

Polygenic Risk Assessment Reveals Pleiotropy between Sarcoidosis and Inflammatory Disorders in the Context of Genetic Ancestry

Supplementary Notes

Caleb A. Lareau, Colby F. DeWeese, Indra Adrianto, Christopher J. Lessard, Patrick M. Gaffney, Michael C. Iannuzzi, Benjamin A. Rybicki, Albert M. Levin, Courtney G. Montgomery

Disease Base	African-American GWAS	European-American GWAS
Asthma	710,508	567,870
Celiac Disease	288,312	265,221
Crohn's Disease	740,852	647,712
Multiple Sclerosis	266,673	245,553
Primary Biliary Cirrihosis	478,738	438,866
Rheumatoid Arthritis	665,886	603,098
Systemic Lupus Erythematous	711,208	644,469
Type 1 Diabetes	349,927	322,393
Ulcerative Colitis	739,856	647,711

Table S1. Number of overlapping variants between summary statistics and Sarcoidosis GWAS datasets.

Disease Base	$P_T = 0.0001$		$P_T = 0.5$	
	p-value	R^2	p-value	R^2
Asthma	0.837	0.00%	0.297	0.05%
Celiac Disease	0.399	0.03%	0.390	0.04%
Crohn's Disease	0.681	0.01%	0.704	0.01%
Multiple Sclerosis	0.841	0.00%	0.120	0.12%
Primary Biliary Cirrihosis	0.043	0.20%	0.121	0.12%
Rheumatoid Arthritis	0.019	0.27%	0.330	0.05%
Systemic Lupus Erythematous	0.196	0.08%	0.570	0.02%
Type 1 Diabetes	0.475	0.02%	0.590	0.01%
Ulcerative Colitis	0.709	0.01%	0.429	0.03%

Table S2. Polygenic risk score results associating African-American Sarcoidosis including MHC in disease bases. The Nagelkerke pseudo- R^2 and pvalue for the association models are shown at two variant thresholds, $P_T = \{0.0001, 0.5\}$

Disease Base	$P_T = 0.0001$		$P_T = 0.5$	
	p-value	R^2	p-value	R^2
Asthma	0.968	0.00%	8.89E-09	2.03%
Celiac Disease	0.222	0.09%	8.21E-09	2.03%
Crohn's Disease	0.183	0.11%	0.243	0.09%
Multiple Sclerosis	0.005	0.49%	0.035	0.28%
Primary Biliary Cirrihosis	0.107	0.16%	2.01E-10	2.43%
Rheumatoid Arthritis	6.87E-06	1.31%	2.50E-17	4.32%
Systemic Lupus Erythematous	0.02	0.33%	0.28	0.07%
Type 1 Diabetes	0.894	0.00%	0.006	0.47%
Ulcerative Colitis	0.291	0.07%	0.282	0.07%

Table S3. Polygenic risk score results associating European-American Sarcoidosis including MHC in disease bases. The Nagelkerke pseudo- R^2 and pvalue for the association models are shown at two variant thresholds, $P_T = \{0.0001, 0.5\}$

Disease Base	$P_T = 0.0001$		$P_T = 0.5$	
	p-value	R^2	p-value	R^2
AA Sarcoidosis with MHC	0.368	0.00%	0.020	0.34%
AA Sarcoidosis without MHC	0.963	0.00%	0.020	0.34%
EA Sarcoidosis with MHC	0.077	0.02%	0.388	0.03%
EA Sarcoidosis without MHC	0.241	0.06%	0.404	0.03%

Table S4. PRS Results associating across Sarcoidosis ethnicities. The Nagelkerke pseudo- R^2 and pvalue for the association models are shown at two variants thresholds, $P_T = \{0.0001, 0.5\}$

As African-Americans have a combined genetic ancestry from both European and African individuals, we hypothesized that the proportion European ancestry would be associated with the pleiotropic risk from European-derived summary statistics across these disorders. To determine the effect of modulating the polygenic risk score by the proportion of European ancestry, we first fit model (1) for each of the PRSs from our 10 disease bases, including European-American sarcoidosis.

$$\text{logit}P(D = 1) = \beta_0 + \beta_1 PRS + \beta_2 EAadmix \quad (1)$$

The statistical significance of the model was computed from a likelihood ratio test against a null model of only β_0 . Next, we compared the proportion of variance explained (from Nagelkerke pseudo- R^2) and the significance of model (2) with the β_3 interaction term.

$$\text{logit}P(D = 1) = \beta_0 + \beta_1 PRS + \beta_2 EAadmix + \beta_3 (PRS * EAadmix) \quad (2)$$

Supplementary Tables (S5-S7) summarize the proportion of variance explained, the significance of models (1) and (2) against the null model from a likelihood ratio test, and the pvalue associated with the β_3 coefficient in model (2) for $P_T = \{0.0001, 0.5\}$.

Disease Base	(1) R^2	(1) pvalue	(2) R^2	(2) pvalue
Asthma	0.02%	0.7973	0.24%	0.1796
Celiac Disease	0.06%	0.5499	0.13%	0.4560
Crohn's Disease	0.03%	0.7476	0.08%	0.6405
Multiple Sclerosis	0.02%	0.8065	0.03%	0.8724
Primary Biliary Cirrhosis	0.21%	0.1172	0.21%	0.2316
Rheumatoid Arthritis	0.28%	0.0549	0.34%	0.0750
Systemic Lupus Erythematosus	0.09%	0.4167	0.54%	0.0115
Type 1 Diabetes	0.05%	0.6155	0.07%	0.6979
Ulcerative Colitis	0.02%	0.7957	0.17%	0.3268
European-American Sarcoidosis	0.18%	0.1655	0.73%	0.0018

Table S5. Summary statistics of models (1) and (2) comparing at the variant threshold $P_T = 0.0001$ used to associate disease status in African-American sarcoidosis.

Disease Base	(1) R^2	(1) pvalue	(2) R^2	(2) pvalue
Asthma	0.24%	0.0873	1.10%	4.62E-05
Celiac Disease	0.05%	0.6121	0.11%	0.5318
Crohn's Disease	0.02%	0.8153	0.07%	0.6774
Multiple Sclerosis	0.16%	0.1880	0.17%	0.3170
Primary Biliary Cirrhosis	0.017%	0.1662	0.74%	0.0016
Rheumatoid Arthritis	0.07%	0.5110	0.10%	0.5815
Systemic Lupus Erythematosus	0.02%	0.8227	0.82%	0.0008
Type 1 Diabetes	0.0006%	0.5662	0.27%	0.1403
Ulcerative Colitis	0.03%	0.7284	0.41%	0.0374
European-American Sarcoidosis	0.06%	0.5497	0.79%	0.0011

Table S6. Summary statistics of models (1) and (2) comparing at the variant threshold $P_T = 0.5$ used to associated disease status in African-American sarcoidosis.

Disease Base	β_1	β_2	β_3	β_1 pvalue	β_2 pvalue	β_3 pvalue
Asthma	-502.56	0.97	104.95	0.279	2.81E-03	0.0074
Celiac Disease	-350.50	-0.01	119.90	0.780	0.789	0.521
Crohn's Disease	4.63	-0.02	7.35	0.959	0.774	0.632
Multiple Sclerosis	1013.52	0.01	20.73	0.549	0.791	0.936
Primary Biliary Cirrhosis	-98.63	-0.21	201.29	0.908	0.126	0.050
Rheumatoid Arthritis	-233.59	0.01	-48.90	0.843	0.522	0.786
Systemic Lupus Erythematosus	-2257.13	-0.39	372.02	0.016	0.0011	0.0003
Type 1 Diabetes	-2201.35	-0.15	327.75	0.054	0.060	0.047
Ulcerative Colitis	323.85	0.10	-119.37	0.574	0.055	0.125
European-American Sarcoidosis	-247.67	-0.32	61.94	0.355	0.046	0.011

Table S7. Estimates of the β coefficients and univariate Wald Test p-values. The regression was computed under model (2) using a variant threshold of $P_T = 0.5$ for all variants including the MHC where available.

Disease Base	Asthma	Celiac Disease	Primary Biliary Cirrhosis	Rheumatoid Arthritis
Asthma		0.271	0.353	-0.348
Celiac Disease			0.030	-0.446
Primary Biliary Cirrhosis				-0.508

Table S8. Polygenic risk score correlations at $P_T = 0.5$ for EA sarcoidosis for the four strongest associations. All models include the MHC were available.

Disease Base	Asthma	Celiac Disease	Primary Biliary Cirrhosis	Rheumatoid Arthritis
Asthma		3.2%	3.3%	4.9%
Celiac Disease			3.0%	4.6%
Primary Biliary Cirrhosis				4.7%

Table S9. Pairwise multiple correlation coefficients at $P_T = 0.5$ for EA sarcoidosis for the four strongest associations. All models include the MHC were available.

ACTN4	ELMO1	LRRC18	RGS7
AFF3	ERAP1	MANBA	RIN3
ARFGEF2	EXTL2	METTL1	RPL3
ARID5B	FCRL3	MFSD6	SCHIP1
ATXN2	FRMD4A	MMEL1	SMC4
BRE	GCKR	NAB1	SPEF2
C1orf106	GNAT2	NCOA5	SPIB
CAPSL	GSDMB	OAS2	STARD3
CD226	ICAM3	PDE4A	STAT4
CD58	ICAM5	PFKFB3	SYNGR1
CD80	IKZF3	PHLDB1	TMEM39A
CDK4	IL12RB2	PLCL1	TNFAIP3
CLEC16A	IRF4	PLCL2	TNIP1
CLK1	KIF21B	PLEKHM1	UGT3A1
CNTLN	KIF5A	PPHLN1	WDFY4
CSE1L	LAYN	PPIL3	WDR37
CTDSP2	LBH	PPM1H	YAF2
CX3CR1	LOC100049716	PRICKLE1	ZBTB20
DGKQ	LOC100288310	PTPN2	ZNF687
DKFZp667F0711	LOC388210	PTPN22	
DNAJC4	LOC730109	RARB	
DOCK2	LRP1	RFTN2	

Table S10. List of all genes inputted for enriched pathway-based set analysis. These genes were mapped from the 238 significant ($FDR < 0.05$) variants determined by GPA between the sarcoidosis, asthma, celiac disease, primary biliary cirrihosis, and rheumatoid arthritis summary statistics. 85 unique genes shown above were mapped from the 238 variants using WebGestalt.

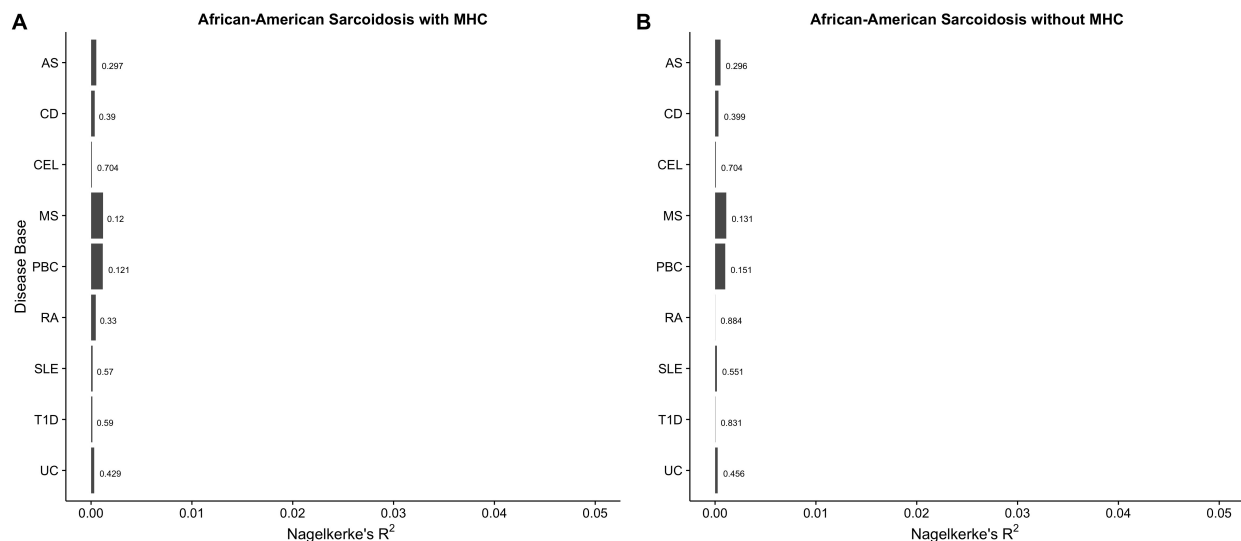


Figure S1. Summary of associations between nine inflammatory conditions and African-American sarcoidosis. (A) with the MHC (except Celiac Disease) and (B) without the MHC at a polygenic risk score variant threshold of $P_T = 0.5$

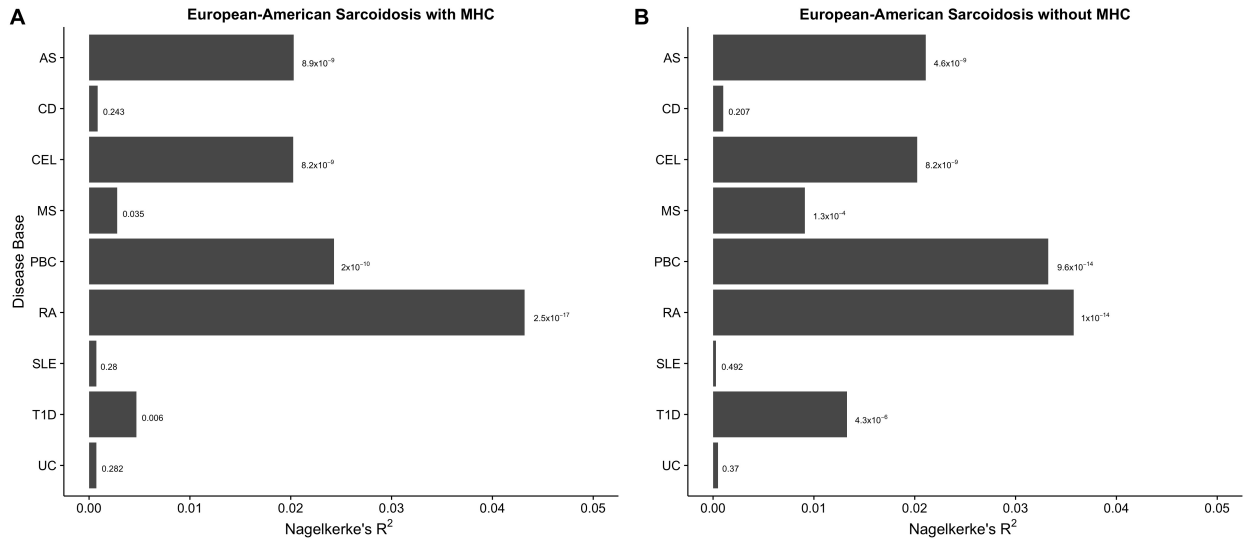


Figure S2. Summary of associations between nine inflammatory conditions and European-American sarcoidosis. (A) with the MHC (except Celiac Disease) and (B) without the MHC at a polygenic risk score variant threshold of $P_T = 0.5$

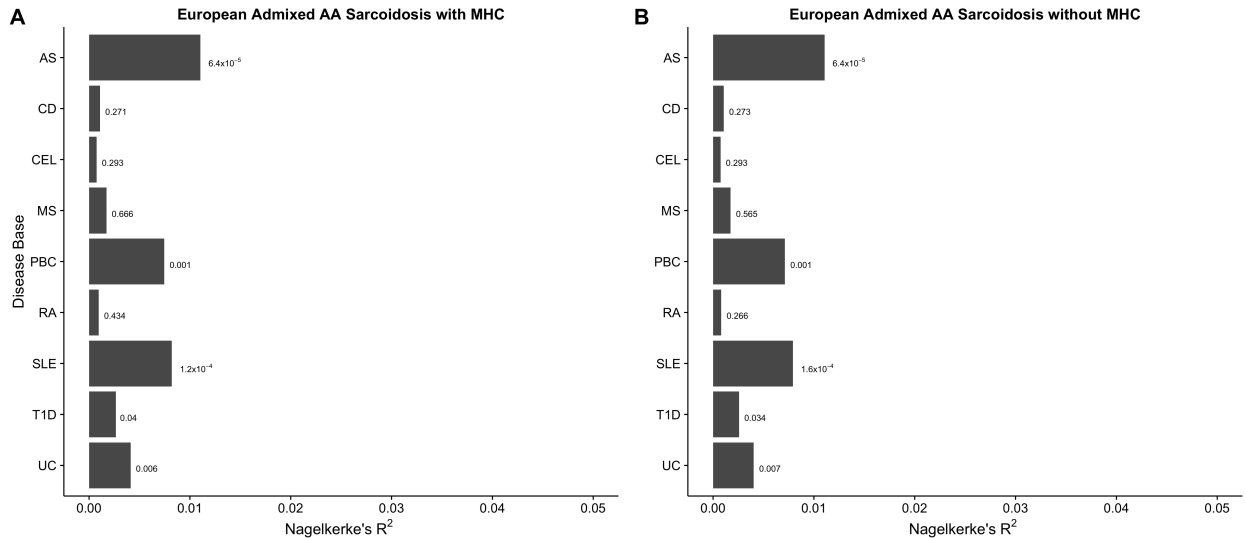


Figure S3. Summary of associations between nine inflammatory conditions and African-American sarcoidosis modified by European Ancestry (Model (2)). Panel (A) shows polygenic risk scores with the MHC (except Celiac Disease) and (B) without the MHC at a variant threshold of $P_T = 0.5$

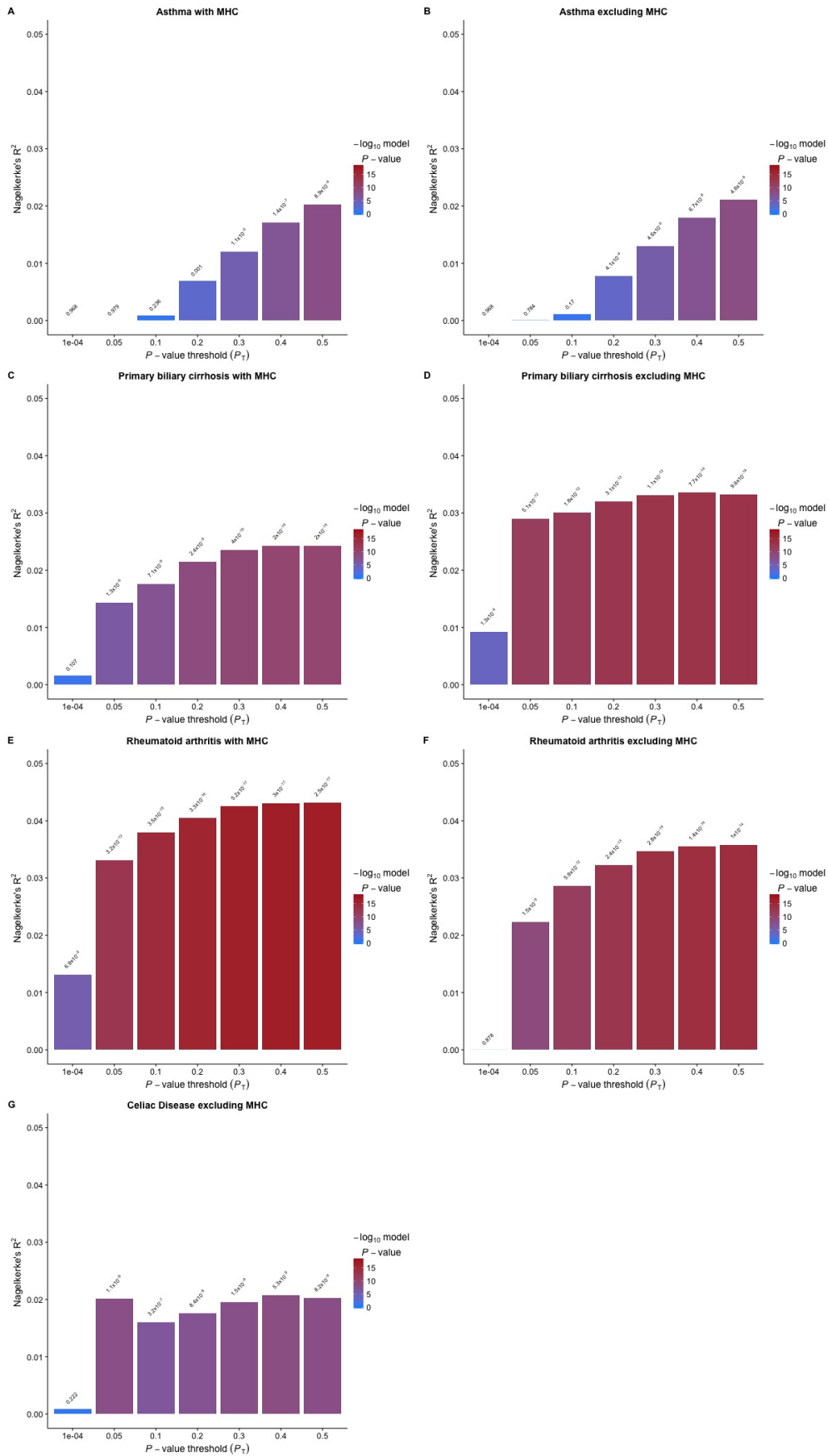


Figure S4. Summary of four most significant disease bases for European-American sarcoidosis with and without the MHC. Note that the MHC was not available in the Celiac Disease summary statistics.

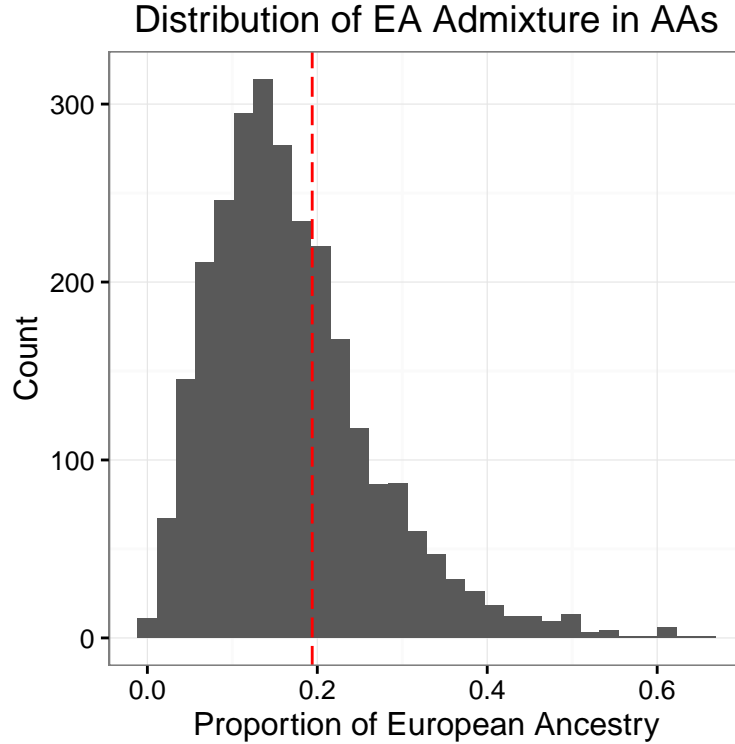


Figure S5. Histogram of the distribution of European Admixture in the African-American Sarcoidosis sample (N = 2,738) determined using LAMP. The red line indicates the cutoff of 0.194 used to define the top tertile of European Ancestry within the African-American cohort.

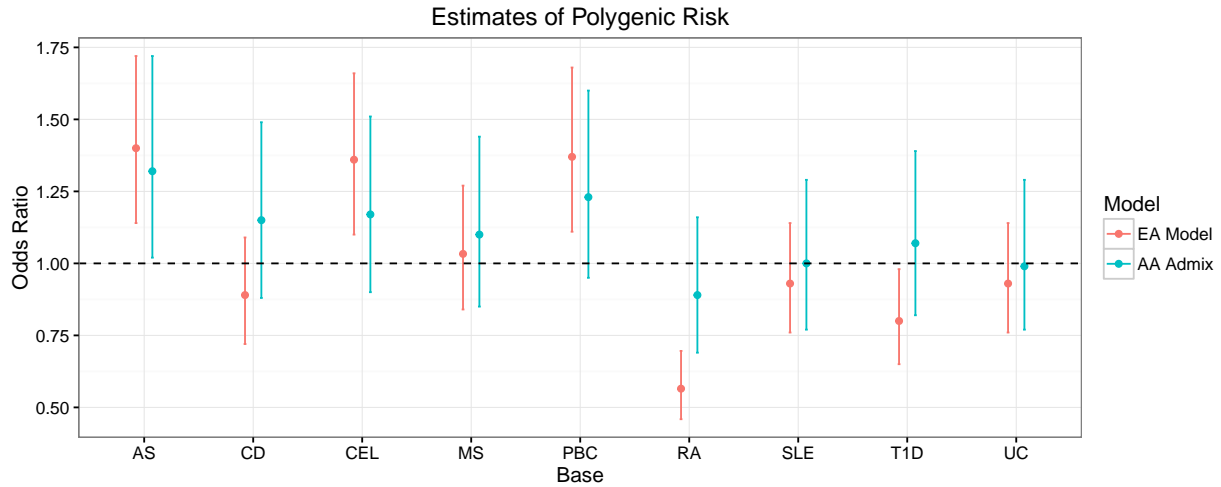


Figure S6. Full effect sizes of 9 polygenic risk bases. The odds ratio (OR) of sarcoidosis cases and controls was computed in the EA model using polygenic risk scores, dichotomized at the median value, as the predictor. In the AA admix model, the individuals in the top third of European admixture in our AA cohort were considered, and the polygenic risk scores were again used as the predictor, dichotomized at the median polygenic risk score of the high European admixture subset. The top third of European admixture from our AA sample contained 414 cases and 495 controls. Figure 3 is a subset of this figure for asthma, celiac disease, primary biliary cirrihosis, and rheumatoid arthritis.