Various chromosomal loci transfer susceptibility to the development of Crohn’s disease and/or ulcerative colitis. The disease-causing gene on one of these loci (IBD1) has been identified as CARD15/NOD2 and certain loss-of-function mutations were linked to the development of Crohn’s disease. The recent data from association studies of CARD15/NOD2 mutations with certain phenotypes of Crohn’s disease are reviewed. These mutations link to early onset ileal and fibrostenotic disease corresponding to the A1/L1 or L3/B2 subgroup of the Vienna classification. The present data on variations in HLA or cytokine genes suggest that these genes are disease modifying rather than disease predisposing. Certainly, inflammatory bowel diseases consist of more than two genotypes and phenotypes. At this stage, predictions on the number of disease causing genes, mutations or environmental factors are impossible. Eur J Gastroenterol Hepatol 15:599–606 © 2003 Lippincott Williams & Wilkins


Keywords: Crohn’s disease, ulcerative colitis, phenotype, Vienna Classification, CARD15/NOD2

*Department of Medicine 4, Division of Gastroenterology and Hepatology, University of Vienna, Austria; ‡Genetic Epidemiology Unit, Departments of Epidemiology & Biostatistics and Clinical Genetics, Erasmus MC Rotterdam, the Netherlands; and *Department of Gastroenterology and Laboratory of Immunogenetics, Vrije Universiteit Medical Centre, the Netherlands.

Correspondence to Dr Christoph Gasche, General Hospital Vienna, Department of Medicine 4, Wahringer Gurtel 18-20, A-1090 Vienna, Austria. Tel: +43 1 40400 4764; fax: +43 1 40400 4735; e-mail: christoph.gasche@akh-wien.ac.at

Definitions and relationship
It has long been known that living organisms inherit certain traits from previous generations. This inheritable information is internally coded, stored in DNA and carried by all living cells. DNA is the matrix for building and maintaining a living creature. Large variations in this matrix, such as the number of genes, exist between species; small variations affecting only a single nucleotide exist between different individuals of a single species. This internally stored information is considered the genotype.

The phenotype of a living organism is the outward, physical expression of the internal, genetic information. The phenotype involves structures, metabolism, different tissues or organs and behaviour of an organism. The phenotype is actually the outcome of genotype–environment interactions on the expression of certain biological signs or symptoms. The phenotype is a classifiable biological group.

The relationship between genotype and phenotype is simple: the genotype codes for a phenotype that is also shaped by environmental factors (Fig. 1). DNA has been identified as the genetic material that is transcribed into RNA and translated into peptides or proteins, which after being folded and processed become the working units of the cell. The functional consequences of protein expression cause the actual expression of a phenotype.

There are several types of variation in the DNA sequence, including insertions and deletions, differences in the length of repetitive sequences, and single, base pair differences. The latter are the most common. They are termed single nucleotide polymorphisms (SNPs), when the variant sequence type has a frequency of at least 1% in the population. SNPs occur approximately once in every 1250 bases [1,2].

Fig. 1

The redhead’s genotype–phenotype relationship. The DNA holds all information to produce a red hair and fair skin phenotype. Various genetic variations on the melanocortin 1 receptor (MC1R) have been identified and linked to the expression of this red hair phenotype [4]. Besides genetic variations of MC1R, sunlight (resembling environment) changes the colour tone of hair and skin. Thus, the certain tone of red hair represents the phenotypic expression of the MC1R genotype under certain environmental conditions (low or high exposure to sunlight).
Different classes of SNPs are considered according to their sites within genes (non-coding, degenerate and non-degenerate sites). Non-conservative missense mutations that have been termed coding SNPs (cSNPs), alter the amino acid sequence of the encoded protein and are found at lower rates and with lower allele frequencies than silent substitutions [3]. This likely reflects selection against harmful alleles during human evolution. Such nucleotide sequence variations may cause functional changes of the encoded proteins (typically, a loss of function), are responsible for phenotypic variations in humans [4] and may cause disease [5]. Diseases are clusters of certain symptoms or abnormalities and described as a disease phenotype. Many aspects of current biological research intend to identify genes and their variations that are causatively related to disease phenotypes (‘genomics’). To a better understanding of the mechanisms of disease, we need to study the function of proteins and changes that are caused by variations (‘proteomics’).

The phenotype of inflammatory bowel diseases
Crohn’s disease and ulcerative colitis are classified as chronic idiopathic inflammatory bowel diseases (IBDs). The two disorders are usually easily separated: Crohn’s disease may involve any part of the digestive tract, whereas ulcerative colitis is restricted to the colon. Besides this anatomical differentiation, both diseases can be separated by endoscopic and histological criteria [6–8]. Some cases of idiopathic colitis cannot be attributed to one or the other disease phenotype. Such patients are clustered under the term ‘indeterminate colitis’. Despite advances in video endoscopy and in physician training, a significant number of IBD patients are still misdiagnosed [9].

The burden of not having a pure disease phenotype is not unique to IBD but a general problem in biomedical research. Mixed disease phenotypes or misclassification cause a loss of power in genotype–phenotype correlations [9]. The phenotype of ulcerative colitis is more homogeneous and simpler than that of Crohn’s disease, because ulcerative colitis is restricted to the large bowel. It usually starts in the rectum and may continuously spread into upper segments of the colon. The disease is classified as proctitis, when inflammation is limited to the rectum; as left sided colitis, when the maximum extent is below the left colonic flexure; and as extended, or pancolitis, when disease reaches beyond the left flexure into the caecum. Occasionally, ulcerative colitis starts as pancolitis and may regress in extension.

The classification of Crohn’s disease is more elusive. Early attempts to classify the disease by its location go back to the 1970s [10]. Inherent differences in disease behaviour (obstructing versus perforating) were first recognized by the Mount Sinai group [11] and later implemented into the first classification system [12]. This system had some shortcomings, however, such as too many categories and subgroups or a fair inter-observer agreement [13]. This was one of the reasons why a simple classification has been developed [14]. After several multinational meetings and interim evaluations it was agreed to limit this classification on three categories, namely ‘age at diagnosis’, ‘disease location’ and ‘disease behaviour’ (Table 1). The group tried to provide stringent definitions (Table 2) allowing reproducible results [15]. The primary intention was to better describe biological homogeneous clusters of disease on a pure observational basis. Both ‘location’ and ‘age at diagnosis’ proved to be stable over time, but the ‘behaviour’ category may relate to ‘age at diagnosis’ and changes with disease progression. It was questioned whether ‘behaviour’ is therefore suitable for such genotype–phenotype correlations [16].

| Table 1 Phenotypic classification of Crohn’s disease [14] |
|------------------------|------------------------|------------------------|
| Age at diagnosis       | A1 - < 40 years        | A2 - > 40 years        |
| Location               | L1 - Terminal ileum    | L2 - Colon             |
|                        | L3 - Ileo-colic        | L4 - Upper GI          |
| Behaviour              | B1 - Non-stricturing non-penetrating | B2 - Stricturing |
|                        |                        | B3 - Penetrating       |
| Further data to be collected | Patient’s name:        | Date of birth:         |
|                        | Sex: female/male       |                        |
|                        | Ethnicity: Caucasian/Black/Asian/other |                |
|                        | Jewish: yes/no/partly  |                        |
|                        | Family history of IBD: first degree relatives/other/none |            |
|                        | Extra-intestinal manifestation: yes/no |                  |

Reproduced from Inflammatory Bowel Disease, with permission from Lippincott, Williams & Wilkins © 2000.
sized that common genes exist among several autoimmune diseases, which may also explain the associations of different autoimmune diseases in individual patients or within families. Indeed, a meta-analysis of 23 genome scans that had been performed in autoimmune diseases and animal models showed considerable overlap of genetic susceptibility loci [18]. These loci fell into 18 clusters containing a large number of genes of known and unknown function and suggesting a possible shared genetic basis among different autoimmune diseases. Among other possibilities, these results indicate that there is no strict genotype–phenotype correlation in autoimmune disorders, but more a shared genetic susceptibility to autoimmunity in general.

When reflecting over the recent data on IBD1 or CARD15/NOD2, this assumption is not valid for Crohn’s disease. IBD1 was the first susceptibility region that had been identified on chromosome 16q12 [19,20]. Both the positional cloning and target gene approach had found three major mutations on the CARD15/NOD2 gene (reviewed by Hugot and colleagues in this issue p. 593), which account for most of the linkage to IBD1 [21,22]. CARD15/NOD2 seems to be a rather disease specific gene that does not confer risk to ulcerative colitis or other autoimmune diseases [19,22–24]. The frequency of variant CARD15/NOD2 alleles in ulcerative colitis is not different from control populations. The transmission of mutant CARD15/NOD2 alleles from parents to children with ulcerative colitis was even lower than expected [23]. These are strong arguments that ulcerative colitis and Crohn’s disease are indeed different disease phenotypes on the basis of a different genotype.

Further studies investigated different distributions of the CARD15/NOD2 mutations within phenotypic subgroups of Crohn’s disease. Early attempts to correlate the CARD15/NOD2 mutations and the Vienna classification of Crohn’s disease showed some trends, but failed to demonstrate a significant association [25]. This was most likely due to the small number of patients studied. Several larger series followed and demonstrated an association of CARD15/NOD2 variants with certain subgroups of Crohn’s disease: firstly, with ileal disease location; secondly, with fibrostenotic behaviour; and, thirdly, probably also with early disease onset [23,24,26–30]. All but one study found less frequent colonic involvement (Table 3). No effect was seen on extra-intestinal manifestations of IBD, spondylarthropathy (without IBD) or family history of IBD [24,31]. Since the genetic tests are stringent, the variations between certain studies are most likely due to vague definitions [26,29,30] or errors in population phenotyping or due to genetic variations between ethnic groups [32].

Data from other IBD linkage regions and from some animal models fit into the same context. Though the

---

**Table 2** Definitions in the Vienna classification of Crohn’s disease [14]

<table>
<thead>
<tr>
<th>Location</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L1</strong></td>
<td>Disease limited to the terminal ileum (the lower third of the small bowel) with or without spill over into caecum.</td>
</tr>
<tr>
<td><strong>L2</strong></td>
<td>Any colonic location between caecum and rectum with no small bowel or upper gastrointestinal (GI) involvement.</td>
</tr>
<tr>
<td><strong>L3</strong></td>
<td>Disease of the terminal ileum with or without spill over into caecum and any location between ascending colon and rectum.</td>
</tr>
<tr>
<td><strong>L4</strong></td>
<td>Any disease location proximal to the terminal ileum (excluding the mouth) regardless of additional involvement of the terminal ileum or colon.</td>
</tr>
<tr>
<td><strong>B1</strong></td>
<td>Inflammatory disease which has never been complicated at any time in the course of disease.</td>
</tr>
<tr>
<td><strong>B2</strong></td>
<td>Strictureing disease is defined as the occurrence of constant luminal narrowing demonstrated by radiological, endoscopic or surgical–pathological methods with prestenotic dilatation or obstructive signs/symptoms without the presence of penetrating disease at any time in the course of disease.</td>
</tr>
<tr>
<td><strong>B3</strong></td>
<td>Penetrating disease is defined as the occurrence of intraabdominal or perianal fistulas, inflammatory masses and/or abscesses at any time in the course of disease. Perianal ulcers are also included. Excluded are post-operative intra-abdominal complications and perianal skin tags.</td>
</tr>
</tbody>
</table>

---

**Table 3** Association of CARD15/NOD2 mutations with ileal/non-colonic location

<table>
<thead>
<tr>
<th>Author and reference</th>
<th>L1 (ileal)</th>
<th>L2 (colonic)</th>
<th>L3 (ileocolonic)</th>
<th>L4 (upper GI)</th>
<th>L5 (upper GI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuthert et al. [23]</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Lesage et al. [24]</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Murillo et al. [25]</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>Ahmad et al. [27]</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Vermeire et al. [28]</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hampe et al. [29]</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>Abreu et al. [26]</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

NI, not investigated; ±, a trend, but not statistically significant; *, population too small to report.
disease causing mutations have not been identified on IBD2 (12p13.2–q24.1), IBD3 (6p21), IBD4 (14q), IBD5 (5q31), IBD6 (19q13), IBD7 (1p36) or IBD8 (16p) [20,33–40], the IBD2 locus is primarily associated with ulcerative colitis [33] and the IBD5 locus with early onset Crohn’s disease [38]. This segment on chromosome 5 encodes for many cytokines such as IL-3, -4, -5 and -13, the colony-stimulating factor 2 and in close proximity IL-12p40. Experimental evidence supporting the importance of this region also comes from other sources. Clustering of loci on the syntonic region in mice for numerous models of autoimmunity, has suggested that genes in this region are implicated in the TH1/TH2 balance [41]. Linkage data derived from a dextran sulphate sodium (DDS) model of colitis localized a lesion-severity gene also to this region [42]. In trinitrobenzene sulphonic acid (TNBS) induced colitis, susceptibility loci on chromosomes 9 (Tnbs1) and 11 (Tnbs2) were found. Tnbs2 again harbours the IL-12p40 gene, a candidate gene because IL-12 is a known central mediator for both experimental colitis and human Crohn’s disease [43].

Apparently, our current knowledge from the CARD15/NOD2 variants indicates that Crohn’s colitis and Crohn’s ileitis are genetically distinct diseases. It is, however, inappropriate to conclude from the available data that Crohn’s colitis and Crohn’s ileitis are different inflammatory diseases of the bowel. We need to keep in mind that there is universal recognition of the contribution of a yet unidentified environmental factor(s) leading to the development and course of Crohn’s disease. The intestinal bacterial flora (see the article by Rath in this issue p. 615) as well as smoking habits have been pinpointed as such contributing factors. Interestingly, no correlations were found between the latter and the presence of CARD15/NOD2 variants [24].

Genotype–phenotype correlations from association studies

In the last decade, a systematic approach to the study of those genes that control chronic inflammation has been addressed [44]. The working hypothesis is that chronic IBD occurs as a result of a deregulated immune response against a yet unknown factor in a genetically predisposed host (Fig. 2). Considering the central role of the immune system in the regulation of the inflammation, most association studies in IBD have focused on genes that participate in the regulation of the immune response, such as the HLA and cytokines. There is, however, little evidence that cytokine or HLA variants are sufficient to develop the disease. It has been argued that these genes are disease modifying rather than predisposing. Inadequate methodology and insufficient genetic–epidemiological methods limit the value of many studies. High throughput chip technol-

The major shortcoming of association studies is that the functional transformation of the majority of cytokine polymorphisms has not been determined yet. For pursuing such candidate gene studies, the gene of interest should participate in the pathogenesis of IBD. Good evidence for the pathogenetic involvement of genes is derived from genetically manipulated models with an IBD-like phenotype [45–47].

Human leucocyte antigen

Variations in the human leucocyte antigen (HLA) class I, class II and transporter gene (TAP) are believed to contribute to the heterogeneity of IBD specifically in ulcerative colitis [48] but also to the course of Crohn’s disease [49,50]. Ocular inflammation accompanying IBD is associated with HLA-B*27, B*58 and HLA-DRB1*0103 [51]. Determining whether an HLA allele is involved in a disease process is difficult. We have to remember that the major histocompatibility complex (MHC) is a region central to immune regulation. Linkage disequilibrium exists between alleles across the MHC. Certainly, additional immunoactive genes that are located within this locus, such as tumour necrosis factor (TNF), lymphotoxin (LT) or nuclear factor of kappa B-like 1.

Tumour necrosis factor

TNF is a potent proinflammatory cytokine. The TNF gene is located on chromosome 6 between HLA-B and DR, within the class III region of the MHC (Fig. 3). This region contains a number of polymorphisms including five microsatellites (a–e) and several SNPs (−238, −308, −857, −863, −1031). A
Genetic map of chromosome 6p. (A) Schematic representation of major genes controlling the inflammatory response of the MHC in the short arm of chromosome 6. The position of the two genes, DQA1 and DQB1, encoding the DQ alpha and DQ beta cell-surface glycoproteins, are given relative to other MHC genes, such as the genes encoding tumour necrosis factor alpha, lymphotoxin and the MICA and MICB genes. (B) Microsatellites (a, b, c, d and e) and the most common haplotypes of biallelic polymorphisms in the LTA and TNF gene. The variability and strong linkage disequilibrium between the HLA alleles (involved in antigen presentation) and genes controlling the inflammatory response (i.e. TNF and complement genes) obscure the analysis of isolated gene polymorphisms in the control of inflammation.
TNF haplotype associated with the previously described HLA-DRB1/DQ5 combination in Crohn’s disease (TNFa2b1c2d4e1) was found in 24% of patients [52]. It has also been found that the phenotype frequencies of the DRB1 alleles in 35 unrelated white Dutch Crohn’s disease patients with proven perianal fistulas showed a striking decrease of the DRB1*03 allele in comparison with healthy controls. The DRB1*03 allele is in strong linkage disequilibrium with a polymorphism at position −308 in the promoter region of the TNF gene (TNF−308*2). As DRB1*03 frequency, but not the closely linked TNF−308*2, was decreased, this suggests recombination between the DRB1 and TNF loci in this group of patients, and may help to define the biological basis of fistula formation [53].

Recently, a novel association of the TNF−857C promoter polymorphism with IBD has been reported for Crohn’s disease in patients not carrying common NOD2 mutations [54]. This polymorphism was associated with higher TNF production and altered OCT1 binding. Studies in Japan have shown that the TNF haplotype (−308A, −238G) is significantly associated with ulcerative colitis [55]. Considering the importance of anti-TNF therapy in the treatment of IBD, it is obvious that studies on TNF haplotypes remain of highest importance. Given the complexity of the inflammatory response, a better understanding of the molecular relation between TNF haplotypes with variation in genes of other proinflammatory or anti-inflammatory cytokines such as IL-1 gene family or IL-10 will provide insight to the immunopathogenesis of IBD.

The IL-1 gene family

The IL-1 gene family is assigned to a 430 kb stretch on chromosome 2q14.2 and is involved in the control of IL-1β, a powerful regulator of immune response and proinflammatory cytokines secretion. It contains three related genes: IL1A, IL1B and IL1 receptor antagonist (IL1RN) [56], all of which are in linkage disequilibrium. Most of the investigators have focused on IL-1β and IL1RN polymorphisms. Three biallelic polymorphisms have been reported on IL1B at positions −31, −511 and +3954. A penta allelic 86 bp microsatellite polymorphism has been reported on IL1RN. Among them the 84 bp polymorphism known as allele 2 (IL1RN*2) was associated with ulcerative colitis [57]. Pooled analysis of the 26 published studies on IL1 gene family variants in IBD showed a significant association between the allele 2 of IL1RN and ulcerative colitis but not with Crohn’s disease (B.Z. Alizadeh, unpublished data). No relation was found between the IL1B variants and ulcerative colitis or Crohn’s disease. The IL-1β converting enzyme and other adaptor genes within the IL-1 pathway that are located on different chromosomes may modify the effect of variant IL1B or IL1RN alleles. Adding to that, the evolutionary ethnic difference in alleles/haplotype frequencies or expressions, which commonly refers as population stratification or population admixture, forestalls the observation of a unique correlation between IL-1 gene family and IBD phenotype [58]. Therefore, the simple statistical approach that is commonly applied in association studies may be insufficient for candidate gene studies of multiple linked polymorphisms in a complex causal pathway. Appropriate research design and analytical methods are needed to study the translation of IL-1 gene family to subgroups of IBD [59].

What is the significance of genotype–phenotype correlations?

It had been argued for a long time that IBD is a single disease, although genotype–phenotype correlations have been put forward to suggest that this is not the case. Nowadays, it is unmistakably clear that even Crohn’s disease is genetically heterogeneous with significant differences within and between different ethnic groups [32]. Further learning about the function of susceptibility genes and their mutations that cause disease will improve our understanding of, and possible therapeutic interventions against, the underlying mechanism of disease. This mechanistic understanding of genetic susceptibility could serve as a lead for identification of the environmental factor(s) that cause the disease. Nevertheless, biological therapies are sprouting these days. Though early studies were negative [60], we anticipate that further knowledge about the function of the mutations that cause disease will allow a genotypic classification and pharmacogenetic approach to therapies that are tailored to the genotype or phenotype.

Annotated references

* Of special interest
** Of outstanding interest

8. Tanaka M, Riddell RH, Saito H, Soma Y, Hidaka H, Kudo H.
How novel genes and disease causing mutations can be identified by positional

This paper describes how a candidate gene approach can result in significant

According to location of inflamed bowel.

Description of the process, how the ALB classification has been developed and
classification is also included.

A creative example of a meta-analysis is presented that actually increases our

Sometimes the real world is hard to understand and so is this work: NOD2/
CARD15 is not involved in the pathogenesis of Crohn’s disease in Japanese
patients. With this knowledge in mind, we should reflect on our current practice
to transferring medical knowledge, specifically the result of drug trials, from one
continent to the other.

Stage genome-wide search in inflammatory bowel disease provides

Two stage genome-wide search in inflammatory bowel disease provides


Sometimes the real world is hard to understand and so is this work: NOD2/
CARD15 is not involved in the pathogenesis of Crohn’s disease in Japanese
patients. With this knowledge in mind, we should reflect on our current practice
to transferring medical knowledge, specifically the result of drug trials, from one
continent to the other.

Two stage genome-wide search in inflammatory bowel disease provides

Two stage genome-wide search in inflammatory bowel disease provides


The authors tested novel TNF polymorphisms in patients with IBD that had been stratified for the presence of NOD2 mutations. Nice functional genetics add to the value of this work.


