Pathogenesis of Crohn’s disease
Ray Boyapati¹,², Jack Satsangi¹,² and Gwo-Tzer Ho*¹,²

Addresses: ¹ Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, UK; ²Gastrointestinal Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, EH4 2XU, UK

* Corresponding author: Gwo-Tzer Ho (gho@ed.ac.uk)

F1000Prime Reports 2015, 7:44 (doi:10.12703/P7-44)

All F1000Prime Reports articles are distributed under the terms of the Creative Commons Attribution-Non Commercial License (http://creativecommons.org/licenses/by-nc/3.0/legalcode), which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

The electronic version of this article is the complete one and can be found at: http://f1000.com/prime/reports/b/7/44

Abstract
Significant progress in our understanding of Crohn’s disease (CD), an archetypal common, complex disease, has now been achieved. Our ability to interrogate the deep complexities of the biological processes involved in maintaining gut mucosal homeostasis is a major over-riding factor underpinning this rapid progress. Key studies now offer many novel and expansive insights into the interacting roles of genetic susceptibility, immune function, and the gut microbiota in CD. Here, we provide overviews of these recent advances and new mechanistic themes, and address the challenges and prospects for translation from concept to clinic.

“I am on the edge of mysteries and the veil is getting thinner and thinner.”

Louis Pasteur

Introduction
CD is a debilitating and incurable chronic inflammatory bowel disease (IBD) affecting more than 2.5 million individuals in the Western world and has an increasing incidence in the developing world [1]. CD is characterized by mucosal ulceration and inflammation, which may occur anywhere along the gastrointestinal tract but most commonly affect the distal small intestine. Distinguishing features include discontinuous, transmural inflammation involving the whole thickness of the bowel wall, and an inflammatory response associated with lymphoid aggregates and granulomas [2]. Current treatments include traditional anti-inflammatory agents (corticosteroids), immunomodulators (thiopurines and methotrexate), biological agents with antibodies directed against tumor necrosis factor (anti-TNF), antibiotics, and surgery. Approximately half of CD individuals will require surgery within 10 years of diagnosis and most will experience a disabling course requiring frequent corticosteroids or escalation in immunosuppressive treatment [3,4]. As the most optimal current medical approach (combination of anti-TNF and thiopurines) is effective in only approximately 50% [5], there remains a significant unmet need for novel therapeutics to prevent, alter the natural history of, and ultimately cure CD.

Although the etiology is complex, the most widely accepted hypothesis purports CD as an immune-mediated condition in genetically susceptible individuals, where disease onset is triggered by environmental factors that perturb the mucosal barrier, alter the healthy balance of the gut microbiota, and abnormally stimulate gut immune responses. These three main factors (genetics, gut immune response, and the microbiota) are influenced by the individual’s environmental exposures or triggers (the ‘exposome’) to engage different submechanisms giving rise to ‘Crohn’s diseases’, a concept which is increasingly replacing the traditional paradigm of ‘Crohn’s disease’ as a singular clinical entity with one dominant mechanism (Figure 1). Advances in these fields have catalyzed a decade of spectacular progress in our understanding of CD: a vast and rapidly expanding field with over 18,000 publications in the last 10 years. Here, we provide overviews on CD genetics, immunology, and microbiology (each an enormous area on its own), focus on the key studies which have underpinned progress in our molecular understanding, set them into context, and discuss how these concepts and biological
The successful genome-wide association studies (GWASs) have provided a rational framework for new mechanistic insights and directions for research in CD.

**Genetics**
The most complete picture is from the recent meta-analysis of 15 IBD scans (including ulcerative colitis, UC), involving a combined total of more than 75,000 cases and controls [6]. Overall, 163 IBD loci that meet genome-wide significance thresholds were discovered; this is substantially more than other complex diseases. Most genetic associations are shared between CD and
UC (110 loci), and 30 loci were specifically associated with CD (Figure 2A). These most strongly and consistently implicate themes involving defective intracellular bacteria killing and innate immunity (CARD15/NOD2, IRGM, IL23R, LRRK2, and ATG16L1) and de-regulated adaptive immune responses, namely the interleukin-23 (IL-23) and T helper 17 (Th17) cell pathway (IL23R, IL12B (encoding IL-12p40), STAT3, JAK2, and TYK2) [7]. Dendritic cells (DCs) followed by CD4 T, natural killer (NK), and NKT cells showed the highest enrichment of these susceptibility gene sets when tested in a panel of immune cell subsets, indicating a major role for these cells in CD pathogenesis [6]. It is noteworthy that these GWASs were based predominantly on North American and European populations; the International IBD genetics consortium is in the advanced stages of an expanded meta-analysis of association studies involving non-Caucasian populations together with the populations studied in Europe and North America [8].

On the basis of the GWAS data, the susceptible loci reported so far contribute only 14% of total disease variance [6], but this may be an underestimate. Targeted deep sequencing of key genetic loci has so far shown a negligible impact of rare genetic variants [9], although more detailed and larger-scale whole-genome sequencing studies will provide clearer insight. It is also pertinent that more sophisticated studies involving integrated multi-omics analysis (with profiling panels such as transcriptomics, metabolomics, and epigenomics) are in progress and are likely to provide new insights. Epigenetics is an emerging area of interest [10] in which genome-wide methylation-association studies have identified differential methylation in a number of GWAS-identified susceptibility genes, including TNF, MIR21, HLA, and NOD2, and the Th17 pathway [11,12]. The immediate challenge is to clarify how these genetic variants influence disease-causative mechanisms in CD. Here, we prioritize our review on NOD2, autophagy, and Th17 immune responses as the three areas most strongly implicated in CD pathogenesis.

**NOD2**

NOD2 is a cytosolic pattern recognition receptor (PRR) that controls immunity against intracellular bacteria. Pre-GWAS fine-mapping studies highlighted the NOD2 gene [13,14] as one of the ‘lowest-hanging fruits’ in terms of genetic susceptibility. Three polymorphisms in this gene (amino-acid substitutions Arg702Trp and Gly908Arg and the frameshift FS1007insC) are present in 40% of Western patients with CD [15] and are all found within the leucine-rich repeat region responsible for the recognition of muramyl dipeptide (MDP), a peptidoglycan component of the bacterial cell wall [16]. However, they are absent in Eastern population groups and have a varied prevalence in different Caucasian populations. Of interest, mutations within the NOD2 gene are causative of Blau syndrome, a granulomatous inflammatory disorder affecting the eyes, skin, and joints [17].

NOD2 is expressed in a limited number of tissues that include intestinal epithelial cells (mainly Paneth cells) and monocyte-derived immune cells residing in the lamina propria [18,19]. In both human and murine studies, defects in NOD2 function can affect microbial sensing [20], Paneth cell function and anti-microbial peptide (AMP) production [21], antigen presentation [22], intracellular bacterial killing [23], and innate immune signaling, such as Toll-like receptor (TLR) function [24] and its regulatory role in turning off IL-23-driven Th17 responses [25]. In a recent study, NOD2 activated by microbiota-derived MDP could also promote intestinal stem cell viability and gut epithelial restitution, thus adding a further dimension to its complex role [26]. Overall, although the mechanisms by which NOD2 CD variants contribute to disease remain an enigmatic area, two major, non-mutually exclusive theories have emerged: (1) NOD2 provides critical host anti-bacterial defense and pro-inflammatory responses (Figure 2D), and (2) NOD2 acts to regulate innate immune responses (Figure 2E) [27]. NOD2 activation after recognition of MDP triggers nuclear factor-kappa-B (NF-κB)-dependent signaling [14] but is relatively weak in this respect compared with other PRRs, such as the TLRs [28]. NOD2 can synergize with other PRRs in differential gene regulation, and this synergy is lost in cells expressing CD variant NOD2 [28,29]. NOD2 plays a key role in amplifying the release of certain pro-inflammatory cytokines in this context, particularly IL-1β, IL-6, and IL-23, from DCs and macrophages [18,30]. In contrast, in its regulatory role, deficiency in NOD2 results in enhanced innate TLR signaling. In mice, TLR-mediated IL-12 production is increased in macrophages and DCs deficient in NOD2 [31]. MDP-mediated suppression of TLR-2 responses is enhanced with the normal NOD2 transgene compared with a frameshift polymorphism [32]. Furthermore, pretreatment of monocyte-derived macrophages with MDP leads to inhibition of pro-inflammatory responses to NOD2, IL-1β, and TLR2 and TLR4 in normal individuals but not of TLR-2- and TLR-4-induced responses in cells from CD patients with frameshift polymorphisms [33,34].

In addition to the direct role in innate immunity, several studies show that NOD2 indirectly modulates the gut microbiota, perhaps linked to defective AMP production by Paneth cells [21,35–38]. In mice, NOD2 deficiency does not result in colitis but in defective
processing of intracellular bacteria such as *Listeria monocytogenes* [18]. In humans, a cohort study found a significant association between NOD2 risk alleles and increased abundance of Enterobacteriaceae [39]. In mice, NOD2 deficiency is also associated with ileal dysbiosis [40–42] but this is not consistently replicated [43,44]. NOD2 facilitates autophagic targeting of bacterial pathogens via binding to the autophagy protein ATG16L1, to be discussed later [22,45].

The NOD2 interactome is incredibly complex (Table 1), and all respective network functions and interactions are potentially important in CD, as they are potential novel therapeutic or ‘druggable’ targets [12,27,46–50]. Overall,
isms in the autophagy genes (Following on from NOD2, the discovery of polymorph-
ous interactions in the autophagy genes [52] provides the clearest insight into the pathogenic sequela. The ATG16L1 protein plays an essential role in triggering all forms of autophagy involving the recruitment of microtubule-associated protein 1 light chain 3 (LC3) to membranes. Complex formation of ATG16L1 with ATG12-ATG5 defines the site of LC3 PE conjugation during autophagosome formation. Virtually all the risk of this locus is exerted by the rs2241880 single-nucleotide polymorphism (SNP) coding for a T300A substitution (present in approximately 50% of the general population with twofold increased risk). Recently, Murthy and colleagues [53] showed that amino acids 296 to 299 constitute a caspase cleavage motif in ATG16L1, and that the T300A variant (T316A in mice) significantly increases ATG16L1 sensitization to caspase-3-mediated processing. Here, death-receptor activation or starvation-induced metabolic stress in human and murine macrophages increased the degradation of the T300A or T316A variants of ATG16L1, resulting in diminished autophagy [53].

Two recent complementary studies demonstrate how a defective autophagic response to bacteria can contribute to CD. Cooney and colleagues [22] showed that autophagy cooperates with NOD2: in response to MDP, NOD2 induces autophagy via receptor-interacting serine/threonine-protein kinase 2 (RIPK2), ATG5, ATG7, and ATG16L1 in DCs. This initiates bacterial handling by direct engulfment and subsequent generation of major histo-compatibility complex (MHC) class II for antigen-specific CD4+ T-cell responses in DCs [22]. In the second study, by Travassos and colleagues [45], NOD2 (and NOD1) was shown to recruit ATG16L1 to the plasma membrane at the bacterial entry site to initiate xenophagy. In mice, genetic knock-in of the T300A mutation results in altered cytokine signaling and decreased anti-bacterial response [54]. In a more recent study, ATG16L1 has been shown to negatively regulate NOD1 and NOD2 inflammatory signaling: interestingly, this occurs independently of its role in autophagy [55]. Hence, ATG16L1 may yet have a more complex role in gut inflammatory response.

In the case of another autophagy gene IRGM, a 20-kb deletion polymorphism immediately upstream is associated with CD [56]. Its mouse ortholog Irgm1 contributes to bacterial killing, and Irgm1-deficient mice exhibit increased susceptibility to infections with *Toxoplasma gondii*, *Salmonella typhimurium*, *L. monocytogenes*, and *Mycobacterium tuberculosis* [57–59]. Human macrophages infected with mycobacteria show increased bacterial survival when transfected with IRGM small interfering RNA (siRNA), indicating a role in the control of intracellular mycobacteria [60]. Interestingly, another variant associated with CD (c.313C>T) results in stronger microRNA-196 binding to IRGM and concomitant decrease in IRGM expression, leading to defective autophagy-mediated control of intracellular replication of CD-associated adherent-invasive *Escherichia coli* (AIEC) [61]. Irgm1 knockout leads to exaggerated colonic and ileal inflammation after dextran sulfate sodium (DSS) administration [62]. Of interest, ileitis is not usually a feature of DSS colitis, which suggests a selective function for Irgm1 here [63]. A role for Irgm1 in interferon (IFN)-dependent cellular homeostasis has been proposed by which Irgm1 provides a feedback signal to protect CD4+**

<table>
<thead>
<tr>
<th>Table 1. NOD2 interactome and functional networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Activation</td>
</tr>
<tr>
<td>- Muramyl dipeptide entry into cells (bacterial secretion systems and direct transportation into cytosol)</td>
</tr>
<tr>
<td>- Ligand-NOD2 interaction</td>
</tr>
<tr>
<td>- Cellular localization (for example, recruitment to the plasma membrane)</td>
</tr>
<tr>
<td>• Signaling (for example, RIPK2 interaction and nuclear factor-kappa-B signaling)</td>
</tr>
<tr>
<td>• Regulation (for example, cytoskeleton regulation, epistatic interactions, autophagy, and degradation)</td>
</tr>
<tr>
<td>• Effects</td>
</tr>
<tr>
<td>- Inflammatory responses</td>
</tr>
<tr>
<td>- Adaptive immune responses</td>
</tr>
<tr>
<td>- Antimicrobial functions</td>
</tr>
<tr>
<td>- Facilitating autophagy and xenophagy</td>
</tr>
<tr>
<td>- Gut homeostasis (barrier function, microbiota, and gut epithelial restitution)</td>
</tr>
</tbody>
</table>

NOD2, nucleotide-binding oligomerization domain containing 2; RIPK2, Receptor-interacting serine/threonine-protein kinase 2

NOD2 occupies a strategic hub at the host-microbial level involving autophagy, IL-23/Th17 responses, and gut homeostasis. Current data show that NOD2 CD variants disrupt these pathways, although we still need to understand their relative importance (for example, which pathway is dominant) in order to rationalize the translational potential of this knowledge.

**Autophagy**

Following on from NOD2, the discovery of polymorphisms in the autophagy genes (*ATG16L1*, *IRGM*, and *LRRK2*) from GWASs in CD has triggered significant research in this hitherto unknown area in IBD. Autophagy is a lysosomal degradation pathway that is essential for cellular survival, differentiation, development, and homeostasis [51]. Autophagy principally serves an adaptive role in triggering all forms of autophagy involving the recruitment of microtubule-associated protein 1 light chain 3 (LC3) to membranes. Complex formation of ATG16L1 with ATG12-ATG5 defines the site of LC3 PE conjugation during autophagosome formation. Virtually all the risk of this locus is exerted by the rs2241880 single-nucleotide polymorphism (SNP) coding for a T300A substitution (present in approximately 50% of the general population with twofold increased risk). Recently, Murthy and colleagues [53] showed that amino acids 296 to 299 constitute a caspase cleavage motif in ATG16L1, and that the T300A variant (T316A in mice) significantly increases ATG16L1 sensitization to caspase-3-mediated processing. Here, death-receptor activation or starvation-induced metabolic stress in human and murine macrophages increased the degradation of the T300A or T316A variants of ATG16L1, resulting in diminished autophagy [53].

Two recent complementary studies demonstrate how a defective autophagic response to bacteria can contribute to CD. Cooney and colleagues [22] showed that autophagy cooperates with NOD2: in response to MDP, NOD2 induces autophagy via receptor-interacting serine/threonine-protein kinase 2 (RIPK2), ATG5, ATG7, and ATG16L1 in DCs. This initiates bacterial handling by direct engulfment and subsequent generation of major histo-compatibility complex (MHC) class II for antigen-specific CD4+ T-cell responses in DCs [22]. In the second study, by Travassos and colleagues [45], NOD2 (and NOD1) was shown to recruit ATG16L1 to the plasma membrane at the bacterial entry site to initiate xenophagy. In mice, genetic knock-in of the T300A mutation results in altered cytokine signaling and decreased anti-bacterial response [54]. In a more recent study, ATG16L1 has been shown to negatively regulate NOD1 and NOD2 inflammatory signaling: interestingly, this occurs independently of its role in autophagy [55]. Hence, ATG16L1 may yet have a more complex role in gut inflammatory response.

In the case of another autophagy gene *IRGM*, a 20-kb deletion polymorphism immediately upstream is associated with CD [56]. Its mouse ortholog *Irgm1* contributes to bacterial killing, and *Irgm1*-deficient mice exhibit increased susceptibility to infections with *Toxoplasma gondii*, *Salmonella typhimurium*, *L. monocytogenes*, and *Mycobacterium tuberculosis* [57–59]. Human macrophages infected with mycobacteria show increased bacterial survival when transfected with *IRGM* small interfering RNA (siRNA), indicating a role in the control of intracellular mycobacteria [60]. Interestingly, another variant associated with CD (c.313C>T) results in stronger microRNA-196 binding to *IRGM* and concomitant decrease in *IRGM* expression, leading to defective autophagy-mediated control of intracellular replication of CD-associated adherent-invasive *Escherichia coli* (AIEC) [61]. *Irgm1* knockout leads to exaggerated colonic and ileal inflammation after dextran sulfate sodium (DSS) administration [62]. Of interest, ileitis is not usually a feature of DSS colitis, which suggests a selective function for *Irgm1* here [63]. A role for *Irgm1* in interferon (IFN)-dependent cellular homeostasis has been proposed by which *Irgm1* provides a feedback signal to protect CD4+
lymphocytes from IFN-γ-mediated death [64], and similar mechanisms may apply to other IFN-γ-responsive cell lineages [65]. Collectively, these findings implicate mouse Irgm1 in the regulation of intracellular pathogens or cellular homeostasis; understanding how human IRGM is regulated will be important in order to apply these findings to CD because IRGM is not known to be IFN-γ-responsive [66].

Of note, the role of hypoxia in autophagy (and indeed other mucosal homeostatic systems) has received much interest. Hypoxia is of particular relevance at the gut epithelium-luminal interface, where a unique steep oxygen gradient from the anaerobic lumen to the richly perfused mucosal layer exists. Hypoxia-inducible factors (HIFs) are transcription factors which regulate the induction of genes responsible for cellular adaptation and survival during hypoxia (reviewed in depth by Colgan and Taylor [67]). Pertinently, gut inflammation is associated with increased levels of hypoxia [68] and with high levels of HIFs in murine colitis [69] and IBD [70]. The HIF response is generally considered protective and recently was shown to drive autophagy via HIF1α [71] and increase xenophagic degradation of AIECs [72]. However, there are complexities as HIF1α has a key role in CEACAM6 expression and thus AIEC invasion (discussed in detail later), suggesting that these CD-associated bacteria may take advantage of hypoxic conditions to colonize the intestinal mucosa [73]. HIF1α regulates many genes involved in epithelial barrier function [74–76], including involvement in mucous [77] and AMP production [78]. In murine colitis models, loss of HIF1α expression had a more severe phenotype whereas increased HIF1α levels were protective [69]. The hypoxia response can be modulated by hydroxylase inhibitors (via activation of HIF) [79–81], hyperbaric oxygen [82], and potentially adjustments of lifestyle factors (for example, cigarette smoking). However, this area is complex as heme oxygenase-1 (HO-1) and its metabolic by-product, carbon monoxide, are protective against inflammation and are induced by gut microbiota [83,84].

Beyond xenophagy, autophagy regulates quality-control apparatus, including those involved in control of cell growth, the cell cycle, DNA and membrane repair, and intracellular organelles, such as mitochondria [85]. Defective autophagy can influence cellular homeostasis at the epithelial barrier level in particular and therefore represents a crucial component of disease initiation. Loss of autophagy leading to Paneth cell dysfunction has been a strong focus [86]; these cells are highly metabolically active and specialized enterocytes in the small bowel responsible for AMP production. Individuals with T300A mutation and mice with knocked down/out ATG16L1 and Irgm1 have abnormal Paneth cell morphology lacking in AMP-containing secretory granules [62,86]. The persistence of apoptotic stimuli in the form of metabolic stress, death-receptor activation, or pathogen infection significantly enhances ATG16L1 cleavage, thereby diminishing basal autophagy. Cadwell and colleagues [87] conceptually demonstrated, in this setting and downstream from this, how ‘triggers’ (in this case, murine norovirus infection) may provoke Paneth cell dysfunction and alter response to DSS colitis toward a CD-like phenotype in mice with hypomorphic ATG16L1 function, exemplifying the host-environment interaction in CD.

**Unfolded protein response and endoplasmic reticulum stress**

Following on from autophagy-related epithelial dysfunction, unresolved endoplasmic reticulum (ER) stress in intestinal epithelial cells (IECs) has also emerged as an important factor that initiates gut inflammation relevant to CD (Figure 2I). ER stress-related genes have been implicated by both GWAS (ORMDL3 [88]) and candidate (XBPI [89] and AGR2 [90]) gene approaches. ER stress is induced by the accumulation of unfolded proteins, and cellular adaptation to ER stress is achieved by the activation of the unfolded protein response (UPR), which is an integrated signal transduction pathway that modulates many aspects of ER physiology [91]. Unresolved ER stress is a hallmark of many chronic diseases, and, at the mucosal interphase, UPR is particularly important for highly secretory cells such as Paneth and goblet cells for AMP and secreted mucous barrier, respectively. Kaser and colleagues [89] showed that genetic deletion of UPR transcription factor XBP1 in the intestinal epithelium resulted in loss of Paneth cell function and, interestingly, the development of small-bowel inflammation in mice. This was associated with substantial ER stress and increased inflammatory responsiveness toward microbial and cytokine stimuli. IECs in IBD generally experience unresolved ER stress, even in the absence of overt mucosal inflammation [92]. Of note, UPR is under the influence of primary (genetic) and secondary (environmental) factors and therefore is pivotal in regulating cellular homeostasis [93].

Interestingly, autophagy also cooperates very closely with UPR: autophagy is induced to counter ER stress [94,95] and thus defective autophagy can similarly result in ER stress [96]. The precise interplay between autophagy and ER stress is complex [97,98] and yet to be fully elucidated. Impairment in either of these processes in IECs results in each other’s compensatory engagement and in severe spontaneous CD-like transmural ileitis if
both mechanisms are compromised in mice [99]. Overall, distinct factors can impair autophagy (increased cleavage of ATG16L1) or overwhelm autophagy (ER stress), and subsequent secondary triggers initiate gut inflammation. These data linking three closely related pathways (NOD2, autophagy, and ER stress) clearly demonstrate how disease causation requires specific and critical interaction(s) between host defects and distinct triggers. Independently, these factors may confer only limited risk.

**Immune response: IL-23/Th17 pathway and IL-10**

The dynamic crosstalk between the gut microbiota, IECs, and mucosal immune cells is essential to maintain intestinal homeostasis [100,101]. In CD, the CD4+ T-cell compartment is the most influential and includes Th1, Th17, and Foxp3 regulatory T (Treg) cells [102]. The first IBD GWAS shifted the focus from the traditional Th1 paradigm to IL-23/Th17 responses in CD. Here, Duerr and colleagues [103] demonstrated that carriage of the glutamine allele of Arg381Gln variant of the IL23R gene confers protection against CD, and associations with several SNPs in IL-23/Th17 genes have been consistently shown.

IL-23 has a key role in both innate and T cell-dependent experimental mouse models of colitis [104,105]. IL-23R signaling in T cells leads to enhanced Th17 response, reduced differentiation of Treg cells, and anti-inflammatory IL-10 production [106] (Figure 2F). IL-23 is not indispensable to Th17 differentiation but rather modulates Th17 effector function and pathogenicity [106–108]. IL-23 signaling is mediated through the engagement of heterodimeric IL-23 (composed of the p19 and shared IL-12p40 subunits) with its heterodimeric receptor (comprising IL-23R and IL-12Rβ1), and signals predominantly through JAK2-STAT3 (both with genetic associations with CD) but can also weakly activate STAT1, STAT4, and STAT5 [109]. IL-12 and IL-23 drive differentiation of CD4+ T cells into Th1 and Th17 cells, respectively. IL-23, secreted by macrophages and DCs, together with IL-6 and transforming growth factor-beta (TGFβ), sustains Th17 responses [110].

The gut microbiota regulates both Th17 and Treg cell responses, which appear to be reciprocally related. Th17 cells are absent in germ-free mice, and human fecal transplant into germ-free mice triggers a Th17 response but not with killed-bacteria extracts [111–113]. In health, intestinal Th17 cells are abundant and likely are important components of mucosal host defense. However, the Th17 signature cytokines (IL-17A, IL-17F, IL-22, and IL-26) [114] are particularly elevated in the intestine and serum of patients with IBD, and Th17 cells with an activated phenotype are present in the gut mucosa and blood of patients with CD [115–118]. Therefore, an unrestrained rather than a primarily pathogenic function for Th17 cells is the likely mediator of CD inflammation.

Recently, two discoveries provided further insights into IL-23-Th17 signaling in CD. Firstly, Buonocore and colleagues [119] described a new subset of innate lymphoid cells (ILCs), which rely on IL-23 to induce Th17 responses and colitis [120]. ILCs are important effectors of innate mucosal immunity and tissue remodeling. These previously unknown cells have a lymphoid morphology but lack antigen receptors and myeloid or DC markers. This subset of ILCs (group 3) is defined by their capacity to produce the cytokines IL-17A or IL-22 or both [120]. ILCs possess the ability to regulate CD4+ T cell responses [121]. Secondly, a previously uncultivable organism, segmental filamentous bacteria (SFB), was found to markedly induce a small-bowel Th17 response and promote Th17-dependent autoimmune disease in mice [122]. These studies are exciting as they demonstrate how other immune cells can contribute to Th17 responses. Although the case for SFB in humans is not clearly established, it is a cogent example of how specific microbial stimuli (in this case, a singular microbe) can preferentially induce a Th17 response and immune-mediated pathology.

On the other hand, Treg cells are constitutively present (mostly in gut-associated lymphoid tissue) and maintain mucosal homeostasis predominantly via IL-10. IL-10-deficient mice develop spontaneous colitis in contact with gut commensal microbiota with a Th1/17 pattern but not in germ-free conditions [123]. Genetic variants of the IL-10 gene are associated with IBD, and, intriguingly, rare mutations resulting in complete loss of function in the IL-10 receptor in humans result in extensive clinical manifestations of CD [124]. Several lines of evidence demonstrate the essential role for the microbiota in regulating mucosal Treg cells relevant to CD (Figure 2G).

Specific clusters of the genus *Clostridium*, subsets of which are reduced in CD and include *Faecalibacterium prausnitzii* [125,126], are potent inducers of mucosal and systemic Treg cell responses [127]. Metabolic products of the microbiota, specifically short chain fatty acids [128] (including from *F. prausnitzii* and polysaccharide A [PSA; from *Bacteroides fragilis*]), can also promote Treg cells and limit the Th17 response [129,130]. Recently, T-cell immunology has indeed taken center stage, although the upstream roles for IECs and antigen-presenting cells (DCs and macrophages) converging on the dialogue between the innate and adaptive immune systems are clearly as important (reviewed in depth [101,131]). Inclusively, the gut microbiota is indispensable in educating and shaping the host immune system.
Defining the role of microbiota in Crohn’s disease: recent progress and emerging challenges

Advances in culture-free techniques, next-generation high-throughput sequencing platforms, and the use of larger and more sophisticated human cohorts have ushered in a dramatic era in understanding the role of the gut microbiota in IBD [132,133]. Progression from shallow small-subunit rRNA gene analysis to whole-genome shotgun sequencing and deep functional characterization has been stimulated by a progressive reduction in the cost of high-throughput technologies and provided unique insights into the community structure, genetic repertoire, metabolic products, and function of the complex gut microbiota (total of 10^{12}, which outnumbers somatic cells 10-fold and is an approximately 150-fold larger gene set than the human complement; reviewed in depth [131,134]). The importance of the gut microbiota in the pathogenesis of CD is strikingly demonstrated clinically where the diversion of fecal stream treats and prevents recurrence of CD [135,136]. Several specific mechanistic hypotheses are broadly based on the microbiota’s effects (both general and specific) on mucosal health (for example, epithelial barrier function) and immune system (as antigenic stimuli, regulators of innate immune function (for example, TLR signaling), and balance of Th1/Th2 cell function). Furthermore, mechanisms sustaining a healthy microbial composition (for example, fucosylation [137,138]) and host-microbial symbiosis and containment (barrier function, for example, AMP and mucus; bacterial killing and mucosal immune response - NOD2, autophagy) are increasingly understood as pathogenetic factors in CD.

Determining the ‘high risk’ microbiota in CD thus represents a major research priority. Reduced complexity and diversity of the commensal gut microbiota are consistently demonstrated in CD (and UC) [125,139–142], although a causal effect for this is not yet clear (Figure 2B). In health, shifts in gut microbial composition can be influenced by a number of factors, including host genetics [143]. In CD, earlier studies have shown that host genetic factors (NOD2 and ATG16L1) and disease location (ileal) are associated with mucosal dysbiosis [144], where there is a decrease in Firmicutes, in particular Faecalibacterium prausnitzii, and an increase in Enterobacteriaceae, especially Escherichia coli. Of the phylum Firmicutes, as discussed earlier, Clostridium subsets (including Faecalibacterium prausnitzii) directly induce colonic Treg cells. Reduced Faecalibacterium prausnitzii levels are found in CD and are associated with risk of post-resection recurrence of ileal CD [126], although a separate study in pediatric CD found increased numbers [145]. Escherichia coli, which has acquired specific virulence or pathogenic factors leading to increased adherence and invasive capability (AIEC, is more prevalent in CD [146–149]. In one study, AIECs were isolated in ileal specimens of 36.4% of CD and 6% of controls [146]. Most AIEC strains associated with CD express type 1 pili variants that increase the interaction between AIEC and ileal epithelial cells via CEACAM6 [150] acting as a receptor (Figure 2B). AIECs induce an epithelial inflammatory response and, when phagocytosed by macrophages, are more resistant to xenophagy and induce a persistent inflammatory response by releasing large amounts of TNFα [151,152]. Several factors control AIEC-epithelial interaction: CEACAM6 expression is associated with inflammation, smoking [153], and epigenetic regulation [73].

More recently, there has been considerable interest in the relatively unexplored fields of the mycobiota (fungal community) and virome in CD. Ott and colleagues [154] found an altered fungal profile in the intestinal mucosa of patients with CD and UC compared with healthy controls; interestingly, in contrast to the microbiome, diversity was increased in CD. Analysis of a de novo pediatric IBD cohort by using next-generation sequencing found a distinct difference in mycobiota composition compared with controls with a Basidiomycota dominance [155]. The potential importance of the virome in CD pathogenesis was shown in animal models, whereby viruses in association with gut bacteria affect intestinal biology, leading to inflammation in genetically susceptible hosts [87]. Recent metagenomic sequencing of virus-like particle preparations from fecal samples demonstrated disease-specific viromes for CD and UC [156]. Fascinatingly, this study found CD to be associated with significant expansion of Caudovirales bacteriophages and a reduction in the relative abundance of bacterial taxa, suggesting a potential role for the virome leading to bacterial dysbiosis.

With powerful molecular tools now at our disposal, a number of challenges have emerged in study design and its potential confounders (fecal versus mucosal microbiome, the effects of host genetics, disease activity/duration/location, and drug treatment). Recent studies have focused on combined approaches encompassing all of these factors, including twin studies (to dissect the relative importance of genetics versus environment) [142,157]. Gevers and colleagues [158] analyzed the mucosal and lumen-associated microbiota in treatment-naive CD. In this largest study to date (approximately 450 patients with CD), analysis of the mucosal-associated microbiome confirmed previous findings [126,159,160] of increases in Enterobacteriaceae and decreases in Bacteroidales, Faecalibacterium, and Clostridiales as well as novel associations with other bacterial species. In contrast to an earlier study [159], fecal analysis was less useful and this will impact on how future studies are conducted [158]. Ileal microbiome signatures were
predictive of CD and were observed even in the absence of overt inflammation [161]. Palm and colleagues [162] adopted a creative approach by using the host immune system (IgA-coated sorting followed by 16sRNA sequencing) to home in on the ‘colitogenic’ microbiota. When a smaller cohort of IBD patients and controls was used, IgA sorting revealed 35 species of bacteria that were abundantly coated with IgA in the IBD samples. Several species were found in both healthy and IBD patients but were only highly coated by IgA in patients with IBD. Of interest, gnotobiotic mice colonized with highly coated IgA+ *B. fragilis* elicited more severe colitis compared with those colonized with a *B. fragilis* strain that was IgA−.

Conceptually, mouse studies show that colitogenic microbiota ‘caused’ or induced by host genetic defects (in these cases, NLRP6 and T-bet deficiency, respectively) can be transmissible in co-housing or cross-fostering experiments, leading to increased susceptibility to induced colitis in genetically intact mice [163,164]. It is conceivable that shared environmental factors—notably diet, smoking, and antibiotic use—can result in a ‘high risk’ microbiota that influences susceptibility to CD, although this has not yet been shown in humans [165–167]. This brings in a new dimension, the ‘exposome’, as a factor in modulating the gut microbiota (Figures 1 and 2C).

**Clinical translation**

In this concluding section, we discuss prospects and challenges in clinical translation in CD, where there is a rich seam of creative opportunities from multiple angles. We briefly discuss mechanistic themes, targets, and potential strategies for translation, which are highlighted in detail in Figure 3.

There is an inexorable shift toward mechanistic and molecular stratification that is likely to change current historic clinical classification and eventually lead to better personalized treatment (Figure 1). Rapid improvements in technology now provide the scale, economy, and computational power to allow multi-layered integrative profiling at a metagenomic level (genomic, epigenomic, microbiomic, metabolomic, and proteomic), and a number of studies are already in progress. Furthermore, previously poorly characterized factors such as time and the exposome will now be incorporated [168]. This will provide further novel insights into variability in major clinical phenotypes (for example, early versus adult-onset

---

**Figure 3. Summary of therapeutic targets, underlying mechanisms, and opportunities for translation in Crohn’s disease**

<table>
<thead>
<tr>
<th>Mechanistic themes</th>
<th>Target</th>
<th>Potential Strategies for Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific microbial stimuli/triggers (e.g. AIECs)</td>
<td>Microbiota</td>
<td>Targeting specific groups of bacteria, e.g. AIECs with antibiotics</td>
</tr>
<tr>
<td>Understanding factors leading to dysbiosis</td>
<td></td>
<td>Identifying the ‘high risk’ microbiota</td>
</tr>
<tr>
<td>Loss of key microbial factors regulating epithelial function; APC and T-cell function</td>
<td></td>
<td>Restoring the ‘healthy’ microbiota – Probiotics, FMT</td>
</tr>
<tr>
<td>Loss of epithelial health and barrier function</td>
<td></td>
<td>Correcting factors regulating microbiota, e.g. Diet</td>
</tr>
<tr>
<td>Impaired mucosa and AMP</td>
<td></td>
<td>Identifying triggers and biomarkers of epithelial dysfunction (e.g. ER- and metabolic stress, HIF1α)</td>
</tr>
<tr>
<td>Switch to innate pro-inflammatory epithelial response</td>
<td></td>
<td>Correcting epithelial barrier dysfunction (e.g. mucus, AMP)</td>
</tr>
<tr>
<td>Endogenous DAMP release and immunogenic triggers</td>
<td></td>
<td>Addressing abnormal epithelial-microbial interactions (e.g. CAECAM16)</td>
</tr>
<tr>
<td>Defective bacterial killing</td>
<td>NOD2/Autophagy</td>
<td>Stratification of patients based on NOD2/autophagy function</td>
</tr>
<tr>
<td>Loss of bacterial tolerance</td>
<td></td>
<td>Re-balancing autophagy function, e.g. mTOR inhibitors</td>
</tr>
<tr>
<td>Abnormal bacterial handling</td>
<td></td>
<td>Novel interacting pathways with NOD2 and autophagy</td>
</tr>
<tr>
<td>Breakdown in innate-adaptive immune crosstalk</td>
<td></td>
<td>Direct mucosal immune-profiling in CD (e.g. in clinical trials)</td>
</tr>
<tr>
<td>Unrestricted Th17 activation</td>
<td>Immune Response</td>
<td>Targeting specific cytokines, signalling pathways and immune-trafficking</td>
</tr>
<tr>
<td>Loss of immunoregulatory response</td>
<td></td>
<td>Mucosal immune-modulation (helminth proteins, retinoic acid) and targeting resolution pathways</td>
</tr>
<tr>
<td>Specific mucosal milieu favouring pro- vs. resolution of inflammation</td>
<td></td>
<td>Cell-based therapy to reset the immune response (autologous stem cell transplantation)</td>
</tr>
<tr>
<td>Understanding variability in drug response, disease progression and clinical course</td>
<td>Personalised medicine</td>
<td>Metagenomic profiling based on immune signatures, genomics, epigenomics, microbiomics, metabolomics and proteomics</td>
</tr>
<tr>
<td>New biological pathways in disease causation</td>
<td></td>
<td>Pharmacogenetic stratification: AEs &amp; response to medication</td>
</tr>
<tr>
<td>Identifying clinical vs. molecular phenotypes of CD (e.g. early vs. adult onset disease, extensive vs. limited, inflammatory vs. stricturing)</td>
<td></td>
<td>Advances in endoscopy and imaging techniques for better stratification</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forward prospective cohorts and studying shared gene-environment factors</td>
</tr>
</tbody>
</table>

**Notes:**
- **AE:** adverse effect; **AIEC:** adherent invasive *Escherichia coli*; **AMP:** anti-microbial peptide; **APC:** antigen-presenting cell; **DAMP:** damage-associated molecular pattern; **ER:** endoplasmic reticulum; **FMT:** fecal microbiota transplantation; **HIF1α:** hypoxia-inducible factor 1α; **NOD2:** nucleotide-binding oligomerization domain containing 2; **Th17:** T helper 17 (cells).
CD and inflammatory versus stricturing) and, critically, treatment response. Several recent studies illustrate tantalizing prospects of how we can move from concept to clinic. Haberman and colleagues [161] demonstrated distinct ileal transcriptome and microbiome signature in early-onset CD. In two different studies, Lee and colleagues showed that CD8+ T-cell immune signatures are better at predicting disease course than traditional clinical or serological markers in IBD [169], and that a genetic variant in the FOXO3A gene influences prognosis rather than susceptibility by modulating the inflammatory response in CD [170]. Recently, pharmacogenetics in IBD has seen novel discoveries, including associations at genome-wide significance thresholds with thiopurine-induced leukopenia and NUDT15 gene [171], and the risk of thiopurine-induced pancreatitis with the HLA-DQA1*02:01-HLA-DRB1*07:01 haplotype [172]. Major advances in imaging [173] provide not only a realistic prospect to further stratify CD but a molecular tool to dissect the inflammatory process. Improved magnetic resonance imaging cross-sectional imaging and endoscopic techniques, including video capsule endoscopy, now provide better disease characterization and monitoring [174,175]. Confocal laser endomicroscopy can detect early epithelial dysfunction in predicting relapse [176] and therapeutic response to biological therapy [177] in IBD.

In targeting the microbiota, antibiotic treatment for AIECs, negating the adherence and invasive properties of AIECs [178], manipulating microbial symbiosis factors and metabolome, repopulating the gut habitat with a healthy microbiota via probiotics, endogenous protective commensals (for example, F. prausnitzii) or fecal microbiota transplantation (FMT) may be achievable in the future [179]. A recent review by Sartor [180] succinctly outlined the therapeutic challenges, including posing the questions of whether commensal microbiota can be permanently altered by our interventions and whether endogenous protective commensals eventually can be used as treatment.

As the ‘low-hanging fruits’, NOD2 and ATG16L1 provide an important focus to identify novel ‘druggable’ biological pathways and targets. The identification of vimentin as a NOD2-interacting protein with a role in AIEC handling as a drug target is one of many examples in CD [181]. Stimulating NOD2 and autophagy signaling is another strategy, but perhaps more likely to be successful in a stratified setting (for example, in those patients with defective NOD2 or autophagy) [182,183]. Autophagy inducers (for example, rapamycin) have been used successfully in case reports [184]. However, a clinical trial has shown that everolimus, a mammalian target of rapamycin (mTOR) inhibitor and autophagy inducer, is not efficacious in CD, again highlighting the case for stratification [185]. Paneth cell dysfunction as a focal point provides targets for both upstream (for example, ER stress and autophagy) and downstream (for example, AMP production) factors. Such a platform is highlighted by a recent study using histologic analysis of Paneth cell phenotypes to divide patients with CD into subgroups with distinct pathognomonic and clinical features [36].

It is unsurprising that targeting or inhibition of the immune/inflammatory response has seen the strongest interest in drug development in CD. The success of anti-TNF agents has provided the primer, although this is likely to be the ‘high water mark’ in this area. Following on closely from IBD genetic discoveries, targeting the IL-23/Th17 pathway (and indeed activated T cells) has had mixed success. Ustekinumab, a humanized immunoglobulin G1κ monoclonal antibody against the shared p40 subunit of IL-12 and IL-23, had modest efficacy [186,187], and briakinumab, another anti-IL-12/23 antibody, failed to show benefit. Targeting Th17 responses via secukinumab (anti-IL-17A) and broadalumab (anti-IL-17 receptor) resulted in worse outcomes [188]. Equally unsuccessful in CD were tofacitinib [189] (a JAK inhibitor that is efficacious in UC [190]), fontolizumab (anti-IFN-γ [191]), and abatacept (a CTL4 inhibitor [192]). A number of potential explanations are offered, although more simply the heterogeneity of immune response in CD may confound these ‘general’ clinical trials. It is clear that a re-evaluation is required. Incorporating in-depth immunological analyses during early-phase clinical development should be exploited to gain important insights and this has been discussed in some detail in the IBD research community [193–195]. Beyond this, major immune themes such as resetting the mucosal immune response (autologous stem cell transplantation or more specific cell-based therapies), exploiting mucosal regulatory factors (for example, microbial, helminthic proteins, and dietary factors), and correcting the mucosal milieu, which favors the resolution of inflammation, are likely to feature more prominently.

Conclusions

In the next 10 years, we envisage major progress in (1) stratifying and addressing disease heterogeneity in CD on the basis of dominant molecular mechanism(s); (2) re-design of clinical trials that will follow from (1), where the ‘one size fits all’ approach to new therapeutics requires major re-thinking [193]; and (3) a shift of focus to the causative factors to prevent disease onset and maintain long-term remission in addition to inhibiting...
the abnormal immune/inflammatory response in CD. This will almost certainly rely on simultaneous targeting of genetic, environmental, microbial, and immune factors. In this review, we have focused on the known/ established disease mechanisms, which are framed by recent landmark studies in genetics, immunology, and microbiology in CD. As discussed earlier, it is beyond the scope of this review to cover CD pathogenesis in its entirety. Pertinently, there remain many virtually unexplored concepts and scientific questions. We are at a fascinating inflection point of discovery in CD research. Ambitious goals, including long-term remission, permanent alteration of natural history, and, indeed, curing CD, are not inconceivable for all patients with CD.

Abbreviations

AIEC, adherent-invasive *Escherichia coli*; AMP, antimicrobial peptide; ATG16L1, autophagy-related 16-like 1 gene; CD, Crohn’s disease; CEACAM6, carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross-reacting antigen); DC, dendritic cell; DSS, dextran sulfate sodium; ER, endoplasmic reticulum; GWAS, genome-wide association study; HIF, hypoxia-inducible factor; IBD, inflammatory bowel disease; IEC, innate lymphoid cell; IRGM, immunity-intestinal epithelial cell; IFN-γ, interferon-gamma; IL, interleukin; ILC, innate lymphoid cell; IRGM, immunity-related GTPase family M; LC3, light chain 3; MDP, muramyl dipeptide; NOD2, nucleotide-binding oligomerization domain containing 2; PRR, pattern recognition receptor; SFB, segmental filamentous bacteria; SNP, single-nucleotide polymorphism; Th17, T helper 17; TLR, Toll-like receptor; TNF, tumor necrosis factor; Treg, regulatory T; UC, ulcerative colitis; UPR, unfolded protein response.

Disclosures

The authors declare that they have no disclosures.

References


127. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007, 447:661-78.


173. Roberts CL, Keita AV, Duncan FR, O’Toole PA, Lawrance IC: Environmental Factors that Influence the Onset and Course of Inflammatory and Infectious Disease to a FOXO3-regulated Pathway. Cell 2013, 155:57-69.


