Genes in inflammatory bowel disease: lessons from complex diseases

Paul Henderson and Jack Satsangi

The major forms of chronic inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC) are characterised by inflammation of the intestinal tract. These complex, lifelong conditions can present at any age with peak incidences in the second and third decades. The natural history of IBD varies worldwide,1 with gradients in incidence, phenotype and recurrence having been demonstrated in Europe and North America.2,3 The exact aetiological factors involved in IBD pathogenesis remain elusive, however it is becoming clear that genetic predisposition, environmental influences and a dysregulated immune response to the intestinal microflora are involved.4–6

Since the first description of Crohn's disease in the early 20th century there have been several lines of epidemiological evidence implicating genetic susceptibility in the pathogenesis of IBD.7,8 The most notable of these include the familial clustering of IBD cases, ethnic differences in disease susceptibility and high concordance rates in twins.9–11 The observation of increased twin concordance was demonstrated in the late 1980s from large studies carried out in Scandinavia and the UK.12 These studies showed that the risk of CD in monozygotic twins is 36% and for UC 16%, in contrast to 4% for both diseases in dizygotic twins.13

Following this strong evidence for heritable factors non-parametric linkage analysis was employed to further delineate possible genes causing IBD susceptibility in the 1990s (Fig 1).14 Linkage analysis involves the study of a limited set of genetic markers across the entire genome and does not explicitly specify the mode of inheritance, looking instead for excessive allele sharing in affected relatives. This method of gene identification has come under scrutiny in recent years, mainly due to newer techniques failing to replicate findings in these early analyses. However, it should be noted that the gene identified in these studies (NOD2/CARD15) remains the strongest individual signal using genome-wide association scanning (GWAS).4

Without a doubt the major breakthrough in not only IBD genetics but complex polygenic disease as a whole was the discovery of NOD2/CARD15 in 1996. This gene was discovered by the French paediatric gastroenterologist Hugot following fine mapping of the IBD1 locus.15 Hugot’s group later identified variants in the leucine-rich repeat region (LRR) of the gene as conferring susceptibility to IBD, something which has been widely replicated in both adult and early-onset disease.16,17 The gene encoded by NOD2/CARD15 has gut expression limited to anti-microbial Paneth cells, with the LRR domain recognising muramyl dipeptide (MDP), a product of bacterial cell wall degradation.18,19 Since these initial studies, however, further work has demonstrated a wide range of functions for this pattern recognition receptor including the regulation of commensal bacteria, the control of certain immune cell populations and viral recognition.20–22

In more recent years the introduction of the Human Genome Project and other international collaborations has allowed the detailed examination of the entire genetic code.23 This information, coupled with the creation of high-throughput genotyping platforms, has led to the ability to analyse up to 500,000 single nucleotide polymorphisms (SNPs) using hypothesis-free GWAS. To date there have been a total of 14 GWAS performed in IBD patients leading to an explosion of multiple candidate genes implicated in IBD pathogenesis.17,24 Before the first major CD meta-analysis published in 2008 by Barrett and colleagues,
Implicated pathogenic pathways in IBD

**Innate immune signalling**
- NOD2
- TLR4

**Autophagy**
- ATG16L1
- IRGM
- LRRK2
- MTMR3

**Endoplasmic reticulum biology**
- XBP1
- CAPN10
- ORMDL3
- ERAP2

**Pathogenic Th17 – T\_\text{reg} balance**
- IL23R
- STAT3
- JAK2
- ICOSLG

**Secondary immune response**
- CCR6
- IL27
- TYK2
- HLA-region
- IL-12B
- MST1
- IL18R1-IL18RAP
- CCL-gene cluster
- XBP1
- TNFRSF6B
- TL1A
- SMAD3
- TAGAP
- FUT2

**Fig 2. Broad overview of the genetic architecture of inflammatory bowel disease giving examples of confirmed and candidate genes implicated in each pathway.**

T\_\text{reg} = regulatory T cell.

only 16 loci had been identified, with the meta-analysis confirming many of these and providing evidence for 21 additional SNPs.24,25 This has been expanded further still through recent work by the International IBD Genetics Consortium with the confirmed loci now numbering 69 in CD alone.26 Although these identified loci will undoubtedly generate decades of future research, it is postulated that they still only represent approximately 25% of the genetic variance, with only the option of deep sequencing and functional studies likely to increase understanding of these susceptibility loci. What has become very clear, however, is that the genetic architecture exposed by these scans has led to a clearer understanding of the disease pathways and mechanisms involved in IBD.

This insight provided by GWAS has uncovered candidate genes involved in pathways broadly involving the control of intestinal barrier function, the mucosal response to luminal bacteria and the secondary downstream immune response (Fig 2). Besides NOD2/CARD15, the two pathways generating the most interest at present are autophagy (the degradation of unwanted cell organelles and the elimination of pathogenic microorganisms) and the IL12–IL23 pathway (involved in pathogenic T cell differentiation).27 SNPs in the regions of the autophagy genes ATG16L1 and IRGM have been widely replicated as conferring susceptibility to CD, with functional work beginning to unravel the functional implications of their encoded proteins.28 The association with CD of a germline variation of the IL23R (interleukin-23 receptor) gene in 2006 provided the initial stimulus to look more closely at the IL17-producing subset of CD4+ Th17 cells.29 Since then risk and protective haplotypes of IL23R have been described with other genes in the Th17 pathway, such as STAT3, JAK2 and ICOSLG also implicated. In addition, the exciting recent discovery that bacterial handling and MHC class-II presentation are regulated by NOD2 autophagy induction through ATG16L1 has ignited the drive to establish a cohesive model of disease pathogenesis,28,29

Not only has the study of genetics led to advances in the understanding of IBD aetiopathogenesis it has also enabled researchers to focus on key areas of potential therapeutic targets. Although gene therapy is unlikely to play a part in IBD management in the near future, several lines of pharmaceutical intervention have already been highlighted from gene discovery. The targeting of the IL12-IL23 pathway has already shown promise with the use of anti-p40 monoclonal antibody for induction of remission in CD.30 Additionally, the use of Sirolimus (which alters cellular autophagy) has been shown to be of use in refractory IBD and in the treatment of experimentally-induced murine colitis.31,32 Although specific drug treatments are still under development, genotype analysis has already been shown to inform clinicians regarding the natural history of IBD and in predicting treatment response to newer biological drugs.33,34 It is hoped that in the coming years the combination of genotyping, biomarkers and detailed phenotyping will allow a tailored approach to clinical management using a validated predictive model.

It can be seen that IBD genetics has come a long way in a short space of time, from twin studies to a clinical tool in less than 30 years. Although there is still a great deal to be learned, gaining unprecedented insight into complex polygenic disease pathophysiology will no doubt lead to rapid discoveries not only in IBD but other similar diseases. Within IBD research there is now a pressing need to explore the gene-environmental interactions especially with relation to CD-associated adherent-invasive Escherichia coli, mycobacterium and cigarette smoke.35,36 These interactions, in addition to the deep sequencing of the candidate genes to unlock the disease-causing mutations, will hopefully lead us even further towards a panacea for this debilitating disease.

**Sources of support**
P.H is funded by a Medical Research Council project grant (No. G0800675).

References

© Royal College of Physicians, 2011. All rights reserved.
Paul Henderson and Jack Satsangi


7 Dalziel TK. Chronic interstitial enteritis. *BMJ* 1913;3(ii):1068–70.


19 International HapMap Project. Database of genes associated with human disease and response to pharmaceuticals [HapMap web site]. www.hapmap.org


Address for correspondence: Dr P Henderson, Department of Child Life and Health, University of Edinburgh, 20 Sylvan Place, Edinburgh EH9 1UW.

Email: paul.henderson2@nhs.net