PART I:
Genes and Phenotype in IBD
Which will take us further in IBD—study of coding variation or epigenetics?

Miles Parkes
Department of Gastroenterology, Addenbrooke's Hospital and University of Cambridge, Cambridge, UK

LEARNING POINTS

- Genome-wide association scans have revealed many genetic risk factors for Crohn’s disease and ulcerative colitis.
- As with environmental risk factors, some of the genetic risk is shared and some is specific to either Crohn’s disease or ulcerative colitis.
- Only about 20% of the variance in heritability has been accounted for by known genetic loci.
- The study of genetic variants is valuable because it reveals insights into disease pathogenesis.
- Increasing evidence suggests that much of the host susceptibility to IBD may be epigenetic, lying at the level of the regulation of gene expression.
- Epigenetic risk is heritable through mitosis and possibly meiosis, and many of the known environmental or lifestyle risk factors may operate at an epigenetic level by influencing gene transcription.

Gene expression: sequence variation versus epigenetic factors

The human genome is thought to encode some 23,000 protein-coding genes, comprising just 1.5% of the total of 3 billion base pairs. Sequence variation can take many forms from single nucleotide polymorphisms (SNPs) to indels (insertion–deletion polymorphisms) to copy number variants, where segments up to thousands of base pairs long can be deleted or duplicated. SNPs are the commonest variant. They occur approximately every 200 base pairs, but less frequently in coding sequence because of potential for adversely affecting protein function and hence incurring negative selection pressure.

Genes comprise exons (the coding sequence) and introns, which are removed prior to mRNA being translated to protein. Gene density varies considerably, with lengthy tracts of noncoding sequence, formerly and erroneously
Genes and Phenotype in IBD

Genetic variation in IBD

What forms of genetic variation contribute to IBD? The answer is likely to be “all of them,” to a greater or lesser extent, perhaps including mechanisms yet to be characterized. Extrapolation from monogenic disease initially suggested that coding variation was likely to be most relevant, and its obvious impact on protein structure and function supported this intuition. Further, the three relatively common causal variants in NOD2, the first IBD gene to be identified, were all coding variants [3]. Thus, early genome-wide genotyping arrays, which could accommodate relatively few SNPs, focussed only on “nonsynonymous” SNPs. Although some interesting results were obtained, particularly in identifying the importance of ATG16L1 and autophagy in CD, the yield was unimpressive [4].

Truly hypothesis-free GWAS studies have followed, interrogating most if not all common variations (allele frequency >5%) genome-wide. Interestingly, the yield from these “proper” GWAS studies has been much greater than from nonsynonymous SNP scans and many lessons have been learned.

One remarkably consistent feature of GWAS studies has been the number of “gene deserts” showing association across a range of complex diseases. The supposition is that these loci contain elements that regulate transcription, and there is now evidence that sequence variation influences transcription for many genes. Thus, epigenetic regulation is itself a heritable trait and may be the key factor contributing to phenotypic variation in humans [5].

Several “gene desert” associations have been seen in IBD: indeed in the first meta-analysis plus replication of CD GWAS studies from the international IBD genetics consortium, 6 out of the 32 confirmed loci mapped to gene deserts. More than this, our now detailed knowledge of all common sequence variations genome-wide allowed us to identify how many of the CD susceptibility loci correlated with any known coding variation. The answer, rather startlingly, was just 9 [1]. To emphasize this point, coding variation has to date been confirmed as causal for just two loci—NOD2 and ATG16L1, with one other at IL23R strongly implicated.

Regulation of gene expression in IBD

Accepting the indirect evidence that regulatory effects are important, is there any direct evidence? The answer is emphatically yes. In the Belgian CD GWAS, the strongest association was seen with a 1.25-Mb gene desert on chromosome 5. Using publicly available expression quantitative trait loci (eQTL) data, Libioulle et al. showed that these same SNPs that showed association with CD also correlated strongly with expression of the prostaglandin receptor gene EP4 270 Kb away [6]. The international CD meta-analysis study identified a number of other such correlations [1], and in its most recent analysis identified association at a DNA methyltransferase gene, emphasizing the importance of epigenetic regulation and its interrelationship with sequence variation in CD susceptibility.

Evidence from basic research corroborates the importance and potential complexity of epigenetic effects. Thus, the toll-like receptor-induced inflammatory response in mouse macrophages is regulated at a gene-specific level by transient chromatin modification, with Th2 “bias” being conferred by a transcriptional regulator of IL-4 called Mina. Highlighting the interplay of sequence variation with epigenetics, production of Mina is itself strongly correlated with SNP haplotypes in its promoter [7].

Identifying correlation between IBD association signals and gene expression hints at functional regulatory elements, but usually does not explain the mechanism. The expectation is that genome-wide assays for DNA methylation, ChIP seq, histone binding, and DNA tertiary structure (e.g., chromatin conformation capture or 3C), will provide some answers over the next few years [8]. They should allow both a better understanding of the mechanisms underlying current GWAS signals and also permit de novo genome-wide studies.
Limitations of current studies of epigenetic mechanisms in IBD

At present, difficulties in defining which cell type to target for expression analyses are limiting. The relevance of this comes from the recognition that many gene regulatory effects are cell-type specific—as seen for the CD-associated allele of IRGM which affects expression in opposite directions in different cell types [9]. Further concerns relate to the confounding effects of inflammation and drug therapy. Nonetheless, the evidence that epigenetic mechanisms are crucial in regulating gene transcription and thereby affecting susceptibility to disease will drive development of the appropriate resources to tackle these questions.

Epigenetic regulation is also significantly influenced by environmental factors, including diet, smoking, and infection—all of which are implicated in IBD pathogenesis. For example, aryl hydrocarbon receptor (AhR) agonists, which are present in substances as varied as cigarette smoke and Brassica vegetables, can strongly influence COX-2 expression. The effect may be related to AhR acting directly as a transcriptional regulator and also by regulating histone acetylation and hence chromatin structure [10]. The AhR also plays a key role in modulating Th17 lymphocyte development through epigenetic mechanisms [11]. The suggestion that some epigenetic regulatory influences may be transmissible through meiosis to the next generation adds particular interest to this story [12].

Conclusions

At present, GWAS studies are being widely deployed not because they provide all the answers, but rather because they are technologically tractable and provide robust and reproducible data. More technologically challenging and complex studies will follow to advance our knowledge of the epigenetic regulation of gene transcription and its contribution to inflammatory disease [13].

The suspicion is that many of the pathways highlighted by GWAS studies will also be flagged as important for IBD pathogenesis by other techniques. A case in point might come from the confirmed association of noncoding SNPs adjacent to IL-10 with IBD [14], plus the recent observation that IL-10 gene expression in antigen-presenting cells is strongly regulated by the histone deacetylase HDAC11 [15]. Perhaps these findings are directly correlated, or maybe the GWAS signal is flagging a pathway influenced by many epigenetic and other mechanisms that themselves regulate IL-10 transcription and thereby influence IBD susceptibility.

All of these studies represent work in progress, and despite recent exciting developments this remains a field in its infancy. Nonetheless, current evidence suggests that for most individuals with IBD, coding variation is likely to have made only a modest contribution to their disease risk, while epigenetic regulation of gene transcription, perhaps influenced by environmental factors such as smoking, bacteria, and diet, play a much more important role.

References

Genes and Phenotype in IBD


