Journal of Autoimmunity xxx (2015) 1–10

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Review article

Understanding inflammatory bowel disease via immunogenetics

Katrina M. de Lange [a], Jeffrey C. Barrett [a, b, c, *]

[a] Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom
[b] European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom
[c] Centre for Therapeutic Target Validation, Wellcome Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom

ARTICLE INFO

Article history:
Received 23 July 2015
Accepted 23 July 2015
Available online xxx

Keywords:
Inflammatory bowel disease
Genetics
Genome wide association study
Therapeutic targets
Microbial sensing

ABSTRACT

The major inflammatory bowel diseases, Crohn's disease and ulcerative colitis, are both debilitating disorders of the gastrointestinal tract, characterized by a dysregulated immune response to unknown environmental triggers. Both disorders have an important and overlapping genetic component, and much progress has been made in the last 20 years at elucidating some of the specific factors contributing to disease pathogenesis. Here we review our growing understanding of the immunogenetics of inflammatory bowel disease, from the twin studies that first implicated a role for the genome in disease susceptibility to the latest genome-wide association studies that have identified hundreds of associated loci. We consider the insight this offers into the biological mechanisms of the inflammatory bowel diseases, such as autophagy, barrier defence and T-cell differentiation signalling. We reflect on these findings in the context of other immune-related disorders, both common and rare. These observations include links both obvious, such as to pediatric colitis, and more surprising, such as to leprosy. As a changing picture of the underlying genetic architecture emerges, we turn to future directions for the study of complex human diseases such as these, including the use of next generation sequencing technologies for the identification of rarer risk alleles, and potential approaches for narrowing down associated loci to casual variants. We consider the implications of this work for translation into clinical practice, for example via early therapeutic hypotheses arising from our improved understanding of the biology of inflammatory bowel disease. Finally, we present potential opportunities to better understand environmental risk factors, such as the human microbiota in the context of immunogenetics.

1. Introduction

The Inflammatory Bowel Diseases (IBD) are a group of inflammatory disorders affecting the gastrointestinal tract. The major types are Crohn's disease (CD) and ulcerative colitis (UC). Neither CD nor UC are fatal, but both are debilitating conditions: affected patients experience a range of symptoms associated with inflammation of the gut, including abdominal pain, fever, vomiting, diarrhoea, rectal bleeding, anaemia and weight loss. There is currently no known cure, but symptoms can be managed through anti-inflammatory steroids or immunosuppressants to reduce inflammation, dietary changes to try and remove environmental triggers, and (in severe cases) surgery to remove damaged portions of the bowel. Although CD and UC share a number of clinical features, there are important distinctions in incidence patterns, disease localization, histopathology and endoscopic features (Table 1) that suggest there are differences in the underlying pathways driving each disease [1,2].

IBD prevalence is currently highest in Europe (UC, 505 per 100,000 persons; CD, 322 per 100,000 persons) and North America (UC, 249 per 100,000 persons; CD, 319 per 100,000 persons) [3]. It is more common in Ashkenazi Jews, who are five to eight times more likely to develop IBD compared to non-Jewish populations [4]. More broadly, global prevalence is rising, with rapid increases in incidence rates occurring as more countries adopt a 'Westernised' lifestyle [5]. Incidence rates are also rising in younger people, placing an increased strain on healthcare resources (particularly as early-onset IBD has been associated with a higher risk of developing colorectal cancer) [6]. Overall, IBD represents a significant global health burden that is of growing concern.

* Corresponding author. Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom.
E-mail address: jb26@sanger.ac.uk (J.C. Barrett).

http://dx.doi.org/10.1016/j.jaut.2015.07.013

Table 1
Clinical and epidemiological features of the two major inflammatory bowel disease subtypes, Crohn's disease and ulcerative colitis [1,2].

	Crohn's disease	Ulcerative colitis
Incidence patterns		
Sex	More common in women than men	Equal rates in men and women
Prevalence rates	CD is more prevalent than UC in developed countries	UC emerged before CD in developed countries, and is more prevalent in still-developing countries
Disease localisation		
Affected areas	Entire gastrointestinal tract (from mouth to anus)	Colon, plus some potential backwash ileitis
Inflammation pattern	May occur as patchy, discontinuous inflammation	Continuous inflammation in the affected area (though sometimes a separate cecal patch)
Histopathology		
Penetrance	Transmural inflammation of the entire bowel wall	Inflammation restricted to the mucosal and submucosal layers (other than in fulminant colitis)
Appearance	Thickened colon wall with granulomas, deep fissures and a cobblestone appearance	Distorted crypt architecture, with shallow erosions and ulcers; granulomas, if present, only around crypts
Serological markers		
	Anti-Saccaromyces cerevisiae antibodies	Anti-neutrophil cytoplasmic antibodies
Complications		
	Fistulas, abdominal mass (typically lower right quadrant), colonic and small-bowel obstructions, stomatitis	Haematochezia, passage of mucus or pus, fulminant colitis and toxic megacolon

2. The early days of IBD genetics

2.1. Twin studies

IBD is characterised by a dysregulated immune response to unknown environmental triggers in a *genetically susceptible* individual, and a heritable component to the disease is well recognised. Early epidemiological observations showed clear familial clustering, which was reflected in high sibling risk ratios (UC, 7–17; CD, 15–42) [7]. Twin studies have since conclusively shown these observations to be due to genetics, rather than shared environmental factors, by comparing disease concordance rates between pairs of monozygotic (MZ) and dizygotic (DZ) twins, with the assumption that both types of twin pair have roughly the same shared environment, and thus variation in concordance can be attributed to genetics. In a large meta-analysis of 6 twin studies the resulting rates of 30.3% vs 3.6% for Crohn's disease (112MZ vs 196DZ), and 15.4% vs 3.9% for ulcerative colitis (143MZ vs 206DZ), support the importance of genetics in IBD risk [8]. Motivated by these findings, there have been a number of studies aimed at identifying the specific genomic loci that explain IBD heritability. Ideally, each of these associated loci would identify a single gene, or indeed a causative genetic variant, to help understand the biological processes of IBD.

2.2. Linkage studies

Initially, IBD genetics studies were by necessity coarse-grained, as the high cost of sequencing and low throughput of genotyping restricted data collection to just a handful of genetic variants within a small number of individuals. The earliest studies therefore measured the patterns of inheritance *within families*: around 300 markers evenly distributed across the genome were sufficient to capture the pattern of DNA inheritance within a family [9]. By tracing the DNA segments that segregate with disease status (such as variant alleles only seen in affected individuals, and not in their unaffected relatives), sections of the genome that confer risk to the disease can be identified. This linkage analysis approach is good for detecting highly penetrant variants (i.e. those that are extremely likely to cause disease whenever present) that segregate well with disease status.

Linkage studies successfully identified many such highly penetrant variants for rare disorders [10–16], and were subsequently applied to a range of more common diseases. In 1996, the first such study in IBD linked a portion of chromosome 16 (dubbed 'IBD1') with Crohn's disease [17], which was successfully replicated in subsequent studies [18–23]. This finding was followed up using more closely packed markers within a small number of genes, and the IBD1 region on chromosome 16 was found to be caused by three disease risk alleles in the gene *NOD2*, whose role in the recognition of bacterial peptidoglycans and subsequent stimulation of an immune response supports its likely association with the development of CD (Fig. 1A) [24–26]. These variants were especially common in Ashkenazi Jews, partially explaining the increased burden of CD in that group. Unfortunately, however, successes like *NOD2* were rare: it remained one of the few robustly replicated genetic risk loci discovered via linkage, not just in IBD, but across common diseases.

2.3. Limitations of linkage studies and the common disease, common variant hypothesis

The widespread disappointment from linkage results among common diseases reflected a fundamental property of their genetic architecture: these diseases did not have a single, highly penetrant genetic cause. Instead, it was proposed by Risch and Merikangas (1996) that complex diseases were driven by the accumulation of many risk factors of only modest effect (the 'common disease, common variant' hypothesis). Finding associations via linkage under this scenario is difficult, as the genetic risk may be spread throughout the genome rather than concentrated in a single locus [27]. An alternative association analysis approach (which tests if the population-level allele frequencies of cases and controls are statistically different) is much more powerful, provided it is possible to choose the right variant to test among the millions known to exist in the human population. Risch and Merikangas (1996) calculated that 17,997 affected sibling pairs would be needed to detect a risk allele with 50% frequency and an odds ratio of 1.5 using linkage, as opposed to just 484 using an association analysis.

3. The GWAS era

3.1. Technological developments that made GWAS possible

Case-control association studies could detect signals too weak to show linkage, but suffered the drawback of needing to know which variants to test. One solution to this problem was to select candidate genes, based on prior biological hypotheses, but this produced

Please cite this article in press as: K.M. de Lange, J.C. Barrett, Understanding inflammatory bowel disease via immunogenetics, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.07.013