The genetics of chronic inflammatory diseases

Graham A. Heap and David A. van Heel

Centre for Gastroenterology, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Whitechapel, London E1 2AT, UK

Received December 22, 2008; Revised and Accepted December 31, 2008

Chronic inflammatory diseases have been at the forefront of the new genome-wide association study era. Conditions such as coeliac disease, type 1 diabetes, Crohn’s disease and ulcerative colitis have all benefited with multiple loci identified and replicated for each condition. As cohort sample numbers increase and researchers collaborate and share cohorts, common susceptibility loci are beginning to emerge between several diseases. Crohn’s disease and coeliac disease both demonstrate considerable overlap in their common genetic susceptibility with other related conditions. These shared loci offer an insight into the biology of the conditions but still present researchers with the problem of attempting to identify the true causal variants.

INTRODUCTION

Chronic inflammatory diseases are defined by long-term inflammatory processes directed at a particular endogenous or exogenous antigen. Within this definition, there is considerable overlap with autoimmune conditions such as multiple sclerosis and type 1 diabetes (T1D), but it can also be extended to cover a range of other conditions that would not normally be termed ‘autoimmune’, such as inflammatory bowel disease, idiopathic pulmonary fibrosis and atherosclerosis. The purpose of this review is not to provide an in-depth analysis of the genetics and pathogenesis of each condition. Instead, it will focus on a small subset of diseases, with particular emphasis on recent advances, what we have learned and how the genetic data have begun to illustrate the defining biology of the conditions.

Chronic inflammatory diseases have been at the forefront of a new genetics revolution primarily because they represent a considerable portion of human morbidity and many are known to be heritable. In the past 3 years, more than 50 genome-wide association studies (GWAS) have been conducted in around 20 different chronic inflammatory conditions ranging from asthma to osteoarthritis (Fig. 1) (1,2). The chronic inflammatory diseases that have been examined by GWAS are complex diseases, conditions where there is a considerable environmental impact on a polygenic susceptibility background. Linkage studies were largely unsuccessful at identifying major components of the heritability of these traits, with many conditions entering the GWAS era with no well-replicated findings. Many of these studies were conducted by collaborations based within a single country, each identifying a few novel, but low risk, alleles. It was soon realized, however, that even greater patient sample numbers would be needed to identify alleles with very small odds ratios. 2008 was the year that saw the successful pooling of large collections of GWAS data between groups to increase statistical power and identify further alleles in conditions such as T1D, type 2 diabetes, Crohn’s disease and colorectal cancer (3–6).

Two common and debilitating gastrointestinal disorders can be used to illustrate the journey that many complex diseases have been following: Crohn’s disease, an inflammatory response to luminal gut bacterial antigens, and coeliac disease, an autoimmune inflammatory response in the small bowel to ingested wheat gluten. At the turn of the millennium, Crohn’s disease genetics had identified a single linkage region on chromosome 16, but no definitive loci had been mapped (7). Today, the loci count stands at 31, with several of these harbouring multiple independent protective and risk variants (5,7). Coeliac disease, on the other hand, identified a major determinant of the heritability in the 1970s, with the identification of an association to the HLA complex, but prior to the GWAS era, other loci have been elusive (8,9).

CROHN’S DISEASE

Crohn’s disease has always been at the forefront of the GWAS era. It was the subject of one of the very first genome-wide SNP association scans in 2005. In what would now be described as a small study, 484 Japanese Crohn’s disease patients were genotyped for 80 000 SNPs, resulting in an
association to a haplotype mapping within the \textit{TNFSF15} gene (10). Since then, concerted efforts have been made to not only replicate this finding, but also to expand cohorts to much larger sizes and genotype many more SNPs in the hope of identifying additional associations. One of the first large-scale genome-wide association scans, published at the end of 2006, was also conducted in Crohn’s disease and identified associations within the \textit{IL23R} gene in two separate case-control cohorts (11). Although the sample size was not considerably different from the first study, the major advance came in the use of high-density SNP arrays, allowing approximately 300,000 haplotype-tagging SNPs, primarily derived from the International HapMap Project, to be genotyped at once. These high-density SNP arrays made a huge impact on complex disease genetics. By the start of 2007, Crohn’s disease geneticists had three genes, \textit{CARD15}, \textit{TNFSF15} and \textit{IL23R}, which could be confidently associated with the condition as a result of association studies and linkage results. This number was raised over the course of 2007 by a series of GWAS scans conducted in North American and European populations (12–16).

A recent meta-analysis, conducted on data from three complete genome-wide scans, replicated association at 11 loci from the individual scans and identified a further 21 loci (5). Despite this, however, estimates suggest that these 31 loci only explain about 20% of the heritability of Crohn’s disease. This figure comes with several caveats. It is often postulated that identifying the missing heritability will be aided by finding the real causal variants, and there now appears to be some evidence for this. For example, a pre-GWAS meta-analysis, investigating studies that directly typed the functional variants of \textit{CARD15} in Crohn’s disease, identified an allelic odds ratio of 4.09 for the 3020insC insertion (17). The original WTCCC GWAS scan reported an allelic odds ratio of 1.29, for association with \textit{CARD15}, in line with the odds ratios for many of the reported associations (14). The recent meta-analysis which was able to impute genotypes from data across two different platforms, however, reported the odds ratio for carriage of a \textit{CARD15} susceptibility allele as 3.99, suggesting that this variant can be well tagged through imputation (5).

\textbf{COMMON GENETICS OF THE INFLAMMATORY BOWEL DISEASES}

Ulcerative colitis and Crohn’s disease have distinct endoscopic, histological and pathological features as well as distinct differential risk factors. For example, smoking is a risk factor for Crohn’s disease, yet is protective for ulcerative colitis (18). However, despite these marked differences, there appears to be some overlap in genetic susceptibility. To date, two genome-wide association scans have been undertaken in ulcerative colitis: a non-synonymous SNP scan and genome-wide SNP scan (19,20). Both identified unique and overlapping disease associations. More interestingly perhaps is the overlap between these findings and associations with Crohn’s disease. By comparing the results obtained from the
two ulcerative colitis genome-wide association scans, together with a replication of Crohn’s disease variants in ulcerative colitis, with the results of the recent Crohn’s disease meta-analysis, it is possible to see that the underlying shared susceptibility is attributable to at least nine distinct regions (1p31, 1q32, 3p21, 5q33, 10p11, 10q21, 10q24, 17q21.2 and 18q11) (19–21).

Crohn’s disease, psoriasis and ankylosing spondylitis share many similarities and can often co-occur (22). It is not surprising, therefore, that all three display a common genetic susceptibility at the IL23R locus (11,23,24). Although useful for prediction of disease risk, the IL23R association also suggests a role for IL23 in the pathogenesis of the three conditions. Both Crohn’s disease lamina propria mononuclear cells and psoriatic lesions display elevated levels of IL-23, a cytokine believed to be involved in the maintenance of a pro-inflammatory subset of T cells, termed Th17 cells (25,26). Sarcoidosis and Crohn’s are both granulomatous inflammatory diseases that share common clinical features, but rarely co-occur. A joint GWAS between the two conditions was able to identify a common susceptibility variant that, after fine mapping, was located to the C10orf67 gene (27). This study elegantly demonstrates that conditions that share clinical features as well as conditions that co-occur can be associated with similar common susceptibility variants. Each condition, however, can also be expected to have unique variants that do not predispose to the related conditions. It is these variants, along with environmental insults, that may predispose an individual to a particular disease.

COELIAC DISEASE

Coeliac disease is a common disorder with an estimated population prevalence of 1% (28). There is a 75% concordance rate within monozygotic twins, making coeliac disease one of the most heritable complex diseases (29). Almost 90% of coeliac disease patients carry the HLA-DQ2.5 haplotype, with many of the remaining patients carrying a related haplotype, HLA-DQ8 (30,31). The HLA heterodimers, created by the associated haplotypes, present gluten peptides to cognate T cells primarily once the peptide has been de-amidated by the enzyme tissue transglutaminase (32). Related, but unassociated, HLA-DQ haplotypes have much lower binding affinities for gluten peptides (32). Estimates place the heritability of coeliac disease attributable to the HLA complex, at ~35% (33). As a result, particular emphasis has been placed on identifying the non-HLA genetic risk variants. In common with many complex diseases, genome-wide linkage studies were fairly unsuccessful at identifying loci that consistently replicated among populations in coeliac disease. Recent data from GWAS have identified eight regions beyond the HLA, which predispose to coeliac disease in three European populations (33,34). The largest allelic odds ratio identified to date, outside of the HLA complex, is 1.34 (3p21) (34). Associations such as these only explain a small proportion of the heritability at present (~3–4%). There is, however, the caveat that only tagging SNPs have been used and the causal variants likely not identified. The variants identified do, however, identify important biological pathways that could be hypothesized to be important in the development of disease. For example, seven of the eight coeliac disease regions identified contain immune genes, involved in T or B cell function. These genes have been identified in a hypothesis-free manner, yet they function in two of the most important cells in the pathogenesis of the condition.

COELIAC DISEASE AND TYPE 1 DIABETES

Many autoimmune diseases share common pathways, and recent GWAS have begun to elucidate the common genetic loci involved (35). In some cases, these common pathways can be inferred from commonly occurring loci in the literature; in the case of coeliac disease, however, this has also been implicitly tested. There has long been an epidemiological link between T1D and coeliac disease. Approximately 4.5% of children and 6% of adults diagnosed with T1D have concurrent coeliac disease (36). In addition, T1D patients often express peripheral blood antibodies to tissue transglutaminase (a diagnostic marker in coeliac disease) (37). T1D results from the autoimmune destruction of the insulin-producing β-islet cells, and clusters within families. The HLA variant that predisposes to coeliac disease, HLA-DQ2, is encoded by the two alleles DQA1*0501 and DQB1*0201 (38). T1D also has very a strong association to the HLA region, with DQB1*0302, DQB1*0201, DRB1*0401 and HLA-B all being implicated (39). One of the strongest T1D associations arises from the two coeliac-associated alleles, HLA-DQ2 and HLA-DQ8, but especially when encoded as a heterozygous genotype (40). Interestingly, this genotype presents only a moderately increased risk for coeliac disease (41). The mechanism behind the association with coeliac disease is well documented, with the heterodimer presenting de-amidated gluten peptides (42). The association with diabetes is less well understood, although it is reasonable to suppose that the unique peptide-binding characteristics result in a diabetogenic peptide being presented to the immune system. To test the hypothesis that common variants may predispose to both coeliac disease and T1D, Smyth et al. (43) genotyped all the currently identified T1D and coeliac disease loci in cohorts of both T1D and coeliac disease patients in an attempt to identify any common susceptibility locus. The results demonstrate the considerable overlap between the two conditions and offer some further explanation for the epidemiological overlap. Out of the eight coeliac disease susceptibility loci, six showed evidence of association with T1D (although the susceptibility allele differed for IL18RAP and TAGAP). Of the 17 T1D loci, eight showed evidence of association in coeliac disease, including CTLA4. The CTLA4 locus received considerable attention as a linkage peak and candidate gene for coeliac disease but replication was inconsistent and as a result, it was never accepted as a true susceptibility locus (44,45). The result is that seven loci are shared between T1D and coeliac disease, which given the relatively small number of total variants identified to date, in each condition, represents a remarkable shared susceptibility. The variants detected above are not just shared between T1D and coeliac; there is also considerable overlap with Graves'
disease, multiple sclerosis and rheumatoid arthritis (Fig. 2).
The greater number of T1D variants identified to date probably reflects the larger cohorts collected rather than a wider genetic susceptibility.

From the data collected to date, the hypothesis has been suggested that there may be a general set of autoimmunity susceptibility genes which are modulated by disease-specific genes (such as \textit{IL12A} in coeliac disease and \textit{INS} in T1D), as well as the host’s HLA status (43). The exact combinations of variants, combined with environmental insults, could determine which disease a person develops.

WHERE NEXT FOR GENETICS

Despite the large numbers of genome-wide studies established to date, most diseases have only managed to explain a tiny additional percentage of the heritability estimates. In an attempt to explain some of this missing heritability, researchers are adopting several complementary strategies. Larger cohorts of cases are being collected, through either further patient recruitment or collaboration. The meta-analysis data generated to date have elegantly demonstrated how increasing the cohort sample size generates additional statistical power to detect smaller and smaller odds ratios (5). Advances in technology and particularly bioinformatics have now made it possible to perform GWAS using common copy number variation probes. Studies such as these are currently ongoing. Many groups are looking to high-throughput sequencing technology, with the aim of sequencing candidate gene regions identified by GWAS, to hopefully identify either the causal or rare variants, depending on the number of samples sequenced. The results of these studies are likely to define the next few years of not just chronic inflammatory diseases genetics, but all complex diseases genetics.

Conflict of Interest statement. None declared.

FUNDING

We would like to acknowledge funding from the Juvenile Diabetes Research Foundation International, the Wellcome Trust and Coeliac UK.

REFERENCES

variants regulating ORMEL3 expression contribute to the risk of childhood asthma. Nature, 448, 470–473.


