Genetic markers and the classification of Inflammatory Bowel Disease

G. Bouma, A.S. Peña

SUMMARY
Crohn’s disease and ulcerative colitis are common causes of gastrointestinal morbidity in the Western World. The exact aetiology is unknown, yet, it is generally accepted that disease occurs as the consequence of dysregulated immune response to one or more exogenous factors in a genetically predisposed host. Epidemiological, family, association, and human genome screening studies have provided overwhelming evidence for a genetic susceptibility to IBD. These studies clearly indicate that IBD is not inherited as a simple Mendelian disorder, but more likely as a complex genetic trait, including gene-gene and gene-environment interactions. Recently, significant progress has been accomplished in understanding the immunological mechanisms that mediate chronic intestinal inflammation, as well as in identifying the genetic abnormalities that underlie these pathologic inflammatory responses. In the current review, we will focus on recent developments in unravelling the genetic basis of IBD and we will discuss the potential clinical applications of these findings.

INTRODUCTION
The chronic inflammatory bowel diseases (IBD), Crohn’s disease (CD) and ulcerative colitis (UC) are common causes of gastrointestinal morbidity in the Western World. The exact aetiology is unknown, yet, it is generally accepted that disease occurs as the consequence of an exaggerated immune response to one or more exogenous factors in a genetically predisposed host. Epidemiological and family studies have provided overwhelming evidence for a genetic susceptibility to IBD. Indeed, as recently reviewed, an increased prevalence of IBD in 1st degree relatives, increased concordance rates in identical versus dizygotic twins and ethnic and geographical differences in prevalence all pointed to the presence of a strong genetic basis underlying IBD. These studies also indicated that IBD is not inherited as a simple Mendelian disorder, but more likely as a complex genetic trait, including gene-gene and gene-environment interactions. This complex, multifactorial basis of IBD has significantly hampered the identification of the responsible genes. Recently however, tremendous progress has been made in identifying the genetic basis of IBD. One major finding, among others, of relevance for clinicians is that these studies provided clear evidence for a widespread belief that IBD is not a single disease entity but rather a heterogeneous disease. These developments pave the way to define clinical phenotypes based on differences in aetiology rather than on clinical appearance. In this regard, the major assumption is that groups of patients exist with different pathogenesis and most important for the clinicians, with different prognosis and response to treatment. In the current review, we will focus on recent developments in unravelling the genetic basis of IBD and we will discuss the potential clinical applications of these findings.

Strategies to identify genes underlying complex genetic disorders
Whole human genome screens
Human genome screens are used to search for alleles that are passed on together with the disease (linkage). They are based on the discovery that in the human genome dinucleotide repeats (or microsatellites) are
present on average every 50 thousand base pairs. The number of repeats represents an allele, and varies between people. These alleles are passed on from parents to their children. Since each parent passes randomly one copy of each chromosome to a child, the chance that two siblings receive the same allele (in this case the allele from a microsatellite marker) is 50%. In contrast, an allele that is in close proximity to a disease gene will, by definition, be passed on from an affected parent to an affected child. Also, two affected siblings that share a disease gene will share the allele of any nearby microsatellite marker. Thus, by ‘screening’ large groups of parent-children combinations, sibling pairs, or other first-degree relatives using a large number of microsatellite markers equally dispersed throughout the genome, the approximate chromosomal location of a disease gene can be identified. This approach has proven very powerful in locating human disease genes, in particular in the identification of mutations underlying monogenic diseases. Since the mid-nineties, this approach has also been applied to complex immune-mediated disorders, including IBD. For monogenic disorders, the technique of linkage analysis often was capable to define the genomic location of disease alleles with sufficient precision to limit the number of candidate genes to a small enough number that allowed the cloning of the mutated gene. In contrast, in complex genetic disorders, the statistical associations with any specific locus in general are not so robust, resulting in genomic segments that may contain hundreds, if not thousands of genes. Obviously, this has deteriorated the final stage gene of identification in complex genetic disorders. One exception however is Crohn’s disease, in which genome-wide linkage efforts recently led to the successful identification of the first susceptibility gene for CD.

**Overlap in autoimmune disease loci**

An important conclusion that can be drawn from linkage studies in complex immune-mediated diseases is that these studies allowed the identification of a group of susceptibility loci on different chromosomes that appear to be involved in multiple autoimmune disorders.

This may imply that ulcerative colitis and Crohn’s disease are part of a wider spectrum of autoimmune disorders and that common genes probably exist among several diseases, which in turn may explain the associations of different autoimmune diseases in the same patients and/or in the same families. Thus, a meta-analysis of 23 genome screens performed in several autoimmune diseases and animal models thereof has shown considerable overlap in susceptibility genes.\(^3\) Human autoimmune diseases included multiple sclerosis, Crohn’s disease, familial psoriasis, and human type-I diabetes. Although as the authors have remarked, there were marked differences between most of these studies in experimental design, patient populations, sample size, markers used, and calculations of results, it appears that in almost all common autoimmune and inflammatory diseases there may be no single genes exerting a predominate effect. These loci fall into 18 clusters and contain a large number of genes of known and unknown function suggesting a possible shared genetic basis among different autoimmune diseases. These results suggest among other possibilities that it may be likely that the genes found at these clusters are involved in primary or secondary regulation of the immune system. This is highly suggestive of a common mechanism that leads to organ-specific autoimmunity in general, whereas other genes confer susceptibility to the specific disease. This may help to explain why patients with an autoimmune disease often have relatives with other autoimmune diseases.\(^3\)

At the same time, there must be disease-specific genes. This is clearly indicated for example by the fact that, as will be discussed below, the *NOD2* mutation is not involved in the susceptibility to ulcerative colitis.

**Genome-wide scanning in inflammatory bowel disease**

Hugot and coworkers published in 1996 the first genome-wide scan in Crohn’s disease.\(^4\) They identified a region of linkage on chromosome 16, which they called *IBD1*. Their findings were almost universally repeated in a large number of data sets, providing strong support that the *IBD1* locus was based on true linkage. Hugot and colleagues continued their gene-hunt by typing the region of linkage with an additional set of microsatellite markers, thereby further narrowing the region of linkage. This allowed the construction of a physical map and finally, the identification of mutations in a gene called *NOD2* (or *CARD15*).\(^5\) Several mutations were found in this gene, including a frameshift mutation at position 3020, resulting in the truncation of the C-terminal 33 amino acids. In addition, 2 missense mutations were identified that lead to substitution of Arg by Trp at position 702, and Gly by Arg at position 908.

Realizing that *NOD2* might be a likely positional candidate within the identified region of linkage, Ogura and colleagues followed a slightly different approach. They analyzed the *NOD2* gene in both a case-control study as well as intrafamilial association study, and were also able to show strong evidence for linkage with Crohn’s disease.\(^6\)
NOD2 confers responsiveness to bacterial components through the leucine-rich repeats which are needed to activate nuclear factor NF-kB by bacterial lipopolysaccharides (LPS). Ogura and colleagues also studied the functional effect of the frameshift mutation. They observed that the mutant NOD2 was capable of strikingly decreased NF-kB activation. This is somewhat paradoxical, given the pronounced inflammatory properties of Crohn’s disease, including activation of NF-kB. Several possibilities have been postulated since to explain these paradoxical findings, however, at present it is unclear how the NOD2 gene is involved in disease pathogenesis.

Despite the fact that the exact role of the NOD2 variants in the pathogenesis remains to be elucidated, the identification of this gene has caused real excitement in the scientific community. It demonstrated that the methodology of linkage analysis is applicable to complex human immune disorders. In addition, since this gene plays a key role in the innate rather than the adaptive immune response, it has stimulated a critical re-evaluation of the role of the innate versus the adaptive immune response in the pathogenesis of IBD.

The NOD2 mutations act in a recessive manner. Thus, heterozygotes only have a slightly increased relative risk to develop the disease, whereas homozygous individuals have a 40 times increased risk. In fact, in most studies thus far, no homozygous individuals have been found in the healthy control population. The slightly increased frequency in heterozygous Crohn’s disease patients as compared to controls is most likely attributable to compound heterozygotes (i.e., individuals carrying one mutation on the one chromosome and another mutation on the other). The mutations in the NOD2 gene are specific for patients with Crohn’s disease. Thus, whereas strong associations are found between this gene and Crohn’s disease, a role of this gene in the pathogenesis of ulcerative colitis has been excluded.

Other loci

The NOD2 mutations account for an estimated 20-25% of Crohn’s disease patients. Thus, there must be other genetic factors that are operative in Crohn’s disease patients as well as in ulcerative colitis patients. Indeed, genome-wide screens have identified several loci that showed significant linkage and have been confirmed in independent groups of patients. These chromosomal regions include loci on chromosome 12 (IBD2), on the short arm of chromosome 6 (IBD3; the MHC region), on chromosome 14 (IBD4) and on chromosome 5 (IBD5). In addition, several other loci have been identified that either did not satisfy the statistical criteria for linkage, or that have not been confirmed yet in independent data sets. Although in some instances the region has been narrowed down to an interval that is theoretically small enough to allow positional cloning, the disease causing mutations remain to be elucidated for these regions.

The search for these additional genes may be facilitated by a complementary approach, i.e., the genetic analysis of animal models of intestinal inflammation. Several experimental models have been identified that resemble, to a certain extent, human Crohn’s disease or ulcerative colitis. Genetic analysis of rodent strains that exhibit differences in susceptibility may theoretically lead to the identification of genetic regions and genes that are pertinent to human IBD as well. This approach has been very fruitful for a large number of complex immunemediated disorders, including asthma, rheumatoid arthritis, and SLE. Despite the fact that a large number of immunologically well-characterized models for intestinal inflammation exist, the genetic analysis of these models has only recently received attention. At least 4 different models of murine colitis have thus far been examined which led to the identification of several chromosomal regions involved in determining susceptibility/resistance in these models. The relevance of these findings can be directly applied to human disease. Thus, for example, we recently identified a region on mouse chromosome 11 that is involved in the susceptibility to colitis induced by the intrarectal application of the haptenating agent, tri-nitro-benzene sulfonic acid. We subsequently were able to demonstrate that interleukin-12B is a logical candidate gene for the association with this chromosomal region, and we are currently involved in determining the relevance of this gene for human Crohn’s disease.

The candidate gene approach

A different approach, in which genetic associations were sought between IBD and genes that were assumed to play a role in the disease pathogenesis preceded the strategy of genome-wide scanning. This approach was based on the assumption that chronic inflammatory bowel diseases occur as the result of a dysregulated immune response against an as yet unknown factor in a genetically predisposed host. Considering the central role of the immune system in the regulation of the inflammation, most association studies in IBD focused on genes that participate in the regulation of the immune and inflammatory response, in particular HLA and cytokine

Genetic markers and the classification of Inflammatory Bowel Disease 315
gene polymorphisms. Indeed, this approach has revealed several statistically significant associations between particular gene polymorphisms and either Crohn’s disease or ulcerative colitis. It must be noted however, that the results of these studies have thus far been highly conflicting and inconclusive. In addition, if and how these genetic variants might exert their role in disease pathogenesis is unknown. Finally, these studies have not been able to be used in the diagnosis, treatment or determining the prognosis of the disease. Although the candidate gene approach has been criticized for not being powerful enough to identify disease genes underlying complex genetic disorders, this approach has recently received renewed attention. As was demonstrated by Ogura and colleagues for NOD2, studying logical candidate genes within a defined region of linkage may be competent to identify the genetic defect(s) once linkage to a chromosomal region has been established.6

Another area in which the candidate approach may be particularly fruitful, is as to how certain genetic variations may influence the course or severity of the disease. Given the recent explosion in understanding the complexity of the genetic factors underlying IBD, the variable outcomes of these studies with regard to overall disease susceptibility may not be surprising. It is feasible to speculate that the differences in outcomes may in part be due to the sampling of patients and controls and that in fact associations that are weak or non-significant in an unselected population may reveal strong significance when evaluated in the appropriate group of patients. Indeed, evidence is accumulating to support the latter. For example, we previously found that the phenotype frequencies of the DRB1*03 allele in an unselected group of unrelated white Dutch Crohn’s disease patients did not differ from the healthy control population. In contrast, when we studied this allele in a group of Crohn’s disease patients with proven perianal fistulas, we found a striking decrease in the frequency of the DRB1*03 allele in this phenotypically selected group of patients.15

Frequently, associations between HLA and cytokine gene polymorphisms revealed stronger statistical associations when studied in patients with more severe or extended disease. This lends support to the hypothesis that some of these genes are involved in determining the severity of the disease rather than overall disease susceptibility. In this respect it is less important to determine whether particular alleles are associated with the disease in general. Even in populations where the association with a particular allele is not significant, the presence of this allele may indicate that these patients may develop a more severe disease.

**Genes and the clinical phenotype of IBD**

As alluded to above, identifying the genes underlying IBD may be helpful in defining subgroups of patients that differ in prognosis and response to therapy. Although the genetic factors identified thus far are currently not clinically applicable for these purposes, some progress has been made in identifying genetic factors that determine either disease phenotype or the response to treatment. We will illustrate this with a few examples.

**NOD2 and HLA in relation to anatomical location**

A recent finding from two independent studies indicated that mutations in NOD2 are strongly associated with the anatomical site of the disease. In these studies it was found that patients carrying the NOD2 mutation predominantly have ileal disease. Thus, as shown by Ahmad and colleagues, all NOD2 homozygous or compound heterozygote patients had ileal disease, with a population attributable risk (PAR) of the NOD2 mutation for ileal disease of 32%.20 Similar findings were obtained in a study from Cuthbert published at the same time.21 Whereas the NOD2 mutations were highly associated with ileal disease, they did not appear to contribute significantly to either colonic or perianal disease.20 The pathophysiological significance of these finding remains to be determined, but it is tempting to speculate that the association of this mutation with ileal disease is related to the quantity or nature of the bacterial flora in the ileum as compared to the colon.

In contrast to the strong association between NOD2 mutations and ileal disease, the presence of colonic disease, with or without ileal involvement, was found to be associated with the classical autoimmune haplotype HLA-A1-B8-DR3 in this study.20 Finally, the primary genetic association with perianal disease was with the haplotype based around MICA*010. This corresponded to a PAR of the MHC region for both colonic and perianal disease of 40%.20

**HLA and cytokine polymorphisms in relation to disease severity**

Genetic factors that show a weak or moderate significance in an unselected group of patients, may exhibit stronger associations when examined in patients with more severe or extensive disease. Realising that patients with the need for colectomy in general can be considered to have severe disease, Roussomoustakaki and colleagues tested this hypothesis in patients undergoing sur-
In support of this possibility, they found that the DRB1*0103 allele was increased in patients with ulcerative colitis undergoing colectomy (14.1% vs. 8.3% among unselected ulcerative colitis patients, and 3.2% among healthy controls). This association was greatest in patients with extensive disease (15.8%) or extraintestinal manifestations (22.8%): mouth ulcers (25.8%), arthritis (27.2%), and uveitis (35.7%). They also confirmed previous observations that allele 2 of a polymorphism in the interleukin-1 receptor antagonist gene was predominantly found in patients with extensive disease (28.6% in extensive disease versus 10.9% in distal disease).

Genotype-phenotype and the response to treatment

Anti-TNF therapy

The response rate to anti-TNF therapy (Infliximab) is 70%. A young age at diagnosis of Crohn’s disease appears to be a good predictor for a better response to infliximab therapy.

Several polymorphisms in the gene encoding TNF and the TNF receptor have been examined, but most studies have been without success. Also, the presence of NOD2/CARD15 mutations was not predictive of treatment outcome with infliximab in Crohn’s disease. Furthermore, multivariate analysis could not identify clinical characteristics that, in combination with NOD2/CARD15 mutations, were associated with response to Infliximab.

Recent observations however suggest that sANCA may identify a Crohn’s disease subgroup with a better response to infliximab and that pANCA and homozygosity for a combination of polymorphisms in the lymphotoxin alpha gene may identify subgroups with a poorer response.

Thiopurine methyltransferase gene polymorphism

Thiopurine methyltransferase (TPMT) is an enzyme that metabolizes both azathioprine and 6-MP. The inhibition of this enzyme might lead to increased levels of 6-thioguanine nucleotides that in turn could result in hazardous myelosuppression.

The TPMT genetic polymorphism has been shown to have a significant role in the therapeutic efficiency of thiopurine drugs used in the treatment of a wide range of diseases. For example, it is known that the TPMT*3A allele with two mutations G460 to A and A719 to G, is associated with TPMT enzymatic deficiency. In 24 children submitted to curative therapy of acute lymphoblastic leukaemia, 4 of them were heterozygous for the TPMT*3A allele. Examination of their clinical histories showed that all four patients exhibited signs of severe hepatic toxicity during treatment.

Severe marrow toxicity occurs in homozygotes but in the majority of cases, however, TPMT genotyping prior to azathioprine therapy would not have predicted myelosuppressive events.

Implications

As illustrated above, several studies support the concept that genetic factors may influence the course and severity as well as the response to treatment in IBD. However, large-scale prospective studies will be necessary to provide formal evidence for this concept, and to determine whether determining these genetic risk factors can be implemented in clinical practice. If successful, these studies may be helpful in the tailoring of an appropriate treatment for particular subgroups of patients. A premise for the success of these studies is to use a solid classification, which should take into account the dynamic process and the natural evolution of the disease; to use a carefully selected and ethnically matched group of healthy controls, and to use the advanced molecular biological typing techniques.

It is to be hoped that functional polymorphisms of cytokine genes and HLA profiles will help the doctor of the future to select and to monitor “old and new” specific therapies. Specific immunomodulatory therapy may not only be more effective for certain patients but also less harmful in particular situations.

REFERENCES


