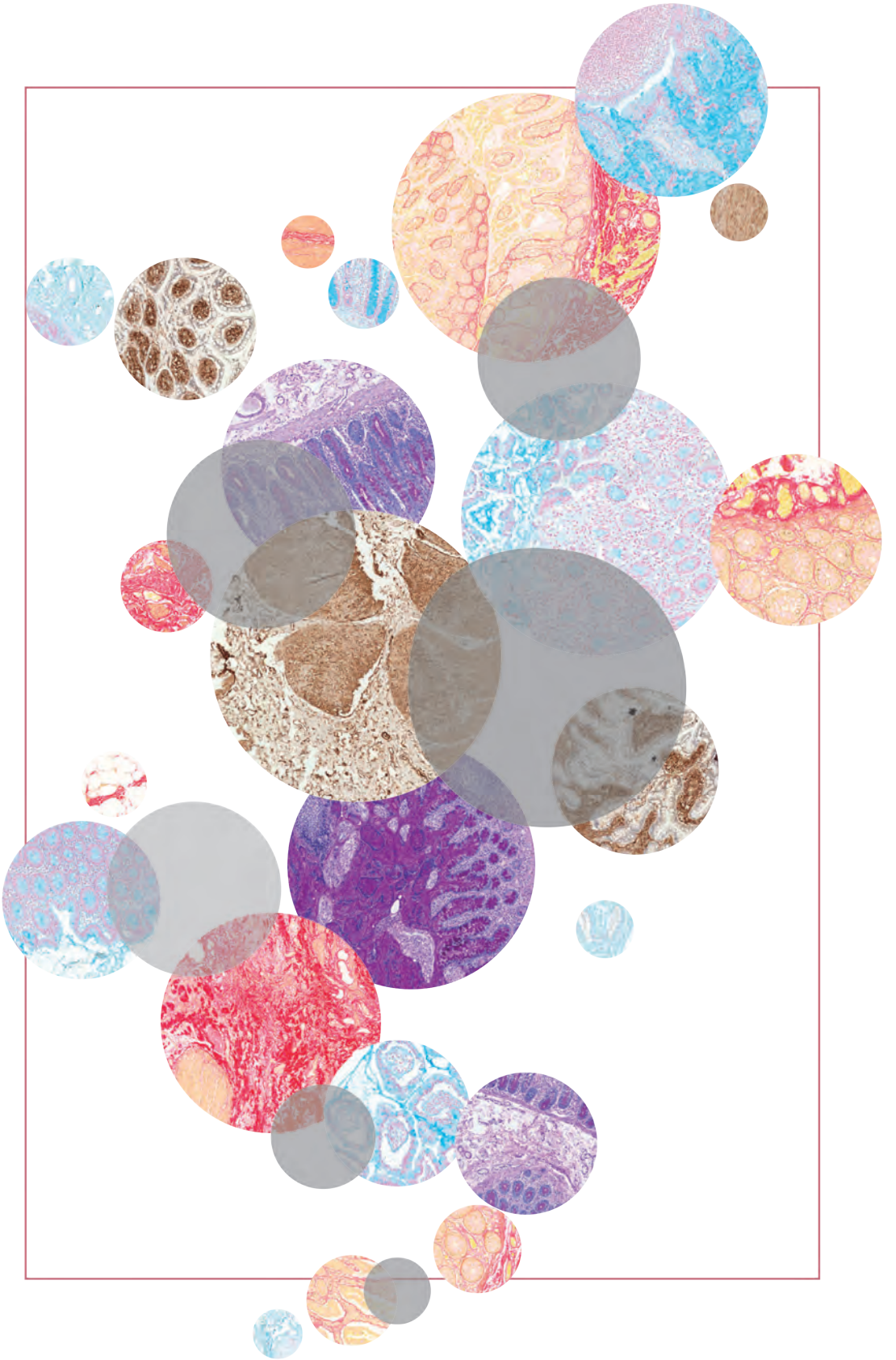
Gut mucosal gene expression in inflammatory bowel disease

Uniken Venema, Werna

DOI:
10.33612/diss.178571637



chapter TWO

The genetic background of Inflammatory Bowel Disease: From correlation to causality

Werna T.C. Uniken Venema, Michiel D. Voskuil, Gerard Dijkstra, Rinse K. Weersma, Eleonora A.M. Festen

* authors share first authorship position

Journal of Pathology, 2017
DOI: 10.1002/path.4817

ABSTRACT

Recent studies have greatly improved our insight into the genetic background of Inflammatory Bowel Disease (IBD). New high throughput technologies and large-scale international collaborations have contributed to the identification of 200 independent genetic risk loci for IBD. However, in most of these loci it is unclear which gene conveys the risk for IBD. More importantly, it is unclear which variant within or near the gene is causal to the disease. Using targeted GWAS, imputation, re-sequencing of risk loci and in silico fine-mapping of densely typed loci, several causal variants have been identified in IBD risk genes, and various pathological pathways have been uncovered. Current research in the field of IBD focuses on the effect of these causal variants on gene expression and protein function. However, more elements than only the genome must be considered to disentangle the multifactorial pathology of IBD. The genetic risk loci identified to date only explain a small part of genetic variance in disease risk. Currently, large multi-omics studies are incorporating such factors, ranging from the gut microbiome to the environment. In this review we present the progress that has been made in IBD genetic research and stress the importance of studying causality to increase our understanding of the pathogenesis of IBD. We highlight important causal genetic variants in the candidate genes *NOD2*, *ATG16L1*, *IRGM*, *IL23R*, *CARD9*, *RNF186* and *PRDM1*. We describe their downstream effects on protein function and their direct effects on the gut immune system. Furthermore, we discuss the future role of genetics in unravelling disease mechanisms in IBD.

INTRODUCTION

Inflammatory Bowel Disease (IBD) is a chronic relapsing immune-mediated disease, characterized by inflammation and ulceration of the gut mucosa. IBD roughly consists of two subtypes: ulcerative colitis (UC) and Crohn's disease (CD). In UC, the disease is generally limited to the colon, whereas in CD, inflammation can occur in the entire gastrointestinal tract. UC patients have continuous inflammation limited to the mucosal layer. In CD, the inflammation is discontinuous and involves all layers of the gut.

The incidence of IBD in the western world is around 1 per 1000 individuals and is increasing[1]. In the clinic, IBD patients present with recurrent episodes of abdominal pain, diarrhea, bloody stools and weight loss. Furthermore, patients can suffer from extra-intestinal manifestations of the skin, joints, eyes, and less frequently in abdominal organs such as the biliary tract[2]. Medical treatments are expensive, and often surgical resection of parts of the intestine is necessary. Although the efficacy of medical therapies is improving, up to 20% of UC patients and 50% of CD patients require surgery within 10 years of diagnosis[3].

The aetiology of IBD is complex, and is most likely based on a combination of microbial, environmental and genetic factors. Smoking and appendectomy, for instance, are known to be risk factors for CD, although both seem to have a protective effect in UC[4-7]. Diet is a shared risk factor for both UC and CD: frequent intake of fast-food predisposes to both diseases[8]. Breastfeeding is associated with lower risk for IBD[9]. Twin studies show that there is an hereditary component in IBD risk, which is stronger in CD than in UC[10].

Hypothesis-driven genetic studies in CD revealed the first susceptibility genes[11,12]. Subsequently, more in-depth and large-scale research on the genetic background of IBD has been carried out through Genome-Wide Association Studies (GWAS). To date, these studies have identified 200 genetic risk loci, or genetic regions, associated with IBD[13]. For the majority of these genetic regions, the disease associated genes or the causal genomic variants or remain unknown. In order to search for disease mechanisms and pathways, current research focuses on identifying these causal genes and causal variants **(Figure 1)**. In the search for these, various strategies can be applied: first, sometime causal variants are identified in the GWAS, as happened in the first GWAS for CD (e.g. R381Q in *IL23R* and T300A in *ATG16L1*)[14,15]; second, when not identified by the original GWAS data analysis, causal variants can be found through re-sequencing of candidate genomic regions surrounding an associated common variant (e.g. IVS11+1C>G in *CARD9*, R179X in *RNF186*)[16-18]; third, in case of the availability of dense genotyping of genomic

regions, causal variants can be identified by in silico fine-mapping analyses (e.g. W620R in *PTPN22* and I170V in *SMAD3*)[19]; and fourth, whole-exome sequencing (WES) and whole-genome sequencing (WGS) of large cohorts of IBD cases and population matched healthy controls are currently being performed by US and UK groups. These studies are expected to yield more in-depth knowledge on causal variants in genomic IBD risk loci[20,21].

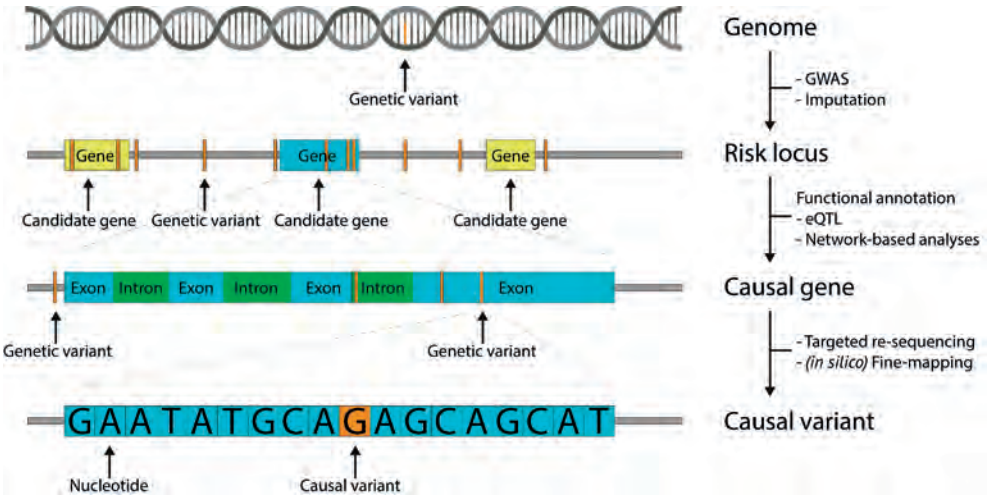
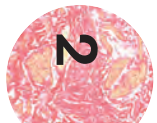


Figure 1. The identification of a causal genetic variant.
From top to bottom. **Genome**: Genome-wide association studies (GWAS) with or without imputation methods identify signals that are highly significantly associated to a certain trait. The identified region is called a risk locus. **Risk locus**: a risk locus on the genome may harbor one or more candidate gene(s) and/or genetic variant(s). Functional annotation may identify the causal gene within the risk locus. **Causal gene**: the gene displayed here contains several introns and exons. Genetic variants can be located in both introns and exons but also up- or downstream of the causal gene. The actual genetic causal variant may still be unknown, and techniques such as re-sequencing or fine-mapping can be used to identify the causal variant. **Causal variant**: an example of nucleotides located in an exon of the candidate gene. The actual causal variant (G: guanine) in orange. This variant can be subjected to functional studies.

This review highlights the recently identified causal genetic variants for IBD, as discovered through the above-mentioned methods (**Table 1**). Furthermore, we describe the disease pathways that have been revealed through these findings. The focus of this review is on adult-onset IBD, as (very) early-onset IBD generally has an oligogenic nature and is reviewed elsewhere[22,23].

Table 1. Overview of causal genetic variants discussed in this review article.
Method refers to the method by which the causal genetic variant was identified. References refer to the first reported discovery. Odds ratio (OR) and minor allele frequencies (MAF) derive from these references.

Method	Genetic variant	Type	Chromosome	Gene	Effect	OR	MAF	Reference
Hypothesis-driven	fs1007insC	Nonsense	16	*NOD2*	Risk for CD	2.1	0.04	Hugot et al., 2001 Ogura et al., 2001
	R702W	Missense	16	*NOD2*	Risk for CD	2.13	0.04	Hugot et al., 2001
	G908R	Missense	16	*NOD2*	Risk for CD	2.32	0.01	Hugot et al., 2001
	rs10870077	Intronic	9	*CARD9*	Risk for IBD	1.21	0.436	Zhernakova et al., 2008
Genome-wide association	R381Q	Missense	1	*IL23R*	Protective for IBD	0.26	0.070	Duerr et al., 2006
	T300A	Missense	2	*ATG16L1*	Risk for CD	1.45	0.533	Hampe et al., 2007
Targeted re-sequencing	G149R	Missense	1	*IL23R*	Protective for IBD	0.60	0.0043	Rivas et al., 2011 Momozawa et al., 2011
	V362I	Missense	1	*IL23R*	Protective for IBD	0.72	0.0152	Rivas et al., 2011 Momozawa et al., 2011
	c.313C>T	Synonymous	5	*IRGM*	Risk for CD	-	0.3037	Brest et al., 2011
	CNVdel	Deletion	5	*IRGM*	Risk for IBD	1.5	0.10	McCarrol et al., 2008
	L105	Synonymous	5	*IRGM*	Risk for CD	-	0.3037	Parkes et al., 2007
	A64T	Missense	1	*RNF186*	Risk for UC	1.49	0.0069	Beaudoin et al., 2013
	R179X	Nonsense	1	*RNF186*	Protective for UC	0.30	0.00-0.0078	Rivas et al., 2015
	c.IVS11+1G>C	Splice-site	9	*CARD9*	Protective for IBD	0.29	0.0071	Rivas et al., 2011
Whole-exome sequencing	S354N	Missense	6	*PRDM1*	Risk for CD	1.23	0.01835	Ellinghaus et al., 2013
	L450F	Missense	6	*PRDM1*	Protective for UC	0.37	0.01135	Ellilnghaus et al., 2013



GENETICS: GWAS STATE OF THE ART

Since the first GWAS was published in 2002[24], many genetic variants, or single nucleotide polymorphisms (SNPs), have been shown to be associated with a specific trait or disease in individuals. Over the entire length of the human genome, SNPs occur with a frequency of approximately one per 300 nucleotides. In GWAS, differences in frequencies of certain SNPs are compared between cases, usually patients with a specific disease, and healthy controls. When the frequency of a SNP is statistically significantly different (usually a $p < 5x10^{-8}$ is considered to be genome-wide significant) between patients and healthy controls, then the region on the genome surrounding this SNP, hereafter called a "risk locus", is associated with the disease. GWAS have become an important tool to study the aetiology of complex inflammatory diseases, including IBD.

Early GWAS and subsequent meta-analyses in CD revealed CD risk loci with candidate genes, encoding proteins involved in immunological pathways such as autophagy (*ATG16L1* and *IRGM*), bacterial handling (*NOD2*) and T cell signaling (*IL23R*)[14,15,25,26]. GWAS and GWAS meta-analyses of UC and CD showed substantial overlap between UC and CD risk loci (*IL23R*, *IL12B*, *NKX2-3* and *MST1*), but also identified some UC specific (*IL10*, *HLA*) and CD specific (*NOD2*, *ATG16L1* and *IRGM*) risk loci[27-31]. Increasing the sample size by combining both genetic and clinical data from databases of the members of the International IBD Genetics Consortium (IIBDGC) substantially increased the number of identified IBD risk loci[25,26,31,32].

Of the above-mentioned IBD risk loci, the majority of risk loci overlap with those of other, mainly immune-mediated diseases[33,34]. This knowledge was applied in the design of the ImmunoChip: a chip composed of all genetic variants correlated to immune-mediated diseases[35,36]. The ImmunoChip project has several limitations, such as lack of dense coverage of all loci, lower sensitivity in non-European ethnic groups, and limited coverage of the genome. Nonetheless it has contributed greatly to our knowledge of IBD risk loci: a meta-analysis of all GWAS and ImmunoChip data from a total of 75,000 individuals increased the number of risk loci to 163[37]. Most of these genetic risk loci contribute to both UC and CD, and the vast majority of loci associated with only one disease showed the same direction of effect in the non-associated disease. This suggests that nearly all biological mechanisms involved in one disease have a similar role in the other. However, some shared loci show a risk effect in opposite directions for each disease. For example, the presence of the risk variant R620W, located in the *PTPN22* gene, increases risk for UC, but is protective for CD[37,38]. The mechanism underlying this inverse risk effect remains yet elusive.

The 163 Immunochip-identified genetic IBD risk loci have all been discovered in cohorts of European descent, presumably introducing a major ethnic bias[37]. Recently, GWAS have been performed in Asian populations in which completely new risk loci for IBD were discovered[39-42]. Subsequent meta-analyses were performed and several risk loci associated with IBD in European populations were found to be associated with IBD also in these non-European cohorts. The largest GWAS and ImmunoChip data meta-analysis in the field of IBD to date combined genotype data from cohorts of different ancestry[13]. This study comprised genetic data from 96,486 individuals including 9,846 individuals of non-European ancestry and allowed for the identification of 38 additional risk loci for IBD, increasing the total number of IBD risk loci to 200. Of these 200 IBD risk loci in cohorts from different ancestry, 138 show overlap with other complex diseases, including other immune-mediated diseases[43]. One of the key findings of this study is that differences in effect size of variants between ethnicities are minimal. It shows that most of the apparent differences in genetic risk between populations can be explained by differences in allele frequencies between populations, or in sample size differences between the cohorts. These findings suggest that trans-ethnicity studies can be used to identify new risk loci in complex diseases.

FROM GWAS "TAG SNPS" TO FUNCTIONAL "CAUSAL" VARIANTS

GWAS has made genetic research much more affordable, as it only covers tag SNPs as markers for genetic loci, representing a proportion of the common genetic variation. However, a large part of the genetic variation is left untagged.

Imputation methods deal with the problem of untagged missing genetic variation, by using the principle of linkage disequilibrium (LD). LD is the linked transmission of multiple SNPs within a genomic region, which can be used to infer the alleles of SNPs that are not directly genotyped. Imputation processes extrapolate SNPs in LD from densely characterized reference panels (such as reference panels from the 1000Genomes project or the Genome of the Netherlands (GoNL) project[44,45]) to a sparsely genotyped study sample. By this means, it is possible to estimate unobserved genotypes with a high accuracy, thus increasing chances of finding true associations[46].

Many risk loci contain more than one gene. To identify the candidate gene in such genomic regions, a number of analyses can be applied **(Figure 1)**. First, network-based analysis tools can calculate the enrichment of connectivity between genes in associated loci and can thereby prioritize specific genes within a risk locus[37,47,48]. Second, expression quantitative trait loci (eQTL) can be used to identify candidate genes[49]. eQTL are genetic

variants whose variation is associated to gene expression levels. If an IBD associated variant or a variant in LD with an IBD associated SNP is a known eQTL, the gene whose expression is regulated by the eQTL is the most likely candidate gene within a genomic region.

When candidate genes are identified using significantly associated SNPs, the actual causal genetic variant for the disease often remains unknown. In order to find this causal variant, genomic regions harbouring these candidate genes can be re-sequenced, revealing the complete genomic sequence of this region. The first causal variants identified using this method were located in genetic regions of *NOD2*, *IL23R* and *CARD9*[16,50].

If genetic associations are identified in densely genotyped genomic regions, high-resolution in silico fine-mapping can be used to identify causal variants. In silico fine-mapping can identify which associated risk variants are statistically convincing causal variants. To date, 18 variants with a >95% probability of causality have been identified using this method[19]. In the last few years, causal genetic variants have been identified through a combination of targeted GWAS, imputation, re-sequencing and fine-mapping. After causal variants have been identified, functional studies are necessary to show how these causal genetic variants lead to disease[17,51].

IBD candidate genes that encode proteins can roughly be categorized in three major disease-associated pathways: regulation of inflammatory responses (e.g. *IL23R*), regulation of autophagy (e.g. *ATG16L1*), and microbial sensing to activate autophagy (e.g. *NOD2*). Genetic variants in genes encoding proteins involved in the above-described pathways have been identified to be associated to IBD, either conferring risk for or protection from the development of IBD[52,53]. It is known that host genetics can alter immunological responses to commensal microbes. Moreover, genetics can influence the gut microbiome[54,55]. Compared to healthy controls, IBD patients show a dysregulated microbiome. This dysregulation may contribute to the disease, as microbes play an important role in regulation of the immune system[56,57].

In the sections below, we will describe current knowledge on important causal genetic variants and illustrate how these variants contribute to the pathogenesis of IBD.

NOD2

From all genetic variants that are assumed to play a role in the pathogenesis in IBD, the most striking are those in the region of the *Nucleotide-Binding Oligomerization Domain Containing 2 (NOD2)* gene, increasing the risk for CD. Three of the most common variants located in *NOD2* (fs1007insC, R702W and G908R) had been identified prior to the GWAS era, by hypothesis-driven methods[11,12]. Re-sequencing and fine-mapping studies confirmed these variants to be causal and identified novel causal variants located in *NOD2*[16,19].

NOD2 plays an important role in the innate immune system by sensing muramyl dipeptide (MDP), a peptidoglycan derived from bacterial lipopoylysaccharides (LPS)[58]. *NOD2* variants conferring the largest genetic risk are located in the leucine-rich repeat (LRR), causing either a frameshift mutation (fs1007insC) leading to a truncated LRR, or amino acid changes (R702W and G908R)[11,12,59]. Cells that express these variants fail to activate NFκB upon stimulation with the NOD2 ligand MDP[60,61]. Although the function of NOD2 and the influence of the genetic variants on NOD2 function have been studied intensively, there is yet no consensus on how this altered NOD2 function eventually leads to disease[62].

Currently, there are two major hypotheses of NOD2-related pathogenesis, based on different aspects of NOD2 function and mechanisms of disease in CD[62]. First, NOD2 has an important role in host defence in epithelial cells (*i.e.* Paneth cells), most likely through its role in the production of α-defensin[63-65]. *NOD2* deficiency is associated with an altered microbiome composition, causing UC-resembling colitis in Dextran Sulfate Sodium (DSS)-treated mice, in which massive entry of luminal bacteria takes place[66-68]. Contrarily, *NOD2* deficiency prevents CD-resembling colitis by supressing inflammation in 2,4,6-Trinitrobenzenesulforic acid (TNBS)- colitis models, having milder entry of luminal bacteria[69]. A second hypothesis is that *NOD2* deficiency leads to colitis because of an inappropriate Toll-like receptor (TLR) stimulation[62,70,71]. In both mouse and human dendritic cells (DC), pre-stimulation with MDP leads to downregulation of TLR ligands and stimulants, which does not happen in *NOD2*-transgenic mice[72,73]. *NOD2*-expressing mice are protected from colitis upon MDP stimulation, whereas *NOD2*-deficient mice are not[73].

It is likely that both alteration of α-defensin expression and TLR downregulation play a role in the pathogenic effects of the *NOD2* risk variants, and that *NOD2* risk variants increase the inflammatory potential of the microbiome through mucosal immunodeficiency, leading to increased inflammation[62].

ATG16L1

Many studies have linked genetic variants in *Autophagy Related 16-Like 1 (ATG16L1)* to CD[14,74,75]. The missense variant T300A is strongly associated with CD in European populations and has remained one of the most significant variants in CD.

ATG16L1 encodes a protein essential for autophagy, a cellular lysosomal degradative process for recycling of proteins in the maintenance of homeostasis[76-78]. Both extra- and intracellular stress, signals such as starvation, growth factor deprivation, endoplasmatic reticulum (ER) stress, and pathogen infection can upregulate autophagy[77].

Numerous studies have investigated the role of ATG16L1 to autophagy and inflammatory signalling[79-86]. ATG16L1 can limit cytokine production, and both mice and patients carrying the *ATG16L1* risk variant have defective microbial clearance and altered cytokine response toward both viruses and bacteria, suggesting the involvement of host-microbe interactions[55,79,81,82,88,97]. In fact, biopsies from inflamed parts of the terminal ileum from CD patients carrying the T300A risk variant contain more *Bacteroides fragilis*, *Fusobacteria* and *Escherichia coli* in comparison to those from patients carrying the protective T300 variant[88]. *B. fragilis* has beneficial immunomodulatory properties and protects against experimental colitis[55,89-94]. The bacterium uses an ATG16L1-dependent pathway which induces regulatory T cells to generate this mucosal tolerance. Human immune cells from CD patients carrying the T300A variant are unable to induce regulatory T cells in response to *B. fragilis*, and show impaired killing of pathogens[55,88].

Only recently the mechanisms underlying the *ATG16L1* T300A variant have been elucidated. Turnover of ATG16L1 is dependent of caspase-3 and -7[95,96]. T300A is located in the cleavage site for these caspases. Cellular stress enhances cleavage, and the presence of T300A leads to increased cleavage resulting in protein instability. This consequently leads to diminished autophagy and increased inflammatory cytokine responses. Smoking is a risk factor for developing CD. Interestingly, it has been shown that cigarette smoke has a bacterium- and *ATG16L1* allele-specific effect on the xenophagic activity of human monocytes. This promotes survival of pathogenic adherent invasive *E. coli* (AIEC), thereby linking a common occurring risk variant of a gene to an environmental risk factor[97].

In summary, *ATG16L1* T300A may contribute to CD by both impaired killing of pathogens and an inability to respond to beneficial microbes.

IRGM

Genetic risk variants in the region harbouring the *Immunity-Related GTPase family M (IRGM)* gene have been strongly associated with CD and, to a lesser extent, to UC[31,37,98-101]. The IRGM gene encodes an autophagy protein that plays an important role in innate immunity against intracellular pathogens like *Mycobacterium tuberculosis*, *Salmonella typhimurium* and AIEC bacteria (associated with CD)[102-104]. The fact that *IRGM* is a distinctly human gene makes it more difficult to study its function[105]. IRGM indirectly activates AMP-activated protein kinase (AMPK), a key regulator of cellular energy homeostasis, stabilizing autophagy regulators (such as ULK1 and ATG14)[106,107]. Furthermore, IRGM directs co-assembly of autophagy machinery, and connects this machinery to bacterial pattern recognition receptors such as NOD2, leading to enhancement of autophagy. This IRGM-NOD2 interaction is thought to prevent excessive inflammation.

The genetic variant L105 that lies immediately upstream of *IRGM*, is strongly associated with CD[98,108]. After re-sequencing exons of *IRGM*, coding sequence-variation within *IRGM* was excluded as a source of this association[98]. A deletion allele of a copy number variant (CNVdel) upstream of *IRGM*, in perfect LD with L105, is associated with altered *IRGM* expression and CD[101,102]. In turn, reduced *IRGM* expression reduces the efficacy of anti-bacterial autophagy.

However, a different study showed that the exonic risk variant *IRGM* rs10065172 (c.313C>T), also in perfect LD with L105 and the CNVdel, could be a causal variant in *IRGM*[109]. c.313C>T alters the microRNA (miRNA) binding site of miRNA-196, a miRNA overexpressed in the inflamed intestinal epithelium in CD patients. Normally miRNA-196 downregulates *IRGM*, but this downregulation is lost in the presence of c.313C>T. Loss of this downregulation results in compromised autophagy of AIEC bacteria in patients carrying the variant. IRGM induces autophagy to eliminate intracellular mycobacteria and this genetic variant is of particular interest since it has been associated with susceptibility to tuberculosis[110-112]. This fact supports the theory that mycobacteria may be important in the pathogenesis of CD[103]. Taken together, these results suggest that the association of IRGM with CD arises from an alteration in *IRGM* regulation affecting the efficacy of autophagy.

INTERPLAY BETWEEN NOD2, ATG16L1 AND IRGM

NOD-receptors detect bacteria and induce an autophagy-mediated immune response. MDP stimulation in primary human DCs activates autophagy, resulting in increased bacterial killing and the generation of $CD4^+$ antigen-specific T cell responses[82]. NOD2 (and NOD1) can recruit ATG16L1 to the plasma membrane at the bacterial entry site[113]. Cells with *NOD2* genetic variants (R702W, G908R and fs1007insC) fail to recruit ATG16L1 to the plasma membrane and show impaired pathogen-induced autophagy and impaired $CD4^+$ antigen-specific responses[82,113,114]. In contrast, there is no consensus regarding NOD2 signalling in the presence of *ATG16L1* T300A. Two independent studies show that MDP-stimulated autophagy is impaired in DC and lymphoblastoid cells from CD patients carrying *ATG16L1* T300A[82,113]. However, in another study using cells from healthy donors, the *ATG16L1* T300A variant impaired NOD2-dependent anti-bacterial function in colonic epithelial cells, but not in monocytic cells[114]. These seemingly conflicting results from functional studies indicate that the manner in which *ATG16L1* and *NOD2* contribute to disease may be tissue- and situation specific. Thus, the interaction between *ATG16L1* and *NOD2* and its contribution to the pathogenesis of IBD may very well be related to the process of autophagy[115], but could also act in an autophagy-independent manner[85].

A third well known IBD risk gene, *IRGM*, may play an additional important role in the interaction between NOD2 and ATG16L1: in human cells IRGM is in complex with NOD2 and ATG16L1, enhancing the assembly of this complex[106]. IRGM stimulates phosphorylation and activation of important autophagy regulators (such as ULK1 and Beclin 1) and connects them to ATG16L1[106,116-118]. Polyubiquitination of IRGM is under control of NOD2 and stabilizes autophagy initiation complexes. NOD2 enhances association of ubiquitination-competent IRGM with ULK1 and Beclin 1, but not in mutant IRGM[106]. In contrast, mutant IRGM remains its ability to bind ATG16L1 equally well as the ubiquitination-competent IRGM. Our increased knowledge on NOD2, ATG16L and IRGM dysfunction in IBD this has up to now not resulted in therapeutic options. Manipulation of the microbiome or small molecules can directly or indirectly modulate autophagy and innate immunity, an effect that could be explored in patients carrying these risk alleles[119].

IL23R

Multiple variants within the locus harbouring the *Interleukin 23 Receptor (IL23R)* gene are strongly associated to IBD[15,25,26]. The *IL23R* variant R381Q, conferring protection against IBD, has been identified through GWAS[15], while the protective G149R and V362I *IL23R* variants have been identified by targeted re-sequencing[16,18]. The *IL23R* risk locus plays a role in an important pathway for IBD: T-helper 17 (Th17) signalling[120]. The *IL23R* gene encodes the IL23 receptor complex (IL23R), which is a receptor for the heterodimeric pro-inflammatory cytokine IL23, composed of p19 and p40 subunits. p40 is also part of the IL12 cytokine[121]. Binding of IL23 by IL23R results in activation of Janus Kinase 2 (JAK2), subsequent phosphorylation of signal transducer and activator of transcription 3 (STAT3) and STAT4, leading to the transcription of other pro-inflammatory cytokines[122]. This increased transcription of pro-inflammatory cytokines causes differentiation of $CD4^+$ T cells into pro-inflammatory Th17 cells, that are critical in antimicrobial defences. Genetic variants in or near the *IL23R* gene are thought to generate an inappropriate immune response to commensal bacteria in the gut, and dysregulation of the IL23-Th17 pathway[120,123]. Moreover, loci within genes encoding proteins downstream in the IL23-Th17 pathway have also been associated with IBD (*JAK2* and *STAT3*)[25].

Several studies have investigated functional consequences of genetic variants in *IL23R*. The R381Q variant, conferring protection against IBD, results in loss of function: cells expressing this variant have decreased levels of phosphorylated STAT3 and STAT4. This consequently leads to the reduced production of pro-inflammatory cytokines with reduced Th17 effector functions[124-127].

As outlined above, the IL23-Th17 pathway plays an important role in mediating intestinal inflammation and is therefore a promising target for IBD medication. Ustekinumab, an anti-IL12/23p40 monoclonal antibody, prevents human IL12 and IL23 from binding to the IL12Rβ1 receptor chain of IL12 (IL12Rβ1/β2) and IL23 (IL12Rβ1/23Rα) receptor complexes on the surface of natural killer (NK) and T cells. A recent phase 3 study in patients with moderate-to-severe CD refractory to anti-Tumour Necrosis Factor alpha (TNFα) showed favourable response and remission to Ustekinumab[128]. Specific blockade of the IL23 pathway through use of a monoclonal antibody against p19 was effective in both the prevention and treatment of a murine model of $CD4^+$ T cell–mediated colitis[129]. Similarly, in a phase 2a trial, CD patients refractory to anti-TNF showed favourable response to anti-IL23p19[130].

CARD9

In 2008, the first *caspase recruitment family member 9 (CARD9)* disease association has been discovered through a custom SNP array for innate immunity. The identified variant rs10870077 in the *CARD9* genetic locus on chromosome 9, confers risk for IBD. Another variant, S12N, in perfect LD with rs10870077, increases expression of *CARD9*[26,131,132]. Moreover, this variant is associated with other immune-mediated diseases such as ankylosing spondylitis, immunoglobulin A nephropathy and primary sclerosing cholangitis[145-147].

The *CARD9* locus encodes an innate immune system pattern recognition receptor, which is necessary for the immune response against bacterial and fungal infections. Currently, all clinical reports concerning humans that have loss of function of CARD9, or deficient *CARD9* expression, show a higher susceptibility to fungal infections than subjects with normal CARD9 function and expression. Also, reduced numbers of inflammatory cells are seen[133-135]. After its genetic association with innate immunity, the mechanisms by which *CARD9* is involved in protection from infections have also been studied intensively. *CARD9*-deficient mice show lower pro-inflammatory cytokine production than *CARD9*-competent mice and are more susceptible to DSS-induced colitis[136-138]. The functional pathway involved is the recognition of microbes through a surface receptor, with subsequent binding of CARD9 and activation of NFκB in macrophages and neutrophils[139]. In addition, CARD9 interacts with NOD2 in response to bacteria[138-140].

Recently, CARD9 was found to be a mediator in the IL22 pathway activation, which initiates immune responses to both bacteria and fungi through antimicrobial peptides[141,142]. CARD9 augments IL22 activity, through microbial stimulation of the aryl hydrocarbon receptor (AHR)[143]. In IBD patients, reduced AHR-ligand production is seen. An AHR-agonist, such as the strain lactobacillus, could counter this reduced ligand production. This creates therapeutic possibilities for personalized medicine[144].

In contrast, more recently a rare splice-site variant, c.IVS11+1G>C, conferring protection against IBD has been identified[16,18]. This variant leads to a truncated protein which causes an ineffective pro-inflammatory response with diminished cytokine production through defective microbial sensing[134].

Altogether, the effects of *CARD9* variants may be bidirectional: c.IVS11+1G>C can cause loss of function of CARD9 resulting in diminished immune responses, whereas the risk variant S12N may lead to overexpression of *CARD9*. This overexpression can result in a hyper reactive immune state and increasing susceptibility to IBD[148].

RNF186

Several GWAS have identified risk loci for UC in the region of the *Ring Finger Protein 186 (RNF186)* gene[29,42,131]. However, the function of this gene, the causal variant(s) within these loci and eventually its consequences for protein function have long remained elusive. The *RNF186* gene encodes the ring finger E3 ligase which localizes at the endoplasmatic reticulum (ER) and regulates ER stress-mediated apoptosis[149]. Deep re-sequencing of GWAS loci recently allowed for the identification of a likely causal variant in *RNF186*: A64T[18]. This variant confers risk for the development of UC. The A64T variant is located in the E3 ubiquitin-protein ligase domain of RNF186[18]. The ligases encoded regulate key adaptors of pro-inflammatory pathways[150-152]. *RNF186* expression is higher in human intestinal tissue as compared to other human tissue, with the highest levels of expression located in the transverse colon[131,153]. *Shigella* infection in mice induces upregulation of RNF186 in the intestinal epithelium[154,155]. A recent study conducting a search for protein truncating variants (*i.e.* loss of function variants), identified a novel genetic variant in *RNF186*, R179X, which confers protection against UC[17]. Interestingly, R179X truncation does not inhibit protein expression, and the mechanism of action of this variant is as yet unclear.

PRDM1

The *Positive Regulatory Domain 1 (PRDM1)* gene on chromosome 6 encodes a zinc finger transcriptional repressor, B Lymphocyte-Induced Maturation Protein 1 (BLIMP1), which is involved in plasma cell differentiation, immunoglobulin secretion, and is strongly upregulated in apoptosis[156-161]. *PRDM1* is mainly expressed in mucosal T cells and plasma cells[51]. However, mutations in *PRDM1* also affect various other immunological cell types such as NK-cells, DC and macrophages. Through WES studies, two missense mutations in *PRDM1* (S354N and L450F) associated with IBD have been identified[51]. The variant S354N confers risk for CD, and L450F is protective for UC. In both healthy controls and CD patients carrying S354N but not L450F, T cells exhibit increased proliferation, cytokine secretion and activation markers upon stimulation. T cells from CD patients also exhibit increased expression of L selectin. L selectin is necessary for peripheral blood lymphocyte migration into the intestinal tissue, which is thought to contribute to the pathogenesis of IBD[162-164]. In addition, the common risk allele of rs7746082, representing the strongest association to CD in this locus, lowers expression of *PRDM1* in ileal biopsies of CD patients[25-51].

Currently, mice with a T cell specific BLIMP1 deficiency are used as a model for CD. These mice develop severe colitis and have more peripheral effector T cells upon colitis induction, as compared to WT mice[165-166].

Moreover, in DC from subjects carrying the common risk allele of rs6911490, representing the strongest association to UC in this locus, *PRDM1* expression is significantly lower as compared with DC from subjects not carrying the risk allele[31,167]. Moreover, these cells show exaggerated immune responses upon bacterial stimuli[167].

Various immune cell subtypes are influenced by *PRDM1*. BLIMP1 has been shown to be an essential transcription factor to activate Th17 cells, the dysregulation of which contributes to the pathogenesis of IBD[168]. Moreover, mouse *PRDM1*-knockout colonic DC show higher pro-inflammatory cytokine production and lower levels of the anti-inflammatory cytokines. Blocking of the pro-inflammatory cytokines with anti-IL1β and anti-IL6 diminishes the exaggerated immune response in the *PRDM1*-knockout cells. This provides future possibilities for targeted treatments[167]. In addition to this, the BLIMP1-dependent production of IL10 protects against inflammation and may be a target for future therapies as well[169].

CONCLUSION AND FUTURE PERSPECTIVES

In this review we have outlined the progress that has been made in the understanding of IBD through the use of genetic studies. We have discussed the role of GWAS in studying the genetics of IBD and also the importance of moving from genetic correlation to causality and from genetic correlation to functional consequences. Further, we have highlighted recently identified causal genetic variants in candidate genes and their downstream effects on protein function. These recent advances greatly contribute to our understanding of how genetic variants can lead to disease. In addition, these recent advances show potential opportunities for future therapies.

The genetic risk loci for IBD identified thus far only explain a small part of the genetic variance in disease risk: 13-13.5% in CD and 7.5-9% in UC[37,170]. Most likely, the number of identified genetic risk loci will grow further over the coming years and within the identified loci a growing number of causal genetic variants will be discovered. The identification of causal variants is extremely important for making a more adequate estimation of the genetic variance explained. Identifying causal variants is also instrumental for drug development and repositioning[171,172]. In order to better understand disease mechanisms and to develop more specific therapies to target inflammatory pathways, the effects of causal variants on protein function will be a principal subject of research in the field of IBD. Identifying causal variants through techniques such as re-sequencing and fine-mapping alone will be insufficient to completely elucidate mechanisms leading to disease.

A promising approach to study the effect of specific candidate genes and causal variants are disease models, such as the TNBS/DSS-treated mouse models and human-derived cells. Studies on the function of genetic variants were initially performed in complete knockout models of the candidate gene, but have more recently been studied in specific causal variant models, which provide a more realistic representation of the clinical effect of these causal variants.

In order to increase the chances of identifying causal variants in IBD risk loci, there are efforts to enlarge numbers of cases in studies. Currently, large numbers of samples of IBD patients in both the US and the UK are sequenced using whole-exome and whole-genome technologies, respectively[20,21]. These efforts will likely result in the identification of additional causal variants.

In some IBD risk loci the causal variant will be extremely difficult to detect. This can be due to variable effect size or effect direction of a variant amongst patients; variants in currently known risk loci can be both protective and risk-promoting (e.g. *CARD9*). Within other risk loci the causal variants are located in non-coding areas of the genome, which makes them difficult to detect. Functional studies on IBD risk loci in which no causal variants have been identified, can be performed by obtaining an approximation of the functional effects within this risk locus. This can be done with eQTL studies, which show the effect of the associated risk variant on gene expression.

As the possibilities to analyse and integrate large data sets are evolving, future research opportunities emerge. Considering the multifactorial origin of IBD, consisting of, amongst others, the microbiome, the environment and the proteome, the study of IBD should be approached more widely. Although on the one hand these factors have a clear effect on disease, they can also be affected by and have an interaction with genetic variants. Using a multi-omics approach, these factors must be incorporated in the study of IBD. Multi-omics studies integrate different levels of data that contribute to disease. Multi-omics studies have so far revealed important interplays between the host's genetics and other factors, such as genetic variation and clinical disease behaviour and the composition of the microbiome[173] (Imhann *et al*, *unpublished data*). GWAS studies have shown genetic susceptibility to adverse drug responses to IBD medication[174,175]. From oncologic immunotherapy studies we have learned that the gut microbiome can affect efficacy of these immunotherapies[176,177]. These findings stress the importance of integrating multi-omics data from large cohorts in order to find therapies targeting patients with specific risk profiles. Multi-omics studies provide a basis for the next step in IBD research: translating findings into clinical practice and more specifically into personalized treatment strategies

The current progress that has been made in IBD research has paved the way for translation of basic scientific findings to clinical practice. With the ongoing identification of causal genetic variants, both functional and multi-omics studies will greatly improve our understanding of IBD and subsequently the treatment of IBD patients over the coming years.

Acknowledgements

R.K. Weersma is supported by a VIDI grant (016.136.308) from the Netherlands Organisation for Scientific research (NWO). E.A.M. Festen is supported by a MLDS Career Development grant (CDG 14-04).

REFERENCES

1. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology. 2012 Jan;142(1):46–54.
2. Vavricka SR, Schoepfer A, Scharl M, et al. Extraintestinal Manifestations of Inflammatory Bowel Disease. Inflamm Bowel Dis. 2015 Aug;21(8):1982–92.
3. Frolkis AD, Dykeman J, Negrón ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. Gastroenterology. 2013 Nov;145(5):996–1006.
4. Lakatos PL, Szamosi T, Lakatos L. Smoking in inflammatory bowel diseases: good, bad or ugly? World J Gastroenterol. 2007 Dec;13(46):6134–9.
5. Mahid SS, Minor KS, Soto RE, et al. Smoking and inflammatory bowel disease: a meta-analysis. Mayo Clin Proc. 2006 Nov;81(11):1462–71.
6. Kaplan GG, Jackson T, Sands BE, et al. The risk of developing Crohn's disease after an appendectomy: a meta-analysis. Am J Gastroenterol. 2008 Nov;103(11):2925–31.
7. Andersson RE, Olaison G, Tysk C, et al. Appendectomy and protection against ulcerative colitis. N Engl J Med. 2001 Mar;344(11):808–14.
8. Niewiadomski O, Studd C, Wilson J, et al. Influence of food and lifestyle on the risk of developing inflammatory bowel disease. Intern Med J. 2016 Jun;46(6):669–76.
9. Klement E, Cohen R V, Boxman J, et al. Breastfeeding and risk of inflammatory bowel disease: a systematic review with meta-analysis. Am J Clin Nutr. 2004 Nov;80(5):1342–52.
10. Gordon H, Trier Moller F, et al. Heritability in inflammatory bowel disease: from the first twin study to genome-wide association studies. Inflamm Bowel Dis. 2015 Jun;21(6):1428–34.
11. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001 May;411(6837):599–603.
12. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001 May;411(6837):603–6.
13. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015 Sep;47(9):979–86.
14. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet. 2007 Feb;39(2):207–11.
15. Duerr RH, Taylor KD, Brant SR, et al. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. Science. 2006 Dec;314(5804):1461–3.
16. Rivas MA, Beaudoin M, Gardet A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet. 2011 Nov;43(11):1066–73.
17. Rivas MA, Graham D, Sulem P, et al. A protein truncating R179X variant in RNF186 confers protection against ulcerative colitis. bioRxiv 2015 Dec; doi: http://dx.doi.org/10.1101/035105.
18. Beaudoin M, Goyette P, Boucher G, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. PLoS Genet. 2013;9(9):e1003723.
19. Huang H, Fang M, Jostins L, et al. Association mapping of inflammatory bowel disease loci to single variant resolution. bioRxiv 2015 Oct. doi: http://dx.doi.org/10.1101/028688
20. Broad Institute | IBD whole-exome sequencing initiative. Available from: https://www.broadinstitute.org/news/5188
21. UK10K Project. Available from: http://www.uk10k.org/
22. Uhlig HH, Schwerd T, Koletzko S, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology. 2014 Nov;147(5):990–1007.

23. Moran CJ, Klein C, Muise AM, et al. Very early-onset inflammatory bowel disease: gaining insight through focused discovery. Inflamm Bowel Dis. 2015 May;21(5):1166–75.

24. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet. 2002 Dec;32(4):650–4.

25. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008 Aug;40(8):955–62.

26. Franke A, B McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010 Dec;42(12):1118-25.

27. Franke A, Balschun T, Karlsen TH, et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet. 2008 Nov;40(11):1319–23.

28. Fisher SA, Tremelling M, Anderson CA, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet. 2008 Jun;40(6):710–2.

29. Silverberg MS, Cho JH, Rioux JD, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet. 2009 Feb;41(2):216–20.

30. Festen EAM, Stokkers PCF, van Diemen CC, et al. Genetic analysis in a Dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk. Am J Gastroenterol. 2010 Feb;105(2):395–402.

31. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet. 2011 Mar;43(3):246–52.

32. IIBDGC. Available from: www.ibdgenetics.org

33. Lees CW, Barrett JC, Parkes M, et al. New IBD genetics: common pathways with other diseases. Gut. 2011 Dec;60(12):1739–53.

34. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet. 2009 Jan;10(1):43–55.

35. Trynka G, Hunt KA, Bockett NA, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011 Dec;43(12):1193–201.

36. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13(1):101.

37. Jostins L, Ripke S, Weersma RK, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012 Oct 31;491(7422):119–24.

38. Parkes M, Cortes A, van Heel DA, et al. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013 Sep;14(9):661–73.

39. Asano K, Matsushita T, Umeno J, et al. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. Nat Genet. 2009 Dec;41(12):1325–9.

40. Juyal G, Negi S, Sood A, et al. Genome-wide association scan in north Indians reveals three novel HLA-independent risk loci for ulcerative colitis. Gut. 2015 Apr;64(4):571–9.

41. Yang S-K, Hong M, Zhao W, et al. Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. Gut. 2014 Jan;63(1):80–7.

42. Yang S-K, Hong M, Zhao W, et al. Genome-Wide Association Study of Ulcerative Colitis in Koreans Suggests Extensive Overlapping of Genetic Susceptibility With Caucasians. Inflamm Bowel Dis. 2013 Apr;19(5):954–66.

43. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014 Jan;42(Database issue):D1001–6.

44. 1000 Genomes Project. Available from: http://www.1000genomes.org/

45. Genome of the Netherlands | GoNL. Available from: http://www.nlgenome.nl/

46. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009 Jun;5(6):e1000529.

47. Raychaudhuri S, Plenge RM, Rossin EJ, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009 Jun;5(6):e1000534.

48. Rossin EJ, Lage K, Raychaudhuri S, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011;7(1):e1001273.

49. Abiola O, Angel JM, Avner P, et al. The nature and identification of quantitative trait loci: a community's view. Nat Rev Genet. 2003 Nov;4(11):911–6.

50. Momozawa Y, Mni M, Nakamura K, et al. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. Nat Genet. 2011 Jan;43(1):43–7.

51. Ellinghaus D, Zhang H, Zeissig S, et al. Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. Gastroenterology. 2013 Aug;145(2):339–47.

52. Liu T-C, Stappenbeck TS. Genetics and Pathogenesis of Inflammatory Bowel Disease. Annu Rev Pathol. 2016 Feb;11:127–48.

53. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011 Jun;474(7351):307–17.

54. Knights D, Silverberg MS, Weersma RK, et al. Complex host genetics influence the microbiome in inflammatory bowel disease. Genome Med. 2014 Jan;6(12):107.

55. Chu H, Khosravi A, Kusumawardhani IP, et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. Science. 2016 May;352(6289):1116-20.

56. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014 Mar;15(3):382–92.

57. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology. 2014 May;146(6):1489–99.

58. Strober W, Murray PJ, Kitani A, et al. Signalling pathways and molecular interactions of NOD1 and NOD2. Nat Rev Immunol. 2006 Jan;6(1):9–20.

59. Hampe J, Cuthbert A, Croucher PJ, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet. 2001 Jun;357(9272):1925–8.

60. Bonen DK, Ogura Y, Nicolae DL, et al. Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. Gastroenterology. 2003 Jan;124(1):140–6.

61. Chamaillard M, Philpott D, Girardin SE, et al. Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases. Proc Natl Acad Sci U S A. 2003 Mar;100(6):3455–60.

62. Strober W, Asano N, Fuss I, et al. Cellular and molecular mechanisms underlying NOD2 risk-associated polymorphisms in Crohn's disease. Immunol Rev. 2014 Jul;260(1):249–60.

63. Wehkamp J, Harder J, Weichenthal M, et al. NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. Gut. 2004 Nov;53(11):1658–64.

64. Wehkamp J, Salzman NH, Porter E, et al. Reduced Paneth cell alpha-defensins in ileal Crohn's disease. Proc Natl Acad Sci U S A. 2005 Dec;102(50):18129–34.

65. Simms LA, Doecke JD, Walsh MD, et al. Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease. Gut. 2008 Jul;57(7):903–10.

66. Petnicki-Ocwieja T, Hrncir T, Lui YJ, et al. Nod2 is required for the regulation of commensal microbiota in the intestine. Proc Natl Acad Sci U S A. 2009;106(37):15813.

67. Rehman A, Sina C, Gavrilova O, et al. Nod2 is essential for temporal development of intestinal microbial communities. Gut. 2011 Oct;60(10):1354–62.

68. Mondot S, Barreau F, Al Nabhani Z, et al. Altered gut microbiota composition in immune-impaired Nod2(-/-) mice. Gut. 2012 Apr;61(4):634–5.

69. Amendola A, Butera A, Sanchez M, et al. Nod2 deficiency is associated with an increased mucosal immunoregulatory response to commensal microorganisms. Mucosal Immunol. 2014 Mar;7(2):391–404.

70. Strober W, Fuss I, Mannon P. The fundamental basis of inflammatory bowel disease. J Clin Invest. 2007 Mar;117(3):514–21.

71. Watanabe T, Kitani A, Murray PJ, et al. Nucleotide Binding Oligomerization Domain 2 Deficiency Leads to Dysregulated TLR2 Signaling and Induction of Antigen-Specific Colitis. Immunity. 2006 Sep;25(3):473–85.

72. Watanabe T, Kitani A, Murray PJ, et al. NOD2 is a negative regulator of Toll-like receptor 2–mediated T helper type 1 responses. Nat Immunol. 2004 Aug;5(8):800–8.

73. Watanabe T, Asano N, Murray PJ, et al. Muramyl dipeptide activation of nucleotide-binding oligomerization domain 2 protects mice from experimental colitis. J Clin Invest. 2008 Feb;118(2):545–59.

74. Prescott NJ, Fisher SA, Franke A, et al. A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn's disease and is independent of CARD15 and IBD5. Gastroenterology. 2007 May;132(5):1665–71.

75. Weersma RK, Zhernakova A, Nolte IM, et al. ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. Am J Gastroenterol. 2008 Mar;103(3):621–7.

76. Mizushima N, Kuma A, Kobayashi Y, et al. Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. J Cell Sci. 2003 May;116(Pt 9):1679–88.

77. He C, Klionsky DJ. Regulation mechanisms and signaling pathways of autophagy. Annu Rev Genet. 2009;43:67–93.

78. Salem M, Ammitzboell M, Nys K, et al. ATG16L1: A multifunctional susceptibility factor in Crohn disease. Autophagy. 2015;11(4):585–94.

79. Saitoh T, Fujita N, Jang MH, et al. Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1beta production. Nature. 2008 Nov;456(7219):264–8.

80. Cadwell K, Liu JY, Brown SL, et al. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature. 2008 Nov13;456(7219):259–63.

81. Cadwell K, Patel KK, Maloney NS, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. Cell. 2010 Jun;141(7):1135–45.

82. Cooney R, Baker J, Brain O, et al. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. Nat Med. 2010 Jan;16(1):90–7.

83. Fujita N, Saitoh T, Kageyama S, et al. Differential involvement of Atg16L1 in Crohn disease and canonical autophagy: analysis of the organization of the Atg16L1 complex in fibroblasts. J Biol Chem. 2009 Nov;284(47):32602–9.

84. Messer JS, Murphy SF, Logsdon MF, et al. The Crohn's disease: associated ATG16L1 variant and Salmonella invasion. BMJ Open. 2013;3(6).

85. Sorbara MT, Ellison LK, Ramjeet M, et al. The protein ATG16L1 suppresses inflammatory cytokines induced by the intracellular sensors Nod1 and Nod2 in an autophagy-independent manner. Immunity. 2013 Nov;39(5):858–73.

86. Conway KL, Kuballa P, Song J-H, et al. Atg16l1 is required for autophagy in intestinal epithelial cells and protection of mice from Salmonella infection. Gastroenterology. 2013 Dec;145(6):1347–57.

87. Kuballa P, Huett A, Rioux JD, et al. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. PLoS One. 2008;3(10):e3391.

88. Sadaghian Sadabad M, Regeling A, de Goffau MC, et al. The ATG16L1-T300A allele impairs clearance of pathosymbionts in the inflamed ileal mucosa of Crohn's disease patients. Gut. 2015 Oct;64(10):1546-52.

89. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014 Mar;157(1):121–41.

90. Honda K, Littman DR. The microbiome in infectious disease and inflammation. Annu Rev Immunol. 2012;30:759–95.

91. Hooper L V, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science. 2012 Jun;336(6086):1268–73.

92. Allan SE, Broady R, Gregori S, et al. CD4+ T-regulatory cells: toward therapy for human diseases. Immunol Rev. 2008 Jun;223:391–421.

93. Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents intestinal inflammatory disease. Nature. 2008 May;453(7195):620–5.

94. Round JL, Mazmanian SK. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proc Natl Acad Sci U S A. 2010 Jul;107(27):12204–9.

95. Murthy A, Li Y, Peng I, Reichelt M, et al. A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. Nature. 2014 Feb;506(7489):456–62.

96. Lassen KG, Kuballa P, Conway KL, et al. Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. Proc Natl Acad Sci U S A. 2014 May;111(21):7741–6.

97. Regeling A, Volders H, Blokzijl T, et al. Cigarette Smoke is an Environmental Factor That Reduces Innate Immune Functions Especially in Individuals With the CD-Associated Risk Variant ATG16L1-T300A. Gastroenterology. 2012;142(5):S687.

98. Parkes M, Barrett JC, Prescott NJ , et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet. 2007 Jul;39(7):830–2.

99. Glas J, Seiderer J, Bues S, et al. IRGM variants and susceptibility to inflammatory bowel disease in the German population. PLoS One. 2013;8(1):e54338.

100. Weersma RK, Stokkers PCF, Cleynen I, et al. Confirmation of multiple Crohn's disease susceptibility loci in a large Dutch-Belgian cohort. Am J Gastroenterol. 2009 Mar;104(3):630–8.

101. Prescott NJ, Dominy KM, Kubo M, et al. Independent and population specific association of risk variants at the IRGM locus with Crohn's disease. Hum Mol Genet. 2010 May;19(9):1828–39.

102. McCarroll SA, Huett A, Kuballa P, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet. 2008 Sep;40(9):1107–12.

103. Singh SB, Davis AS, Taylor GA, et al. Human IRGM induces autophagy to eliminate intracellular mycobacteria. Science. 2006 Sep;313(5792):1438–41.

104. Lapaquette P, Glasser A-L, Huett A, et al. Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly. Cell Microbiol. 2010 Jan;12(1):99–113.

105. Bekpen C, Marques-Bonet T, Alkan C, et al. Death and resurrection of the human IRGM gene. PLoS Genet. 2009 Mar;5(3):e1000403.

106. Chauhan S, Mandell MA, Deretic V. IRGM Governs the Core Autophagy Machinery to Conduct Antimicrobial Defense. Mol Cell. 2015;58(3):507–21.

107. Chauhan S, Mandell MA, Deretic V. Mechanism of action of the tuberculosis and Crohn disease risk factor IRGM in autophagy. Autophagy. 2016 Feb 27;12(2):429–31.

108. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun;447(7145):661–78.

109. Brest P, Lapaquette P, Souidi M, et al. 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. 2011 Mar;43(3):242–5.

110. King KY, Lew JD, Ha NP, et al. Polymorphic allele of human IRGM1 is associated with susceptibility to tuberculosis in African Americans. PLoS One. 2011;6(1):e16317.

111. Song JH, Kim SY, Chung KS, et al. Association between genetic variants in the IRGM gene and tuberculosis in a Korean population. Infection. 2014 Aug;42(4):655–60.

112. Lu Y, Li Q, Peng J, Zhu Y, et al. Association of autophagy-related IRGM polymorphisms with latent versus active tuberculosis infection in a Chinese population. Tuberculosis. 2016 Mar;97:47–51.

113. Travassos LH, Carneiro LAM, Ramjeet M, et al. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. Nat Immunol. 2010 Jan;11(1):55–62.

114. Homer CR, Richmond AL, Rebert NA, et al. ATG16L1 and NOD2 interact in an autophagy-dependent antibacterial pathway implicated in Crohn's disease pathogenesis. Gastroenterology. 2010 Nov;139(5):1630–41.

115. Ramjeet M, Hussey S, Philpott DJ, et al. "Nodophagy": New crossroads in Crohn disease pathogenesis. Gut Microbes. 2010;1(5):307–15.

116. Kim J, Kundu M, Viollet B, et al. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. Nat Cell Biol. 2011 Feb;13(2):132–41.

117. Kim J, Kim YC, Fang C, et al. Differential regulation of distinct Vps34 complexes by AMPK in nutrient stress and autophagy. Cell. 2013 Jan;152(1-2):290–303.

118. Egan DF, Shackelford DB, Mihaylova MM, et al. Phosphorylation of ULK1 (hATG1) by AMP-activated protein kinase connects energy sensing to mitophagy. Science. 2011 Jan 28;331(6016):456–61.

119. Shaw SY, Tran K, Castoreno AB, et al. Selective modulation of autophagy, innate immunity, and adaptive immunity by small molecules. ACS Chem Biol. 2013 Dec 2;8(12):2724–33.

120. Abraham C, Cho J. Interleukin-23/Th17 pathways and inflammatory bowel disease. Inflamm Bowel Dis. 2009 Jul;15(7):1090–100.

121. Oppmann B, Lesley R, Blom B, et al. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. Immunity. 2000 Nov;13(5):715–25.

122. Parham C, Chirica M, Timans J, et al. A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rbeta1 and a novel cytokine receptor subunit, IL-23R. J Immunol. 2002 Jun;168(11):5699–708.

123. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. Nat Rev Immunol. 2008 Jun;8(6):458–66.

124. Sarin R, Wu X, Abraham C. Inflammatory disease protective R381Q IL23 receptor polymorphism results in decreased primary $CD4^+$ and $CD8^+$ human T-cell functional responses. Proc Natl Acad Sci U S A. 2011 Jun;108(23):9560–5.

125. Pidasheva S, Trifari S, Phillips A, et al. Functional studies on the IBD susceptibility gene IL23R implicate reduced receptor function in the protective genetic variant R381Q. PLoS One. 2011;6(10):e25038.

126. Di Meglio P, Di Cesare A, Laggner U, et al. The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans. PLoS One. 2011;6(2):e17160.

127. Sivanesan D, Beauchamp C, Quinou C, et al. IL23R (Interleukin 23 Receptor) Variants Protective against Inflammatory Bowel Diseases (IBD) Display Loss of Function due to Impaired Protein Stability and Intracellular Trafficking. J Biol Chem. 2016 Apr;291(16):8673–85.

128. Sandborn W, Gasink C, Blank M, et al. O-001. A Multicenter, Double-Blind, Placebo-Controlled Phase3 Study of Ustekinumab, a Human IL-12/23P40 mAB, in Moderate-Service Crohn's Disease Refractory to Anti-TNFα: UNITI-1. Inflamm Bowel Dis. 2016 Mar;22 Suppl 1:S1.

129. Elson CO, Cong Y, Weaver CT, et al. Monoclonal anti-interleukin 23 reverses active colitis in a T cell-mediated model in mice. Gastroenterology. 2007 Jun;132(7):2359–70.

130. Sands BE, Chen J, Penney M, et al. A randomized, double-blind placebo-controlled phase 2a induction study of MEDI2070 (anti-p19 antibody) in patients with active Crohn's disease who have failed anti-TNF antibody therapy. J Crohns Colitis. 2015;9(S15-S16).

131. McGovern DPB, Gardet A, Törkvist L, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet. 2010 Apr;42(4):332–7.

132. Zhernakova A, Festen EM, Franke L, et al. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. Am J Hum Genet. 2008 May;82(5):1202–10.

133. Alves de Medeiros AK, Lodewick E, Bogaert DJA, et al. Chronic and Invasive Fungal Infections in a Family with CARD9 Deficiency. J Clin Immunol. 2016 Mar;36(3):204–9.

134. Cao Z, Conway KL, Heath RJ, et al. Ubiquitin Ligase TRIM62 Regulates CARD9-Mediated Anti-fungal Immunity and Intestinal Inflammation. Immunity. 2015 Oct;43(4):715–26.

135. Hung C-Y, Castro-Lopez N, Cole GT. Card9- and MyD88-Mediated Gamma Interferon and Nitric Oxide Production Is Essential for Resistance to Subcutaneous Coccidioides posadasii Infection. Infect Immun. 2016 Apr;84(4):1166–75.

136. Sokol H, Conway KL, Zhang M, et al. Card9 mediates intestinal epithelial cell restitution, T-helper 17 responses, and control of bacterial infection in mice. Gastroenterology. 2013;145(3):591–601.

137. Zhao X-Q, Zhu L-L, Chang Q, et al. C-type lectin receptor dectin-3 mediates trehalose 6,6'-dimycolate (TDM)-induced Mincle expression through CARD9/Bcl10/MALT1-dependent nuclear factor (NF)-κB activation. J Biol Chem. 2014 Oct;289(43):30052–62.

138. Hsu Y-MS, Zhang Y, You Y, et al. The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. Nat Immunol. 2007 Feb;8(2):198–205.

139. Ismail IH, Dronyk A, Hu X, et al. BCL10 is recruited to sites of DNA damage to facilitate DNA double-strand break repair. Cell Cycle. 2016 Jan;15(1):84–94.

140. Van Limbergen J, Wilson DC, Satsangi J. The genetics of Crohn's disease. Annu Rev Genomics Hum Genet. 2009 Jan;10:89–116.

141. Zenewicz LA, Flavell RA. IL-22 and inflammation: leukin' through a glass onion. Eur J Immunol. 2008 Dec;38(12):3265–8.

142. Ouyang W, Kolls JK, Zheng Y. The Biological Functions of T Helper 17 Cell Effector Cytokines in Inflammation. Immunity. 2008 Apr;28(4):454–67.

143. Zelante T, Iannitti RG, Cunha C, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity. 2013 Aug;39(2):372–85.

144. Lamas B, Richard ML, Leducq V, et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nat Med. 2016 Jun;22(6):598-605.

145. Kiryluk K, Li Y, Scolari F, et al. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. Nat Genet. 2014 Nov;46(11):1187–96.

146. Pointon JJ, Harvey D, Karaderi T, et al. Elucidating the chromosome 9 association with AS; CARD9 is a candidate gene. Genes Immun. 2010 Sep;11(6):490–6.

147. Janse M, Lamberts LE, Franke L, et al. Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for IL2, REL, and CARD9. Hepatology. 2011 Jun;53(6):1977–85.

148. Ruland J. CARD9 signaling in the innate immune response. Ann N Y Acad Sci. 2008 Nov;1143:35–44.

149. Wang P, Wu Y, Li Y, et al. A novel RING finger E3 ligase RNF186 regulate ER stress-mediated apoptosis through interaction with BNip1. Cell Signal. 2013 Nov;25(11):2320–33.

150. Natoli G, Chiocca S. Nuclear ubiquitin ligases, NF-kappaB degradation, and the control of inflammation. Sci Signal. 2008 Jan;1(1):pe1.

151. Zhao W, Wang L, Zhang M, et al. E3 ubiquitin ligase tripartite motif 38 negatively regulates TLR-mediated immune responses by proteasomal degradation of TNF receptor-associated factor 6 in macrophages. J Immunol. 2012 Mar;188(6):2567–74.

152. Jin W, Chang M, Sun S-C. Peli: a family of signal-responsive E3 ubiquitin ligases mediating TLR signaling and T-cell tolerance. Cell Mol Immunol. 2012 Mar;9(2):113–22.

153. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015 May;348(6235):648–60.

154. Fernandez M-I, Regnault B, Mulet C, et al. Maturation of paneth cells induces the refractory state of newborn mice to Shigella infection. J Immunol. 2008 Apr;180(7):4924–30.

155. Li X, Madison BB, Zacharias W, et al. Deconvoluting the intestine: molecular evidence for a major role of the mesenchyme in the modulation of signaling cross talk. Physiol Genomics. 2007 May;29(3):290–301.

156. Ochiai K, Muto A, Tanaka H, et al.. Regulation of the plasma cell transcription factor Blimp-1 gene by Bach2 and Bcl6. Int Immunol. 2008 Mar;20(3):453–60.

157. Tanaka H, Muto A, Shima H, et al. Epigenetic Regulation of the Blimp-1 Gene (Prdm1) in B Cells Involves Bach2 and Histone Deacetylase 3. J Biol Chem. 2016 Mar;291(12):6316–30.

158. Martins G, Calame K. Regulation and functions of Blimp-1 in T and B lymphocytes. Annu Rev Immunol. 2008;26:133–69.

159. John SA, Garrett-Sinha LA. Blimp1: a conserved transcriptional repressor critical for differentiation of many tissues. Exp Cell Res. 2009;315(7):1077–84.

160. Tellier J, Shi W, Minnich M, et al. Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response. Nat Immunol. 2016 Mar;17(3):323–30.

161. Alanazi IO, Ebrahimie E. Computational Systems Biology Approach Predicts Regulators and Targets of microRNAs and Their Genomic Hotspots in Apoptosis Process. Mol Biotechnol. 2016 Jul;58(7):460–79.

162. Steeber DA, Tang ML, Zhang XQ, et al. Efficient lymphocyte migration across high endothelial venules of mouse Peyer's patches requires overlapping expression of L-selectin and beta7 integrin. J Immunol. 1998 Dec;161(12):6638–47.

163. Wagner N, Löhler J, Tedder TF, et al. L-selectin and beta7 integrin synergistically mediate lymphocyte migration to mesenteric lymph nodes. Eur J Immunol. 1998 Nov;28(11):3832–9.

164. Salmi M, Granfors K, MacDermott R, et al. Aberrant binding of lamina propria lymphocytes to vascular endothelium in inflammatory bowel diseases. Gastroenterology. 1994 Mar;106(3):596–605.

165. Fu S-H, Lin M-H, Yeh L-T, et al. Targeting tumour necrosis factor receptor 1 assembly reverses Th17-mediated colitis through boosting a Th2 response. Gut. 2015 May;64(5):765–75.

166. Martins GA, Cimmino L, Shapiro-Shelef M, et al. Transcriptional repressor Blimp-1 regulates T cell homeostasis and function. Nat Immunol. 2006 May;7(5):457–65.

167. Kim SJ, Goldstein J, Dorso K, et al. Expression of Blimp-1 in dendritic cells modulates the innate inflammatory response in dextran sodium sulfate-induced colitis. Mol Med. 2014;20:707–19.

168. ain R, Chen Y, Kanno Y, et al. Interleukin-23-Induced Transcription Factor Blimp-1 Promotes Pathogenicity of T Helper 17 Cells. Immunity. 2016;44(1):131–42.

169. Montes de Oca M, Kumar R, de Labastida Rivera F, et al. Blimp-1-Dependent IL-10 Production by Tr1 Cells Regulates TNF-Mediated Tissue Pathology. PLoS Pathog. 2016 Jan;12(1):e1005398.

170. Chen G-B, Lee SH, Brion M-JA, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Hum Mol Genet. 2014 Sep;23(17):4710–20.

171. Colij V, Festen EA, Albers R, et al. Drug repositioning in inflammatory bowel disease based on genetic information. Inflamm Bowel Dis. 2016 [In Press].

172. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014 Feb;506(7488):376–81.

173. Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet. 2016;387(10014):156–67.

174. Heap GA, Weedon MN, Bewshea CM, et al. HLA-DQA1-HLA-DRB1 variants confer susceptibility to pancreatitis induced by thiopurine immunosuppressants. Nat Genet. 2014;46(10):1131–4.

175. Yang S-K, Hong M, Baek J, et al. A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. Nat Genet. 2014;46(9):1017–20.

176. Vétizou M, Pitt JM, Daillère R, et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science. 2015 Nov;350(6264):1079–84.

177. Sivan A, Corrales L, Hubert N, et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science. 2015 Nov;350(6264):1084–9.

SUPPLEMENTARIES

Sequence variant nomenclature used in this review. Please note that we used nomenclature widely used in IBD literature though not always HGVS approved.

Substitution: a sequence change where, compared to a reference sequence, one nucleotide is replaced by one other nucleotide.
Example at DNA level: c.313C>T, substitution of base C (cytosine) for T (thymine) at position 313. (c. refers to a substitution in coding DNA).
Example at protein level: R381Q, substitution of amino acid R (arginine) for Q (glutamine) at position 381.

Splice-site mutation: mutation at specific site at which splicing takes place during the processing of precursor messenger RNA into mature messenger RNA.
Example at DNA level: c.IVS11+1G>C, IVS11 refers to intervening sequence (i.e. intron) 11, this mutation indicates a substitution of G (guanine) for C (cytosine) at position 1 of the given intron.

Insertion: a sequence change where, compared to the reference sequence, one or more nucleotides are inserted
Example: c.3020insC: a C (cytosine) nucleotide insertion at position 3020 in coding DNA. This may or may not lead to a frame-shift mutation.

Frame-shift mutation: a sequence change between the translation initiation (start) and termination (stop) codon where, compared to a reference sequence, translation shifts to another reading frame.
Example: fs1007insC, the mutation c.3020insC, in exon 11 of the *NOD2* gene, results in a frameshift at the second nucleotide of codon 1007 and a leucine-1007-to-proline substitution in the tenth leucine-rich repeat, followed by a premature stop codon. fs1007insC is widely used to denote this mutation in IBD literature.

Copy-number variations (CNV): deletions and duplicates of chromosomal segments.
Example: CNVdel, denotes in this review a deleted allele of a 20kb CNV upstream of *IRGM*.

rs number: reference SNP cluster identification tag assigned by the NCBI.
*Example: r*s1120926, this rs number refers to a G (guanine) to A (adenine) corresponding to c.1142G>A at DNA level and R381Q at protein level.