The history of genetics in inflammatory bowel disease

Weronica E. Ek[a], Mauro D'Amato[a], Jonas Halfvarson[b]

Karolinska Institutet, Stocholm; 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

Ann Gastroenterol 2014; 27(4): 1-10

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,

Departments of [a]Biosciences and Nutrition, Karolinska Institutet, Stockholm (Weronica E. Ek, Mauro D'Amato); [b]Internal medicine, Division of Gastroenterology, Örebro University Hospital and School of Health and Medical Sciences, Örebro University, Örebro (Jonas Halfvarson), Sweden

Conflict of Interest: None

Correspondence to: Jonas Halfvarson, MD, PhD, Department of Internal medicine, Division of Gastroenterology, Örebro University Hospital S-70185 Örebro, Sweden, Tel.: +46 738 082361, Fax: +46 19 6021774, e-mail: jonas.halfvarson@orebroll.se

Received 22 May 2014; accepted 15 July 2014

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

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

| Population | Proband with UC | | | | Proband with CD | | | |
| | First-degree relatives with UC | | First-degree relatives with IBD | | First-degree relatives with CD | | First-degree relatives with IBD | |
	n	%	n	%	n	%	n	%
Finnish [19]	436	11.3	436	13.8	257	10.9	257	15.6
Swedish [20]					963	5.7	963	5.7
Swedish [21]					1048	6.9	1048	6.9
Danish [22]	504	7.5	504	8.1	133	2.2	133	5.2
Welsh [23]					139	5.0	139	9.3
UK [24]	469	6.2	469	6.8	424	9.4	424	10.4
UK [25]					Not stated		433	11.5
Belgian [26]					640	13.6	640	14.5
Dutch [27]					400	8.0	400	9.5
French [28]					1316	7.5	1316	8.4
Canadian [29]					1000	8.7	Not stated	
USA [30]	316	15.5	316	15.8	522	15.1	522	16.7
USA [31]					554	12.2	Not stated	
USA [14]	269	7.1	269	8.6	258	7.4	258	14.0
USA [32]	101	8.9	101	13.9	80	16.2	80	22.5
In total	6.2-15.5%		6.8-15.8%		2.2-16.2%		5.2-22.5%	

CD, Crohn's disease; UC, ulcerative colitis

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 in 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

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 [14,33,43,44,49,50].

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

Table 2 Proband concordance (percentage) in different twin cohorts

Cohorts	Crohn's disease		Ulcerative colitis	
	Monozygotic twins	Dizygotic twins	Monozygotic twins	Monozygotic twins
Swedish, 1988 [51] (n=80)	58.3	3.8	6.6	0
Swedish, follow-up, 2000 [57] (n=80)	62.5	3.8	18.8	0
Swedish, 2011 [58] (n=229)	38.5*	2.0*	14.6*	8.0*
Danish, 2000 [52] (n=103)	58.3	0	18.2	4.5
Danish follow up, 2005 [59] (n=103)	63.6	3.6	18.2	4.5
German, 2008 [55] (n=189)	52.4	6.7	27.9	3.1

*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 needs 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 monogenetic 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 (Leu1007finsC), 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 major 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, was 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 *ILR23R* 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, DNMT3A, 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 annotating associated regions using a gene relationship across implicated loci pathway analysis, expression quantitative trait loci data and correlations with non-synonymous SNPs, several candidate genes, including *IL1R2, IL8RA-IL8RB, IL7R, IL12B, DAP, PRDM1, JAK2, IRF5, GNA12* and *LSP1*, were identified, providing potentially important insights into disease pathogenesis [95].

Only a year later, the International IBD Genetics Consortium (IIBDGC) expanded the breadth of these analyses reporting one of the largest meta-analyses ever performed, based on 75,000 IBD (UC and CD) cases and controls including data from 15 different GWAS for UC and CD. This study also included additional typing on the immunochip, an array specifically designed to capture variation at 200 known risk loci for 12 common autoimmune diseases [35]. From this study, 71 new causative regions were identified, which brought the total number of independent IBD risk loci to 163, far more than reported for any other complex disease. For the 163 confirmed IBD loci, 110 appear to be relevant to both CD and UC, 23 show risk effects that are UC-specific and the remaining 30 are CD-only loci. The individual gene and causal variant are yet to be explored in many of these loci and even less is known about their functional implications. Fig. 1 summarizes a number of the confirmed common and disease specific loci, represented by lead gene name, according to pathway. Interestingly, 43 of these 53 disease specific loci show the same direction of effect in both CD and UC, suggesting that nearly all of the biological mechanisms involved in one disease play some role in the other. One intriguing exception is *NOD2*, which still represents the strongest causal gene in CD but shows significant protective effects in UC, an observation that may reflect biological differences between the two diseases.

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 $T_H 17$ cells and in several other diseases. $T_H 17$ 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 $T_H 1$ and $T_H 2$ –mediated diseases, respectively [86]. *TNF*-signaling genes (*TNFRSF9, TNFRSF14,* and *TNFSF15*) 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
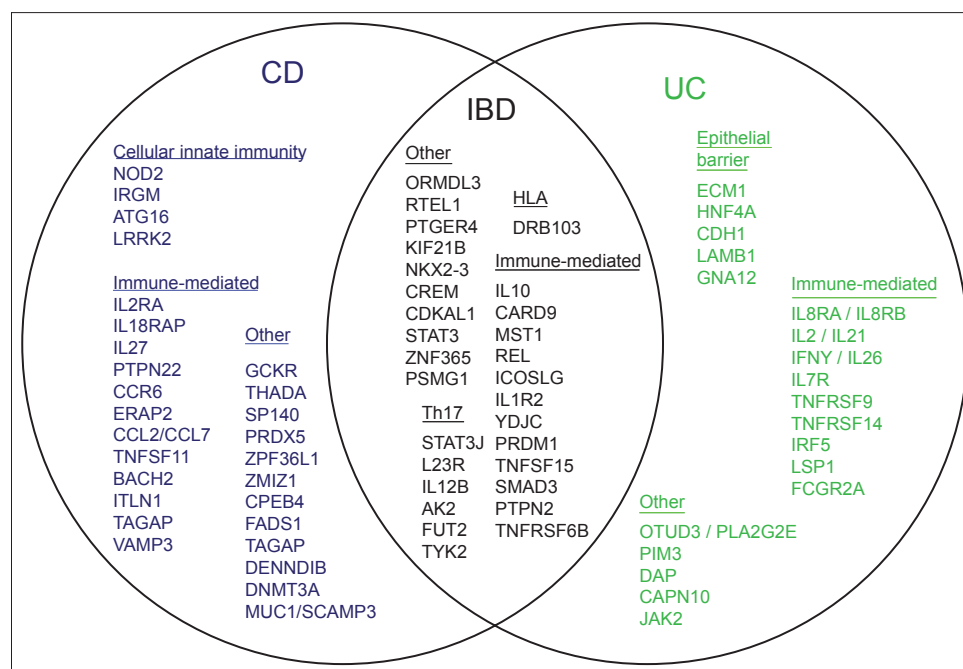

Figure 1 Inflammatory bowel disease (IBD) loci, represented by lead gene name, according to pathway. Loci associated with inflammatory bowel disease are shown in black, Crohn's disease (CD) in blue and ulcerative colitis (UC) in green

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. At the present, a transition from GWAS-style date (which only studies a subset of common variation) to complete genome sequencing is enabled by the decreased cost of sequencing [105]. Sequencing has the potential to enable the wide spectrum of variation beyond the common alleles targeted by GWAS. Possible contributions of other molecular mechanisms of heritability, such as epigenetics, are still largely undiscovered and need to be explored. Similarly, the pronounced phenotypic similarities within multiplex families and especially within concordant monozygotic twin pairs with CD suggest that genetics also influences the phenotype of the disease. Patients with IBD have so far not benefited from the scientific progress within molecular genetics, but clinical applications 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 archetypal complex diseases.

References

1. Economou M, Pappas G. New global map of Crohn's disease: Genetic, environmental, and socioeconomic correlations. *Inflamm Bowel Dis* 2008;**14**:709-720.
2. Binder V. Genetic epidemiology in inflammatory bowel disease. *Dig Dis* 1998;**16**:351-355.
3. Kyle J. Crohn's disease in the northeastern and northern Isles of Scotland: an epidemiological review. *Gastroenterology* 1992;**103**:392-399.
4. Lapidus A. Crohn's disease in Stockholm County during 1990-2001: an epidemiological update. *World J Gastroenterol* 2006;**12**:75-81.
5. Gunesh S, Thomas GA, Williams GT, Roberts A, Hawthorne AB. The incidence of Crohn's disease in Cardiff over the last 75 years: an update for 1996-2005. *Aliment Pharmacol Ther* 2008;**27**:211-219.
6. Chouraki V, Savoye G, Dauchet L, et al. The changing pattern of Crohn's disease incidence in northern France: a continuing increase in the 10- to 19-year-old age bracket (1988-2007). *Aliment Pharmacol Ther* 2011;**33**:1133-1142.
7. Molinie F, Gower-Rousseau C, Yzet T, et al. Opposite evolution in incidence of Crohn's disease and ulcerative colitis in Northern France (1988-1999). *Gut* 2004;**53**:843-848.
8. Sood A, Midha V, Sood N, Bhatia AS, Avasthi G. Incidence and prevalence of ulcerative colitis in Punjab, North India. *Gut* 2003;**52**:1587-1590.
9. Yang SK, Hong WS, Min YI, et al. Incidence and prevalence of ulcerative colitis in the Songpa-Kangdong District, Seoul, Korea, 1986-1997. *J Gastroenterol Hepatol* 2000;**15**:1037-1042.
10. Yao T, Matsui T, Hiwatashi N. Crohn's disease in Japan: diagnostic criteria and epidemiology. *Dis Colon Rectum* 2000;**43**:S85-S93.
11. Yang SK, Yun S, Kim JH, et al. Epidemiology of inflammatory bowel disease in the Songpa-Kangdong district, Seoul, Korea, 1986-2005: a KASID study. *Inflamm Bowel Dis* 2008;**14**:542-549.
12. Sandler E. Epidemiology of inflammatory bowel disease. In: Targan SR, Shanahan F (editors): Inflammatory bowel disease: from bench to bedside. Williams & Wilkins: Baltimore; 1994, pp. 5-30.
13. Roth MP, Petersen GM, McElree C, Feldman E, Rotter JI. Geographic origins of Jewish patients with inflammatory bowel disease. *Gastroenterology* 1989;**97**:900-904.
14. Yang H, McElree C, Roth MP, Shanahan F, Targan SR,

Rotter JI. Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. *Gut* 1993;**34**:517-524.

15. Niv Y, Abuksis G, Fraser GM. Epidemiology of Crohn's disease in Israel: a survey of Israeli kibbutz settlements. *Am J Gastroenterol* 1999;**94**:2961-2965.

16. Odes HS, Fraser D, Hollander L. Epidemiological data of Crohn's disease in Israel: etiological implications. *Public Health Rev* 1989;**17**:321-335.

17. Allchin WH. Ulcerative colitis. *Proc Roy Soc Med* 1909;**2**:59-75.

18. Kirsner JB, Spencer JA. Family occurrences of ulcerative colitis, regional enteritis, and ileocolitis. *Ann Intern Med* 1963;**59**:133-144.

19. Halme L, Turunen U, Helio T, et al. Familial and sporadic inflammatory bowel disease: comparison of clinical features and serological markers in a genetically homogeneous population. *Scand J Gastroenterol* 2002;**37**:692-698.

20. Monsen U, Brostrom O, Nordenvall B, Sorstad J, Hellers G. Prevalence of inflammatory bowel disease among relatives of patients with ulcerative colitis. *Scand J Gastroenterol* 1987;**22**:214-218.

21. Monsen U, Bernell O, Johansson C, Hellers G. Prevalence of inflammatory bowel disease among relatives of patients with Crohn's disease. *Scand J Gastroenterol* 1991;**26**:302-306.

22. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen IA, Binder V. Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991;**324**:84-88.

23. Mayberry JF, Rhodes J, Newcombe RG. Familial prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *Br Med J* 1980;**280**:84.

24. Probert CS, Jayanthi V, Hughes AO, Thompson JR, Wicks AC, Mayberry JF. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut* 1993;**34**:1547-1551.

25. Satsangi J, Rosenberg WMC, Jewell DP. The prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *Eur J Gastroenterol Hepatol* 1994;**5**:413-416.

26. Peeters M, Nevens H, Baert F, et al. Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology* 1996;**111**:597-603.

27. Weterman IT, Pena AS. Familial incidence of Crohn's disease in The Netherlands and a review of the literature. *Gastroenterology* 1984;**86**:449-452.

28. Carbonnel F, Macaigne G, Beaugerie L, Gendre JP, Cosnes J. Crohn's disease severity in familial and sporadic cases. *Gut* 1999;**44**:91-95.

29. Freeman HJ. Familial Crohn's disease in single or multiple first-degree relatives. *J Clin Gastroenterol* 2002;**35**:9-13.

30. Farmer RG, Michener WM, Mortimer EA. Studies of family history among patients with inflammatory bowel disease. *Clin Gastroenterol* 1980;**9**:271-277.

31. Bayless TM, Tokayer AZ, Polito JM, 2nd, Quaskey SA, Mellits ED, Harris ML. Crohn's disease: concordance for site and clinical type in affected family members--potential hereditary influences. *Gastroenterology* 1996;**111**:573-579.

32. Roth MP, Petersen GM, McElree C, Vadheim CM, Panish JF, Rotter JI. Familial empiric risk estimates of inflammatory bowel disease in Ashkenazi Jews. *Gastroenterology* 1989;**96**:1016-1020.

33. Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004;**126**:1504-1517.

34. Bridger S, Lee JC, Bjarnason I, Jones JE, Macpherson AJ. In siblings with similar genetic susceptibility for inflammatory bowel disease, smokers tend to develop Crohn's disease and non-smokers develop ulcerative colitis. *Gut* 2002;**51**:21-25.

35. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119-124.

36. Nimmo ER, Prendergast JG, Aldhous MC, et al. Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflamm Bowel Dis* 2012;**18**:889-899.

37. Ahmad T, Satsangi J, McGovern D, Bunce M, Jewell DP. Review article: the genetics of inflammatory bowel disease. *Aliment Pharmacol Ther* 2001;**15**:731-748.

38. Bennett RA, Rubin PH, Present DH. Frequency of inflammatory bowel disease in offspring of couples both presenting with inflammatory bowel disease. *Gastroenterology* 1991;**100**:1638-1643.

39. Laharie D, Debeugny S, Peeters M, et al. Inflammatory bowel disease in spouses and their offspring. *Gastroenterology* 2001;**120**:816-819.

40. Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989;**170**:2-6; discussion 16-19.

41. Stange EF, Travis SP, Vermeire S, et al. European evidence based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *Gut* 2006;**55**(Suppl 1):i1-i15.

42. Satsangi J, Grootscholten C, Holt H, Jewell DP. Clinical patterns of familial inflammatory bowel disease. *Gut* 1996;**38**:738-741.

43. Colombel JF, Grandbastien B, Gower-Rousseau C, et al. Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* 1996;**111**:604-607.

44. 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.

45. Annese V, Andreoli A, Astegiano M, et al. Clinical features in familial cases of Crohn's disease and ulcerative colitis in Italy: a GISC study. Italian Study Group for the Disease of Colon and Rectum. *Am J Gastroenterol* 2001;**96**:2939-2945.

46. Polito JM, 2nd, Childs B, Mellits ED, Tokayer AZ, Harris ML, Bayless TM. Crohn's disease: influence of age at diagnosis on site and clinical type of disease. *Gastroenterology* 1996;**111**:580-586.

47. Akolkar PN, Gulwani-Akolkar B, Heresbach D, et al. Differences in risk of Crohn's disease in offspring of mothers and fathers with inflammatory bowel disease. *Am J Gastroenterol* 1997;**92**:2241-2244.

48. Zelinkova Z, Stokkers PC, van der Linde K, Kuipers EJ, Peppelenbosch MP, van der Woude CP. Maternal imprinting and female predominance in familial Crohn's disease. *J Crohns Colitis* 2012;**7**:771-776.

49. Franchimont D, Belaiche J, Louis E, et al. Familial Crohn's disease: a study of 18 families. *Acta Gastroenterol Belg* 1997;**60**:134-137.

50. Cottone M, Brignola C, Rosselli M, et al. Relationship between site of disease and familial occurrence in Crohn's disease. *Dig Dis Sci* 1997;**42**:129-132.

51. Tysk C, Lindberg E, Järnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988;**29**:990-996.

52. Orholm M, Binder V, Sorensen TI, Rasmussen LP, Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 2000;**35**:1075-1081.

53. Subhani J, Montgomery SM, Pounder RE, Wakefield AJ. Concordance rates of twins and siblings in inflammatory bowel disease (IBD). [abstract]. *Gut* 1998;**42**(Suppl I):A40.

54. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ* 1996;**312**:95-96.

55. Spehlmann ME, Begun AZ, Burghardt J, Lepage P, Raedler A, Schreiber S. Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. *Inflamm Bowel Dis* 2008;**14**:968-976.

56. Ng SC, Woodrow S, Patel N, Subhani J, Harbord M. Role of genetic and environmental factors in British twins with inflammatory

bowel disease. *Inflamm Bowel Dis* 2012;**18**:725-736.

57. Halfvarson J, Bodin L, Tysk C, Lindberg E, Järnerot G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003;**124**:1767-1773.

58. Halfvarson J. Genetics in twins with Crohn's disease: less pronounced than previously believed? *Inflamm Bowel Dis* 2011;**17**:6-12.

59. Jess T, Riis L, Jespersgaard C, et al. Disease concordance, zygosity, and NOD2/CARD15 status: follow-up of a population-based cohort of Danish twins with inflammatory bowel disease. *Am J Gastroenterol* 2005;**100**:2486-2492.

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

61. Bengtson MB, Aamodt G, Vatn MH, Harris JR. Concordance for IBD among twins compared to ordinary siblings--a Norwegian population-based study. *J Crohns Colitis* 2010;**4**:312-318.

62. Gasche C, Scholmerich J, Brynskov J, et al. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;**6**:8-15.

63. Halfvarson J, Jess T, Bodin L, et al. Longitudinal concordance for clinical characteristics in a Swedish-Danish twin population with inflammatory bowel disease. *Inflamm Bowel Dis* 2007;**13**:1536-1544.

64. Medzhitov R, Schneider DS, Soares MP. Disease tolerance as a defense strategy. *Science* 2012;**335**:936-941.

65. Weinrieb IJ, Fineman RM, Spiro HM. Turner syndrome and inflammatory bowel disease. *N Engl J Med* 1976;**294**:1221-1222.

66. Schinella RA, Greco MA, Cobert BL, Denmark LW, Cox RP. Hermansky-Pudlak syndrome with granulomatous colitis. *Ann Intern Med* 1980;**92**:20-23.

67. Couper R, Kapelushnik J, Griffiths AM. Neutrophil dysfunction in glycogen storage disease Ib: association with Crohn's-like colitis. *Gastroenterology* 1991;**100**:549-554.

68. Lloyd-Still JD. Crohn's disease and cystic fibrosis. *Dig Dis Sci* 1994;**39**:880-885.

69. Compton RF, Sandborn WJ, Yang H, et al. A new syndrome of Crohn's disease and pachydermoperiostosis in a family. *Gastroenterology* 1997;**112**:241-249.

70. Jayson MI, Bouchier IA. Ulcerative colitis in patients with ankylosing spondylitis. *Proc R Soc Med* 1968;**61**:340-341.

71. Yates VM, Watkinson G, Kelman A. Further evidence for an association between psoriasis, Crohn's disease and ulcerative colitis. *Br J Dermatol* 1982;**106**:323-330.

72. Minuk GY, Lewkonia RM. Possible familial association of multiple sclerosis and inflammatory bowel disease. *N Engl J Med* 1986;**314**:586.

73. Cottone M, Cappello M, Puleo A, Cipolla C, Filippazzo MG. Familial association of Crohn's and coeliac diseases. *Lancet* 1989;**2**:338.

74. Lees CW, Barrett JC, Parkes M, Satsangi J. New IBD genetics: common pathways with other diseases. *Gut* 2011;**60**:1739-1753.

75. Asakura H, Tsuchiya M, Aiso S, et al. Association of the human lymphocyte-DR2 antigen with Japanese ulcerative colitis. *Gastroenterology* 1982;**82**:413-418.

76. Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;**379**:821-823.

77. Satsangi J, Parkes M, Louis E, et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;**14**:199-202.

78. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;**11**:241-247.

79. Cooney R, Jewell D. The genetic basis of inflammatory bowel disease. *Dig Dis* 2009;**27**:428-442.

80. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**:599-603.

81. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603-606.

82. Bonen DK, Cho JH. The genetics of inflammatory bowel disease. *Gastroenterology* 2003;**124**:521-536.

83. Riis L, Vind I, Vermeire S, et al. The prevalence of genetic and serologic markers in an unselected European population-based cohort of IBD patients. *Inflamm Bowel Dis* 2007;**13**:24-32.

84. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;**316**:889-894.

85. Yamazaki K, McGovern D, Ragoussis J, et al. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet* 2005;**14**:3499-3506.

86. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;**314**:1461-1463.

87. Fisher SA, Tremelling M, Anderson CA, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat Genet* 2008;**40**:710-712.

88. Matsuda A, Suzuki Y, Honda G, et al. Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene* 2003;**22**:3307-3318.

89. Goyette P, Lefebvre C, Ng A, et al. Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis. *Mucosal Immunol* 2008;**1**:131-138.

90. Henderson P, van Limbergen JE, Wilson DC, Satsangi J, Russell RK. Genetics of childhood-onset inflammatory bowel disease. *Inflamm Bowel Dis* 2011;**17**:346-361.

91. Imielinski M, Baldassano RN, Griffiths A, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet* 2009;**41**:1335-1340.

92. Kugathasan S, Baldassano RN, Bradfield JP, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet* 2008;**40**:1211-1215.

93. Amre DK, Mack DR, Morgan K, et al. Investigation of reported associations between the 20q13 and 21q22 loci and pediatric-onset Crohn's disease in Canadian children. *Am J Gastroenterol* 2009;**104**:2824-2828.

94. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;**42**:1118-1125.

95. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011;**43**:246-252.

96. Bettelli E, Korn T, Oukka M, Kuchroo VK. Induction and effector functions of T(H)17 cells. *Nature* 2008;**453**:1051-1057.

97. Bernstein CN, Wajda A, Blanchard JF. The clustering of other chronic inflammatory diseases in inflammatory bowel disease: a population-based study. *Gastroenterology* 2005;**129**:827-836.

98. Cohen R, Robinson D, Jr., Paramore C, Fraeman K, Renahan K, Bala M. Autoimmune disease concomitance among inflammatory bowel disease patients in the United States, 2001-2002. *Inflamm Bowel Dis* 2008;**14**:738-743.

99. Weng X, Liu L, Barcellos LF, Allison JE, Herrinton LJ. Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern california-managed care organization. *Am J Gastroenterol* 2007;**102**:1429-1435.

100. Coenen MJ, Trynka G, Heskamp S, et al. Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum Mol Genet* 2009;**18**:4195-4203.

101. Hinks A, Martin P, Flynn E, et al. Investigation of type 1 diabetes and coeliac disease susceptibility loci for association with juvenile idiopathic arthritis. *Ann Rheum Dis* 2010;**69**:2169-2172.

102. Smyth DJ, Plagnol V, Walker NM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* 2008;**359**:2767-2777.

103. Zhernakova A, Stahl EA, Trynka G, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet* 2011;**7**:e1002004.

104. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet* 2013;**14**:661-673.

105. Kilpinen H, Barrett JC. How next-generation sequencing is transforming complex disease genetics. *Trends Genet* 2013;**29**:23-30.

106. Radford-Smith G, Doecke JD, Lees CW, et al. Clinical and Molecular Characterization of Medically Refractory Acute, Severe Colitis: Preliminary Results From the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) Immunochip Study. *Gastroenterology* 2013;**144**:S-470.