New developments in genetics of myositis

Simon Rothwell\textsuperscript{a}, Janine A. Lamb\textsuperscript{b}, and Hector Chinoy\textsuperscript{c}

Purpose of review
This article reviews the advances that have been made in our understanding of the genetics of the idiopathic inflammatory myopathies (IIM) in the past 2 years, with a particular focus on polymyositis, dermatomyositis and inclusion body myositis.

Recent findings
Two large human leukocyte antigen (HLA) imputation studies have confirmed a strong association with the 8.1 ancestral haplotype in clinical subgroups of myositis and suggest multiple independent associations on this haplotype. Risk in these genes may be due to specific amino acid positions within the peptide-binding grooves of HLA molecules. A large genetic study in 2566 IIM patients revealed associations such as \textit{PTPN22}, \textit{STAT4}, \textit{UBE2L3} and \textit{BLK}, which overlap with risk variants reported in other seropositive autoimmune diseases. There is also evidence of different genetic architectures in clinical subgroups of IIM. Candidate gene studies in the Japanese and Chinese populations have replicated previous IIM associations which suggest common aetiology between ethnicities.

Summary
International collaborations have facilitated large genetic studies in IIM that have revealed much about the genetics of this rare complex disease both within the HLA region and genome-wide. Future approaches, such as sequencing and trans-ethnic meta-analyses, will advance our knowledge of IIM genetics.

Keywords
genetics, human leukocyte antigen, idiopathic inflammatory myopathies

INTRODUCTION
The idiopathic inflammatory myopathies (IIM) are a spectrum of rare autoimmune diseases clinically characterized by muscle weakness with heterogeneous systemic organ involvement. Clinically, they are subclassified as dermatomyositis, polymyositis, inclusion body myositis (IBM) and increasingly recognized immune-mediated necrotizing myopathy (IMNM). IIM are thought to be complex genetic diseases, initiated by immune activation following specific environmental events in genetically predisposed individuals. Because of the rarity of these diseases, a lack of research has meant that treatment is largely borrowed from other autoimmune diseases with varying degrees of efficacy. Therefore, research into the genetics of these diseases may lead to more effective treatment, prognosis or accurate stratification into research studies.

Together with an emphasis on coordinated case ascertainment in IIM, the advent of high-throughput genetic approaches has enabled investigations of sufficient size to conduct statistically significant genetic analysis, and recent studies have enabled us to further understand the genetic architecture of these rare diseases. A timeline of the landmark studies published in IIM genetic research are shown in Fig. 1. This article reviews the advances that have been made in the past 2 years in our understanding of the genetics of IIM, with a particular focus on polymyositis and dermatomyositis, and potential future directions of research in this rare disease.

MAJOR HISTOCOMPATIBILITY COMPLEX ASSOCIATIONS IN MYOSITIS
The major histocompatibility complex (MHC), also known as the human leukocyte antigen (HLA)
region, has been shown consistently to be the strongest genetic risk factor for autoimmune disease. In IIM, the strongest association is with the 8.1 ancestral haplotype (8.1 AH), a large common haplotype in Caucasian populations that confers susceptibility to many other autoimmune or immune-mediated diseases [1,2]. Although there is strong evidence for association with the 8.1 AH in IIM, it is not clear which gene or genes contribute to disease pathogenesis, as strong linkage disequilibrium within this haplotype means multiple genes are associated strongly with disease.

Recently, two large studies focused on the HLA region in IIM and clinical subgroups of disease. Both have used HLA imputation to impute classical HLA alleles from single nucleotide polymorphism (SNP) genotyping; the first using genome-wide association studies (GWAS) data and 1710 IIM patients [3,4], and the second using 2566 partially overlapping IIM samples using high-density SNP data from the Immunochip study [5] (see section: Genome-Wide Associations in Polymyositis and Dermatomyositis). Interestingly, while in both studies the most associated variants were alleles of the 8.1 AH, when stratifying cohorts by clinical subgroup, there were conflicting results about which gene had the strongest association. Miller et al. [4] reported that in dermatomyositis, the strongest association was with HLA-DRB1/C303:01, while in polymyositis it was with HLA-B/C308:01. This contrasts with Rothwell et al. [5] where in dermatomyositis, HLA-B/C308:01 was the most associated, and in polymyositis it was HLA-DRB1/C303:01. In both instances, these two alleles were associated at similar levels of significance, making it difficult to differentiate between these genes. Rothwell et al. [5] used stepwise conditional

---

**KEY POINTS**

- Two large HLA imputation studies have confirmed a strong association with the 8.1 AH in clinical subgroups of myositis and suggest multiple independent associations on this haplotype.
- A large genetic study in IIM patients revealed multiple non-HLA associations, which overlap with risk variants reported in other seropositive autoimmune diseases.
- Candidate gene studies in the Japanese and Han Chinese populations have replicated previous IIM associations in the Caucasian population suggesting common aetiology between ethnicities.
- Future approaches, such as sequencing and trans-ethnic meta-analyses, and utilizing existing national and international biorepositories, will advance our knowledge of IIM genetics.
analysis to show that both HLA-DRB1*03:01 and HLA-B*08:01 were independently associated in IIM and polymyositis, suggesting that multiple genes on this 8.1 AH may contribute to risk. Indeed, the association with the HLA region is strongest when multiple alleles of the 8.1 AH were included [4**]. Other autoimmune diseases such as rheumatoid arthritis (RA), inflammatory bowel disease and psoriasis have shown that there can be multiple independent risk factors within the MHC region [6–8], and it may be that there are additional risk factors in IIM that we are currently underpowered to detect. The strongest associations with the HLA are found when stratifying by serotype [9], and this was confirmed in patients positive for anti-Jo1 antibodies with a strong association to HLA-DRB1*03:01 [4**]. Further work is required to correlate genotype with the rarer serological subtypes in IIM.

Many studies have attributed risk in the MHC region to the classical class I and class II genes described above, that are involved in antigen presentation and processing. MHC class III genes are structurally and functionally different, coding for proteins involved in regulation of the immune system, and therefore are also candidate genes for association in autoimmune disease. Previous studies have implicated nuclear factor-kB (NF-kB) [10] and TNF-α [11] in polymyositis and dermatomyositis, and NOTCH4 in IBM [12]; however, these variants all reside, or are in linkage with the risk 8.1 AH and therefore may not be independently associated with IIM. A recent study [13*] investigated potential associations with gene copy number variations of complement genes in 105 juvenile dermatomyositis (JDM) patients. Although the role of complement in IIM is not well defined, in dermatomyositis there is evidence of complement-induced vascular damage and muscle fibre ischaemia. Many components of the complement system are encoded within the MHC, and notably the 8.1 AH is strongly associated with a genetic deficiency of complement C4 because of the presence of only a single copy of C4B but the absence of C4A. The authors found that C4A deficiency and HLA-DRB1*03:01 were both risk factors for JDM; however, the strongest effect was concurrent presence of DRB1*03:01 and C4A-deficiency [13*]. Whether the association with C4A deficiency is independent of HLA-DRB1*03:01 is currently unclear; however, both likely result in a permissive background to development of autoimmune disease.

HUMAN LEUKOCYTE ANTIGEN AMINO ACID ASSOCIATIONS

It has been hypothesized that risk within MHC class I and class II genes can be explained by differences in the structure of the peptide-binding pocket affecting the ability to bind antigenic peptides. Key amino acid positions within HLA genes may be responsible for the risk in these genes. For example, in anti-Cyclic Citrullinated Peptide positive RA, specific amino acids in position 11/13 and 74 of HLA-DRB1 and in position 9 of both HLA-B and HLA-DPB1 explain most of the risk within the MHC region [6]. In IIM, two positions in HLA-DRB1 were significantly associated with myositis and its clinical subgroups [5*]. Of these two positions, position 74 lies within the peptide-binding groove, suggesting functional relevance; an arginine at this location is highly associated and may explain the risk of association with HLA-DRB1*03:01 (Fig. 2).

As mentioned previously, the strongest association within the MHC region is with the 8.1 AH, an extended haplotype in which the strong linkage disequilibrium makes it difficult to identify causal genes in the locus. Although many of the largest studies in IIM have been conducted in Caucasian populations, other populations can have unique HLA haplotypes in IIM. For example, HLA-DRB1*08:03 confers risk of IIM in the Japanese population [14], while HLA-DQA1*01:04 and HLA-DRB1*07 alleles are associated with an increased risk of dermatomyositis in the Chinese population [15]. Although these haplotypes differ, they may share common features that confer risk. For example, a study in 2014 [16] constructed a pan-Asian reference panel to impute HLA alleles and amino acids to investigate risk for RA in Asian populations. Although the HLA risk alleles also were found to be unique in different populations, significant amino acid associations were consistent across
Myositis and myopathies

Asian and Caucasian populations. In IIM, it will be interesting to examine whether risk alleles in other populations share the same risk amino acids as in Caucasian populations, such as an arginine at position 74 of HLA-DRB1. Such validation may suggest functionality of these variants.

GENOME-WIDE ASSOCIATIONS IN POLYMYOSITIS AND DERMATOMYOSITIS

As many autoimmune diseases share common associations within the MHC, it is hypothesized that other genes independent of this region will contribute to specific disease susceptibility. There is remarkable genetic overlap between autoimmune diseases; however, these associations are commonly of small effect sizes and require large sample sizes to detect. GWAS compare the frequency of hundreds of thousands of genetic variants between cases and controls in a hypothesis free manner. GWAS of 1187 dermatomyositis patients conducted in 2013 confirmed the strong genetic component in the MHC region, and suggested that there are shared genetic risk factors between IIM and established genetic risk loci for other autoimmune disorders [3]. Follow-up studies have therefore tended to focus on this genetic overlap, while increasing sample sizes and including additional subgroups of IIM. A candidate gene study by Jani et al. [17] in 2014 sought to extend the findings of the original dermatomyositis GWAS by genotyping additional autoimmune variants not captured in the GWAS, and by including 410 polymyositis patients. In this study, the finding that associations with BLK and TYK2 may be dermatomyositis specific highlighted that there may be genetic heterogeneity between polymyositis and dermatomyositis.

A recent study [5] sought to investigate the genetic overlap between IIM and other autoimmune diseases with a sample size large enough to potentially identify differences between the different subgroups of polymyositis and dermatomyositis. Two thousand five hundred and sixty-six Caucasian patients were recruited through the Myositis Genetics Consortium (MYOGEN) and genotyped using the Immunochip. The Immunochip is a high-density SNP array covering 186 established autoimmune-associated loci. To be expected, the HLA was the most associated region, with PTPN22 also reaching genome-wide significance. When including associations that reached a more conservative level of significance \(P = 2.25 \times 10^{-5}\), there was a large overlap of associations that had previously been associated with particular autoimmune diseases. For example, PTPN22, STAT4, UBE2L3, BLK and HLA class II genes are associated with seropositive rheumatic diseases such as RA, systemic lupus erythematosus, Sjögren’s syndrome and systemic sclerosis [18]. This provides evidence of key pathogenic mechanisms in IIM, as these genes are all involved in activation of the adaptive immune system directed against autoantigens.

Stratification by polymyositis \(n = 931\) and dermatomyositis \(n = 1360\) suggested that there may be risk loci specific to these subgroups of IIM. For example, a variant in PTPN22 was associated with polymyositis, and is involved in T cell signalling. A variant in BLK was associated with dermatomyositis and is known to be involved in B cell activation. These associations suggest different genetic architectures underpinning these subgroups and functionally fit with our current understanding of the pathogenesis of IIM. Other associated genes such as RGS1 and IL18R1 in polymyositis, and GSDMB in dermatomyositis suggest novel mechanisms that may differentiate between these diseases. Evidence that there are both shared and distinct genetic risk factors within subgroups of polymyositis and dermatomyositis underlines the importance of ongoing sample collection, to enable studies with greater statistical power in more homogenous subgroups of IIM.

There is still further research needed to elucidate the mechanisms behind these associations, and their function in disease pathogenesis. Some associations, such as the risk variant in PTPN22, result in an amino acid change that directly affects the function of the protein. Commonly, however, disease associations fall in intergenic regions wherein their function is unknown. It is likely that they fall within gene regulatory regions affecting the expression of genes in cell types crucial to disease pathogenesis. An interesting observation from the Immunochip analysis was that fewer genes in the dermatomyositis subgroup reached suggestive significance than in the polymyositis subgroup, suggesting that the Immunochip explains less genetic risk in dermatomyositis. Indeed, post-hoc analysis on these data (Rothwell, unpublished data) supports this hypothesis. Genome-wide complex trait analysis (GCTA) uses all SNPs to estimate the total amount of phenotypic variance explained by the array [19]. In polymyositis, SNPs on the Immunochip explain 8.3% of the phenotypic variance, whereas in dermatomyositis, this is only 5.5%. That the Immunochip explains less genetic variance in dermatomyositis may be due to selected content of the Immunochip favouring genes involved in polymyositis, heterogeneity of phenotypes present within dermatomyositis or a weaker genetic influence compared with other autoimmune diseases.
GENETIC ASSOCIATIONS IN NON-CAUCASIAN POPULATIONS

Reassuringly, many genetic associations reported in Caucasian studies in IIM have been replicated in other ethnic populations suggesting a common aetiology. In the Han Chinese population, candidate gene studies have replicated associations with CCL21, BLK and PLCLI1 that have been reported in Caucasian IIM patients [20–22], as well as pan-autoimmune risk loci TNFAIP3 and IRRS [23]. In the Japanese population, associations with STAT4 and BLK also have been replicated [24,25]. In a rare disease, replication of statistically ‘suggestive’ associations such as those described above, along with biological rationale and evidence of functionality, may allow us to interpret these results with more confidence. This evidence of common aetiology between populations suggests that there may be an opportunity to conduct larger trans-ethnic association studies to increase power, and also to break down large population-specific haplotype blocks for identification of causal variants.

GENETICS OF INCLUSION BODY MYOSITIS

Most research into the genetics of IBM has been conducted using candidate gene studies focusing on genes known to be involved in neurodegenerative diseases such as Alzheimer’s disease, or genes previously implicated in the autosomal-recessive or dominant form of IBM known as hereditary IBM. However, because of small sample sizes, these have failed to find significant common associations. A well-studied locus is the APOE region, a risk locus for Alzheimer’s disease. Multiple studies [26–28] have implicated APOE, as well as TOMM40, a gene in linkage disequilibrium. A recent study [29] sought to replicate these findings in a larger IBM cohort (n = 158), and although genotyping APOE and TOMM40 showed no significant associations with risk of developing IBM, a potential association with later onset of symptoms was reported. Two hundred and fifty-two IBM patients were also included in the IIM Immunochip study described above, and analysis of these data was presented at a recent rheumatology meeting [30]. A suggestive association with CCR5 in IBM suggests that immunorelated genes may have a role in the aetiology of IBM, and confirms that multiple HLA haplotypes are associated with disease.

Because of the rarity of IBM, it is difficult to ascertain the sample sizes needed for GWAS, therefore next generation sequencing could be an approach to detect rare, potentially functional variants of large effect size. A recent study [31*] sequenced 38 candidate genes in 79 IBM patients. The authors identified 27 rare missense-coding variants, including mutations in VCP, a gene known to be associated with hereditary IBM, demonstrating that sequencing can be a clinically useful method of detecting potentially causal variants in IBM. Studies are currently underway taking a hypothesis free approach and sequencing exomes of a large number of IBM patients [32]. In a disease wherein the aetiology is unknown, this strategy could be successful in identifying novel variants and/or pathways involved in disease pathogenesis.

CONCLUSION

In IIM, substantial genetic risk resides within the MHC; however, large studies are beginning to reveal associations outside this region that overlap with other seropositive autoimmune diseases suggesting common aetiologies and pathways. In addition, there is evidence of non-HLA associations that differentiate between clinical subgroups of IIM that may be useful for future research that leads to patient benefit in the clinic. Although advances in technology and analytical methods, such as the Immunochip and HLA imputation, have been invaluable, much of the success can be put down to increasing sample sizes possible because of consortia such as MYOGEN. A recent review highlights the large number of biorepositories of myositis samples that potentially could be utilized in future genetic studies [33*]. Further research will require more homogenous and larger cohorts, which may be facilitated by collaborations and trans-ethnic meta-analysis between international consortia.

Acknowledgements

None.

Financial support and sponsorship

This work was funded by an MRC Partnership Grant (MR/N003322/1).

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest

• of outstanding interest

Myositis and myopathies


This large HLA imputation study was conducted on subgroups of IM. It confirms a strong association with alleles of the 8.1 AH and suggests that the strongest effects were seen when multiple alleles of the 8.1 AH were included together. The strongest association was seen in anti-Jo-1 autoantibody-positive patients indicating that genetic studies should be conducted on clinically homogeneous subgroups.


This is the largest genetic study to date in IM and reveals associations overlapping with other seropositive autoimmune diseases, as well as suggesting associations with amino acid positions within risk HLA loci.


The first study to suggest a genetic association with copy number variants of complement genes in JDM patients.


30. Rothwell S, Cooper RG, Lundberg IE, et al. Largest genetic study to date in sporadic inclusion body myositis confirms the human leukocyte antigen as the most associated region and suggests a role for C-C chemokine receptor type 5. Rheumatology 2016; 55:48.


The first next generation sequencing study published in IBM. The investigators reported rare variants in the VCP gene, demonstrating that sequencing can be a clinically useful method in detecting rare/novel variants in IBM.


A thorough review of existing clinical registries and biorepositories in IM that could potentially be utilized in future genetic studies.