Inflammatory bowel disease: Progress toward a gene

David A van Heel MRCP1,2, Jack Satsangi DPhil MRCP1,2, Alisoun H Carey PhD3, Derek P Jewell DPhil FRCP2

The pathogenesis of ulcerative colitis (UC) and Crohn’s disease (CD) is still unknown, but the importance of genetic susceptibility has been clearly shown by epidemiological data from family and twin studies. Linkage studies have identified two susceptibility loci for inflammatory bowel disease (IBD) on chromosomes 12 and 16. Importantly, these linkages have been replicated by independent investigators, and studies of positional candidates within these regions continue, together with fine mapping strategies. Regions of ‘suggestive’ linkage on chromosomes 1, 3, 4, 6, 7, 10, 22 and X have also been reported in individual studies. Other important candidate genes investigated include the interleukin-1 receptor antagonist, MUC3 and genes of the human leukocyte antigen (HLA) system. The apparently conflicting data in different studies from around the world may be explained by ethnic differences, case mix and genetic heterogeneity. Replicated class II HLA associations include HLA DRB1*0103 and DR2 (DRB1*1502), involved in UC susceptibility, and HLA DRB1*03 and DR4 as resistance alleles for CD and UC respectively.

Animal studies have provided insights from targeted mutations and quantitative trait locus analysis. The goals of continuing research include narrowing the regions of linkages and analysis of candidate genes, and possibly the application of newly developed methods using single nucleotide polymorphisms. Advances in IBD genetics hold the potential to provide knowledge about the disease pathogenesis at the molecular level, with ensuing benefits for clinical practice.

Key Words: Colitis; Crohn’s disease; Genetics; Inflammatory bowel disease; Review; Ulcerative colitis.
Genes are unequivocally shown to be involved in disease susceptibility when there is a consistent Mendelian segregation pattern or a chromosomal abnormality. Although these features have not been found in IBD, genetic susceptibility has been strongly suggested by ethnic differences in disease prevalence, familial clustering of disease, twin concordance rates and, most recently, data from molecular studies in humans and in animal models.

Ethnicity and IBD: Certain ethnic groups have a consistently higher risk of developing IBD. Ashkenazi Jews have a two- to fourfold increased risk of IBD (especially CD) compared with the 150 to 200/100,000 observed in Northern Europe (1-10). The risk of developing IBD is lower in Sephardic and Oriental Jews than in Ashkenazim. The lowest overall prevalence is reported in Afro-Caribbeans (11,12). The contribution of genetic susceptibility rather than entirely environmental factors to these ethnic differences is further suggested by the persistently high risk in Ashkenazi Jews regardless of their worldwide location (13). However, the relative importance of genetic and environmental influences is difficult to establish from these data (see also 'twin studies').

Familial clustering: A positive family history of IBD is found in 13% to 36% of patients (14-19). The highest prevalence of family history in IBD is found in early onset cases. The Cleveland Clinic (a tertiary referral centre) reported a 29% family history for IBD in UC patients and 35% in CD patients when the disease was diagnosed before the age of 21 years. More than half of the affected relatives were first degree relatives (parents, children or siblings) (16). This study is complemented by data from Johns Hopkins, where a 30% prevalence of family history was found in CD patients younger than 20 years of age at diagnosis, double that of those over 20 years of age (20). Conversely, patients with a positive family history tend to be younger at the time of diagnosis than sporadic cases (14,21). Patients with early onset symptoms may be a group for whom there is a higher genetic contribution to disease risk.

Familial clustering can be measured by lambda s (λ s), the ratio of risk to siblings compared with the general population risk. Measures in IBD, including two recent studies from the United Kingdom, suggest that λ s for CD is 15 to 35, and for UC is 7 to 12 (19,22-24). By way of comparison, λ s is less than 3 for hypertension, 15 for type 1 diabetes and 450 for a monogenic condition such as cystic fibrosis. The level of familial relative risk is related to the ease of mapping susceptibility loci (25). A comparatively high λ s in IBD compared with that in other complex diseases may explain why it has been possible to replicate IBD susceptibility loci.

Clinical patterns of familial IBD: Consecutive CD patients were enrolled in a recent study of 554 individuals at Johns Hopkins University. Seventeen per cent had a relative affected with CD (12% for first degree relatives) and 5% had a relative affected with UC. Eighty-six per cent of families were concordant in at least two members for disease site and 82% of the 60 families analyzed were concordant for clinical type (classified as inflammatory, strictureing or fistulizing). These authors also reported that 18 of 22 monozygotic twins with CD reported in the literature were concordant for disease site (26). In Oxford, affected IBD sibling pairs were concordant for disease type in 82%, disease extent in 76% and extraintestinal manifestations in 84% (27). Concordance for CD location and type has also been shown by other centres (24) and increases with the number of relatives affected within the family (21). Investigators from St Marks Hospital showed that families with three or more affected first degree relatives were concordant for CD or UC status (28). These studies suggest that multiply-affected families segregate a specific clinical phenotype of IBD, rather than being at an overall increased risk of all types of IBD.

A younger age of onset in the affected offspring than in the affected parent was seen in a study of 54 IBD families (27). Other investigators (29) have shown that the younger member of second-generation first degree relative CD pairs also had disease onset at an earlier age. The tendency for successive generations to develop more severe and earlier onset disease, genetic anticipation, has been shown in some monogenic diseases. Although these data might provide preliminary evidence for genetic anticipation in CD, it is an unlikely explanation. The effect may be due to bias from length of follow-up, earlier diagnosis in the offspring of an affected parent or perhaps the inheritance of a greater genetic load (28).

Twin studies: Twin studies have been widely used in human genetics to study the relative contributions of genes and environment. Di- and monozygotic twins share half and nearly all their genes, respectively (excepting mitochondrial DNA and antibody/ T cell receptor repertoire). Unless separated at birth, they also share to a large extent the same environment, which is, therefore, controlled in epidemiological studies. In a Swedish study (30), 80 same-sex twin pairs (in whom at least one individual had IBD) were studied. In the UC group, one of 18 monozygotic and none of 20 dizygotic or unknown zygosity twins were concordant, with eight of 18 monozygotic and one of 26 dizygotic twins concordant for CD. Similar data were found in the British twin study (31), which identified twins from the national patient support group, and a Danish study (32), which used a national twin register. Combining these data gave overall UC mono- and dizygotic concordance rates of 7% and 2%, respectively, compared with rates of 33% and 4% for CD. The higher degree of heritability for CD than UC is consistent with the family study data. In addition, the scarcity of reports of
mixed CD/UC monozygotic twins fits with the high intrafamilial disease concordance (33).

Models of disease inheritance: Many studies have been carried out of models of inheritance of IBD. Complex segregation analysis has been used and has found evidence for a major dominant or additive disease gene in 10% of UC families (34, 35). Both of these studies suggested that the sporadic model was very unlikely. Studies in CD have suggested that between 7% and 30% of patients may be homozygous for a recessive disease gene (although this may also be explained by a multifactorial model) (35, 36). Multiple genes acting independently have also been proposed (37).

Although good evidence exists for concordance of disease type within families, CD and UC do occur within the same family. This is most strikingly shown by the presence of mixed CD and UC sibling pairs in linkage studies (38, 39). Thus CD and UC might be a group of oligogenic diseases sharing some susceptibility genes but differing at others (Figure 1). Several disease genes would act together to produce a genotype-dependent susceptibility risk. A positive family history confers a greater risk than smoking, diet, measles exposure or other environmental factors (40-42).

It is possible to construct a model incorporating both genetic and environmental risk factors. An environmental insult, such as a nonsteroidal anti-inflammatory drug or enteric pathogen, is the initial factor leading to impairment of the intestinal mucosal barrier. An inflammatory response is then triggered through exposure of the epithelium and lamina propria to luminal antigen. Nonsusceptible individuals then use the inflammatory response to clear antigen from the cellular compartment and repair the epithelial and mucosal defences. In susceptible individuals, impairment of the mucosal barrier, increased intestinal permeability, inappropriately increased inflammatory and immunological responses or failure to downregulate these responses may lead to chronic intestinal inflammation (43). Disruption of genes involved at any point in this pathway may lead to increased susceptibility to IBD. Understanding the complex interaction between environmental factors and genes encoding susceptibility or specific phenotype is one of the major challenges in IBD research.

LINKAGE STUDIES AND SUSCEPTIBILITY LOCI
The mode of inheritance of IBD is unknown, and classical parametric linkage analysis (using a dominant or recessive model), therefore, is unsuitable. Nonparametric methods (‘model free’) of linkage analysis have been developed using affected sibling or relative pairs, in order to overcome difficulties. Highly polymorphic microsatellite (simple sequence repeat) markers at average intervals of 0.4 cM throughout the genome (44) and the development of rapid semiautomated genotyping techniques have made these studies possible. Affected families are genotyped and analyzed to identify genetic markers where the degree of allele sharing between affected individuals exceeds that expected by chance (greater than 0.5 for sibling pairs). A marker close to a disease gene is less likely to be changed by meiotic recombination than one far away. If the same disease gene is involved in both affected siblings then the siblings will share the same parental allele at the nearby marker. A subchromosomal region can then be identified that is genetically linked (ie, co-inherited) with an as yet unknown disease susceptibility gene or genes. To adequately cover the whole genome, a set of about 300 markers are needed (10 cM intervals). Lander and Kruglyak (45) argued that, due to the statistical problems associated with the multiple tests involved, the threshold for a genome-wide screen should be one false positive in 20 genome scans (significant linkage, maximum logarithm10 of odds [lod] score 3.6 or higher for sibling pair studies, P<2×10⁻⁵). Confirmation of linkage, in a new dataset, would then require P<0.01. Suggestive linkage, likely to occur by chance once in every genome scan, would be defined by a lod score of 2.2 or greater for sibling pair studies, P<7×10⁻⁴). These criteria are conservative and may miss weaker effects, but the cautionary tale from early studies in psychiatric genetics (many nonreplicable linkage findings using weaker significance levels) must be remembered.

Two main susceptibility loci for IBD have been identified and replicated by several independent groups. These are on chromosome 16 (pericentromeric) and chromosome 12q13-14 (Table 1).

**Initial reports of strongest linkage on chromosomes 12 and 16:** The first genome screen using affected sibling pair methods was in CD patients, albeit with only 25 families (40 sibling pair equivalents) (46). This is a small number of families for a complex disease with genetic heterogeneity and is likely...
TABLE 1 Linkage studies in inflammatory bowel disease (IBD)

<table>
<thead>
<tr>
<th>Published study</th>
<th>Study design</th>
<th>Study size</th>
<th>Phenotype</th>
<th>Ethnicity</th>
<th>IBD1, chromosome 16 (maximum lod)</th>
<th>IBD2, chromosome 12q (maximum lod)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hugot et al (46) and Lesage et al (47)</td>
<td>Genome wide</td>
<td>112 asp equivalents</td>
<td>CD only</td>
<td>European, white</td>
<td>3.17 (CD, D16S409 to D16S419)*</td>
<td>Negative</td>
</tr>
<tr>
<td>Satsangi et al (38) and Parkes et al (143)</td>
<td>Genome wide</td>
<td>186 asp</td>
<td>81 asp CD, 64 asp UC</td>
<td>European, white</td>
<td>2.6 (singlepoint, CD, D16S411)</td>
<td>5.47 (singlepoint, IBD, D12S83)*</td>
</tr>
<tr>
<td>Ohmen et al (49), Yang et al (50)</td>
<td>Replication: 1,3,7,12,16</td>
<td>117 asp</td>
<td>75 asp CD, 23 asp UC</td>
<td>North American, 33% Jewish</td>
<td>2.41 (Non-jewish CD, D16S411)</td>
<td>NPL=1.9 (CD only, not specified)</td>
</tr>
<tr>
<td>Cavanaugh et al (51)</td>
<td>Replication: 16</td>
<td>Mainly 54 multiplex families</td>
<td>54 multiplex CD only, further 19 mixed IBD</td>
<td>Australian (80% European descent)</td>
<td>6.3 (CD, D16S409 to D16S53)</td>
<td>Negative</td>
</tr>
<tr>
<td>Mirza et al (53)</td>
<td>Replication: 1,16</td>
<td>70 kindreds</td>
<td>32 kindreds UC only, rest mixed IBD</td>
<td>Mostly Northern European</td>
<td>2.02 (UC only, D16S3120)</td>
<td>2.79 (IBD, D12S1724 to D12S590)</td>
</tr>
<tr>
<td>Duer et al (54)</td>
<td>Replication: 12</td>
<td>208 arp</td>
<td>80 arp CD, 22 arp UC</td>
<td>European Caucasoid descent, 22% Jewish</td>
<td>1.69 (CD only, D16S679)</td>
<td>Negative</td>
</tr>
<tr>
<td>Cho et al (60) and Brant et al (144)</td>
<td>Genome wide</td>
<td>297 arp</td>
<td>175 arp CD, 26 arp UC</td>
<td>American, 37% Jewish</td>
<td>71% white, 27% Jewish</td>
<td>Negative</td>
</tr>
<tr>
<td>Rioux et al (58)</td>
<td>Replication: 3,7,12,16</td>
<td>200 arp</td>
<td>114 arp CD, 36 arp UC</td>
<td>Mostly Northern European</td>
<td>1.7 (CD only, D16S407)</td>
<td>1.82 (CD only, D12S599)</td>
</tr>
<tr>
<td>Vermeire et al (59)</td>
<td>Replication: 3,7,12,16</td>
<td>79 asp</td>
<td>CD mainly</td>
<td>Belgian, white</td>
<td>0.9</td>
<td>2.08</td>
</tr>
<tr>
<td>Akolkar et al (57)</td>
<td>Replication: 12,16</td>
<td>68 arp</td>
<td>CD only</td>
<td>Most Jewish</td>
<td>1.71 (CD only, D16S407)</td>
<td>1.82 (CD only, D12S599)</td>
</tr>
<tr>
<td>Hampe et al (39) and Curran et al (134)</td>
<td>Genome wide</td>
<td>353 asp</td>
<td>162 asp CD, 114 asp UC</td>
<td>White Northern European</td>
<td>1.71 (CD only, D16S407)</td>
<td>1.82 (CD only, D12S599)</td>
</tr>
</tbody>
</table>

*Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.

Initial genome screen microsatellite markers for a putative susceptibility locus, meeting criteria for significant linkage (reference 45). Other suggestive linked loci are shown where the criteria for suggestive linkage are met (see text). Confirmitory studies for a locus where P<0.01, or described as negative if P>0.05. Multipoint logarithm of odds (lod) scores are given where possible. arp affected relative pairs; asp affected sibling pair; CD Crohn’s disease; UC Ulcerative colitis; NPL Nonparametric linkage score.
heterogeneity or that the size of the region may still be too great for detection by association study with infrequently spaced markers. A further study in Northern European patients (32 kindreds with UC only, and 38 mixed IBD kindreds) at the chromosome 16 locus reported linkage in the UC group only (53). However the P value of 0.02 obtained with D16S3120 does not quite meet the strict criteria for replication, and again the UC group was small.

Investigators from Pittsburgh (54) have also confirmed linkage to the chromosome 12 region in 127 CD- and 122 UC-affected relative pairs. Both CD and UC pairs contribute equally. The linkage method of ‘affected relative pair analysis’ used has more power when a trait allele is rare and less power when a trait allele is common compared with affected sibling pair analysis (35). The transmission disequilibrium test was positive at D12S83, but this result was not significant after statistical correction for the multiple comparisons made. These investigators have more recently published data replicating the region on chromosome 7 found initially in the Oxford genome scan (56).

Three further studies aimed at replicating the previously reported regions of linkage show differing results. A study from New York has also replicated the chromosome 12 locus in predominantly Jewish CD sibpairs (57). Canadian patients (114 CD and 36 UC sibling pairs) from 161 Toronto kindreds showed no linkage to the chromosome 12 or 16 loci or the initial Oxford chromosome 3 and 7 regions (58). These loci have also not been replicated in 76 CD sibpairs from 91 Belgian white families (59). One explanation for these apparently conflicting results is the ethnic and case mix in the groups studied. Genetic heterogeneity, where nonoverlapping sets of gene mutations all result in intestinal inflammation (Figure 1), is likely to be a further factor.

The third genome-wide screen in 174 American families with 175 CD-affected relative pairs and 26 UC relative pairs was performed in Chicago (60). Over one-third of families were of Ashkenazi Jewish descent. A denser set of 377 microsatellite markers allowed examination of regions not well covered in previous screens. Three novel susceptibility loci were identified - at chromosome 3q with a broad region of linkage in all families, at chromosome 1p with linkage in non-Ashkenazi families and at chromosome 4q with greatest linkage in the mixed CD and UC subset of families. The chromosome 16 linkage was replicated, with greatest linkage again in CD families and no linkage in UC families. This linkage proved greatest in the Ashkenazim. An epistatic interaction was observed between the chromosome 1 and 16 loci. No confirmation was provided for the chromosome 3p, 7 or 12 loci identified by the Oxford genome scan, although again the number of UC studied was much smaller than for CD. Interestingly, these investigators have found pedigrees in which an individual has a young age of onset of disease to demonstrate greater linkage to their chromosome 1p, 3q and the replicated chromosome 16 loci (61). This observation fits well with the family study data suggesting a greater genetic component to IBD susceptibility for this group.

In the most recent, and fourth genome-wide analysis in IBD, 162 CD- and 114 UC-affected sibling pairs were genotyped from 268 white northern European families (39). The chromosome 16 linkage was replicated for CD only, and the chromosome 12 locus was replicated in CD and IBD overall, but did not quite reach the significance criteria for sibling pairs with UC only. The two loci were predicted to be additive in effect rather than interact epistatically. The chromosome 4 linkage reported previously was confirmed (60). Novel susceptibility loci were reported on chromosomes 1, 6, 10, 22 and X, although again these did not meet Lander’s strict criteria for first genome scan data. The chromosome 1 locus found is 100 cM q telomeric to the previous reported chromosome 1 locus, and thus does not overlap (60). Linkage to the HLA region was shown in this study (see below). Interestingly, given the association between Turner’s syndrome and IBD (62-67), linkage to the X chromosome has been reported. The Belgian group have also found linkage to the X chromosome (68).

In summary, strong evidence for linkage exists for chromosomes 12 and 16 (Table 1). The chromosome 16 linkage has been demonstrated by six independent groups and appears to be most strongly linked to CD. Evidence for linkage to chromosome 12 has been provided by five groups. Chromosome 12 appears to be linked to both CD and UC. The limited number of UC sibling pairs in many of the studies may mean that sufficient power was not present to detect linkage to chromosome 12, although alternative explanations include ethnic and case mix differences and genetic heterogeneity. Several other putative susceptibility loci have been identified, but these await wider replication.

**METHODS UNDERLYING CANDIDATE GENE ASSOCIATION STUDIES**

Many new polymorphisms have been described in genes involved in the immune and inflammatory response over the past decade, and several candidates have been studied in IBD patients. Genes outside as well as within the replicated regions of linkage have been suggested from studies of expression pattern, function or animal models. Association studies are more powerful tools than linkage analysis in these studies, and because linkage disequilibrium operates over 1 cM at most, a positive association with a gene or marker may define the true susceptibility gene (69,70).

Case-control association studies compare allele frequencies in affected patients with that in a set of matched controls. Spurious associations can occur due to population stratification — where the study and control groups are drawn from genetically distinct subpopulations and unrecognized ethnic differences in allele and/or disease frequencies exist. Incorrect matching of cases and controls may also occur. To avoid these problems, family-based studies have been designed using genotypes from an affected individual and two parents. This allows the nontransmitted parental alleles to be used as internal well matched controls. One such method, the transmission disequilibrium test (TDT), has been widely used in complex genetics (71,72) but so far in only two studies in IBD. Future studies in IBD are likely to use the TDT.
TABLE 2
Positional candidate genes for inflammatory bowel disease (IBD) on chromosomes 12 and 16

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Gene</th>
<th>Name</th>
<th>Possible role in IBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>STAT6</td>
<td>Interleukin 4 transcription factor</td>
<td>Regulation of Th1/Th2 cells</td>
</tr>
<tr>
<td>12</td>
<td>ITGA5</td>
<td>Alpha 5 integrin</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td>12</td>
<td>ITGB7</td>
<td>Beta 7 integrin</td>
<td>Leucocyte adhesion</td>
</tr>
<tr>
<td>12</td>
<td>MMP18</td>
<td>Matrix metalloproteinase</td>
<td>Degrades extracellular matrix</td>
</tr>
<tr>
<td>12</td>
<td>IFNG</td>
<td>Interferon gamma</td>
<td>Th1 pathway</td>
</tr>
<tr>
<td>12</td>
<td>NRAMP2</td>
<td>Natural resistance associated macrophage protein 2</td>
<td>Intestinal iron transport</td>
</tr>
<tr>
<td>16</td>
<td>ITGAX</td>
<td>Leucocyte surface antigen (alpha-X integrin)</td>
<td>Leucocyte adhesion</td>
</tr>
<tr>
<td>16</td>
<td>IL4R</td>
<td>Interleukin 4 receptor</td>
<td>Inflammatory cell regulation</td>
</tr>
<tr>
<td>16</td>
<td>SPN</td>
<td>Sialophorin</td>
<td>T cell activation</td>
</tr>
<tr>
<td>16</td>
<td>CD19</td>
<td>B lymphocyte antigen CD19</td>
<td>B lymphocyte function</td>
</tr>
</tbody>
</table>

Th T helper

more widely as families are collected due to its power and robust nature.

A further problem of association studies is that there may well be no association to find, and so each test performed carries an independent risk of a false positive result (69). A Bonferroni correction has to be applied, and unfortunately P values are often still presented uncorrected. The best test of an association is replication by different investigators in a separate patient group.

POSITIONAL CANDIDATE GENES FOR THE WIDELY REPLICATED CHROMOSOME 12 AND 16 SUSCEPTIBILITY LOCI

Identification of genes within a region of linkage (positional candidates) has been made simpler by the publication of a physical map containing transcript markers for nearly half of all predicted human disease genes (73). Sequence tagged sites have been taken from cDNA sequences and ordered by radiation hybrid mapping, along with a framework of microsatellite markers. The accompanying website provides the map position information along with data on known gene function and homologies with other species (74). Promising positional candidates identified in this way for chromosomes 12 and 16 are shown in Table 2.

Interferon-gamma (IFN-γ) (chromosome 12q14) is a key T helper (Th)-1 cytokine, and abnormalities of this pathway are well described in animal models and in humans with IBD. An intragenic microsatellite marker was typed in 506 simplex families (ie, patients and parents) and 133 affected relative pairs. A nonsignificant lod score of 0.67 was obtained for CD, although this sample was less than the 353 sibling pairs used by these authors in their recent genome scan replicating the chromosome 12 linkage. The transmission disequilibrium test was negative for CD, UC and IBD overall (75). The genomic sequence of IFN-γ is only 6 Kb, and any mutation is highly likely to be in linkage disequilibrium with the polymorphic intragenic repeat. The authors propose that abnormalities related to IFN-γ observed in IBD are likely to be a secondary phenomenon.

Data have been reported in abstract form for two other candidates on chromosome 12. Vitamin D has multiple immune actions, including the suppression of lymphocyte proliferation and inhibition of cytokine production. The TaqI t allele and tt genotype of the vitamin D receptor gene (chromosome 12q13) were significantly associated with CD (76). This genotype has been reported to be of functional importance in several infectious diseases. Natural resistance-associated macrophage protein (NRAMP)2 (chromosome 12q13) is involved in intestinal iron transport and has homology to NRAMP1 (discussed earlier). Three restriction fragment length polymorphisms were assessed, and the gene excluded as a candidate by linkage and TDT analysis. No novel mutations were found by direct sequencing of the gene in 33 CD patients (77).

Positional candidates on chromosome 16 include the CD19 gene and CD43 (sialophorin) gene, involved in B lymphocyte function and intracellular adhesion molecule-1 (ICAM-1) interactions, respectively. Direct sequencing of the coding sequence, including intron-exon junctions, was carried out in 10 CD patients and two healthy controls. Two single nucleotide polymorphisms were identified in each gene, three of which were silent mutations and one a single base amino acid missense mutation in CD19 exon 3. The TDT was negative for the CD19 missense mutation in 78 CD families. Although rare alleles may have been missed, the CD19 and CD43 genes did not appear to be involved in the genetic predisposition to CD in these families (78).

HUMAN LEUKOCYTE ANTIGEN TYPE AND IBD

Among the first candidate genes to be studied in IBD were those of the human leukocyte antigen (HLA) system, involved in antigen processing and presentation. HLA class II molecules present partially digested antigen to the T cell receptor and play a central role in the immune response. HLA-DP, -DQ and -DR code for the alpha and beta chains of the class II molecule. HLA-DP and -DQ genes are polymorphic for both chains. HLA-DR is only polymorphic for the beta chain, and the beta1 chain is the most polymorphic of all. Initial studies based on serology were superseded by the increased number of polymorphisms detectable by DNA-based genotyping. The most consistent results have been seen in Japan, where the most genetically homogenous population is found. Conflicting results may be explained by genetic heterogeneity, small sample sizes, the lack of statistical correction for multiple testing and the limited number of HLA polymorphisms examined in some studies as well as by ethnic differences between populations.

UC: HLA-DRB1 and -DQB1 have been studied in white patients, with linkage and haplotype sharing observed in 29 affected UC sibling pairs (but not in CD sibling pairs) (79).
Allelic association was observed with the rare DRB1*103 allele (8.3% UC versus 2.1% controls) and DRB1*12. The DRB1*0301 DQB*0201 (DR3 DQ2) haplotype was reduced in females, particularly in those with distal UC. This haplotype was also predictive of extensive UC in both sexes in the 175 UC patients studied. HLA linkage had not been confirmed in earlier studies (80,81). In the much larger genome-wide scans in affected sibling pairs, only one study found significant linkage with IBD to the HLA region (39). However, a nonsignificant distortion toward allele sharing in UC was observed for the same marker (D6S276) in the Oxford genome scan (38). In contrast to affected sibling pair linkage methods, allelic association studies have more power to detect small effects once the precise region of interest has been identified (70). The increased frequency of the rare DRB1*0103 allele has been confirmed, again in extensive UC and also in patients undergoing colectomy (82). The more common DRB1*15 was also increased in these 98 UC patients investigated. DRB1*0103 was present in 14.1% of UC patients undergoing colectomy and ileoanal pouch anastomosis versus 3.2% in controls, and was shown to be predictive of extensive disease and extraintestinal manifestations in a separate series from Oxford (83). UC patients in California had a statistically significant positive association with the HLA-DR2 allele (84), but this has not been shown in several other studies in white patients (85,86). Many of these conflicting studies have relied on serological rather than molecular methods (87-91).

Japanese UC patients also have an increased frequency of DR2 (92,93-95). The molecularly defined DRB1*1502 allele of serological DR2 was present in 49.2% of 59 UC patients versus 17.6% in controls, and this genotype carried an increased need for steroid treatment (93). The DRB1*1502 association has also been observed in patients from Turkey (96). However, this allele is rare in non-Jewish European whites (5% versus 57% in Japan). Thus, HLA class II associations with UC clearly differ between populations and emphasize the necessity to define ethnicity when studying populations.

CD: Californian CD families have shown increased haplotype sharing in multiplex families (97). However, other studies that have found evidence for linkage of IBD to the HLA locus have found the effect to be greater for UC (38,39). One of the largest association studies of 344 white CD patients found a negative association with DRB1*03 (6.5% CD versus 13.8% controls) and increased DRB1*01,07 (98). The finding that DRB1*03 may be a resistance allele and the positive association with DRB1*07 have been replicated (99). The DRB3*0301/DRB1*1302 haplotype has also been associated with CD in white patients (100-102). A marked decrease in DRB1*03 was found in fistulizing CD (103). A positive DQB1*0501 association with CD has been found in two studies (84,98). Japanese studies have reported associations with DQB1*04 and DQA1*0102 as a resistance allele (104). However, other studies in CD have found no or conflicting disease associations.

Arthropathy: The peripheral arthropathy associated with IBD has been divided into pauciarticular large joint and bilateral symmetrical types (105). Molecular genetic studies of HLA in these patients have provided preliminary evidence for genes controlling IBD arthropathy. DR103 is significantly increased in IBD patients, with large joint type 1 arthropathy being found in 40% versus 3% controls (106). Associations with class I HLA-B27 and B35 were also noted for type I arthropathy. In contrast, B44 was increased in the small joint type 2 IBD arthropathy.

The major histocompatibility complex (MHC) plays a relatively minor role in IBD compared with that in autoimmune disease such as celiac disease and insulin-dependent diabetes, confirmed by the weak evidence for linkage compared to other susceptibility loci (see below). Replicated positive associations have been obtained with HLA DRB1*0103 in whites with extensive UC and DR2 (DRB1*1502) in Japanese UC patients. DRB1*03 is replicated as a resistance allele in white CD patients, and DR4 protects against UC. A recent meta-analysis of 27 class II HLA studies confirmed these findings (107). HLA associations are more consistently implicated in UC than CD. HLA genes may play a disease-modifying role, rather than influencing IBD susceptibility per se, as shown by the association studies analyzing disease extent, character and arthropathy. These associations may reflect direct T cell responses to class II presented peptides, a T cell-driven autoimmune effect triggered by an initial pathogen or even linkage disequilibrium between the HLA allele and a functional gene nearby.

**OTHER CANDIDATE GENES**

Many other genes have been studied in IBD on the basis of potential function alone. The interleukin (IL)1 receptor antagonist (IL-1RA) (chromosome 2q) is a potent anti-inflammatory protein, implicated in intestinal inflammation from animal studies. An initial case-control study in UC and CD found an association in UC patients with allele 2 of an intragenic variable number of tandem repeats (VNTR) (35% versus 24% controls) (108). In the Oxford study (109), no significant differences among CD patients, UC patients and controls were initially seen. Analysis of subgroups in UC showed allele 2 to be increased in antineutrophil cytoplasmic antibody (ANCA)-positive UC patients, and more frequent in extensive disease than distal colitis (83). IL-1 receptor antagonist genes were not associated with UC overall in a Dutch study, but allele 2 of IL-1RA was increased in extensive disease (82). This did not retain significance after correction for multiple comparisons. The IL-1 and IL-1RA genes are in linkage disequilibrium, and while no association with UC overall has been shown by some groups (110), dissection of the interplay between genes of the IL-1 system has found positive associations (111). Studies of key cytokine genes involved in inflammation must also be able to explain the intestinal specificity of inflammation (or the lack of inflammation in other tissues) because it is possible that some abnormalities observed in cytokine levels are a secondary rather than an initiating step.

ICAM-1 (chromosome 19p13) is involved in neutrophil...
adhesion, and two exonic polymorphisms have been tested in UC and CD patients. No association was found for CD or UC overall. After stratification by ANCA status, ANCA-positive CD and ANCA-negative UC had an increased frequency of allele R241 (112). Codon 241 is in a functionally important domain of ICAM-1, and this interesting association awaits replication by other groups. NRAMP1 (chromosome 2q35), involved in macrophage regulation, has been implicated in susceptibility to tuberculosis and other infections (113). Two close polymorphic microsatellite markers were examined for associations in CD and UC patients. Two haplotypes found in CD patients were significantly different from controls (114). Again this merits replication and screening of the gene for possible mutations. Beta1 and beta2 kinin receptor genes (chromosome 1q43) have also been studied by one group because of evidence for kinin-mediated gastrointestinal inflammation (115). A polymorphism in the beta1 receptor promoter region was significantly different between IBD patients and healthy controls. This too awaits confirmatory studies by others.

The gene encoding tumour necrosis factor alpha (TNF) (chromosome 6p21) has been studied because of well described alterations of TNF production in IBD, and because of the efficacy of anti-TNF monoclonal antibodies in CD (116). Many TNF polymorphisms have been described, of which the –308 and –238 promoter mutations have been suggested to be functional (117). As for the IL-1RA polymorphisms, data from several published studies and analyses of subgroups have been discordant (108,109,118). Investigators in Los Angeles, California found a TNF microsatellite haplotype (a2b1c2d4e1) in 24% of CD patients versus 7% of controls, although a French group did not (119,120). This haplotype is in linkage disequilibrium with HLA-DR and DQ alleles, and dissection of the pathological mutation (versus polymorphism) in this region of chromosome 6 is needed. Of interest, linkage of CD to the MHC region has also been reported by these Californian investigators. Interestingly, one study (103) has reported a low frequency of HLA-DRB1*03 and a normal frequency of TNF -308 in a subgroup of CD patients with perianal fistulas. Recombination events in this subgroup of patients were proposed by the authors to play a role.

TAP1 and TAP2 genes (chromosome 6p21) are involved in the transfer of antigenic peptides before their association with class I HLA molecules. No difference in TAP1 and TAP2 alleles was found among UC patients, CD patients and controls (121). A significant (corrected) decrease in TAP2 AA allele frequency was found in CD patients who did not respond to steroid therapy. MHC class I-like gene-A (MIC-A) is an atypical MHC class I molecule, found in the same genomic region, expressed on the intestinal epithelial surface. It is thought to interact with gamma-delta T cells and be upregulated by cellular stress. MIC-A allele*007 was significantly associated with UC compared with controls (32% versus 11%), and this was confirmed in a second independent cohort by these investigators (122).

Chemokines are released by intestinal epithelium, playing a major role in recruitment and activation of inflammatory cells. The chemokine receptor CCR-5 (chromosome 3p21) has a 32 base pair deletion found to confer resistance to human immunodeficiency virus, and is close to a linked marker D3S1573 (38). Forty-four per cent of CD patients and 31% of UC patients versus 13% of controls possessed this functional deletion (123).

Abnormalities of intestinal mucin expression have been observed in IBD patients and the cotton-top tamarin model of colitis. The MUC3 gene (chromosome 7q22) lies close to a microsatellite D7S669, linked to IBD with a lod score of 3.08 in the Oxford genome scan (38). Polymorphisms of VNTRs within the MUC3 gene in 75 Japanese patients with UC and 168 Japanese controls were analyzed. When the frequency of patients carrying one or two rare VNTR alleles was compared with that of controls, a significant increase was found in the Japanese patients (odds ratio 2.72). Similar results were found in white patients. Rare alleles of MUC3 may confer predisposition to UC, although these were found in only 22% of UC patients studied (124).

GENETIC INSIGHTS FROM ANIMAL MODELS

Lessons have been gained from spontaneous, chemically-induced and genetically engineered animal models of IBD. Many studies have used inbred animal strains, and in these models development of intestinal inflammation is strain dependent. In the dextran sulphate sodium (DSS) model, for instance, C3H/HeJ/Bir mice are highly susceptible and display spontaneous colitis, while the C57/black 6 strain is resistant (125). Similar findings apply in gene-targeted models of IBD, in addition to chemically induced colitis and are furthermore seen in rat models.

Intestinal inflammation has been observed in mice with diverse targeted gene alterations including alpha-beta T cell receptor, keratin 8, TNF, transforming growth factor-beta and IL-2, -7 and -10 genes (126-130). In another model, mice with a targeted deletion of the G protein G(i2 develop IBD resembling UC, and further develop adenocarcinoma of the colon (131). Interestingly, the human homologue of the Gai2 gene (chromosome 3p21), which encodes an inhibitory G protein expressed in many tissues, is located within a region of linkage identified in one IBD genome scan (38). The presence of gut bacterial flora has been shown to be a key element in the development of inflammation in animal models. IL-2 and IL-10 knockout mice develop IBD but not under germ-free conditions (132). These various genetically engineered models develop colitis predominantly via a Th1 T cell-mediated pathway, and less commonly via a Th2 pathway. That diverse immunological defects all lead to IBD via one of two final common pathways suggests that many candidate genes might be involved in human IBD.

A genome-wide screen for quantitative trait loci in the DSS mouse model was carried out using both susceptible and resistant strains (125). Two of the loci were further evaluated in congenic strains. Susceptibility loci were found on mouse chromosomes 2 and 5. Reduced susceptibility was observed in congenic mice carrying alleles from chromosomes...
genes, IL-10 receptor and IFN- 
congenic chromosome 9 interval include CD3 subunit 
that multiple genes control susceptibility to DSS colitis in 
mice. Candidate genes with human homologues within the 
congenic chromosome 9 interval include CD3 subunit 
genomes, IL-10 receptor and IFN- 
2 and 9 from resistant mouse strains. The authors concluded 
that multiple genes control susceptibility to DSS colitis in mice. Candidate genes with human homologues within the 
congenic chromosome 9 interval include CD3 subunit 
genomes, IL-10 receptor and IFN- 
Animal models are thus proved useful in studying specific 
targeted gene mutations and susceptibility loci for induced 
disease as well as the more traditional physiological and drug 
treatment studies. Insights into immunological mechanisms of 
human IBD, as well as potential human candidate genes 
have been gained. Studies of gene function in animal 
induced mutants will follow from identification and character- 
ization of candidate genes in human IBD.

OVERVIEW AND DIRECTIONS 
FOR FUTURE WORK

IBD genetics have advanced rapidly in the past five years. Earlier data were limited to epidemiological and serological 
HLA association studies. Molecular biological techniques have progressed through genome-wide linkage studies to 
widespread replication of regions of linkage and the analysis of positional candidate genes. In particular, IBD is unique 
among complex genetic disease in having two non-HLA sus- 
ceptibility loci reproduced by several investigators using dif- 
f erent patient sets.

The relatively small role of HLA genotypes in IBD sus- 
ceptibility is becoming clearer. HLA DRB1*0103 and DR2 
(DRB1*1502) are involved in UC susceptibility in white 
and Japanese populations, respectively, while DRB1*03 and 
DR4 appear to be resistance alleles for CD and UC. Studies 
in arthropathy and of different clinical types of IBD suggest 
that HLA may have a greater role in modifying IBD pheno- 
type, than on overall disease susceptibility.

The two regions of linkage on chromosomes 12 and 16 
both span over 20 million base pairs, regions that may each 
contain several hundred genes. Identification of susceptibil- 
ity genes is thus a formidable prospect, and strategies pro- 
poused include analysis of subgroups, fine mapping, meta- 
analysis, and a systematic analysis of all candidates (Figure 2). Fine mapping of these loci using a dense set of microsatel- 
ite markers is being attempted, with some promising initial 
results (49). In contrast to simple Mendelian traits where the 
gene always lies within the maximal region of linkage, in 
complex diseases the gene may well lie outside the maximal 
region (133). In support of this argument, the positive studies 
for chromosome 12 found three peaks of linkage, each ap- 
proximately 10 cM apart. It is also possible that more than 
one susceptibility gene may reside within the region (50,54,134). The power of linkage studies to narrow a region is 
limited. Indeed, over 700 sibling pairs may be needed to 
narrow down to 1 cM a locus causing a twofold increased risk 
(135). An international IBD genetics consortium has been 
set up to carry out a meta-analysis of linkage data for this rea- 
son, and should report later in the year (136). Genetic het- 

erogeneity within CD and UC contributes to these 
difficulties. Analysis of specific subgroups, eg, early age of on- 
set or large families, may allow identification of patients with 
a stronger genetic susceptibility.

New developments in association studies, such as the 
TDT, provide more statistically robust methods of testing 
than the traditional case-control study. Two studies have 
suggested linkage disequilibrium for a microsatellite marker 
D12S83 on chromosome 12 (48,54). However recent statis- 
tical modelling suggests that useful levels of linkage disequi- 
librium in an outbred population may extend over very small 
regions (3 kB on average, although with wide regional varia- 
tion) (47). Microsatellite markers may be unable to provide 
sufficient resolution, being too widely spaced, and other 
markers such as single nucleotide polymorphisms (SNPs) 
may hold more promise. An alternative strategy is to identify 
all the potential genes within the large regions of linkage and 
apply methods for detection of novel mutations. The more 
likely candidates may be identified on the basis of known or 
predicted function and tissue specific expression. Systematic 
analysis of positional candidate genes, starting with the most 
plausible biological candidates, is likely to progress in paral- 
lel with attempts to narrow the regions of linkage.

Single nucleotide polymorphisms are mainly diallelic 
polymorphisms that occur on average every 1 Kb throughout 
the genome. Although they are less informative than micro- 
satellites they are more frequent and mutually more sta-
low investigation of potential environmental trigger factors. Clinical practice may include stratification of patients by genotype and potential response to drug therapy. For example, susceptibility to the myelosuppressive effects of azathioprine is now partly understood at the molecular level (142). Prognostication using genotype may be possible, and initial data for HLA type in IBD-associated arthropathy and HLA DR103 in prediction of need for surgery in UC are promising. Genetic counselling, both of individual patients and their relatives, may eventually become possible. Tremendous progress has been made in IBD genetics over the past five years, and there is great optimism that further advances will follow rapidly.

ACKNOWLEDGEMENTS: David van Heel is a Medical Research Council Clinical Training Fellow, and Jack Satsangi a Medical Research Council Clinician Scientist.


139. To affinity ...and beyond! Nat Genet 1996;14:367-70.


ERRATA


On page 208, under the heading “Twin studies”, the sentence, “In the UC group, one of 18 monozygotic and none of 20 dizygotic or unknown twins were concordant...” should have read “…one of 16 monozygotic and none of 20 dizygotic or...”. Later in the same paragraph, the sentence “Combining these data gave overall UC mono- and dizygotic concordance rates of 7% and 2%, respectively…” should have read “…concordance rates of 13% and 2%, respectively.

The errors are regretted.
Submit your manuscripts at
http://www.hindawi.com