Sex Differences in the Genetics of Sarcoidosis Across European and African Ancestry Populations

Ying Xiong

Karolinska Institutet

Susanna Kullberg

Karolinska Institutet

Lori Garman

Oklahoma Medical Research Foundation

Nathan Pezant

Oklahoma Medical Research Foundation

David Ellinghaus

Christian Albrechts Universität zu Kiel: Christian-Albrechts-Universität zu Kiel

Vasiliki Vasila

Karolinska Institutet

Anders Eklund

Karolinska Institutet

Benjamin Rybicki

Henry Ford Health System

Michael Iannuzzi

Staten Island University Hospital

Stefan Schreiber

Christian-Albrechts-Universität zu Kiel

Joachim Müller-Quernheim

University of Freiburg Hospital: Universitätsklinikum Freiburg

Courtney Montgomery

Oklahoma Medical Research Foundation

Johan Grunewald

Karolinska Institutet

Leonid Padyukov

Karolinska Institutet

Natalia V. Rivera (✉ natalia.rivera@ki.se)

Karolinska Institutet <https://orcid.org/0000-0002-1298-1243>

Research

Keywords: sarcoidosis, genome-wide association study, meta-analysis, genetic variants, single nucleotide polymorphisms, autoimmune disease

Posted Date: April 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-382787/v1>

Abstract

Background: Sex differences in the susceptibility of sarcoidosis are unknown. The study aims to identify sex-dependent genetic variations in two sarcoidosis clinical phenotypes: Löfgren's syndrome (LS) and non- Löfgren's syndrome (non-LS).

Methods: A meta-analysis of genome-wide association studies was conducted in Europeans and African Americans, totaling 10,103 individuals from three population-based cohorts, Sweden (n = 3,843), Germany (n = 3,342), and the United States (n = 2,918), followed by replication look-up in the UK Biobank (n = 387,945). A genome-wide association study based on ImmunoChip data consisting of 141,000 single nucleotide polymorphisms (SNPs) was conducted in males and females in each cohort, respectively. The association test was based on logistic regression using the additive model in LS and non-LS independently. Additionally, gene-based analysis, expression quantitative trait loci (eQTL) assessments, and enrichment analysis were performed to discover functionally relevant mechanisms related to biological sex.

Results: In LS sarcoidosis, we identified various sex-dependent genetic variations (798 SNPs in males and 703 SNPs in females). Genetic findings in sex groups were explicitly located in the extended major histocompatibility complex. In non-LS, we detected 16 SNPs in males and 38 in females, primarily localized to the MHC class II region. Additionally, the *ANXA11* gene, a well-documented locus in sarcoidosis, was associated exclusively with non-LS males. Gene-based, eQTL assessment and enrichment analyses revealed distinct sex-dependent genomic loci and gene expression variation in the sex groups.

Conclusions: Our findings provide new evidence of the existence of sex-dependent genetic variations underlying sarcoidosis genetic architecture. These findings suggest a sex bias in molecular mechanisms of sarcoidosis.

Background

Sarcoidosis is a systemic disease with unknown etiology, characterized by non-caseating granulomas[1]. In ninety percent of patients, sarcoidosis affects the lungs and lymphatic system[2]. However, sarcoidosis patients experience heterogeneous manifestations and clinical outcomes, including affected organs, that appear to be associated with biological sex[3, 4]. Additionally, the disease course seems to be positively modulated during pregnancy, resulting in low disease activity and even remission. Still, this ameliorating effect is lost, as postpartum flares occur after delivery[5, 6].

Sarcoidosis is characterized by two clinical phenotypes, Löfgren's syndrome (LS) and non-Löfgren's syndrome (LS). LS is an acute form of the disease characterized by periarticular swelling around the ankles, erythema nodosum and/or bilateral periarticular swelling around the ankles[7]. Typically, females with LS often exhibit erythema nodosum, while males with LS have symptoms of bilateral ankle arthritis. Patients with LS usually have a good prognosis[8]. LS often occurs in individuals of European ancestry and is relatively rare in individuals of African ancestry. In non-LS, patients tend to have a heterogeneous disease course, often characterized by an insidious onset, unrelenting disease course, and a high risk for clinically organ impairment, predominantly pulmonary fibrosis[7, 9]. Patients with fibrotic sarcoidosis have markedly decreased survival compared with the general population[10, 11].

On a global scale, sarcoidosis incidence and prevalence vary based on sex, age, geography, and ethnicity[12]. Sarcoidosis is prevalent worldwide, commonly affecting individuals aged 40–55 years, with males diagnosed earlier in age than females[13].

To date, the underpinnings for clinical differences and disease course in sarcoidosis between males and females are unknown. However, evidence-based factors, including genetics, epigenetics, and environmental exposures, have been

suggested[14–17]. Few studies have focused on the sex differences in genetic variation of sarcoidosis, despite most studies adjusting for sex in their analyses[18]. An exception is a genome-wide admixture scan conducted in African Americans stratified by sex that identified several genetic variants associated with sarcoidosis exclusively in females[19].

Nevertheless, no genetic investigation has been conducted to examine sex differences in genetic variations of sarcoidosis in European and African ancestry populations together. Therefore, in this study, we set out to delineate the genetic architecture of sarcoidosis in males and females by conducting a meta-analysis of genome-wide association studies based on Immunochip data in European and African Americans using three population-based cohorts from Sweden, Germany, and the United States. Moreover, given the availability of genome-wide data sets in other populations of the same ancestry, a replication analysis was performed using GWAS summary statistics on sarcoidosis conducted in the UK Biobank, consisting of 500,000 individuals[20].

Methods

Samples and study design

The samples were from three independent population-based cohorts consisting of sarcoidosis cases and controls from Sweden, Germany, and the United States. Study protocols of all studies had been approved by respective local institutional boards. All participants provided written informed consent for the study. Further details of the three cohorts are described in previous studies[18, 21, 22].

Genotyping and quality control

Genotyping for sarcoidosis patients in the Swedish cohort was performed using the Illumina Immunochip platform. Further details are described in [18]. Genotypes for the German cohort were quality controlled, as described in [21]. Genotyping for the African American cohort (US-AA) was performed at the OMRF using the Illumina HumanOmni1-Quad array for ~1.1M variants across the genome. Of these, 121,988 were in common with the Immunochip platform and carried forward for analysis. Further details on genotyping and quality control filtering are described in [22].

Statistical Analysis

Association analysis of LS and non-LS in sex groups

In each sex group, an additive model using logistic regression was applied to examine the association between single nucleotide polymorphisms (SNPs) and sarcoidosis, mainly LS and non-LS, independently. To address the genetic effects, we implemented three models with different strategies for adjusting confounding factors. Model 1 was a crude model with no covariate adjustment. Model 2 was based on the genetic effects adjusted for age. Model 3 was based on the genetic effects adjusted for age and the first four principal components (PCs). PCs were derived from principal component analysis (PCA) that was performed using EIGENSTRAT[23] software on a pruned genotyped data set. Pruning of genotypes was performed using the pruning function as implemented in PLINK version 1.07 software. For each cohort, the first 4 PCs were used, which explained at least 50% of the genetic variation. An association test was performed in PLINK across all models. P-value significance was defined by the genomic control P-value (P_{GC}) at $<5 \times 10^{-8}$. A suggestive p-value of $5 \times 10^{-5} < P_{GC} < 5 \times 10^{-8}$ was applied to identify potential associations[24-27].

Meta-analysis of genome-wide association studies

The meta-analysis of GWAS was conducted in sarcoidosis LS and non-LS independently for each sex group. In order to analyze the genetic effects across the three population-based cohorts (Swedish, German, and US African Americans), genetic variants were detected via the inverse variance weighting (IVW) method. This method calculates effect size estimates and standard error as implemented in the METAL software[28]. A primary meta-analysis was performed based on the inclusion of SNPs with $P_{GC} < 0.05$ across all three cohorts. Additionally, a secondary GWAS meta-analysis was performed with SNPs with $P_{GC} < 0.05$ in the Swedish cohort and $0.05 < P_{GC} < 0.1$ in the German cohort and US-AA cohorts.

In the LS sex groups, the meta-analysis was based on pooling GWAS associations of the Swedish and German cohorts. The US-AA cohort was excluded from the analysis because LS is rare in African Americans. In the non-LS sex groups, the meta-analysis was conducted in a 2-step process. First, we pooled together the GWAS results from the Swedish and German cohorts and performed a meta-analysis. Second, we pooled together the results from the first step and analyzed them together with the GWAS results of the US-AA cohort. Of note, we considered the summary statistics of Model 1 from all three cohorts as an input for the meta-analysis. The combined genetic effect size (defined by Beta), standard error (SE), meta P-value (P_{meta}), total variability in effect size, also known as the heterogeneity index (I^2), and heterogeneity P-value were calculated. SNPs with $P_{meta} < 5 \times 10^{-8}$ were defined as genome-wide significant. Genetic variants with $P_{meta} < 5 \times 10^{-5}$ were considered to be suggestive associations.

Replication look-up of meta-GWAS findings in the UK Biobank

A replication look-up was conducted using GWAS summary statistics on sarcoidosis conducted in males and females, separately, in the UK Biobank[20]. The UK Biobank (<https://www.ukbiobank.ac.uk/>) is an excellent health resource that can be used to study risk factors of diseases since it contains data of various diseases and risk factors collected from 500,000 individuals aged 40-69 years who were recruited 2006-2010[29]. Briefly, a genome-wide association study was independently conducted on self-reported sarcoidosis in 194,153 females (361 cases and 193,792 controls) and 166,988 males (344 cases and 166,644 controls) of European ancestry, respectively. A linear regression model was applied on typed and imputed genotypes totaling 13.7 million quality-controlled variants and adjusted by covariates, including the first 20 PCs, age, and age².

Of note, given the scarcity of further phenotypic characterization of sarcoidosis (LS and non-LS) in the UK Biobank, we pooled together LS and non-LS associated genetic variants with $P_{meta} < 5 \times 10^{-5}$ from each meta-analysis GWAS and extracted these variants from the UK Biobank GWAS summary statistics.

Further details on phenotype characterization, imputation, and GWAS analysis are available at https://github.com/Nealelab/UK_Biobank_GWAS. The significance threshold for replication was set at a nominal $P < 0.05$, with Bonferroni correction; that is, nominal p-value divided by the pooled number of genetic variants considered.

Gene-based analysis

To identify genomic loci of interest, a gene-based analysis using the Versatile Gene-based Association Study (VEGAS) software version 2[30] was conducted using the GWAS meta-analysis results as the primary input. SNPs with $P_{meta} < 5 \times 10^{-8}$ and $< 5 \times 10^{-5}$ were selected for the analysis. The significance level was defined by gene-based $P < 2 \times 10^{-6}$.

Expression quantitative trait loci (eQTLs) analysis

To explore the biological function of genetic variants identified in the GWAS meta-analysis with $P_{meta} < 5 \times 10^{-8}$, cross-tissue eQTL enrichment analysis including immune cells, lung tissue, and whole-blood was conducted using EUGENE[31,

32] software. An SNP associated with an eQTL trait was defined at $P < 0.05$.

Enrichment and pathway analysis

In order to elucidate molecular mechanisms, enrichment and pathway analysis were conducted based on SNPs with $P_{\text{meta}} < 5 \times 10^{-8}$ and using MetaCore™ software. A secondary analysis using GWAS results of Model 3 in the Swedish cohort was conducted by including SNPs with $P_{\text{GC}} < 5 \times 10^{-5}$.

Results

Association results

The genome-wide association analysis was conducted in males and females, independently, in each cohort, i.e., Swedish (Supplementary Tables S1-S4), German (Supplementary Table S5-S8), and US-African American (Supplementary Tables S9-S10). The association results for these cohorts and supplementary Tables S1-S10 are reported in Additional File 1. The details of the discovery cohorts as well as the replication cohort in the UK Biobank, are shown in Table 1 A and B, respectively.

Meta-analysis of genome-wide association studies

Meta-GWAS on LS sex groups. The meta-analysis on LS sex groups performed by pooling together GWAS results from the Swedish and German cohorts revealed 766 SNPs in males and 683 SNPs in females with $P_{\text{meta}} < 5 \times 10^{-8}$. The lead SNPs were rs3130288 (OR=4.83, 95%CI=3.65-6.38, $P_{\text{meta}} = 1.80 \times 10^{-28}$) located in the 5' of *ATF6B* in LS males, and rs9268219 (OR=3.69, 95%CI=2.87-4.74, $P_{\text{meta}} = 2.04 \times 10^{-24}$) located in an intron of *C6orf10* in LS females. Several SNPs located outside the MHC and with $P_{\text{meta}} < 5 \times 10^{-5}$ were also identified, including rs116711574 (OR=0.24, 95%CI: 0.13-0.47, $P_{\text{meta}} = 2.06 \times 10^{-5}$) in an intron of *HMP19* on chromosome 5 in LS males, and rs17713228 (OR=0.29, 95%CI: 0.17-0.48, $P_{\text{meta}} = 2.21 \times 10^{-6}$) located 68kb 3' of *ZNF804A* on chromosome 2 in LS females. Regional association plots for the lead meta-GWAS SNPs are illustrated in Figure 1 A-B. The forest plots highlighting the overall genetic effects in the LS sex groups are shown in Figure 2 A-B. The top 25 significant meta-GWAS SNPs in males and females are listed in Tables 2-3, respectively. The meta-GWAS results for LS males and females at $P_{\text{meta}} < 5 \times 10^{-8}$ are provided in Supplementary Tables S11-S12 in Additional File 2. Manhattan plots are shown in Supplementary Figure S9 A-B. Q-Q plots are shown in Supplementary Figure S9 C-D.

Results from the secondary meta-GWAS analysis including SNPs with $P_{\text{GC}} < 0.05$ in the Swedish cohort and SNPs with $0.05 < P_{\text{GC}} < 0.1$ in the German cohort identified 32 SNPs in LS males and 20 SNPs in LS females with $P_{\text{meta}} < 5 \times 10^{-8}$. The leading SNPs were rs3129763 (OR=2.05, 95%CI: 3.32-2.61, $P_{\text{meta}} = 6.74 \times 10^{-15}$) located 14kb 5' of *HLA-DQA1* in LS males, and rs2517985 (OR=0.48, 95%CI:0.38-0.60, $P_{\text{meta}} = 5.12 \times 10^{-11}$) located in an intron of *CCHCR1* in LS females. Results of secondary meta-GWAS for LS males and females are shown in Supplementary Tables S13-S14 in Additional File 2.

Meta-GWAS on non-LS sex groups. The meta-analysis on non-LS sex groups in the European cohorts (Sweden and Germany) identified 67 SNPs in males and 97 SNPs in females with $P_{\text{meta}} < 5 \times 10^{-8}$. The most significant SNP was rs1964995, located 36kb from the 3' of *HLA-DRB5*, and was the top signal for both sexes (OR=0.57, 95%CI:0.50-0.64, $P_{\text{meta}} = 3.92 \times 10^{-18}$ in males, and OR=0.58, 95%CI:0.51-0.65, $P_{\text{meta}} = 5.00 \times 10^{-20}$ in females). Complete results of the meta-GWAS analysis in the European cohorts for non-LS males and females are provided in Supplementary Tables S15-S16 in Additional File 3.

The meta-GWAS in the three cohorts comprising European and African ancestries identified 14 SNPs in males and 36 SNPs in females with $P_{meta} < 5 \times 10^{-8}$. The top signal was rs1964995 and was the same for both sexes (OR=0.56, 95%CI:0.50-0.63, $P_{meta}=1.46 \times 10^{-22}$ in males, OR=0.66, 95%CI:0.60-0.72, $P_{meta}=2.86 \times 10^{-19}$ in females). However, in non-LS females, rs1964995 had a high heterogeneity index with $P < 0.05$, so the next top signal was rs7195 (OR=1.44, 95%CI:1.33-1.57, $P_{meta}=7.64 \times 10^{-18}$) located in an intron of *HLA-DRA*. The top 25 significant meta-GWAS SNPs are listed in Tables 4-5, respectively. Forest plots highlighting the overall genetic effects in the non-LS sex groups are shown in Figure 3 A-D. Regional association plots for the lead SNPs are illustrated in Figure 4 A-B. The complete lists of meta-GWAS results in the non-LS sex groups are provided in Supplementary Tables S17-S18 in Additional File 3. Manhattan plots are shown in Supplementary Figure S10 A-B, respectively. Q-Q plots are shown in Supplementary Figure S10 C-D.

The secondary meta-analysis consisting of SNPs with $P_{GC} < 0.05$ in the Swedish and German cohorts and SNPs with $0.05 < P_{GC} < 0.1$ in the US-AA cohort revealed two genome-wide SNPs in non-LS males and females, respectively. Chiefly, in non-LS males rs2395153 (OR=0.69, 95%CI:0.61-0.77, $P_{meta}=4.77 \times 10^{-10}$) located 5.9kb from the 5' of *C6orf10* and rs1049550 (OR=0.67, 95%CI: 0.60-0.76, $P_{meta}=1.98 \times 10^{-11}$) located in *ANXA11* were found. In non-LS females, rs9275224 (OR=1.29, 95%CI:1.19-1.40, $P_{meta}=1.46 \times 10^{-9}$) located 25kb from the 5' of *HLA-DQB1* and rs4502931 (OR=0.79, 95%CI:0.73-0.86, $P_{meta}=3.29 \times 10^{-8}$) located 5.9kb from the 5' of *BTNL2* were observed. The complete lists of the secondary GWAS-meta results are shown in Supplementary Tables S19-S20 in Additional File 3.

Replication look-up of meta-GWAS SNPs in the UK Biobank

Given the lack of further characterization (i.e., LS and non-LS) on the sarcoidosis phenotype in the UK Biobank, we focus on the sex groups for the replication look-up. In the male group, 1,253 SNPs with $P_{meta} < 5 \times 10^{-5}$ in LS and non-LS were pooled together. Out of this set, 1,224 (97.6%) unique SNPs were identified and extracted from the UK Biobank using the GWAS summary statistics conducted in males. The lookup identified 723 SNPs with Bonferroni p-value of $< 4.0 \times 10^{-5}$ (0.05/1,224). Moreover, after excluding SNPs with heterogeneity p-value < 0.05 , 715 SNPs were kept (Supplementary Tables S21 in Additional File 4).

In the female group, 1,356 SNPs with $P_{meta} < 5 \times 10^{-5}$ were pooled together from the GWAS meta-analyses of LS and non-LS. Out of this set, 1,316 (97%) unique SNPs were identified and extracted from the UK Biobank GWAS summary statistics conducted in females. The lookup identified 454 SNPs with Bonferroni p-value of $< 3.79 \times 10^{-5}$ (0.05/1,316). Of note, since 9 SNPs had a heterogeneity p-value < 0.05 , these SNPs were excluded from the replication set, 445 SNPs were kept (Supplementary Table S22 in Additional File 4).

Gene-based analysis

Results from the gene-based analysis revealed different genomic loci in LS and non-LS sex groups. Mainly, 101 genes in LS males and 93 genes in LS females were identified at gene-based P-value $< 2 \times 10^{-6}$. Genes shared (n=91) among LS males and females were located in the MHC region. These genes ranged from *ZNF184* to *PSMB9* (chr6: 27371788-32827628). The top 15 significant genes in the LS sex groups are shown in Tables 6-7. Complete results are shown in Supplementary Tables S23-S24 in Additional File 5.

In the non-LS sex groups, three genes (i.e., *HLA-DRA* and *STK19* and *ANXA11*) in non-LS males and five genes (i.e., *BTNL2*, *C6orf10*, *HLA-DRA*, *PPT2-EGFL8*, and *PPT2* in non-LS females were identified. Significant genes in the non-LS sex groups are shown in Tables 8–9. A Venn diagram illustrating significant loci between LS and non-LS sex groups is depicted in Supplementary Figure S11.

Expression quantitative trait loci (eQTLs) enrichment

eQTL enrichment using eQTL data of immune cells, whole blood, and lung on SNPs associated with LS and non-LS sex groups at $P_{\text{meta}} < 5 \times 10^{-8}$ identified several *cis*-eQTL SNPs (Supplementary Table S25 in Additional File 6). Specifically, the number of SNPs mapping immune cells, whole blood, and lung eQTL data were (40, 88, and 24) in LS males, (36, 79, and 19) in LS females, (4, 7, and 2) in non-LS males and (5, 1, and 3) in non-LS females. The lists of eQTL SNPs in LS and non-LS sex groups are shown in Supplementary Tables S26-S29 in Additional File 6.

A Venn diagram highlighting overlaps and differences in *cis*-eQTL SNPs in LS sex groups is shown in Figure 5 A-C. eQTL datasets are listed in Appendix 1 in the Supplementary Materials.

Gene enrichment and pathway analysis

In the LS, we identified significant pathway maps common to both sex groups and those specific to one group. Two pathways maps of interest were (1) immune response on the induction of the antigen presentation machinery by IFN-gamma (males: $P = 2.39 \times 10^{-12}$, 1.16×10^{-14} , 2.04×10^{-14} ; females: $P = 1.38 \times 10^{-10}$, 1.80×10^{-13} , 5.88×10^{-13}) and (2) maturation and migration of dendritic cells in skin sensitization (males: $P = 2.64 \times 10^{-13}$, 1.31×10^{-11} , 4.51×10^{-10} ; females: $P = 2.06 \times 10^{-13}$, 2.77×10^{-12} , 2.54×10^{-10}) (Supplementary Tables S30-S31 in Additional File 7).

In non-LS, pathways maps were only detected in whole blood eQTLs in males and in immune cells in females. Among these, pathway maps related to (1) immune escape mechanisms in prostate cancer ($P = 6.27 \times 10^{-3}$) in males and (2) immune response induction of the antigen presentation machinery by IFN-gamma ($P = 2.39 \times 10^{-8}$) in females were identified (Supplementary Tables S32-S33 in Additional File 7).

Moreover, a cross-comparison analysis showed various common and sex-specific pathway maps in LS. Results of the comparative analysis are shown in Supplementary Tables S34-S35 in Additional File 7.

Discussion

In the present study, we focused on investigating sex differences in sarcoidosis phenotypes Löfgren's syndrome (LS) and non-Löfgren's syndrome (non-LS). This work is the first investigation for characterizing the genetic architecture of sarcoidosis stratified by biological sex.

In brief, we analyzed genetic associations on LS and non-LS in sex groups conducted in three population-based cohorts, i.e., Sweden, Germany, and the United States. The cohorts included individuals of different ethnicities, i.e., white European and black African ancestry. The genetic analysis was based on a genome-wide association study (GWAS) approach in individual cohorts followed by a GWAS meta-analysis and a replication look-up in the UK Biobank.

The GWAS meta-analysis in LS and non-LS sex groups highlighted sex-specific genetic variations in white European and black African sample populations.

In LS, we identified distinct genetic associations clustering in the extended MHC (xMHC)[33] in both sex groups. Such finding highlights the relevance of the xMHC in the LS's susceptibility, as we previously[18]. Nevertheless, this study refines our previous work by pinpointing sex-dependent alleles and their significant differences in beta-estimates for the SNPs associated. In the non-LS, genetic associations in the sex groups were wide-ranging within the MHC class II subregion and other chromosomes. Notably, the genetic architecture of non-LS appears to be more complex and likely to have numerous genetic associations across the genome. Genetic findings were also replicated in the UK Biobank and self-reported sarcoidosis data.

Strikingly, across all the cohorts, we observed differences in association patterns among male and female groups, suggesting sex-bias in the genetic architecture of sarcoidosis. These findings are also supported by the observed sex differences in gene expression variation marked by eQTL signals in tissues and cells related to sarcoidosis. However, the functional mechanisms and the genetic effects on sex-dependent gene expression and disease risk remain to be elucidated.

In the LS sex groups, out of the total genetic associations found across the three genetic models, 322 variants were consistent with previous findings[18]. Nevertheless, old and new genetic associations were explicitly identified in males and females, providing further insights into LS's genetic architecture. Principally, *ATF6B* in LS males and *C6orf10* in LS females shall be further investigated.

In non-LS, results showed significant sex-specific associations located in the MHC classical class III and II subregions and the *ANXA11* locus. Such findings were substantiated with findings in the gene-based, eQTL analysis, and pathway analysis. Further studies in *HLA-DRB* are crucial for determining the genetic architecture of non-LS as this locus is often associated with autoimmune disorders and holds many clues regarding sex-dependent risk alleles. It is well-documented that autoimmunity has a profound sex bias, as autoimmune diseases predominantly affect females; however, the mechanisms underlying sex selection in autoimmunity are not entirely understood[34]. Interestingly, the *ANXA11* gene, a well-established for sarcoidosis susceptibility and initially identified by Hofmann et al.[35], was associated with non-LS males. The *STK19* gene in non-LS males is also worth looking into as it encodes a serine/threonine kinase and plays a role in transcriptional regulation[36]. In non-LS females, *C6orf10*, *HLA-DRA*, and *PPT2* are significant loci that require further investigation to explain their functional role in females affected by sarcoidosis. Chiefly, *PPT2* is noteworthy as its product catalyzes the hydrolysis of long-chain thioester-linked fatty acyl groups[37], which may be implicated in hormonal processes.

Our findings provide new insights into the genetic architecture of sarcoidosis. Notably, the differences observed in the genetic susceptibility of sex groups in LS and non-LS may likely explain the well-known correlations between sex and the incidence, prevalence, age of onset, symptoms, severity of the disease, as well as drug reactions[18, 38–41].

Future genetic and omics studies shall be conducted to further characterize the genetic architecture of sarcoidosis in the sex groups, bearing in mind the inclusion of multiple population cohorts with considerable large sample sizes, equally balanced sex ratios, and relevant ancestry groups. An ongoing effort meeting these requirements is the MESARGEN consortium (<https://mesargen.wordpress.com/>).

Limitations of the study. We used the Illumina Immunochip version 2 platform in our analysis - a customized design based on the investigation of major autoimmune and inflammatory diseases[42]. Therefore, we limited our investigations to immune-mediated regions, leaving out many genetic variants implicated in sarcoidosis.

Perspectives And Significance

In this work, we revealed sex differences in the genetic architecture of sarcoidosis, particularly in clinical phenotypes LS and non-LS. Our work provides further knowledge that the genetic architectures of LS and non-LS are not only distinct but that biological sex plays a role in determining the underlying genetic architecture.

Abbreviations

LS - Löfgren's syndrome

Non-LS - non- Löfgren's syndrome

SNPs - single nucleotide polymorphisms

eQTL - expression quantitative trait loci

UK Biobank – United Kingdom Biobank

ANXA11 – Annexin A11 gene

GWAS – Genome-wide association study

US-AA – United States African Americans

PC – Principal component

PCA – principal component analysis

IVW - inverse variance weighting

OR – odds ratio

CI – confidence interval

Declarations

Ethics approval and consent to participate

This study was approved by the Stockholm Ethical Review Board, Stockholm, Sweden (DNR: 2007/340-32, 2018/2425-32, and 2018/1991-32 (2018-10-15)). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institution and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for publication

Not applicable.

Availability of data and material

Please contact the author for data requests.

Competing interests

The authors declare no competing interests.

Funding

This project is funded by the Swedish Heart-Lung Foundation awarded to NVR (Grant No. 20170664, 20200505, and 20200506), SK (Grant No. 20200163), and JG (Grant No. 20190478); Karolinska Institutet Foundation awarded to NR (Grant No. FS-2018:0007); Swedish Research Council awarded to LP (Grant No. 2018-02884) and JG (Grant No. 2019-01034). The project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through grant FI 1935/1-1. The project received further infrastructure support from the DFG under Germany's Excellence Strategy – EXC 2167-390884018 and the PopGen Biobank (Kiel, Germany; Ref Nr. 2018-032). The popgen 2.0 network (P2N) is supported by a grant from the German Federal Ministry for Education and Research (01EY1103). US-AA sarcoidosis data and specimen collections were funded by grants to MCI (N01-HR056067, U01-HL060263, R01-HL054306) and BAR (R56-

AI072727, R01-HL092576). The AA cohort investigators are funded via grants from the Foundation for Sarcoidosis Research (Chicago, IL) and the National Institutes of Health (R01HL113326, P30 GM110766, U54GM104938-06).

Authors' contributions

NVR and LP designed and implemented the study. YX conducted all the genetic analyses in the Swedish cohort and meta-analysis among all cohorts. YX and NVR drafted the manuscript. SK, JG, and AE recruited patients, obtained samples, and clinically evaluated patients for phenotype characterization in the Swedish cohort. DE conducted genetic analyses in the German cohort. SS and JMQ recruited patients, obtained samples and clinically evaluated patients for phenotype characterization in the German cohort. BR and MI recruited patients and obtained samples, and CM provided genotype data and oversaw analyses conducted by LG and NP of the US-AA cohort.

All authors revised the manuscript and provided feedback.

Acknowledgments

The authors thank all patients and the personnel involved in this study.

The computations and genomic data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Authors' information

Respiratory Medicine Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Ying Xiong, Susanna Kullberg, Vasiliki Vasila, Anders Eklund, Johan Grunewald, and Natalia V. Rivera

Rheumatology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Leonid Padyukov and Natalia V. Rivera

Center of Molecular Medicine, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Johan Grunewald, Leonid Padyukov, and Natalia V. Rivera

Department of Respiratory Medicine, Theme Inflammation and Infection, Karolinska University Hospital, Stockholm, Sweden

Susanna Kullberg, Anders Eklund, and Johan Grunewald,

Genes and Human Disease, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

Lori Garman, Nathan Pezant, and Courtney G. Montgomery

Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

David Ellinghaus and Stefan Schreiber

Department of Public Health Sciences, Henry Ford Health System, Detroit, MI, USA

Benjamin A. Rybicki

Staten Island University Hospital, Zucker School of Medicine, Northwell/Hofstra University, Staten Island, NY, USA

Michael C. Iannuzzi

Clinic of Internal Medicine I, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

Stefan Schreiber

Department of Pneumology, Medical Center, University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

Joachim Müller-Quernheim

References

1. Valeyre D, Prasse A, Nunes H, Uzunhan Y, Brillet P-Y. Müller-Quernheim J: **Sarcoidosis** *The Lancet*. 2014;383(9923):1155–67.
2. Sauer WH, Stern BJ, Baughman RP, Culver DA, Royal W. High-Risk Sarcoidosis. Current Concepts and Research Imperatives. *Annals of the American Thoracic Society*. 2017;14(Supplement_6):437–44.
3. Ungprasert P, Crowson CS, Matteson EL. Influence of Gender on Epidemiology and Clinical Manifestations of Sarcoidosis: A Population-Based Retrospective Cohort Study 1976–2013. *Lung*. 2017;195(1):87–91.
4. Birnbaum AD, Rifkin LM. **Sarcoidosis: Sex-Dependent Variations in Presentation and Management**. *Journal of ophthalmology* 2014, **2014**:236905.
5. Hadid V, Patenaude V, Oddy L, Abenhaim HA. Sarcoidosis and pregnancy: obstetrical and neonatal outcomes in a population-based cohort of 7 million births. *Journal of perinatal medicine*. 2015;43(2):201–7.
6. Cozier YC, Berman JS, Palmer JR, Boggs DA, Wise LA, Rosenberg L. Reproductive and hormonal factors in relation to incidence of sarcoidosis in US Black women: The Black Women's Health Study. *Am J Epidemiol*. 2012;176(7):635–41.
7. Grunewald J, Grutters JC, Arkema EV, Saketkoo LA, Moller DR, Muller-Quernheim J: **Sarcoidosis**. *Nat Rev Dis Primers* 2019, **5**(1):45.
8. Grunewald J, Eklund A. Löfgren's Syndrome. *Am J Respir Crit Care Med*. 2009;179(4):307–12.
9. Patterson KC, Strek ME. Pulmonary fibrosis in sarcoidosis. Clinical features and outcomes. *Annals of the American Thoracic Society*. 2013;10(4):362–70.
10. Nardi A, Brillet PY, Letoumelin P, Girard F, Brauner M, Uzunhan Y, Naccache JM, Valeyre D, Nunes H. Stage IV sarcoidosis: comparison of survival with the general population and causes of death. *Eur Respir J*. 2011;38(6):1368–73.
11. Jeny F, Uzunhan Y, Lacroix M, Gille T, Brillet PY, Nardi A, Bouvry D, Planes C, Nunes H, Valeyre D. Predictors of mortality in fibrosing pulmonary sarcoidosis. *Respiratory medicine*. 2020;169:105997.
12. Mateyo K, Thomeer M. Sarcoidosis around the Globe. *Semin Respir Crit Care Med*. 2017;38(4):393–403.
13. Grunewald J, Grutters JC, Arkema EV, Saketkoo LA, Moller DR. Müller-Quernheim J: **Sarcoidosis**. *Nature Reviews Disease Primers*. 2019;5(1):45.
14. Pollard KM. Gender differences in autoimmunity associated with exposure to environmental factors. *J Autoimmun*. 2012;38(2–3):J177–86.

15. da Costa CH, Silva VL, Fabricio-Silva GM, Usnayo M, Rufino R, Porto LC. HLA in a cohort of Brazilian patients with sarcoidosis. *Hum Immunol*. 2013;74(10):1326–32.
16. Mallampalli MP, Davies E, Wood D, Robertson H, Polato F, Carter CL. Role of environment and sex differences in the development of autoimmune diseases: a roundtable meeting report. *J Womens Health (Larchmt)*. 2013;22(7):578–86.
17. Arkema EV, Cozier YC. Epidemiology of sarcoidosis: current findings and future directions. *Therapeutic Advances in Chronic Disease*. 2018;9(11):227–40.
18. Rivera NV, Ronninger M, Shchetynsky K, Franke A, Nothen MM, Muller-Quernheim J, Schreiber S, Adrianto I, Karakaya B, van Moorsel CH, et al. High-Density Genetic Mapping Identifies New Susceptibility Variants in Sarcoidosis Phenotypes and Shows Genomic-driven Phenotypic Differences. *Am J Respir Crit Care Med*. 2016;193(9):1008–22.
19. Rybicki BA, Levin AM, McKeigue P, Datta I, Gray-McGuire C, Colombo M, Reich D, Burke RR, Iannuzzi MC. A genome-wide admixture scan for ancestry-linked genes predisposing to sarcoidosis in African-Americans. *Genes Immun*. 2011;12(2):67–77.
20. **UK Biobank (2007) Protocol for a large-scale prospective epidemiological resource. Stockport: UK Biobank Coordinating Centre.** <https://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf>.
21. Fischer A, Ellinghaus D, Nutsua M, Hofmann S, Montgomery CG, Iannuzzi MC, Rybicki BA, Petrek M, Mrazek F, Pabst S, et al. Identification of Immune-Relevant Factors Conferring Sarcoidosis Genetic Risk. *Am J Respir Crit Care Med*. 2015;192(6):727–36.
22. Adrianto I, Lin CP, Hale JJ, Levin AM, Datta I, Parker R, Adler A, Kelly JA, Kaufman KM, Lessard CJ, et al. Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PloS one*. 2012;7(8):e43907.
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
24. Steve E, John B, Dorothée D, Annette L, Anne B, Paul M, Alexandra Z, Eli S, Sebastien V, Kate M, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet*. 2012;44(12):1336.
25. Beecham A, Patsopoulos N, Xifara D, Davis M, Kempainen A, Cotsapas C, Shah T, Spencer C, Booth D, Goris A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet*. 2013;45(11):1353.
26. David E, Hansjörg B, Jorge E-G, Elke R, Anja M, Ingo M, Norbert H, Heidi S, Natalija N, Sven M, et al. High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. *Nat Genet*. 2013;45(7):808.
27. Carmona Fd M, Sarah I, Martín J-E, TJ c, Vaglio, Eyre A, Bossini-Castillo S, Castañeda L S, Cid Maria c, Hernández-Rodríguez J et al: **A Large-Scale Genetic Analysis Reveals a Strong Contribution of the HLA Class II Region to Giant Cell Arteritis Susceptibility.** *The American Journal of Human Genetics* 2015, **96**(4):565–580.
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190–1.
29. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M. **UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.** *PLoS medicine* 2015, 12(3).
30. Mishra A, Macgregor S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Research Human Genetics*. 2015;18(1):86–91.
31. Ferreira MA, Jansen R, Willemsen G, Penninx B, Bain LM, Vicente CT, Revez JA, Matheson MC, Hui J, Tung JY, et al. Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. *J Allergy Clin Immunol*. 2017;139(4):1148–57.

32. Franke L, Jansen RC. eQTL analysis in humans. *Methods in molecular biology* (Clifton NJ). 2009;573:311.
33. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC Jr, Wright MW, et al. Gene map of the extended human MHC. *Nature reviews Genetics*. 2004;5(12):889–99.
34. Rubtsova K, Marrack P, Rubtsov AV. Sexual dimorphism in autoimmunity. *J Clin Investig*. 2015;125(6):2187–93.
35. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, Schurmann M, Muller-Quernheim J, Krawczak M, Rosenstiel P, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*. 2008;40(9):1103–6.
36. Yin C, Zhu B, Zhang T, Liu T, Chen S, Liu Y, Li X, Miao X, Li S, Mi X, et al. Pharmacological Targeting of STK19 Inhibits Oncogenic NRAS-Driven Melanomagenesis. *Cell*. 2019;176(5):1113–27.e1116.
37. Soyombo AA, Yi W, Hofmann SL: **Structure of the human palmitoyl-protein thioesterase-2 gene (PPT2) in the major histocompatibility complex on chromosome 6p21.3.** *Genomics* 1999, **56**(2):208–216.
38. Dempsey OJ, Paterson EW, Kerr KM, Denison AR. **Sarcoidosis** *BMJ*. 2009;339:b3206.
39. Miedema JR, Kaiser Y, Broos CE, Wijsenbeek MS, Grunewald J, Kool M. Th17-lineage cells in pulmonary sarcoidosis and Lofgren's syndrome: Friend or foe? *J Autoimmun*. 2018;87:82–96.
40. Grunewald J, Eklund A. Sex-specific manifestations of Lofgren's syndrome. *Am J Respir Crit Care Med*. 2007;175(1):40–4.
41. Fischer A, Valentonyte R, Nebel A, Nothnagel M, Muller-Quernheim J, Schurmann M, Schreiber S. Female-specific association of C-C chemokine receptor 5 gene polymorphisms with Lofgren's syndrome. *Journal of molecular medicine*. 2008;86(5):553–61.
42. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Res Ther*. 2011;13(1):101–1.

Tables

Table 1. Sample size in association analysis in Sweden, Germany and United States (A); Sample size in replication look-up in the UK Biobank (B)

A.

	LS		non-LS	
	Males	Females	Males	Females
Swedish cohort	166 cases vs. 849 HC	132 cases vs. 2,236 HC	263 cases vs. 855 HC	181 cases vs. 2245 HC
German cohort	24 cases vs. 1750 HC	40 cases vs. 1,692 HC	470 cases vs. 1750 HC	732 cases vs. 1,692 HC
US-AA cohort	-	-	322 cases vs. 497 HC	951 cases vs. 1148 HC

HC: healthy controls.

B.

Males	Females
344 cases vs. 166,644 controls	361 cases and 193,792 controls

Table 2. The top 25 significant SNPs ($P_{\text{meta}} < 5\text{e-}8$) of GWAS meta-analysis on LS males in the Swedish and German cohorts and replication look-up in the UK Biobank

SNP	Swedish	German	UK Biobank
rs1117200	1.2e-8	1.5e-8	1.8e-8
rs1117201	1.1e-8	1.4e-8	1.7e-8
rs1117202	1.0e-8	1.3e-8	1.6e-8
rs1117203	9.5e-9	1.2e-8	1.5e-8
rs1117204	9.0e-9	1.1e-8	1.4e-8
rs1117205	8.5e-9	1.0e-8	1.3e-8
rs1117206	8.0e-9	9.5e-9	1.2e-8
rs1117207	7.5e-9	9.0e-9	1.1e-8
rs1117208	7.0e-9	8.5e-9	1.0e-8
rs1117209	6.5e-9	8.0e-9	9.5e-9
rs1117210	6.0e-9	7.5e-9	9.0e-9
rs1117211	5.5e-9	7.0e-9	8.5e-9
rs1117212	5.0e-9	6.5e-9	8.0e-9
rs1117213	4.5e-9	6.0e-9	7.5e-9
rs1117214	4.0e-9	5.5e-9	7.0e-9
rs1117215	3.5e-9	5.0e-9	6.5e-9
rs1117216	3.0e-9	4.5e-9	6.0e-9
rs1117217	2.5e-9	4.0e-9	5.5e-9
rs1117218	2.0e-9	3.5e-9	5.0e-9
rs1117219	1.5e-9	3.0e-9	4.5e-9
rs1117220	1.0e-9	2.5e-9	4.0e-9
rs1117221	5.0e-10	1.5e-9	3.5e-9
rs1117222	4.0e-10	1.0e-9	3.0e-9
rs1117223	3.0e-10	5.0e-10	2.5e-9
rs1117224	2.0e-10	4.0e-10	2.0e-9
rs1117225	1.0e-10	3.0e-10	1.5e-9

SNP	CHR	BP (hg19)	CA/NCA	Gene (Nearby genes)	Cohorts	OR (95%CI)	P _{meta}	P _{UK- Biobank}
rs3130288	6	32096001	A/C	<i>ATF6B</i>	SWE, GER	4.83 (3.67- 6.36)	1.80E- 28	2.28E-15
rs9268219	6	32284108	A/C	<i>C6orf10</i>	SWE, GER	0.21 (0.16- 0.28)	8.56E- 28	4.49E-14
rs3129927	6	32333827	A/C	<i>C6orf10</i>	SWE, GER	0.21 (0.16- 0.28)	8.83E- 28	1.31E-13
rs2395149	6	32325562	A/G	<i>C6orf10</i>	SWE, GER	4.69 (3.56- 6.17)	8.83E- 28	1.26E-13
rs1265757	6	32302382	A/G	<i>C6orf10</i>	SWE, GER	4.68 (3.56- 6.17)	9.19E- 28	1.26E-13
rs9268235	6	32290208	A/G	<i>C6orf10</i>	SWE, GER	4.68 (3.56- 6.17)	9.55E- 28	-
rs389884	6	31940897	A/G	<i>STK19</i>	SWE, GER	0.21 (0.16- 0.28)	1.07E- 27	3.57E-16
rs3135394	6	32408497	A/G	<i>HLA-DRA</i>	SWE, GER	0.22 (0.17- 0.29)	1.11E- 27	6.53E-12
rs9268177	6	32274882	A/C	<i>C6orf10</i>	SWE, GER	4.61 (3.49- 6.05)	1.62E- 27	3.98E-14
rs7775397	6	32261252	A/C	<i>C6orf10</i>	SWE, GER	0.22 (0.17- 0.29)	1.62E- 27	3.91E-14
rs1150753	6	32059867	A/G	<i>TNXB</i>	SWE, GER	0.21 (0.16- 0.28)	1.72E- 27	2.92E-15
rs3129950	6	32358201	C/G	<i>(BTNL2)</i>	SWE, GER	4.56 (3.46- 5.99)	3.30E- 27	2.20E-13
rs1270942	6	31918860	A/G	<i>CFB</i>	SWE, GER	0.22 (0.17- 0.29)	4.15E- 27	1.46E-16
rs3132971	6	32230256	A/C	<i>(C6orf10)</i>	SWE, GER	0.22 (0.17- 0.29)	4.20E- 27	4.15E-14
rs519417	6	31878433	A/G	<i>C2</i>	SWE, GER	4.53 (3.42- 5.99)	8.33E- 27	4.19E-16
rs497309	6	31892484	A/C	<i>C2</i>	SWE, GER	0.22 (0.17- 0.30)	9.39E- 27	1.31E-16
rs1794282	6	32666526	A/G	<i>(HLA-DQB1)</i>	SWE, GER	4.49 (3.42- 5.93)	1.80E- 26	3.04E-12
rs3135353	6	32392877	A/G	<i>(HLA-DRA)</i>	SWE, GER	3.91(3.03- 5.05)	2.49E- 26	1.59E-10
rs3130679	6	31807540	A/G	<i>C6orf48</i>	SWE, GER	0.23 (0.17- 0.30)	2.52E- 26	3.06E-15
rs558702	6	31870326	A/G	<i>C2</i>	SWE, GER	4.43 (3.35- 5.81)	3.13E- 26	5.00E-16
rs3132450	6	31596138	A/G	<i>PRRC2A</i>	SWE, GER	0.23 (0.17- 0.30)	3.38E- 26	3.96E-15

rs9267531	6	31636742	A/G	<i>CSNK2B</i>	SWE, GER	0.23 (0.17- 0.30)	4.09E- 26	3.81E-15
rs3117582	6	31620520	A/C	<i>(BAG6)</i>	SWE, GER	0.23 (0.17- 0.30)	4.09E- 26	3.83E-15
rs3130490	6	31739120	A/C	<i>C6orf27</i>	SWE, GER	4.51 (3.42- 5.99)	4.81E- 26	2.62E-15
rs915652	6	31749142	A/G	<i>VAR5</i>	SWE, GER	4.36 (3.32- 5.75)	5.24E- 26	5.46E-15

SNP: single nucleotide polymorphism; CHR: chromosome; BP: base pairs; hg19 = human genome assembly version 19; CA: coded allele; NCA: non-coded allele; SWE: Sweden; GER: Germany; OR: odds ratio; CI: confidence interval; Pmeta: meta-P value; P_{UK-Biobank}: P-value for association in replication look-up in the UK Biobank

Table 3. The top 25 significant SNPs (Pmeta < 5e-8) of GWAS meta-analysis on LS females in the Swedish and German cohorts and replication look-up in the UK Biobank

SNP	CHR	BP	CA/NCA	Gene (Nearby genes)	Cohorts	OR (95%CI)	P _{meta}	P _{UK-Biobank}
rs9268219	6	32284108	A/C	<i>C6orf10</i>	SWE, GER	0.27 (0.21- 0.35)	2.04E- 24	3.02E-12
rs3129927	6	32333827	A/C	<i>C6orf10</i>	SWE, GER	0.28 (0.21- 0.35)	2.85E- 24	3.31E-13
rs2187668	6	32605884	A/G	<i>HLA-DQA1</i>	SWE, GER	3.51 (2.76- 4.48)	3.49E- 24	-
rs3129950	6	32358201	C/G	<i>(BTNL2)</i>	SWE, GER	3.66 (2.84- 4.70)	4.20E- 24	7.05E-13
rs2395149	6	32325562	A/G	<i>C6orf10</i>	SWE, GER	3.58 (2.79- 4.59)	1.09E- 23	3.35E-13
rs9268235	6	32290208	A/G	<i>C6orf10</i>	SWE, GER	3.57 (2.78- 4.58)	1.38E- 23	-
rs1265757	6	32302382	A/G	<i>C6orf10</i>	SWE, GER	3.57 (2.78- 4.58)	1.38E- 23	3.33E-13
rs3130288	6	32096001	A/C	<i>ATF6B</i>	SWE, GER	3.56 (2.78- 4.57)	1.49E- 23	1.05E-13
rs7775397	6	32261252	A/C	<i>C6orf10</i>	SWE, GER	0.28 (0.22- 0.36)	1.52E- 23	9.55E-13
rs9268177	6	32274882	A/C	<i>C6orf10</i>	SWE, GER	3.57 (2.78- 4.58)	1.53E- 23	9.28E-13
rs2854275	6	32628428	A/C	<i>HLA-DQB1</i>	SWE, GER	3.47 (2.72- 4.42)	1.54E- 23	8.91E-11
rs3132510	6	31172151	A/G	<i>(HCG27)</i>	SWE, GER	0.28 (0.22- 0.36)	1.89E- 23	2.69E-11
rs3129716	6	32657436	A/G	<i>(HLA-DQB1)</i>	SWE, GER	0.29 (0.23- 0.37)	2.75E- 23	8.74E-11
rs3132971	6	32230256	A/C	<i>(C6orf10)</i>	SWE, GER	0.28 (0.22- 0.36)	4.50E- 23	9.45E-13
rs3117577	6	31727474	A/G	<i>MSH5</i>	SWE, GER	0.29 (0.23- 0.37)	4.94E- 23	1.43E-13
rs2856674	6	32659645	A/G	<i>(HLA-DQB1)</i>	SWE, GER	0.29 (0.23- 0.37)	4.95E- 23	7.38E-11
rs3130544	6	31058340	A/C	<i>(C6orf15)</i>	SWE, GER	3.48 (2.72- 4.46)	4.98E- 23	1.44E-09
rs3131618	6	31434621	A/G	<i>(HCP5)</i>	SWE, GER	0.28 (0.22- 0.36)	5.59E- 23	2.35E-11
rs3130985	6	31085356	A/G	<i>CDSN</i>	SWE, GER	3.48 (2.72- 4.46)	5.79E- 23	6.13E-10
rs3094222	6	31081434	A/G	<i>(C6orf15)</i>	SWE, GER	0.29 (0.22- 0.37)	6.06E- 23	2.89E-09
rs3135394	6	32408497	A/G	<i>HLA-DRA</i>	SWE, GER	0.29 (0.23- 0.37)	6.30E- 23	7.65E-13

rs3132449	6	31626013	A/G	(<i>APOM</i>)	SWE, GER	3.44 (2.69- 4.40)	6.53E- 23	1.52E-13
rs1150753	6	32059867	A/G	<i>TNXB</i>	SWE, GER	0.29 (0.22- 0.37)	6.63E- 23	3.83E-13
rs3131378	6	31725285	A/G	<i>MSH5</i>	SWE, GER	0.29 (0.23- 0.37)	8.10E- 23	1.36E-13
rs389884	6	31940897	A/G	<i>STK19</i>	SWE, GER	0.29 (0.22- 0.37)	8.84E- 23	4.72E-13

SNP: single nucleotide polymorphism; CHR: chromosome; BP: base pairs; hg19 = human genome assembly version 19; CA: coded allele; NCA: non-coded allele; SWE: Sweden; GER: Germany; OR: odds ratio; CI: confidence interval; Pmeta: meta-P value; PUK-Biobank: P-value for association in replication look-up in the UK Biobank.

Table 4. The top significant SNPs ($P_{\text{meta}} < 5e-8$) from the GWAS meta-analysis on non-LS males in the Swedish, German, and US-AA cohorts and replication look-up in the UK Biobank

SNP	CHR	BP	CA/NCA	Gene (Nearby genes)	Cohorts	OR (95%CI)	P _{meta}	P _{UK-Biobank}
rs1964995	6	32449411	A/G	(<i>HLA-DRB5</i>)	SWE, GER, AA	1.77 (1.58-1.99)	1.46E-22	1.45E-08
rs9271588	6	32590953	A/G	(<i>HLA-DQA1</i>)	SWE, GER, AA	1.57 (1.41-1.75)	1.64E-16	6.74E-07
rs7754768	6	32420179	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	0.66 (0.60-0.73)	1.80E-15	1.54E-07
rs2239802	6	32411846	C/G	<i>HLA-DRA</i>	SWE, GER, AA	1.54 (1.38-1.71)	1.88E-14	2.65E-01
rs4530903	6	32581889	A/G	(<i>HLA-DQA1</i>)	SWE, GER, AA	0.45 (0.36-0.56)	5.32E-13	1.41E-01
rs1953600	10	81911725	A/G	(<i>ANXA11</i>)	SWE, GER, AA	0.68 (0.60-0.76)	8.50E-12	2.78E-04
rs477515	6	32569691	A/G	(<i>HLA-DRB1</i>)	SWE, GER, AA	0.69 (0.61-0.78)	2.23E-09	3.56E-06
rs389883	6	31947460	A/C	<i>STK19</i>	SWE, GER, AA	0.71 (0.63-0.79)	3.99E-09	6.74E-10
rs9405108	6	32438648	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	0.70 (0.62-0.79)	6.27E-09	4.22E-06
rs2395185	6	32433167	A/C	(<i>HLA-DRA</i>)	SWE, GER, AA	0.70 (0.62-0.79)	6.27E-09	4.31E-06
rs9268923	6	32432835	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	0.70 (0.62-0.79)	6.27E-09	7.01E-06
rs9268969	6	32434349	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	0.70 (0.62-0.79)	6.82E-09	4.24E-06
rs9368726	6	32438542	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	1.43 (1.27-1.61)	7.50E-09	4.21E-06
rs9268853	6	32429643	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	1.43 (1.27-1.61)	7.94E-09	3.90E-06

SNP: single nucleotide polymorphism; CHR: chromosome; BP: base pairs; hg19 = human genome assembly version 19; CA: coded allele; NCA: non-coded allele; SWE: Sweden; GER: Germany; OR: odds ratio; CI: confidence interval; P_{meta}: meta-P value; PUK-Biobank: P-value for association in replication look-up in the UK Biobank.

Table 5. The top 25 significant SNPs ($P_{\text{meta}} < 5e-8$) from the second-step GWAS meta-analysis on non-LS females in the Swedish, German, and US-AA cohorts and replication look-up in the UK Biobank

SNP	CHR	BP	CA/NCA	Gene (nearby genes)	Cohorts	OR (95%CI)	P _{meta}	I ² (%)	P _{het}	P _{UK-Biobank}
rs1964995	6	32449411	A/G	(<i>HLA-DRB5</i>)	SWE, GER, AA	1.52 (1.39- 1.66)	2.86E- 19	91.8	4.93E- 04	1.45E- 08
rs2395153	6	32345595	C/G	(<i>C6orf10</i>)	SWE, GER, AA	0.66 (0.60- 0.72)	1.17E- 18	75.1	4.53E- 02	1.35E- 06
rs7195	6	32412539	A/G	<i>HLA-DRA</i>	SWE, GER, AA	1.44 (1.33- 1.57)	7.64E- 18	0	6.23E- 01	3.13E- 07
rs8084	6	32411035	A/C	<i>HLA-DRA</i>	SWE, GER, AA	1.43 (1.31- 1.55)	6.04E- 17	0	6.17E- 01	-
rs2294878	6	32367795	A/C	<i>BTNL2</i>	SWE, GER, AA	0.71 (0.65- 0.77)	2.00E- 15	75	4.53E- 02	-
rs4373382	6	32350868	A/C	(<i>C6orf10</i>)	SWE, GER, AA	1.41 (1.29- 1.54)	7.78E- 15	87.9	4.05E- 03	-
rs7754768	6	32420179	A/G	(<i>HLA-DRA</i>)	SWE, GER, AA	0.72 (0.67- 0.79)	1.86E- 14	40.6	1.94E- 01	1.54E- 07
rs3817973	6	32361111	A/G	(<i>BTNL2</i>)	SWE, GER, AA	0.72 (0.66- 0.78)	4.67E- 14	89.7	1.86E- 03	-
rs2076529	6	32363955	A/G	<i>BTNL2</i>	SWE, GER, AA	1.40 (1.28- 1.52)	4.67E- 14	89.7	1.86E- 03	-
rs4424066	6	32354428	A/G	(<i>BTNL2</i>)	SWE, GER, AA	1.39 (1.28- 1.52)	6.44E- 14	89.5	1.99E- 03	-
rs9268472	6	32355605	A/G	(<i>BTNL2</i>)	SWE, GER, AA	0.72 (0.66- 0.78)	7.24E- 14	90.3	1.31E- 03	-
rs2076530	6	32363816	A/G	<i>BTNL2</i>	SWE, GER, AA	1.38 (1.27- 1.51)	1.46E- 13	88.9	2.71E- 03	-
rs6913309	6	32339840	A/T	(<i>C6orf10</i>)	SWE, GER, AA	0.71 (0.64- 0.78)	4.23E- 12	0	6.41E- 01	-
rs2269425	6	32123639	A/G	<i>PPT2</i>	SWE, GER, AA	0.62 (0.54- 0.71)	9.75E- 12	21.1	2.60E- 01	5.15E- 03
rs3129727	6	32679690	A/G	(<i>HLA-DQA2</i>)	SWE, GER, AA	2.11 (1.68- 2.65)	1.61E- 10	51.4	1.51E- 01	-
rs6457617	6	32663851	A/G	(<i>HLA-DQB1</i>)	SWE, GER, AA	0.77 (0.71- 0.83)	2.41E- 10	85.7	8.12E- 03	-

rs9268145	6	32257284	A/C	(<i>C6orf10</i>)	SWE, GER, AA	1.53 (1.34- 1.76)	6.48E- 10	0	3.29E- 01	-
rs6910071	6	32282854	A/G	<i>C6orf10</i>	SWE, GER, AA	1.53 (1.34- 1.76)	6.48E- 10	0	3.29E- 01	-
rs9268362	6	32333341	A/G	<i>C6orf10</i>	SWE, GER, AA	1.53 (1.34- 1.76)	6.68E- 10	0	3.27E- 01	-
rs3115572	6	32220484	C/G	(<i>NOTCH4</i>)	SWE, GER, AA	1.29 (1.19- 1.40)	8.33E- 10	41.3	1.92E- 01	-
rs3817963	6	32368087	A/G	<i>BTNL2</i>	SWE, GER, AA	1.38 (1.25- 1.54)	1.53E- 09	0	6.31E- 01	-
rs9268482	6	32367777	A/T	<i>BTNL2</i>	SWE, GER, AA	1.38 (1.24- 1.54)	1.86E- 09	0	5.27E- 01	-
rs9268474	6	32357165	A/G	(<i>BTNL2</i>)	SWE, GER, AA	1.38 (1.24- 1.54)	1.94E- 09	0	5.35E- 01	-
rs8180664	6	32347490	A/G	(<i>C6orf10</i>)	SWE, GER, AA	0.72 (0.65- 0.80)	1.94E- 09	0	5.35E- 01	-
rs2076525	6	32370616	A/G	<i>BTNL2</i>	SWE, GER, AA	1.38 (1.24- 1.54)	1.94E- 09	0	5.35E- 01	-

SNP: single nucleotide polymorphism; CHR: chromosome; BP: base pairs; hg19 = human genome assembly version 19; CA: coded allele; NCA: non-coded allele; SWE: Sweden; GER: Germany; OR: odds ratio; CI: confidence interval; P_{meta} : meta-P value; I^2 = heterogeneity index; P_{het} : P-value for heterogeneity statistic $P_{\text{UK-Biobank}}$: P-value for association in replication look-up in the UK Biobank.

Table 6. Top 15 significant gene-based results of SNPs of meta-analysis associated with LS in males

CHR	Gene	nSNPs	Start	Stop	P-val	Best-SNP	SNP-pval
6	<i>ABCF1</i>	1	30539169	30559309	1.00E-06	rs3132610	1.11E-18
6	<i>ABT1</i>	1	26597170	26600278	1.00E-06	rs4634439	4.36E-09
6	<i>AGER</i>	1	32148744	32152099	1.00E-06	rs3134940	7.12E-17
6	<i>AGPAT1</i>	3	32135982	32145888	1.00E-06	rs3131297	4.02E-15
6	<i>APOM</i>	1	31620186	31625987	1.00E-06	rs3117582	4.09E-26
6	<i>ATAT1</i>	1	30594618	30614600	1.00E-06	rs9262132	4.61E-19
6	<i>ATF6B</i>	1	32083044	32096017	1.00E-06	rs3130288	1.80E-28
6	<i>ATP6V1G2</i>	2	31512227	31514625	1.00E-06	rs9267488	1.13E-25
6	<i>ATP6V1G2-DDX39B</i>	6	31497995	31514625	1.00E-06	rs9267488	1.13E-25
6	<i>BAG6</i>	6	31606804	31620477	1.00E-06	rs3117583	4.72E-22
6	<i>BTN3A2</i>	7	26365386	26378548	1.00E-06	rs9393714	1.42E-08
6	<i>BTNL2</i>	1	32361115	32374907	1.00E-06	rs2395158	1.28E-25
6	<i>C2</i>	3	31865561	31913451	1.00E-06	rs519417	8.33E-27
6	<i>C6orf10</i>	50	32260474	32339689	1.00E-06	rs9268219	8.56E-28

CHR: chromosome; Start: start position of genes; Stop: stop position of genes; P-val: gene-based p-value

Table 7. Top 15 significant gene-based results of meta-analysis SNPs associated with LS females

CHR	Gene	nSNPs	Start	Stop	P-val	Best-SNP	SNP-pval
6	<i>ABCF1</i>	1	30539169	30559309	1.00E-06	rs3132610	2.25E-17
6	<i>AGER</i>	1	32148744	32152099	1.00E-06	rs3134940	8.41E-15
6	<i>AGPAT1</i>	3	32135982	32145888	1.00E-06	rs3134947	5.13E-13
6	<i>APOM</i>	1	31620186	31625987	1.00E-06	rs3117582	1.57E-22
6	<i>ATAT1</i>	1	30594618	30614600	1.00E-06	rs9262132	3.40E-18
6	<i>ATF6B</i>	1	32083044	32096017	1.00E-06	rs3130288	1.49E-23
6	<i>ATP6V1G2</i>	2	31512227	31514625	1.00E-06	rs9267488	1.51E-21
6	<i>ATP6V1G2-DDX39B</i>	5	31497995	31514625	1.00E-06	rs9267488	1.51E-21
6	<i>BAG6</i>	2	31606804	31620477	1.00E-06	rs3117583	1.99E-15
6	<i>BTNL2</i>	1	32361115	32374907	1.00E-06	rs2395158	1.58E-19
6	<i>C2</i>	4	31865561	31913451	1.00E-06	rs497309	1.13E-22
6	<i>C6orf10</i>	50	32260474	32339689	1.00E-06	rs9268219	2.04E-24
6	<i>C6orf136</i>	1	30614815	30620987	1.00E-06	rs9262135	4.05E-18
6	<i>C6orf15</i>	1	31078999	31080332	1.00E-06	rs2233974	1.40E-19

CHR: chromosome; Start: start position of genes; Stop: stop position of genes; P-val: gene-based p-value

Table 8. Gene-based results of significant meta-analysis SNPs associated with non-LS males

CHR	Gene	nSNPs	Start	Stop	P-val	Best-SNP	SNP-pval
6	<i>HLA-DRA</i>	1	32407618	32412823	1.00E-06	rs2239802	1.88E-14
6	<i>STK19</i>	1	31938951	31949223	1.00E-06	rs389883	3.99E-09
10	<i>ANXA11</i>	1	81914879	81965433	1.00E-06	rs1049550	1.98E-11

CHR: chromosome; Start: start position of genes; Stop: stop position of genes; P-val: gene-based p-value

Table 9. Gene-based results of significant meta-analysis SNPs associated with non-LS females

CHR	Gene	nSNPs	Start	Stop	P-val	Best-SNP	SNP-pval
6	<i>BTNL2</i>	9	32361115	32374907	1.00E-06	rs2294878	2.00E-15
6	<i>C6orf10</i>	2	32260474	32339689	1.00E-06	rs6910071	6.48E-10
6	<i>HLA-DRA</i>	2	32407618	32412823	1.00E-06	rs7195	7.64E-18
6	<i>PPT2-EGFL8</i>	1	32121775	32136062	1.00E-06	rs2269425	9.75E-12
6	<i>PPT2</i>	1	32121775	32136062	1.00E-06	rs2269425	9.75E-12

CHR: chromosome; Start: start position of genes; Stop: stop position of genes; P-val: gene-based p-value

Figures

Figure 1

Regional association plots for the lead SNPs of GWAS meta-analysis on LS males (A) and females (B)

Figure 2

Forest plots of lead SNPs from the GWAS meta-analysis on LS males (A) and females (B) in Swedish and German cohorts

Figure 3

Forest plots of lead SNPs from the GWAS meta-analysis on non-LS males and females in European cohorts (Sweden and, Germany) (A – B) and in multi-ethnic cohorts (Sweden, Germany, and US-AA) (C – D)

Figure 4

Regional association plots for the lead SNPs of GWAS meta-analysis on non-LS in males (A) and females (B)

Figure 5

Venn diagram of *cis*-SNPs in sex groups on LS derived from SNPs of GWAS meta-analysis

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile2.pdf](#)
- [AdditionalFile1.pdf](#)
- [Additionalfile3.pdf](#)
- [AdditionalFile4.pdf](#)
- [AdditionalFile5.pdf](#)
- [AdditionalFile6.pdf](#)
- [AdditionalFile7.pdf](#)