ABSTRACT: It is now clear that the major identified risk factor for the inflammatory bowel diseases (IBDs) is a positive family history. Furthermore, the available data in spouses and twins indicate that the genetic susceptibility is due in large measure to shared familial predisposition. This emphasizes the importance of identifying the actual familial susceptibilities. Given the data for immunopathogenetic etiologies in the genesis of IBD, the logical candidate genes are those that involve the immune system. Data for several of these gene marker systems have been considered confusing or inconclusive. When approached with the concept that IBD is a genetically heterogeneous group of disorders, patterns are beginning to emerge for the human lymphocyte antigen class II region genes on chromosome 6, and the complement C3 gene on chromosome 19. Available data do not yet implicate the immunoglobulin or T cell receptor genes, but further studies are needed, especially for the latter. Firm identification of genetic susceptibilities will require the study of an adequate number of families, which is being facilitated by the establishment of an IBD family-based cell line bank. Identification of the genes that predispose to IBD will allow the study of natural history from susceptibility to clinical disease and, when understood, will provide new approaches to disease therapy and even prevention. Can J Gastroenterol 1990;4(7):261-266

Key Words: Genetic, Heterogeneity, Immunogenetics, Inflammatory bowel disease

Susceptibilités immunogénétiques dans les maladies inflammatoires de l'intestin

RESUME: Il est maintenant clairement reconnu que des antécédents familiaux positifs constituent le facteur de risque majeur des maladies inflammatoires de l'intestin (MII). De plus, les données portant sur les conjoints et les jumeaux indiquent que la sensibilité génétique est due en large partie à une prédisposition

Division of Medical Genetics, Departments of Medicine and Pediatrics, Cedars-Sinai Medical Center, and the UCLA School of Medicine, Los Angeles, California, USA

Correspondence and reprints: Dr Jerome I Rotter, Director, Division of Medical Genetics, Cedars-Sinai Medical Center, ASB 3, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA

THE MAJOR IDENTIFIED RISK FACTOR for the inflammatory bowel diseases (IBDs) – Crohn's disease and ulcerative colitis – is genetic susceptibility. It is the purpose of this paper first to document this statement; second, to review the progress in identifying certain of the predisposing genes; and third, to propose strategies for maximizing the success of this effort.

However, before attempting any of the above, it is useful to emphasize the importance of identifying such susceptibility genes. Currently therapy for Crohn's disease and ulcerative colitis is supportive and nonspecific. To develop truly specific therapies for the various forms of IBD, a much better understanding will be needed of the etiopathogenetic pathways from genetic susceptibility to clinical disease. Knowledge of the specific genes will suggest specific etiologic mechanisms. Possibly even more important, it is quite likely when a patient presents clinically with IBD that the immunopathogenetic mechanisms have been in place for a considerable time. To proceed to disease prevention then, this preclinical natural history must be characterized, and this will only be possible when the means exist to identify
EVIDENCE FOR GENETIC FACTORS

First, what is the evidence for genetic susceptibility to IBD? A number of studies have reported an increased frequency of positive family histories of IBD among cases of IBD compared to controls, with a range of 10 to 20% (1,2). While these data clearly indicate an increased family aggregation, there are methods of reporting family history data that provide greater information than just the presence or absence of a positive family history. Data can be reported as empiric recurrence risks to various classes of relatives, eg, the actual risks to siblings and offspring (Table 1) (1). These empiric data are necessary for genetic counseling and mathematical modelling. However, since IBD consists of diseases with variable and late ages of onset, the most useful data for genetic counseling and modelling are empiric risks that have been adjusted for the ages of the relatives at risk, thus creating lifetime age-adjusted risks. These have recently been reported in one study, but much more data of this type are needed (Table 1).

Even with these data, however, certain conclusions can be drawn. First, epidemiologically, having a first degree relative with IBD is clearly both the best established and well known risk factor for IBD, with a frequency of disease in first degree relatives 30- to 100-fold greater than that in the general population. Second, if this aggregation is due to genetic factors, then at least one simple mode of inheritance can be rejected. The observation that age-adjusted empiric risks for siblings and offspring are of the same magnitude are incompatible with a straightforward recessive model of inheritance, even with reduced penetrance. Third, the familial aggregation data indicate that at least some forms of Crohn's disease and ulcerative colitis are genetically related, since the increased risk in many families is for both ulcerative colitis and Crohn's, regardless of the form of the disease in the index case.

Familial aggregation, of course, can have both genetic and environmental causes. There are at least four lines of evidence that indicate that increased familial aggregation has a significant genetic component. One is the prevalence of IBD among spouses of IBD patients (2,3). The available data indicate that the prevalence in spouses is equivalent to that in the general population. The second line of evidence is the numerous instances of affected relatives whose disease onset is completely separated geographically and temporally from other affected family members (4). The third line of evidence is the identification of several genetic syndromes that feature an increased frequency of IBD, including several immunodeficiency disorders, Turner's syndrome and the Hermansky-Pudlak syndrome (3,5,7).

Finally, the fourth line of evidence is the increased concordance of IBD in monozygotic twins (who share all genes) versus dizygotic twins (who are genetically equivalent to siblings) (Table 2) (3,8). What is also noteworthy regarding the latter is that the concordance rate in dizygotic twins is of the same order of magnitude as that in siblings. Thus the twin data not only indicate that genetic factors account for the majority of the familial aggregation, but that even within families it is the shared genes rather than shared environment that determines who eventually develops IBD.

It is clear, however, that genetic susceptibility to IBD is quite complex. The family aggregation does not fit any simple Mendelian mode of inheritance. Theoretical possibilities include Mendelian susceptibility with reduced penetrance, multifactorial/polygenic inheritance (many genes of small effect acting together in an additive fashion), multilocus (two or more major genes acting together), and genetic heterogeneity (different genetic forms of the disease, each of which could be due to any of the above genetic mechanisms). Distinguishing clearly between these possibilities will require identification of the actual genes.
METHODS OF IDENTIFYING SUSCEPTIBILITY GENES

In order to fully use the information that the major risk factors for IBD are genetic susceptibilities, the actual genes that provide that susceptibility must be identified. Fundamentally there are two basic paradigms for identifying genes that predispose to any disease (9). One is to start from the known pathophysiology of the disorder, also referred to as the phenotype, and work down towards identification of the actual gene. An excellent example of this approach is the work of Goldstein and Brown (10) in identifying defects in the gene for the low density lipoprotein receptor as the genetic defect in familial hypercholesterolemia, a dominant form of coronary heart disease. Similar approaches in IBD include ongoing studies within families of such pathophysiologic abnormalities as autoantibodies, intestinal permeability and colonic mucins (11,12).

The other basic paradigm is to start with known genes, also referred to as the genotype, and work upward towards identifying the responsible gene or genes by using the known genes as gene markers in population association or family linkage studies. One of the greatest recent successes of this approach was the identification of the gene for cystic fibrosis (13). The gene marker approaches will be discussed further.

The gene marker approach takes advantage of known variation of identified genes. The location of the gene is called the gene locus. The variant forms of the gene are called alleles. If variation of a gene is common, it is called polymorphic. There are two basic types of studies. In one, the frequency of specific alleles are compared between cases with a disease, ie, IBD, and controls. This is termed a "association study". In the other, the alleles of the marker gene are traced through a family and compared with the transmission of clinical disease. If the marker gene and clinical disease are transmitted together (co-segregate), this is termed "linkage".

There are also two categories of gene markers. One is random markers that have no hypothesized relationship to the disease under study. With modern molecular techniques, a sufficient number of these exist to include all human chromosomes. This is a 'brute force' strategy for gene localization, but it has had notable success for a number of Mendelian genetic disorders, cystic fibrosis already being mentioned as an outstanding example.

However, the success of the random gene approach has been limited to Mendelian disorders. As was already noted, familial aggregation of the various forms of IBD is considerably more complex. An alternative approach is to use gene markers from genes of known function, termed a 'candidate' (as opposed to random) gene approach. Given the large amount of data for immunopathogenic processes in the various forms of IBD, the logical candidate genes are those involving the immune system. The immunologic candidate genes that are being actively explored in IBD are listed in Table 3, along with their chromosomal locations. Many of these genes are responsible for the diversity of the immune system's responses. This list is by no means all-inclusive, and additional possibilities include various immunologic mediators and growth factors such as the interleukins.

THE HLA GENE COMPLEX

There have been a number of studies of the human lymphocyte antigen (HLA) region and IBD, and the results are often considered equivocal and confusing (3). The majority of these studies have had a number of limitations. Most have been association studies, and the majority have studied only the HLA class I loci (the A, B and C loci). However, it would be anticipated that the class II HLA genes (DR, DQ and DP), responsible for development of the immune response to an antigen, would be of greater interest. In fact, when the available DR typing studies are reviewed in the aggregate, some interesting patterns are apparent. Of the six reported studies of HLA-DR and ulcerative colitis, in four there is an increased frequency of the DR2 allele among ulcerative colitis patients (Table 4) (14-19). Not all of the differences reach statistical significance, and the strongest result is observed in the Japanese study. The data in the aggregate suggest that there may well be a subset of ulcerative colitis with a relationship to DR2, that the frequency of this subset varies among different populations, and that it may have its highest frequency among the more homogeneous Japanese population. This relationship to DR2 may also explain the familial association between IBD and multiple sclerosis (20). With regard to Crohn's disease, less of a pattern has emerged. Of six reported

### Table 3

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA complex (includes class II genes such as DR, DQ and DP)</td>
<td>6</td>
</tr>
<tr>
<td>Complement (C2, C4, C6)</td>
<td>6</td>
</tr>
<tr>
<td>Complement (C3)</td>
<td>19</td>
</tr>
<tr>
<td>GM</td>
<td>14</td>
</tr>
<tr>
<td>KM</td>
<td>2</td>
</tr>
<tr>
<td>T cell receptor</td>
<td></td>
</tr>
<tr>
<td>Alpha chain</td>
<td>14</td>
</tr>
<tr>
<td>Beta chain</td>
<td>7</td>
</tr>
<tr>
<td>HLA Human lymphocyte antigen: GM immunoglobulin heavy chain allotype</td>
<td></td>
</tr>
<tr>
<td>KM immunoglobulin light chain allotype</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Study (reference)</th>
<th>Ulcerative colitis</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asakura et al (1982)* (14)</td>
<td>70% (28/40)</td>
<td>31% (16/51)</td>
</tr>
<tr>
<td>Smolen et al (1982) (15)</td>
<td>37% (11/30)</td>
<td>20% (25/125)</td>
</tr>
<tr>
<td>McConnell et al (1983) (16)</td>
<td>46% (14/31)</td>
<td>30% (15/49)</td>
</tr>
<tr>
<td>Burnham et al (1981) (17)</td>
<td>24% (18/75)</td>
<td>36% (179/500)</td>
</tr>
<tr>
<td>Cottone et al (1985) (18)</td>
<td>23% (11/46)</td>
<td>31% (52/169)</td>
</tr>
<tr>
<td>Caruso et al (1985) (19)</td>
<td>32% (13/41)</td>
<td>17% (25/151)</td>
</tr>
</tbody>
</table>

*P<0.01
association studies with HLA-DR, a clear association was reported in only one study with HLA-DR4, again Japanese (3,21). However, of interest is that molecular studies of the HLA region have tentatively observed that Caucasian patients with Crohn's disease who by serologic typing are HLA-DR4 had a significantly higher frequency of a specific HLA-DQ DNA fragments (22). These data suggest that there are possible subsets of Crohn's disease, one of which has an HLA association, is disproportionately represented in the Japanese population, and may be identified in the more heterogeneous Caucasian population by a specific molecular DR-DQ haplotype.

The more powerful method to examine the relationship of a genetic locus to a disease is that of genetic linkage. In linkage studies one asks whether the gene marker and the disease are transmitted together within families. Linkage studies are important because linkage may be observed even when population associations are absent. More importantly, if linkage is found (with or without an association), then a major gene for that disease has been identified.

Linkage studies are conducted in families with multiple affected members. In the case of classic genetic diseases that are inherited in a Mendelian mode, extensive parametric linkage analysis methods and programs have been developed (these are called parametric because certain parameters, the mode of inheritance, are known). In diseases such as IBD that are not inherited in Mendelian fashion and for which the mode of inheritance is unknown, other analytical methods have been developed. These methods are termed 'nonparametric' and basically use for linkage assessment only the information obtained from the affected individuals in the pedigree. The most common method is the 'sibpair' method and its various extensions. This method has been applied most extensively for HLA-associated diseases, as the approach is most powerful when a marker system is highly polymorphic, true for the HLA gene complex. However, the sibpair method alone is not very powerful in the face of genetic heterogeneity, ie, linked and unlinked forms. More recently, new methods of nonparametric linkage analysis methods are being developed in which data derived from information on marker concordance in more distant pairs of blood relatives affected with the disease can be used (23, 24).

Several HLA haplotype linkage studies have been conducted in families with multiple affected members with IBD. Of seven studies of HLA linkage (ie, testing co-segregation of entire HLA haplotypes and IBD together within family groups), four observed increased haplotype sharing among affected siblings, while three could find no evidence for haplotype sharing (3).

The difficulty with these prior studies is that methods using sibpair data alone are not sensitive to the problem of disease heterogeneity. If there are forms of IBD linked to the HLA region and others which are not, then nonparametric linkage methods which depend on sibpairs alone may not be powerful enough to identify linkage. A family haplotype study, applying a nonparametric linkage method which does not use sibpairs alone, but other more distant relatives as well, does provide evidence for linkage of HLA with IBD, as does the re-analysis of the above family data (23). In the light of these latter analyses, it is possible that the prior inconclusive results with sibpair data alone reflect heterogeneity of genetic susceptibility. Thus these latter linkage results support the concept raised in association studies of genetic heterogeneity. These data suggest that IBD, and even Crohn's disease and ulcerative colitis, are not single diseases, and that there is a subset of patients defined by a combination of serologic and molecular methods in which the disease is linked to the HLA complex (and also a subset which does not involve the HLA region).

THE COMPLEMENT SYSTEM AND IBD

The complement system plays a major role in the immune response. The association of IBD with rare complement deficiencies suggests that its role may even be fundamental, at least in a subgroup of IBD patients (2,3). Recent evidence supporting this concept includes the observation of enhanced local production of complement components in the small intestine of patients with Crohn's disease (25). Furthermore, the recent proposal that the pathogenesis of Crohn's disease is mediated by multifocal gastrointestinal inflammation is also consistent with a complement-mediated process (26).

The available genetic data, though still meager, suggest that genetic susceptibility provided by complement genes may differ among Crohn's patients, ie, genetic heterogeneity. This includes both functional and gene marker studies. In a study of Crohn's disease patients and their clinically unaffected first degree relatives, Elmgren and co-workers (27) observed that 38% of cases and 18% of relatives showed abnormal generation of chemotactic activity and decreased use of C3 (the third component of complement) by the alternative complement pathway. All of the relatives with C3 abnormalities were related to probands with similar abnormalities, suggesting that: the abnormalities are not simply secondary to the Crohn's disease; the abnormalities are familial; and since these occurred only in some families and not others, these abnormalities predispose in only a subset of Crohn's disease. These abnormalities were not found in ulcerative colitis patients. Using a gene marker association approach, the same investigative group reported an association of a C3 polymorphism in a series of Crohn's disease patients (28). In this series, the 'P' and 'FS' phenotypes occurred significantly more often in Crohn's disease patients than in ulcerative colitis patients or normal controls. Further arguing for heterogeneity, the gene frequency of the 'P' allele was 0.33 in Crohn's disease patients, while 0.23 in all Crohn's disease patients, 0.18 in ulcerative colitis patients, and 0.17 in healthy volunteers. The author has preliminary unpublished data of a similar nature in an entirely separate Caucasian popula-
tion. These data not only suggest that there may be genetic etiologic differences between ulcerative colitis and Crohn's disease, but also between Crohn's disease with and without colonic involvement.

**OTHER IMMUNOGENETIC LOCI**

Other candidate loci include genes that determine variability in both immunoglobulins and T cell receptors. The available data are not large, and have consisted mainly of association studies (29,30). The available data do not as yet clearly implicate these genetic regions. But it must be emphasized that appropriate family linkage studies need to be conducted and analyzed by methods which can test for linkage in the face of genetic heterogeneity.

**DIRECTIONS OF GENETIC STUDIES**

It should be apparent from the data reviewed herein that to increase significantly present understanding of the genetics of the various diseases that comprise IBD, we must improve our understanding of the actual genetic susceptibilities. There are a number of directions of genetic studies that will add to current genetic knowledge. These are outlined in Table 5. These strategies will be greatly facilitated by the establishment of certain resources. One of the most important of these resources is a cell bank consisting of cell lines and DNA from IBD patients, families and appropriate controls. In fact, such a resource is being established at Cedars-Sinai Medical Center in Los Angeles, with the support of the National Foundation for Ileitis and Colitis. This resource should facilitate gene marker and linkage studies. The identification of specific genes for IBD will suggest new etiologic mechanisms and innovative therapies. In addition, with the identification of the specific genes that predispose to IBD, the investigative community will then be able to conduct prospective studies of disease natural history and determine what converts genetic susceptibility into clinical disease. This should lead eventually to the development of therapies for disease prevention for what are currently among the most problematic of the gastrointestinal disorders.

**ACKNOWLEDGEMENTS:** This study was supported by National Institute of Health Inflammatory Bowel Disease Center Grant DK62600, and grants from the Stuart Foundations and the National Foundation for Ileitis and Colitis.

**TABLE 5**

Strategies for genetic studies of inflammatory bowel disease

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establish a family cell line bank as a DNA resource</td>
<td>Large number of families with multiple members affected with IBD</td>
</tr>
<tr>
<td>Explore specific candidate genes in association and linkage studies</td>
<td>Case control and family material, specific genes cloned</td>
</tr>
<tr>
<td>Study pathophysiological abnormalities in relatives</td>
<td>Families of affected</td>
</tr>
<tr>
<td>Use genetic and subclinical markers to identify relatives at risk for studies of natural history</td>
<td>Family members of affected studied over time</td>
</tr>
</tbody>
</table>

**REFERENCES**


