Strong epidemiological evidence for a genetic contribution to pathogenesis in inflammatory bowel disease (IBD) has stimulated efforts to identify susceptibility genes for both of its major clinical forms, Crohn’s disease and ulcerative colitis. Genome scans for linkage have indicated multiple regions of interest, but replication of these has been limited. The detection of linkage on chromosome 16 (IBD1) led to the unequivocal identification of the NOD2 gene (now called CARD15) as a susceptibility gene for Crohn’s disease. This seminal discovery has provided proof of principle for positional cloning and candidate gene approaches to identify IBD genes. It has also led to useful strategic insights in complex disease genetics, and generated new directions in the investigation of the molecular pathways to pathogenesis. Linkage and association studies have also provided strong support for IBD susceptibility genes on chromosomes 5q31, 6p21 and 19p, while loci of interest at 3p, 3q and 14q require further follow-up. Although important obstacles to further progress will need to be overcome, the successes of the past 2 years suggest that a detailed description of the genetic basis of inflammatory bowel disease is a realistic goal.

INTRODUCTION

Inflammatory bowel disease (IBD) is a disorder characterized by chronic inflammation of the gastrointestinal tract. There are two clinical subtypes, Crohn’s disease (CD) and ulcerative colitis (UC). CD can affect any part of the intestine, and is associated with discontinuous, transmural lesions of the gut wall, whereas in UC inflammation is confined to the colon and rectum, and lesions are continuous and superficial. The annual incidence of UC and CD in Western countries is about 10 and 6 per 100 000, respectively (1). Treatments for IBD include anti-inflammatory drugs, targeted immunosuppression with agents such as anti-tumour necrosis factor α (TNF-α), and surgery (2). The molecular basis of pathogenesis of IBD is not yet clear, but contributory factors may include persistent bacterial infection, a defective mucosal barrier, and an imbalance in the regulation of the intestinal immune response (3). Epidemiological and genetic research has provided firm evidence for the existence of genetic determinants of susceptibility to IBD, and raised expectations that the identification of IBD susceptibility genes may lead to a clearer understanding of pathogenesis, and ultimately better treatment. This review will focus on some of the new developments in the genetics of IBD. Additional information on the genetic and immunological basis of IBD can be found in other recent reviews (4,5).

GENETIC EPIDEMIOLOGY

The sibling recurrence risks ($\lambda_s$) for IBD are in the range of 15–35 for CD (6–10) and 6–9 for UC (9,10). The wide range of estimates is likely to be due to differences in study design. Some studies did not distinguish between CD and UC in either probands or relatives, others included all first degree relatives ($\lambda_r$, not $\lambda_s$), cases were ascertained from various sources, and all estimates are sensitive to the increasing incidence rates, particularly for CD. A genetic contribution to CD is confirmed by twin studies, and a recent Danish study found concordance rates in monozygotic (MZ) and dizygotic (DZ) twins of 58 and 0% for CD, and 18 and 5% for UC (11). Twin and family studies both support a stronger genetic contribution to CD than UC. The increased risk of UC in relatives of CD cases, and vice versa is widely accepted, and supports the existence of common genes for both phenotypes. However, few published studies exist to support this hypothesis: Orholm et al. (10) showed significantly increased prevalence of UC in relatives of CD cases, but not in reverse, and lower rates were found by Meucci et al. (9).

LINKAGE STUDIES

Linkage studies of IBD have focused on affected sibling pairs or small families, using non-parametric methods of linkage...
analysis for CD, UC and IBD. Ten genome-wide linkage searches have been published since 1996 (12–21). These results are summarized in Figure 1, which shows regions of suggestive linkage (LOD score > 2.2, or P-value < 7.4 × 10^{-5}) and significant linkage (LOD score > 3.6 or P-value < 2.2 × 10^{-5}) (22). For each genome-wide scan, LOD scores or P-values were extracted from multipoint analysis, where performed, for IBD, CD or UC.

As in most complex diseases, the results from genome-wide scans have been disappointing: only five significant LOD scores have been achieved, with little replication across the scans. Evidence for linkage in the pericentromeric region of chromosome 16 was identified in the first IBD scan (12). This was named IBD1, and was later shown to contain the CARD15 susceptibility gene for CD (see below). Other scans confirmed this (P < 0.01) (22), although, in general, evidence for linkage in genome scans was weak. Evidence for this region has been stronger in replication studies, and a consortium study obtained an LOD score of 5.8 to this region (23). Linkage to chromosome 5q (IBD5) has been confirmed by association studies (see below), although linkage was detected in only one scan, stratified by age of diagnosis (17), and in a small scan of Jewish families (16). The significant linkage to chromosome 19p (17) was confirmed (P < 0.01) by two other studies (14,18), and occurs primarily in families who lack CARD15 mutations, but carry the high risk IBD5 haplotype (21). Linkage to chromosome 14q has been detected in two studies (16,18). Of the significant linkage regions, only the chromosome 12 result is in doubt, with most evidence for linkage from UC families; the LOD score was 3.7 in 138 UC families from the US/UK (24), and 1.2 in the consortium data set of 89 UC families (23).

Nine further regions with suggestive linkage were identified, with little replication across studies (Fig. 1). Notably, the linkage to 1p36 was also found in four American Iraqi Chaldean families with mainly UC (LOD score = 3.0) (25). Linkage to the major histocompatibility complex (MHC) region of chromosome 6p (IBD3) is well-established in both CD and UC (26–28). The linkage to chromosome 3p is supported by candidate region studies (29,30), although results cover a broad region of over 50 cM. The linkage to chromosome 3q (14) has been confirmed in the recent Oxford genome scan (LOD = 2.1) (21).

The lack of consistency of genome-wide linkage results may be due to low power, since studies had an average of 100 CD affected relative pairs (range 19–175), and even fewer UC pairs (six studies, mean 29.5, range 20–114). Further progress in linkage analysis may be achieved through meta-analysis and larger or pooled studies, particularly to identify UC-specific genes. Two meta-analyses of linkage scans have been performed using the Genome Scan Meta-analysis method, which ranks the LOD scores obtained in each study within 30 cM bins, then sums the ranks to provide an overall summary statistic for linkage within each bin (31). An analysis of six scans supported linkage on chromosomes 5q, 6p and 16cen (19). A more extensive meta-analysis of the 10 genome-wide linkage studies from Figure 1 showed evidence of linkage to 2q, 3q, 5q, 6p, 7q and 16 for IBD, and also 17q and 19p for CD (32). Stratification of linkage by genetic or clinical features may also increase power. For example, several studies of linkage in CD families that do not carry the three major CARD15 mutations show linkage proximal or distal to CARD15 (21,33–35); linkage to the HLA region occurs mainly in families carrying CARD15 mutations (35), and mainly in brother–brother pairs with CD (36). However, these analyses are susceptible to multiple testing of small data sets and replication is required.

In summary, evidence for linkage to the regions of 16cen ([IBD1]), 5q31 ([IBD5]), 6p21 ([IBD3]) and 19p is particularly strong. Other regions with confirmed linkage, and worthy of follow-up studies, are 3p, 3q and 14q. Further regions highlighted by individual scans or the meta-analysis may be strengthened by larger studies, mega-analysis of current genotype data from multiple groups or denser genotyping in these regions.

THE CARD15/NOD2 GENE (IBD1)

Discovery of NOD2

The well-replicated linkage of Crohn’s disease to chromosome 16 stimulated efforts to identify the relevant susceptibility gene. The two main strategies employed, positional cloning and mutation screening of functional candidate genes in the region of linkage, both led to the identification of mutations in the NOD2 gene. Hugot et al. (33) detected a modest association of two alleles of a microsatellite polymorphism, D16S3136, with CD. They then sequenced the bacterial artificial chromosome containing D16S3136, and identified coding sequences of a gene which contained multiple single-nucleotide polymorphisms (SNPs), many of which were strongly associated with CD. This gene, originally named NOD2, was cloned independently by Ogura et al. (37) as a homologue of NOD1 and shown to be involved in the activation of nuclear factor-κB (NF-κB). Since NOD2 mapped within the region of strongest linkage to IBD, this group screened it as a candidate for CD and detected association of a frameshift mutation 3020insC (1007fs) with CD (38). Our British/German group also screened NOD2 for mutations, and detected the association of 1007fs with CD (39). The strength of these independent replications, coupled with the predicted structure and function of the encoded protein (see below), provided firm evidence that NOD2 was IBD1. The presence of two caspase recruitment domains in NOD2 led to a change in nomenclature from NOD2 to CARD15.

Population genetics of NOD2/CARD15

The work of Hugot et al. (33) and subsequent studies (40–43) showed that three rare SNPs (R702W, G908R and 1007fs) in CARD15 were independently associated with CD (haplotypes B2, B3 and B4 in Fig. 2). These disease susceptibility alleles (DSAs) all arose on the 268S allele of a common SNP (P268S) that shows strong association with CD as a single marker, but the 268S haplotype that lacks any of those three DSAs is not associated with CD (haplotype B1 in Fig. 2) (33,41). The combined allele frequency of the three DSAs is about 18–20% in European CD cases and 6–7% in controls. At least one CARD15 DSA is present in 30–40% of northern European CD patients, and in about 14% of controls (41,44). Thus mutations in CARD15 are neither necessary nor sufficient for the development of CD. A gene dosage effect was evident, with
most studies showing an odds ratio of 2–4 in heterozygotes and 20–40 in homozygotes or compound heterozygotes for these three mutations. A meta-analysis of the three CARD15 mutations in 29 studies (J.P. Ioannidis, personal communication) found a risk of 2.7-fold for CD in non-Jewish Caucasian heterozygotes (95% CI 2.3–3.3) and 20-fold in homozygotes (95% CI 12–35), with a population attributable risk (PAR) for CD of 26.3%. Risks were lower in Jewish individuals and the PAR was only 12.7%. Multiple other rare non-synonymous mutations were detected in CARD15 (33,44), but their status as DSAs is unresolved. The fact that, collectively, they have been observed in 13% of CD cases but in only 6% of controls (44), and that some of them show an impaired response in functional assays (45), suggests that a subset of them are genuine DSAs. At least one rare mutation that does not show altered activity in the current functional assays for CARD15 (R703C) is strongly associated with CD (46) (K. King et al., submitted for publication), which suggests that such assays may not assess all aspects of CARD15 function (see below). Large differences in the contribution of CARD15 mutations to the development of CD in different populations are apparent: the three DSAs are absent in CD patients from Japan and Korea (47,48) and no other CARD15 SNPs were associated with CD in Korean patients (48). Thus CARD15 does not appear to contribute to the risk of CD in these populations. Interestingly, a 268S haplotype that lacks any of the three main DSAs but contains an additional variant (IVS8 + 158; JW1) is associated with CD in the Ashkenazi Jewish population but not in non-Jews, which suggests the presence of an additional DSA on this haplotype in Jews (49).

Genotype–phenotype correlations

Numerous studies have analysed the effect of the CARD15 genotype on the phenotype of IBD. Neither the known DSAs nor rare variants of CARD15 are associated with ulcerative colitis, the other major clinical form of IBD (33,39). Six studies, each of more than 200 CD patients, have shown that the presence of CARD15 mutations is associated with the ileal form of CD (40–44,50). Ileal involvement is associated with an earlier presentation of CD, and two of these studies also found an association between mutation of both CARD15 alleles and earlier onset CD (40,44) or stenosis (44,50). However, analysis of disease behaviour is complicated by variation in clinical classification schemes and by changes in disease behaviour with time (4). The reason for the association of CARD15 mutations with ileal disease is not known, but two recent studies have shown that CARD15 is strongly expressed in the Paneth cells of...
the terminal ileum (51,52). It is possible that partial or complete lack of CARD15 function may lead to loss of the ability of Paneth cells to respond to bacterial components. There is as yet no evidence that CARD15 genotype influences the response of patients to any of the major therapies for IBD, such as anti-TNF-α (infliximab) (53,54).

**CARD15 function**

The predicted structure of the CARD15 protein (Fig. 2) offers some clues to its function. It contains two caspase recruitment domains (CARDs), a central nucleotide binding oligomerization domain and 10 carboxy-terminal leucine-rich repeats (LRRs). LRR domains in other proteins such as disease-resistant R proteins in plants and the Toll-like receptor family recognize pathogen-associated molecular patterns and lead to induction of the innate immune response. Current evidence suggests that CARD15 interacts with a serine threonine kinase, RICK (or RIP2), which leads to the activation of NF-κB (reviewed in 5). NF-κB is a major transcriptional regulator of pro-inflammatory cytokines, including TNF-α, that are involved in intestinal inflammation. Its activation by CARD15 in vitro is stimulated by bacterial components, particularly muramyl dipeptide (55,56), and may be mediated via the LRR domain. The 1007fs mutation, which results in loss of the C-terminal LRR, is associated with reduced activation of NF-κB in vitro (38). The two other common CD-associated mutations also show reduced NF-κB activity, but to a lesser extent (45). The G908R mutation is located in the LRR domain, but R702W lies upstream of this (Fig. 2). The mechanism whereby loss of CARD15 activity leads to persistent inflammation of the intestinal epithelium is not yet known (see 5,57 for a detailed discussion). Mice lacking Nod2 showed no signs of intestinal pathology, and their macrophages responded normally to Toll-like receptor agonists (58). Surprisingly, the Nod2−/− mice had an enhanced survival to systemically induced endotoxic shock. These data will fuel the debate as to whether the LRRs in CARD15 are involved directly in pathogen recognition (59). Other murine models of colitis have been reviewed recently (5).

**ASSOCIATION STUDIES AT OTHER IBD LOCI**

**IBD3 (6p21)**

The strong evidence of linkage of markers at the IBD3 locus on chromosome 6p21 to both clinical forms of IBD suggests that major susceptibility genes for CD and UC are located in this region. The evidence for linkage is strongest in the vicinity of the MHC, which, given its central role in the immune response, has long been considered an excellent candidate locus for IBD. Many association studies of the MHC region in IBD have been reported over the past 30 years, but a consistent picture has been slow to emerge (60). Recently, independent association of DRB1*0701 and Cw*0802 with CD was reported, while the classic autoimmune haplotype A1-B8-DR3 was associated with colonic disease (40). HLA DRB1*0103 was found to be associated with both the colonic form of CD and with UC (61). The −857C SNP in the TNFA gene has been reported to be associated with UC and CARD15-negative CD in both family-based and case–control studies (62).
**IBD5 (5q31)**

The linkage on chromosome 5q31–33 (IBD5) to CD in Canadian families led to a systematic search for association of polymorphisms at this locus with CD (63). This identified a common risk haplotype over a region of 250 kb which contained multiple SNPs that were in almost complete linkage disequilibrium (LD) with each other, and were strongly associated with CD. SNP screening of the known genes in the region in a small number of CD patients and association analysis did not identify a disease-causing mutation, although the locus contains a cytokine gene cluster. The association of IBD5 with CD was replicated in large Northern European cohorts, but it was not associated with UC (64,65). The frequency of the risk haplotype was higher in CD patients with one or two CARD15 mutations than in patients with no CARD15 mutations or in controls (64), which suggests epistatic interaction between IBD5 and CARD15. Epistasis between CARD15 and IBD5 has also been detected in UC (66,67). There is also evidence of association of IBD5 with early onset CD (64) and with perianal CD (68). Relative risks of 1.5–2.0 are conferred by this locus in homozygotes for the risk haplotype (64,67). The extensive LD across the region has complicated efforts to find the causal mutations and identify the susceptibility gene; analysis of this association in non-European populations may prove useful if they show more limited LD.

**Other candidate genes**

The multidrug resistance gene, MDR1, encodes the membrane transport protein P-glycoprotein-1. It has been proposed as a candidate gene for IBD on the basis of its location within the region of linkage detected on chromosome 7q (13), its possible function as a barrier to xenobiotics in intestinal epithelial cells and the development of colitis in mdr1a knockout mice (reviewed in 69). The silent SNP C3435T in MDR1 is associated with lower intestinal P-glycoprotein expression, and the T allele has been reported to be associated with UC but not with CD (69). This finding was not replicated in two other studies (70,71). However, association between the alanine allele of the Ala893Ser/Thr tri-allelic SNP and IBD has been reported in both case-control and family-based tests of association (71). Replication of this finding in an adequately powered study is a priority. The NOD1/CARD4 gene has been investigated for association with IBD, based on its structural and functional similarity to NOD2/CARD15, but no association with CD, UC or IBD was detected (72). There are numerous other reports of association of SNPs in a wide variety of candidate genes with IBD or one of its sub-phenotypes. Many of these studies involve small numbers of cases, and have not been replicated.

**PAST LESSONS, FUTURE PROSPECTS**

**How many IBD loci?**

The data from replication studies, meta-analyses of genome scans and association studies provide strong collective support for the existence of at least four IBD loci, on chromosomes 5q31, 6p21, 16 and 19p. Further genome scans, particularly of UC families, may detect novel loci, provided that they are of adequate power. Thus far, both of the loci for which strong association has been detected (CARD15, 5q31) primarily increase the risk of CD. This is consistent with the epidemiological studies, which suggest a stronger genetic contribution to CD than UC. This may allow easier detection of genes conferring susceptibility to CD. A mega-analysis of genotype data from existing scans, with stratification by CARD15 or other loci, has the potential to detect new loci or to confirm others that currently have only modest support. It is also likely that additional loci with small effect sizes exist which may not be detectable by linkage analysis (73), but which may emerge from candidate gene or genome-wide association studies.

**Associations—true or false?**

Many of the published associations of genetic variants with IBD have not been replicated in follow-up studies. In part this may be attributed to a lack of adherence to good-practice guidelines for association studies. These include analysis of an adequate sample size to detect a small effect, independent replication of a primary finding in a second sample, and a check for Hardy–Weinberg equilibrium of genotypes in the control sample. There should also be some correction for multiple testing in subgroup analyses such as testing alleles, genotypes or haplotypes, or testing in various phenotypic categories. The field would be well served by attention to the principles of sound design, analysis and interpretation of association studies (74). Conversely, it is often difficult to publish negative studies such as lack of replication of a reported association or a lack of association with SNPs in strong candidate genes. A possible solution to this would be for specialist journals in the IBD field to publish abstracts of competently conducted negative studies, with an electronic link to the full article.

**The NOD2/CARD15 experience**

The identification of CARD15 as a susceptibility gene for CD was an important breakthrough not only for the insights that it will provide in time into the pathogenesis of IBD, but also as a proof of principle for positional cloning and candidate gene analysis approaches to complex disease genetics. A degree of good fortune was involved, since all three major disease susceptibility alleles, although individually rare (frequency <4%), arose on the same more ancient haplotype. Thus all common SNPs on that haplotype showed strong association with Crohn’s disease. Also, the degree of risk associated with the variants was sufficiently large to detect linkage in sib pair studies of adequate power, and to detect association in modest sample sizes. The extent of LD around the CARD15 gene, and the association of one of three common haplotypes with disease (42) suggest that a systematic association study based on LD maps and haplotype-tagging SNPs (75,76) might also have been successful in identifying CARD15.

The advances in IBD genetics over the past 2 years have provided considerable encouragement that the challenges posed by complex disease genetics are not intractable, and that
additional susceptibility genes will be found. The extensive linkage disequilibrium at two of the well-established IBD loci (5q31 and 6p21) may complicate efforts to identify the causal variants and susceptibility genes, and other difficulties such as small effect sizes and genetic and/or allelic heterogeneity may await us at these and other loci. Nonetheless, substantial progress has been made, and this, combined with global initiatives to identify and map genetic variation in the human genome (77), gives grounds for optimism that further significant progress in the genetics of IBD is likely within the next few years.

ACKNOWLEDGEMENTS

Our work is supported by the Wellcome Trust, the Charitable Foundation of Guy’s and St Thomas’ Hospitals, the Engineering and Physical Sciences Research Council, and the European Union Fifth Framework Programme. We apologize to colleagues whose work could not be cited due to space constraints.

REFERENCES


