. Gastroenterology 111, 604-607, PubMed ID: 96374272
18 Bayless, T.M. et al. (1996) Crohn's disease: concordance for site and clinical type in affected family members—potential hereditary influences. Gastroenterology 111, 573-579, PubMed ID: 96374268
19 Peeters, M. et al. (1996) Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. Gastroenterology 111, 597-603, PubMed ID: 96374271
20 Kirsner, J.B. (1973) Genetic aspects of inflammatory bowel disease. Clin Gastroenterol 2, 557-571
21 Orholm, M. et al. (1991) Familial occurrence of inflammatory bowel disease. N Engl J Med 324, 84-88, PubMed ID: 91074235
22 Lander, E.S. (1996) The new genomics: global views of biology. Science 274, 536-539, PubMed ID: 97066815
23 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. Science 273, 1516-1517, PubMed ID: 96379554
24 Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11, 241-247, PubMed ID: 96042570
25 Hugot, J.P. et al. (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 379, 821-823, PubMed ID: 96172856
26 Ohmen, J.D. et al. (1996) Susceptibility locus for inflammatory bowel disease on chromosome 16 has a role in Crohn's disease, but not in ulcerative colitis. Hum Mol Genet 5, 1679-1683, PubMed ID: 97049979
27 Parkes, M. et al. (1996) Susceptibility loci in inflammatory bowel disease. Lancet 348, 1588, PubMed ID: 97108576
28 Cho, J.H. 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 Natl Acad Sci U S A 95, 7502-7507, PubMed ID: 98301604
29 Brant, S.R. et al. (1998) American families with Crohn's disease have strong evidence for linkage to chromosome 16 but not chromosome 12. Gastroenterology 115, 1056-1061, PubMed ID: 99014189
30 Cavanaugh, J.A. et al. (1998) Analysis of Australian Crohn's disease pedigrees refines the localization for susceptibility to inflammatory bowel disease on chromosome 16. Ann Hum Genet 62, 291-298, PubMed ID: 99123682
31 Curran, M.E. et al. (1998) Genetic analysis of inflammatory bowel disease in a large European cohort supports linkage to chromosomes 12 and 16. Gastroenterology 115, 1066-1071, PubMed ID: 99014191
32 Hampe, J. 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, 808-816, PubMed ID: 99162204
33 Annese, V. et al. (1999) Genetic analysis in Italian families with inflammatory bowel disease supports linkage to the IBD1 locus—a GISC study. Eur J Hum Genet 7, 567-573, PubMed ID: 99368836
34 Hugot, J.P. et al. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411, 599-603, PubMed ID: 21279172
35 Ogura, Y. et al. (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411, 603-606, PubMed ID: 21279173
36 Hampe, J. et al. (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 357, 1925-1928, PubMed ID: 21319311
37 Cavanaugh, J. (2001) International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. Am J Hum Genet 68, 1165-1171, PubMed ID: 21205595
38 Satsangi, J. et al. (1996) Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nat Genet 14, 199-202, PubMed ID: 96438860
39 Duerr, R.H. et al. (1998) Linkage and association between inflammatory bowel disease and a locus on chromosome 12. Am J Hum Genet 63, 95-100, PubMed ID: 98299803





**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

40 Akolkar, P. et al. (1998) Fine mapping of regions linked to Crohn's disease on chromosomes 12 and 16. Am J Hum Genet 63, A279
41 Ma, Y. et al. (1999) A genome-wide search identifies potential new susceptibility loci for Crohn's disease. Inflamm Bowel Dis 5, 271-278, PubMed ID: 20046082
42 Hampe, J. et al. (2000) Fine mapping of the chromosome 3p susceptibility locus in IBD. Gastroenterology 118, A1802
43 Paavola, P. et al. (2000) Suggestive evidence for linkage to chromosome 3p21 in Finnish IBD families. Gastroenterology 118, A1819
44 Rioux, J.D. et al. (2000) Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 66, 1863-1870, PubMed ID: 21060603
45 Parkes, M. et al. (2000) The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. Am J Hum Genet 67, 1605-1610, PubMed ID: 20530245
46 Rioux, J.D. et al. (1998) Absence of linkage between inflammatory bowel disease and selected loci on chromosomes 3, 7, 12, and 16. Gastroenterology 115, 1062-1065, PubMed ID: 99014190
47 Vermeire, S. et al. (2000) Exclusion of linkage of Crohn's disease to previously reported regions on chromosomes 12, 7, and 3 in the Belgian population indicates genetic heterogeneity. Inflamm Bowel Dis 6, 165-170, PubMed ID: 20415761
48 Lesage, S. et al. (2000) Genetic analyses of chromosome 12 loci in Crohn's disease. Gut 47, 787-791, PubMed ID: 20530576
49 Suarez, B., Hampe, C. and Eerdewegh, P.V. (1994) Problems of replicating linkage claims in psychiatry. In Genetic Approaches to Mental Disorders (Gershon, E. and Cloniger, C., eds), pp. 23-46, American Psychiatric Press
50 Mandal, D.M. et al. (1999) Environmental covariates: effects on the power of sib-pair linkage methods. Genet Epidemiol 17, S643-648, PubMed ID: 20065375
51 Satsangi, J. et al. (1996) Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet 347, 1212-1217, PubMed ID: 96206417
52 Vermeire, S. et al. (2001) Evidence for inflammatory bowel disease of a susceptibility locus on the X chromosome. Gastroenterology 120, 834-840, PubMed ID: 21153317
53 Mahler, M. et al. (1999) Genetic analysis of susceptibility to dextran sulfate sodium-induced colitis in mice. Genomics 55, 147-156, PubMed ID: 99134292
54 Duerr, R.H. et al. (2000) High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. Am J Hum Genet 66, 1857-1862, PubMed ID: 21060604
55 Boehnke, M. (1994) Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. Am J Hum Genet 55, 379-390, PubMed ID: 94311321
56 Kruglyak, L. and Lander, E.S. (1995) High-resolution genetic mapping of complex traits. Am J Hum Genet 56, 1212-1223, PubMed ID: 95243240
57 Feakes, R. et al. (1999) Exploring the dense mapping of a region of potential linkage in complex disease: an example in multiple sclerosis. Genet Epidemiol 17, 51-63, PubMed ID: 99254763
58 Cardon, L.R. and Bell, J.I. (2001) Association study designs for complex diseases. Nat Rev Genet 2, 91-99, PubMed ID: 21147228
59 Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52, 506-516, PubMed ID: 93190982
60 Martin, E.R. et al. (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67, 146-154, PubMed ID: 20307186
61 Cho, J.H. et al. (2000) Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. Hum Mol Genet 9, 1425-1432, PubMed ID: 20275631
62 Brant, S.R. et al. (2000) Linkage heterogeneity for the IBD1 locus in Crohn's disease pedigrees by disease onset and severity. Gastroenterology 119, 1483-1490, PubMed ID: 20565554
63 Olavesen, M.G. et al. (2000) Analysis of single-nucleotide polymorphisms in the interleukin-4 receptor gene for association with inflammatory bowel disease. Immunogenetics 51, 1-7, PubMed ID: 20130126
64 Feder, J.N. et al. (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13, 399-408, PubMed ID: 96331279
65 Sakuntabhai, A. et al. (1999) Mutations in





**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

ATP2A2, encoding a Ca2+ pump, cause Darier disease. Nat Genet 21, 271-277, PubMed ID: 99178263
66 Asakura, H. et al. (1982) Association of the human lymphocyte-DR2 antigen with Japanese ulcerative colitis. Gastroenterology 82, 413-418, PubMed ID: 82096121
67 Sugimura, K. et al. (1993) Analysis of genes within the HLA region affecting susceptibility to ulcerative colitis. Hum Immunol 36, 112-118, PubMed ID: 93216486
68 Kobayashi, K. et al. (1990) HLA-DR, DQ and T cell antigen receptor constant beta genes in Japanese patients with ulcerative colitis. Clin Exp Immunol 80, 400-403, PubMed ID: 90322735
69 Roussomoustakaki, M. et al. (1997) Genetic markers may predict disease behavior in patients with ulcerative colitis. Gastroenterology 112, 1845-1853, PubMed ID: 97322021
70 Orchard, T.R., Wordsworth, B.P. and Jewell, D.P. (1998) Peripheral arthropathies in inflammatory bowel disease: their articular distribution and natural history. Gut 42, 387-391, PubMed ID: 98238207
71 van Heel, D.A. et al. (2001) Transmission disequilibrium testing confirms the association of the TNFalpha -1031C allele with Crohn's disease. Gut 48 (Suppl 1), A319
72 Negoro, K. et al. (1999) Crohn's disease is associated with novel polymorphisms in the 5'-flanking region of the tumor necrosis factor gene. Gastroenterology 117, 1062-1068, PubMed ID: 20005906
73 Kuster, W. et al. (1989) 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 32, 105-108, PubMed ID: 89205580
74 Orholm, M. et al. (1993) Investigation of inheritance of chronic inflammatory bowel diseases by complex segregation analysis. Bmj 306, 20-24, PubMed ID: 93169141
75 Kruglyak, L. and Lander, E.S. (1996) Limits on fine mapping of complex traits. Am J Hum Genet 58, 1092-1093, PubMed ID: 96216162
76 Podolin, P.L. et al. (1998) Localization of two insulin-dependent diabetes (Idd) genes to the Idd10 region on mouse chromosome 3. Mamm Genome 9, 283-286, PubMed ID: 98191727
77 Parkes, M. et al. (1998) Precision mapping of chromosome 12 linkage in IBD: evidence of a haplotype association. Gastroenterology 114, A1058
78 Hampe, J. et al. (1998) The interferon-gamma gene as a positional and functional candidate gene for inflammatory bowel disease. Int J Colorectal Dis 13, 260-263, PubMed ID: 99086997
79 van Heel, D., Carey, A. and Jewell, D. (2001) Novel polymorphism in the beta 7 integrin gene: family based association studies in IBD. Gut 48 (Suppl 1), A48
80 Woolley, A.T. et al. (2000) Direct haplotyping of kilobase-size DNA using carbon nanotube probes. Nat Biotechnol 18, 760-763, PubMed ID: 20347558
81 Vermeire, S. et al. (2000) Response to anti-TNF treatment is associated with the TNFalpha-308*1 allele. Gastroenterology 118, A3595
82 Marion, J. et al. (2000) TNF microsatellite polymorphism does not predict response to infliximab in patients with Crohn's disease. Gastroenterology 118, A3596
83 Colombel, J.F. et al. (2000) Genotypic analysis of thiopurine S-methyltransferase in patients with Crohn's disease and severe myelosuppression during azathioprine therapy. Gastroenterology 118, 1025-1030, PubMed ID: 20295034
84 Sartor, R.B. (1995) Microbial factors in the pathogenesis of Crohn's disease, ulcerative colitis and experimental intestinal inflammation. In Inflammatory Bowel Disease (4th edn) (Kirsner, J.B. and Shorter, R.G., eds), pp. 96-124, Williams and Wilkins
85 Hugot, J.P. et al. (1994) Linkage analyses of chromosome 6 loci, including HLA, in familial aggregations of Crohn disease. G.E.T.A.I.D. Am J Med Genet 52, 207-213, PubMed ID: 95100264
86 Yang, H. et al. (1998) Linkage and association between Crohn's disease and a putative locus on chromosome 12. Gastroenterology 114, G4581
87 Forabosco, P. et al. (2000) Combined segregation and linkage analysis of inflammatory bowel disease in the IBD1 region using severity to characterise Crohn's disease and ulcerative colitis. On behalf of the GISC. Eur J Hum Genet 8, 846-852, PubMed ID: 20546215





**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press

## Further reading, resources and contacts

The following texts/articles provide excellent background on mapping susceptibility genes in complex traits:

Strachan, T. and Read, A. (1996) Human Molecular Genetics, BIOS Scientific Publishers Limited

Kirsner, J. (1999) Inflammatory Bowel Disease, W.B. Saunders

Lander, E.S. and Schork, N.J. (1994) Genetic dissection of complex traits. Science 265, 2037-2048, PubMed ID: 94378000

Cardon, L.R. and Bell, J.I. (2001) Association study designs for complex diseases. Nat Rev Genet 2, 91-99, PubMed ID: 21147228

Nature 2001 Vol. 409 (6882), http://www.nature.com/cgi-taf/dynapage.taf?file=/nature/journal/v409/n6822/index.html

The websites of the National Association for Colitis and Crohn's disease, UK, and the Crohn's and Colitis Foundation of America provide patient support and information on research funding.

http://www.nacc.org.uk

http://www.ccfa.org

The British Society of Gastroenterology and American Gastroenterology Association websites include information on publications, meetings, research and clinical practice for gastroenterologic physicians and scientists.

http://www.bsg.org.uk

http://www.gastro.org

The US National Center for Biotechnology Information's Online Mendelian Inheritance in Man website catalogues human genes and genetic disorders.

http://www.ncbi.nlm.nih.gov/Omim (UC, MIM 191390; IBD1, MIM 266600; IBD2, MIM 601458).

Wellcome Trust Centre for Human Genetics, Oxford, UK, includes links to sites related to complex-disease genetics.

http://www.well.ox.ac.uk

D. Jewell departmental website:

http://www.medicine.ox.ac.uk/gastro

## Features associated with this article

### Figure

Figure 1. Algorithm for identifying susceptibility genes for inflammatory bowel disease (fig001mpo).

### Tables

Table 1. Guidelines for the interpretation of linkage results derived from genome scans in complex traits as proposed by Lander and Kruglyak (tab001mpo).
Table 2. Concordance for Crohn's disease and ulcerative colitis in monozygotic and dizygotic twins in which at least one of the pair has inflammatory bowel disease (tab002mpo).
Table 3. Worldwide linkage data for the inflammatory bowel disease candidate regions on chromosomes 12 and 16 (tab003mpo).

## Citation details for this article

Miles Parkes and Derek Jewell (2001) Ulcerative colitis and Crohn's disease: molecular genetics and clinical implications. Exp. Rev. Mol. Med. 19 November, http://www-ermm.cbcu.cam.ac.uk/0100391Xh.htm



**Accession information:** (01)00391-Xa.pdf (short code: txt001mpo); 19 November 2001
ISSN 1462-3994 ©2001 Cambridge University Press

https://doi.org/10.1017/S146239940100391X Published online by Cambridge University Press