Chapter 5
The Genetics of Crohn's Disease

Andre Franke and Miles Parkes

Abstract This chapter summarises progress in understanding the genetic basis of Crohn's disease (CD). It starts with a brief review of family studies for CD epidemiology and then summarises findings of the so-called "linkage era". Given the success of genome-wide association studies (GWAS) in terms of identifying CD susceptibility loci, the focus of this chapter is on the key GWAS studies and their main results. These have demonstrated association with multiple Th17 pathway components and strongly implicated defects in innate immunity, particularly in autophagy and the handling of intracellular bacteria, as playing key roles in CD pathogenesis. Besides GWAS for adult-onset CD, paediatric-onset GWAS are discussed. Although paediatric-onset CD presents with more extensive disease and rapid progression compared to adult-onset CD, genetic studies have shown marked molecular similarities between the two disease forms. Not only have single GWAS contributed to completing the molecular map of CD genetics, but also systematic cross-phenotype analyses and meta-analyses of several CD GWAS, both of which are discussed in the current chapter. Lastly, the first sequencing studies for CD as well as future challenges are described.

A. Franke, Ph.D. (✉)
Institute of Clinical Molecular Biology, Christian–Albrechts–University of Kiel, Schittenhelmstraße 12, 24105 Kiel, Schleswig-Holstein, Germany
e-mail: a.franke@mucosa.de

M. Parkes, M.A., M.B.B.S., F.R.C.P., D.M.
IBD Research Unit, Department of Medicine, University of Cambridge, Cambridge, UK
e-mail: miles.parkes@doctors.org.uk

M. D'Amato and J.D. Rioux (eds.), *Molecular Genetics of Inflammatory Bowel Disease*,
DOI 10.1007/978-1-4614-8256-7_5,

Introduction

This chapter summarises progress in understanding the genetic basis of Crohn's disease (CD). It starts with a brief review of family studies for CD epidemiology and then summarises findings of the so-called "linkage era". Given the success of genome-wide association studies (GWAS) in terms of identifying CD susceptibility loci, the focus of this chapter is on the key GWAS studies and their main results. These have demonstrated association with multiple Th17 pathway components and strongly implicated defects in innate immunity, particularly in autophagy and the handling of intracellular bacteria, as playing key roles in CD pathogenesis. Besides GWAS for adult-onset CD, paediatric-onset GWAS are discussed. Although paediatric-onset CD presents with more extensive disease and rapid progression compared to adult-onset CD, genetic studies have shown marked molecular similarities between the two disease forms. Not only have single GWAS contributed to completing the molecular map of CD genetics, but also systematic cross-phenotype analyses and meta-analyses of several CD GWAS, both of which are discussed in the current chapter. Lastly, the first sequencing studies for CD as well as future challenges are described.

Over the last 5 years, genetic studies have provided major new insights regarding key pathogenic mechanisms underlying Crohn's disease and ulcerative colitis. To date these have mostly been based on genome-wide association studies, but newer genomics technologies are now beginning to complement GWAS findings and add to our understanding of the molecular genetic universe of inflammatory bowel disease [1]. Ultimately the improved understanding of IBD pathogenic mechanisms, including clues regarding environmental factors, can help in the design of improved therapies and development of better preventative strategies in individuals identified as being at risk.

IBD has for many years been recognised to result from a complex interaction of genetic susceptibility with environmental risk factors, producing dysregulation of the mucosal immune system and an inflammatory response targeting the gut flora. Epidemiological studies have highlighted the contribution of smoking and factors which perturb the epithelial barrier and have hinted at the importance of dysbiosis, for example, relating to childhood exposure to antibiotics and the potential importance of enteric infection as a trigger for IBD. Likewise, in the immunological heyday of the 1990s, multiple immune and cytokine pathways were noted to be abnormal in IBD. However, analysis of these entities is difficult—particularly in separating cause from effect in individuals in whom IBD has already developed. An attraction of studying germline genetic variation is the ability to say with certainty what came first.

Family Studies and Genetic Epidemiology

Epidemiological and family studies, in the 1980s and 1990s, provided convincing evidence for a genetic contribution to IBD susceptibility [2–11]. The increased risk can be quantitated by the λ_s ratio, which describes the increased risk to siblings

compared to the background population risk. For Crohn's disease this is 17–35 and for UC is 8–15. By comparison, the λ_s for coronary artery disease is ~3, and for type 1 diabetes (T1D) it is ~15, but with ~50 % of this accounted for by the strong MHC association—early indicators that IBD would prove fertile ground for hunting down non-MHC susceptibility genes. Twin and family studies and their importance in the initial identification of the genetic component of IBD are reported and discussed in "Chapter 2."

Early Molecular Genetic Studies

Early approaches to identifying susceptibility genes for IBD hinged mainly on candidate gene studies. The main success in this regard came with the identification of association between UC and specific alleles of the MHC class 2 region. This was initially identified in Japanese cohorts—Asakura et al. reporting *HLA-DR2* to be associated with UC, with Sugimura et al. subsequently identifying the *DRB1*1502* allele as responsible for this [12, 13]. Association between UC and the MHC was subsequently identified in Europeans, with the *DRB1*0103* allele implicated in both severe UC and extra-intestinal manifestations of IBD [14, 15]. Interestingly, the contribution of the MHC to Crohn's disease susceptibility appears to be modest in Europeans and largely confined to the subgroup of patients with exclusively Crohn's colitis [16]. This illustrates how different CD is to classic autoimmune diseases and supports the clinical impression that colon-only CD may actually be closer to UC than to small bowel CD in terms of its aetiology.

Genome-wide linkage studies using large panels of affected sibling pairs dominated the IBD genetic field in the late 1990s (see Fig. 5.1, [17]). Their yield was relatively modest and reflected the lack of statistical power of this technique for detecting the loci of modest effect size that we now know typifies complex disease. *NOD2* represents the exception, which proved this rule, and was identified within a large linkage interval in the peri-centromeric region of chromosome 16 using a combination of positional cloning and positional candidate gene analysis [6, 8]. Much has been written about NOD2 in the last 10 years, and it will be discussed in "Chapter 10". Suffice to say the discovery of NOD2 provided the first major insight into the critical contribution of innate immunity to CD pathogenesis. NOD2 is an intracellular receptor for muramyl dipeptide, a ubiquitous component of bacterial cell walls. Both the original report from Hugot et al. and subsequent detailed resequencing studies have identified the fact that the CD risk variants cluster in the portion of the gene encoding the "leucine rich region" which recognises and binds muramyl dipeptide. Carriage of a single risk variant confers an odds ratio (OR) of 1.5–3, while in homozygotes the OR rises to 17–40—with the association being almost exclusively with small bowel CD. East Asian populations lack the CD-associated coding variants in *NOD2*—perhaps explaining in part the lower prevalence of CD in East Asia and the relative preponderance of UC. Functional analyses have implicated a variety of mechanisms by which NOD2 variants might predispose to CD—including aberrant activation of NF-κB, altered modulation of

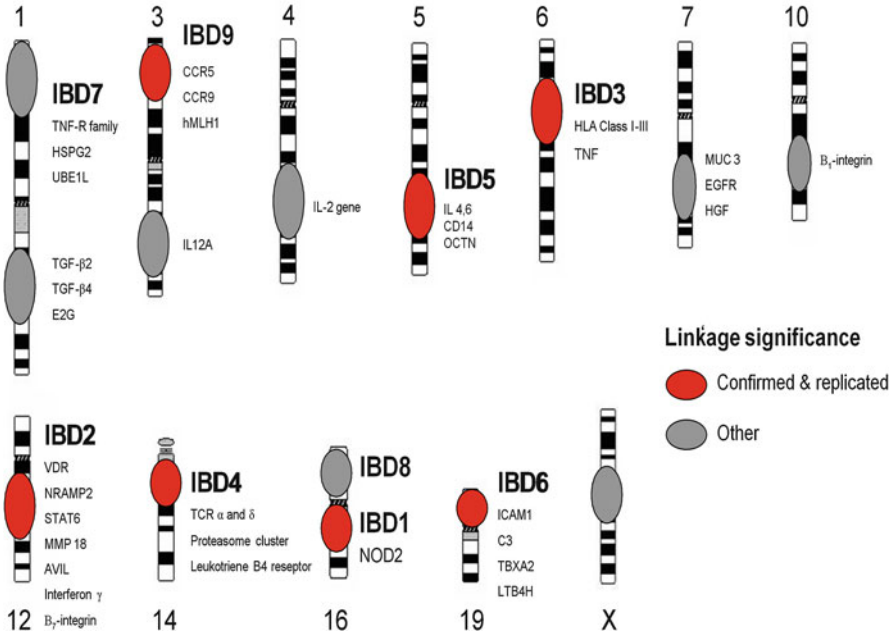

Fig. 5.1 The major focus of genetic research in the mid-1990s was the identification of disease susceptibility loci using GWAS. A large number of markers distributed across the genome are typed in individuals from multiply affected families. Markers that are inherited together with a disease in a family are used to define these regions. Genome-wide scans have identified susceptibility loci on a number of chromosomes. These include *IBD1*, the first and most consistently replicated IBD locus identified on chromosome 16, which is exclusive to CD. This area harbours the *NOD2* gene. Another replicated area of linkage termed *IBD3* is found on chromosome 6. This harbours the HLA region. Modified from Ahmad T, Satsangi J, McGovern D, Bunce M, Jewell DP. Review article: the genetics of inflammatory bowel disease. Alimentary pharmacology & therapeutics. 2001;15(6):731–48. Epub 2001/05/31. With kind permission from Elsevier Limited

TLR signalling, disrupted Paneth cell function with reduced mucosal defensin production and defective autophagy [18]. All highlight the central role played by NOD2 in innate immunity and regulation of the cellular response to bacteria, and it may be precisely because of NOD2's pleiotropic roles in innate immune responses that its mutation exerts such a powerful effect in predisposing to Crohn's disease.

GWAS

Introduction

Genome-wide association study (GWAS) technology allows an unbiased survey of the genome for regions showing association with IBD. By identifying genes, which map within the association intervals, particularly if they are further implicated by,

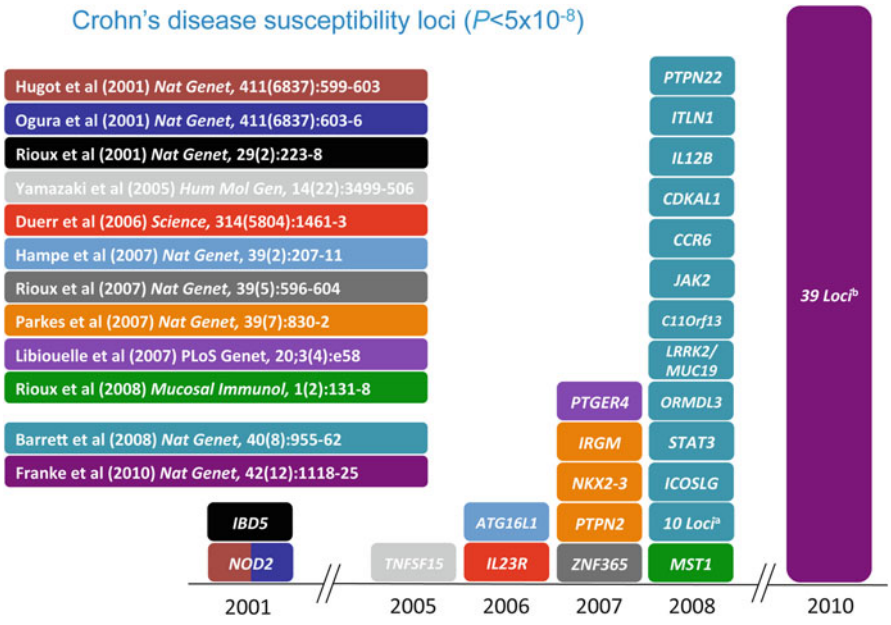

Fig. 5.2 Exponential increase of disease-gene findings for CD

for example, association with coding variants or correlation of genotype with gene expression, inferences can be drawn regarding their primary contribution to IBD pathogenesis. Downstream studies are still required to delineate precise causal variants and explore the allelic spectrum by sequencing and fine mapping, to understand the impact of associated variants on gene expression in relevant cell types and to explore their functional impact on cell biology, immunology and microbial interaction. However, it was GWAS, which have revolutionised our understanding of IBD genetic susceptibility. For more details on the GWAS method, see also "Chapter 4".

GWAS studies have proven successful in IBD on a number of levels—but particularly in terms of the large number of loci identified which show confirmed association with IBD susceptibility, the new pathogenic insights gained and the understanding of the molecular genetic relationship between CD and UC. The index GWAS studies have now been complemented by large GWAS meta-analyses and replication studies from the International IBD Genetics Consortium (IIBDGC). These have included tens of thousands of IBD patients globally and have led to the most recent identification of a total of 163 confirmed IBD susceptibility loci to date [7] (see Fig. 5.2) [19].

Overview of Large-Scale CD Association Studies

Several large-scale association studies (with at least 10,000 SNPs per screen) have been carried out in the last years and enhanced significantly our understanding on the causes of CD. Table 5.1 summarises the study design and most important results

Table 5.1 Overview of large-scale CD association studies

Study	References	Platform	Disease	Ancestry	New loci
Yamazaki et al.	[21]	73 k	CD	Japan	1 (*TNFSF15*)
Duerr et al.	[22]	Illumina 300 k	Ileal CD	North America	1 (*IL23R*)
Hampe et al.	[23]	Non-synonymous SNP 20 k	CD/UC	Germany	1 (*ATG16L1*)
Franke et al.	[60]	Affymetrix [92,387]	CD	Germany	1 (*NELL1*)
Rioux et al.	[24]	Illumina [304,413]	Ileal CD	North America	1 (10q21.1)
Libioulle et al.	[25]	Illumina [302,451]	CD	France/ Belgium	1 (*5p13.1*, upstream of *PTGER4*)
WTCCC	[26]	Affymetrix [469,557]	CD	Europe	9
Parkes et al.	[27]	Affymetrix [469,557]	CD	UK	4
Raelson et al.	[29]	Perlegen [164,279]	CD	Québec founder population, Germany	–
Franke et al.	[41]	Affymetrix [83,360]	CD and SA	Germany	1 (*10p12.2*, shared CD/ SA)
Barrett et al.	[46]	Affymetrix and Illumina [635,547] (imputed)	CD	UK, North America, France/ Belgium	19
Franke et al.	[40]	Affymetrix and Illumina [953,241] (imputed)	CD	Europe	39
McGovern et al.	[32]	Illumina [304,825]	CD	Europe	6
Wang et al.	[44]	Illumina 550 k	CD/UC/T1D	Caucasian	6
Festen et al.	[43]	Affymetrix and Illumina [471,504]	CD and CelD	Europe	7 (shared CD/ Cel)
Ellinghaus et al.	[45]	Affymetrix, Illumina and Perlegen [1,116,213] (imputed)	CD and PS	Europe, North America	1 (SOCS1)
Kenny et al.	[61]	Affymetrix and Illumina [1,060,934] (imputed)	CD	Ashkenazi Jews	5
Kugathasan et al.	[36]	Illumina 550 k	Early-onset IBD	North America, Italy	2 (20q13, 21q22)
Imielinski et al.	[38]	Illumina 550 k	Early-onset IBD	North America, Scotland, Italy	5

of all large association studies published so far. The next sections briefly go into detail by highlighting key studies.

Early Studies

The first ever GWAS was published in 2005, for age-related macular degeneration [20], followed 6 months later by the first GWAS for CD [21]. Here, Yamazaki and colleagues successfully genotyped 72,738 SNPs in 94 Japanese CD patients and 752 healthy control individuals. One thousand eight hundred and eighty-eight significantly associated SNPs were then further tested in a Japanese replication sample. Several SNPs in the *TNFSF15* gene region on chromosome 9q32 were successfully replicated, and a subset of SNPs was then further tested in a study sample from the UK in which the locus was also replicated. The neighbouring *TNFSF8* gene was excluded as a candidate gene by dense SNP fine mapping, which showed that both genes are located on distinct LD blocks. *TNFSF15* encodes the protein TL1A (TNF ligand-related molecule 1A), a tumour necrosis factor (TNF) family member. It is important to note that despite the small sample size (hence low statistical power) and low SNP coverage (more than 500,000 SNPs should be genotyped and then subjected to imputation for adequate genomic coverage according to current standards)—Yamazaki and colleagues identified a novel CD candidate gene that subsequently replicated in both Asian- and Caucasian-descent patients and controls. *TNFSF15* thus represents the first IBD gene to be identified by GWAS, and it is notable that its contribution to disease susceptibility is not restricted to a single ethnic group.

Over a year later, a larger association study was reported for Crohn's disease by an American research group. Duerr and colleagues [22] tested 308,332 SNPs in 547 ileal CD patients (non-Jewish, European ancestry) and 548 healthy controls. Besides the known risk locus *NOD2*, a coding variant in the *IL23R* gene was identified as significantly associated. This coding polymorphism is non-synonymous—that is, it affects the open reading frame of the transcript leading to a change of a single amino acid in the resulting IL23R protein. In the original study, the glutamine variant of the SNP Arg381Gln was found to be significantly less common than the arginine allele, with an allelic frequency of 1.9 % in the non-Jewish patients with ileal CD vs. 7.0 % in non-Jewish controls. The glutamine allele appeared to protect against development of CD in both non-Jewish [OR = 0.26] and Jewish [OR = 0.45] case–control cohorts. Following this, several other groups replicated the Arg381Gln association finding, and additional independent variants were identified at the *IL23R* locus that is associated with both CD and UC.

Also in 2006, a genome-wide candidate SNP association study was published by Hampe et al. [23]. In their study, 7,159 informative non-synonymous SNPs were tested in 735 healthy controls and 368 CD patients. The best-associated 72 SNPs from the screening stage were then tested for association in 380 independent CD trios, 498 CD singleton cases and 1,032 controls. Disease association of rs2241880 in the autophagy-related 16-like 1 gene (*ATG16L1*) was replicated in these samples

and confirmed in a UK panel. By haplotype and regression analysis, the authors found that marker rs2241880, a coding SNP (T300A), carries virtually all the disease risk exerted by the *ATG16L1* locus. This study implicated the autophagy pathway in CD pathophysiology for the first time, with the *ATG16L1* association being replicated later by several other groups [24] (see also "Chapter 12").

The "GWAS Era"

Beginning of 2007, Rioux and colleagues published the first high-density GWAS study in CD with >300,000 SNPs under study, these being examined in 998 ileal CD cases and 1007 healthy controls [24]. This study replicated the *ATG16L1* finding—additionally demonstrating the importance of this protein in the autophagy pathway by several in vitro studies. Recent functional studies in ATG16 hypomorphic mice have shown abnormal Paneth cell morphology and elegantly demonstrated the complex interaction between genetic susceptibility, environmental stressors, intact gut flora and the need for a particular (noroviral) trigger to elaborate the full phenotype of intestinal inflammation—thereby perhaps beginning to approach the complexity seen in human IBD. Two other CD GWAS studies were also published in 2007. A French–Belgian team also used the Illumina 300K SNP array [25] and identified a CD-associated region at *5p13.1*. This localises to a 250 kb linkage disequilibrium (LD) block, which maps to a 1.25 Mb gene desert. Despite the lack of protein-coding genes in this interval, Libioulle et al. nevertheless showed that the Crohn's disease-associated alleles in this gene desert correlate with quantitative expression levels of the prostaglandin receptor EP4, encoded by *PTGER4*. This gene resides closest to the associated region is but still 270 kb away from the most associated SNP. This nicely demonstrates that disease-associated variants can affect regulatory regions, which in consequence can influence the expression of distant genes. This and other related studies led to systematic genome-wide SNP-expression correlation analyses, known as eQTL (expression quantitative trait locus) mapping studies. Although the expression patterns of these eQTL studies were not always measured in the disease-relevant tissue(s) of a particular phenotype—in most cases lymphoblastoid cell line resources were exploited—these analyses significantly increased the knowledge on the influence of genetic variation on gene expression in general, besides generating several plausible hypotheses for noncoding disease-associated variants identified by GWAS.

The UK CD GWAS was also published in 2007 as part of the Wellcome Trust Case Control Consortium [26] and set new standards for association studies. The WTCCC included seven common diseases—amongst them CD with ca. 2000 patients—comprising altogether 14,000 patients and 3,000 shared controls. Besides the statistical and technological advance, this so far largest study for CD identified four novel disease loci (*3p21–22*, *5q33.1*) [*IRGM*], *10q24* [*NKX2-3*] and *18p11* [*PTPN2*]; follow-up described in [27] and replicated the *IL23R*, *NOD2*, *5p13.1*, *5q31*, *ATG16L1* and the previously found intergenic region on *10q21.1* [24]. For the novel *3p21* locus, Goyette and colleagues later showed that the R689C variant

(rs3197999) in *MST1* is the most likely causative variant at this locus [28]. *MST1* encodes macrophage-stimulating protein (MSP), a protein regulating the innate immune responses to bacterial ligands. R689C is predicted to interfere with MSP binding to its receptor, suggesting a role for this gene in the pathogenesis of IBD. Dense fine mapping of the *3p21* locus was also carried out in the study by Raelson et al. [29]. For the novel WTCCC study locus on *5q33.1*, Parkes and colleagues implicated variants at the *IRGM* gene locus as the strongest associated signals in this region [27]. IRGM belongs to the p47 immunity-related GTPase family. Its mouse homologue, LRG-47 (encoded by Lrgm), critically controls intracellular pathogens by autophagy, and Lrgm−/− mice show markedly increased susceptibility to *Toxoplasma gondii* and *Listeria monocytogenes*. Consistent with this, IRGM induces autophagy—another demonstration of the importance of this pathway in CD aetiology—and thereby controls intracellular *Mycobacterium tuberculosis* in human macrophages. McCarroll et al. identified a 20 kb deletion polymorphism upstream of *IRGM* carried by ~40 % of the Caucasian population, which correlates with expression of *IRGM* and which is in complete LD with the neighbouring GWAS lead SNP rs13361189. Whether this structural variant is causal remains a topic of debate, particularly as Prescott et al. subsequently reported the finding of an insertion–deletion ("indel") polymorphism in the 5′UTR of *IRGM* which disrupts a transcription factor-binding site, and most recently Brest et al. reported that an associated synonymous SNP in the coding sequence of *IRGM* alters a microRNA-binding domain, hence affecting mRNA stability and gene translation [30]. Corroborating the expression data suggesting that the Crohn's disease-associated IRGM variants result in reduced gene expression, Lapaquette et al. reported that knockdown of IRGM by siRNA in human macrophages permitted a substantial increase in the number of adherent *E. coli* able to survive within these macrophages and suggested that this effect was specific for the LF82 serotype which this group have demonstrated plays a key role in triggering Crohn's disease [31]. While the complexity inherent in the multiple other genes and loci associated with Crohn's disease susceptibility must not be ignored, this example nicely illustrates one potentially important pathway all the way from associated gene variant to impact on gene expression and functional impact on innate immunity, allowing a recognised environmental agent to exert its IBD-predisposing effect.

One of the most recent GWAS for CD is the study by McGovern and colleagues who analysed >300,000 SNPs in 896 CD cases and 3,204 healthy controls, all of Caucasian descent [32]. Besides replicating 21 previously known loci, they identified suggestive associations with genes involved in tight junctions/epithelial integrity (*ASHL*, *ARPC1A*), innate immunity (*EXOC2*), dendritic cell biology [*CADM1* (*IGSF4*)], macrophage development (*MMD2*), TGF-β signalling (*MAP3K7IP1*) and *FUT2*. The association at the *FUT2* locus was then further replicated and is of particular interest as the gene product is a physiological trait that regulates gastrointestinal mucosal expression of blood group A and B antigens. About 20 % of Caucasians are so-called nonsecretors who do not express ABO antigens in saliva as a result of being homozygous for the nonsense variant of the *FUT2* W134X SNP (rs601338). No excess of heterozygotes in CD were observed compared to controls, which the

authors speculated is in line with the proposed hypothesis that the *FUT2* association is "driven" by an association between nonsecretor status and *FUT2*. The W143X SNP displays evidence of being maintained by strong selective pressure. A large body of evidence suggests that this maintenance may be because of numerous trade-offs surrounding host–microbe interactions. For example, nonsecretors are resistant to infection with the Norwalk (Noro) and respiratory viruses but are more susceptible to duodenal ulcers, rheumatic fever and cholera. Furthermore, the breast milk of secreting mothers provides protection against *Campylobacter jejuni* to their offspring by exploiting the binding affinity of the bacterium to fucosyloligosaccharides. In a follow-up study by Rausch and colleagues [33], it was observed that the *FUT2* genotype explained substantial differences in microbial community composition, diversity and structure in biopsies of the large intestine. They further identified several bacterial species displaying disease-by-genotype associations. These findings indicated that alterations in resident microbial communities may in part be explained by the variety of host susceptibilities surrounding nonsecretor status and that *FUT2* is an important genetic factor influencing host–microbial diversity. Future studies that examine the genotype–microbiome interactions (which seem to play a crucial role in CD aetiology) at genome-wide levels are likely following soon. It has already been demonstrated in mice that a very early interaction between the host and bacteria is necessary to allow a normal response to inflammatory stimuli later in life [34]—supporting the hygiene hypothesis for chronic inflammatory diseases. While the genome seems to partially influence which bacteria will "like" or "dislike" their host, the epigenome seems to be the "mediator" and the "memory" of the various interactions (for more details on host–microbe interactions, see "Chapter 14").

GWAS for Early-Onset CD

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. Implicit in this strategy is the assumption that these patients represent a more severe, more genetically influenced group of affected individuals. 15–20 % of IBD patients present in childhood or adolescence with epidemiological and natural history studies clearly demonstrating a rising incidence in this age group. Although early-onset disease is characterised by particular phenotypic features, such as more extensive disease at onset and rapid progression, two recent genome-wide association studies (GWAS) carried out exclusively in this age group have demonstrated marked genetic similarities to adult disease [35].

In the first study, which used a case–control panel that was a subset of that used in the second, Kugathasan et al. performed a GWAS using DNA from 1,011 individuals with paediatric-onset IBD (647 CD and 317 UC) and 4,250 matched controls [36]. They replicated several known loci from non-paediatric association studies (*NOD2*, *IL23R*, *HLA*, *TNFSF15*) and identified two novel disease-associated loci, *20q13* and *21q22*. Although the authors were unable to pinpoint the causal gene in the *20q13* region, they considered the *TNFRSF6B* gene the most compelling

candidate based on the critical role of specific polymorphisms within genes involved in the TNF pathway in the pathogenesis of IBD. It is of interest that the protein product for *TNFRSF6B* acts as a decoy receptor (protein called DCR3) in preventing FasL-induced cell death, and a resistance to FasL-dependent apoptosis has previously been shown for T lymphocytes in CD. The authors also observed that the mean serum DCR3 concentration was significantly increased in individuals with IBD carrying the major allelic variants compared to IBD carrying the minor allelic variants. The *21q22* signal resides in a small region of LD that harbours no genes, with the nearest gene being *PSMG1* (proteasome assembly chaperone 1). A Canadian study for early-onset CD (410 patients) later replicated the *20q13* but not the *21q22* finding [37].

In the second GWAS for early-onset IBD, Imielinski and colleagues [38] analysed 550,000 SNPs in 3,426 affected individuals (1,636 CD, 724 UC, 53 unclassified) and 11,963 genetically matched controls, recruited through international collaborations in Europe and North America. The authors identified five new regions associated with early-onset IBD susceptibility, including *16p11* near the cytokine gene *IL27* (rs8049439), *22q12* (rs2412973), *10q22* (rs1250550), *2q37* (rs4676410) and *19q13.11* (rs10500264). The scan also detected associations at 23 of 32 loci previously implicated in adult-onset CD and at 8 of 17 loci implicated in adult-onset UC, highlighting the close pathogenetic relationship between early- and adult-onset IBD.

Essers and colleagues later demonstrated, by analysing 35 common established CD susceptibility loci in early-onset patients (average age of onset of 11.7 years), that paediatric patients do not carry significantly more risk alleles than adult CD patients [39]. This does not exclude the hypothesis though that adult and/or paediatric patients carry significantly different risk variants which are yet unknown. Future large-scale resequencing studies will most likely clarify this hypothesis. As described elsewhere in the text, a small handful of studies have already identified monogenic forms of severe early-onset inflammatory bowel disease. It should be noted here that out of the aforementioned two and five novel loci identified by Kugathasan et al. and Imielinski et al., respectively, none and two were genome-wide significant in the below-mentioned meta-analysis for CD [40], including mostly adult-onset patients, respectively.

Combined Analyses with Other Phenotypes

As predicted from their close clinical relationship, many key susceptibility loci are shared between Crohn's disease and ulcerative colitis. However, a less expected finding prior to the GWAS era was the extent to which overlap would be seen for multiple loci across many immune-mediated diseases (discussed in more detail in "Chapter 7").

Given this overlap, one topic which continues to be the source of significant debate is what statistical thresholds are appropriate for "claiming" association between a given locus and, for example, CD if the same locus has already demonstrated genome-wide significant association with UC or indeed with another

immune-mediated disease. Using the conventional genome-wide significance threshold of $p < 5 \times 10^{-8}$ is clearly overly conservative given the markedly high prior probability for association. This issue has been addressed in more detail in the recent report of the analysis of Immunochip data from the IIBDGC [7].

Some of the overlap between immune-mediated diseases reflects known co-morbidities—for example, Crohn's disease and psoriasis, both appearing often in the same patient, share a significant portion of their genetic risk map. Therefore, in some instances, researchers have combined more than one disease in GWAS to identify such shared risk genes or to search for differences. To this end, CD GWAS data sets were combined with GWAS for sarcoidosis [41]; one novel shared locus on *10p12.2*, replicated recently in [42]; celiac disease [43]; *TAGAP* and *PUS10* as novel shared loci, T1D and UC [44]; e.g., T1D risk loci HLA, *PTPN22*, *IL27*, *IL18RAP* and *IL10* are protective for CD and psoriasis [45]; seven shared non-HLA loci plus *SOCS1* on *16p13* as a novel CD risk locus. Ongoing research efforts of the "Immunochip Consortium" are aiming at combining and jointly analysing GWAS data sets for most known autoimmune diseases, an effort that will enlarge the genetic risk map even further.

Meta-Analyses of CD GWAS Studies

While the index GWAS studies were able to identify loci conferring (in complex disease genetic terms) larger effect sizes, they were, in retrospect, underpowered to detect the many more loci that confer an OR of disease of <1.2. Reliable identification of such loci requires analysis of substantially larger sample sets. This becomes possible with subsequent work from the IIBDGC which has conducted two meta-analyses of CD GWAS studies [40, 46] and undertaken a large collaborative experiment using the Immunochip [7]. Each of these studies has identified multiple new CD susceptibility loci of progressively smaller effect size. While critics might argue that these loci exert such a weak effect that their impact on CD pathogenesis is negligible, a more rounded view is that these loci very much help to "join the dots", aiding informatics analyses such as GRAIL in identifying causal genes and helping to define entire pathways where index GWAS studies highlighted just single components. The latter is particularly important where proteins participate in several distinct pathways or where their function has previously been only partially elucidated—such as had been the case for *IRGM*, which some authorities dismissed as a pseudogene until the evidence from CD GWAS corroborated its putative function in autophagy. An additional and potentially important point is that the biological impact may be out of proportion to the strength of the genetic association signal for a particular variant. This is illustrated by the modest association between HMG-CoA reductase gene variants and hypercholesterolaemia, and yet this gene product represents the target of a class of drugs called statins, one of the most effective therapies for reducing cholesterol population-wide.

A detailed discussion of all the CD genes and loci identified in these three IIBDGC studies—amounting to in excess of 100 independent CD susceptibility

loci—is beyond the scope of this chapter, and interested readers are referred to the original publications for all details [40, 47]. Below we focus on some of the key themes and key pathways identified.

The Franke et al. meta-analysis published in 2010 comprised GWAS data on 6,333 cases and 15,056 controls and followed up the top association signals in 15,694 cases, 14,026 controls and 414 parent–offspring trios. It identified 30 new susceptibility loci meeting genome-wide significance. 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*.

These meta-analyses have demonstrated association between CD and an ever-increasing number of loci encoding IL23/Th17 pathway components, such as *TYK2*, *JAK2*, *STAT3*, *ICOSLG* and *CCR6*. Intuitively the IBD-associated variants in *IL23R* might be predicted to exert their effect on adaptive immunity via CD4+ and Th17 pathways. There has been increasing evidence for a major role of Th17 cells in IBD pathogenesis in recent years, significantly spurred by the genetic evidence. However, caution is required before jumping to this conclusion as polymorphisms in *IL23* pathway genes may also impact innate immunity. For example, Buonocore et al. recently reported the accumulation of CD3 negative, IL23-responsive innate lymphoid cells in the colon, these being capable of producing IL17 and IFN-γ and mediating innate colitis in mice [48]. Production of Th17 cytokines by analogous cells in humans appeared higher in colons from IBD cases vs. controls. Additional functional interrogation is required to explore the role of Th17 pathways in mediating mucosal homeostasis and microbial interaction, particularly in light of a recent clinical study demonstrating that anti-IL17 antibody therapy leads to worse clinical outcomes than placebo when trialled in Crohn's disease.

The region encoding interleukin-10 was also found to be associated with CD in the Franke et al. meta-analysis. This locus was originally identified as associated with IBD in a German GWAS in ulcerative colitis, association being documented with noncoding variants upstream of the *IL10* gene [49]. Additional interest in this pathway derived from a study, which identified mutations in the interleukin-10 receptor as causing an extreme form of IBD in infants. The latter study, which was based on analysis of two consanguineous families by exome sequencing, identified homozygous mutations in both the *IL10RA* and *IL10RB* genes as abrogating interleukin-10 signalling and leading to severe intestinal inflammation [50, 51]. IL-10 is known to play an important regulatory role in immune homeostasis, and *IL10* knockout mice represent one of the best animal models of IBD (see also "Chapter 3"). Attempts at manipulating IL-10 signalling for therapeutic benefit in IBD have so far been unsuccessful—but this probably reflects the need for a better understanding of the functional impact of the disease-predisposing variants and improved timing and targeting of any IL-10-based treatment.

Another exome-sequencing study this time in a large panel of IBD cases identified association, with a rare variant in the *CARD9* gene. *CARD9* mediates signalling between pattern recognition receptors such as NOD2 and the pro-inflammatory transcription factor NF-κB, and the genomic interval in which it lies was originally

implicated in the Barrett et al. CD meta-analyses. Here, the lead SNP was a common noncoding variant associated with increased disease risk, and in eQTLs data sets, this SNP was found to correlate with strongly increased expression of the *CARD9* gene. The subsequent exome-sequencing study by Rivas et al. neatly mirrored this by identifying a novel low-frequency splice site variant in *CARD9* which impairs the function of the protein and is associated with reduced disease risk [52]. From these findings, we can see that the genetic universe of IBD encompasses the whole allele frequency spectrum from common variants, identified in GWAS studies, low-frequency variants, identified by targeted exome sequencing in large panels, and private mutations identified by whole exome sequencing of extreme cases in families, as exemplified by the *IL10* receptor mutations.

SMAD3 showed association with CD in the Franke et al. meta-analysis. Phosphorylated following TGF-β signalling through its receptor, the SMAD3 protein complexes with SMAD4 and is then translocated to the nucleus to modulate target gene expression and exert broadly immunosuppressive effects. Involvement of TGF-β signalling pathways in CD pathogenesis has recently been corroborated by the findings from Immunochip, which identify association with several other components of the TGF-β pathway including SMAD7, SMURF1 and FURIN [7]. These all play important roles in the TGF-β-mediated induction of Foxp3+ regulatory T cells, with SMAD3 deficiency reciprocally enhancing Th17. TGF-β is also critically involved in epithelial restitution in the gut, where its potent pro-fibrogenic effects may be relevant both to mucosal repair and intestinal stricture formation which is a hallmark of Crohn's disease [53].

Association between CD and variants at the *ERAP2* locus is intriguing, particularly as the CD-associated variant correlates strongly with *ERAP2* gene expression in published eQTL data sets. Regulated by NF-κB, *ERAP2* encodes one of two human endoplasmic reticulum aminopeptidases, which work in concert to trim peptides for presentation on MHC class I and hence critically affect antigen presentation to T cells. Ankylosing spondylitis and psoriasis are also associated with this locus, but with a pattern of associated variants more closely implicating *ERAP1*. Given the close clinical relationship between Crohn's disease and ankylosing spondylitis, and the strong association of HLA-B27 with the latter but not with the former, the divergent association of these closely related molecules is intriguing and will refocus interest on the MHC class I associations in Crohn's disease.

As already indicated above, the NF-κB pathway is another for which multiple components show genetic association with CD susceptibility. NF-κB is a master transcriptional regulator of multiple cytokine genes involved in the inflammatory response and controls epithelial integrity and mucosal immune homeostasis in the presence of gut microflora [54]. Association with the *TNFAIP3* gene locus was first seen in Wang et al. comparative GWAS analysis [44] and is now recognised as a pleiotropic locus associated with multiple other immune-mediated diseases including rheumatoid arthritis, SLE, celiac disease, T1D and psoriasis. This gene encodes the ubiquitin-editing protein A20 protein, a TNF-α-inducible zinc finger protein thought to limit NF-κB-mediated immune responses. Recent data from the

Immunochip experiment have identified a number of other key constituents of the NF-κB pathway as being associated with CD, including *REL* on chromosome 2, *RELA* on chromosome 11 and *NF-κB* itself on chromosome 4.

The first meta-analyses of CD and UC GWAS studies undertaken by IIBDGC treated each form of IBD as a separate phenotype. Even from these analyses, which used overly conservative thresholds for association in view of the higher "priors" for loci meeting genome-wide significance in one phenotype, it was evident that multiple susceptibility loci were shared between the two phenotypes [40, 47]. This applied, for example, to multiple components of the Th17 pathway, namely, CARD9, IL-10 and ICOS ligand as well as many others. Given the phenotypic overlap between CD and UC, such overlap is not surprising. This issue has been evaluated more thoroughly in the analysis of Immunochip data by the IIBDGC, in which it is clear that over 100 of the 163 loci which are associated at genome-wide significance with some form of IBD contribute to both phenotypes [7]—albeit in some instances the association with one form is stronger than with the other. As interesting, and perhaps even more informative with regard to specific pathogenic mechanisms, is the identification of loci, which appear disease-specific. Par excellence this applies to *NOD2* but also to *ATG16L1*, both of which are associated exclusively with CD—indeed *NOD2* variants are actually modestly protective against UC. Evidently the innate immune mechanisms of processing of intracellular bacteria mediated by the protein products of these genes critically influence IBD phenotype. Corroborating this, functional studies of both have implicated a variety of potentially pathogenic mechanisms, including processing of bacterial antigens and disruption of Paneth cells, which are key mediator of innate immunity in the ileum.

It is intriguing that variants in genes linked to epithelial barrier function seem to be specifically associated with UC and not Crohn's disease—the converse of *NOD2* and the autophagy genes. While many complex explanations might exist, these observations correlate nicely with UC being confined to the superficial layers of the colon, while the transmural inflammation of Crohn's disease is caused by defects in cellular innate immunity and bacterial handling in the deeper layers of the lamina propria and beyond.

Future Challenges

To date based on loci meeting genome-wide significance thresholds, approximately 25 % of the heritability of CD has been accounted for. This proportion can be increased significantly based on reanalysis of GWAS data if one accounts for the many hundreds of additional common variants that individually contribute only a tiny fraction of overall variance of disease risk [55]. Nonetheless, accounting for the additional "missing heritability", or at least enough of it to allow substantial progress towards understanding pathogenic mechanisms in CD, is a significant challenge to the IBD genetic community, and various approaches are being adopted.

One of the most important of these is rare variant studies—attempting to identify (possibly highly penetrant) low-frequency and rare variants associated with CD susceptibility. It is already apparent from the relatively limited sequencing efforts conducted to date that rare variants, for example, within GWAS loci, contribute independently to disease risk. The question is whether and to what extent this applies more broadly to loci, which lie outside GWAS intervals and which within multiply affected families make a significant contribution to disease risk. The analytic and logistic challenges are substantial, but high-throughput whole-genome sequencing efforts are underway to identify new variants to take forward into large-scale association studies.

Additional genomics methodologies are also being applied, increasingly at a genome-wide level. This applies to expression analyses and their correlation with germline genetic variation. Such studies are particularly helpful when undertaken in separated cells rather than homogenised tissue or mixed cell populations, due to the now good evidence that gene-regulatory mechanisms are often cell-type specific [56]. A challenge to the IBD community is to ensure that such analyses are undertaken in cell types of potential relevance to CD—including leukocyte subsets and intestinal epithelial cells, but not forgetting less "obvious" candidates such as stromal cells and stem cells.

Genome-wide epigenetic analyses are also increasingly being applied [57–59] to better understand how core regulatory mechanisms such as methylation, histone binding, chromatin remodelling and microRNAs might affect CD susceptibility. The same issues regarding the need to conduct such studies in separated cell types pertain, but efforts are now underway to tackle the logistic challenges involved and understand the interaction between germline variation and environmental factors (infection, gut microbiota, smoking and diet) with such epigenetic mechanisms.

A major aspiration for genetic studies is that they will provide substantial insights regarding environmental triggers and drivers of CD. This is already playing out, and the overlap between susceptibility loci for CD and mycobacterial infection has been strongly highlighted in the Immunochip analysis [7]. Genomic and particularly sequencing technologies are increasingly being used to characterise the microbiota in CD vs. health and seek bacterial and viral triggers for IBD.

As the genetic story unfolds, so interest will refocus on the functional impact of the implicated genetic mechanisms and the extent to which they can be manipulated or subjugated to therapeutic benefit. Preventative strategies may also become relevant at the point that we are better able to identify, by prediction algorithms, those at significant absolute risk of developing disease. Further goals commensurate with clinical translation include detailed pharmacogenetic studies and prognostic modelling to separate patients destined to run a severe disease course from those less likely to have trouble. Clinical translation must remain the ultimate goal for the CD genetic community, and with the exciting progress that has been made over the last 5 years in understanding the pathogenesis of CD, the hope is that CD will be one of the first common diseases that can be cured.

References

1. Vavricka SR, Schoepfer AM, Bansky G, Binek J, Felley C, Geyer M et al (2011) Efficacy and safety of certolizumab pegol in an unselected Crohn's disease population: 26-week data of the FACTS II survey. Inflamm Bowel Dis 17(7):1530–1539
2. Bayless TM, Tokayer AZ, Polito JM II, Quaskey SA, Mellits ED, Harris ML (1996) Crohn's disease: concordance for site and clinical type in affected family members—potential hereditary influences. Gastroenterology 111(3):573–579, Epub 1996/09/01
3. Colombel JF, Grandbastien B, Gower-Rousseau C, Plegat S, Evrard JP, Dupas JL et al (1996) Clinical characteristics of Crohn's disease in 72 families. Gastroenterology 111(3):604–607, Epub 1996/09/01
4. Farmer RG, Michener WM, Mortimer EA (1980) Studies of family history among patients with inflammatory bowel disease. Clin Gastroenterol 9(2):271–277
5. Halfvarson J (2011) Genetics in twins with Crohn's disease: less pronounced than previously believed? Inflamm Bowel Dis 17(1):6–12, Epub 2010/09/18
6. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J et al (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411(6837):599–603, Epub 2001/06/01
7. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY et al (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491(7422):119–124, Epub 2012/11/07
8. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R et al (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411(6837): 603–606, Epub 2001/06/01
9. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen TI, Binder V (1991) Familial occurrence of inflammatory bowel disease. N Engl J Med 324(2):84–88, Epub 1991/01/10
10. Peeters M, Nevens H, Baert F, Hiele M, de Meyer AM, Vlietinck R et al (1996) Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. Gastroenterology 111(3):597–603, Epub 1996/09/01
11. Satsangi J, Grootscholten C, Holt H, Jewell DP (1996) Clinical patterns of familial inflammatory bowel disease. Gut 38(5):738–741, Epub 1996/05/01
12. Asakura H, Tsuchiya M, Aiso S, Watanabe M, Kobayashi K, Hibi T et al (1982) Association of the human lymphocyte-DR2 antigen with Japanese ulcerative colitis. Gastroenterology 82(3):413–418, Epub 1982/03/01
13. Sugimura K, Asakura H, Mizuki N, Inoue M, Hibi T, Yagita A et al (1993) Analysis of genes within the HLA region affecting susceptibility to ulcerative colitis. Hum Immunol 36(2):112–118, Epub 1993/02/01
14. Orchard TR, Chua CN, Ahmad T, Cheng H, Welsh KI, Jewell DP (2002) Uveitis and erythema nodosum in inflammatory bowel disease: clinical features and the role of HLA genes. Gastroenterology 123(3):714–718, Epub 2002/08/29
15. Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI et al (1996) Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet 347(9010):1212–1217, Epub 1996/05/04
16. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ et al (2008) Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 40(6):710–712, Epub 2008/04/29
17. Ahmad T, Tamboli CP, Jewell D, Colombel JF (2004) Clinical relevance of advances in genetics and pharmacogenetics of IBD. Gastroenterology 126(6):1533–1549, Epub 2004/05/29
18. Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M (2005) Genetics of Crohn disease, an archetypal inflammatory barrier disease. Nat Rev Genet 6(5):376–388, Epub 2005/04/30
19. http://www.ibdgenetics.org/
20. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C et al (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308(5720):385–389, Epub 2005/03/12

21. Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D et al (2005) Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. Hum Mol Genet 14(22):3499–3506, Epub 2005/10/14
22. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ et al (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 314(5804):1461–1463, Epub 2006/10/28
23. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K et al (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 39(2):207–211, Epub 2007/01/04
24. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A et al (2007) Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 39(5):596–604, Epub 2007/04/17
25. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D et al (2007) Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet 3(4):e58
26. WTCCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661–678, Epub 2007/06/08
27. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA et al (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 39(7):830–832, Epub 2007/06/08
28. Goyette P, Lefebvre C, Ng A, Brant SR, Cho JH, Duerr RH et al (2008) Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis. Mucosal Immunol 1(2):131–138, Epub 2008/12/17
29. Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P et al (2007) Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. Proc Natl Acad Sci U S A 104(37):14747–14752, Epub 2007/09/07
30. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V et al (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat Genet 43(3):242–245, Epub 2011/02/01
31. Lapaquette P, Bringer MA, Darfeuille-Michaud A (2012) Defects in autophagy favour adherent-invasive Escherichia coli persistence within macrophages leading to increased pro-inflammatory response. Cell Microbiol 14(6):791–807, Epub 2012/02/09
32. McGovern DP, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M et al (2010) Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Hum Mol Genet 19(17):3468–3476, Epub 2010/06/24
33. Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S et al (2011) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc Natl Acad Sci U S A. 2011 Nov 22;108(47):19030–19035. doi: 10.1073/pnas.1106408108, Epub 2011/11/08
34. Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A et al (2012) Microbial exposure during early life has persistent effects on natural killer T cell function. Science 336(6080): 489–493, Epub 2012/03/24
35. Henderson P, van Limbergen JE, Wilson DC, Satsangi J, Russell RK (2011) Genetics of childhood-onset inflammatory bowel disease. Inflamm Bowel Dis 17(1):346–361, Epub 2010/09/15
36. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL et al (2008) Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet 40(10):1211–1215, Epub 2008/09/02
37. Amre DK, Mack DR, Morgan K, Fujiwara M, Israel D, Deslandres C et al (2009) Investigation of reported associations between the 20q13 and 21q22 loci and pediatric-onset Crohn's disease in Canadian children. Am J Gastroenterol 104(11):2824–2828, Epub 2009/07/23

38. Imielinski M, Baldassano RN, Griffiths A, Russell RK, Annese V, Dubinsky M et al (2009) Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet 41(12):1335–1340, Epub 2009/11/17
39. Essers JB, Lee JJ, Kugathasan S, Stevens CR, Grand RJ, Daly MJ (2009) Established genetic risk factors do not distinguish early and later onset Crohn's disease. Inflamm Bowel Dis 15(10):1508–1514, Epub 2009/03/27
40. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T et al (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet 42(12):1118–1125, Epub 2010/11/26
41. Franke A, Fischer A, Nothnagel M, Becker C, Grabe N, Till A et al (2008) Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. Gastroenterology 135(4):1207–1215
42. Cozier YC, Ruiz-Narvaez EA, McKinnon CJ, Berman JS, Rosenberg L, Palmer JR (2012) Fine-mapping in African-American women confirms the importance of the 10p12 locus to sarcoidosis. Genes Immun 13(7):573–578, Epub 2012/09/14
43. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G et al (2011) A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. PLoS Genet 7(1):e1001283, Epub 2011/02/08
44. Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S et al (2010) Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Hum Mol Genet 19(10):2059–2067, Epub 2010/02/24
45. Ellinghaus D, Ellinghaus E, Nair RP, Stuart PE, Esko T, Metspalu A et al (2012) Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. Am J Hum Genet 90(4):636–647, Epub 2012/04/10
46. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD et al (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40(8):955–962, Epub 2008/07/01
47. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD et al (2011) Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet 43(3):246–252, Epub 2011/02/08
48. Buonocore S, Ahern PP, Uhlig HH, Ivanov II, Littman DR, Maloy KJ et al (2010) Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. Nature 464(7293): 1371–1375, Epub 2010/04/16
49. Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G et al (2008) Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 40(11):1319–1323, Epub 2008/10/07
50. Glocker EO, Frede N, Perro M, Sebire N, Elawad M, Shah N et al (2010) Infant colitis—it's in the genes. Lancet 376(9748):1272, Epub 2010/10/12
51. Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schaffer AA, Noyan F et al (2009) Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N Engl J Med 361(21): 2033–2045, Epub 2009/11/06
52. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK et al (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43(11):1066–1073, Epub 2011/10/11
53. Di Sabatino A, Jackson CL, Pickard KM, Buckley M, Rovedatti L, Leakey NA et al (2009) Transforming growth factor beta signalling and matrix metalloproteinases in the mucosa overlying Crohn's disease strictures. Gut 58(6):777–789, Epub 2009/02/10
54. Nenci A, Becker C, Wullaert A, Gareus R, van Loo G, Danese S et al (2007) Epithelial NEMO links innate immunity to chronic intestinal inflammation. Nature 446(7135):557–561, Epub 2007/03/16
55. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90(1):7–24, Epub 2012/01/17

56. Fu J, Wolfs MG, Deelen P, Westra HJ, Fehrmann RS, Te Meerman GJ et al (2012) Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet 8(1):e1002431, Epub 2012/01/26
57. Cooke J, Zhang H, Greger L, Silva AL, Massey D, Dawson C et al (2012) Mucosal genome-wide methylation changes in inflammatory bowel disease. Inflamm Bowel Dis 18(11):2128–2137, Epub 2012/03/16
58. Hasler R, Feng Z, Backdahl L, Spehlmann ME, Franke A, Teschendorff A et al (2012) A functional methylome map of ulcerative colitis. Genome Res 22(11):2130–2137, Epub 2012/07/25
59. Nimmo ER, Prendergast JG, Aldhous MC, Kennedy NA, Henderson P, Drummond HE et al (2012) Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. Inflamm Bowel Dis 18(5):889–899, Epub 2011/10/25
60. Franke A, Hampe J, Rosenstiel P, Becker C, Wagner F, Hasler R et al (2007) Systematic association mapping identifies NELL1 as a novel IBD disease gene. PLoS One 2(8):e691, Epub 2007/08/09
61. Kenny EE, Pe'er I, Karban A, Ozelius L, Mitchell AA, Ng SM et al (2012) A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. PLoS Genet 8(3):e1002559, Epub 2012/03/14