

Chapter 6

Genetics of Ulcerative Colitis

Mauro D'Amato, Catherine Labbé, and John D. Rioux

Abstract While it has been known for several decades that susceptibility to ulcerative colitis (UC), one of the two major forms of inflammatory bowel disease (IBD), is in part inherited, it is only recently that considerable progress has been made in identifying some of the key genetic determinants playing a role in predisposition to disease. The advent of genome-wide association studies (GWAS) has provided geneticists with an extremely powerful tool for the identification of causative genes in complex diseases. To date, 163 IBD risk loci have been identified, and functional insight obtained from genes mapping to these associated regions has highlighted the importance of immune function and gut interaction with bacteria, as well as the maintenance of epithelial barrier in UC pathogenesis. The functional characterization of the remaining genes, as well as the identification of rare causative variants in these associated regions, represents a considerable challenge for the future, but promises to unravel novel biological pathways that may be amenable to exploitation for the design of innovative therapeutic strategies in UC.

Ulcerative colitis (UC) was described for the first time in 1859 by British physician Sir Samuel Wilks [1]. One hundred and fifty years of scientific studies and careful observation of its clinical manifestations and yet the pathogenesis of this debilitating inflammatory bowel disease (IBD) remains, for a large part, a mystery. Only in

M. D'Amato, Ph.D. (✉)

Department of Biosciences and Nutrition, NOVUM, Karolinska Institutet,
Hälsövägen 7-9, 14183 Stockholm, Sweden
e-mail: Mauro.damato@ki.se

C. Labbé, M.Sc., Ph.D.

Montreal Heart Institute, Montréal, QC, Canada

Université de Montréal, Montréal, QC, Canada

J.D. Rioux

Université de Montréal and the Montreal Heart Institute, Montréal, QC, Canada

the last decade have tremendous improvements been made in the knowledge of the molecular pathways underlying disease onset and chronic inflammation. These are primarily attributable to large-scale genetic studies that aimed to identify genetic risk factors in UC and led to the discovery of specific genes and associated biological pathways contributing to disease susceptibility.

Early UC genetic studies focussed on the major histocompatibility complex (MHC) located on chromosome 6p; linkage and association to this locus represented the first discoveries of genetic susceptibility to UC [2]. Following this came a series of mostly unsuccessful targeted association studies, unsuccessful owing to limited power to detect significant association due to the number of markers and the size cohorts tested. The technical innovations and discoveries that followed the publication of the complete sequence of the human genome set the stage for the advent of genome-wide association studies (GWAS) [3, 4]. Today, recent GWAS, meta-analyses of GWAS and replication studies testing very large patient cohorts obtained through extensive international collaborations have identified 133 genomic regions associated to UC susceptibility, which cumulatively contribute to explain 7.5 % of disease variance [5–7].

The MHC and UC Susceptibility

In 1967, the MHC region was first identified to be involved in immune-mediated disease when the expression of HLA-B antigens was found to be higher in patients with Hodgkin's lymphoma [8]. Variations in the MHC locus have since then been associated to autoimmune, inflammatory and also infectious disease, confirming this locus to be of paramount importance in immunity. Traditionally, the MHC is split into class I, class II and class III regions, each class containing multiple genes, including human leukocyte antigen (HLA)-encoding genes as well as non-HLA-encoding genes. The sequencing of the 3.44 Mb MHC region identified 224 different genes [9]. Class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II (*HLA-DP*, *HLA-DQ*, *HLA-DR*) genes are involved in antigen presentation and stimulation of T helper cells (T_H cells). Class III region genes encode several proteins with immune functions, including components of the complement system, cytokines and heat shock proteins. The MHC region exhibits striking sequence variability; for example, the *HLA-B* gene is the most polymorphic gene in the human genome, with over 2,000 alleles already identified in different populations [10]. Another particular feature of the MHC region is the presence of haplotypes with very strong and extensive linkage disequilibrium (LD), which often limit the exact localization of causative association signals in classical genetic studies.

Since the 1970s, several linkage and association studies have tried to verify whether specific HLA alleles are implicated in UC susceptibility. More often than not, these early studies produced conflicting results. This can mostly be explained by inadequate coverage of genetic variability in the region, by different typing methodologies and by low statistical power owing to limited cohort sizes. In 1999,

Stokkers et al. published a meta-analysis of several previously reported association studies of HLA alleles, in order to estimate the contribution of the HLA class II alleles to IBD. Specifically, they analyzed 18 association studies of the MHC region in UC [2]. In the pooled results, HLA-DR2, HLA-DR9 and HLA-DRB1*0103 were associated with UC risk, while a protective effect of allele HLA-DR4 was detected. A similar effort published by Fernando et al. a decade later aimed to review 35 years of MHC genetic research and pool the results in an updated meta-analysis [11]. This particular study examined the association of the MHC alleles to six different diseases, and 37 independent studies were considered for UC [2]. This new meta-analysis confirmed the association of UC to HLA-DRB1*0103 and HLA-DRB1*1502 and detected a weak association to the MHC class I related gene A (*MICA*), which had been previously reported [12]. The study identified novel associations to HLA-DR5, HLA-A19 and HLA-A24.

In 2009, the International MHC and Autoimmunity Genetics Network (IMAGEN) published an extensive association study of the MHC region [13]. A total of 1,472 single nucleotide polymorphisms (SNPs) covering common MHC variation were genotyped in over 10,000 DNA samples from cohorts of patients suffering from one of seven immune-related diseases. The cohort for UC consisted of 667 patients from Italy, and three different signals associated to increased UC risk were identified, namely, in the *HLA-DRB1* gene (with HLA-DRB1*1101 possibly responsible for the association signal), the *NOTCH4* gene region and a third locus containing the *BAT8*, *C2*, *RDBP* and *SKIV2L* genes.

As of today, the MHC is still one of the strongest and most consistently replicated association signals in UC. This remains the case even within the context of the more recent genome-wide SNP-based association studies. One of the first in UC was a genome-wide scan of non-synonymous SNPs from Fisher et al. [14]. This study identified a 400-kb haplotype block containing *BTNL2* as well as loci *HLA-DQA1*, *HLA-DRA*, *HLA-DRB5* and *HLA-DRB1*. The first true GWAS (>100,000 SNPs) of UC to find an association to the MHC was of IBD patients with early-onset disease [15]. Association to the MHC has since been detected in every subsequent GWAS and meta-analyses of GWAS. However, as for many other immune-mediated diseases with association in the MHC, the complexity of the region makes it difficult to pinpoint the actual causal gene(s) or variant(s) for UC.

Limited Success of Early Targeted Associated Studies

Before the GWAS era, association studies to candidate genes outside the MHC were attempted, but they generally resulted in limited success and no locus was consistently replicated. These early association studies were generally performed in small cohorts from different populations and ethnic groups, and the coverage of genetic variability in the queried genes or genomic regions was often low. Two loci that stand out, however, are the *ABCB1* locus and *IL2-IL21* locus as multiple studies have shown consistent results.

The ATP-binding cassette subfamily B member 1 gene (*ABCB1*), also known as the multidrug resistance gene 1 (*MDR1*), encodes P-glycoprotein 170. This transmembrane protein is an efflux transporter pump that regulates the flow of substances in and out of the cell. It is highly expressed in epithelial cells of the intestinal tract and influences the response to many drugs, including glucocorticoids used in the treatment of IBD [16]. *Abcb1* knockout mice develop severe spontaneous colitis that can be cured with antibiotics, which suggested that intestinal flora is necessary to maintain inflammation in mouse models of IBD [17]. Several candidate gene studies from the last decade have confirmed that genetic variation in the *ABCB1* gene is associated with UC. While a meta-analysis of six targeted association studies published until 2006 concluded that the T allele of *ABCB1* SNP rs1045642 is significantly associated with UC risk [18], a more recent meta-analysis of 16 association studies suggested the association with this allele is only marginal [19]. In addition, this locus was not among the 163 genetic risk factors identified in the most recent GWAS of IBD, suggesting that if truly associated with UC, its effect is likely to be quite modest [6].

Before being described in UC, the locus 4q27 had already been characterized as a well-established immune-related locus, because of its association with celiac disease, type 1 diabetes, Grave's disease, SLE, rheumatoid arthritis and psoriasis [20–24]. This locus contains the cytokine *IL2* and *IL21* genes, which are both functionally interesting candidates. *IL21* is involved in the differentiation of T_H17 cells implicated in the pathogenesis of IBD [25], and its expression is increased in intestinal tissue from UC patients compared to controls [26], while *il2* knockout mice develop an IBD-like condition [27]. The first study to implicate the *IL2-IL21* locus in IBD susceptibility was published by Festen et al. [28], who reported the genotyping of four SNPs in a three-stage association study on a total of 3,195 UC cases from North America, the Netherlands and Italy. Association signals at all four SNP sites reached genome-wide significance. The identified locus covers a 480 kb region of high linkage disequilibrium (LD) and contains *IL2*, *IL21* and two other genes, namely, *TENR* and *KIAA1109*, which are less plausible biological candidates to play a role in UC. Soon after, the associations were replicated in German and Spanish studies [29, 30] and later replicated in a large meta-analysis of GWAS [5]. Additionally, the locus is also associated in Chinese populations where LD is less extensive than in Caucasian populations, allowing for the identification of two apparently independent association signals for the *IL2* and *IL21* genes [31].

The GWAS Revolution

The limited success of targeted gene-candidate association studies made it clear that a new strategy would be needed to allow for the identification of genes underlying this complex disease. Success in hypothesis-driven association studies, such as gene-candidate studies, is restricted by the limited knowledge of disease molecular pathogenesis. Early association studies were also plagued by the low statistical

power of small cohorts to identify risk genes with modest effect size, which often resulted in false-positive findings. Knowledge of the complete sequence of the human genome, availability of markers covering most of the common variation in the human genome and microarray technology for mass parallel genotyping allowed the introduction of GWAS. Access to large cohorts and stringent statistical methods contributed to the success of GWAS in identifying complex disease genes by international teams of collaborators.

Since 2008, five groups published independent GWAS on UC [14, 15, 32–38]. The findings from these GWAS and from their first comprehensive meta-analysis (Table 6.1) recently led to identification of 47 loci associated with UC susceptibility [5]. This number has recently been brought to 133 (23 UC-specific loci and 110 loci associated with both CD and UC), thanks to a major collaborative effort from the International IBD Genetics Consortium (IIBDGC), who studied genetic information coming from a total of 38,565 IBD patients (17,865 UC and 20,700 CD) and 37,747 controls from 17 participating countries around the world [6].

Historically, the first GWAS in IBD was performed in 2006 for Crohn's disease, when CD-risk variants in the interleukin 23 receptor gene (*IL23R*) were identified [39]. Replication was performed in cohorts of both CD and UC patients, and *IL23R* was found to be a gene common to both IBD subphenotypes [39], with the coding variant Arg381Gln conferring protection against both CD and UC. *IL23R* represents one of the first genes, outside of the MHC, to be successfully associated to UC, and has opened up new avenues for potential therapeutic exploitation (see “Chapters 11 and 15” for more detail).

The first large-scale association study specifically targeting UC was published in 2008 by a UK group [14], who studied 10,886 non-synonymous SNPs and MHC tag SNPs. The screening cohort was composed of 905 UC patients recruited in the UK and 1,465 controls from the 1958 British Birth Cohort [40], and replication was performed in an independent cohort of similar size and origin. This scan resulted in the identification of three UC loci, namely, the 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 NF- κ B activation [41]; and the macrophage stimulating gene *MST1*, previously associated to CD and UC [42] and confirmed as a CD locus [43].

The first true GWAS targeting the entire genome in UC was published in 2008 by a German group [37]. The genotyping of 355,262 SNPs (post-quality control) was performed in a screening cohort composed of 1,167 UC cases from Germany and Norway and 777 controls, while three replication cohorts included patients from Germany, the UK and Belgium, for a total of 1,855 UC and 3,091 controls. Again, association signals in the MHC region were detected, together with additional independent loci containing the anti-inflammatory interleukin 10 (*IL10*) gene and the *ARPC2* gene, respectively. This group also reported an analysis of 50 Crohn's disease risk loci in UC [44], which revealed variants in *BSN*, *IL12B*, *NKX2-3*, *PTPN2*, *KIF21B*, *CCNY*, *HERC2*, *STAT3* and 10q21.2 (intergenic) loci to be relevant also to UC.

Table 6.1 Genome-wide association studies of ulcerative colitis

Study	Type	Cohort size (screen)	Ethnic origin	Main findings	New Loci ^a	Replicated loci
[14]	GWAS scan of nsSNPs	905 UC 1,465 HC	UK	ECM1 MST1	3	MHC
[15]	GWAS of early-onset	317 UC 4,250 HC	North America	20q13 ^b 21q22 ^b	2	IL23R MHC
[37]	GWAS	1,167 UC 777 HC	Germany Norway	IL10 ARP2C	2	MHC
[34]	GWAS	1,052 UC 2,571 HC	North America	1p36 12q15	2	IL23R MHC
[35]	GWAS	2,361 UC 5,417 HC	UK	HNF4A 16q22 LAMB1	3	23
[36]	GWAS	749 UC 2,031 HC	Japan	FCGR2A 13q12 SLC26A3	3	3
[38]	GWAS of early-onset IBD	724 UC 6,158 HC	North America Italy Scotland	Suggestive association to 2q37	1	8
[33]	GWAS	1,043 UC 1,703 HC	Germany	IL17REL ^a 7q22 ^a	2	MHC
[32]	Meta-analysis	2,693 UC 6,791 HC	North America Sweden	5p15 2p16 ORMDL3	3	14
[5]	Meta-analysis	6,687 UC 19,718 HC	All cohorts of Caucasian ancestry	16 % of heritability explained	29	18
[6]	Meta-analysis	10,920 UC 15,977 HC	All cohorts of Caucasian ancestry	IBD genetic architecture shaped by host-microbe co-evolution	55	44

UC ulcerative colitis, HC healthy controls, IBD inflammatory bowel diseases, GWAS genome-wide association study

^aDid not reached genome-wide significance in first screen and was genotyped in replication stage after inclusion criteria were lowered

^bSignificant results from IBD (CD+UC) cohorts

The second GWAS in UC was performed by a North American group [34]. Genotyping for the screening stage of this study was performed on 1,052 UC patients and 2,571 controls from Caucasians of European ancestry, while the replication phase was performed on cohorts from North America (768 UC patients and 721 controls) and Italy (619 UC and 394 controls). Confirming the association with MHC and IL23R, this GWAS also identified two new UC loci: 1p36 (RNF186, OTUD3, PLA2G2E) and 12q15 (*INFG*, *IL26*, *IL22*).

The third UC GWAS was performed by the UK IBD Genetics Consortium [35], who genotyped a total of 4,682 UC patients and 10,235 controls and identified three new loci on 20q13, 16q22 and 7q31. Each of these loci contains candidate genes that are biologically relevant in the context of inflammation. Locus 20q13.12 maps to the 3' untranslated region of gene hepatocyte nuclear factor 4 α (*HNF4A*), which is involved in cell adherence, and rare variants have been associated to maturity-onset diabetes of the young [45, 46]. Locus 16q22 encodes several genes including *CDH1*, the gene encoding glycoprotein E-cadherin that is involved in cell adhesion in the intestinal epithelium and whose expression is reduced in the inflamed colon of UC patients [47]. The strongest candidate gene for locus 7q31 is *LAMB1* encoding laminin β 1 subunit. Laminins are present in the lamina propria of the intestines where they help in anchoring the epithelial layer, and expression of laminins is decreased in the colonic mucosa of UC patients [48]. While this study also replicated previously associated UC loci (including 1p36, the MHC, *IL23R*, *MST1* and *NKX2-3*), further data from the UK IBDGC group implicated nine CD loci in UC including *KIF21B*, *IL18RAP*, *IL12B*, *JAK2* and *STAT3* [49].

The fourth UC GWAS (and the first executed in a non-Caucasian population) was performed by a Japanese group [36] on a total of 1,384 cases and 3,057 controls (index and replication cohorts), with three additional UC loci identified: *FCGR2A*, 13q12.13 (intergenic) and *SLC26A3* (7q31). The *FCGR2A* gene encodes a member of the family of immunoglobulin Fc receptor genes that is expressed on platelets, macrophages and neutrophils where it is involved in the phagocytosis of IgG-coated particles [50]. The *SLC26A3* gene is located in locus 7q31, which was previously identified by the UK group, but the association signal appears to somewhat differ in the two populations (stronger for *LAMB1* and *SLC26A3*, respectively, in the UK and Japanese cohorts).

In 2010, the German group carried out an extension of their first GWAS [33], with the goal to increase coverage of the genome-wide variability by using a combination of a new genome-wide SNP genotyping chip (the Affymetrix 6.0) that allowed more SNPs to be studied and of novel imputation approaches. In addition to the known MHC signal, association to new loci, *IL17REL* and 7q22 (near *SMURF1*, *KPNA7*), was identified. *IL17REL* encodes interleukin 17 receptor E-like, a poorly characterized member of the IL17 receptor gene family with no known ligand, which may bind specific IL17 cytokines to drive T_H17 inflammatory responses [51]. In an effort to further characterize the function of these loci in UC pathogenesis, the authors studied expression quantitative trait loci (eQTLs) by looking at expression signatures related to the associated SNPs at both loci. Differential immune gene expression was detected, with *IL17REL* variants modulating the expression of the *IL17RE* receptor, the cytokine *CSF3* and the modulator of T cell response *CD276*, while alleles at SNP rs7809799 in the 7q22 locus modulated the expression of the cytokine *IL1F10*, the transcription regulator of B cell development *FOXP1* and the MHC-associated gene *BTN3A*.

Early-Onset Ulcerative Colitis

In an effort to identify modest effect genes that had not been identified in previous adult-onset studies, a large collaborative GWAS of early-onset IBD was performed [15]. The rationale presented was that it should be easier to identify signals previously missed in (predominantly) adult-onset studies because of the clearer ascertainment of patients (early-onset IBD patients usually present extensive colitis) and stronger family history in early-onset IBD. All patients were diagnosed before turning 19 years old, were of European ancestry and were ascertained in various centres across the USA and one centre in Rome, Italy. The genotyping of the screening cohort (1,011 early-onset IBD patients including 317 UC patients) and replication of association signals led to the identification of two new IBD loci on chromosome 20q13 and 21q22. These loci were found as general IBD loci as they were found to be associated to both CD and UC. The 20q13 locus harbours several candidate genes, among which the authors highlighted *TNFRSF6B* because of the importance of the TNF pathway in IBD pathogenesis [52] and their detection of higher *TNFRSF6B* expression in colonic biopsies in relation to mucosal inflammation. The 21q22 locus is intergenic but the nearest gene codes for the proteasome assembly chaperone 1 (*PSMG1*), which showed a modestly increased expression in IBD patients compared to control.

A follow-up early-onset IBD GWAS was also published by the same group [38], including genotyping of adult-onset loci in their cohorts. While no new UC risk locus was identified in the study, this effort led to the association of two CD loci with UC: *ICOSLG* (21q22) and *ORMDL3* (17q12). *ICOSLG* is the ligand of ICOS expressed on activated T cells [53], while *ORMDL3* is involved in cellular Ca^{2+} homeostasis and appears to be a general inflammation and multi-diseases locus, as it was previously associated to asthma [54].

Meta-analyses: The Power of Many

The next phase in the evolution of association studies in UC was underlined by the activities of the IIBDGC. This consortium comprises several research groups from 16 different countries, who joined forces by sharing data and expertise in order to increase cohort size and improve statistical power to detect modest effect loci in large meta-analysis efforts [55].

The first meta-analysis in UC pooled GWAS data from the previously mentioned North American study, along with new North American and Swedish GWAS [32]. This first UC meta-analysis included a total of 2,693 UC patients and 6,791 controls for the discovery phase, as well as 2,009 UC patients and 1,580 controls for the replication phase. A total of 13 loci achieved genome-wide significance including four new loci: 1q21 (*FCGRA2*—previously identified in the Japanese GWAS), 2p16 (*REL-PUS10*), 17q12 (*ORMDL3*) and 5p15. Locus 2p16 contains the RNA chaperon gene *PUS10* and the gene encoding C-Rel (REL), one of the NF- κ B

proteins, while for the 5p15 locus, the nearest gene *CEP72* encodes a regulator of microtubule organization during mitosis [56]. Additionally, 14 loci that were previously either formally or suggestively identified were replicated in this study, including *IL23R*, *IL12B*, *TNFRSF15*, *IBD5* and *IL10*. Ten traditional CD loci were also identified in this UC meta-analysis including *IRGM*, *STAT3*, *CCL2-CCL7* and 21q21. Following the publication of this first UC meta-analysis, the authors estimated that less than 10 % of UC genetic variance could be explained by the associated loci known at the time [32].

The GWAS meta-analyses of CD and UC that followed had a dramatic impact on our knowledge of IBD genetic risk factors and again are the result of the efforts of the IIBDGC. Specifically, at the beginning of 2011, there were 18 loci significantly associated with UC identified by a few different independent studies. The same year, an IIBDGC meta-analysis of all these studies, including a total of 16,000 UC cases and 32,000 controls, discovered 29 additional loci, bringing the number of known UC loci to 47 and the estimated heritability they explain to 16 %. In order to identify relevant causal gene(s) at each locus, the authors used literature mining (GRAIL), eQTL databases, 1,000 genome sequencing data to search for correlated non-synonymous SNPs and physical proximity within the loci. Only a year later, the IIBDGC expanded the breadth of these analyses reporting one of the largest meta-analyses ever performed, based on 75,000 IBD cases and controls and including data from 15 different GWAS and additional typing on the immunochip, an array specifically designed to capture variation at 200 known risk loci for 12 common autoimmune diseases [6]. In this study, 71 new causative regions were identified, which brought the total tally to 163 independent IBD risk loci, far more than reported for any other complex disease. Of these, 110 appear to be relevant to both Crohn's disease and ulcerative colitis, while 23 show risk effects that are UC specific (the remaining 30 are CD-only loci). Fundamental contribution from this study was also the realization 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 defence against infection in IBD.

In addition to identifying hundreds of regions associated with increased IBD risk, most recent meta-analyses have also provided key information as to the potential biological pathways involved in UC pathogenesis. Among other methods, functional annotation of the identified UC genes allows to cluster them in pathways and clarify the molecular origin of the disease. In the latest meta-analysis from IIBDGC, where the highest number of UC loci was identified, the most significantly enriched Gene Ontology [57] term was regulation of cytokine production, specifically IFN- γ , IL-12, TNF- α and IL10 signalling. T, B and NK cell activation was the next most significant, and strong enrichment was also seen for response to molecules of bacterial origin and for the JAK-STAT signalling pathway. These pathways add to those already suspected to be relevant from individual gene functions such as, for instance, transcriptional regulation (*PRDMI*, *IRF5* and *NKX2-3*) and intestinal barrier functions (*GNAI2*, *HNF4A* and *LAMB1*). A list of prioritized candidate UC-risk genes and their functional properties is reported in Fig. 6.1.

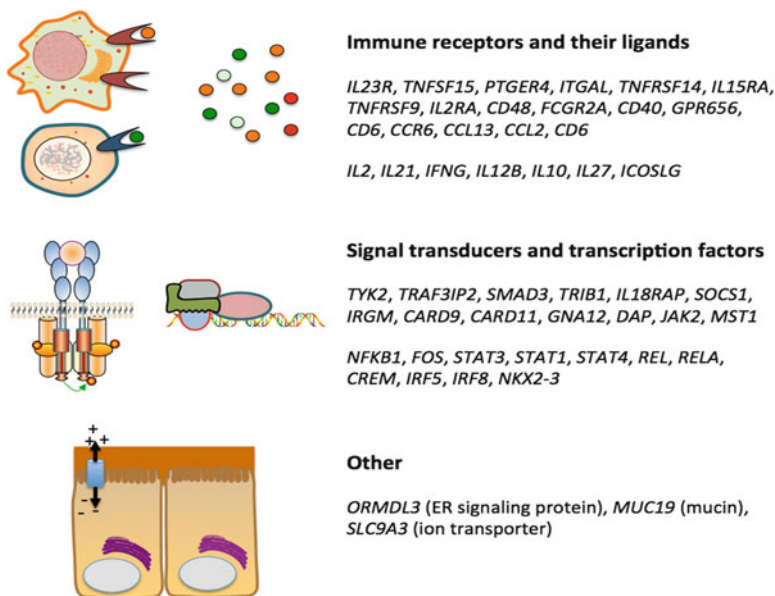


Fig. 6.1 UC risk genes grouped by function. The list of genes and corresponding biological functions is based on the prioritization at their respective risk loci based on the results included in the most recent UC meta-analysis [6]

Many UC Genes Are General IBD Risk Factors

Overlapping phenotype and aggregation of both CD and UC in some families have suggested that part of the genetic risk is common to both diseases. Many genetic studies have tested the association of Crohn's disease loci to UC and vice versa (as mentioned earlier in this chapter). The large meta-analyses of CD and UC, respectively, by Franke et al. and Anderson et al. enabled one of the first systematic analyses of loci shared by these two diseases. Specifically, Anderson et al. identified common loci by comparing the results of these large GWAS studies in CD and UC: common loci were defined when genome-wide significant loci for one disease ($P < 5 \times 10^{-8}$) reached $P < 1 \times 10^{-4}$ in the other disease, and at least 28 shared loci were identified by this means [5]. Among those, many are involved in T cell differentiation (e.g. 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*), which is involved in the maintenance of T_H17 cells and has been involved in several other diseases (see "Chapters 7 and 11"). T_H17 cells are thought to coordinate defence against specific pathogen and mediate inflammation [58], and the original identification of *IL23R* as an IBD risk factor shattered the paradigm of CD and UC being primarily T_H1 - and T_H2 -mediated diseases, respectively [39]. Genes involved in TNF signalling are also well represented among

common IBD genes (*TNFRSF9*, *TNFRSF14* and *TNFSF15*), and they encode proteins with various immune effects including systemic inflammation and activation of inflammatory transcription factor NF- κ B. As mentioned, the latest meta-analysis has brought the number of common CD-UC shared loci to 110 [6], while 23 appear to be the risk loci with UC-specific genetic effects. Interestingly, however, most of these UC-specific genes show the same direction of effect in Crohn's disease, suggesting that nearly all of the biological mechanisms involved in one disease have 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.

Associated Loci Differ Between Ethnic Groups

With the exception of a Japanese study [58], all UC GWAS carried out to date have been performed on Caucasians of European descent. Considering the interethnic differences in genomic architectures that reflect genetic drift, mutations and evolutionary factors, it is likely that some risk variants may be relevant in some populations but not others. One of the ultimate goals of current genetic studies is their exploitation for the design of novel therapeutic approaches, and it is therefore crucial to understand what portion of the genetic risk is detectable and hence also relevant in individual populations. The increasing number of association studies performed in cohorts of individuals of non-European descent, some of these GWAS, allows for some initial information to be evaluated in this context. The risk gene *IL23R*, for example, might not be associated in all ethnic groups, as it was not detected in the Japanese GWAS of UC [36] and in a targeted association study based on a Korean case-control cohort [59]. This may be at least in part due to the fact that some European *IL23R* variants are less common or absent in the Japanese population, like in the case of the protective *IL23R* 381Q allele (1 % compared to 7 % in Caucasians). Evidence for similar interethnic differences exists for other IBD loci, for instance, *NOD2* and *ATG16L1* [60, 61, 62], and large-scale analyses are currently at the core of IIBDGC interethnic studies that aim to shed further light on this important question.

Many Risk Factors Are Still to Be Discovered

Surprisingly, the total number of 133 UC risk loci (a number by far higher than in most other complex diseases) is still estimated to cumulatively account only for a minor portion of disease variance, namely, 7.5 % [6]. It should be noted that several additional loci narrowly miss the genome-wide-significance threshold of significance ($P < 5 \times 10^{-8}$) and thus are not considered to be validated loci; however, this observation suggests that other risk loci are still to be discovered. It has been proposed that additional loci or causative rare variants may not be captured by GWAS

platforms that primarily test common variants. Deep re-sequencing of risk loci is a way to identify these rare variants, and next-generation sequencing (NGS) technologies have greatly reduced per sample costs and process time, thus allowing such studies to be feasibly performed on adequate numbers of cases and controls. The first large-scale re-sequencing study performed in IBD was published in October 2011 [63]. Fifty six CD loci were re-sequenced in two DNA pools from 350 CD cases and 350 healthy controls for variants discovery, and follow-up genotyping of 115 low-frequency SNPs (non-synonymous, nonsense or splice-site variants) was performed in 16,054 CD and 12,153 UC patients and 17,575 healthy controls. The study identified a protective splice-site variant in the *CARD9* gene, and additional IBD susceptibility variants were also detected in *IL23R*, *CUL2*, *PTPN22* and *Clorf106*. In total, the authors estimated that 1–2 additional percent of CD heritability (and likely less for UC) is explained by these rare variants. These findings appear to be very modest but they give an important insight into complex disease genetics, where common variants of modest effect and rare variants of higher penetrance coexist in the same genes. Additional UC-focussed re-sequencing efforts will likely allow for the identification of other rare risk variants that may help to both pinpoint exact causative genes (in multiple-gene loci) and improve our understanding of the functional consequences of genetic variation at specific UC risk loci.

The Future of UC Genetics

The last few years have seen tremendous progress and important discoveries in UC genetics; however, most of the known disease heritability is still to be explained. As mentioned, one favoured hypothesis is that the remaining heritability is to be found in new risk loci and in additional rare variants located in known and yet to be identified UC loci. One of the obstacles to these discoveries is the statistical power that will be required, given that current loci have required tens of thousands of samples. Additional independent GWAS and the pooling of these new GWAS into increasingly bigger meta-analyses will certainly lead to the identification of new loci. Re-sequencing studies that aim to identify rare causal variants will be essential companions to these GWAS, and the analysis of news markers such as copy number variations (CNVs) will hopefully provide additional insight into disease heritability. In addition, long-distance regulatory variants need to be more carefully evaluated given that some associated loci are gene deserts, as well as individual SNPs for their potential role in altering epigenetic signatures in DNA methylation, histone acetylation, microRNA binding, etc. Genotyping cohorts of different ethnic background will be helpful in fine-mapping-associated loci. Genetic variation patterns are often different in different populations, and, for example, African populations have more diversity and the linkage disequilibrium blocks are generally shorter than in population of European ancestry, thus allowing for the identification of new or refined association signals.

Finally, there is great need for integrative genomic and system biology approaches where multilayer data are collectively analyzed and exploited to decipher the mysteries of the genome. Together with the avalanche of data coming from several GWAS and their meta-analysis, the considerable challenge of having to understand the role of each and individual genetic risk factor from hundreds of loci has emerged. In this endeavour, large-scale functional studies into the transcriptome and the proteome are now indispensable. The amount of work still ahead in UC genetics seems to be no less than what it was some years ago, but we certainly now have a very solid foundation of knowledge upon which we can build.

References

1. De Dombal FT (1968) Ulcerative colitis: definition, historical background, aetiology, diagnosis, natural history and local complications. *Postgrad Med J* 44(515):684–692, Epub 1968/09/01
2. Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ (1999) HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. *Gut* 45(3):395–401, Epub 1999/08/14
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921, Epub 2001/03/10
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al (2001) The sequence of the human genome. *Science* 291(5507):1304–1351, Epub 2001/02/22
5. 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
6. 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
7. 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
8. Amiel JL (1967) Study of the leukocyte phenotypes in Hodgkin’s disease. In: Curtoni ES, Mattiuz PL, Tosi RM (eds) *Histocompatibility testing*. Munksgaard, Copenhagen, pp 79–81
9. The_MHC_Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 401(6756):921–923, Epub 1999/11/30
10. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ et al (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 31(1):311–314, Epub 2003/01/10
11. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM et al (2008) Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 4(4):e1000024, Epub 2008/04/26
12. Orchard TR, Dhar A, Simmons JD, Vaughan R, Welsh KI, Jewell DP (2001) MHC class I chain-like gene A (MICA) and its associations with inflammatory bowel disease and peripheral arthropathy. *Clin Exp Immunol* 126(3):437–440, Epub 2001/12/12
13. Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MM, Green T et al (2009) Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci U S A* 106(44):18680–18685, Epub 2009/10/23
14. 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

15. 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
16. Ho GT, Moodie FM, Satsangi J (2003) Multidrug resistance 1 gene (P-glycoprotein 170): an important determinant in gastrointestinal disease? *Gut* 52(5):759–766, Epub 2003/04/15
17. Panwala CM, Jones JC, Viney JL (1998) A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, *mdr1a*, spontaneously develop colitis. *J Immunol* 161(10):5733–5744, Epub 1998/11/20
18. Annese V, Valvano MR, Palmieri O, Latiano A, Bossa F, Andriulli A (2006) Multidrug resistance 1 gene in inflammatory bowel disease: a meta-analysis. *World J Gastroenterol* 12(23):3636–3644, Epub 2006/06/15
19. Zintzaras E (2012) Is there evidence to claim or deny association between variants of the multidrug resistance gene (MDR1 or ABCB1) and inflammatory bowel disease? *Inflamm Bowel Dis* 18(3):562–572, Epub 2011/09/03
20. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M et al (2007) A genome-wide association study for celiac disease identifies risk variants in the region harboring *IL2* and *IL21*. *Nat Genet* 39(7):827–829, Epub 2007/06/15
21. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V et al (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39(7):857–864, Epub 2007/06/08
22. Zhernakova A, Alizadeh BZ, Bevova M, van Leeuwen MA, Coenen MJ, Franke B et al (2007) Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *Am J Hum Genet* 81(6):1284–1288, Epub 2007/11/14
23. Sawalha AH, Kaufman KM, Kelly JA, Adler AJ, Aberle T, Kilpatrick J et al (2008) Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus. *Ann Rheum Dis* 67(4):458–461, Epub 2007/08/28
24. Liu Y, Helms C, Liao W, Zaba LC, Duan S, Gardner J et al (2008) A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet* 4(3):e1000041, Epub 2008/03/29
25. Yang L, Anderson DE, Baecher-Allan C, Hastings WD, Bettelli E, Oukka M et al (2008) *IL-21* and *TGF-beta* are required for differentiation of human *T(H)17* cells. *Nature* 454(7202):350–352, Epub 2008/05/13
26. Monteleone G, Monteleone I, Fina D, Vavassori P, Del Vecchio Blanco G, Caruso R et al (2005) Interleukin-21 enhances T-helper cell type I signaling and interferon-gamma production in Crohn's disease. *Gastroenterology* 128(3):687–694
27. Sadlack B, Merz H, Schorle H, Schimpl A, Feller AC, Horak I (1993) Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. *Cell* 75(2):253–261, Epub 1993/10/22
28. Festen EA, Goyette P, Scott R, Annese V, Zhernakova A, Lian J et al (2009) Genetic variants in the region harbouring *IL2/IL21* associated with ulcerative colitis. *Gut* 58(6):799–804, Epub 2009/02/10
29. Glas J, Stallhofer J, Ripke S, Wetzke M, Pfenning S, Klein W et al (2009) Novel genetic risk markers for ulcerative colitis in the *IL2/IL21* region are in epistasis with *IL23R* and suggest a common genetic background for ulcerative colitis and celiac disease. *Am J Gastroenterol* 104(7):1737–1744, Epub 2009/05/21
30. Marquez A, Orozco G, Martinez A, Palomino-Morales R, Fernandez-Arquero M, Mendoza JL et al (2009) Novel association of the interleukin 2-interleukin 21 region with inflammatory bowel disease. *Am J Gastroenterol* 104(8):1968–1975, Epub 2009/05/28
31. Shi J, Zhou L, Zhernakova A, Qian J, Zhu F, Sun G et al (2011) Haplotype-based analysis of ulcerative colitis risk loci identifies both *IL2* and *IL21* as susceptibility genes in Han Chinese. *Inflamm Bowel Dis* 17(12):2472–2479, Epub 2011/06/08
32. McGovern DP, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD et al (2010) Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet* 42(4):332–337, Epub 2010/03/17

33. Franke A, Balschun T, Sina C, Ellinghaus D, Hasler R, Mayr G et al (2010) Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat Genet* 42(4):292–294, Epub 2010/03/17
34. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V et al (2009) Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet* 41(2):216–220, Epub 2009/01/06
35. Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A et al (2009) Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* 41(12):1330–1334, Epub 2009/11/17
36. Asano K, Matsushita T, Umeno J, Hosono N, Takahashi A, Kawaguchi T et al (2009) A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat Genet* 41(12):1325–1329, Epub 2009/11/17
37. 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
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. 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
40. Power C, Elliott J (2006) Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* 35(1):34–41, Epub 2005/09/13
41. Matsuda A, Suzuki Y, Honda G, Muramatsu S, Matsuzaki O, Nagano Y et al (2003) Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene* 22(21):3307–3318, Epub 2003/05/23
42. 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
43. 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
44. Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T et al (2008) Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet* 40(6):713–715, Epub 2008/04/29
45. Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ et al (1996) Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* 384(6608):458–460, Epub 1996/12/05
46. Battle MA, Konopka G, Parviz F, Gaggl AL, Yang C, Sladek FM et al (2006) Hepatocyte nuclear factor 4alpha orchestrates expression of cell adhesion proteins during the epithelial transformation of the developing liver. *Proc Natl Acad Sci U S A* 103(22):8419–8424, Epub 2006/05/23
47. Karayiannakis AJ, Syrigos KN, Efstathiou J, Valizadeh A, Noda M, Playford RJ et al (1998) Expression of catenins and E-cadherin during epithelial restitution in inflammatory bowel disease. *J Pathol* 185(4):413–418, Epub 1998/11/26
48. Schmehl K, Florian S, Jacobasch G, Salomon A, Korber J (2000) Deficiency of epithelial basement membrane laminin in ulcerative colitis affected human colonic mucosa. *Int J Colorectal Dis* 15(1):39–48, Epub 2000/04/15
49. Anderson CA, Massey DC, Barrett JC, Prescott NJ, Tremelling M, Fisher SA, et al (2009) Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship. *Gastroenterology* 136(2):523–529. e3. Epub 2008/12/11
50. Huang ZY, Chien P, Indik ZK, Schreiber AD (2011) Human platelet FcgammaRIIA and phagocytes in immune-complex clearance. *Mol Immunol* 48(4):691–696, Epub 2010/12/21
51. Wu B, Jin M, Zhang Y, Wei T, Bai Z (2011) Evolution of the IL17 receptor family in chordates: a new subfamily IL17REL. *Immunogenetics* 63(12):835–845, Epub 2011/07/07

52. van Heel DA, Udalova IA, De Silva AP, McGovern DP, Kinouchi Y, Hull J et al (2002) Inflammatory bowel disease is associated with a TNF polymorphism that affects an interaction between the OCT1 and NF(κ)B transcription factors. *Hum Mol Genet* 11(11): 1281–1289, Epub 2002/05/23
53. Hutloff A, Dittrich AM, Beier KC, Eljaschewitsch B, Kraft R, Anagnostopoulos I et al (1999) ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* 397(6716):263–266, Epub 1999/02/04
54. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S et al (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448(7152):470–473, Epub 2007/07/06
55. <http://www.ibdgenetics.org/>
56. Oshimori N, Li X, Ohsugi M, Yamamoto T (2009) Cep72 regulates the localization of key centrosomal proteins and proper bipolar spindle formation. *EMBO J* 28(14):2066–2076, Epub 2009/06/19
57. <http://www.ncbi.nlm.nih.gov/>
58. Bettelli E, Korn T, Oukka M, Kuchroo VK (2008) Induction and effector functions of T(H)17 cells. *Nature* 453(7198):1051–1057
59. Yang SK, Jung Y, Hong M, Kim H, Ye BD, Lee I et al (2011) No association between TNFSF15 and IL23R with ulcerative colitis in Koreans. *J Hum Genet* 56(3):200–204, Epub 2011/01/14
60. Mahurkar S, Banerjee R, Rani VS, Thakur N, Rao GV, Reddy DN et al (2011) Common variants in NOD2 and IL23R are not associated with inflammatory bowel disease in Indians. *J Gastroenterol Hepatol* 26(4):694–699, Epub 2010/12/16
61. Yamazaki K et al (2002) Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease. *J Hum Genet* 47:469–472
62. Yamazaki K et al (2007) Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients. *J Hum Genet* 52:575–583
63. 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