Rothwell Simon (Orcid ID: 0000-0003-2123-9902)
Miller Frederick W (Orcid ID: 0000-0003-2831-9593)
Rider Lisa G. (Orcid ID: 0000-0002-6912-2458)
Lundberg Ingrid E. Elisabeth (Orcid ID: 0000-0002-6068-9212)
Selva-O'Callaghan Albert (Orcid ID: 0000-0003-2823-9761)
Machado Pedro M. (Orcid ID: 0000-0002-8411-7972)
Reed Ann (Orcid ID: 0000-0002-4521-7647)
Doria Andrea (Orcid ID: 0000-0003-0548-4983)
De Paepe Boel (Orcid ID: 0000-0001-9403-4401)
Wedderburn Lucy R (Orcid ID: 0000-0002-7495-1429)
Lamb Janine A (Orcid ID: 0000-0002-7248-0539)

Genome-wide imputation identifies novel associations and localises signals in idiopathic inflammatory myopathies

Simon Rothwell PhD[1], Christopher I. Amos PhD[2], Frederick W. Miller MD PhD[3], Lisa G. Rider MD[3], Ingrid E. Lundberg MD PhD[4], Peter K. Gregersen MD[5], Jiri Vencovsky MD PhD[6], Neil McHugh MD FRACP FRCPath[7], Vidya Limaye MBBS FRACP PhD[8], Albert Selva-O'Callaghan MD PhD[9], Michael G. Hanna[10], Pedro M. Machado MD PhD FRCP[10,11], Lauren M. Pachman MD[12], Ann M. Reed MD[13], Øyvind Molberg MD PhD[14], Olivier Benveniste MD PhD[15], Pernille Mathiesen MD PhD[16], Timothy Radstake MD PhD[17], Andrea Doria MD[18], Jan L. De Bleecker MD PhD[19], Boel De Paepe PhD[19], Britta Maurer MD[20], William E. Ollier PhD[21], Leonid Padyukov MD PhD[4], Terrance P. O'Hanlon PhD[3], Annette Lee PhD[5], Lucy R. Wedderburn MBBS PhD FRCP MRCPCH[22,23], Hector Chinoy PhD FRCP BMBS[24-26], Janine A. Lamb DPhil[27*] and the Myositis Genetics Consortium[28]

1. Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
2. Baylor College of Medicine, Houston, TX, USA
3. Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, National Institutes of Health, Bethesda, MD, USA
4. Division of Rheumatology, Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
5. The Robert S Boas Center for Genomics and Human Genetics, The Feinstein Institute, Manhasset, NY, USA
6. Institute of Rheumatology and Department of Rheumatology, 1st Medical Faculty, Charles University, Prague, CZ
7. Department of Pharmacy and Pharmacology, University of Bath, Bath, UK
8. Rheumatology Unit, Royal Adelaide Hospital and Discipline of Medicine, Adelaide University, Adelaide, Australia
9. Internal Medicine Department, Vall d'Hebron General Hospital, Universitat Autonoma de Barcelona, Barcelona, Spain
10. Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London, UK
11. Centre for Rheumatology, UCL Division of Medicine, University College London, London, UK
12. Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
13. Department of Pediatrics, Duke University, Durham, NC, USA
14. Department of Rheumatology, Oslo University Hospital, Oslo, Norway
15. Department of Internal Medicine and Clinical Immunology, Pitié-Salpêtrière Hospital, Paris, France
16. Paediatric Department, Slagelse Hospital and Paediatric Rheumatology Unit, Rigshospitalet, Copenhagen, Denmark
17. Department of Rheumatology and Clinical Immunology, University Medical Center, Utrecht, the Netherlands
18. Rheumatology Unit, Department of Medicine, University of Padova, Padova, Italy
19. Department of Neurology, Ghent University, Ghent, Belgium
20. Department of Rheumatology and Immunology, University Hospital, Bern, CH
21. Manchester Metropolitan University, School of Healthcare Sciences, Manchester, UK
22. NIHR Biomedical Research Centre at Great Ormond Street Hospital, London, United Kingdom
23. Arthritis Research UK Centre for Adolescent Rheumatology, UCL Great Ormond Street Institute of Child Health, University College London, London, UK
24. National Institute for Health Research Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, The University of Manchester, Manchester, UK
25. Department of Rheumatology, Salford Royal Hospital, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Science Centre, Salford, UK
26. Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/art.42434

27. Epidemiology and Public Health Group, Division of Population Health, Health Services Research & Primary Care, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
28. Contributing Myositis Genetics Consortium members listed in the supplementary materials

* Corresponding author: Janine A. Lamb
Epidemiology and Public Health Group
University of Manchester
Manchester
UK
Janine.Lamb@manchester.ac.uk

Abstract (226 words)

Objectives: The idiopathic inflammatory myopathies (IIM) are heterogeneous diseases, thought to be initiated by immune activation in genetically predisposed individuals. In this study we imputed variants from the Immunochip array using a large reference panel to fine-map associations and identify novel associations in IIM.

Methods: We analysed 2,565 Caucasian IIM samples collected through the Myositis Genetics Consortium (MYOGEN) and 10,260 ethnically-matched controls. We imputed 1,648,116 variants from the Immunochip array using the Haplotype Reference Consortium panel and conducted association analysis on IIM, and clinical and serological subgroups.

Results: The human leukocyte antigen (HLA) locus was consistently the most significantly associated region. Four non-HLA regions reached genome-wide significance, three in the whole IIM cohort (*SDK2* and *LINC00924* - both novel, and *STAT4*), with evidence of independent variants in *STAT4*, and *NAB1* in the polymyositis (PM) subgroup. We also found suggestive evidence of association with loci previously associated with other autoimmune rheumatic diseases (*TEC* and *LTBR*). We identified more significant associations than those previously reported in IIM, for *STAT4* and *DGKQ* in the total cohort, for *NAB1* and *FAM167A-BLK* loci in PM, and *CCR5* in inclusion body myositis. We found enrichment of variants among DNase I hypersensitivity sites and histone marks associated with active transcription within blood cells.

Conclusions: We report novel and strong associations in IIM and PM, and localise signals to single genes and immune cell types.

Main Body (1999 words)

Introduction

The idiopathic inflammatory myopathies (IIMs) are a heterogeneous group of rare autoimmune diseases primarily characterised by muscle weakness with extra-muscular manifestations. The strongest genetic risk for IIM resides in the human leukocyte antigen (HLA) region, although non-HLA associations have also been reported. To date, the largest genetic association studies have been conducted in Caucasian populations through the Myositis Genetics Consortium (MYOGEN) [1] and subsequent meta-analyses [2]. However a genome-wide association study (GWAS) in clinically amyopathic dermatomyositis in the Japanese population [3], and candidate gene studies in Japanese and Chinese populations have also identified significant genetic risk factors for IIM [4–6].

We previously published a genetic association study on 90,536 genetic variants from the Immunochip, a targeted array containing coverage of 186 established autoimmune susceptibility loci [1]. In this follow-up study we re-analysed the IIM Immunochip dataset after imputation of 1,648,116 variants to identify novel associations and to facilitate fine mapping of risk regions reported in IIM and clinical and serological subgroups.

Methods

Samples

Caucasian IIM samples were collected through MYOGEN [1]. IIM cases were included if they fulfilled probable or definite Bohan and Peter classification criteria for polymyositis (PM), juvenile PM (JPM), dermatomyositis (DM) or juvenile DM (JDM), and Griggs, or European Neuromuscular Centre, or Medical Research Council criteria for inclusion body myositis (IBM) [7]. Immunochip control data from 12 countries was provided by four disease consortia. Anti-Jo1 autoantibodies were detected using either immunoprecipitation (IP) or line blot using methods described previously [8].

Genotyping and Imputation

Analysis was conducted on the existing Illumina Immunochip Array data consisting of 2,565 IIM patients [1]. Clinical subgroup analysis was conducted on PM (n=903), DM (n=817), JDM (n=508), IBM (n=252) and patients with anti-Jo1 autoantibodies (n=311). Controls were taken from the same pool of controls, but matched with cases 4:1 based on principal components analysis (PCA) coordinates [8]. Variants were imputed with the Michigan Imputation Server using the Haplotype Reference Consortium panel (HRC) (Version r1.1 2016). The HRC consists of 64,940 haplotypes of predominantly European ancestry. Poorly imputed genotypes (r^2<0.5), SNPs deviating from Hardy Weinberg Equilibrium in controls (p<0.001) and those with a low MAF (MAF<0.01) were removed. After stringent SNP quality control (QC), we analysed 1,648,116 variants, of which 120,734 were directly genotyped SNPs. Population stratification was assessed by the genomic inflation factor (λ) scaled to 1000 cases and 1000 controls. Including the top three principal components as covariates was sufficient to control for population differences (λ_{1000} = 1.04).

Statistical Analysis

Association analysis was conducted on gene dosages using SNPTEST. Linkage disequilibrium (LD) between SNPs was calculated using PLINK v1.90 The first three principal components were included as covariates and used in a logistic regression analysis using an additive model. Forward stepwise logistic regression was used to test for independent effects conditional on the variant of interest.

Genome-wide significance is defined as $p<5\times10^{-8}$. Suggestive significance is defined as $p<2.25\times10^{-5}$, based on Bonferroni correction for the number of independent haplotype blocks on the original genotyping array. Regions were defined as novel if there was no genome-wide or suggestive evidence of association in previous genetic analyses. Regional association plots were generated using LocusZoom.js [9].

Functional Analysis

We report the functional effect of the most strongly associated SNP in the locus from dbSNP's predicted functional effect. GARFIELD analysis (GWAS Analysis of Regulatory and Functional Information Enrichment with LD correction) was used to characterise the cellular and regulatory contribution of the associated variants [10]. 95% credible SNP sets were calculated using the 'gwas-credible-sets' package implemented in LocusZoom.js [9] and annotated with functional information from public databases including GTEx expression and splicing QTLs, and regulatory information from the UCSC genome browser.

Results

Regions of interest associated with IIM and clinical subgroups are presented in Table 1 and Figure 1. SNPs are reported if they reach genome-wide significance, or reach suggestive significance and the locus has previous statistical evidence of association in an autoimmune rheumatic disease in the NHGRI-EBI GWAS Catalog. All associations reaching a suggestive significance threshold of $p<2.25x10^{-5}$ are included in the supplementary data. A Manhattan plot of the combined IIM analysis is shown in Figure 1. Variants reaching suggestive significance are coloured green if directly genotyped, to differentiate between imputed variants in grey/black. Manhattan plots for subgroup analyses, and regional association plots for non-HLA associations are included in supplementary figures 1-19.

For IIM as a whole and all clinical subgroups the HLA region is the most significant genetic risk factor. Three non-HLA regions, *STAT4* (Signal Transducer and Activator of Transcription 4) ($p=1.38x10^{-8}$, OR=0.81, 95% CI= 0.75-0.87), *SDK2* (Sidekick Cell Adhesion Molecule 2) ($p=1.46x10^{-8}$, OR=1.15, 95% CI=1.08-1.23) and *LINC00924* (Long Intergenic Non-Protein Coding RNA 924) ($p=1.9x10^{-8}$, OR=0.84, 95% CI=0.79-0.89), reached genome-wide significance in the total IIM cohort, and one non-HLA region, *NAB1* (NGFI-A Binding Protein 1) ($p=1.96x10^{-8}$, OR=1.41, 95% CI=1.24-1.60), reached genome-wide significance in the PM subgroup. This is the first time these loci have reached genome-wide significance in IIM. After conditioning on the SNP with the lowest p-value in these regions, there were no additional independent effects reaching genome-wide significance. However, an independent intronic variant in *STAT4* remained suggestively significant in IIM (rs6752770, $p=6.1x10^{-6}$, OR=1.11, 95% CI=1.04-1.19) (supplementary figure 7).

The 95% credible SNP sets for the 4 regions reaching genome-wide significance are included in the supplementary data, along with predicted deleteriousness and functionality using CADD and regulomeDB respectively. For the *STAT4* and *NAB1* regions, a substantial proportion of the 95% posterior probability for credible SNPs could be attributed to a single SNP; rs4853540 (*STAT4*) maps to an enhancer for *STAT1*, and rs6733720 (*NAB1*) is a significant sQTL for *NAB1* in 26 tissues ($p<1.6 \times 10\text{-}6$) and a significant eQTL for several genes, including *NAB1* and *GLS* (glutaminase), in multiple tissues (supplementary data).

The *LINC00924* loci is a novel association in autoimmune disease, reaching genome-wide significance. Most variants in these regions were imputed (coloured grey/black), explaining the lack of association in the prior IIM Immunochip study, where the original SNP may have been removed during QC[1]. Notably, we also see suggestive evidence of association with *TEC* (Tyrosine-Protein

Kinase Tec) and *LTBR* (Lymphotoxin Beta Receptor), which have previously been associated with autoimmune rheumatic diseases [11–13]. Associations with the *STAT4* and *DGKQ* (Diacylglycerol Kinase Theta) loci are more significant than in the original IIM Immunochip study, as are associations previously reported in IIM clinical subgroups, such as *NAB1* and *FAM167A-BLK* (Family with sequence similarity 167, member A - B lymphocyte kinase) loci in PM and *CCR5* (C-C chemokine receptor type 5) in IBM (Table 1).

We used GARFIELD to assess whether IIM associated variants are enriched in regulatory elements of specific cell types [10]. We found enrichment of variants among DNase I hypersensitivity (DHS) sites and histone marks associated with active transcription within blood cells (supplementary figure 20 and supplementary data). Specifically, IIM variants were most enriched within DHS hotspots of primary CD19+ B-cells ($p=1.1x10^{-16}$) and CD3+ T-cells ($p=3x10^{-15}$).

Discussion

By using imputation to identify and fine-map genetic associations in IIM we found three new genome-wide associations in the combined IIM cohort: *STAT4*, *SDK2* and *LINC00924*. Conditional analysis revealed evidence of independent associations in the *STAT4* region. For the whole IIM group and clinical subgroups, the HLA region is the most significant genetic risk factor. The strongest genetic risk in this region was for the anti-Jo1 subgroup, despite a sample size of only 331 patients. In the PM subgroup, we report one novel genome-wide association with *NAB1*. This is the first time this dataset has been used for genome-wide imputation, and the first time it has been stratified by adult and juvenile-onset myositis subgroups, and by individuals with anti-Jo1 autoantibodies.

We defined an associated region by proximity to the nearest gene. For example, in the *LINC00924* region, the strongest association is intergenic, lying approximately 150Kb from this long intergenic non-coding (LINC) RNA, which has been associated with a number of traits such as coronary heart disease and ischemic stroke. However, it may be that these associations are influencing a different gene or regulatory element lying further away from the most associated SNP. It is worth noting that the lead variant in the *LINC00924* locus was removed during QC in the original Immunochip analysis. In other instances, such as in the *SDK2* region, the most associated variants lie intronic of the gene. *SDK2* is a member of the immunoglobulin superfamily although the specific function is not yet known. In both instances there is no previous evidence of genetic association with rheumatic diseases. A limitation of this study is the lack of a replication cohort due to the rarity of IIM, although sample recruitment is ongoing within the MYOGEN consortium.

For some regions, we found suggestive novel associations with prior evidence in autoimmune disease, such as *TEC*, a tyrosine kinase involved in T-cell signalling and activation, and with *LTBR*, a signalling receptor expressed on myeloid cells. In other instances we strengthened previous associations with IIM (*STAT4* and *DGKQ*). In addition, it is interesting to note the suggestive association with *PLCL1* (Inactive phospholipase C-like protein 1) ($p=1.45x10^{-5}$, OR=1.14, 95% CI=1.07-1.21) in the combined IIM analysis. Variants in *PLCL1,* a gene encoding the intracellular signalling molecule PLCL1*,* were originally reported in the MYOGEN DM GWAS, although it was not identified in the subsequent Immunochip analysis. We have identified more significant associations than were previously reported in IIM clinical subgroups, such as *NAB1* and *FAM167A-BLK* loci in PM and *CCR5* in IBM, which also have prior evidence of association in rheumatic disease. Although the study was comprised of the same cohort as previous studies, this analysis can be viewed as a fine mapping experiment; imputation from a large reference panel allows better coverage to localise further signals. Indeed, we found that many associations can be localised to single genes or credible SNPs with high posterior probability, likely due to the coverage of coding genes on the Immunochip resulting in high quality imputation. We note that the Immunochip is a targeted array, and therefore imputation coverage is not genome-wide.

Functional analysis of variants reaching suggestive significance in IIM was conducted using GARFIELD. This method uses functional annotation from primary tissues and cell lines from the ENCODE and Roadmap Epigenomics projects. As expected from data generated using the Immunochip, associated variants in IIM were enriched in regulatory elements of blood cells. In particular, the strongest relative enrichment was seen in regions of open chromatin in CD19+ B cells and CD3+ T cells. The role of T and B cells is well known in IIM through muscle immunohistopathology and the presence of autoantibodies, and our findings suggest their contribution to disease pathology may be genetically encoded.

Increasing the number of cases in an analysis should statistically strengthen suggestive associations from previous studies. A recent genome-wide meta-analysis in four seropositive rheumatic diseases revealed several novel loci, for example, *NAB1*, *DGKQ* and *YDJC* (YdjC Chitooligosaccharide Deacetylase Homolog), where IIM contributed to the association. Using additional IIM cases to those included in the meta-analysis, we were able to identify these associations in IIM only, and in some cases, such as *NAB1* in PM, attribute these associations to specific clinical subgroups of IIM. The lead variant rs6733720 in *NAB1* is a significant sQTL for *NAB1* in 26 tissues and an eQTL for several genes, including *NAB1* and *GLS*, in different tissues. These *NAB1* variants are also are in moderate LD (r^2=0.66) with a variant previously identified in SSC and RA, which acts as an eQTL for *NAB1* expression in lymphoblastoid cell lines [2].

Although we found a number of interesting associations in the PM subgroup, one limitation is the potential heterogeneity within this subgroup. The Bohan and Peter criteria do not differentiate between PM and immune mediated necrotizing myopathy. In addition, we included patients classified as PM that have autoantibodies against tRNA-synthetases, although anti-synthetase syndrome is increasingly recognised as a separate entity. Some previous genetic studies in IIM combined adult- and juvenile-onset DM for analysis. This study has stratified DM by age of onset to investigate non-HLA associations. However, there do not seem to be strong signals that differentiate these clinical subgroups. Our previous work has shown that stratifying IIM cohorts by autoantibody status may increase power to detect genetic associations within the HLA region [8]. Therefore, we also analysed a subgroup of patients with anti-Jo1 autoantibodies; however, we did not find any significant associations outside the HLA region. Although the anti-Jo1 subgroup is thought to be more clinically homogeneous, there may be a lack of power to detect associations with only 331 patients in the analysis. For this reason, we did not investigate genetic associations with other, less common autoantibodies. To our knowledge, this is the largest genetic association study investigating non-HLA genes in patients with anti-Jo1 autoantibodies. A recent study targeted a number of SNPs in the *IL1B* (Interleukin 1 beta) locus in a Mexican cohort of 154 anti-synthetase positive IIM patients and found an association with a synonymous SNP in *IL1B* [14]. We could not replicate this association in our Caucasian analysis (rs1143634, p=0.1). In the original Immunochip analysis *PTPN22* (Protein Tyrosine Phosphatase Non-Receptor Type 22) was the only non-HLA region to reach genome-wide significance. In this analysis, *PTPN22* does not reach genome-wide significance. This disparity may be due to the more accurate matching of cases to controls, as it is known that there is a wide variation of allele frequency among different European populations of the PTPN22 R620W risk polymorphism [15].

In summary, we have used imputation to identify and fine-map genetic associations in IIM. We report four new genome-wide associations in IIM and PM, and report associations reaching suggestive significance that have previously been associated with autoimmune rheumatic disease. These risk variants are functionally enriched in relevant immune cells, expanding our knowledge of the genetic architecture of IIM.

Acknowledgements.

We would like to thank Drs. Elaine Remmers and Michael Ombrello for their outstanding advice. Additionally, we would like to thank all the patients and healthy controls for their essential collaboration.

This research was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. JV acknowledges the support from the Czech Ministry of Health - Conceptual Development of Research Organization 00023728 (Institute of Rheumatology).

BDP and JDB acknowledge the European Reference Network for Rare Neuromuscular Diseases network (ERN EURO-NMD). LRW acknowledges support from Versus Arthritis (21593, 21552), Wellcome trust (085860), Myositis UK, Cure JM Foundation, Remission Charity and the NIHR Biomedical research Centre at GOSH. JAL and HC acknowledge support from the Medical Research Council UK (MR/N003322/1) and Myositis UK. HC is supported by the National Institution for Health Research Manchester Biomedical Research Centre Funding Scheme. PMM and MGH acknowledge support from the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. IEL received grants from the Swedish Research Council and Stockholm Regional Council (ALF).

A full list of MYOGEN collaborators is included in the supplementary materials

References

1 Rothwell S, Cooper RG, Lundberg IE, *et al.* Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. *Ann Rheum Dis* 2016;**75**:1558–66. doi:10.1136/annrheumdis-2015-208119

2 Acosta-Herrera M, Kerick M, González-Serna D, *et al.* Genome-wide meta-analysis reveals shared new loci in systemic seropositive rheumatic diseases. *Ann Rheum Dis* 2019;**78**:311–9. doi:10.1136/annrheumdis-2018-214127

3 Kochi Y, Kamatani Y, Kondo Y, *et al.* Splicing variant of WDFY4 augments MDA5 signalling and the risk of clinically amyopathic dermatomyositis. *Ann Rheum Dis* 2018;**77**. doi:10.1136/annrheumdis-2017-212149

4 Sugiura T, Kawaguchi Y, Goto K, *et al.* Positive association between STAT4 polymorphisms and polymyositis/dermatomyositis in a Japanese population. Ann. Rheum. Dis. 2012;**71**:1646–50. doi:10.1136/annrheumdis-2011-200839

5 Chen S, Wang Q, Wu Z, *et al.* Genetic association study of TNFAIP3, IFIH1, IRF5 polymorphisms with polymyositis/dermatomyositis in Chinese Han population. *PLoS One* 2014;**9**:e110044. doi:10.1371/journal.pone.0110044

6 Chen S, Wu W, Li J, *et al.* Single nucleotide polymorphisms in the FAM167A-BLK gene are associated with polymyositis/dermatomyositis in the Han Chinese population. *Immunol Res* 2015;**62**:153–62. doi:10.1007/s12026-015-8646-0

7 Rothwell S, Cooper RG, Lundberg IE, *et al.* Immune-Array Analysis in Sporadic Inclusion Body Myositis Reveals HLA-DRB1 Amino Acid Heterogeneity Across the Myositis Spectrum. *Arthritis Rheumatol* 2017;**69**:1090–9. doi:10.1002/art.40045

8 Rothwell S, Chinoy H, Lamb JA, *et al.* Focused HLA analysis in Caucasians with myositis identifies significant associations with autoantibody subgroups. *Ann Rheum Dis* 2019;**78**:996–1002. doi:10.1136/annrheumdis-2019-215046

9 Boughton AP, Welch RP, Flickinger M, *et al.* LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics* 2021;**37**:3017–8. doi:10.1093/BIOINFORMATICS/BTAB186

10 Iotchkova V, Ritchie GRS, Geihs M, *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet* 2019;**51**:343–53. doi:10.1038/s41588-018-0322-6

11 Evans DM, Spencer CCA, Pointon JJ, *et al.* Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 2011;**43**:761–7. doi:10.1038/ng.873

12 Hinks A, Cobb J, Marion MC, *et al.* Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet* 2013;**45**:664–9. doi:10.1038/ng.2614

13 Okada Y, Wu D, Trynka G, *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;**506**:376–81. doi:10.1038/nature12873

14 Ponce-Gallegos MA, Ramos-Martínez E, García-Carmona A, *et al.* Genetic Susceptibility to Antisynthetase Syndrome Associated With Single-Nucleotide Variants in the IL1B Gene That Lead Variation in IL-1β Serum Levels. *Front Med* 2020;**7**:547186. doi:10.3389/fmed.2020.547186

15 Burn GL, Svensson L, Sanchez-Blanco C, *et al.* Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? FEBS Lett. 2011;**585**:3689–98. doi:10.1016/j.febslet.2011.04.032

16 Bentham J, Morris DL, Cunninghame Graham DS, *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015;**47**:1457–64. doi:10.1038/ng.3434

17 López-Isac E, Acosta-Herrera M, Kerick M, *et al.* GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat Commun* 2019;**10**:1–14. doi:10.1038/s41467-019-12760-y

18 Wang YF, Zhang Y, Lin Z, *et al.* Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat Commun* 2021;**12**:1–13. doi:10.1038/s41467-021-21049-y

19 López-Isac E, Smith SL, Marion MC, *et al.* Combined genetic analysis of juvenile idiopathic arthritis clinical subtypes identifies novel risk loci, target genes and key regulatory mechanisms. *Ann Rheum Dis* 2020;**80**:321–8. doi:10.1136/annrheumdis-2020-218481

Table 1. Loci associated with IIM

Gene region	Subgroup	SNP	Chr	Position	Minor Allele	MAF Cases	MAF Controls	P-value	OR (95% CI)	Function†	Overlap
HLA-DRA	IIM	rs9268813	6	32424594	C	0.24	0.11	6.69×10^{-120}	2.49 (2.30-2.69)		Multiple
*STAT4**	IIM	rs4853540	2	191917317	T	0.19	0.22	1.38×10^{-8}	0.81 (0.75-0.87)	Intronic	RA [13], SLE [16], SSc [17]
SDK2	IIM	rs7209879	17	71532097	T	0.37	0.34	1.46×10^{-8}	1.15 (1.08-1.23)	Intronic	
LINC00924	IIM	rs8040452	15	96197257	T	0.37	0.42	1.9×10^{-8}	0.84 (0.79-0.89)	Intergenic	
PHTF1-PTPN22	PM	rs6679677	1	114303808	A	0.12	0.1	1.57×10^{-7}	1.3 (1.19-1.43)	Exonic	RA [13], SLE [16]
DGKQ	IIM	rs6599390	4	956047	A	0.3	0.34	1.64×10^{-7}	0.84 (0.78-0.89)	Intronic	Seropositive RD [2], SSc [17]
UBE2L3-YDJC	IIM	rs11089637	22	21979096	C	0.19	0.16	1.23×10^{-6}	1.23 (1.13-1.33)	Intergenic	RA [13], JIA [12], SLE [16]
TEC	IIM	rs80105690	4	48155618	T	0.09	0.07	6×10^{-6}	1.27 (1.14-1.42)	Intronic	RA [13]
PLCL1	IIM	rs1518359	2	198847383	T	0.52	0.49	1.45×10^{-5}	1.14 (1.07-1.21)	Intronic	SLE [18]
LTBR	IIM	rs11064180	12	6523249	T	0.38	0.41	1.58×10^{-5}	0.87 (0.81-0.92)	Intergenic	JIA [12], AS [11]
HLA-DQB1	PM	rs3129716	6	32657436	C	0.29	0.13	1.54×10^{-54}	2.65 (2.35-3.00)		Multiple
NAB1	PM	rs6733720	2	191516020	G	0.22	0.17	1.96×10^{-8}	1.41 (1.24-1.60)	Intronic	Seropositive RD [2], SSc [17]
PTPN22	PM	rs2476601	1	114377568	A	0.15	0.1	1.25×10^{-6}	1.46 (1.26-1.70)	Exonic	RA [13], SLE [16]
FAM167A-BLK	PM	rs17799348	8	11333521	T	0.33	0.39	1.72×10^{-6}	0.77 (0.69-0.85)	Intergenic	Seropositive RD [2], SSc [17], RA [13], SLE [16]
HLA-B	DM	rs7748141	6	31288877	C	0.23	0.12	2.70×10^{-28}	2.17 (1.89-2.49)		Multiple
HLA-DRA	JDM	rs12204922	6	32451613	C	0.3	0.18	7.99×10^{-16}	1.91 (1.63-2.23)		Multiple
HLA-DQB1	IBM	rs4713570	6	32626040	T	0.57	0.24	3.9×10^{-45}	4.15 (3.39-5.09)		Multiple
CCR5	IBM	rs41490645	3	46410137	C	0.08	0.16	5.96×10^{-7}	0.45 (0.32-0.63)	Upstream Variant	JIA [19]
HLA-DRA	Jo-1	rs9268813	6	32424594	C	0.39	0.11	5.01×10^{-65}	5.15 (4.21-6.32)		Multiple

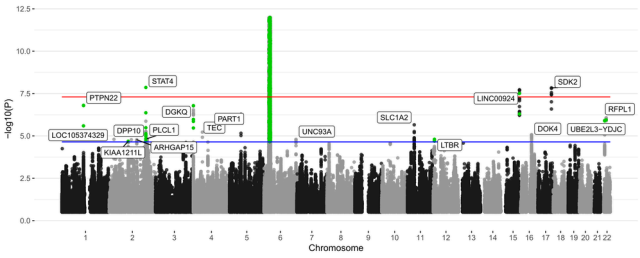
Loci reported at genome-wide significance ($p < 5 \times 10^{-8}$), except for loci with previous statistical evidence of association in an autoimmune rheumatic disease (RD) in the NHGRI-EBI GWAS Catalog, which were reported at suggestive significance ($p < 2.25 \times 10{-}5$). AS – Ankylosing Spondylitis, Chr – chromosome, JIA – Juvenile Idiopathic Arthritis, MAF – minor allele frequency, OR – Odds ratio, RA – Rheumatoid Arthritis, SLE – Systemic Lupus Erythematosus, SSc - Systemic Sclerosis, 95% CI – 95% confidence interval

** Locus with evidence of independent effects*

† Function taken from dbSNP's predicted functional effect

Figure Legend

Figure 1. *Manhattan plot of the total IIM (n=2565) association analysis. Red line represents genome-wide level of significance ($p < 5\times 10^{-8}$); blue line represents suggestive significance calculated from the original coverage from the Immunochip array ($p < 2.25 \times 10^{-5}$). SNPs reaching $P < 2.25 \times 10^{-5}$ that were directly genotyped are coloured in green to differentiate between imputed variants. For visualisation purposes the Y axis has a cut-off in the HLA region (chromosome 6 25–35Mb) of $p < 1 \times 10^{-1}$*



ART_42434_Figure 1a.tiff