IBD GENETICS REVIEW

Genetics of inflammatory bowel disease: Current status and future directions

Thomas D Walters MBBS FRACP1, Mark S Silverberg MD PhD2

The genetic analysis of complex diseases has become a mainstream biomedical research goal. The methods for genetically dissecting human diseases have now evolved into an integrated science combining human genetics, functional genomics, high-throughput experimental technology and computational techniques. The inflammatory bowel diseases (IBDs) feature prominently in this field, and the realm of IBD genetics is continually and exponentially increasing. The present review will summarize and put into perspective the breadth of IBD genetics as it has evolved over the past 10 years, and will give the reader some sense of anticipation as to where it may be leading us.

EPIDEMIOLOGY – EVIDENCE FOR A GENETIC ETIOLOGY

A variety of epidemiological data suggest that genetic factors are intimately involved in the pathogenesis of IBD. For instance, IBD demonstrates a familial pattern of disease with a much higher disease frequency in first-degree relatives of affected individuals compared with the general population (1-4). The degree of ‘heritability’ or \( \lambda_s \) (the relative risk of the sibling of an affected individual) is estimated to range from 15 to 35 for Crohn’s disease (CD) and six to nine for ulcerative colitis (UC) (5-9). Twin studies strengthen the argument for a genetic basis for IBD, with a much higher rate of disease concordance demonstrated in monzygotic twins compared with dizygotic twins, particularly in CD (10-13). Marked ethnic and racial differences in disease prevalence have been noted, particularly in Ashkenazi Jews, and are also in keeping with genetic etiology (14-22).

IBD is said to be a complex genetic disorder with the disease appearing to occur in susceptible individuals carrying one or more risk alleles after exposure to one or more environmental triggers. Because there is an increased risk of UC in CD-affected families and vice versa, a genetic model is proposed in which CD and UC are related diseases sharing some susceptibility loci but perhaps differing at others, which results in diverging disease courses (23,24). Additional genetic ‘hits’ or environmental factors may determine the particular phenotype of the individual. The chosen ‘candidate’ is usually based on findings from an earlier linkage study putting that gene within the susceptibility region identified in a prior GWS (a positional candidate) or by the purported function of the gene such as tumour necrosis factor-alpha (TNF-\( \alpha \)) (a functional candidate). These approaches are certainly complementary, and ideally, both approaches are required in tandem, as was illustrated with the identification of the first IBD susceptibility gene described below. More recently, single nucleotide polymorphisms (SNPs) have been used in the place of microsatellite markers for the purpose of performing genome-wide association studies. SNPs represent sites along the genome in which a single base pair variation occurs from person to person where the least frequent allele has a frequency of 1% or greater. Many SNPs occur within genes, with some representing variations that alter gene function and may even represent the genetic lesion of interest.

GENE IDENTIFICATION TECHNIQUES

The quest for IBD susceptibility genes has used two broad approaches: linkage analysis using multiple-affected families and association using affected individuals in a case-control format. Genome-wide scans (GWSs) were traditionally used only with linkage analysis but recent progress has allowed a genome-wide approach by case-control association studies. Given the almost infinite number of possible genetic candidates, it is not surprising that the first major advances in this area were generated by linkage analysis: a ‘hypothesis-free’ technique in which the entire genome can be evaluated to look for areas of genetic susceptibility before trying to narrow down the search to specific genes. A genome-wide linkage analysis usually requires the ascertainment of numerous ‘sibling pair’ families, that is, two affected siblings and their parents (whether affected or not). Genotyping of all the members of each family is performed for polymorphic DNA microsatellite markers (approximately 300 microsatellite markers), which are located at intervals throughout the genome. Linkage analysis generates log of odds (LOD) scores which are measures of excess sharing of the same allele between affected siblings, suggesting a correlation between inheritance of disease and inheritance of that particular allele (Figure 1). This allows for the narrowing down of the search for susceptibility genes to a more specific part of the genome. Numerous genome-wide linkage studies in IBD have been published since 1996, identifying at least nine susceptibility loci (IBD1 – IBD9) thought to contain causal genes (Table 1).

On the other hand, candidate gene analysis to some extent is hypothesis driven. It is an ‘association test’ that examines differences in the allele frequencies of the nominated gene in affected patients compared with controls. A variety of different statistical techniques can be used and the controls can be ‘population-based’ controls who must be carefully matched, or family-based controls who are often the parents of the affected individual. The chosen ‘candidate’ is usually based on findings from an earlier linkage study putting that gene within the susceptibility region identified in a prior GWS (a positional candidate) or by the purported function of the gene such as tumour necrosis factor-alpha (TNF-\( \alpha \)) (a functional candidate). These approaches are certainly complementary, and ideally, both approaches are required in tandem, as was illustrated with the identification of the first IBD susceptibility gene described below. More recently, single nucleotide polymorphisms (SNPs) have been used in the place of microsatellite markers for the purpose of performing genome-wide association studies. SNPs represent sites along the genome in which a single base pair variation occurs from person to person where the least frequent allele has a frequency of 1% or greater. Many SNPs occur within genes, with some representing variations that alter gene function and may even represent the genetic lesion of interest.
...evidence to date suggests that the three risk alleles within the colonic disease (27,30,35,42,43,51,52). In summary, the variants are associated with earlier age of onset (27,51,52) and negatively associated with purely associated with fibrostenotic disease behavior, ileal disease location (27-30,35,42,52-54). They are also suggested to be associated with fibrostenotic disease behaviour, ileal disease location (27-30,35,42,52-54). They are also suggested to be associated with fibrostenotic disease, (29,30,34,36,51,52,55) and negatively associated with pure colonic disease (27,30,35,42,43,51,52). In summary, the evidence to date suggests that the three risk alleles within the CARD15/NOD2 gene are correlated with CD which has an earlier onset, involves the ileum and is frequently fibrostenotic.

OTHER IMPORTANT IBD LOCI

To date, CARD15/NOD2 is the only confirmed IBD susceptibility gene identified. There are a number of other likely candidate genes and loci that have been described, although details remain less clear, and their roles less well characterized than CARD15/NOD2.

The IBD5 locus on chromosome 5q31

The IBD5 locus on chromosome 5q31 has received much attention recently. Aside from CARD15/NOD2, it is the next most relevant region for IBD susceptibility. The locus was initially identified as significant by two GWSSs (56,57). Subsequent analyses refined the locus to a 250 kb risk haplotype (58). The association of this haplotype with IBD has been widely replicated in a number of independent populations (59-62). The IBD5 risk haplotype has been principally associated with CD, although there have been some suggestions of a weak association with UC as well. Polymorphisms within the organic cation transporter (OCTN1 and OCTN2) genes in the region have been recently put forward as the causative variant based on a combination of functional and genetic evidence (63). OCTN1 and OCTN2 are transporters that mediate transmembrane transport of carnitine and other organic cations (64,65). Although putative mechanisms have been suggested (65), no direct evidence is available to date to explain the role the variants of these genes could play in the pathobiology of IBD. Despite the efforts of several large centres, there has been an inability to replicate the findings that OCTN1 and OCTN2 polymorphisms are independently associated with CD, and these additional studies (66-69) support the possibility that the OCTN1 and OCTN2 SNPs are simply part of the extended haplotype in the region. This region is particularly difficult to fine map because it contains a significant degree of linkage disequilibrium, making it difficult to discern a causative allele from a marker allele co-inherited with the disease-causing allele (58,70).

Phenotypically, the IBD5 locus has been associated with earlier onset disease (56) as well as perianal disease (61,67). While most commonly reported with CD, an association with UC has been reported (62) with the strength of both associations possibly increased in the presence of a known CARD15/NOD2...
The major histocompatibility complex region (IBD3) on chromosome 6p

Linkage of IBD (both UC and CD) to the IBD3 region on chromosome 6 has been confirmed in a variety of GWSs (56,57,72-74). In a recent meta-analysis (75) of 10 such scans, it was the only locus that achieved genome-wide significance. The IBD3 region contains the major histocompatibility complex genes (also referred to as the human leukocyte antigen [HLA]). The HLA complex is divided into three regions (class I, class II and class III) and contains a total of 224 densely packed highly polymorphic gene loci (76). Not surprisingly, the study of this area has proved very challenging, with varying conflicting results. Several linkage and association studies (56,73,74,77-80) have implicated the HLA region in both IBD susceptibility and phenotype, although it has not been consistently confirmed in replication studies (81). A recent meta-analysis of 20 studies (82) highlighted both positive and negative associations with a variety of class II DRB1 alleles. For example, DRB1*1502 is associated with UC across a variety of different populations. Despite extremely variable background prevalence (from less than 1% in Northern Europeans to approximately 25% in Japanese) the relative risk in each ethnic group is surprisingly similar (two to 4.5) (80,83-86). This pattern suggests the allele is a susceptibility variant. However, genetic variation may not only contribute to susceptibility, but may also modify disease phenotype. Indeed, multiple-affect families show surprising concordance for disease phenotype, including age of onset, disease location, behaviour, need for surgery and extraintestinal manifestations (3,87-90). It is possible that HLA plays a greater role in determining final disease phenotype than initial disease susceptibility (76). For instance, DRB1*07 is the most consistently replicated association with CD. It is specifically associated with ileal disease in the absence of a major CARD15/NOD2 mutation (27,28,82,91). In comparison, DRB1*0103 is a rare allele that has been associated with both UC and CD. In UC, it is the most consistently replicated HLA association. In UC, the association has been reported across a variety of ethnic groups in combination with various haplotypes, and is particularly strong in patients with extensive or severe disease (80,83,92-95). In CD, the allele is strongly associated with isolated colonic disease (27,28,91,96). Interestingly, this allele (along with HLA-B*27) has also been associated with uveitis (97). In addition, DRB1*0103 (as well as B*27 and B*35) has been associated with type I peripheral arthropathy (acute, self-limiting, pauciarticular large- joint arthritis associated with IBD relapses) while HLA-B*44 has been associated with type II arthritis (symmetrical, seronegative, small-joint arthropathy unrelated to disease activity) (98).

Many of the genes in this region are also involved in the immune response and include the genes for TNF, lymphotoxin alpha and the heat shock proteins. Of these, TNF is certainly the most intensely studied. Associations have been found with CD and promoter polymorphisms at positions –308, –857, –863 and –1031; however, their functional significance remains unclear (27,99-106).

Drosophila discs large homologue 5 (DLG5) and IBD susceptibility

The pericentromeric region of chromosome 10 was identified as a potential IBD susceptibility locus in a European GWS. More recently, the region was refined to a haplotype block that contained only one gene: drosophila discs large homologue 5 (DLG5) (107). The data suggested an overall association for DLG5 with IBD susceptibility rather than just CD, with a relative risk of approximately 1.5 (107). These findings have been, at least, partially replicated in some centres (108,109), but not in the majority of studies (67,110,111). DLG5 is an attractive susceptibility candidate because it is putatively involved in maintaining epithelial integrity, and thus, its dysfunction would be consistent with an etiological model of impaired barrier function. However, much work remains to be performed to ascertain the role of this gene in IBD susceptibility.

Multidrug resistance-1 gene and IBD susceptibility

The multidrug resistance-1 (MDR1) gene is a biologically plausible candidate gene for a number of reasons. First, MDR1-deficient mice are known to spontaneously develop enterocolitis when maintained in a specific pathogen-free environment (112). The gene's product, P-glycoprotein 170, is highly expressed in various epithelial surfaces including the intestine. Langemann et al (113) recently demonstrated a marked downregulation of this product in the colonic tissue of UC patients but not CD patients. Finally, the MDR-1 gene maps to chromosome 7q22 in a region that was previously identified by a GWS as being suggestive for linkage to IBD (114). To date, the most consistently reported association is with UC (115-118). Ho et al (115) described two haplotypes, one associated with disease susceptibility and the other disease protective. Current data suggest a specific association with extensive UC (115). The exact physiological role of this protein within the gut remains controversial. A variety of other case-control studies (119,120) have failed to detect any association between this gene variant and UC.

Clinical implications

The field of IBD genetics is in its infancy despite relatively rapid success in a number of areas. The integration of genetic testing into the clinic is still premature, however, it will not be long before genetic testing may become an important part of the initial workup of a patient with suspected or known IBD (Table 2).

The use of genetic testing to predict disease in presymptomatic patients is still not possible due to the relative lack of sensitivity and specificity of CARD15/NOD2 testing. The same argument applies to genetic testing for diagnosis of IBD and the early presentation of IBD patients. It is more plausible that in the future, a panel of genetic and possibly serum markers will be tested to provide a measure of how likely a person is to develop IBD and what a patient's diagnosis is after symptoms have begun. Prognostic testing in individuals diagnosed with IBD is an area where molecular testing of genetic and serum markers...
**Walters and Silverberg**

**TABLE 2**

<table>
<thead>
<tr>
<th>Potential application of genetic testing in inflammatory bowel disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predication of disease susceptibility</td>
</tr>
<tr>
<td>Presymptomatic testing to detect ‘at risk’ individuals</td>
</tr>
<tr>
<td>Diagnosis</td>
</tr>
<tr>
<td>Prognosis</td>
</tr>
<tr>
<td>Choice of therapy and response to therapy</td>
</tr>
<tr>
<td>Disease prevention with specific interventions</td>
</tr>
<tr>
<td>Do not smoke?</td>
</tr>
<tr>
<td>Avoid unnecessary nonsteroidal anti-inflammatory drugs?</td>
</tr>
<tr>
<td>Probiotics administration?</td>
</tr>
<tr>
<td>Gene therapy?</td>
</tr>
</tbody>
</table>

may have the most potential current value. For example, it is known that CARD15/NOD2-positive individuals are more likely to have ileal disease and fibrostenotic disease, and are potentially more likely to proceed to an early ileal resection (121). If this is confirmed, then testing may allow the identification of those at risk of this complication and potentially target earlier or more advanced therapies to these individuals to prevent such complications. Again, a panel of genetic markers will likely be more useful for this indication. This type of predictive testing is evolving in serological marker testing where combinations of markers are associated with more aggressive small-bowel CD with a predilection to advanced disease behaviour (such as fibrostenotic or internal penetrating disease). Indeed, it is possible that the presence of such serum markers is genetically mediated and that ultimately a combination of genetic and serum marker testing will predict the course of the disease.

Pharmacogenetics is the study of how genetic variation influences an individuals’ response to therapy (122,123). The hypothesis is that characterization of a specific genetic polymorphism will predict drug response and/or toxicity. It has been estimated that genetic variation can account for 20% to 90% of variability in drug disposition and effect (124). A classic example of the role of pharmacogenetics in IBD is the utility of thiopurine methyltransferase (TPMT) genotyping and its in vivo expression levels. TPMT is one of the main degradation enzymes for these drugs, and mutations within this gene have been associated with toxic side effects. However, the appropriate clinical application of this knowledge is still debated. While one might hypothesize that knowledge of the genotype could avoid toxicity, it is important to note that TPMT mutations appear to only account for approximately 10% to 27% of observed toxic reactions (125-128). Recent work with MDR-1 introduced the potential ability to predict steroid resistance with the suggestion that there is an overexpression of MDR-1 in the peripheral blood lymphocytes of steroid-resistant IBD patients (129). Hoffmeyer et al (130) have demonstrated an association with a specific variant of this gene (SNP C3235T) and its in vivo expression levels. In the future, it may be possible to predict a patient’s steroid responsiveness from a genetic test (122,131). Conversely, as discussed earlier, an association was demonstrated with underexpression of MDR-1 and the development of colitis. Pharmacogenetic studies are being undertaken in virtually all current IBD therapies, however, to date, there are minimal data available that would yield an impact upon clinical care.

**FUTURE DIRECTIONS**

While each individual genetic risk factor identified thus far for IBD accounts for very little of the disease’s overall heritability, the interaction between these and other genes as well as between genes and environmental risk factors is likely to play a very important role in disease pathogenesis and outcome. With advances in genetic technology leading to genome-wide association testing about to be completed in IBD, it is much more likely that all of the important genetic variants that contribute to IBD susceptibility will be identified. This will initiate an era of tremendous promise in the field of IBD research with the unique opportunity to make real advances in understanding the cause of IBD and to enable us to reach the goal of disease prevention and possibly a cure. While this is surely to take many years of painstaking work, the discoveries in IBD genetics will provide many ways to improve the lives of those living with IBD and make more tools available to those treating IBD. With improvements in diagnostic and prognostic testing as well as the development of novel drugs or drug tools to enable us to use existing therapies more effectively, the field of IBD genetics places us at the dawn of a new horizon in our efforts to better understand and manage IBD.

**REFERENCES**


Walters and Silverberg


Genetics of inflammatory bowel disease


Submit your manuscripts at http://www.hindawi.com