The history of genetics in inflammatory bowel disease

Weronica E. Ek, Mauro D’Amato, Jonas Halfvarson

Karolinska Institutet, Stockholm; Örebro University Hospital and School of Health and Medical Sciences, Örebro University, Örebro, Sweden

Abstract

The influence of genetics in the etiology of inflammatory bowel disease (IBD) was initially demonstrated by epidemiological data, including differences in prevalence among different ethnic groups, familial aggregation of IBD, concordance in twins, and association with genetic syndromes. These early observations paved the way to molecular genetics in IBD, and culminated in the identification of nucleotide-binding oligomerization domain containing 2 (NOD2) gene as an IBD risk gene in 2001. As in other complex diseases, the advent of Genome Wide Association studies has dramatically improved the resolution of the IBD genome and our understanding of the pathogenesis of IBD. However, the complexity of the genetic puzzle in IBD seems more pronounced today than ever previously. In total, 163 risk genes/loci have been identified, and the corresponding number of possible causal variants is challenging. The great majority of these loci are associated with both Crohn’s disease and ulcerative colitis, suggesting that nearly all of the biological mechanisms involved in one disease play some role in the other. Interestingly, a large proportion of the IBD risk loci are also shared with other immune-mediated diseases, primary immunodeficiencies and mycobacterial diseases.

Keywords Inflammatory bowel disease, genetics, genetic epidemiology

Introduction

There has been great progress within the field of molecular genetics in inflammatory bowel disease (IBD). The early epidemiological findings, and especially the observed concordance within twin pairs, were the major triggers that commenced the successful era of molecular genetics in IBD. Even today, these key observations can bring important pieces to our understanding of the pathogenesis of IBD.

Genetic epidemiology

Ethnic differences

IBD has been associated with considerable geographic and ethnic differences in incidence and prevalence [1]. Generally, the incidence of both Crohn’s disease (CD) and ulcerative colitis (UC) has gradually increased since the Second World War, especially in northern Europe and North America, where the highest incidence rates have been reported [2-7]. In several areas with traditionally low occurrence of IBD, such as Asia and Africa, increasing figures have been reported in more recent years [8-11]. Although historical differences could be influenced by a number of factors, including different types of biases, these shifts in the risk of developing IBD within limited period of times can barely be explained by changes in the genome, but rather provide evidence for the importance of exposure to environmental factors in the disease pathogenesis. On the other hand, the higher risk of CD in Jews, and especially in Ashkenazi Jews, seem to persist irrespective of geographic location or time period [12-16], suggesting that there actually might be ethnic differences in the genetic predisposition to IBD.

Family studies

The familial nature of IBD was first recognized in 1909 [17,18]. Since then, aggregation of cases of IBD in families has been widely confirmed, with 5-23% of patients with IBD having an affected first-degree relative (Table 1). Families with multiply affected individuals, so-called multiplex families, are most often concordant for disease type, i.e. either CD or UC [2], although mixed families are reported in approximately one-fourth of the families. This was an early argument for a model in which some genetic variants...
are disease specific and some variants are common to both UC and CD, whereby environmental factors might influence disease phenotype [33,34]. In recent years, molecular genetics has confirmed that certain genetic variants are shared by CD and UC and other variants are disease-specific [35]. The recent reported associations between methylation of specific regions of the genome and CD [36] may indicate that environmental factors could influence disease phenotype in IBD by epigenetic modifications of the genome.

### Risk for relatives

The greatest risk for developing IBD is having a relative with the disease [22,25,26]. On the whole, the estimated relative risk to a sibling of a patient with CD is 13-36, and the corresponding figures are 7-17 for UC [37]. However, from a clinical perspective patients ask for the absolute risk and not the relative risk of IBD in relatives, especially offspring. There are limited studies addressing this clinically relevant question, and quoted absolute risks differ between the studies. Overall, a lifetime risk of developing IBD for first-degree relatives of a CD patient is 4.8-5.2% for non-Jews and 7.8% for Jews [14,26,32]. The corresponding figures for first-degree relatives of a patient with UC are 1.6% for non-Jews and 5.2% for Jews [14]. Similarly, the age-corrected risk for offspring of a CD patient developing IBD is reported to be 0-10.4% in non-Jews and 7.4-15.8% in Jews. The equivalent estimates for offspring of a patient with UC are 11% and 2.9-7.4%, respectively [14,26,32].

The risk of IBD in an offspring increases dramatically if both parents suffer from IBD. Case series have estimated the risk of IBD in the offspring to be 33-52%, depending on follow up [38,39].

### Phenotypic similarities within families with IBD

IBD has traditionally been categorized as CD, UC, or IBD unclassified [40,41]. However, there are great heterogeneities within the three different diagnoses, suggesting the existence of subgroups. Epidemiological studies from the mid-1990s of familial IBD suggest that there could be a genetic basis for these different subgroups. In general, a high degree of clinical similarities of IBD has been observed within multiplex families, and data are especially striking for CD. Several groups have reported concordance for location of disease and some also for disease behavior, age at diagnosis, extraintestinal manifestations and number of bowel resections [26,31,42-44].

Less is known about concordance within families with multiply affected members with UC. Some agreement for disease extent has been reported [42,45], although the observed concordance rates were higher within families with CD, which supports a smaller contribution of genetics in UC than in CD. It cannot be ruled out

---

**Table 1** Studies of first-degree relatives in a proband with inflammatory bowel disease (IBD)

<table>
<thead>
<tr>
<th>Population</th>
<th>Proband with UC</th>
<th>Proband with CD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First-degree relatives with UC</td>
<td>First-degree relatives with IBD</td>
</tr>
<tr>
<td></td>
<td>n %</td>
<td>n %</td>
</tr>
<tr>
<td>Finnish [19]</td>
<td>436 11.3</td>
<td>436 13.8</td>
</tr>
<tr>
<td>Swedish [20]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish [21]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish [22]</td>
<td>504 7.5</td>
<td>504 8.1</td>
</tr>
<tr>
<td>Welsh [23]</td>
<td>139 5.0</td>
<td></td>
</tr>
<tr>
<td>UK [24]</td>
<td>469 6.2</td>
<td>469 6.8</td>
</tr>
<tr>
<td>UK [25]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgian [26]</td>
<td>640 13.6</td>
<td></td>
</tr>
<tr>
<td>Dutch [27]</td>
<td>400 8.0</td>
<td></td>
</tr>
<tr>
<td>French [28]</td>
<td>1316 7.5</td>
<td></td>
</tr>
<tr>
<td>Canadian [29]</td>
<td>1000 8.7</td>
<td></td>
</tr>
<tr>
<td>USA [30]</td>
<td>316 15.5</td>
<td>316 15.8</td>
</tr>
<tr>
<td>USA [31]</td>
<td>554 12.2</td>
<td></td>
</tr>
<tr>
<td>USA [14]</td>
<td>269 7.1</td>
<td>269 8.6</td>
</tr>
<tr>
<td>USA [32]</td>
<td>101 8.9</td>
<td>101 13.9</td>
</tr>
<tr>
<td>In total</td>
<td>6.2-15.5%</td>
<td>6.8-15.8%</td>
</tr>
</tbody>
</table>

CD, Crohn’s disease; UC, ulcerative colitis
that the observed concordance in disease characteristics within families with IBD is an effect of shared environment within the families, rather than of genetic predisposition. In contrast to the findings in families with IBD, no similarities were identified in a study of married couples with CD [39].

**Familial and sporadic IBD**

Based on the findings in family studies, it has even been proposed that familial IBD could be a different entity than sporadic disease. However, the evidence for phenotypic differences between these two groups is sparse, and data are inconsistent. An earlier age at onset for familial cases of IBD than for patients without any family history of IBD is probably the most robust observation [14,28,43,46]. Predominance of female cases [21,43,44] and predominant transmission from mother to child has also been described, especially in CD [21,43,44,47,48]. Based on these findings, a female sex-specific epigenetic inheritance pattern for CD has been proposed, which could contribute to the family-specific risk in CD [48]. Beyond younger age at diagnosis and predominance of female cases, differences in disease location, behavior, extraintestinal manifestations, and disease severity between familial and sporadic disease have been proposed. However, interpretation of the data is difficult, since definitions differ between the studies, univariate analyses have often been employed and associations between different clinical characteristics exist [14,33,43,49,50].

**Twin studies**

In 1988 Tysk et al published the first unbiased study showing a higher concordance rate in monozygotic twin pairs than in dizygotic twin pairs with CD, reflecting the influence of genetics in the disease pathogenesis [51]. Since then, data on twins with IBD have been reported from the United Kingdom, Denmark, and Germany [52-56]. The design of the studies differs between countries, with the Scandinavian studies being based on the national twin registry in each country [51,52]. In contrast, the British and German studies were set up by calls for twins with IBD using advertisements and newsletters distributed to members of the national patient organizations and physicians within each country [53-56]. Since the pair concordance rate, simply reflecting the proportion of concordant pairs, varies with the thoroughness of ascertainment, the proband concordance should be used for comparisons between different studies. In the pivotal study by Tysk et al proband concordance rates of 58% and 4% in monozygotic and dizygotic twins with CD, respectively, were observed, reflecting the pronounced genetic predisposition [51]. The corresponding figures for twins with UC were 6% and 0%, respectively. Orholm et al later confirmed these findings in the Danish cohort, where proband concordance rates of 58% and 0% were observed in monozygotic and dizygotic twins with CD, respectively [52]. Similarly, the corresponding rates in twins with UC were 18% and 4%, respectively. Follow ups of the two Scandinavian cohorts, extending the observation period in previous healthy twin siblings, had only marginal effects on the concordance rates, since just a few additional twins had been diagnosed during the extended observation (Table 2).

The population-based data from the Scandinavian twin registries are supported by the German and British twin studies [53-56]. Overall the concordance is higher in monozygotic than in dizygotic pairs, and the difference is more pronounced in CD than in UC [60], although the British study does not include any information on proband concordance rates. Recently, the extraordinarily high concordances in monozygotic twins have been questioned. The very short observation periods between the year of birth in national twin cohorts and the year of study might have biased the inclusion towards twins with early onset disease and possibly with a more aggressive disease course, disease phenotypes that could be associated with a more pronounced genetic predisposition [58]. In contrast to some of the previous family studies in IBD, none of the original twin studies have standardized their findings according to age. Similarly, the follow-ups of the Scandinavian cohorts have studied twin pairs included in the original publications only, not extending their analyses to potential new twin pairs within the total background population in each country. By rerunning the Swedish hospital discharge records, the proband concordance rate for CD in monozygotic twins was reduced to 38.5% after excluding twin pairs with an age difference greater than 2 years, and the difference to dizygotic twins became more pronounced [58].

### Table 2 Proband concordance (percentage) in different twin cohorts

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>Crohn’s disease</th>
<th>Ulcerative colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monozygotic twins</td>
<td>Dizygotic twins</td>
</tr>
<tr>
<td>Swedish, 1988 [51] (n=80)</td>
<td>58.3</td>
<td>3.8</td>
</tr>
<tr>
<td>Swedish, follow-up, 2000</td>
<td>62.5</td>
<td>3.8</td>
</tr>
<tr>
<td>[57] (n=80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish, 2011 [58] (n=229)</td>
<td>38.5*</td>
<td>2.0*</td>
</tr>
<tr>
<td>Danish, 2000 [52] (n=103)</td>
<td>58.3</td>
<td>0</td>
</tr>
<tr>
<td>Danish follow up, 2005</td>
<td>63.6</td>
<td>3.6</td>
</tr>
<tr>
<td>[59] (n=103)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German, 2008 [55] (n=189)</td>
<td>52.4</td>
<td>6.7</td>
</tr>
</tbody>
</table>

n, number of twin pairs, *Analyses restricted to twin pairs born 1886-1958 (n=179)
register with the Swedish twin registry and restricting the analyses to twins born during the original studied period, that is, 1886-1958, the proband concordance rates in monozygotic and dizygotic twins with CD were corrected to 38% and 2%, respectively [58]. The corresponding figures for twins with UC were 15% and 8%, respectively.

Reported heritability, providing an estimate of the relative contribution of genetics to disease etiology, is very high, especially in CD, and within the same range as other complex diseases with a pronounced genetic contribution. However, the figures need to be interpreted with caution, since the methods for calculating the heritability vary with the studies, and the confidence intervals when reported are wide. A higher relative risk for concordant disease has also been observed in dizygotic twins than in ordinary siblings [61]. This would point towards the contribution of shared internal intrauterine factors and/or shared external childhood environment, although the conclusion should be treated with some caution, since the results are based on small numbers and a less robust method.

 Concordance in clinical characteristics has also been observed within twin pairs where both twins are affected by IBD. This was first reported in the follow up of the original Swedish twin study [57]. Using the Vienna classification [62], a remarkable phenotypic similarity was observed statistically within twin pairs, considering age at diagnosis, location of the disease, and possible progress in disease location, in spite of the limited number of concordant monozygotic twins with CD. Although the similarity in disease behavior, defined as non-stricturing non-penetrating disease, stricturing disease or penetrating disease, was of borderline significance only, concordance in disease behavior has been confirmed more recently in the combined Swedish-Danish twin cohort [63]. The observation of concordance in disease behavior also includes the entity perianal disease. Beyond the similarity in clinical characteristics at diagnosis of CD, phenotypic concordance in disease behavior and location has also been reported longitudinally 10 years after diagnosis. These findings point towards a pronounced genetic impact on clinical characteristics. However, comparison with concordant dizygotic twin pairs with CD could not be performed, due to low numbers. Thus, it cannot be ruled out that the phenotypic similarity within monozygotic twin pairs concordant for CD is due to shared internal intrauterine factors and/or shared external childhood environment rather than to genetic predisposition. Also still to be explored is the possible influence of disease tolerance, that is, differences in susceptibility to tissue damage, in contrast to disease resistance [64].

 In contrast to the high degree of similarity within pairs with CD, phenotypic concordance has been observed for age at diagnosis and symptomatic onset only in monozygotic twins where both twins are affected by UC [57,59,63].

Associated syndromes and diseases with well recognized genetic susceptibility

Early observed associations of IBD with genetically determined syndromes, including Turner syndrome [65], Hermansky-Pudlak syndrome [66], glycogen storage disease Ib [67], cystic fibrosis [68], and pachydermoperiostosis [69], provided additional epidemiological evidence for a role of genetics in IBD. An increased prevalence of IBD has also been observed in other inflammatory disorders with strong evidence of genetic susceptibility, like ankylosing spondylitis [70], psoriasis [71], multiple sclerosis [72], and celiac disease [73]. Recent molecular studies have also revealed shared genetic architecture between several chronic immune-mediated diseases and additional data are awaited [74].

Molecular genetics

Genetics plays an important role in susceptibility to a wide variety of complex human diseases (diseases regulated by many genes as well as the environment). Almost 100 years ago, R.A. Fischer and others reconciled the discrete Mendelian inheritance of individual genes with the continuous distribution of complex heritable traits. While familiar clustering is observed for several of these diseases, only a few diseases seem to strictly follow Mendel’s law of inheritance. Instead, the majority of diseases involve the action of many genes (as well as non-genetic or environmental factors). Thousands of mutations in single genes have been found to cause ‘Mendelian disorders’, while attempts to find single causative genes for complex diseases have been relatively unsuccessful. The mode of inheritance clearly demonstrates that CD and UC are not simple monogenic Mendelian disorders but genetically complex diseases. The observed concordance in twin studies and families with IBD were the major triggers leading to the application of molecular genetic approaches in IBD.

Early linkage studies

Early association studies in the 1980s, using functional candidate genes, focused mainly on the HLA genes and progressed rather slowly, showing fairly disappointing results [75]. Subsequent genome wide scanning based on linkage studies, using microsatellite markers of tri- or tetranucleotide repeats, identified over-proportional shared regions of the chromosomes in affected relative pairs. In 1996, the first two genome scans using this strategy were published [76,77]. The first region of replicated linkage, meeting the Lander and Kruglyak criteria for significant linkage [78], was located on chromosome 16 (IBD1) [37]. Subsequent studies identified and replicated areas of significant linkage on additional chromosomes subsequently designated as IBD1-9 [37,79]. Some disease loci were disease specific, like IBD1 showing linkage in CD only, and others showed association with both CD and UC.

Nucleotide-binding oligomerization domain containing 2 (NOD2) gene

In 2001, two independent groups identified the first CD susceptibility gene, NOD2 gene within the IBD 1 locus [80,81].
Hugo et al. described the identification of NOD2 using positional cloning strategy, while Ogura et al. used positional plus functional candidate gene approach based on its location within the IBD1 locus and structural homology with NOD1. The NOD2 consists of two amino-terminal caspase recruitment domains (CARDs), a centrally located nucleotide binding domain and multiple leucine rich repeats (LRRs) at its carboxy-terminal end. Hugo et al. found independent associations with CD for three different polymorphisms in NOD2, all three located in or near the LRR region [80]. These three variants comprise the frameshift mutation (Leu1007InsC), which causes a truncated protein transcript, and two non-synonymous polymorphisms (Arg702Trp and Gly908Arg).

The identification of NOD2 was a major breakthrough and the independent associations between the three single nucleotide polymorphisms (SNPs) and CD have been widely replicated. The great majority of studies suggest a gene-dosing effect. Carriage of one copy of the risk alleles confers to a modest increased risk of developing CD, i.e. 2- to 4-fold. However, having two copies, homozygote or combine heterozygote, is associated with a 20- to 40-fold increased risk. The prevalence of the three main coding polymorphisms varies throughout the world, both in the healthy background population and in individuals affected by CD [82,83]. Highest prevalence rates have been reported from some parts of Europe and USA, with up to 40% of patients with CD carrying at least one of the polymorphisms. On the other hand, lower mutation rates have been reported from Northern Europe, including Scandinavia and Scotland, and NOD2 mutations seem to be almost absent in Asian countries like Japan, Korea and China [79,83].

**Genome-wide association studies**

The principal insight learned from the early genetic studies in IBD was that there is no single gene (nor even a very small number) but rather a large number of involved genes. Linkage studies (previously described), lose power rapidly with decreasing effect of the associated genetic variant. So if the genetic basis of IBD consists of dozens or hundreds of small effects, then linkage would never have the power to discover them. Genome-wide association studies (GWAS), where one compares the allele frequency of a particular variant between unrelated cases and controls, were the next step to unravel the genetic architecture of complex diseases like IBD. GWAS have successfully been used to discover loci with tiny effects, thereby confirming the increasing suspicion that individual common risk alleles generally show very weak effects on disease risk [84].

While the index GWAS identified loci conferring (in complex disease genetics terms) larger effect sizes they were, in retrospect, underpowered to detect loci that confer an odds ratio (OR) of disease of <1.2. Reliable identification of such loci requires analysis of substantially larger sample sets. In order to gain power to index GWAS, individual scans were combined into a GWAS meta-analysis. These studies, which typically consist of many thousand individuals, confirmed the suspicion that a large number of additional common alleles of small effect were to be identified.

**GWAS for IBD**

During the last years, IBD has seen tremendous success in the identification of disease susceptibility alleles. Large GWAS meta-analyses of CD and UC, that followed a few independent GWAS, have dramatically increased our knowledge of IBD genetic risk factors. Presently, 163 susceptibility loci are described in the published literature [35].

Historically, only 6 months after the first GWA study was published, Yamazaki and colleagues published the first association study for CD. This study only included 72,738 SNPs and identified several associated SNPs in the TNFSF15 gene [85]. One year later, a larger association study was published for CD. Besides the previously known risk locus NOD2, this study identified CD-risk variants in the interleukin 23 receptor (IL23R) gene. This was further replicated in cohorts of both CD and UC patients that confirmed IL23R to be a gene common for both IBD subphenotypes [86]. The first large-scale association study specifically targeting UC was published in 2008 [87]. This study identified three loci associated to UC, namely the major histocompatibility complex (MHC) region (which confirmed previous findings), the gene encoding the extracellular matrix protein 1 (ECM1), expressed in the small and large intestine and involved in nuclear factor (NF)-κB activation [88], and the macrophage stimulating gene MST1, previously shown to be associated to both CD and UC [89].

Studying early-onset presentations of complex disease is appealing to geneticists because of the expectation that these efforts have a higher chance of identifying novel risk variants that have not been identified in adult-onset studies. Early onset IBD shows more extensive disease at onset and rapid progress. IBD presents during childhood or adolescence in 15-20% of patients [90]. Two recent GWAS carried out exclusively in this age group have demonstrated genetic similarities between early- and adult-onset IBD [91,92]. In the first study, which was a subset of the second GWA study, Kughasan et al. replicated several known loci from previous adult-onset studies (NOD3, IL23R, HLA, TNFSF15) and identified two novel disease-associated loci, 20q13 and 21q22 [92]. The authors were not able to pinpoint the causal gene in the 20q13 region but considered TNFRSF6B to be the most compelling candidate based on the critical role of specific polymorphisms within genes involved in the TNF pathway in the pathogenesis of IBD. The 21q22 signal resides in a small region of LD that harbors no genes, with the nearest gene being PSMG1. Later, a Canadian study for early-onset CD replicated the 20q13 locus but not the 21q22 finding [93]. A follow-up early-onset IBD GWAS was published by the same group and identified five new loci associated with early-onset IBD, including 16p11 close to the cytokine gene IL27 [91]. This study also replicated 23 of 32 loci previously implicated in adult-onset CD and 8 of 17 loci implicated in adult-onset UC, which highlights the close pathogenetic relationship between early- and adult-onset IBD.
In 2011, GWAS had identified 41 loci significantly associated to CD [87]. When combining previous GWAS into a meta-analysis, including a total of 22,027 CD cases and 29,082 controls, another 30 significantly associated loci were found, which adds up to a total of 71 loci, explaining 23.3% of the estimated heritability for CD [94]. Following in silico analyses and manual curation a number of positional candidate genes were identified as being of interest, including SMAD3, ERAP2, IL10, IL2RA, TYK2, FUT2, DNM3TA, DENND1B, BACH2 and TAGAP [94]. A few independent studies had at the same time identified 18 loci significantly associated to UC. This was followed by a meta-analysis including previous independent studies, including a total of 16,000 cases with UC and 32,000 controls. In total, 29 additional loci for UC were identified from this effort, increasing the number of known UC loci to 47, with an estimated heritability explained of 16%. After this effort, increasing the number of known UC loci to 47, with an estimated heritability explained [87].

Many IBD-associated genes are involved in T-cell differentiation (for example cytokines IL21, IL10, IFNG, and cytokine receptor IL7R). Some of them are more specifically associated with the IL23R pathway (IL23R, JAK2, STAT3, IL12B, and PTPN2), involved in the maintenance of T₄,17 cells and in several other diseases. T₄,17 cells are thought to coordinate defence against specific pathogens and mediate inflammation [96], and the original identification of IL23R as an IBD risk factor shattered the paradigm that CD and UC were primarily T₄,1 and T₄,2-mediated diseases, respectively [86]. TNF-signaling genes (TNFRSF9, TNFRSF14, and TNFRSF15) are also well represented among IBD genes. These genes encode proteins with various immune effects including systemic inflammation and activation of inflammatory transcription factor NF-κB. As predicted from

Figure 1 Inflammatory bowel disease (IBD) loci, represented by lead gene name, according to pathway. Loci associated with inflammatory bowel disease are shown in black, Crohn's disease (CD) in blue and ulcerative colitis (UC) in green.
their close clinical relationship, many susceptibility loci have been shown to be shared between CD and UC.

The additional typing on the immunochip, which is a custom-made chip that includes 200,000 SNPs relevant to multiple different immune-mediated diseases including UC and CD, discovered that a large portion of IBD risk loci are shared with other complex immune-mediated diseases (particularly ankylosing spondylitis and psoriasis), primary immunodeficiencies and mycobacterial disease, pointing to an essential role for host factors involved in defense against infection in IBD [35]. These findings point towards shared pathogenic mechanisms between 'distinct' diseases and confirmed previous reported epidemiologic overlap between IBD and other immune-mediated diseases [97-99], and observed genetic overlap in GWA-based studies [100-103]. However, the resolution of the immunochip study revealed that there is a high degree of complexity of this genetic architecture. The genetic overlap does not necessarily consist of a shared loci for which the same SNP or haplotype confer to increased risk for more than one immune-mediated disease. Instead, the same SNP or haplotype might confer to increased risk for one disease but may be protective for another, alternatively the overlap might be caused by different haplotypes within the loci [104]. The associations with gene coding for proteins involved in autophagy and innate immunity point towards the importance of defective processing of intracellular bacteria in CD. The genetic overlap between CD and susceptibility to infection with Mycobacterium leprae is particularly intriguing. A total of seven out of eight susceptibility loci, including NOD2, IL23R, RIPK2 and TNFSF15, for infection with Mycobacterium leprae have also been associated with CD, although with genetic effects in different directions for some of these associations. It is still to be explored whether this shared immunogenetic risk profile underpins a true causative role for mycobacteria in CD, or rather represent the result of convergent evolutionary adaptations to several pathogens.

**Future aspects**

In spite of the dramatic progress within molecular genetics in recent years, some of the key epidemiological observations are still unanswered. Based on analyses of the DNA sequence, the observed high heritability in IBD is only partly understood. Future fine-mapping efforts of identified loci are awaited. Up until now, it has been difficult to establish any firm genotype-phenotype associations beyond NOD2, but analyses addressing possible contribution of other loci based on data generated by typing on the immunochip are under way. Similarly, genetic models for predicting disease diagnosis, CD vs. UC, or even phenotypic subcategories of these two diagnoses are expected within the near future. Preliminary data also suggest that genetic analyses might be used for prediction models of treatment response in the future. The IIBDGC's study on the risk of colectomy in acute severe UC might be the most promising example, where an association between rs2403456 (11p15.3) and colectomy has been reported [106]. Future longitudinal studies with periodic measurements in subjects at high risk, that is siblings and offspring below or around the peak age of onset of IBD, as well of treatment naïve, newly diagnosed patients will probably become important tools to elucidate the genetic and environmental interactions underlying these archetypical complex diseases.

**References**

14. Yang H, McElree C, Roth MP, Shanahan F, Targan SR,


49. Lee JC, Lennard-Jones JE. Inflammatory bowel disease in 67 families each with three or more affected first-degree relatives. Gastroenterology 1996;111:587-596.


60. Brant SR. Update on the heritability of inflammatory bowel disease: the importance of twin studies. Inflamm Bowel Dis 2011;17:1-5.


