Title

Genome-wide association study identifies multiple HLA loci for sarcoidosis susceptibility.

Journal

Human Molecular Genetics, 32(16)

Authors

Liao, Shu-Yi
Jacobson, Sean
Hamzeh, Nabeel
et al.

Publication Date

2023-08-07

DOI

10.1093/hmg/ddad067

Peer reviewed

OXFORD

Human Molecular Genetics, 2023, Vol. 32, 16, 2669–2678
https://doi.org/10.1093/hmg/ddad067
Advance access publication date 3 July 2023
Association Studies Article

Genome-wide association study identifies multiple HLA loci for sarcoidosis susceptibility

Shu-Yi Liao[1,2,3], Sean Jacobson[1], Nabeel Y. Hamzeh[4], Daniel A. Culver[5], Briana Q. Barkes[1], Margerate Mroz[1], Kristyn Macphail[1], Karin Pacheco[1,2,3], Divya C. Patel[6], Yasmine Wasfi S.[7], Laura L. Koth[8], Carl D. Langefeld[9,10], Sonia Leach[1], Elizebeth White[1], Courtney Montgomery[11], Lisa A. Maier[1,2,3], Tasha Fingerlin E.[2,3,12,*] and GRADs Investigators[13]

[1]Department of Medicine, National Jewish Health, Denver, CO 80206, USA
[2]Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[3]Colorado School of Public Health, University of Colorado Denver–Anschutz Medical Campus, Aurora, CO 80045, USA
[4]Department of Medicine, University of Iowa, Iowa City, IA 52242, USA
[5]Department of Medicine, Cleveland Clinic, Cleveland, OH 44195, USA
[6]Department of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, University of Florida, Gainesville, FL 32610, USA
[7]Johnson & Johnson, Spring House, PA 19034, USA
[8]Department of Medicine, University of California-San Fransisco, San Fransisco, CA 94143, USA
[9]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC 27101, USA
[10]Wake Forest University School of Medicine, Center for Precision Medicine, Winston-Salem, NC 27101, USA
[11]Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA
[12]Department of Immunology and Genomic Medicine, National Jewish Health, Denver, CO 80206, USA
[13]Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS)

*To whom correspondence should be addressed at: Department of Medicine, National Jewish Health, 1400 Jackson Street, Denver, CO 80206, USA. Tel: +1 3033982487; Email: fingerlint@njhealth.org

Abstract

Sarcoidosis is a complex systemic disease. Our study aimed to (1) identify novel alleles associated with sarcoidosis susceptibility; (2) provide an in-depth evaluation of HLA alleles and sarcoidosis susceptibility and (3) integrate genetic and transcription data to identify risk loci that may more directly impact disease pathogenesis. We report a genome-wide association study of 1335 sarcoidosis cases and 1264 controls of European descent (EA) and investigate associated alleles in a study of African Americans (AA: 1487 cases and 1504 controls). The EA and AA cohort was recruited from multiple United States sites. HLA alleles were imputed and tested for association with sarcoidosis susceptibility. Expression quantitative locus and colocalization analysis were performed using a subset of subjects with transcriptome data. Forty-nine SNPs in the HLA region in HLA-DRA, -DRB9, -DRB5, -DQA1 and BRD2 genes were significantly associated with sarcoidosis susceptibility in EA, rs3129888 was also a risk variant for sarcoidosis in AA. Classical HLA alleles DRB1*0101, DQA1*0101 and DQB1*0501, which are highly correlated, were also associated with sarcoidosis. rs3135287 near HLA-DRA was associated with HLA-DRA expression in peripheral blood mononuclear cells and bronchoalveolar lavage from subjects and lung tissue and whole blood from GTEx. We identified six novel SNPs (out of the seven SNPs representing the 49 significant SNPs) and nine HLA alleles associated with sarcoidosis susceptibility in the largest EA population. We also replicated our findings in an AA population. Our study reiterates the potential role of antigen recognition and/or presentation HLA class II genes in sarcoidosis pathogenesis.

Introduction

Sarcoidosis is a complex systemic disease affecting between 45 and 300/100000 persons in the United States (1). While environmental and genetic susceptibility factors have been associated with sarcoidosis (2–4), the driving risk factors for disease are still largely undefined. Previous genome-wide association studies (GWAS) have consistently implicated the HLA region as harboring genetic risk loci based on single nucleotide polymorphisms (SNPs) (5–12), while previous studies have implicated specific HLA alleles such as DRB1*1101, DRB1*1501 and DQB1*0602 (13,14). The strong role of the HLA region for sarcoidosis disease risk is consistent with exposure and immune-based associations observed in patients, as well as the heterogeneous disease course.

Most sarcoidosis genetic studies have focused on genotype data, limiting understanding of potential functional genetic features associated with disease status. Studying the impact of risk alleles on gene expression can provide the first step in understanding their potential biological impact (15). This is especially true for complex diseases since the majority of genetic variants robustly associated with these diseases fall in non-coding regions of the genome (16). While the function of non-coding regions was unknown in the past, there is substantial evidence that many of them influence disease risk through regulatory effects, including those that impact gene expression (15). Hence, integrating genetic and transcriptomic data may identify loci with a more direct or functional effect on disease pathogenesis.

To identify genetic risk factors for disease and study their effects on gene expression, we conducted a genome-wide association study in European American (EA) population that includes follow-up of risk alleles on gene expression in peripheral blood as well as lung cells. We confirm the strong role of the HLA region in sarcoidosis risk and demonstrate both novel SNPs and

Received: December 3, 2022. **Revised:** March 30, 2023. **Accepted:** April 19, 2023

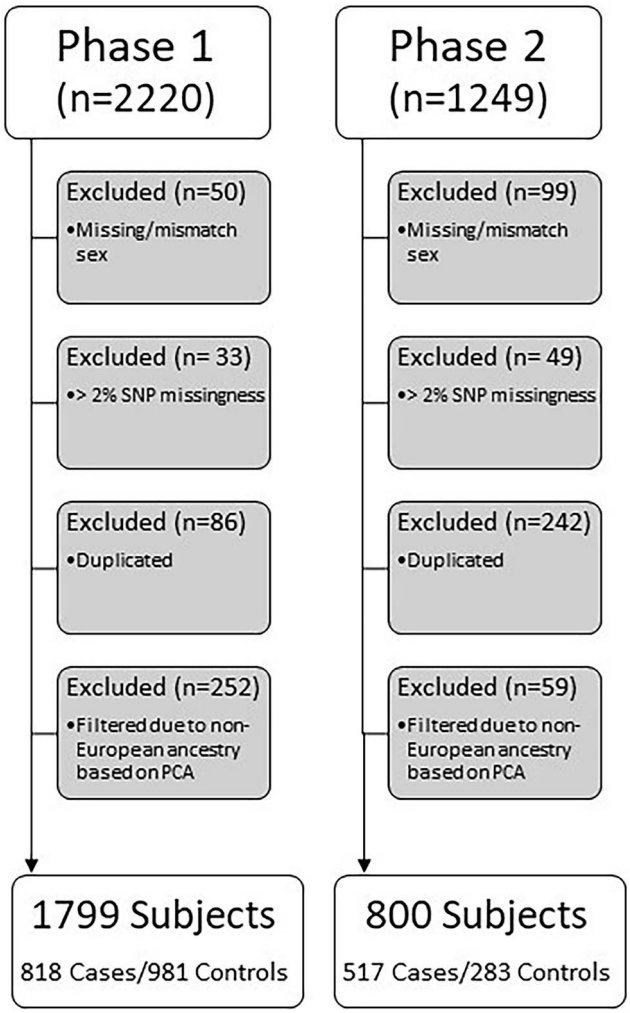

Figure 1. CONSORT flow chart for the GWAS samples. This figure outlines the exclusion of the subject in Phase 1 and Phase 2, respectively.

classical HLA alleles associated with disease and influencing HLA expression. We replicated some of these findings in a large African American (AA) population.

Results

Descriptive analysis of GWAS in EA

We enrolled 3141 subjects genotyped on the Illumina array. After quality control (Fig. 1), we included n = 2599 subjects (1335 cases and 1264 controls) in the analysis (Table 1); 818 cases (49% female/51% male) and 981 controls (37% female/63% male) in Phase 1 and 517 cases (53% female/47% male) and 283 controls (71% male/29% female) in Phase 2. Totally, 1088 out of the 1355 cases (80%) of sarcoidosis cases have detailed organ assessment (17). The mean age was 54 ± 11 (SD). The cases' organ involvement ranked from highest to lowest was as follows: lung (94%), cardiac (14%), skin (12%), ocular (7%), non-thoracic lymph nodes (6%), bone/joint (5%), spleen (5%), calcium metabolism (5%), small fiber neuropathy (4%), renal (4%), ear/nose/throat (3%), liver (2%), parotid/salivary gland (2%), bone marrow (2%), muscle (2%) and neurological (2%). For expression quantitative locus (eQTL) analyses, paired genotype-transcription data was available for 136 genotype-BAL transcription and 193 genotype-PBMC transcription samples.

GWAS identifies HLA region associations with sarcoidosis

The meta-analysis of Phases 1 and 2 data identified 49 SNPs reaching GWAS significance ($P < 5 \times 10^{-8}$) with 527 878 non-imputed and 10 002 112 imputed SNPs tested. The genomic control parameter (λ_{GC}) estimate for Phases 1 and 2 was 1.02 and 1.05, respectively. The r^2 LD plot of the 49 SNPs is shown in Supplementary Material, Figure S1 in addition to the top seven SNPs representing all significant SNPs ($r^2 > 0.70$ between that SNP and at least one other genome-wide significant SNP). Those top seven SNPs were rs9269233, rs9271346, rs35656642, rs28589559, rs9276935, rs3129888 and rs71549283 (Table 2), located across the HLA Class II region in *HLA-DRA, -DRB9, -DRB5, -DQA1* and *BRD2 genes on* chromosome 6. The remaining significant SNPs are shown in Supplementary Material, Table S1. With the much-reduced Phase 2 sample size compared with Phase 1, the Phase 2 *P*-value is not nominally significant, although effect sizes (odds ratios) were comparable, and both contribute proportionally to meta-analysis *P*-values. The Manhattan plot for SNP associations is shown in Figure 2A, and Figure 2B and Supplementary Material, Figure S2 show the locus-specific plot for all significant SNPs highlighting the seven top SNPs. For the seven top SNPs, we used a stepwise approach to adjust for other SNPs (Supplementary Material, Table S2) and found rs9269233, rs9276935, rs28589559 and rs3129888 still nominally significant after adjustment (all *P*s < 0.01). Other top SNPs with $P < 5 \times 10^{-5}$ are shown in Supplementary Material, Table S3. In the AA cohort, rs3129888 was also significantly associated with increased risk of sarcoidosis ($P < 0.007$ [0.05/seven SNPs tested], Table 2). These findings suggest that rs3129888 may be a common SNPs associated with disease in both EA and AA. Other SNPs are in the same direction (either protective or increased risk) between EA and AA but may have more influence in EA (significant in EA but not in AA). In the UK Biobank European population, we found two out of the seven SNPs were significantly associated with sarcoidosis with $P < 0.007$ (0.05/7): rs35656642 ($P = 6.67 \times 10^{-8}$) and rs28589559 (2.25×10^{-4}). When we compared our results to previously identified loci from other GWAS studies, we found SNPs in *ANXA11* and *BTNL2* were nominally significant (Supplementary Material, Table S4) in the meta-analysis ($P < 0.005$).

Classic HLA alleles are associated with sarcoidosis

We evaluated classic HLA variants and found nine HLA alleles were significantly associated with sarcoidosis susceptibility ($P < 0.00011$, Table 3 and Supplementary Material, Table S5). Three HLA alleles had a *P*-value$<5 \times 10^{-8}$: DRB1*0101, DQA1*0101 and DQB1*0501. All three HLA alleles were highly correlated and protective against sarcoidosis. We found three previously reported HLA alleles, DRB1*1101, DRB1*1501 and DQB1*0602, significantly associated with increased risk of sarcoidosis (Table 3).

Two GWAS SNPs in the HLA region are statistically independent of HLA allele associations

Although the *P*-values were attenuated slightly, each genome-wide significant SNP remained associated with sarcoidosis after adjustment for each of the three most significant HLA risk alleles (Table 4). We then adjusted for all three most significant HLA risk alleles; the associations were attenuated but remain significant for rs9269233 ($P = 2.15 \times 10^{-7}$) and rs9276935 ($P = 2.21 \times 10^{-7}$), and the effects of each SNP on odds ratios were largely unchanged

Table 1. Sample size and sex for Phase 1 and Phase 2[a] (all European American)

	All (N = 2599)		Phase 1 (N = 1799)		Phase 2 (N = 800)	
Sex	**Sarcoidosis (n = 1335)**	**Control (n = 1264)**	**Sarcoidosis (n = 818)**	**Control (n = 981)**	**Sarcoidosis (n = 517)**	**Control (n = 283)**
Female, n (%)	671 (50%)	565 (45%)	399 (49%)	364 (37%)	272 (53%)	201 (71%)
Male, n (%)	664 (50%)	699 (55%)	419 (51%)	617 (63%)	245 (47%)	82 (29%)

[a]Phase 2 DNA samples became available after the Phase 1 group had been genotyped (see Materials and Methods).

Table 2. Summary of top 7 genome-wide significant SNPs of the genome-wide association study of sarcoidosis (*P*-value $< 5 \times 10^{-8}$)

					European American Cohort								African American Cohort		
					Phase 1			Phase 2			Meta-analysis				
SNP	**Chr**	**Position**	**Minor allele**	**Nearest gene (+/-25 K)**	**MAF case**	**OR (95% CI)**	***P*-value**	**MAF case**	**OR (95% CI)**	***P*-value**	**OR (95% CI)**	***P*-value**	**MAF case**	**OR (95% CI)**	***P*-value**
rs9269233	6	32 451 762	A	HLA-DRB9	0.38	1.73 (1.48,2.01)	5.67E-13	0.34	1.08 (0.85,1.36)	0.54	1.50 (1.32,1.71)	2.32E-10	0.32	1.09 (0.81,1.47)	0.56[a]
rs9271346	6	32 583 468	C	HLA-DQA1	0.27	1.71 (1.45,2.02)	1.53E-10	0.24	1.14 (0.89,1.47)	0.30	1.51 (1.32,1.74)	3.73E-09	0.24	1.19 (1.05,1.34)	0.01
rs35656642	6	32 583 610	A	HLA-DQA1	0.32	0.66 (0.58,0.76)	5.73E-09	0.34	0.84 (0.67,1.05)	0.12	0.71 (0.63,0.80)	1.19E-08	0.29	0.87 (0.78,0.98)	0.02
rs28589559	6	32 587 716	T	HLA-DQA1	0.08	0.50 (0.40,0.63)	2.90E-10	0.13	0.91 (0.67,1.23)	0.55	0.62 (0.52,0.74)	2.46E-08	0.13	0.93 (0.80,1.08)	0.33
rs9276935	6	32 936 441	C	BRD2 (inside gene)	0.05	0.51 (0.39,0.68)	1.26E-06	0.05	0.56 (0.37,0.84)	0.006	0.53 (0.42,0.66)	2.72E-08	0.01	0.87 (0.55,1.40)	0.57
rs3129888	6	32 411 726	G	HLA-DRA (inside gene)	0.27	1.63 (1.38,1.91)	3.71E-09	0.25	1.15 (0.90,1.48)	0.26	1.47 (1.28,1.68)	3.21E-08	0.27	1.36 (1.20,1.54)	1.52E-06
rs71549283	6	32 505 038	A	HLA-DRB5	0.33	0.61 (0.51,0.72)	9.70E-09	0.34	0.83 (0.63,1.10)	0.20	0.66 (0.57,0.77)	4.22E-08	0.20	0.86 (0.64,1.17)	0.34[a]

The Genome Reference Consortium Human Build 37 (GRCh37) was used for genome reference. The minor allele is the allele modeled as an additive effect. SNP: single nucleotide polymorphism; Chr: chromosome; MAF: minor allele frequency; OR: odds ratio; 95% CI: 95% confidence interval. [a]The results of rs9269233 and rs71549283 in the African American cohort did not pass the imputation quality control. The result was obtained from the sequencing data on 932 subjects.

Table 3. Association between HLA alleles and sarcoidosis

	Phase 1				Phase 2					
	Dosage frequency		Univariate results		Dosage frequency		Univariate results		Meta-analysis	
HLA allele	Cases	Controls	OR (95% CI)	*P*-value	Cases	Controls	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Novel alleles*										
DRB1*0101	0.07	0.18	0.35 (0.26, 0.48)	1.46E-11	0.09	0.23	0.31 (0.20, 0.47)	3.51E-08	0.34 (0.26, 0.43)	2.27E-18
DQA1*0101	0.12	0.22	0.53 (0.42, 0.68)	2.61E-07	0.12	0.29	0.34 (0.24, 0.50)	1.09E-08	0.47 (0.38, 0.57)	1.09E-13
DQB1*0501	0.13	0.24	0.61 (0.49, 0.75)	5.45E-06	0.12	0.30	0.39 (0.28, 0.55)	6.03E-08	0.53 (0.45, 0.64)	1.37E-11
Previously reported alleles										
DRB1*1101	0.15	0.11	1.57 (1.13, 2.18)	7.05E-03	0.14	0.12	1.58 (0.91, 2.73)	0.10	1.57 (1.19, 2.08)	1.67E-03
DRB1*1501	0.40	0.27	1.56 (1.30, 1.86)	1.03E-06	0.35	0.29	1.20 (0.90, 1.61)	0.21	1.45 (1.25, 1.69)	1.75E-06
DQB1*0602	0.39	0.26	1.57 (1.31, 1.88)	1.11E-06	0.34	0.27	1.23 (0.91, 1.66)	0.19	1.47 (1.26, 1.72)	1.53E-06

OR: odds ratio; 95% CI: 95% confidence interval. *P-value $< 5 \times 10^{-8}$, remaining significant HLA alleles ($P < 0.00011$) are listed in Supplementary Material, Table S5.

after adjustment (Supplementary Material, Table S6). The results of adjusting all nine HLA risk alleles are shown in Supplementary Material, Table S7.

eQTL and colocalization analyses demonstrate an association between SNPs and gene expression

Colocalization analysis was conducted across five genes: *HLA-DRA*, *-DRB9*, *-DRB5*, *-DQA1* and *BRD2*. We found no significant colocalization when we assumed only one casual SNP in the region. The model assuming two casual SNPs in the region demonstrated that rs3 135 387 colocalized with expression levels of *HLA-DRA* in PBMC (CLPP = 0.003), BAL cell (CLPP = 0.002), lung tissue from GTEx (CLPP = 0.002) and whole blood from GTEx (CLPP = 0.001). Other nearby SNPs, rs3129888 and rs3 135 390, within the *HLA-DRA* region also showed significant colocalization posterior probability with PBMC gene expression (both SNPs CLPP = 0.002, Supplementary Material, Fig. S3) and lung tissue from GTEx (both SNPs CLPP = 0.001). A comprehensive cis-eQTL search of whole blood and lung from GTEx is shown in Table 5. In brief, rs9269233

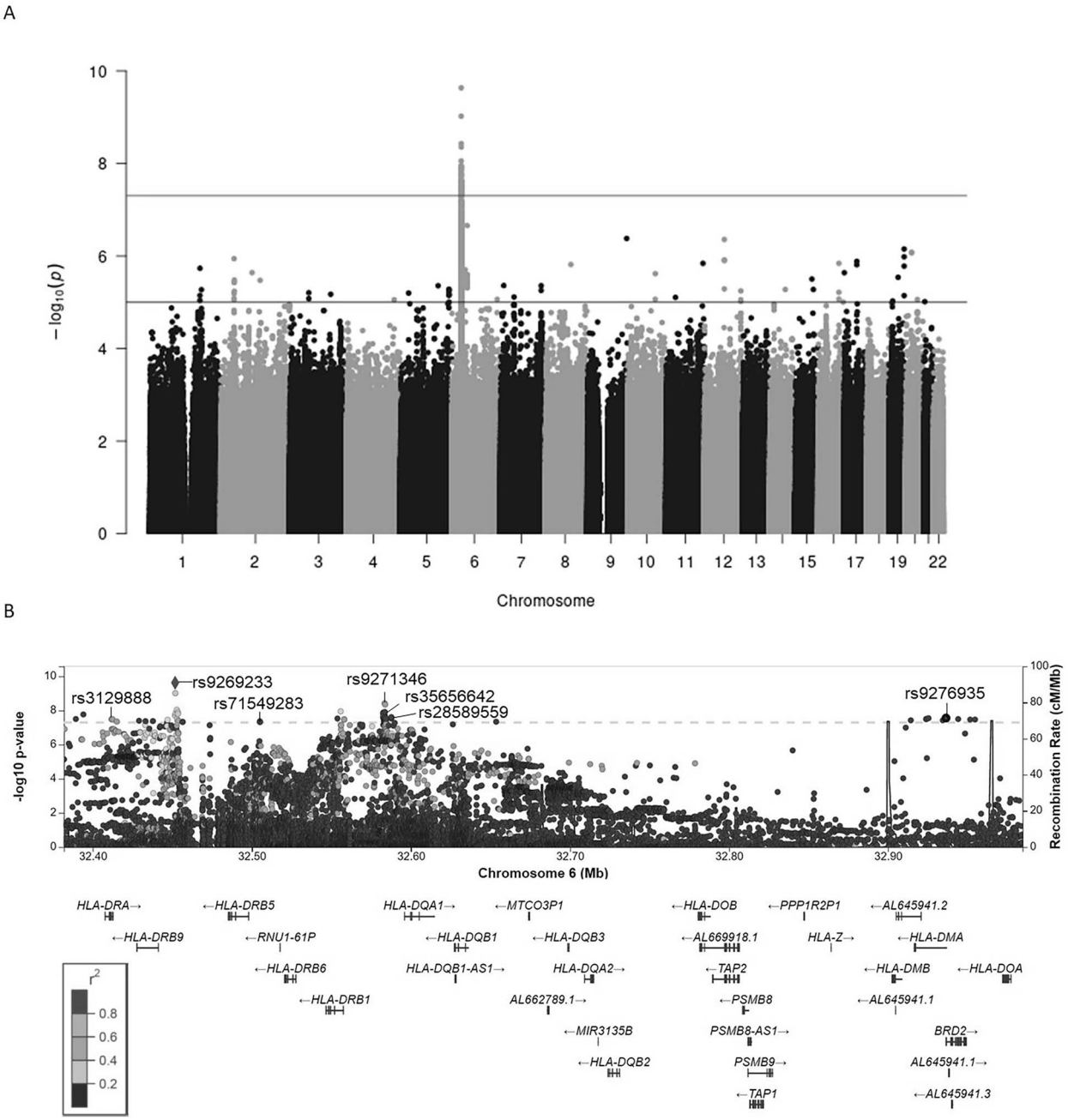

Figure 2. (**A**) Manhattan plot for SNP associations, with significant associations in the HLA region of chromosome 6. Upper line: genome-wide significance line ($P = 5 \times 10^{-8}$). Lower line: ($P = 10^{-5}$). (**B**) Locus-zoom plot for significant SNPs on chromosome 6 (position: 32381726–32 984 689); rs9269233 is the most significant SNP and r^2 values with other significant SNPs are low.

Table 4. *P*-values of the three genome-wide significant SNPs adjusted by HLA alleles (one at a time)

					Without HLA adjustment	**DRB1*0101**	**DQA1*0101**	**DQB1*0501**
SNP	Chr	Position	Minor allele	Nearest gene (+/-25 K)	*P*-value	*P*-value	*P*-value	*P*-value
rs9269233	6	32 451 762	A	HLA-DRB9	2.32E-10	2.22E-07	2.41E-07	1.35E-07
rs9271346	6	32 583 468	C	HLA-DQA1	3.73E-09	1.85E-07	5.93E-08	2.17E-08
rs35656642	6	32 583 610	A	HLA-DQA1	1.19E-08	7.22E-05	4.81E-05	1.87E-05
rs28589559	6	32 587 716	T	HLA-DQA1	2.46E-08	0.03	0.02	3.88E-03
rs9276935	6	32 936 441	C	BRD2 (inside gene)	2.72E-08	2.43E-07	1.46E-07	1.49E-07
rs3129888	6	32 411 726	G	HLA-DRA (inside gene)	3.21E-08	1.38E-06	1.45E-06	1.22E-06
rs71549283	6	32 505 038	A	HLA-DRB5	4.22E-08	4.26E-05	3.96E-05	4.66E-05

SNP: single nucleotide polymorphism; Chr: chromosome; MAF: minor allele frequency; OR: odds ratio; 95% CI: 95% confidence interval.

Table 5. Genotype-tissue expression portal (GTEx) *Cis*-eQTL in lung and whole blood for associated SNPs

SNP	Nearest gene (+/-25 K)	Common for lung and whole blood	Unique for lung[a]	Unique for whole blood[a]
rs9269233	HLA-DRB9	HLA-DQA2, HLA-DQB1, HLA-DQB1-AS1, HLA-DRB1, HLA-DRB5, HLA-DRB6	LY6G5C, STK19B	C4A, CYP21A2, HLA-DQB2, HLA-DRB9
rs9271346	HLA-DQA1	CYP21A2, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB1-AS1, HLA-DQB2, HLA-DRB1, HLA-DRB5, HLA-DRB6, HLA-DRB9	HCG23, XXbac-BPG154L12.4	LY6G5B, TNXB
rs35656642	HLA-DQA1	HLA-DQA1, HLA-DQA2, HLA-DRB5, HLA-DRB6, LY6G5B	HCG23, LY6G5C, XXbac-BPG154L12.4	CYP21A1P, HLA-DRB1
rs28589559	HLA-DQA1	HLA-DQA1, HLA-DQA2, HLA-DQB2, HLA-DRB6, HLA-DRB9	HLA-DOB, NOTCH4, TAP2	HLA-DQB1, PBX2
rs3129888	HLA-DRA (inside gene)	C4A, CYP21A2, HLA-DQA2, HLA-DQB1, HLA-DQB1-AS1, HLA-DRB5, HLA-DRB6, HLA-DRB9, STK19B	HCG23, HLA-DQA1, HLA-DRB1, XXbac-BPG154L12.4	HLA-DQB2

rs9276935 were not eQTLs in lung or whole blood tissue; rs71549283 was not found in the GTEx database as of 5/31/2022. [a]Unique gene with its gene expression affected by the specific SNPs only in the specific tissue. For example. rs9269233 only affects *STK19B* gene expression in the lung but not in the whole blood.

demonstrated significant eQTL for *HLA-DRB9* expression in whole blood, while rs9271346, rs35656642 and rs28589559 were significant eQTLs for expression of *HLA-DQA1* in whole blood and lung. The colocalization analysis for the three genome-wide significant HLA alleles showed that DRB1*0101 was the most significantly associated with sarcoidosis susceptibility and with expression levels in PBMC and BAL among sarcoidosis patients for *DRB1, DQA1, DQB1* and *DRB9* (all CLPPs>0.001). The DRB1*0101 risk allele was positively associated with PBMC *DQB1* gene expression ($P = 0.03$, Fig. 3A) and negatively associated with BAL *DRB9* gene expression ($P = 0.03$, Fig. 3B).

Discussion

We identified 49 SNPs ($P < 5 \times 10^{-8}$) associated with sarcoidosis in this largest EA GWAS of sarcoidosis, 7 of which represented distinct linkage disequilibrium groups; one of these SNPs, rs3129888, was also associated ($P < 0.007$) with sarcoidosis in an AA GWAS. All SNPs are on Chromosome 6 in the well-known HLA risk region. While we evaluated previously identified GWAS SNPs *a priori* and found an association ($P < 0.005$) with *ANXA11* and *BTNL2* (HLA region), we found no other SNPs outside the HLA region associated with sarcoidosis. Using colocalization and eQTL analysis, we found rs3 135 387 colocalizes with PBMC, BAL, lung tissue and whole blood gene expression of *HLA-DRA* in our EA population. HLA DRB1*0101, DQA1*0101 and DQB1*05*01 were the most significant HLA alleles ($P < 5 \times 10^{-8}$) associated with sarcoidosis among all 9 significant HLA alleles, with the other 6 alleles demonstrating a $P < 0.00011$, while DRB1*0101 was also associated ($P < 0.05$) with expression of PBMC *HLA-DQB1* and BAL *HLA-DRB9*. These results suggest that in sarcoidosis, the HLA region is likely functional, impacting gene expression and disease development.

The most significant SNP associated with sarcoidosis in our study, rs9269233, is between *HLA-DRB9* and *HLA-DRB5*. This association with rs9269233 was modestly attenuated after adjustment for three most significant sarcoidosis-related HLA alleles (adjusted $P = 2.15 \times 10^{-7}$), suggesting that rs9269233 is an independent risk allele for sarcoidosis in the region. The A allele of this SNP showed increased risk of sarcoidosis, has not been reported in other studies and was not significant in AA cohort. Potentially, this may indicate that sarcoidosis pathogenesis differs between EA and AA, and in fact, other studies have found different risk variants in EA and AA subjects (18). In a previous GWAS (18), rs1 964 995, also located between *HLA-DRB9* and *HLA-DRB5* (r^2 with rs9269233 = 0.53), showed a protective effect in non-Lofgren sarcoidosis vs. controls in a white Swedish cohort. Interestingly, rs1 964 995 increases the risk of rheumatoid arthritis (RA) in AA, although not in EA (19). rs9269233 has been found to be an eQTL for *HLA-DRB9* gene expression in whole blood in our analysis using the GTEx database. In addition to rs9269233, the presence of at least one DRB1*0101 allele was significantly associated with *HLA-DRB9* gene expression in BAL cells. The finding of an association with BAL gene expression may suggest processes that are occurring at the site of granulomatous inflammation or even the initial site of potential antigenic exposure; however, peripheral blood gene expression may represent signals that are being transmitted between organs or that reflect complex regulation systemically (20), some overlapping with those in the lung. *HLA-DRB9* is a pseudogene that is transcribed into RNA but does not encode proteins. However, pseudogenes can regulate other protein-coding genes (21); our findings suggest this may be the case with DRB9 in our population. In previous studies, *HLA-DRB* loci were found to group into five major haplogroups (DR1, DR8, DR51, DR52 and DR53), which differ by alleles at functional DRB genes (DRB3, DRB4 or DRB5) and DRB pseudogenes (DRB2, DRB6, DRB7, DRB8 or DRB9) (22). The HLA-DRs are associated with various diseases, including sarcoidosis (23), which is usually thought to have implications for antigen presentation. Our data suggests that these associations may reflect gene expression regulation that is not currently known to be functional but may impact immune responses in sarcoidosis.

Two other significant SNPs associated with sarcoidosis, rs9271346 and rs35656642, are both near *HLA-DQA1* (<25 K base pairs). rs9271346 was nominally significant in AA and EA in a previous study (5), with the same allele associated with risk in each, although it did not reach genome-wide significance ($P = 0.007$ in AA and 8.63×10^{-6} in EA). rs35656642 is a novel risk variant for sarcoidosis that has not been described previously, and we found the same allele associated with sarcoidosis in AA and also replicated in the UK Biobank European population ($P = 6.67 \times 10^{-8}$, with seven SNPs tested). Both SNPs are also eQTLs in lung tissue and whole blood based on publicly available datasets, although we did not find associations in lung BAL cells in our study. A study found rs2 187 668 near gene *HLA-DQA1* significantly associated with Lofgren's syndrome (18), while another reported three SNPs (rs28609302, rs9 273 113 and rs9 272 594) in this region associated with ocular sarcoidosis vs. controls

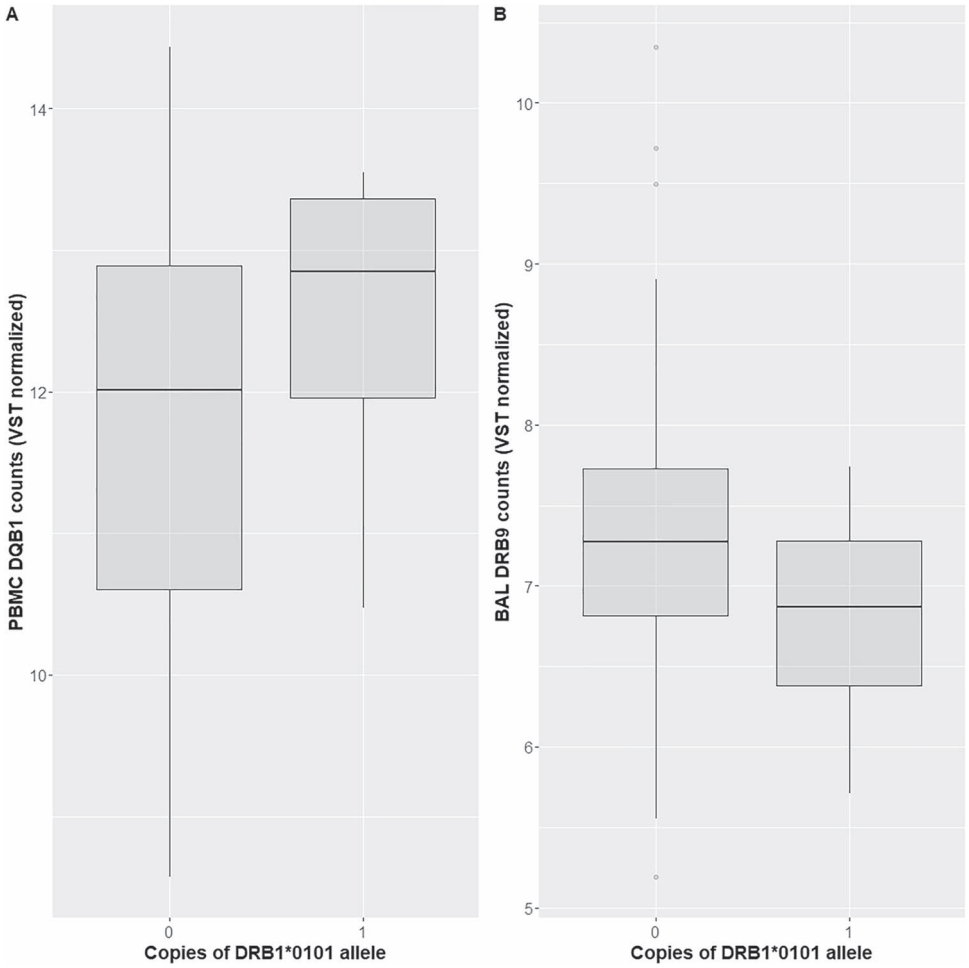

Figure 3. (**A**) PBMC DQB1 gene expression (variance stabilizing transformation [VST] normalized counts) and presence/absence of DRB1*0101 allele. Individuals with one DRB1*0101 allele have higher PBMC DQB1 gene expression compared with those with none. (**B**) BAL DRB9 gene expression (VST normalized counts) and presence/absence of DRB1*0101 allele. Individuals with one DRB1*0101 allele have higher BAL DRB9 gene expression.

in US AA(2×10–7 to 6×10^{-6}) (24). Of note, a subpopulation from the AA cohort was used in our study. We found low LD between those previously reported SNPs and the SNPs in our EA population (absolute r^2 0.06–0.58), indicating that the findings in our EA population are unlikely simple replications of previously identified SNPs.

Except for rs3129888 ($P = 1.52 \times 10^{-6}$), significant SNPs in EA were not significantly associated with sarcoidosis in AA. The rs3129888 G allele, located in an *HLA-DRA* intron, increased sarcoidosis risk. This finding is consistent with a study on Lofgren's syndrome in Sweden, Germany and a US AA population (18). Our US EA samples are distinct from the European cohort, but our AA population largely overlaps with the US AA population (18).

We found nine HLA alleles significantly associated with sarcoidosis ($P < 0.00011$). Among them, the top three HLA alleles are highly correlated with each other. DRB1*0101 and DQB1*0501 showed a protective effect on sarcoidosis risk, consistent with a previous study from the UK, Netherlands and Japan (25). In this previous study, DRB1*0101 was protective against lung-predominant sarcoidosis, Lofgren's syndrome and uveitis. In an AA study, DRB1*0101 also was protective for sarcoidosis risk (26). DQA1*0101 demonstrated protection in a Korean sarcoidosis population, similar to our study, although with statistical significance when adjusted for multiple comparisons (27). These results may suggest different genetic effects by ancestry. The sarcoidosis patients in our study included all subtypes of disease, although pulmonary involvement was the most prevalent. As previous studies have demonstrated the importance of alveolar CD4+ T cells in sarcoidosis and their association with specific HLA variants, presumably responding to antigen presentation HLA Class II (28), our finding of significant association with HLA class II alleles is not surprising. When we searched the database (29) from another large study using 13 835 EA individuals from five US sites of the Electronic Medical Records and Genomics (eMERGE) network (using the International Classification of Diseases code as the definition for diseases), DRB1*0101, DQB1*0501 and DQA1*0101 were all protective for sarcoidosis but increased risk of RA (Supplementary Material, Table S8). These opposite genetic effects are consistent with a study of pleiotropy between sarcoidosis and RA, which demonstrates a higher RA polygenic risk score associated with a protective effect of sarcoidosis (30), and another epidemiologic study demonstrating a lower prevalence of RA in a British sarcoidosis cohort vs. the general population (31). While RA and sarcoidosis are both inflammatory diseases, this may imply that their disease pathogenesis is distinct and drivers of one disease protect from development of the other. Each of the previously reported sarcoidosis HLA risk alleles, DRB1*1101, DRB1*1501 and DQB1*0602, showed a nominal increase in the risk of sarcoidosis in our study. DRB1*1101 allele was previously associated with increased risk of sarcoidosis in

AA and European descent individuals (14). DRB1*1501, which is in high LD with DQB1*0602, has been associated with increased risk of severe pulmonary sarcoidosis in individuals of European descent (13). Of note, these HLA alleles were not the strongest risk alleles in our study, and this might be owing to sub-populations (e.g. race/ethnicity) or varied phenotypes in our cohort compared with others. For example, our cohort is likely a mixture of different sarcoidosis phenotypes (e.g. cardiac, ocular, cutaneous, etc.) as we did not restrict enrollment to specific phenotypes. Despite a large sample size overall, the number of subjects with other phenotypes or organ involvement besides lung is small, limiting power for stratified analysis. Future GWAS studies would benefit from enrolling much larger numbers of patients and focusing on specific sarcoidosis phenotypes or those with specific organ involvement to explore genetic drivers of sarcoidosis manifestations, including neurological, cardiac or specific pulmonary phenotypes of sarcoidosis.

There are limitations to our study. First, we included a heterogeneous population of sarcoidosis patients with various phenotypes; this could reduce our study's power to identify novel SNPs versus European studies focused on Lofgren's syndrome. Regardless, we found HLA associations linked to other sarcoidosis phenotypes in previous studies, like DRB1*1101 and *1501. Second, while we have organ assessment on 80% of patients, at this time we do not have information on comorbidities or disease status over time (e.g. progression or remission), limiting our ability to examine their impact or association with our GWAS results. Third, we have gene expression data available on only a subset of participants for the eQTL analyses, impacting power to identify other eQTLs in our study. To help mitigate this limitation, we used the GTEx database to enhance our evaluation of associations between SNPs and gene expression.

In summary, our findings provide convincing evidence that HLA alleles are an important contributor to risk of sarcoidosis not only in our EA population but also in an AA population with genotyping already available. In addition to traditional GWAS SNPs and imputed HLA variants, we also explored how these genotypes are associated with gene expression in both PBMC and BAL cells using integrated analysis and demonstrated showed potential gene expressions impacted by these risk variants. Our future studies will explore these potential variants/genes and their mechanistic implications.

Materials and Methods

Study design and population

This GWAS used a two-phase approach to identify genome-wide significant SNPs associated with sarcoidosis; additional details are present in Supplemental Methods. DNA samples from sarcoidosis cases and controls were obtained from National Jewish Health (NJH), University of California, San Francisco, Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) consortium (17) and Cleveland Clinic Foundation. Sarcoidosis case definition was based on the ATS/ERS/WASOG statement (32). After quality control (see in the following), 818 cases and 981 controls (Phase 1) and 517 cases and 283 controls (Phase 2), all self-reported EA ancestry (Table 1 and Fig. 1) reflected the majority of the race/ethnicities seen in our clinics. A subset of study participants was part of GRADS study, and all with peripheral blood mononuclear cell (PBMC) and some with bronchoalveolar lavage (BAL) cell RNA sequencing data available (17,33). We tested significant SNPs identified in our EA two-phase approach in an AA study population. The AA GWAS summary statistics were obtained from a published study (5) with updated imputations since publication of those data.

Genome-wide genotyping

DNA was extracted from whole blood using the PAXgene Blood DNA kit. Genotyping was performed using the Illumina HumanOmini 2.5 BeadChip to interrogate ~2.4 million markers. The markers were derived from the 1000 Genomes Project (34), including all three HapMap phases, 19 K SNPs across the MHC and over 41 K non-synonymous SNPs. Genotyping was conducted at Hudson Alpha Biotechnology Institute (Huntsville, AL https://hudsonalpha.org/).

Genotype quality control

Samples were projected onto ancestry-informative principal components using 1000 genomes (34) populations to derive the SNP loadings for the PCs principal components (PCs) using Genome-wide complex trait analysis (GCTA) (35). We prioritized SNPs with minor allele frequency (MAF) > 0.03 and Hardy–Weinberg Equilibrium $P > 0.001$ in cases and controls evaluated separately and < 10% missing data.

Imputation of additional genotypes and HLA variants for the EA population

We imputed genotypes using combined case and control discovery samples for all 1000 genomes SNPs. We imputed classical HLA alleles using R package HLA Genotype Imputation with Attribute Bagging (HIBAG) (36).

RNA sequencing and quality control

Total RNA was extracted, and RNA sequencing was conducted as outlined (33). We followed a similar quality control procedure for both PBMC and BAL samples and removed RNA samples with an unmapped read rate > 20% and mitochondrial read rate > 0%. We removed outlier samples through PC analysis; the EA individuals who also had gene expression data available were included in the expression quantitative trait analyses.

Statistical analysis

Single-SNP association test and meta-analysis

We tested for association between each SNP and sarcoidosis using Snptest (v2) (37) as described previously (38) assuming an additive model with genotype dose and adjusted for sex and three PCs. To obtain an overall measure of association with sarcoidosis, we performed a meta-analysis of Phase 1 and Phase 2 using summary statistic data and the weighted inverse normal method (39) as implemented in the software METAL (40). Genome-wide significance was defined as meta-analysis $P < 5 \times 10^{-8}$. Genome-wide significant SNPs identified in our EA population were tested in the AA population. Statistical significance for these SNPs was defined as $P < 0.05$/(number of significant SNPs in EA). We compared our results with UK Biobank European sarcoidosis GWAS results (cases were defined as physician-diagnosed sarcoidosis) from Neale's laboratory (http://www.nealelab.is/uk-biobank). Statistical significance for these SNPs was defined as $P < 0.05$/number of SNPs tested). Finally, we also compared our GWAS results to previously identified loci in other studies including SNPs in *ANXA11* (rs1049550, rs1 953 600, rs2 573 346, rs2 784 773) (10), *RAB23* (rs1040461) (9), *C100RF67* (rs1398024) (41), *OS9* (rs1050045) (8), *CCDC88B* (rs479777) (7), *BTNL2* (rs2076530) (42) and *NOTCH4* (rs715299) (5). Statistical significance for these *a priori* SNPs was defined as $P < 0.005$ (0.05/10 SNPs).

Classic HLA alleles analysis

We used logistic regression models to test for association between dosage of each imputed HLA allele and sarcoidosis. Given strong *a priori* associations with the HLA region, we used $P < 0.00011$(0.05/448 HLA alleles tested) as statistical significance. We compared our HLA results to previously identified HLA alleles in other studies, including DRB1*1101, DRB1*1501 and DQB1*0602 (13,14). We used a $P = 0.016$(0.05/3 alleles) to determine statistical significance for *a priori* alleles found highly associated with sarcoidosis in prior studies.

Conditional models

To assess the independence of single-SNP effects from HLA risk alleles, we computed multivariable logistic regression models where HLA risk alleles were included as covariates in the model, and each SNP, one at a time, was tested for association (i.e. association adjusted for HLA risk alleles). We then adjusted all HLA alleles reach significant level of $P < 5*10^{-8}$ or $P < 0.00011$, respectively.

eQTL and colocalization analysis

For those sarcoidosis cases with gene expression data from GRADS, we performed colocalization analysis using eCAR (43) to identify variants with evidence for colocalization of disease and *cis* eQTL associations. The algorithm estimates the posterior probability that the same variant is casual in both GWAS and eQTL studies while accounting for linkage disequilibrium (LD). We included all SNPs within the defined gene boundary of the significant SNP in the analysis (+/-25 K base pairs) and tested for association between those SNPs and gene expression. The threshold for significance was set as a colocalization posterior probability (CLPP) > 0.001, as suggested by eCAR (43). The same approach was applied to imputed HLA alleles. We tested for association between HLA alleles and gene expression in two tissues (BAL and PBMC). In addition, using GRADS data, we conducted a comprehensive cis-eQTL search using the publicly available database, Genotype-Tissue Expression (GTEx) (44,45). The GTEx project was supported by the Common Fund of the Office of the Director of National Institutes of Health, NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. The data used for analyses described in this manuscript were obtained from the GTEx Portal on 05/31/2022 with the significant threshold defined as Q-value of 0.05 per GTEx database. Finally, we expanded our co-locolization analysis into lung and whole blood tissue using GTEx eQTL results.

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

Data availability

The full summary statistics are available upon making a request to the corresponding author.

Funding

We appreciate the grant support from National Institutes of Health Grants U01 HL112707, U01 HL112694, U01 HL112695, U01 HL112696, U01 HL112702, U01 HL112708, U01 HL112711, U01 HL112712, UL1 RR029882, UL1 RR025780, R01 HL110883, R01 HL114587 and R01HL114587; Clinical and Translational Science Institute Grant U54 9 UL1 TR000005; Centers for Disease Control and Prevention National Mesothelioma Virtual Bank for Translational Research Grant 5 U24 OH009077 and Foundation for Sarcoidosis Research (FSR).

Authors' contributions

TEF had full access to all data in the study and took full responsibility for the integrity of the data and the accuracy of the data analysis. SL, TEF and LAM wrote the manuscript. SL, TEF and LAM developed the analysis plan, and TEF supervised SL and SJ with the data analysis. TEF and LAM designed the study. CM provided the data collection and analysis of African American cohort. NYH, DAC, BB, P.M., KM, KP, DP, YSW, LK, CDL, SL (Leach) and EW were involved in the sample enrollment, processing and data collection. TEF and LAM supervised the research. We would like to thank all the participants of this study, as well as for the administrative support that we received from NJH and research coordination assistance from Christina Riley.

References

1. Erdal, B.S., Clymer, B.D., Yildiz, V.O., Julian, M.W. and Crouser, E.D. (2012) Unexpectedly high prevalence of sarcoidosis in a representative U.S. metropolitan population. *Respir. Med.*, **106**, 893–899.
2. Liu, H., Patel, D., Welch, A.M., Wilson, C., Mroz, M.M., Li, L., Rose, C.S., Van Dyke, M., Swigris, J.J., Hamzeh, N. *et al.* (2016) Association between occupational exposures and sarcoidosis: an analysis from death certificates in the United States, 1988-1999. *Chest*, **150**, 289–298.
3. Fingerlin, T.E., Hamzeh, N. and Maier, L.A. (2015) Genetics of sarcoidosis. *Clin. Chest Med.*, **36**, 569–584.
4. Rybicki, B.A., Iannuzzi, M.C., Frederick, M.M., Thompson, B.W., Rossman, M.D., Bresnitz, E.A., Terrin, M.L., Moller, D.R., Barnard, J., Baughman, R.P. *et al.* (2001) Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). *Am. J. Respir. Crit. Care Med.*, **164**, 2085–2091.
5. Adrianto, I., Lin, C.P., Hale, J.J., Levin, A.M., Datta, I., Parker, R., Adler, A., Kelly, J.A., Kaufman, K.M., Lessard, C.J. *et al.* (2012) Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PLoS One*, **7**, e43907.
6. Cozier, Y.C., Ruiz-Narvaez, E.A., McKinnon, C.J., Berman, J.S., Rosenberg, L. and Palmer, J.R. (2012) Fine-mapping in African-American women confirms the importance of the 10p12 locus to sarcoidosis. *Genes Immun.*, **13**, 573–578.
7. Fischer, A., Schmid, B., Ellinghaus, D., Nothnagel, M., Gaede, K.I., Schurmann, M., Lipinski, S., Rosenstiel, P., Zissel, G., Hohne, K. *et al.* (2012) A novel sarcoidosis risk locus for Europeans on chromosome 11q13.1. *Am. J. Respir. Crit. Care Med.*, **186**, 877–885.
8. Hofmann, S., Fischer, A., Nothnagel, M., Jacobs, G., Schmid, B., Wittig, M., Franke, A., Gaede, K.I., Schurmann, M., Petrek, M. *et al.* (2013) Genome-wide association analysis reveals 12q13.3-q14.1 as new risk locus for sarcoidosis. *Eur. Respir. J.*, **41**, 888–900.
9. Hofmann, S., Fischer, A., Till, A., Muller-Quernheim, J., Hasler, R., Franke, A., Gade, K.I., Schaarschmidt, H., Rosenstiel, P., Nebel, A. *et al.* (2011) A genome-wide association study reveals evidence of association with sarcoidosis at 6p12.1. *Eur. Respir. J.*, **38**, 1127–1135.

10. Hofmann, S., Franke, A., Fischer, A., Jacobs, G., Nothnagel, M., Gaede, K.I., Schurmann, M., Muller-Quernheim, J., Krawczak, M., Rosenstiel, P. *et al.* (2008) Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat. Genet.*, **40**, 1103–1106.
11. Levin, A.M., Iannuzzi, M.C., Montgomery, C.G., Trudeau, S., Datta, I., Adrianto, I., Chitale, D.A., McKeigue, P. and Rybicki, B.A. (2014) Admixture fine-mapping in African Americans implicates XAF1 as a possible sarcoidosis risk gene. *PLoS One*, **9**, e92646.
12. Levin, A.M., Iannuzzi, M.C., Montgomery, C.G., Trudeau, S., Datta, I., McKeigue, P., Fischer, A., Nebel, A. and Rybicki, B.A. (2013) Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun.*, **14**, 13–18.
13. Voorter, C.E., Drent, M. and van den Berg-Loonen, E.M. (2005) Severe pulmonary sarcoidosis is strongly associated with the haplotype HLA-DQB1*0602-DRB1*150101. *Hum. Immunol.*, **66**, 826–835.
14. Rossman, M.D., Thompson, B., Frederick, M., Maliarik, M., Iannuzzi, M.C., Rybicki, B.A., Pandey, J.P., Newman, L.S., Magira, E., Beznik-Cizman, B. *et al.* (2003) HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am. J. Hum. Genet.*, **73**, 720–735.
15. Nica, A.C. and Dermitzakis, E.T. (2013) Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.*, **368**, 20120362.
16. Ward, L.D. and Kellis, M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.*, **30**, 1095–1106.
17. Moller, D.R., Koth, L.L., Maier, L.A., Morris, A., Drake, W., Rossman, M., Leader, J.K., Collman, R.G., Hamzeh, N., Sweiss, N.J. *et al.* (2015) Rationale and design of the genomic research in Alpha-1 antitrypsin deficiency and sarcoidosis (GRADS) study. sarcoidosis protocol. *Ann Am Thorac Soc*, **12**, 1561–1571.
18. Rivera, N.V., Ronninger, M., Shchetynsky, K., Franke, A., Nothen, M.M., Muller-Quernheim, J., Schreiber, S., Adrianto, I., Karakaya, B., van Moorsel, C.H. *et al.* (2016) High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *Am. J. Respir. Crit. Care Med.*, **193**, 1008–1022.
19. Danila, M.I., Laufer, V.A., Reynolds, R.J., Yan, Q., Liu, N., Gregersen, P.K., Lee, A., Kern, M., Langefeld, C.D., Arnett, D.K. *et al.* (2017) Dense genotyping of immune-related regions identifies loci for rheumatoid arthritis risk and damage in African Americans. *Mol. Med.*, **23**, 177–187.
20. Koth, L.L., Solberg, O.D., Peng, J.C., Bhakta, N.R., Nguyen, C.P. and Woodruff, P.G. (2011) Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am. J. Respir. Crit. Care Med.*, **184**, 1153–1163.
21. Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L. and Carter, D.R. (2011) Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA*, **17**, 792–798.
22. Handunnetthi, L., Ramagopalan, S.V., Ebers, G.C. and Knight, J.C. (2010) Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes Immun.*, **11**, 99–112.
23. Abe, S., Yamaguchi, E., Makimura, S., Okazaki, N., Kunikane, H. and Kawakami, Y. (1987) Association of HLA-DR with sarcoidosis. correlation with clinical course. *Chest*, **92**, 488–490.
24. Garman, L., Pezant, N., Pastori, A., Savoy, K.A., Li, C., Levin, A.M., Iannuzzi, M.C., Rybicki, B.A., Adrianto, I. and Montgomery, C.G. (2021) Genome-wide association study of ocular sarcoidosis confirms HLA associations and implicates barrier function and autoimmunity in African Americans. *Ocul. Immunol. Inflamm.*, **29**, 244–249.
25. Sato, H., Woodhead, F.A., Ahmad, T., Grutters, J.C., Spagnolo, P., van den Bosch, J.M., Maier, L.A., Newman, L.S., Nagai, S., Izumi, T. *et al.* (2010) Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. *Hum. Mol. Genet.*, **19**, 4100–4111.
26. Dawkins, B.A., Garman, L., Cejda, N., Pezant, N., Rasmussen, A., Rybicki, B.A., Levin, A.M., Benchek, P., Seshadri, C., Mayanja-Kizza, H. *et al.* (2022) Novel HLA associations with outcomes of mycobacterium tuberculosis exposure and sarcoidosis in individuals of African ancestry using nearest-neighbor feature selection. *Genet. Epidemiol.*, **46**, 463–474.
27. Sikorova, K., Moon, S.J., Yoon, H.Y., Strnad, A., Song, J.W. and Petrek, M. (2022) HLA class II variants defined by next generation sequencing are associated with sarcoidosis in Korean patients. *Sci. Rep.*, **12**, 9302.
28. Judson, M.A. (2019) A sarcoidosis clinician's perspective of MHC functional elements outside the antigen binding site. *Hum. Immunol.*, **80**, 85–89.
29. McCarty, C.A., Chisholm, R.L., Chute, C.G., Kullo, I.J., Jarvik, G.P., Larson, E.B., Li, R., Masys, D.R., Ritchie, M.D., Roden, D.M. *et al.* (2011) The eMERGE network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genet.*, **4**, 13.
30. Lareau, C.A., DeWeese, C.F., Adrianto, I., Lessard, C.J., Gaffney, P.M., Iannuzzi, M.C., Rybicki, B.A., Levin, A.M. and Montgomery, C.G. (2017) Polygenic risk assessment reveals pleiotropy between sarcoidosis and inflammatory disorders in the context of genetic ancestry. *Genes Immun.*, **18**, 88–94.
31. Rajoriya, N., Wotton, C.J., Yeates, D.G., Travis, S.P. and Goldacre, M.J. (2009) Immune-mediated and chronic inflammatory disease in people with sarcoidosis: disease associations in a large UK database. *Postgrad. Med. J.*, **85**, 233–237.
32. Costabel, U. and Hunninghake, G.W. (1999) ATS/ERS/WASOG statement on sarcoidosis. sarcoidosis statement committee. American thoracic society. European respiratory society. world association for sarcoidosis and other granulomatous disorders. *Eur. Respir. J.*, **14**, 735–737.
33. Vukmirovic, M., Yan, X., Gibson, K.F., Gulati, M., Schupp, J.C., DeIuliis, G., Adams, T.S., Hu, B., Mihaljinec, A., Woolard, T.N. *et al.* (2021) Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. *Eur. Respir. J.*, **58**, 2002950.
34. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
35. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
36. Zheng, X., Shen, J., Cox, C., Wakefield, J.C., Ehm, M.G., Nelson, M.R. and Weir, B.S. (2014) HIBAG–HLA genotype imputation with attribute bagging. *Pharmacogenomics J*, **14**, 192–200.
37. Marchini, J. and Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet*, **11**, 499–511.
38. Fingerlin, T.E., Zhang, W., Yang, I.V., Ainsworth, H.C., Russell, P.H., Blumhagen, R.Z., Schwarz, M.I., Brown, K.K., Steele, M.P., Loyd, J.E. *et al.* (2016) Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for autoimmunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet.*, **17**, 74.

39. Fingerlin, T.E., Murphy, E., Zhang, W., Peljto, A.L., Brown, K.K., Steele, M.P., Loyd, J.E., Cosgrove, G.P., Lynch, D., Groshong, S. *et al.* (2013) Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.*, **45**, 613–620.
40. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
41. Franke, A., Fischer, A., Nothnagel, M., Becker, C., Grabe, N., Till, A., Lu, T., Muller-Quernheim, J., Wittig, M., Hermann, A. *et al.* (2008) Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. *Gastroenterology*, **135**, 1207–1215.
42. Valentonyte, R., Hampe, J., Huse, K., Rosenstiel, P., Albrecht, M., Stenzel, A., Nagy, M., Gaede, K.I., Franke, A., Haesler, R. *et al.* (2005) Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat. Genet.*, **37**, 357–364.
43. Hormozdiari, F., van de Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B. and Eskin, E. (2016) Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.*, **99**, 1245–1260.
44. Consortium, G.T. (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
45. Consortium, G.T. (2015) Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.