Allele frequency spectrum of known ankylosing spondylitis associated variants in a Swedish population

A Mathioudaki, J Nordin, A Kastbom, P Söderkvist, P Eriksson, J Cedergren, K Lindblad-Toh & JRS Meadows

View supplementary material

Published online: 25 Jun 2021.

Submit your article to this journal

Article views: 718

View related articles

View Crossmark data

Allele frequency spectrum of known ankylosing spondylitis associated variants in a Swedish population

A Mathioudaki[1], J Nordin[2], A Kastbom[3], P Söderkvist[4], P Eriksson[3], J Cedergren[3], K Lindblad-Toh[2,5], JRS Meadows[2]

[1]Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
[2]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
[3]Department of Rheumatology, and Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
[4]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
[5]Broad Institute of MIT and Harvard, Cambridge, MA, USA

Objective: The genetic predisposition to ankylosing spondylitis (AS) has been most widely studied in cohorts with European ancestry. However, within Europe, disease prevalence is higher in Sweden. Given this, we aimed to characterize known AS susceptibility variants in a homogeneous Swedish data set, assessing reproducibility and direction of effect.

Method: The power to detect association within an existing Swedish targeted sequencing study (381 controls; 310 AS cases) was examined, and a set of published associations (n = 151) was intersected with available genotypes. Association to disease was calculated using logistic regression accounting for population structure, and *HLA-B27* status was determined with direct polymerase chain reaction genotyping.

Results: The cases were found to be 92.3% *HLA-B27* positive, with the data set showing ≥ 80% predictive power to replicate associations, with odds ratios ≥ 1.6 over a range of allele frequencies (0.1–0.7). Thirty-four markers, representing 23 gene loci, were available for investigation. The replicated variants tagged *MICA* and *IL23R* loci ($p < 1.47 \times 10^{-3}$), with variable direction of effect noted for gene loci *IL1R1* and *MST1*.

Conclusion: The Swedish data set successfully replicated both major histocompatibility complex (MHC) and non-MHC loci, and revealed a different replication pattern compared to discovery data sets. This was possibly due to population demographics, including *HLA-B27* frequency and measured comorbidities.

Ankylosing spondylitis (AS) is a highly heritable chronic inflammatory disease of the axial skeleton and sacroiliac joints, with unresolved aetiology and no cure. Apart from the unprovoked and painful axial inflammation, AS is often accompanied by a combination of comorbidities, and is more prevalent in men than women (3:1 ratio) (1). Despite symptomatic treatment, advanced disease is characterized by new and irreversible bone formation at the inflamed sites.

Much of the current insight into the genetic predisposition to AS comes from large European ancestry-based genome-wide or tailored single-nucleotide polymorphism (SNP) scans, utilizing anywhere from 15 000 to over a million markers, and many thousands of cases and matched controls [e.g. (2–4)]. These studies and others have revealed more than 140 disease associate loci, which together explain in excess of 30% of the genetic heritability (3). The strongest association to disease comes from the human leucocyte antigen-B27 (*HLA-B27*) locus. This locus signal is consistently replicated across studies, and in concordance with this, AS is often more frequent in populations with high *HLA-B27* occurrence. For populations with European ancestry, the general frequency of *HLA-B27* is 6–9% (5) and the standardized AS prevalence is 0.55% (6); Sweden is enriched for *HLA-B27* frequency but reduced for AS prevalence [10–12% *HLA-B27* (7), 0.18% prevalence (8)].

In 2012, a targeted liquid capture, the ImmunoArray, was designed to cover a set of over 1800 genes and their regulatory regions. Selection was based on their involvement in known immune pathways and processes (9). This array was subsequently used to sequence multiple immunological disease sets, including AS. The goal of the current study was to assess the Swedish AS set's ability to replicate European ancestry-derived AS association signals, and by doing so, to place Swedish

JRS Meadows, Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, 75123 Uppsala, Sweden.
E-mail: jennifer.meadows@imbim.uu.se

disease and predisposition patterns in a broader population context.

Method

Sample collection and *HLA-B27* typing

Genotyped *HLA-B27* carrier status (10), positive or negative, was used to select mainly *HLA-B27*-positive cases for inclusion in a set of 691 individuals with self-reported Swedish ancestry, comprising 381 healthy controls (mean ± sd age 63.2 ± 15.7 years; 66% male) and 310 AS cases diagnosed according to the New York criteria (mean age 51.9 ± 11.9 years; 73% male). Overlapping comorbidities were diagnosed in 215 cases: uveitis (n = 114), peripheral joint inflammation (n = 133), psoriasis (n = 23), and gut involvement (n = 40). All participants gave written informed consent, with ethical approval granted from the regional committee of Linköping and in accordance with the Declaration of Helsinki (Dnr 2010/182-32; 98110).

Literature mining for published associated variants

SNPs were considered for replication if they met the following criteria: (i) they were reported in peer-reviewed journal articles with a publication date before 2018; (ii) the publication consisted of more than 2000 markers in population cohorts of British or Han Chinese origin; and (iii) variants had p-value < 1.0×10^{-5} in the discovery set (Supplementary Table S1). Information was retrieved from the main text and supplementary texts. All genomic coordinates were reported in hg19.

Genetic analysis

The Swedish AS genotypes available for comparison were generated through the Illumina resequencing of a custom ImmunoArray (data, doi:10.17044/scilifelab.13027256). In brief, the ImmunoArray targeted the coding and non-coding regions [3′- and 5′-untranslated regions (UTRs), splice sites, promoters] of more than 1800 genes with known involvement in immune pathways (9). PLINK v1.9 (11) was used to extract and prune the variants common to both the Swedish and literature sets (call rate > 85% and minor allele frequency > 0.03).

Logistic regression with 10 eigenvectors [PLINK v1.9 (11)] was used to assess associations, with a Bonferroni significance threshold (i.e. 0.05/markers common to both data sets; Supplementary Table S2). Linkage disequilibrium (LD; r^2) between markers on the same chromosome was also calculated [PLINK v1.9 (11); Supplementary Table S3].

The power to detect associations in the Swedish data set was set at 80%, and the range of detectable odds ratios (ORs) was calculated for 310 cases and 381 controls, allele frequency 0.1–0.7, with an additive model and Bonferroni

significance threshold (1.47×10^{-3}; genpwr package in R v.3.5.0).

The full method is given in the supplementary material.

Results

The intersection of the literature set and the Swedish data set revealed that 23% of the markers (34/151) were shared. These tagged 23 unique gene loci, including *MICA, ERAP1, ERAP2, IL23R*, and *FCGR2A* (Supplementary Table S2). The ImmunoArray (9) target space influenced marker availability, with 74% of the literature set located in untargeted and so unavailable regions, e.g. intronic *ERAP2* variant rs2910686 and gene desert variants rs2310173 in 2q11.2 and rs6556416 in 5q33.3. Within the literature set, 26 gene loci were represented by multiple variants (Supplementary Table S1), and so the unavailability of some markers was not equivalent to loss of the locus, e.g. *IL23R, ILR1*, and *ERAP1* (Supplementary Table S2). Eight variants were present in the target space but failed filtering quality control thresholds including deviation from Hardy–Weinberg equilibrium, e.g. markers tagging gene loci *HLA-B* and *FCGR2A*.

The majority of the 34 markers available conferred small risk effect sizes (21 SNPs, OR = 1.09–1.57) over a range of risk allele frequencies (RAF = 0.08–0.77) (Supplementary Table S2). At 80% power, the Swedish data set would be able to identify SNPs with an OR ≥ 1.6 over an RAF range of 0.1–0.7 (Supplementary Figure S1). This equates to the known effect sizes of *IL23R* or major histocompatibility complex (MHC) loci (Supplementary Table S2). In practice, for the Swedish data set (381 controls, 13.0% *HLA-B27* positive; 310 AS cases, 92.3% *HLA-B27* positive), three markers were found to be significantly associated with AS (risk: *MICA* rs9266825 and rs1051792; protective: *IL23R* rs11209026; p-value < 1.47×10^{-3}, Table 1; LD matrix, Supplementary Table S3). This pattern matched the expectation for ORs and power.

We further examined the associated alleles that did not replicate, and noted that when comparing allele frequencies between Swedish and published case or control frequencies, all controls, and most cases (81%), were within a 5% frequency span (Supplementary Table S2). The largest allele frequency difference was noted for both *MICA* case variants (> 12%). For three variants (representing genes *IL1R1* and *MST1*; Figure 1, Supplementary Figure S2), the direction of SNP effect was reversed (measured as OR), with minor, if any, overlap in 95% confidence intervals. Here, the differences in case allele frequency differences were slight (1–4%).

Discussion

Through the replication of both MHC (*MICA*) and non-MHC loci (*IL23R*), and the dissection of disease-associated allele frequencies, the Swedish AS replication

Table 1. Published associations replicated in the Swedish set.

Locus	Marker	Allele (R/NR)*	p†	OR (95% CI)‡
IL23R	rs11209026	A/G	9.7×10^{-4}	0.4 (0.2–0.7)
MICA	rs1051795	G/A	8.2×10^{-32}	12.7 (8.3–19.4)
	rs9266825	C/A	3.1×10^{-32}	12.6 (8.3–19.2)

*Risk (R) and non-risk (NR) allele.
†Replication p-value.
‡Odds ratio (OR) and 95% confidence interval (CI) for the risk allele.

study placed the genetic predisposition of this data set in the context of previously published populations of European origin. This analysis illustrated that while a smaller data set may have reduced power to detect published variants, it may still have the power to reveal disease association differences between different populations.

The replication success of this study was driven by sample size, the number of available markers, and the contribution of population demographics. The ImmunoArray was designed in 2012 (9) based on published immune-related loci for Sjögren's syndrome, systemic lupus erythematosus, Addison's disease, anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis, myositis, and AS, as well as their extended associated gene pathways. For that reason, markers unique to AS identified after that date were unlikely to be covered by the design. Another issue was replication sample size. In selecting predominantly *HLA-B27*-positive AS cases (92.3%), the goal was to increase power by reducing background heterogeneity. This may yet prove successful for the identification of novel alleles given the inclusion of additional control individuals, such as those from the 1000 genomes of Sweden (12).

The *IL23R* and the *ERAP1* loci were among the first non-MHC risk loci discovered for AS (2). While the former locus was replicated here, the latter was not, even though the extended *ERAP1-ERAP2* block was
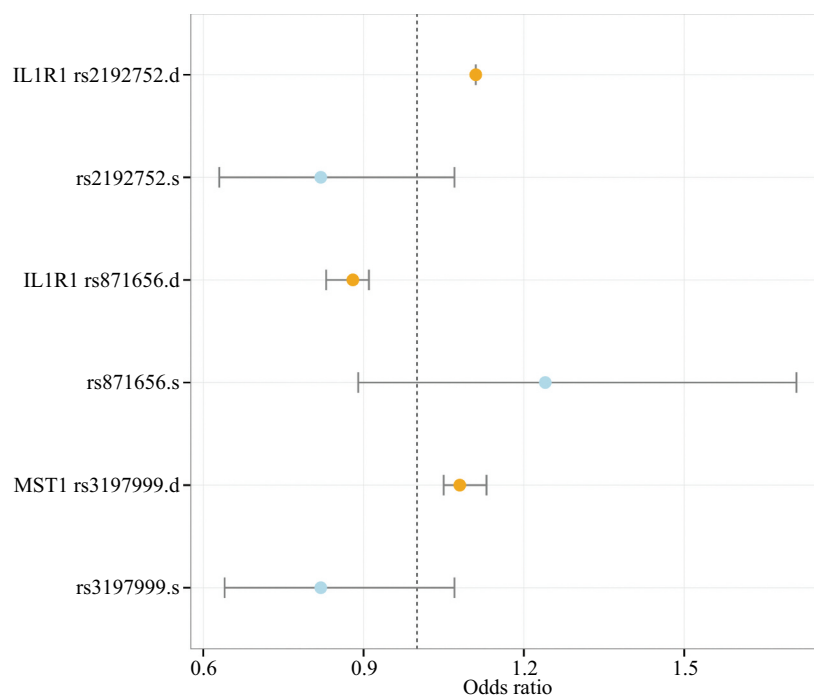

Figure 1. Discordant allele effect sizes (odds ratio and 95% confidence interval) between the published discovery (rsID.d) and the Swedish AS (rsID.s) data sets.

represented by nine variants (Supplementary Table S2). This locus is generally well replicated in data sets of sufficient power (e.g. rs30187 = 3.4×10^{-10}–1.3×10^{-41}; OR across studies ~1.10–1.30; Supplementary Table S2). The protein change conferred by rs30187-T, p.Lys528Arg, has been posited as a loss of function variant, slowing the aminopeptidase activity of this protein [13]. A 2016 publication noted that the missense variant rs30187 had a larger effect in a set of AS cases also with acute anterior uveitis (OR = 1.46, 95% CI 1.35–1.58), in contrast to AS cases lacking this comorbidity [14]. The Swedish data set was not temporal, and given that such single comorbidities were infrequent (total cases, n = 310; AS only, n = 95; AS plus uveitis, n = 53), and that the frequency of uveitis increases with AS duration [15], a similar analysis was not possible. However, we noted that the direction of effect was similar between the published and Swedish data sets (Supplementary Table S2), but that the locus p-value ranking revealed that the 3′-end of *ERAP1* (rs17482078 and rs10050860) was more associated with AS. In cases such as these, the value of the Swedish cohort may lie in the population enrichment of, as yet, unidentified alleles and the downstream dissection of haplotypes. This could indicate which part of the protein or regulatory locus influences disease in a given population background.

Conclusion

The Swedish data set described here had the power to replicate loci both within and outside the MHC region. An examination of allele frequencies indicated the potential of differential genetic architecture between European populations, and suggests that novel disease-associated variants may be found in additional data sets with varied demographics and *HLA-B27* profiles.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplementary material

Supplemental data for this article can be accessed here.

References

1. Lee W, Reveille JD, Davis JC, Learch TJ, Ward MM, Weisman MH. Are there gender differences in severity of ankylosing spondylitis? Results from the PSOAS cohort. Ann Rheum Dis 2007;66:633–8.
2. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, Craddock N. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007;39:1329–37.
3. International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet 2013;45:730–8.
4. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 2016;48:510–8.
5. Khan MA. HLA-B27 and its subtypes in world populations. Curr Opin Rheumatol 1995;7:263–9.
6. Braun J, Listing J, Sieper J. Reply. Arthritis Rheum 2005;52:4049–50.
7. Bjelle A, Cedergren B, Dahlqvist SR. HLA B27 in the population of northern Sweden. Scand J Rheumatol 1982;11:23–6.
8. Exarchou S, Lindström U, Askling J, Eriksson JK, Forsblad-d'Elia H, Neovius M, et al. The prevalence of clinically diagnosed ankylosing spondylitis and its clinical manifestations: a nationwide register study. Arthritis Res Ther 2015;17:118.
9. Eriksson D, Bianchi M, Landegren N, Nordin J, Dalin F, Mathioudaki A, et al. Extended exome sequencing identifies BACH2 as a novel major risk locus for Addison's disease. J Intern Med 2016;280:595–608.
10. Bon MAM, van Oeveren-Dybicz A, van den Bergh FAJTM. Genotyping of HLA-B27 by real-time PCR without hybridization probes. Clin Chem 2000;46:1000–2.
11. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:559.
12. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet 2017;25:1253–60.
13. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, 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.
14. Robinson PC, Leo PJ, Pointon JJ, Harris J, Cremin K, Bradbury LA, et al. The genetic associations of acute anterior uveitis and their overlap with the genetics of ankylosing spondylitis. Genes Immun 2016;17:46–51.
15. Robinson PC, Robinson PC, Robinson PC, Brown MA, Brown MA, Brown MA. Genetics of ankylosing spondylitis. Mol Immunol 2014;57:2–11.