Epidemiological data, detailed molecular studies and recent genome-wide association studies strongly suggest that ulcerative colitis (UC) and Crohn’s disease (CD) are related polygenic diseases that share some susceptibility loci, but differ at others. To date, there are more than 50 confirmed inflammatory bowel disease genes/loci, a number that is widely anticipated to at least double in the next 2 years. Germline variation in IL23R, IL12B, JAK2 and STAT3 is associated with inflammatory bowel disease susceptibility, consistent with the newly described role for IL23 signaling and Th17 cells in disease pathogenesis. Several genes involved in different aspects of bacterial handling are defective only in CD, including NOD2 and the autophagy genes ATG16L1 and IRGM. IL10 and ECM1 are associated with UC, while inherited variation at the HLA region is related to an inflammatory colonic phenotype. The application of genome-wide association studies to inflammatory bowel disease has been successful in defining the genetic architecture of CD and UC and in delivering genuinely novel and important insights into disease pathogenesis. This has unearthed a plethora of attractive targets for the development of future therapeutics. Insights into the natural history of these complex diseases will follow and may enable appropriate patient selection for early aggressive therapy with the view to modifying the disease course.

**Keywords:** autophagy • Crohn’s disease • genetics • genome-wide association study • GLI1 • IL23 • inflammatory bowel disease • innate immunity • NOD2 • Th17 • ulcerative colitis

After birth and during early life, the gut lumen is colonized by an estimated $10^{14}$ bacteria. Thereafter, intestinal homeostasis is crucial for efficient energy extraction from foodstuffs and protection from pathogens. The single-layered epithelium separates the intestinal lumen and the mucosal immune system. Renewing itself every 5–7 days, the epithelium consists of stem cells that differentiate into enterocytes, goblet cells, enteroendocrine cells, Paneth cells and microfold cells (Figure 1). Differentiation and maintenance of these epithelial cells is achieved via a complex network of interacting signaling pathways (WNT, Notch, hedgehog, ephrin and BMP genes). Goblet cells produce the mucus layer at the epithelial surface that acts as a first barrier of defence. Paneth cells secrete specialized antimicrobial peptides known as defensins in man (cytidelins in mouse). Microfold cells and lamina propria dendritic cells (DCs) can directly sample intraluminal microbial components. A combination of these ancient innate immune responses, presentation of antigen to naïve CD4 T cells in the lamina propria with subsequent conditioning into Th17 or Treg subsets (in draining mesenteric lymph nodes) and appropriate epithelial responses to injury (both via direct epithelial signals and indirect signals back from the lamina propria immune cells), serves to maintain intestinal homeostasis.

In the context of this highly complex environment at the mucosal interface, it is perhaps not surprising that an estimated one in 200 people in western Europe and Northern America develop a form of chronic inflammatory bowel disease (IBD), with disruption of intestinal homeostasis and subsequent chronic inflammation [1]. The precise etiologic and pathogenetic mechanisms underpinning the pathogenesis of Crohn’s disease (CD; MIM 26600) and ulcerative colitis (UC; MIM 191390) remain uncertain. However, the presently available data overwhelmingly...
support a hypothesis centered around a dysregulated host immune response to intestinal bacteria (commensal and/or pathogenic) in genetically susceptible individuals.

Factors known to disturb intestinal homeostasis that probably play a role in IBD pathogenesis include alterations in epithelial barrier function, innate immune cells (including macrophages and DCs), lymphocyte function (imbalance between regulator and effector cell function) and stromally derived factors (e.g., TGF-β) [2]. DCs activated by either pathogenic bacteria or disturbed epithelium translocate to the mesenteric lymph node where they instruct naïve T cells to adopt a proinflammatory phenotype. Recent data demonstrates the importance of IL17-producing T cells (Th17) in intestinal inflammation. Th17 cells are involved in clearance of pathogens not dealt with by Th1 or Th2 cells and are potent inducers of tissue inflammation [3].

**Genetic architecture of inflammatory bowel disease**

In addition to the importance of environmental factors (e.g., luminal flora and cigarette smoke), there are considerable epidemiological data that implicate genetic susceptibility in the pathogenesis of CD and UC. Most notably, these include the familial prevalence of IBD, concordance rates in twin pairs and ethnic differences in disease susceptibility. It was the studies of twin pairs that provided the strongest impetus towards further investigation of genetic susceptibility in IBD. Three studies have been carried out in Europe, including Tysk’s important review of the Swedish Twin Registry in 1988 [4–6]. The data from these studies, in Sweden, Denmark and the UK, combine to provide powerful evidence for the role of both genetic and environmental factors in disease susceptibility. The concordance rates for CD in monozygotic and dizygotic twin pairs from these studies are estimated at 37 and 7%, respectively; in UC, the equivalent results are 10 and 3%. The relative risk of developing CD in a first degree relative is 5–35 and for UC 10–15 [7].

Based on these epidemiological data, international teams have been searching for IBD susceptibility genes over the past 15 years (Figure 1). The establishment of a linkage map of the human genome using informative microsatellite markers in the 1990s paved the way for hypothesis-free scanning for loci of association in monogenic and complex genetic disorders. Using this model, nine IBD susceptibility loci (designated IBD1–9) were discovered and replicated to a varying extent. Some of these loci appear to be relatively specific for CD (e.g., IBD1 on 16q [OMIM 266600]) [8,9] and UC (e.g., IBD2 on 12q [OMIM 601458]) [10–12], whereas others are associated with IBD as a whole (e.g., IBD3 on 6p [OMIM 604519]) [13–15]. Despite much initial promise from these genome-wide linkage studies and the 2001 landmark discovery of NOD2 as a CD susceptibility gene, subsequent progress was largely frustratingly slow, and gene ‘discoveries’ were notoriously difficult to consistently replicate.

However, the advent of genome-wide association studies (GWAS) in the past 2–3 years has completely changed the landscape and unparalleled insights into disease pathogenesis have followed (Boxes 1 & 2). While these insights have been and remain the primary objective of these studies, medium- to long-term benefits for patients will include novel drug discovery and advanced predictive models of disease natural history. An emerging theme of complex disease genetics in the past 2 years is of multiple disease variants, each conferring small effects (odds ratios [OR] of <1.20 have been consistently and convincingly replicated in several different studies). In this time period, there has been a number of high-profile GWAS in CD and, later, UC (Table 1) that have, to date, yielded over 50 IBD disease genes (loci) (Table 2). A provisional scheme describing the genetic architecture of IBD is emerging, with many genes conferring susceptibility to CD (and UC) and other genes being specific to either CD or UC (Figure 2). The genes involved in bacterial recognition/innate immunity (e.g., NOD2 and the autophagy genes ATG16L1 and IRGM) are specific to CD; the ECM1 barrier gene and IL10 are specific to UC; the IL23 pathway genes (IL23R, IL12B, JAK2 and STAT3) are common to CD and UC; whereas the HLA region probably confers risk to colonic IBD (i.e., UC and CD colitis) [16–25].

**Insights into disease pathogenesis arising from gene discovery**

While the application of IBD susceptibility gene discovery to diagnostic and prognostic systems remains distant, the key findings are already bearing considerable insights into disease pathogenesis. It is not possible to cover all of the genes in great depth; rather, we will discuss several of the major themes emerging: the role of the IL23 pathway/Th17 signaling in IBD, defective bacterial handling in CD, IL10 signaling in UC, developmental pathways disrupted in IBD and the role of tyrosine phosphatases in disease pathogenesis. Brief summaries of other implicated genes are provided in Box 3.

**IL23 & Th17 signaling**

Duerr and colleagues performed the first large-scale GWAS in IBD (a much more limited Japanese study was arguably the first GWAS) [26], analyzing over 300,000 single nucleotide polymorphisms (SNPs) on the Illumina® HumanHap300 Genotyping BeadChip (Table 1) [18]. After applying Bonferroni correction for multiple testing, only three SNPs remained significant at the 0.05 level. Two of these SNPs were in NOD2 (rs2066843 and rs2076756), effectively providing proof-of-principle for this technique in discovering CD susceptibility genes. The third was a nonsynonymous SNP (rs11209026 Arg381Gln) in the IL23R gene. Several other SNPs within IL23R were subsequently shown to be associated with ileal CD, including some that were independent of the rs11209026 variant. It was proposed that the functional significance of these multiple variants within IL23R may be partly due to differential splicing, as at least six alternatively spliced mRNAs of IL23R are described [18]. Duerr and colleagues demonstrated replication of the IL23R association in an independent case–control association study of non-Jewish and Jewish ileal CD patients and in a family-based association study. Further replication studies for IL23R have been published in adult and pediatric populations in the UK [27–29], France/Belgium [23], Quebec Founder Population [30], Italy [31,32], Canada [33], Holland [34,35], Spain [36], Germany [37], Finland [38], Hungary [39] and Brazil [40].
There was no consistent genotype–phenotype substratification apparent in these studies [28,41,42]. IL23R has also been confirmed as a UC susceptibility gene [19,20,25].

In addition, the CD meta-analysis (Box 2) has confirmed three additional members of the IL23 signaling pathway as susceptibility genes [47]. IL12B, JAK2 and STAT3 were all shown to be associated with CD, and subsequently with UC [19,20], confirming that defects in IL23 signaling confer risk to IBD as a whole.

The timing of the initial IL23R gene discovery study in CD was notable, as it coincided with detailed murine studies of bacteria-induced intestinal inflammation, demonstrating a critical role for IL23 in innate immune pathology [43,44]. The proinflammatory cytokine IL12 has long been thought to be a critical element in the development of pathogenic Th1 CD4+ effector cells. The discovery, in 2000, that IL23 is a heterodimer of p19 and the IL12p40 (IL12B) subunit has also been confirmed with anti-p19 monoclonal antibody significantly decreased IL-10 expression in the inflamed murine intestine. Equivalent findings have been demonstrated with anti-IL10R antibody treatment of H. hepaticus-induced intestinal inflammation as well as decreased proinflammatory cytokines (TNF-α, IFNγ, IL6, IL1β and IL17) in the intestine. Equivalent findings have been demonstrated with anti-IL10R antibody treatment of H. hepaticus-induced T cell-sufficient hosts and CD4+ T-cell transfer into various Rag2−/− recipients [44]. Combining these data demonstrate the crucial role for IL23, but not IL12, in intestinal inflammation, via sustained activation of adaptive and/or innate immune mechanisms.

Another related CD susceptibility gene, CCR6, is expressed by Th17 cells in the memory cell compartment of peripheral blood mononuclear cells (PBMCs) [50]. CCR6 was identified as the CCL20 receptor in 1997 [52]. CCL20 is the only CCR6-triggering chemokine and CCL20 is unable to elicit a biological response through other known chemokine receptors. Together, CCL20 and CCR6 are involved in the maturation of DCs [53]. They play a role in the recruitment of immature DCs and their precursors to sites of potent antigen entry. The CCR6−/− mouse lacks any gross abnormalities in any major organ [54,55]. However, it has underdeveloped Peyer’s patches, and CD11b+ myeloid DCs, known for their functional CCR6 expression, are absent from the subepithelial dome. CCR6−/− mice also have increased subpopulations of T cells in the intestinal mucosa. CCL20 expression, constitutively at low levels in human intestinal epithelium, is induced in inflamed colonic tissue from patients with IBD [56,57], as well as in PBMCs of patients with active

Figure 1. Inflammatory bowel disease pathogenesis.
Inflammatory bowel disease susceptibility genes. APC: Antigen-presenting cell; DC: Dendritic cell; E: Enterocyte; G: Goblet cell; Hh: Hedgehog; MDP: Muramyl dipeptide; MØ: Macrophage; P: Paneth cell; RA: Retinoic acid; TLR: Toll-like receptor.
UC compared with healthy controls [58]. CCL20 has been demonstrated to have antibacterial activity comparable to β-defensins [59]. Furthermore, human β-defensins have been reported to be non-chemokine ligands for CCR6, mediating the in vitro chemotactic activity of β-defensins for immature DCs and T cells.

Defects in bacterial handling in Crohn’s disease

One of the most consistent themes to arise from a primarily genetic approach is of defective bacterial handling in CD. Gene discovery and replication studies have confirmed that NOD2, ATG16L1 and IRGM are associated with CD but not UC. This has redirected research efforts into innate immunity and autophagy.

NOD2

Fine-mapping of the IBD1 region [9] and two positional candidate gene studies identified NOD2 (formerly CARD15) as a CD susceptibility gene in 2001 [60–62]. Structural changes are induced in the leucine-rich repeat region of NOD2 by two SNPs (rs2066845/Gly908Arg and rs2066844/Arg702Trp) and an insertion mutation (rs2066847/3020insC) that leads to a frameshift substitution at Leu1007. The 3020insC mutation results in a premature stop codon and a truncated protein lacking part of the last leucine-rich repeat near the C terminus of the protein. While NOD2 responds to muramyl dipeptide (MDP; the minimal bacterial motif of peptidoglycan, found in the cell wall of Gram-positive and Gram-negative bacteria), no study has yet shown whether direct binding occurs. Unfolding of the NOD2 protein leads to oligomerization and exposure of the CARD domains, providing a platform for RIP2 recruitment and triggering of NF-κB, p38 and Erk MAP kinases.

There has been much debate arising out of apparently conflicting experimental data as to whether the NOD2 mutations represent a ‘loss-of-function’ or ‘gain-of-function’ phenotype. Two different animal models (NOD2+/− and NOD22939C) published simultaneously in Science in 2005 added extra insight but failed to resolve this unsettling paradox [63,64]. Neither mutant developed chronic intestinal inflammation spontaneously; indeed, neither model demonstrated any abnormality in intestinal microstructure. However, when stressed with dextran sodium sulfate (DSS), the NOD22939C mice behaved very differently to their wild type (WT) littermates, developing much more severe colitis with increased mortality (35 vs 0%, respectively). However, the Karin group was unable to breed this NOD2 knock-in mouse (NOD22939C) onto a pure Black-6 background [64]. These in vivo data from a mixed background, therefore, need to be interpreted with some caution. NOD2+/− mice responded less well than WT mice to challenge with oral Listeria monocytogenes [63], an observation that may be related to decreased Paneth cell antimicrobial activity. The Strober lab has argued a gain-of-function model, whereby CD-associated NOD2 mutations result in increased Toll-like receptor (TLR)2-mediated activation of NF-κB signaling and hence excess secretion of Th1 cytokines [65,66]. The same group has subsequently demonstrated a possible therapeutic extension to this work, as addition of exogenous MDP protected WT mice (but not NOD2 knock-out or frame-shift mutant animals) from trinitrobenzene sulfonic acid and DSS colitis by downregulating multiple TLR responses (not just TLR2) [67].

NOD2 is constitutively expressed by macrophages, DCs and Paneth cells [68], the latter being specialized epithelial cells located in the small intestine that secrete antimicrobial peptides, including α-defensins (HDS5 and HD6) into the base of the crypt. Patients with ileal CD have decreased HD5 and HD6 expression, particularly those carrying NOD2 mutations [69], although this latter observation has not replicated in a subsequent study from Australia [70]. PBMCs from CD patients with NOD2 mutations have defective cytokine production and NF-κB activity in response to MDP [71], lending further support to the loss-of-function hypothesis.
Box 1. Genome-wide association studies.

• Up until approximately 2 years ago, researchers were becoming increasingly frustrated by the limitations of genome-wide linkage studies, fine-mapping of these large regions and candidate gene approaches to describing the inherited risk of complex genetic diseases. Despite a few notable exceptions, including NOD2 in CD, progress had become slow and the field muddied by the inability of new disease ‘susceptibility genes’ to consistently replicate in different populations (e.g., DLG5 in CD). It was clear that a radical new approach was required. However, before GWASs could be implemented, a number of scientific and technological hurdles had to be overcome [14]. The first step followed quickly on from the premature announcement by the popular media of the sequencing of the human genome in 2000, with the discovery of over 10\(^{10}\) SNPs. Second was the generation of the Hapmap [120], a high-resolution map of genome variation that utilized measures of the association between SNPs (linkage disequilibrium) to define haplotypes. This provided a means by which ~70% of variation across the genome (in populations of European ancestry) could be described by genotyping 300–500 k SNPs. Third was the development of high-throughput genotyping platforms with the capability of sequencing up to 1 \(\times\) 10\(^6\) SNPs simultaneously on a single chip, all at a reasonable cost [154]. The two commonly used platforms to date are the Illumina\(^\text{®}\) 317k (used in the North American CD GWAS) [18] and the Affymetrix\(^\text{TM}\) 500k (WTCCC study) chips [16]. Other studies have taken a more limited (but more cost-effective) initial approach by typing just nsSNPs (~14k). While much less extensive in their coverage of the genome, these studies have the advantage that any ‘hits’ have almost immediate functional potential (although we now recognize that a large number of causative SNPs function without altering the amino acid sequence). This approach has been successfully employed by the Schreiber group in CD and the UK Inflammatory Bowel Disease Genetics Consortium in UC [19,22].

• When considering GWAS design, it is widely agreed that a large sample size of cases and controls is critical to enable sufficient power to detect association of variants with a modest increase in disease risk at a genome-wide significance threshold (p < 5 \(\times\) 10\(^{-8}\) for combined gene discovery and replication cohorts). The GWAS approach is particularly effective at detecting relatively common variants, as the ability to detect an association at a particular SNP decreases with the frequency of its minor allele (reduction in frequency = reduction in power). Several studies have adopted different approaches to improve study power by attempting to enrich for specific disease-predisposing alleles. Two common approaches have been to minimize population heterogeneity (successfully adopted in the North American CD GWAS [ileal disease only]) [18] and to focus on extreme/familial cases (e.g., early-onset IBD; North American UC GWAS [excluding proctitis]) [25,155]. The true value of such approaches has yet to be determined and is partially cast into doubt by the high success of CD in the WTCCC study, where this was not adopted [16,156].

• Selection of appropriate HCs remains controversial. The WTCCC study demonstrated the success of a common pool of controls for comparison with seven different complex diseases. The WTCCC HCs consisted of both a population-based birth cohort (1958) and opportunistic blood donors. Comparison of these two sources was reassuring in the minimal impact that ascertainment, selection and survival biases had on genotype distributions. Another issue with control populations that will decrease study power is that in most studies, they are not extensively screened for latent disease. For most relatively rare complex diseases, this is not such a big issue as it can be circumvented by sufficiently increasing sample size. For common traits, selection of extreme phenotypes (e.g., very obese or very tall) can also minimize this error.

• Once the cohort has been collected and genotyped on the chosen platform, a series of important quality control steps are implemented prior to a case–control association study. The most critical next step is to distinguish between artefact and a true positive result through validation of positive hits by replication in suitably powered independent cohorts of cases and controls [157]. The hard challenge following this is fine mapping (a particular challenge in tight linkage disequilibrium) to determine causative mutations, followed by functional studies to understand the true biology behind the increased disease risk (or protection). McCarthy and colleagues have recently published a very thorough review of the technical and statistical aspects of all stages of GWAS design, implementation and follow-up [158].

• In the past 2 years, more than 300 replicated associations have been reported in over 70 common diseases and traits [158]. Along with CD, particular success has been obtained in Type II diabetes mellitus (~20 loci) [159] and height (>40 loci) [160,161], a number of common cancers (including colorectal) [162–164] and other continuous traits (e.g., obesity) [165–168].

CD: Crohn’s disease; GWAS: Genome-wide association study; HC: Healthy control; IBD: Inflammatory bowel disease; nsSNP: Nonsynonymous single nucleotide polymorphism; SNP: Single nucleotide polymorphism; UC: Ulcerative colitis; WTCCC: Wellcome Trust Case-Control Consortium.

A recent study has demonstrated decreased production of bacterially-induced IL10 in primary human monocytes from CD patients with the 3020insC mutation [72]. 3020insC was shown to block human IL10 promoter activity. This was not seen in the equivalent murine mutation (2939insC), potentially explaining the lack of a CD phenotype in the NOD2<sup>2939insC</sup> mouse. The 3020insC mutation also prevented NOD2 from forming a tri-molecular complex with IL10 and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1).

**NLRP3**

The NLRP3 gene (1q44) encodes cryopyrin, a protein that controls the inflammation and thus regulates activation of caspase-1 and IL1β. NLRP3, similar to NOD2, is a CATERPILLER gene, in that it has both NOD and leucine-rich repeat domains. Gain-of-function mutations in the NOD domain of NLRP3 cause Muckle–Wells syndrome, familial cold autoinflammatory syndrome and neonatal-onset multisystem inflammatory disease, three hereditary periodic fever syndromes [73,74]. Villani and colleagues demonstrated association of NLRP3 with CD using a tagging SNP candidate gene approach [75]. A total of six SNPs were associated with CD; the lead SNP reaching genome-wide levels of significance.

The effect of the six CD-associated mutations on NLRP3 expression was analyzed in PBMCs [75]. The rs4353135 genotype was significantly associated with altered expression, with homozygosity at the risk allele conferring low expression levels. In a separate assay in cultured monocytes, homozygosity of the risk allele at a different SNP (rs6672995) was associated with decreased
levels of IL1β following stimulation with lipopolysaccharide (LPS) (p = 0.0059). Finally, significant elevation of NLRP3 expression was demonstrated in human CD biopsies (fold change: 4.08; p < 0.0028) and acute trinitrobenzene sulfonic acid-induced colitis (fold change: 9.38; p < 0.0009).

**Autophagy**

Perhaps the most exciting and novel insight to arise from gene discovery in CD is that defective autophagy may play a critical role in disease pathogenesis. Two autophagy genes are now known to be associated with CD: ATG16L1 and IRGM.

Autophagy, the process by which cells digest parts of their own cytoplasm for removal, functions as a homeostatic ‘house-keeping’ mechanism, a pathogenic process in a variety of cancers and neurodegenerative disorders and as an innate immune mechanism against intracellular bacteria [76]. It is a degradative, membrane-based system that plays a critical role in the maintenance of intracellular homeostasis. The defect in autophagy that is present in CD results in the accumulation of non-degraded intracellular components, which can lead to the development of chronic inflammation.

**ATG16L1**

The ATG16L1 gene was originally identified as a CD gene in a German nonsynonymous SNP GWAS and has subsequently been confirmed in UK, North American and French/Belgian GWASs, as well as in a handful of independent replication cohorts [16,22–24,77,78]. The results of logistic regression and haplotype analysis suggested that the CD risk from ATG16L1 was confined to the G allele of rs2241880 (Thr300Ala). It was noteworthy that the variant confers an amino acid change at the evolutionary conserved position 300 of the N terminus, located in exon 9, which is translated into the same reading frame of all six known splice variants of ATG16L1.

Hampe and colleagues demonstrated expression of ATG16L1 mRNA and protein in the colon, small intestine, intestinal epithelial cells and leukocytes. However, they did not detect any difference in protein expression in intestinal tissue between CD and healthy controls [22]. We have demonstrated downregulation of ATG16L1 mRNA in colonic CD biopsies compared with healthy controls in our large microarray dataset [79].

Three different murine models of ATG16L1 deficiency have recently been described in two publications in Nature [80,81]. Cadwell and colleagues generated two mouse lines with gene-trap mutations in Atg16l1 (Atg16l1<sup>HM1</sup> and Atg16l1<sup>HM2</sup>), where Atg16l1 was expressed at 23–37% of expected levels [80]. Intestinal autophagy was shown to be defective in these mice despite normal intestinal morphology. The transcriptional signature of Atg16l1 deficiency was specific to Paneth cells, with disruption of the granule exocytosis pathway (decreased lysozyme in mucus, aberrant disorganized granules and reduced granule number). Furthermore, CD patients with a homozygous Thr300Ala mutation had a strikingly concordant Paneth cell phenotype, with disorganized granules, reduced granule formation, diffuse cytoplasmic staining of lysozyme and positive staining for leptin [80].

Saitoh and coworkers generated mice that lack the entire coiled-coil domain of Atg16l1 [81]. As these Atg16l1<sup>−−</sup> mice died on the first postnatal day, a chimeric mouse was generated by transplanting Atg16l1<sup>−−</sup> fetal liver cells into lethally irradiated
CD45.1+ mice. Under specific pathogen-free conditions, these chimeric mice did not develop spontaneous colitis. However, when stressed with DSS, the 7-day mortality rate was 100% (compared with 0% in WT mice), and large areas of ulceration with severe inflammation were noted in the distal colon. A robust increase in IL1β expression was determined upon stimulation of Atg16l1−/− macrophages with LPS and noninvasive Gram-negative bacteria (e.g., Escherichia coli) but not with invasive strains (e.g., Salmonella typhimurium) [81]. Stimulation with TLR3, TLR4, TLR7, and TLR9 ligands (but not TLR2 or TLR5) also induced IL1β production. Together, these data indicate that reductions in basal autophagy induce IL1β overproduction. The relevance of this was demonstrated in vitro as injection of neutralizing anti-IL1β antibodies significantly improved mortality rate and weight loss in the DSS-stressed chimeric animals.

Two observations suggest that the phenotypes of ATG16L1 and NOD2 mutation are distinct. First, Atg16l1−/− macrophages have a normal inflammatory cytokine response to MDP [81]. Second, Atg16l1−/− animals have no defect in handling oral Listeria monocytogenes challenge (in contrast to NOD2−/− animals) [80].

IRGM

IRGM was identified as a CD gene in the Wellcome Trust Case–Control Consortium (WTCCC) GWAS (Box 2). Two nonfunctional SNPs (rs13361189 and rs4958847) flanking IRGM on chromosome 5q33.1 were strongly associated with CD. Resequencing of the major IRGM exon failed to identify the causative mutation. Subsequently, a 20-kb deletion was identified immediately upstream of IRGM and in perfect linkage disequilibrium with rs13361189 [82]. The frequency of the deletion in a North American reference population was 10%, compared with 15% in CD patients (OR: 1.6; p < 0.01) and 14% in UC (OR: 1.4; p < 0.05). It is intriguing that the deletion was associated with UC in this study, given the absence of association noted for IRGM SNPs in previous studies [19,20]. McCarroll and colleagues went on to demonstrate in vitro that the CD risk (deletion) and protective (reference) haplotypes differentially activate IRGM expression in distinctive cellular contexts [82].

The immunity-related guanosine triphosphatase (IRG or p47 GTPase) genes are critical to innate immunity against intracellular pathogens [83]. While there are 23 complete IRG genes in mice, only three have been identified in man [83,84]. Of these,
Table 2. Inflammatory bowel disease genes/loci discovered (up to April 2009).

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Locus</th>
<th>Genes in locus (n)</th>
<th>Best candidate(s)</th>
<th>GWAS/MA/C GWAS/UC</th>
<th>CD</th>
<th>UC</th>
<th>Other diseases</th>
<th>Best SNP</th>
<th>P (discovery)</th>
<th>P (replication)</th>
<th>P (combined)</th>
<th>OR</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrett et al. (2008)</td>
<td>1p13</td>
<td>7</td>
<td>PTPN22; MAGI3</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td>T1DM; RA</td>
<td>rs2476601</td>
<td>1.81 × 10⁻⁵</td>
<td>0.000101</td>
<td>1.46 × 10⁻⁸</td>
<td>1.31</td>
<td>[17]</td>
</tr>
<tr>
<td>Duerre et al. (2006)</td>
<td>1p31</td>
<td>NA</td>
<td>IL23R</td>
<td>GWAS_CD</td>
<td>Y</td>
<td>Y</td>
<td>Ank spond; psoriasis</td>
<td>rs11209026 (Arg381Gln)</td>
<td>5.05 × 10⁻⁹</td>
<td>7.95 × 10⁻⁴ plus 1.32 × 10⁻¹⁰</td>
<td>6.62 × 10⁻⁹</td>
<td>0.26</td>
<td>[18]</td>
</tr>
<tr>
<td>Silverberg et al. (2009)</td>
<td>1p36</td>
<td>0</td>
<td>OTUD3; PLA2G2E</td>
<td>GWAS_CD</td>
<td>NT</td>
<td>Y</td>
<td></td>
<td>rs20044447</td>
<td>6.8 × 10⁻¹⁰</td>
<td>2.3 × 10⁻⁵</td>
<td>5.1 × 10⁻¹³</td>
<td>0.73</td>
<td>[25]</td>
</tr>
<tr>
<td>Fisher et al. (2008)</td>
<td>1q21</td>
<td>1</td>
<td>ECM1</td>
<td>nsGWAS_UC</td>
<td>N</td>
<td>Y</td>
<td></td>
<td>Rs7377240</td>
<td>1.3 × 10⁻⁴</td>
<td>0.013 and 0.038</td>
<td>2.3 × 10⁻⁶</td>
<td>1.08</td>
<td>[19]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>1q23</td>
<td>2</td>
<td>ITLN1</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs274910</td>
<td>3.5 × 10⁻⁷</td>
<td>0.000481</td>
<td>1.46 × 10⁻⁹</td>
<td>1.14</td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>1q24</td>
<td>0</td>
<td></td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs9288679</td>
<td>4.02 × 10⁻⁷</td>
<td>0.000321</td>
<td>1.53 × 10⁻⁹</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1q32</td>
<td>3</td>
<td></td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs11584383</td>
<td>2.34 × 10⁻⁶</td>
<td>0.000201</td>
<td>1.43 × 10⁻¹¹</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>Silverberg et al. (2009)</td>
<td>1q32</td>
<td>1</td>
<td>IL10</td>
<td>GWAS_UC</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs3024505</td>
<td>1.43 × 10⁻⁵</td>
<td>1.35 × 10⁻¹²</td>
<td>NA</td>
<td>1.46</td>
<td>[25]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>2p23</td>
<td>21</td>
<td>GCKR</td>
<td>MA_CD</td>
<td>M</td>
<td>N</td>
<td>T1DM; CO</td>
<td>rs780094</td>
<td>3.82 × 10⁻⁶</td>
<td>0.00381</td>
<td>3.14 × 10⁻⁷</td>
<td>1.08</td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>2p16</td>
<td>1</td>
<td>NLRP3</td>
<td>MA_CD</td>
<td>M</td>
<td>M</td>
<td></td>
<td>rs13003464</td>
<td>3.44 × 10⁻⁶</td>
<td>0.00565</td>
<td>4.60 × 10⁻⁶</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2q11</td>
<td>5</td>
<td></td>
<td>MA_CD</td>
<td>N</td>
<td>N</td>
<td></td>
<td>rs917997</td>
<td>2.16 × 10⁻⁶</td>
<td>0.0049</td>
<td>2.22 × 10⁻⁶</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>Franke et al. (2008)</td>
<td>2q35</td>
<td>NT</td>
<td>ARPC2</td>
<td>GWAS_UC</td>
<td>NT</td>
<td>Y</td>
<td></td>
<td>rs12612347</td>
<td>8.42 × 10⁻⁶</td>
<td>0.0002</td>
<td>NA</td>
<td>1.18</td>
<td>[21]</td>
</tr>
<tr>
<td>Hampe et al. (2007)</td>
<td>2q37</td>
<td>NA</td>
<td>ATG16L1</td>
<td>nsGWAS_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs2241880 (Thr300Ala)</td>
<td>1.6 × 10⁻⁴</td>
<td>2.7 × 10⁻⁵</td>
<td>1.45</td>
<td>[22]</td>
<td></td>
</tr>
<tr>
<td>Parkes et al. (2007)</td>
<td>3p21</td>
<td>35</td>
<td>MST1</td>
<td>GWAS_WTCCC</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>rs9858542</td>
<td>7.71 × 10⁻⁷</td>
<td>0.01058</td>
<td>4.91 × 10⁻⁸</td>
<td>1.17</td>
<td>[24]</td>
</tr>
<tr>
<td>Librioulle et al. (2007)</td>
<td>5p13.1</td>
<td>0</td>
<td>PTGER4</td>
<td>GWAS</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs1373692</td>
<td>4.1 × 10⁻⁸</td>
<td>3.7 × 10⁻⁴</td>
<td>2.1 × 10⁻¹²</td>
<td>1.46</td>
<td>[23]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>5q31</td>
<td>5</td>
<td>IBD5 (gene?)</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs2188962</td>
<td>4.58 × 10⁻⁹</td>
<td>3.52 × 10⁻¹¹</td>
<td>2.32 × 10⁻¹⁸</td>
<td>1.25</td>
<td>[17]</td>
</tr>
<tr>
<td>Parkes et al. (2007)</td>
<td>5q33</td>
<td>3</td>
<td>IRGM</td>
<td>GWAS_WTCCC</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs13361189</td>
<td>4.38 × 10⁻⁸</td>
<td>0.000664</td>
<td>2.09 × 10⁻¹⁰</td>
<td>1.38</td>
<td>[24]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>5q33</td>
<td>1</td>
<td>IL12B</td>
<td>MA_CD</td>
<td>Y</td>
<td>Y</td>
<td>Psoriasis</td>
<td>rs10045431</td>
<td>8.80 × 10⁻⁹</td>
<td>3.66 × 10⁻⁶</td>
<td>3.86 × 10⁻¹³</td>
<td>1.11</td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>6p25</td>
<td>1</td>
<td>SLC22A23</td>
<td>MA_CD</td>
<td>M</td>
<td>N</td>
<td></td>
<td>rs17309827</td>
<td>2.08 × 10⁻⁶</td>
<td>0.0391</td>
<td>2.74 × 10⁻⁶</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6p25</td>
<td>1</td>
<td>LYRM4</td>
<td>MA_CD</td>
<td>M</td>
<td>Y</td>
<td></td>
<td>rs12529198</td>
<td>7.08 × 10⁻⁷</td>
<td>0.0192</td>
<td>6.96 × 10⁻⁷</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6p22</td>
<td>1</td>
<td>CDKAL1</td>
<td>MA_CD</td>
<td>Y</td>
<td>Y</td>
<td>T2DM; psoriasis</td>
<td>rs6908425</td>
<td>2.52 × 10⁻²</td>
<td>0.000278</td>
<td>8.96 × 10⁻¹⁰</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Fisher et al. (2008)</td>
<td>6p21</td>
<td>Several</td>
<td>MHC</td>
<td>CG and GWAS_UC</td>
<td>Y</td>
<td>Y</td>
<td>Multiple</td>
<td>rs660895</td>
<td>3.1 × 10⁻⁶</td>
<td>0.0035</td>
<td>2.8 × 10⁻⁸</td>
<td>0.73</td>
<td>[19]</td>
</tr>
</tbody>
</table>

Where genes are associated with CD and UC, the p-values are given from the index gene discovery study; this is CD for all the genes presented here.

*Colonic CD only.

*Early-onset disease studied only.

CD: Crohn’s disease; CG: Candidate gene; CD: Coeliac disease; GWAS: Genome-wide association study; IBD: Inflammatory bowel disease; MA: Meta-analysis; NA: Not associated; N: Not available; ns: nonsynonymous; NT: Not tested; OR: Odds ratio; PED: Pediatric; RA: Rheumatoid arthritis; SNP: Single nucleotide polymorphism; T1DM: Type 1 diabetes mellitus; T2DM: Type 2 diabetes mellitus; UC: Ulcerative colitis; WTCCC: Wellcome Trust Case–Control Consortium; Y: Definite association.
### Table 2. Inflammatory bowel disease genes/loci discovered (up to April 2009) (cont.).

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Locus</th>
<th>Genes in locus (n)</th>
<th>Best candidate(s)</th>
<th>GWAS/MA/CG</th>
<th>CD</th>
<th>UC</th>
<th>Other diseases</th>
<th>Best SNP</th>
<th>$P_{\text{discovery}}$</th>
<th>$P_{\text{replication}}$</th>
<th>$P_{\text{combined}}$</th>
<th>OR</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrett et al. (2008)</td>
<td>6q21</td>
<td>0</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>rs7746082</td>
<td>$3.70 \times 10^{-6}$</td>
<td>$7.70 \times 10^{-6}$</td>
<td>$2.44 \times 10^{-10}$</td>
<td>1.17</td>
<td>[17]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>6q25</td>
<td>3</td>
<td>MA_CD</td>
<td>M</td>
<td>N</td>
<td></td>
<td></td>
<td>rs7758080</td>
<td>$7.28 \times 10^{-6}$</td>
<td>0.044</td>
<td>$8.78 \times 10^{-6}$</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>6q27</td>
<td>0</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>rs2301436</td>
<td>$3.30 \times 10^{-7}$</td>
<td>$3.26 \times 10^{-7}$</td>
<td>$1.04 \times 10^{-12}$</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>7p12</td>
<td>0</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>rs1456893</td>
<td>$4.92 \times 10^{-5}$</td>
<td>$1.10 \times 10^{-5}$</td>
<td>$4.60 \times 10^{-9}$</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>8q24</td>
<td>3</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>rs1551398</td>
<td>$4.90 \times 10^{-6}$</td>
<td>0.000109</td>
<td>$4.50 \times 10^{-9}$</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>9p24</td>
<td>3</td>
<td>MA_CD</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>rs10758669</td>
<td>$6.80 \times 10^{-7}$</td>
<td>0.00043</td>
<td>$3.46 \times 10^{-9}$</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>Yamazaki et al. (2005)</td>
<td>9q32</td>
<td>2</td>
<td>TNFSF15 (TL1A)</td>
<td>GWAS and MA_CD</td>
<td>Y</td>
<td>NT</td>
<td></td>
<td>rs4263839</td>
<td></td>
<td></td>
<td>$1.71 \times 10^{-14}$</td>
<td>2.17</td>
<td>[26]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>10p11</td>
<td>3</td>
<td>MA_CD</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>rs17582416</td>
<td>$8.48 \times 10^{-6}$</td>
<td>$2.53 \times 10^{-5}$</td>
<td>$1.79 \times 10^{-9}$</td>
<td>1.16</td>
<td>[17]</td>
</tr>
<tr>
<td>Rioux et al. (2007)</td>
<td>10q21</td>
<td>3</td>
<td>ZNF365</td>
<td>GWAS and MA_CD</td>
<td>Y</td>
<td>NT</td>
<td></td>
<td>rs224136</td>
<td>$7.9 \times 10^{-6}$</td>
<td>$2.9 \times 10^{-7}$</td>
<td>NA</td>
<td>0.6</td>
<td>[77]</td>
</tr>
<tr>
<td>Parkes et al. (2007)</td>
<td>10q24</td>
<td>1</td>
<td>NKK2.3</td>
<td>GWAS_WTCCC</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>rs10883365</td>
<td>$1.41 \times 10^{-8}$</td>
<td>0.00373</td>
<td>$3.71 \times 10^{-10}$</td>
<td>1.18</td>
<td>[24]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>11q13</td>
<td>1</td>
<td>C11orf60</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs7927894</td>
<td>$1.43 \times 10^{-7}$</td>
<td>0.000732</td>
<td>$1.32 \times 10^{-9}$</td>
<td>1.16</td>
<td>[17]</td>
</tr>
<tr>
<td>12q12</td>
<td>3</td>
<td>LRRK2; MUC19</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs11175593</td>
<td>$1.33 \times 10^{-7}$</td>
<td>0.000165</td>
<td>$3.08 \times 10^{-10}$</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>12q13</td>
<td>1</td>
<td>GLI1</td>
<td>CG</td>
<td>M</td>
<td>M</td>
<td></td>
<td>rs2228226</td>
<td>0.0026</td>
<td>0.042</td>
<td>0.0002</td>
<td>1.19</td>
<td>[79]</td>
</tr>
<tr>
<td>Silverberg et al. (2009)</td>
<td>12q15</td>
<td>3</td>
<td>IFN-γ, IL26, IL22</td>
<td>GWAS_UC</td>
<td>NT</td>
<td>Y</td>
<td></td>
<td>rs1558744</td>
<td>$5.5 \times 10^{-10}$</td>
<td>$1.5 \times 10^{-4}$</td>
<td>$2.5 \times 10^{-12}$</td>
<td>1.35</td>
<td>[25]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>13q14</td>
<td>3</td>
<td>NOD2</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs3764147</td>
<td>$1.61 \times 10^{-7}$</td>
<td>$1.33 \times 10^{-7}$</td>
<td>$2.08 \times 10^{-13}$</td>
<td>1.25</td>
<td>[17]</td>
</tr>
<tr>
<td>16q12</td>
<td>NA</td>
<td></td>
<td>CG and MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs2066847</td>
<td>$1.49 \times 10^{-14}$</td>
<td>0.0135</td>
<td>$2.98 \times 10^{-24}$</td>
<td>3.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17q12</td>
<td>4</td>
<td>CCL2, CCL7, CCL11, CCL8</td>
<td>MA_CD</td>
<td>M</td>
<td>N</td>
<td></td>
<td>rs9911804</td>
<td>4.02 $\times 10^{-6}$</td>
<td>0.0135</td>
<td>$1.07 \times 10^{-6}$</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>17q21</td>
<td>17</td>
<td>ORMDL3</td>
<td>MA_CD</td>
<td>Y</td>
<td>NT</td>
<td>Asthma</td>
<td>rs2872507</td>
<td>$2.12 \times 10^{-6}$</td>
<td>0.000292</td>
<td>$5.00 \times 10^{-9}$</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>17q21</td>
<td>4</td>
<td>STAT3</td>
<td>MA_CD</td>
<td>Y</td>
<td>NT</td>
<td></td>
<td>rs744166</td>
<td>$5.94 \times 10^{-6}$</td>
<td>9.15 $\times 10^{-6}$</td>
<td>$6.82 \times 10^{-12}$</td>
<td>1.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parkes et al. (2007)</td>
<td>18p11</td>
<td>1</td>
<td>PTPN2</td>
<td>GWAS_WTCCC</td>
<td>Y</td>
<td>M</td>
<td>T1DM; CO</td>
<td>rs2542151</td>
<td>$4.56 \times 10^{-8}$</td>
<td>0.0479</td>
<td>$3.16 \times 10^{-8}$</td>
<td>1.15</td>
<td>[24]</td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>18q11</td>
<td>0</td>
<td>GPX4</td>
<td>MA_CD</td>
<td>M</td>
<td>N</td>
<td></td>
<td>rs8098673</td>
<td>$3.18 \times 10^{-6}$</td>
<td>0.0443</td>
<td>$2.88 \times 10^{-6}$</td>
<td>1.05</td>
<td>[17]</td>
</tr>
<tr>
<td>19p13</td>
<td>2</td>
<td>TNFRSF6B</td>
<td>MA_CD</td>
<td>M</td>
<td>N</td>
<td></td>
<td>rs4807569</td>
<td>$1.16 \times 10^{-8}$</td>
<td>0.00347</td>
<td>$2.12 \times 10^{-9}$</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13</td>
<td>9</td>
<td>MA_CD</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>rs2315008</td>
<td>$6.30 \times 10^{-8}$</td>
<td>0.01749</td>
<td>NA</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>21q21</td>
<td>0</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>rs1736135</td>
<td>$2.06 \times 10^{-5}$</td>
<td>$4.58 \times 10^{-5}$</td>
<td>$7.40 \times 10^{-9}$</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>21q22</td>
<td>1</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>rs762421</td>
<td>$1.08 \times 10^{-5}$</td>
<td>$1.59 \times 10^{-5}$</td>
<td>$1.41 \times 10^{-11}$</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Barrett et al. (2008)</td>
<td>21q22</td>
<td>0</td>
<td>PSMG1</td>
<td>MA_CD</td>
<td>Y</td>
<td>N</td>
<td></td>
<td>rs2836878</td>
<td>$6.01 \times 10^{-8}$</td>
<td>0.01131</td>
<td>NA</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Kaser et al. (2004)</td>
<td>22q12</td>
<td>NA</td>
<td>XBP1</td>
<td>CG</td>
<td>M</td>
<td>M</td>
<td></td>
<td>rs35873774</td>
<td>0.0011</td>
<td>0.53 plus $1.6 \times 10^{-5}$</td>
<td>$1.6 \times 10^{-5}$</td>
<td>0.74</td>
<td>[56]</td>
</tr>
</tbody>
</table>

Where genes are associated with CD and UC, the p-values are given from the index gene discovery study; this is CD for all the genes presented here.
*C: Colonic CD only.
*Early-onset disease studied only.
CD: Crohn’s disease; CG: Candidate gene; CO: Coeliac disease; GWAS: Genome-wide association study; IBD: Inflammatory bowel disease; M: Evidence for association not definitive; MA: Meta-analysis; N: Not associated; NA: Not available; ns: nonsynonymous; NT: Not tested; OR: Odds ratio; PED: Pediatric; RA: Rheumatoid arthritis; SNP: Single nucleotide polymorphism; T1DM: Type 1 diabetes mellitus; T2DM: Type 2 diabetes mellitus; UC: Ulcerative colitis; WTCCC: Wellcome Trust Case–Control Consortium; Y: Definite association.
IRGM is the human homologue of murine IRGMI (Lrg-47), a key mediator of IFNγ-induced autophagy [85]. Human IRGM is constitutively expressed and contains no IFN-inducible elements in the promoter, leading to initial reports that humans lacked a p47 resistance system [84]. However, despite being unresponsive to IFNγ, IRGM plays a critical role in IFNγ-induced or conventionally induced autophagy in human macrophages [85]. siRNA knockdown of IRGM in human macrophages infected with mycobacteria leads to defective autophagy, with increased bacterial survival compared with control cells [85]. Lrg-47 knock-out mice (Lrg-47−/−) fail to control Mycobacterium tuberculosis infection (succumbing after several weeks following aerosol or intravenous inoculation), despite the formation of well-organized granulomas replete with infiltrating lymphocytes [86]. Lrg-47−/− macrophages have defective bacterial killing, with impaired maturation of M. tuberculosis-containing phagosomes; Lrg-47 is recruited to these vesicles in WT cells [85]. Similarly, Lrg-47−/− mice infected with Mycobacterium avium are unable to control bacterial replication; although surviving the acute illness, they succumb 11–16 weeks later [87]. The phenotype of these animals suggests a regulatory role for Lrg-47 on lymphocyte survival, as they have a profound systemic anemia and lymphopenia, with marked lymphocyte deficiency in mycobacterium granulomas [87].

**Developmental genes**

Another theme emerging from genetic studies is that signaling pathways critical to normal mammalian gut development are dysregulated in IBD pathogenesis. The key molecular pathways involved in the development of the GI tract are the hedgehog, BMP, Notch, WNT/β-catenin, hox, sox and eph/ephB signaling pathways [88]. Patterned gene expression within the endoderm and surrounding mesoderm regulates the morphogenesis, differentiation and boundaries of the developing gut [89]. This is a highly complex and tightly regulated process with multiple overlapping signaling pathways directing cell fate and tissue patterning. Gene discovery studies have demonstrated a role for NK2 transcription factor-related, locus 3 (NKX2.3) and glioma-associated oncogene homolog 1 (GLI1; a key transcriptional regulator of the hedgehog signaling pathway) in IBD.

**NKX2.3**

The NKX2.3 gene was identified as a CD susceptibility gene in the WTCCC study [16,24], with early replication in a German tagging SNP study [20] and subsequent association with UC [19,20]. NKX2.3 is a member of the NKX family of homeodomain-containing transcription factors implicated in cell type specification, and maintenance of differentiation in a number of different tissue types. In mice, NKX2.3 is expressed in the gut mesoderm and branchial arches from E9.5, persisting in the hindgut until adulthood [90]. During later stages of murine development, expression is confined to the inner ring of the gut mesoderm (smooth muscle), immediately underlying the mucosal endodermal epithelium [91]. In the adult, highest levels of expression are noted in the ileum, with lowest levels of expression in the duodenum and rectum. In addition to smooth muscle expression, NKX2.3 is expressed in numerous mesenchymal cells in the intestinal lamina propria and in endothelial cells of small blood vessels.

A significant proportion of NKX2.3−/− mice die from acute intestinal malabsorption within 2 weeks of birth [91,92]. In the early postnatal period, these animals have steatorrhea and bloody diarrhea, with lipid-staining vesicles predominant in the intestinal mucosa. While TNF-α expression was increased in one study of mutant intestine, no acute inflammation was detected at any postnatal stage [91]. Those mice that survive this period have reduced intestinal length and intestinal distension secondary to hyperplasia (not dilatation), but otherwise appear healthy as adults and are fertile [91]. Meanwhile, heterozygotes are viable and appear to develop normally [91,92]. The homozygous mutant animals have severe defects in both the spleen and intestine. Of note, one in three of these mice are asplenic; the spleens in the remaining animals are approximately tenfold smaller than WT animals and demonstrate severe structural abnormalities [91,92]. They lack a marginal zone, have abnormal segregation of B and T lymphocytes and a paucity of macrophages in the red pulp. They are noted to have a smaller cecum than WT mice, from E15.5 onwards [91]. In the small intestine, villus formation is delayed during development; however, normal differentiation of epithelial cell types is present in surviving adult mutants [92]. Furthermore, these animals have significant abnormalities of gut-associated lymphoid tissue, with fewer and smaller Peyer’s patches and lymphoid aggregates than WT mice. Moreover, MadCAM-1 expression is significantly down-regulated (~tenfold reduction) in areas that normally express NKX2.3, providing a potential mechanism by which they have deranged lymphocyte homing [91]. The importance of lymphocyte homing via MadCAM-1 in IBD pathogenesis has been clearly demonstrated as testified by the clinical efficacy of the anti-α4 monoclonal antibody, natalizumab, in CD [93].

**GLI1**

We have shown that germ-line variation in GLI1 (within IB2 on 12q13) is associated with IBD [94]. This was demonstrated in three independent northern European populations (Scotland, England and Sweden), but awaits further independent replication. A nonsynonymous SNP (rs2228226) in a highly conserved region of GLI1, next to a known transactivation domain, was associated with IBD and was functionally defective in vitro [94]. GLI1 is a major transcriptional regulator of the hedgehog signaling pathway. Hedgehog, through exclusively paracrine signaling (epithelium to mesenchyme) [95], plays critical roles in GI tract development, homeostasis and disease [96]. It is also centrally involved in chronic injury, inflammation and repair in several different organs including muscle [97], liver [98,99] and lung [100,101]. Sonic hedgehog is critical for T-lymphocyte development [102], myeloid cell maturation in the spleen [103] and peripheral human CD4+ T-cell activation [104,105], and is a direct target of NF-κB [106].
The GLI1 gene discovery has led to important potential insights into disease pathogenesis. GLI1 mRNA (and by inference hedgehog pathway activity) was shown to increase along the length of the healthy adult colon in man (mirroring the increasing luminal bacterial load in the distal colon) and was reduced in all forms of colonic inflammation studied (UC, colonic CD and non-IBD inflammation) [94]. GLI1-/- mice develop normally, in contrast to lines deficient in all other hedgehog components [96,107]. However, inflammatory challenge (3% DSS) to mice with a 50% reduction in functional GLI1 (GLI1+/-LacZ) led to early, severe colitis with substantial mortality and significant upregulation of IL12, IL17 and IL23 compared with WT littermates [94]. Furthermore, myeloid CD11b- and CD11c-positive cells were identified as direct targets of hedgehog signaling in the lamina propria following inflammatory challenge [94]. Subsequently, Zacharias and colleagues have demonstrated that downregulation of hedgehog signals in the small intestine (overexpression of the pan-hedgehog inhibitor HHIP under the villin promoter) results in robust, spontaneous ileal inflammation and IgA-positive dermatitis [108]. Thus, the body of evidence to date strongly suggests that epithelial hedgehog serves to dampen lamina propria immune responses in the steady state. This may provide options for therapeutic intervention given the intense biotechnology interest in small-molecule hedgehog agonists and antagonists that are in late preclinical and early clinical testing for a variety of different indications [201].

Intracellular tyrosine phosphatases

PTPN2

The PTPN2 gene, identified as a CD susceptibility gene in the WTCCC study [16,24], encodes the T-cell protein tyrosine phosphatase (TCPTP) that is a key negative regulator of inflammatory responses. TCPTP is located intracellularly, containing no transmembrane domains [109]. It is expressed in all tissues and at all stages of development. There are two splice variants in humans (48 kDa: endoplasmic reticulum targeted; 45 kDa: nuclear and cytoplasm targeted) but only one is mice (45 kDa). Expression is ubiquitous, but highest in cells of hematopoietic origin; the TCPTP knock-out mouse dies of severe anemia by postnatal weeks 3–5 [110]. Interestingly, this mouse, although apparently normal through embryonic development, develops a severe systemic inflammatory/wasting disease characterized by running, hunched posture, diarrhea, splenomegaly (caused by macrophage infiltrates) and weight loss [110]. The severe immunosuppression is characterized by defective T-cell-dependent B-cell responses. TCPTP is an important regulator of colony stimulating factor (CSF)-1 and mononuclear phagocyte development (Tppp-/- show increased granulocyte macrophage precursors) [111]. Furthermore, the Tppp-/- mice show widespread lymphocytic infiltrates in non-lymphoid tissues, correlating with increased IFNγ, TNF-α, IL12 and nitric oxide, and increased LPS sensitivity [112]. Increased cytokine production is detected as early as day 3 postpartum, preceding the onset of systemic disease (symptoms develop at 10–14 days postpartum), suggesting that this is a primary abnormality in these animals. IL12/IL23-p40 expression is massively upregulated in the liver and, to a lesser extent, in the spleen at day 3. Intraperitoneal injection of 4 µg LPS to 20-day-old Tppp-/- mice led to the development of septic shock, not seen in WT littermates, with massive induction of serum IFNγ.

T-cell protein tyrosine phosphatase has been shown to dephosphorylate distinct substrates, including the insulin receptor [113], EGF receptor (EGFR) [114], JAK1 [115], JAK3 [115], STAT1 [116], STAT3 [117] and, most recently, STAT6 [118,119]. As a result, it can regulate signaling pathways that are induced by various growth factors (e.g., EGF) and cytokines (e.g., TNF-α). In the case of TNF-α, MAPK signaling is suppressed by TCPTP (it interacts with TRAF2 to dephosphorylate and, hence, inactivate Src tyrosine kinases) [120]. TCPTP-deficient cells show enhanced IL6 production.

Figure 3. Genetic architecture of inflammatory bowel disease. The figure is limited to genes with definitive evidence of association in IBD from Table 2. *UC genes where association has been inadequately tested in CD. +CD genes not fully tested in UC.

CD: Crohn’s disease; IBD: Inflammatory bowel disease; UC: Ulcerative colitis.
**Box 3. Other confirmed inflammatory bowel disease genes/loci of interest.**

**ARPC2**
The microbial equivalent of this gene associates with the GWAS at the protein level. In man, mutant WAS causes Wiskott-Aldrich Syndrome; UC has been reported as the first manifestation of this disease [169]. The WAS protein is involved in the regulation of Treg cells [170].

**ECM1**

In the UK nsSNP GWAS, the strongest novel association was with a locus on 1q21.2 containing ECM1, MRPS21, PRPF3 and TARS2 [19]. While additional tagging at ECM1 provided evidence of association in two nsSNPs within the gene, conditional regression analysis showed that these did not fully explain the association. Therefore, further fine mapping and deep resequencing is required to determine whether the causative mutations lie within ECM1 or elsewhere in the haplotype block. Despite this, ECM1 remains an excellent biological candidate for a UC susceptibility gene. It is expressed throughout the intestine, interacts with the basement membrane, inhibits MMP9 and strongly activates NF-κB. It is noteworthy that the two strongest ECM1 markers (rs3737240 and rs13294) were also associated with ankylosing spondylitis [171], given the overlap between these two diseases clinically.

**Gene deserts and PTGER4**

One of the most interesting and unexpected phenomena that has arisen from GWASs is the strong association of SNPs in large gene deserts (frequently > 1 Mb from any known gene). It is thought that at least some of these areas are related to long-range, cis-acting regulatory elements that influence gene expression. One of the first such regions described in CD was on 5p13.1. Within a 1.25 Mb gene desert, Libioulle and colleagues identified association in a 250 kb region [23]. A bioinformatics-based approach demonstrated that several of the CD-associated SNPs altered expression of prostaglandin receptor EP4 (PTGER4). PTGER4-knockout mice develop severe colitis when stressed with DSS; this is not seen with other prostanoid receptor knockout animals [172]. Furthermore, an EP4-selective agonist was seen to ameliorate DSS-induced colitis in WT animals.

**HLA region**

There is a strong association between genes of the HLA region (6p21), involved in regulating the immune response, with colonic IBD (i.e., UC and CD colitis). The association with UC has been described for over 10 years. The most consistent association was with the rare HLA DRB1*0103 allele at the HLA class II region [173–175]. In the recent UK nsSNP GWAS in UC, MHC tagging SNPs were genotyped in the discovery cohort [19]. The design of this study, with high marker density in this region, demonstrated that the association was confined to a 400 kb haplotype block (containing BTNL2, HLA-DRA, HLA-DRB5, HLA-DRB1 and HLA-DQA1) and does not extend beyond the recombination hotspots at either end. Stratification of BTNL2 genotypes by DRB1*1502 status demonstrated clear residual association at BTNL2, in contrast to a previous Japanese study [176]. Most recently, the German UC GWAS demonstrated genome-wide levels of significance at three SNPs in this region [21].

**IFNγ, IL26 and IL22**

In the recent North American UC GWAS, association was demonstrated at 12q15 (independent signals at rs1558744 and rs2870946) [25]. This region is particularly noteworthy as it maps to the edge of the IBD2 region and contains IFNγ, IL26 and IL22. These genes are 44–137 kb from rs1558744, while rs2870946 is located in IL26. IL26 and IL22 are both secreted by Th17 cells; as a result, they can induce IL10-producing Treg cells from naive CD4 T cells [177].

**Intelectin-1**

Intelectin-1 (ITLN1) is a lectin that recognises sugar motifs specific to bacterial cells walls. ITLN1 is expressed in the gut, specifically Paneth cells, and induced by parasitic infection in a STAT6-dependent manner [182]. It will therefore be of interest to see whether ITLN1 expression is decreased in the terminal ileum or colon from smokers, potentially decreasing pathogen clearance.

**LRRK2**

LRRK2 is implicated in the pathogenesis of Parkinson’s disease, is expressed in myeloid cells and is strongly implicated in autophagy [183].

**TNFSF15 (TL1A)**

In 2005, Yamazaki and colleagues from Japan reported the first GWAS in CD [26]. This study was limited in its genome coverage and had a very small index cohort. However, the major result, association of TNFSF15 with CD, has stood the test of time, robustly replicating in the CD meta-analysis [17]. TL1A expression is increased in macrophages and T cells in inflamed CD ileum and colon [184]. It is produced by antigen-presenting cells in response to FcyR signaling. Murine colitis is ameliorated by administration of neutralising TL1A antibodies by suppressing both Th1 and Th17 responses [185].

CD: Crohn’s disease; DSS: Dextran sodium sulfate; GWAS: Genome-wide association study; IBD: Inflammatory bowel disease; nsSNP: Nonsynonymous single nucleotide polymorphism; SNP: Single nucleotide polymorphism; UC: Ulcerative colitis; WT: Wild-type.
Box 3. Other confirmed inflammatory bowel disease genes/loci of interest.

X-box-binding protein 1 (XBP1)
The XBP1 gene is an excellent biological candidate for IBD susceptibility based on expression, function and position. It is located on 22q12 in a previously identified linkage region [186–188]. The complex haplotype structure of XBP1 and deep resequencing in 564 IBD patients and 282 healthy controls strongly indicated that multiple, rare, private SNPs contribute to the association with IBD [189], although this observation awaits independent replication. Several of these rare nsSNPs demonstrated decreased transactivating function in a series of in vitro assays. Endoplasmic reticulum stress and XBP1 mRNA expression was increased in the inflamed and noninflamed terminal ileum of patients with CD and inflamed and noninflamed colon of patients with CD and UC. Mice with deletion of Xbp1 specifically in intestinal epithelial cells (Xbp1lox/loxVCre) demonstrate increased endoplasmic reticulum stress and mild, patchy, spontaneous mucosal inflammation. These animals were almost completely devoid of Paneth cells, in addition to a minor defect in small intestinal (but not colonic) goblet cells. Oral challenge with Listeria monocytogenes was impaired, with significantly increased numbers of bacteria retrieved in feces and the liver, but not spleen compared with WT animals. Additionally, the knockout animals were more susceptible to DSS colitis than WT animals [189].

ZNF365
ZNF365 is a zinc finger protein that is a cause of susceptibility to uric acid nephrolithiasis in a founder population in Sardinia [190].

CD: Crohn's disease; DSS: Dextran sodium sulfate; GWAS: Genome-wide association study; IBD: Inflammatory bowel disease; nsSNP: Nonsynonymous single nucleotide polymorphism; SNP: Single nucleotide polymorphism; UC: Ulcerative colitis; WT: Wild-type.

production in response to TNF-α, an effect blocked with inhibition of ERK1/2 activation [120]. TCPTP dephosphorylates nuclear STAT6 (PTPN1 does the same to cytoplasmic STAT6), inducing the expression of IL4 target genes [118]. The 45-kDa TCPTP recognises mutant EGFR (a common rearrangement with a truncated protein with an in-frame deletion of 267 amino acids that is detected in glioblastoma, breast, lung and prostate cancers) as a cellular substrate and dephosphorylates it, leading to decreased MAPK, ERK2 and PI-3K levels, suppressing the tumorigenicity of mutant EGFR-expressing glioblastoma cells [121].

PTPN22
Evidence for the association of the PTPN22 gene with CD was provided in the meta-analysis by Barrett et al. [17]. It is of some considerable interest that the lead SNP (R620W) has been associated with a wide range of conditions with prominent humoral autoimmunity (including Type 1 diabetes mellitus, rheumatoid arthritis, autoimmune thyroid disease, myasthenia gravis, systemic sclerosis, vitiligo, Addison’s disease and alopecia areata), yet in CD, this SNP is actually protective [122]. PTPN22 is expressed in many hematopoietic cell types, notably T cells. The 620W autoimmune risk allele behaves as a gain-of-function mutation as it results in a phosphatase with higher catalytic activity and more potent negative regulation of T-cell activation [123,124]. By contrast, knockout mice (Lyp is the mouse ortholog of PTPN22) have increased T-cell activation in combination with an increased production of IL6 [125]. The functional effects have yet to be studied in CD patients.

IL10
As many as ten historic studies (1999–2006) in multiple different populations have analyzed the IL10 promoter (at -592 [rs1800872], -819 [rs1800871] and -1082 [rs1800896]) for evidence of association with UC and CD, with very variable results. A German UC GWAS has recently provided definitive evidence of association at IL10, but not with any of these 3′ promoter SNPs [21]. Their lead IL10 SNP was rs3024505, located 1 kb downstream of the 3′ untranslated region. Following an IL10 tagging SNP (n = 22) supplementary study across the various replication panels, two additional intronic SNPs were associated with UC. Although the three associated SNPs were in perfect linkage disequilibrium, regression analysis suggested they might act independently of each other. Of note, rs3024505 was also associated with German CD, albeit weakly. The location of this SNP – 79bp from a highly conserved stretch of DNA at the 3′ end of IL10 – is of some interest, at it contains a putative AP1 binding site.

IL10−/− mice develop spontaneous colitis in specific pathogen-free conditions owing to defective counter-regulatory anticytokine responses. Indeed, these animals provide one of the oldest (first reported in 1993), most widely used and best characterized genetic animal models of colitis. Physiologically important defects in IL10 signaling in lamina propria mononuclear cells from UC patients were described in 1995 [126], and by 1998, very early clinical trials of subcutaneous recombinant IL10 produced some evidence of benefit in UC patients [127]. For some reason, the Phase II trials of this protein (rHuIL-10) were in CD and when they failed to demonstrate efficacy, this avenue was quietly dropped from further investigation [128–129]. However, given these new genetic data, further exploration of IL10 delivery should be explored as a therapeutic avenue in UC, perhaps utilizing Larry Steidler’s ingenious delivery mechanism via recombinant IL10-producing Lactococcus lactis [130].

Other considerations/implications from gene discovery

Genotype–phenotype associations
The common NOD2 mutations are associated with ileal disease and a fibro-stenosing phenotype [131,132]. The HLA region is associated with colonic IBD, and the DRB1*0103 allele with extensive disease, the need for surgery and extraintestinal manifestations. GWASs have, to date, largely defined association with IBD generally, with CD or UC specifically, or with early-onset IBD. Further than this, there has been a real paucity of consistent genotype–phenotype correlations. There are two possible explanations for this. Either the large GWAS cohorts have been inadequately phenotyped and/or they are underpowered for subanalysis. Certainly a large number of the CD genes/loci identified in the meta-analysis have yet to be subjected to formal genotype–phenotype association. Further collaborative studies designed to specifically address this issue are underway.
Ethnic variations in IBD gene associations

Aside from Yamazaki’s initial small-scale GWAS, all the large genetic experiments in recent years have been performed in Caucasian cohorts. As a result, ethnic variation has yet to be studied in much detail. However, the data presently available suggest this to be an avenue ripe for future study as clear differences in some genes have been identified not only between Asian and European populations but also within populations of European descent. NOD2 mutation carriage rates are seen to range widely (0–50%) and are highest in central Europe [60,133]. The contribution of NOD2 to disease susceptibility is relatively lower in northern European (Scottish [134,135] and Scandinavian [136,137]) populations, where the population-attributable risk ranges from 7.9 (early-onset Scottish CD) to 11.4% (Swedish CD). In Japanese [138], Chinese [139] and South Korean [140] populations, these NOD2 mutations (Gly908Arg, Arg702Trp and Leu1007fsincC) are absent.

TNF-SF15 is the only IBD gene discovery originating in an Asian population to date (Box 3) [26]. This was an early GWAS in a small-index cohort, suggesting that larger studies may yield additional disease genes that may be specific to those of Asian descent. The TNF-SF15 association was confirmed in the Caucasian CD meta-analysis [17] and was replicated in other Asian populations [141]. Several other confirmed IBD genes from European populations have been studied in Asian cohorts. Yamazaki and colleagues from Japan have demonstrated...
associations at NFKX2.3, IL12B and ZNF-365, but not at IL23R, ATG16L1, 3p21, 5q23 (IBD5), IRGM or PTPN2 [142,143]. The 5p13 gene desert was not polymorphic. The lack of association at IBD5 has been previously reported [144]. However, it is difficult to draw firm conclusions regarding the other loci reported by Yamazaki owing to lack of power (n = 484 CD). PTPN22 has not been directly studied in CD populations other than Caucasians; however, Asians rarely carry the 620W variant that is associated with autoimmune diseases but protective for CD [122].

Gene–environment interactions
Barrett estimates that the currently identified susceptibility genes account for only approximately 20% of the genetic contribution to CD susceptibility [17]. One argument as to why this may be an underestimate is that it takes no account of gene–environment interactions, which probably act synergistically to increase disease risk. The environmental stimuli that lend themselves to study include cigarette smoke and commensal/pathogenic bacteria.

The study of the commensal flora, enterohedrater and intra-cellular bacteria in CD patients with varying genotypes (e.g., NOD2, ATG16L1 and IRGM) will be of great interest. Not only may this provide specific clues as to disease etiology, but it may serve to direct personalized therapy (e.g., antibiotics) depending on genotype. However, such studies are difficult as they require large cohorts of homogenous patients with specific genotypes early in disease course and preferably off treatment. In the meantime, modeling of gene–environment interactions in genetically modified animals has proved instructive, as best illustrated by the study of Il10-/- mice and Hla-B27 transgenic rats [145]. These animals develop colitis in specific pathogen-free conditions, but not when kept completely germ-free. Selective colonization with different bacterial species critically affects aspects of disease phenotype, notably severity and anatomical location. In Il10-/- mice mono-associated with Enterococcus faecalis or E. coli, a progressive chronic colitis develops, although the regional distribution and kinetics of this colitis varies with the bacteria. Mice monoassociated with Pseudomonas fluorescens or Bacteroides vulgatus do not develop colitis. By stark contrast, Hla-B27 rats monoassociated with B. vulgatus develop an aggressive colitis, but show no inflammation with E. coli. Clearly, the interaction of both genetic factors (Il10-/- or Hla-B27) and environmental factors (E. coli or B. vulgatus) is fundamentally important for the establishment of colitis in these animal models. It is highly probable that a similar phenomenon occurs in human disease, but this awaits formal testing.

Expert commentary & five-year view
The major evolving paradigm for complex disease genetics is that the biological insights gained (and subsequent abundance of novel therapeutic targets) will turn out to have far greater importance than the relatively small contribution to disease risk conferred by each individual mutation or locus. We speculate that in 5 years’ time, the genetic architecture of IBD will be completely described, with around 200 disease genes/variants accounting for the genetic contribution to disease susceptibility. At this stage, an ‘IBD chip’ will become a realistic possibility that may enable physicians to better diagnose patients, subcategorize (e.g., IBD1, IBD2 and IBD3) and personalize treatment accordingly. However, these data will not be used in isolation; rather, they will be incorporated into existing and future clinical models of disease risk. We will discuss the short- and long-term future of genetic studies and the implications for therapeutics and modeling of disease natural history.

Future of genetic studies
There are several explanations as to why the known IBD disease genes/loci only explain a fraction of the observed familial aggregation [17,146]. First, the present GWAS provide only relatively limited surveys of potential sequence variation and provide little or no information on rare alleles. Second, many ‘hits’ from the index studies are surrogate markers for the true disease causative mutations that may confer greater risk. Third, the risk at some loci will be attributed to multiple, independent mutations. In addition, very few studies into copy number variants have been performed on a genome-wide scale; it remains to be seen whether these will add significantly to the genetic risk. For several or all of these reasons, we are probably currently underestimating the known genetic risk already accounted for. However, the shortfall that remains is (and almost certainly should be) presently limiting the early application of genetics to determining individual disease risk.

The immediate future involves more GWASs in different subphenotypes, meta-analyses of existing individual GWASs (UC and UC plus CD), fine-mapping of loci (a major ongoing phase of the WTCCC follow-up studies, with CD prioritised for deep resequencing of associated loci) and detailed molecular and functional studies to truly understand the biology of various disease-associated mutations (Figure 2).

Implications for therapeutics
There is an urgent, unmet therapeutic need in IBD, as many of the current therapies are limited by poor efficacy, unacceptable toxicity and poor patient acceptance. Several of the pathogenic insights arising from gene discovery described previously lend themselves to potential therapeutic intervention. For example, the association of multiple IL23-signaling components with IBD strongly suggests that this pathway should be prioritized for drug discovery, not least since proof of concept has already been provided in animal models. These data make the critical re-evaluation of early clinical studies using anti-p40 monoclonal antibody therapy in CD of paramount importance [147]. Anti-p19 monoclonal antibodies that are likely to provide a more specific anti-inflammatory intervention in IBD are currently in development.

Manipulating levels of intestinal autophagy in CD is an intriguing therapeutic option as several drugs targeting mTOR (and therefore increasing autophagy) are already in clinical use. Sirolimus is frequently used as an immunosuppressant post-solid organ transplantation, and sirolimus-eluting coronary artery stents are notable for their low restenosis rates. A recent single case report has described the clinical use of sirolimus in a patient
with severe CD, resulting in a marked and sustained improvement in symptoms and endoscopic appearances [148]. In animal models, rapamycin has been demonstrated to be as effective as cyclosporine in reducing experimental chronic colitis induced by DSS [149]. These observations led to the prospect of early clinical trials of these agents to induce autophagy in patients with CD. However, owing to the wide-ranging effects of autophagy in both homeostatic and pathogenic roles in a variety of organs and tissues, not to mention in the GI tract itself, it is likely that more specific targeting of defined autophagy pathways will be required and delivery mechanisms will need to be developed to limit unwanted or dangerous systemic side effects.

In addition, the inclusion of GWAS technology to present IBD therapeutics (e.g., thiopurines and anti-TNF agents) may elucidate genetic factors that predict both clinical response and adverse events. The latter concept has recently been successfully tested in flucloxacillin-induced liver injury, effectively providing proof-of-principle to this approach [150]. Similar success in IBD could rapidly bring pharmacogenetics further into the clinical arena, mirroring the limited success of TPMT genotyping (or measurement of enzymatic activity) presently adopted in many centers prior to commencing azathioprine therapy [151].

**Implications for natural history**

In CD, physicians have lately taken a lead from rheumatologists and have been exploring the use of an aggressive policy of early combined immunosuppression (thiopurine plus an anti-TNF agent) in an attempt to modify the natural history of IBD (i.e., limit progression from inflammatory changes to stricturing or penetrating phenotype) [152]. This policy has shown early promise, but adoption into clinical practice will remain limited until we are able to accurately determine which patients will run an aggressive disease course. At least 30% of patients with CD will have a fairly mild disease course; exposing these patients early to toxic therapies is not likely to prove an acceptable option [153].

Present measures to predict an aggressive disease course are relatively crude and include: early age of disease onset, cigarette smoking and extensive disease. It is widely anticipated that a combination of prospectively and accurately phenotyped patients, DNA, serum, stool and/or intestinal biopsies will allow a detailed accurate predictive model based on clinical data and biomarkers. This reality is at least 5 years away. However, it is in this context that a future ‘IBD chip’ may revolutionize disease risk modeling and allow IBD genetics to enter clinical practice after over 20 years of promise.

**Financial & competing interests disclosure**

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

**Key issues**

- Genome-wide linkage studies in the 1990s culminated in the identification of NOD2 as a Crohn’s disease (CD) susceptibility gene in 2001, a landmark gene discovery for complex disease genetics.
- Genome-wide association studies (GWAS) have recently become possible owing to:
  - The completion of the Human Genome Project;
  - The HapMap project that described the ‘blocks’ (haplotypes) of genetic material that were typically inherited together;
  - Technological advances in microarray manufacturing, processing and analysis.
- Since December 2006, there have been seven GWAS and one meta-analysis published in CD, three GWAS published in ulcerative colitis (UC) and one GWAS in early-onset inflammatory bowel disease (IBD).
- There are presently around 50 confirmed IBD susceptibility genes/loci.
- In CD, it has been estimated that this accounts for only around 20% of inherited contribution to disease susceptibility.
- The genetic architecture of IBD is being unraveled; some genes confer susceptibility to IBD generally while others are specific to CD or UC.
- Several key insights into IBD pathogenesis have arisen primarily from gene discovery:
  - Innate immunity/autophagy and defective bacterial handling in CD (NOD2, ATG16L1 and IRGM);
  - Th17/IL23 signaling in IBD (IL23R, IL12B, JAK2, STAT3 and CCR6);
  - Intracellular tyrosine phosphatases in IBD (PTPN2 and PTPN22);
  - Developmental genes in IBD (NKKX2.3 and GLI1);
  - Defective barrier function in UC (ECM1).
- These insights will lead to novel therapeutics for IBD (e.g., monoclonal antibodies targeting IL23 signaling).
- The short- to medium-term follow-up of gene discovery are:
  - Further GWAS (e.g., UC, early-onset IBD) and meta-analyses;
  - Fine-mapping of disease gene/loci;
  - Function of causal variants in vitro and in vivo;
  - Detailed correlation between genotype and phenotype (e.g., innate immune genes and intracellular bacteria);
  - Pharmacogenetics studies.
References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest


20. The North American group’s report the first comprehensive GWAS in CD, identifying IL23R as a susceptibility gene


22. The Schreiber laboratory report a definite association of IL10 in the first full GWAS in UC.


24. The French/Belgian GWAS in CD uncovers a very strong association in a large gene desert for the first time.


27. The North American group’s report the first comprehensive GWAS in UC reports an association within a region of the IBD2 locus containing the IL22, IL26 and IFN-γ genes.


Lakatos PL, Szamosi T, Szilvasi A et al. ATG16L1 and IL23 receptor (IL23R) genes are associated with disease susceptibility in Hungarian CD patients. Dig. Liver Dis. 40(11), 867–873 (2008).


Baba M, Imai T, Nishimura M et al. Identification of CCR6, the specific receptor for a novel lymphocyte-directed CC chemokine LARC, J. Biol. Chem. 272(23), 14893–14898 (1997).


• Published at the same time as reference 61 in *Nature*, this article reported NOD2 to be a CD susceptibility gene. This was a landmark discovery, not just for IBD genetics but for complex disease genetics as a whole.

• Published at the same time as reference 60 in *Nature*, this article reported NOD2 to be a CD susceptibility gene. This was a landmark discovery, not just for IBD genetics but for complex disease genetics as a whole.

73 Agostoni L, Martinon F, Burns K et al. NALP3 forms an IL-1β-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* 20(3), 319–325 (2004).

• In this candidate gene study, the NLRP3 gene was associated with CD at genome-wide levels of significance.

• This important paper describes the *in vivo* and *in vitro* phenotype of ATG16L1 deficiency.

• This important paper describes the *in vivo* and *in vitro* phenotype of ATG16L1 deficiency.

• The association with IRGM is attributed to a 20-kb deletion that alters the expression of IRGM in a cell-specific manner.

An important role for the hedgehog signaling pathway in IBD pathogenesis is demonstrated via GLI1 gene discovery, human expression studies and in vivo models.

Paracrine hedgehog signaling in stomach and intestine: new roles for hedgehog in gastrointestinal tract: implications for development, homeostasis, and disease.

Postnatal recapitulation of embryonic hedgehog pathway in response to skeletal muscle ischemia.

Accumulation of hedgehog-responsive progenitors parallels alcoholic liver disease severity in mice and humans.

The hedgehog pathway regulates remodeling responses to biliary obstruction in rats.

Hedgehog signaling modulates activation of and cytokine production by human peripheral CD4+ T cells.


Contribution of forms of Crohn's disease.

with familially occurring and complicated phenotypic characteristics of severe disease.

CARD15

childhood-onset Crohn's disease: et

Russell RK, Drummond HE, Nimmo EE

heterogeneity within Europe?

Arnott ID, Nimmo ER, Drummond HE

Am.

patients with inflammatory bowel disease.

Lesage S, Zouali H, Cezard JP

bowel disease.

The contribution of CARD15 mutations to two ethnically divergent populations.

CARD15 haplotype structure and association to Crohn's disease in the Chinese population.

NOD2/CARD15 mutational analysis and association scans.

Cardon LR and 5p13.1 loci with

Kugathasan S, Baldassano RN, Bradfield JP et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease.

McCarthy MI, Abecasis GR, Cardon LR et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges.


www.expert-reviews.com


175 Duerr RH, Chesney JJ. Associations between HLA-DR alleles and subsets of ulcerative colitis defined by extent of colitis. Gastroenterology (113) 2–7 (1997).


Websites
201 Registry of federally and privately supported clinical trials conducted in the USA and around the world. www.clinicaltrials.gov

202 Database of genes associated with human disease and response to pharmaceuticals. www.hapmap.org

Affiliations
• Charlie W Lees, MRCP, PhD
  Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK
  Tel.: +44 131 651 1027
  Fax: +44 131 651 1085
  charlie.lees@ed.ac.uk

• Jack Satsangi, FRCP, DPhil
  Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK
  Tel.: +44 131 651 1807
  Fax: +44 131 651 1085
  j.satsangi@ed.ac.uk