THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A genome-wide admixture scan for ancestry-linked genes predisposing to sarcoidosis in African-Americans

OPEN ACCESS

A Genome-wide Admixture Scan for Ancestry-linked Genes Predisposing to Sarcoidosis in African Americans

Benjamin A. Rybicki[1,*], **Albert M. Levin**[1], **Paul McKeigue**[2], **Indrani Datta**[1], **Courtney Gray-McGuire**[3], **Marco Colombo**[2], **David Reich**[4], **Robert R. Burke**[5], and **Michael C. Iannuzzi**[6]

[1]Department of Biostatistics and Research Epidemiology, Henry Ford Hospital, Detroit, MI [2]Public Health Sciences Section, University of Edinburgh Medical School, Edinburgh, Scotland [3]Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK [4]Department of Genetics, Harvard University Medical School, Boston, MA [5]Division of Pulmonary Critical Care Medicine, Henry Ford Hospital, Detroit, M [6]Department of Medicine, Upstate Medical University, Syracuse, NY

Abstract

Genome-wide linkage and association studies have uncovered variants associated with sarcoidosis, a multi-organ granulomatous inflammatory disease. African ancestry may influence disease pathogenesis since African Americans are more commonly affected by sarcoidosis. Therefore, we conducted the first sarcoidosis genome-wide ancestry scan using a map of 1,384 highly ancestry informative single nucleotide polymorphisms genotyped on 1,357 sarcoidosis cases and 703 unaffected controls self-identified as African American. The most significant ancestry association was at marker rs11966463 on chromosome 6p22.3 (ancestry association risk ratio (aRR)= 1.90; p=0.0002). When we restricted the analysis to biopsy-confirmed cases, the aRR for this marker increased to 2.01; p=0.00007. Among the eight other markers that demonstrated suggestive ancestry associations with sarcoidosis were rs1462906 on chromosome 8p12 which had the most significant association with European ancestry (aRR=0.65; p=0.002), and markers on chromosomes 5p13 (aRR=1.46; p=0.005) and 5q31 (aRR=0.67; p=0.005), which correspond to regions we previously identified through sib pair linkage analyses. Overall, the most significant ancestry association for Scadding stage IV cases was to marker rs7919137 on chromosome 10p11.22 (aRR=0.27; p=2×10^{-5}), a region not associated with disease susceptibility. In summary, through admixture mapping of sarcoidosis we have confirmed previous genetic linkages and identified several novel putative candidate loci for sarcoidosis.

Introduction

Sarcoidosis, a multiorgan granulomatous inflammatory disease, likely results from an exaggerated T cell response to an airborne antigen 1. The first genome-wide linkage study of sarcoidosis in 63 German families found linkage of sarcoidosis to chromosome 6p212. Subsequently, a novel sarcoidosis risk variant at rs2076530 was identified, which predicted a premature splice site at exon 5 of the class II gene *BTNL2*, a B7 family co-stimulatory molecule3. The increased sarcoidosis risk associated with this variant was replicated in another German population 4, by our group in European Americans 5, and in a Dutch and

*Corresponding author .

Conflict of Interests The authors declare no conflict of interests.

British population 6. We also studied variation at exon 5 of *BTNL2* in African-American sarcoidosis cases and controls and found variants other than rs2076530 associated with sarcoidosis 5.

Additional genome-wide studies have confirmed that sarcoidosis is a complex genetic disease with multiple loci having modest effects. In the Sarcoidosis Genetic Analysis (SAGA) study 7 of 516 African-American sib pairs with at least one affected sib and 359 pairs with both sibs affected, our most significant linkage peaks spanned chromosomes 5p15.2 and 5q11.2 8. A second analysis of this same study population using disease sub-phenotypes found that chest radiographic resolution had one of the highest LOD scores among the pulmonary and extrathoracic involvement phenotypes analyzed 9. The latest evidence for a genetic predisposition to sarcoidosis comes from the first sarcoidosis genome-wide association scan (GWAS), conducted in a German population, that had its most significant disease association map to Annexin A11 (*ANXA11*), a gene with complex and essential functions in several biological pathways, including cell apoptosis and proliferation 10.

Study of the genetics of sarcoidosis in African Americans, who are more commonly and severely affected by this disease 11, suggests that focusing on genes of African ancestry may uncover novel risk variants. A higher incidence of sarcoidosis in African Americans compared with whites has long been noted 12,13 with an over three-fold increased incidence of sarcoidosis in African Americans compared with whites, 35.5 per 100,000 vs. 10.9 per 100,000 11. Other studies outside of the United States that compared sarcoidosis incidence in populations of West African ancestry with whites further support a role of African genes in sarcoidosis susceptibility. In South Africa, sarcoidosis incidence was reported as 23.2 per 100,000 in blacks and 3.7 per 100,000 in Whites 14. In Great Britain, Edmondstone and Wilson 15 reported an annual sarcoidosis incidence of 19.8 per 100,000 in blacks and 1.5 per 100,000 in whites. The eight to ten fold higher incidence of sarcoidosis in blacks in the South African and Great Britain studies is notable because of decreased European ancestry in these two populations compared with African Americans. Average European admixture proportions are about 10 percent in African Caribbeans living in Great Britain compared with 18 percent for African Americans 16. Among the disease outcomes in sarcoidosis that have a strong link with African ancestry, persistence of radiographic evidence for disease among patients of African descent appears to be a consistent finding 15. In a multi-site case control study, after two years of observation, African-American cases were 80% more likely to have a worsening chest x-ray, although this finding was not statistically significant 17. In general, self reported African ethnicity is strongly related to increased sarcoidosis risk and more severe disease outcomes.

The recent development of genome-wide ancestry informative marker panels 18,19 now makes scanning the genome for disease genes linked to ancestry in African-American populations feasible. Mapping by admixture linkage disequilibrium (MALD) involves screening the genome of individuals of mixed ancestry who have a disease or trait of interest for chromosomal regions that have a greater percentage of alleles than expected from a parental population. The theory behind MALD is that affected individuals are more likely to carry alleles from the higher risk parental population in chromosomal segments that also harbor disease loci. Advantages to using admixture mapping over genome-wide association mapping include the ability to find effects where there is marked allelic heterogeneity across ethnic groups 20 or where a risk allele is fixed in either parental population. Admixture mapping can also more readily detect risk effects emanating from genes due to rare copy number variants than can a dense SNP map relying only on allelic association. MALD in African-American populations has been used to localize susceptibility loci for prostate

cancer 21, kidney disease 22, peripheral artery disease 23, type 2 diabetes 24, obesity 25,26, interleukin-6 levels 27 and blood lipid levels 28.

To determine whether African ancestry is linked with sarcoidosis genetic susceptibility, we conducted a whole genome admixture scan using a panel of ancestry-informative markers in a large sample of African-American sarcoidosis cases and unaffected controls. We hypothesized that ancestry-linked genes may predispose to either disease susceptibility and/ or severity, as measured by radiographic progression.

Results

Overall Results

The MALD scan estimated associations between increased or decreased African ancestry based on 1,384 SNPs typed on a sample of African-American unrelated sarcoidosis cases (n=1,026) and controls (n=316). The mean African ancestry estimate of the analytic sample was 82.6% with sarcoidosis cases having a slightly higher level of African ancestry than controls (82.9% vs. 81.5%; p=0.03). We detected nine distinct chromosomal regions that had suggestive ancestry associations with sarcoidosis: four with excess African and five with excess European ancestry (table 1). These regions were identified based on affected-only test p values < 0.01 and the same ancestry differential based on the ancestry association test (figure 1). For consistency, the ancestry risk ratios (aRRs) estimated from the more powerful affected-only test are reported in the text. While none met the genome-wide significance level of 10^{-5} 29, the most prominent ancestry linkage peak located at chromosome 6p24.3-6p12.1 had as its most significant marker rs11966463 with an aRR of 1.90 and a Z score of 3.78 (p=0.0002), indicative of increased risk of sarcoidosis associated with excess African ancestry at this locus. The second most prominent peak was on chromosome 17p13.3-17p13.1 with marker rs8070464 having an aRR of 1.68 (Z = 3.51; p=0.0005) also indicative of increased risk of sarcoidosis associated with excess African ancestry. The other two regions that showed excess African ancestry among sarcoidosis cases were chromosomes 4q31.21-4q34.1 and 5p13.3-5p13.2. The five regions that showed excess European ancestry among sarcoidosis cases were chromosomes 2p13.3-2p13.1, 2q35-2q26.3, 5q23.1-5q31.2, 6q23.3-6q25.2 and 8p21.3-8p11.21. Chromosome 8p21.3-8p11.21 was the most significant of the five (p=0.002), but all were comparable in magnitude. In general, ancestry association Z-scores were, as expected, of lower magnitude than affected-only test Z scores, but in the same direction with the same peak SNPs (figure 1). Of note, was a 10 mB region on chromosome Xq24-Xq25, which had the most statistically significant affected-only test p value, but none of the SNPs in this peak attained our ancestry association p value criterion of 0.05. Analyses were rerun for the subset of 910 unrelated cases with a definite (i.e., histologically confirmed) diagnosis of sarcoidosis, and the results were nominally changed (table 2). In these analyses, the most significant finding remained at rs11966463 and increased in significance to p=0.00007. Results in the definite case group for the other eight regions were nominally changed with the possible exception of chromosome 8p21.3-11.21, where the aRR increased from 0.67 to 0.77 and the p value increased from 0.003 to 0.06.

Results in Demographic and Phenotypic Subsets

Of the nine sex-specific aRRs shown in table 3, only the sex-specific aRRs in the 5q23.1-5q31.2 region were statistically significantly different from each other (p=0.003). Female cases had ancestry risk ratios of greater magnitude than male cases for six of the nine ancestry peaks, with the only exceptions being chromosomes 5p13.3-5p13.2, 6q23.3-6q25.2, and 17p13.3-17p13.1 where the female-specific risk ratios were of slightly lower magnitude than for males. For four of the six peaks where the female-specific aRR

was of greater magnitude (chromosomes 2q35-2q36.3, 4q31.21-4q34.1, 5q23.1-5q31.2 and 6p24.3-6p12.1), the excess ancestry risk appeared to be almost exclusive to female subjects. In addition to the loci identified in the unstratified analysis, marker rs2476230 on chromosome 13q32.2-13q33.3 had an elevated aRR (1.60; p=0.003) only in female subjects and marker rs1131877 on chromosome 14q32.2-14q32.33 had an elevated aRR (2.63; p=0.002) only in male subjects.

When stratifying cases by chest radiographic phenotype and analyzing using the affected-only test, seven additional distinct ancestry peaks that were detected, including chromosomes 2p24.3-2p16.1, 4q21.1-4q24, 10p12.2-10q11.23, 16q22.1-1623.2, 18p11.32-18p11.31, 18q23 and 19p12-19q13.41 (table 4). Of the nine original ancestry peaks found in the full case sample, four (2p13.3-2p13.1, 5p13.3-5p13.2, 6q23.3-6q25.2 and 17p13.3-17p13.1) had aRRs of much greater magnitude and more significant associations in the more severe Scadding stage IV case group, despite this case subset being less than 20 percent of the size of the full case sample. A formal test for heterogeneity revealed three loci with significant differences in aRRs between the extreme radiographic phenotypes of stage IV sarcoidosis and radiographically resolved disease (table 5). All three loci were European ancestry-associated loci in the full case group with much stronger associations with European ancestry in stage IV cases. For example, in stage IV chest radiograph cases, chromosome 10p11.22 had the aRR of greatest magnitude (aRR=0.27; $p=2 \times 10^{-5}$) for any of the phenotypic subsets analyzed.

Finally, we ran several validity checks on our results. The main analysis for disease susceptibility was redone using ANCESTRYMAP 30 with resulting African and European ancestry associations virtually identical to what we found using ADMIXMAP. We also redid the main analysis for 100 resamples of our full dataset, in which we relaxed the sampling criteria such that any individual irrespective of affection status could be randomly selected from a family. Analysis of these 100 different samples of unrelated individuals showed essentially the same patterns of ancestry associations across the genome.

Exclusion Map

We prepared an exclusion map for the inclusive (definite and highly probable; n=1,026 cases) case definition (supplementary figure 1) based on the affected-only test results. At least 90% of the genome can be excluded as having a European effect on risk of 2.12 or more, and at least 90% can be excluded as having an African effect on risk of 2.10 or more. Further, among the subset of definite cases (n=910), at least 90% of the genome can be excluded as having a European effect on risk of 2.23 or more, and at least 90% can be excluded as having an African effect on risk of 2.19 or more. The power of the subanalysis for definite cases is less than that for all cases because of the smaller sample size.

Discussion

We used admixture mapping to search the genome for sarcoidosis susceptibility genes linked to ancestry in African Americans. The chromosomal regions we identified complement and confirm previous genomic regions identified by past genome-wide linkage 31,8 and association 10 studies in African Americans and whites. The chromosome 5p13.3-5p13.2, 5q23.1-5q31.2 ancestry linkage signals overlap with the chromosome 5 sarcoidosis linkages we previously reported in African American sib pairs 8. Moreover, the chromosome 5p and 5q regions have also been implicated in Crohn's disease. The chromosome 5p13 region linked to Crohn's disease has been characterized as a 1.25 megabase gene desert, but disease associated variants in this region are associated with expression levels of *PTGER4* 32, the closest gene that maps to this region. Chromosome 5q31 contains the cytokine gene cluster with unique variants associated with Crohn's

disease 33 and psoriasis 34,35. The chromosome 6p24.3-6p12.1 ancestry peak overlaps with the chromosome 6p21 sarcoidosis linkage signal originally reported by Schurmann et al. 31. We also found novel regions containing putative sarcoidosis susceptibility loci on chromosomes 2p13.3-2q12.1, 2q35-2q36.3, 4q31.21-4q34.1, 6q23.3-6q25.2, 8p21.3-8p11.21 and 17p13.3-17p13.1.

Of the nine ancestry linkage peaks we identified in the overall risk analysis, the African ancestry peak on chromosome 6p24.3-6p12.1 was the most statistically significant (p=7 × 10^{-5} for rs11966463 in definite cases) and was close to the genome-wide significance level of 10^{-5} for admixture mapping 29. This region spans 50 megabases and includes the entire major histocompatibility complex on chromosome 6p21. Variants of the class II HLA-DR and DQ genes are located in this region and known to be associated with sarcoidosis susceptibility and progression 36,37,38,39. In addition, the exon 5 SNP rs2076530 of the BTNL2 gene 200 kilobases telomeric from HLA-DRB1 was shown to be the primary disease-associated variant in a German sample that was driving the linkage signal previously identified in this sample 3. Increased allelic heterogeneity across the HLA region in African Americans 40,41 was the likely explanation for why we replicated this association in the ACCESS white case-control sample but not in two different African-American samples 5. The numerous previously reported associations of genetic variation in the HLA region with sarcoidosis and the linkage to this region suggest that the prominent African ancestry peak at 6p22 we observed in African-American sarcoidosis cases could be due to underlying disease causing genes in the HLA region.

To appropriately estimate disease-ancestry linkage on the X chromosome, we had to account for the known unequal contribution of male and female European founders to the present day African-American gene pool 42,43. In our study sample, we estimated ψ, the female/ male odds ratio for European versus African founders, to be 3.1, which falls within the range of ψ estimates (2.3-4.3) derived from European admixture estimates based on mitochondrial and Y chromosome lineages in other African-American samples. Despite the incorporation of ψ into the estimation of X chromosome admixture proportions, we still found two broad peaks on the X chromosome at which the observed African ancestry on the X chromosome of cases was much higher than the expected value given autosomal admixture proportions and ψ. A similar excess of African ancestry was observed over the same regions in controls, however, suggesting that the frequency of African ancestry in these regions is higher than elsewhere as a result of drift or selection for African derived DNA on the X chromosome in African Americans rather than a true disease association.

We failed to find evidence for increased African or European ancestry influencing sarcoidosis susceptibility on chromosome 10q22.3. This is the region containing the *ANXA11* (annexin A11) gene with a common nonsynonymous SNP (rs1049550, T > C, R230C) found to be the most strongly associated sarcoidosis SNP in a German GWAS of sarcoidosis susceptibility 10. None of the *ANAX11* sarcoidosis associated variants show significant allelic diversity between African and European populations. This highlights the main limitation of admixture mapping in that putative disease risk variants with similar frequencies in the ancestral African and European populations that comprise the current day African-American population are undetectable by admixture mapping.

A female preponderance in sarcoidosis cases is well documented 11,44 and some evidence suggests that female cases may also incur greater granuloma burden in early stage disease 45. In terms of underlying genetic liability, polygenic complex disease theory is that individuals with a lower population risk who contract the disease in question (e.g., males for sarcoidosis) have on average more genetic liability 46. We previously found a greater familial risk associated with African-American male sarcoidosis cases compared with

females 47, but the ACCESS case-control study consisting of both white and African-American subjects found no difference in sex-specific familial risk 48. Others have reported sex-specific effects of alleles predisposing to sarcoidosis susceptibility or disease manifestations 49,50,51,52. Our sex-stratified analyses showed that some ancestry linkage signals were consistent for males and females, whereas others were sex-specific. Ancestry signals on chromosomes 2q35-2q36.3, 4q31.21-4q34, 5q23.1-5q31.2, 6p24.3-6p12.1, and 13q32.2-13q33.3 appeared to be exclusive to females, whereas only one significant signal was exclusive to males, on chromosome 14q32.2-14q32.33, however the aRR at this locus at 2.63 was of the greatest magnitude of the sex-specific ancestry signals. Given the 3:1 ratio in female to male sarcoidosis cases in our analytic dataset, it is unclear whether the abundance of significant ancestry signals in female-only analyses simply reflected a greater statistical power or a more robust genetic effect. The etiology of sex differences in sarcoidosis remains unknown, but since female and male sex hormones affect the immune system in different ways 53, one might suspect that some sex-specific genetic effects underlying sarcoidosis risk exist. In terms of genetic risk linked with ancestry, our results suggest that focusing on female cases may provide a more direct path toward sarcoidosis gene finding.

Consideration of phenotype plays a critical role in the search for sarcoidosis susceptibility genes. For instance, the acute form of sarcoidosis known as Lofgren's syndrome has a strong association with a distinct HLA subtype 54. Other studies have shown clear differences in genetic predisposition to acute and chronic forms of sarcoidosis 55,37. In a reanalysis of the SAGA sib pair linkage data according to phenotype 9, we showed that phenotypes defined by specific pulmonary manifestations of disease and affected organs had linkage signals that were different from those based on disease status alone, but more importantly the linkage signals we found were of greater magnitude suggesting a less diluted genetic effect.

Persistence of radiographic defined lung disease is a sarcoidosis phenotype that appears to have a strong association with African ancestry 15,17,56. We analyzed case subsets defined by radiographic persistence of disease. For those with radiographic persistence, we also analyzed the subset of cases with evidence of fibrotic (stage IV) disease. In doing so, we found that several loci had aRRs for stage IV disease that were of greater magnitude than any found for disease susceptibility with the strongest ancestry association signal at chromosome 10p12.2-10q11.23 (aRR=0.27; p=0.00002). Stage IV cases were more than three times more likely to carry two European alleles at this locus in comparison to none (i.e. two African alleles). A recent combined genome-wide association study of Crohn's disease and sarcoidosis cases found the SNP with the strongest association at chromosome 10p12.2 57. The results of these phenotype-stratified analyses show the genetic heterogeneity of sarcoidosis and the need to stratify cases into more homogeneous subsets in gene finding efforts.

Our study had several limitations. The African-American sample we analyzed was recruited from sites across the United States, which likely contributed to a greater variation in European admixture proportions than would be observed if cases were drawn from a more circumscribed geographic area 43. Although African admixture proportions varied between 28 and 99% within the study sample, variation in average African admixture was similar across study sources. Ascertainment schemes also varied across the three source populations, with two of the three studies ascertaining on families. In practice, having a majority of cases with relatives allowed Mendelian inconsistent genotypes to be removed resulting in a more accurate analytic set of genetic data. Admixture mapping cannot detect disease variants with similar allele frequencies in the two parental populations, but disease variants differentially fixed in the parental populations may only be detected through admixture mapping. In addition, admixture mapping is robust to allelic heterogeneity, an

issue that arose in trying to replicate *BTNL2* associations in African-American samples 5. The greater HLA diversity in African-Americans may have limited our ability to find sarcoidosis linkage to this region using traditional methods, but the strong African ancestry signal on chromosome 6p22.2 suggests African alleles in this region may indeed play a role in sarcoidosis susceptibility.

We did find one African ancestry-disease associated locus at chromosome 6p24.3-6p12.1 that was close to the genome-wide significance level of 10^{-5} 29 in cases with histologically confirmed disease, but our results suggest that multiple ancestry-linked loci exist harboring genes with alleles of both African and European origin that predispose to sarcoidosis. Next steps will include prioritizing ancestry-linked regions for fine mapping based on level of association and consistency of results in phenotypic subsets. While several ancestry peaks were found at chromosomal regions where known sarcoidosis genes exist, most were in regions that, if due to true genetic effects, may reveal novel sarcoidosis genes that help us better understand and treat this perplexing disease. Loci identified by this ancestry scan may also be relevant in other immune-related diseases, particularly where African-American race is a disease risk factor. Finally, understanding the effects of ancestry on locus-specific risk should aid in future sarcoidosis gene finding efforts in African Americans, such as genome-wide association studies, which will attempt to isolate modest genetic effects possibly confounded by ancestry.

MATERIALS AND METHODS

Study Sample and Case Phenotyping

DNA samples and subject data were collected from three previously conducted studies: 1) a multi-site case control study 58; 2) a multi-site affected sib-pair study 7 and; 3) a single institution family-based study 38. The current study, as well as the three previous studies all underwent human subjects review with the original informed consents allowing usage of data and genetic material for future studies. All study participants self-identified as being African-American, with the single-institution study drawing participants from the Detroit, Michigan area, whereas the two multi-site studies drew participants from across the United States. Figure 2 depicts the initial sample composition submitted for genotyping and how the final analytic sample of 1,357 sarcoidosis cases and 703 controls was derived. Of the 2,494 samples genotyped, 15 were excluded because of poor genotyping quality. Another 15 samples were excluded because the reported sex did not match genotyped sex. A total of 108 samples were found to be duplicated across studies (i.e., the same individual participated in more than one study), and therefore the duplicate sample(s) with the poorer genotyping quality was dropped from the analytical dataset. Nine samples were dropped due to allele-sharing with another sample that indicated unreported monozygotic twin status. Finally, 10 samples were dropped either for having no genes of European or African ancestry or for having evidence of the first European admixture within the last generation and, therefore, uninformative for MALD as they posses no admixed gametes.

Among the 1,357 cases, 717 (52.8%) were members of nuclear families with one or more sibs and/or parents genotyped. Because of the different ascertainment schemes for the two family studies that contributed to the final analytic sample, sibships included both single and multiple affected individuals. When available, parents were genotyped, but approximately half of the nuclear families (50.7%) had no genotype data on parents.

The diagnostic criteria for sarcoidosis varied among the different study populations from which the samples were derived. Overall, for an individual to be classified as a sarcoidosis case, he or she had to satisfy the diagnostic criteria for definite or highly probable sarcoidosis used in the SAGA study 9. In short, definite cases had histological confirmation

of non-caseating granulomas and evidence of disease involvement in either the thorax or two or more other organ systems. Highly probable cases lacked histologic confirmation, but had characteristic chest radiographs (bilateral symmetrical hilar adenopathy) and either a history of erythema nodosum or at least two years observation during which time no other disease was found to explain radiographic abnormalities.

Where possible, cases were phenotyped as to the persistence or absence of radiographic evidence for lung disease two years after date of diagnosis. The procurement of these data were largely done retrospectively with the exception of a subset of cases from the ACCESS study where study protocol dictated a two-year follow-up exam for cases recruited during the first two years of enrollment 58. For cases that presented with a Scadding stage IV chest radiograph (evidence of lung fibrosis or scarring), no follow-up chest x-ray was needed for phenotyping since a stage IV x-ray indicates permanent changes. Table 6 summarizes the demographic and clinical characteristics of the analytic sample.

Marker Panel Description

We used the Illumina Bead Station platform to assay 1,509 SNPs previously identified as highly informative for distinguishing West African versus European ancestry. The panel of markers is described in documentation on the Illumina website (http://www.illumina.com/products/african_american_admixture_panel.ilmn). After performing both automated and manual genotype calling and quality checks using the Illumina BeadStudio software, we removed data on 49 markers with poor genotype clustering. An additional 23 markers were removed due to a high level (>5%) of missing genotype data (n=8) or genotype frequencies significantly different (p<10^{-5}) than expected under Hardy-Weinberg equilibrium (n=15). Finally, we dropped 53 markers that were in significant (p<0.0001) linkage disequilibrium with one or more adjacent markers. This resulted in a final analytic map of 1,384 markers with an average spacing of 2.9 cM. The median inter-marker distance was 2.4 cM with 95 percent of the markers located no farther than 7.2 cM from an adjacent SNP.

Statistical Methods

Data were analyzed using the ADMIXMAP statistical program (http://homepages.ed.ac.uk/pmckeigu/admixmap/) 29. ADMIXMAP uses a hybrid of Bayesian and traditional modeling to compare observed vs. expected ancestry across the genome. Both affected-only and ancestry association (i.e., case-control) implementations were used. All models were adjusted for age and sex. Prior allele frequencies for the African and European ancestral populations were obtained from the International HapMap project's 59 West AfricanYoruba and European samples, respectively. Analyses were performed on the entire sample and on case subsets stratified by sex, chest radiographic phenotype and disease status (i.e., definite vs. highly probably cases).

As the current versions of admixture mapping methods do not model family data, an unrelated set of cases and controls had to be selected for analysis. The majority of the controls selected for the analyses were derived from the ACCESS portion of the study sample. In addition, for families containing only unaffected sibs (due to lack of DNA from affected individuals in the original pedigree); the unaffected sib with the oldest age of enrollment was selected as a control. Sarcoidosis cases selected for analysis included all ACCESS cases as well as the proband case from each family from the Henry Ford and SAGA studies. For instances in which the proband was not genotyped (due to a lack of DNA or removed as part of the sample quality control process), the case with the earliest age of diagnosis was selected. Stratified analyses used the same case selection criteria (proband or earliest age at diagnosis). This sampling procedure yielded a unique set of 1,026 cases and 316 controls for analysis.

Analyses of mitochondrial and Y chromosomal lineages among unadmixed founders who contributed to the African-American gene pool has shown that the female/male sex ratio was lower in Europeans than in Africans [42,43]. The average European admixture proportion is thus lower on X chromosomes than on autosomes, since male founders contribute half as many X chromosomes as female founders. To correctly model X chromosome admixture proportions, we extended the statistical model to specify a global parameter ψ for the female/male odds ratio in European versus African founders. For any given value of autosomal admixture proportions realized by the Markov chain Monte Carlo sampler in a given individual, the X chromosome admixture proportions can be calculated as follows. We specify a 2×2 frequency table for the most recent unadmixed ancestors of the individual, in which the rows and columns index population ancestry and sex respectively. The row marginal frequencies are the autosomal admixture proportions and the column marginal frequencies reflect the required equal numbers of male and female ancestors.

The cell frequencies are calculated from the odds ratio ψ and the marginal frequencies. The X chromosome European admixture proportions are calculated as a weighted average of the frequencies of European ancestry among females and males, with respective weights in the ratio 2 to 1. The model is specified with a Gaussian prior for $\log_e \psi$, and ψ was sampled from its posterior distribution using a Metropolis algorithm. We allowed ψ to be estimated from our data by specifying a weak prior (i.e., low precision) for the value of $\log_e \psi$, and set the prior mean to the null value of zero (i.e., equal proportion of male and female African founders).

Initial analyses were conducted using the affected-only test because of its greater statistical power, but were followed by the ancestry association test to validate affected-only test results. Significant regions satisfied all of the following criteria: (1) at least one SNP had an affected-only p-value < 0.01 (i.e. absolute Z-score > 2.576); (2) more than 3 SNPs had affected-only p-values < 0.05; and (3) at least one SNP had an ancestry association p-value < 0.05. All locus-specific Z-scores (for both ancestry association and affected-only tests), ancestry odds ratios (aORs; estimated based on the ancestry association test), and ancestry risk ratios (aRRs; estimated from the affected-only test) were calculated with respect to African ancestry, with the aRRs interpreted as the ratio of the risks at the locus of interest associated with carriage of two African alleles versus carriage of two European alleles. Therefore, positive Z-scores and corresponding aRRs > 1 are interpretable as increased risk associated with increased locus-specific African ancestry and negative Z-scores and corresponding aRRs < 1 are interpretable as decreased risk associated with increased locus-specific African ancestry (i.e. increased risk associated with increased European ancestry). To formally test for differences in the results by sex and the extremes of the radiographic phenotype (stage IV radiographic persistent sarcoidosis vs. radiographically resolved disease), we computed a test of heterogeneity of the aRRs estimated from the affected-only analyses from each stratum as the difference in the log affected-only odds ratios (θ) divided by the information (V) for each locus. For strata A and B, the test of heterogeneity was expressed as follows:

$$(\theta_A - \theta_B) / \sqrt{1/V_A - 1/V_B}$$

As with the locus specific affected-only and ancestry association tests, the test of heterogeneity is also distributed as a standard normal deviate. P-values were two-sided for all tests performed.

Additionally, we prepared an exclusion map for African and European risk loci based on the affected-only test results from all sarcoidosis cases by the method described by Hoggart et

al.29 to show which regions of the genome could be excluded from harboring a locus with a specific aRR effect size at the traditional LOD score exclusion threshold of −2 (i.e. 100 times more likely under the null aRR of one in comparison to a specific aRR different than one). Briefly, given the observed score and information from the affected-only test for each SNP used in the analysis, the log-likelihood function (expressed in term of the $\log_e(\text{aRR})$) was solved for the $\log_e(\text{aRR})$ at the value of $\log_e(1/100)$ (i.e. 100 times more likely under the null of no ancestry association). As the log-likelihood is quadratic in $\log_e(\text{aRR})$, there are two possible values for the $\log_e(\text{aRR})$, one greater than and one less than zero. Because the analysis was performed with respect to increased risk associated with increased locus specific African ancestry, the value greater than zero corresponds to the exclusion threshold $\log_e(\text{aRR})$ at or above which we can exclude an African ancestry increased risk locus, and the value less than zero corresponds to the exclusion threshold $\log_e(\text{aRR})$ at or below which we can exclude a African ancestry decreased risk locus or, alternatively, a European ancestry increased risk locus.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. ATS Board of Directors and by the ERS Executive Committee. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG). Am J Respir Crit Care Med. 1999; 160:736–755. [PubMed: 10430755]

2. Schurmann M, Lympany PA, Reichel P, Muller-Myhsok B, Wurm K, Schlaak M, et al. Familial sarcoidosis is linked to the major histocompatibility complex region. Am J Respir Crit Care Med. 2000; 162:861–864. [PubMed: 10988096]

3. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. Nat Genet. 2005; 37:357–364. [PubMed: 15735647]

4. Li Y, Wollnik B, Pabst S, Lennarz M, Rohmann E, Gillissen A, et al. BTNL2 gene variant and sarcoidosis. Thorax. 2006; 61:273–274. [PubMed: 16517590]

5. Rybicki BA, Walewski JL, Maliarik MJ, Kian H, Iannuzzi MC. The BTNL2 Gene and Sarcoidosis Susceptibility in African Americans and Whites. Am J Hum Genet. 2005; 77:491–499. [PubMed: 16080124]

6. Spagnolo P, Sato H, Grutters JC, Renzoni EA, Marshall SE, Ruven HJ, et al. Analysis of BTNL2 genetic polymorphisms in British and Dutch patients with sarcoidosis. Tissue Antigens. 2007; 70:219–227. [PubMed: 17661910]

7. Rybicki BA, Hirst K, Iyengar SK, Barnard JG, Judson MA, Rose CS, et al. A sarcoidosis genetic linkage consortium: the sarcoidosis genetic analysis (SAGA) study. Sarcoidosis Vasc Diffuse Lung Dis. 2005; 22:115–122. [PubMed: 16053026]

8. Iannuzzi MC, Iyengar SK, Gray-McGuire C, Elston RC, Baughman RP, Donohue JF, et al. Genome-wide search for sarcoidosis susceptibility genes in African Americans. Genes Immun. 2005; 6:509–518. [PubMed: 15951742]

9. Rybicki BA, Sinha R, Iyengar S, Gray-McGuire C, Elston RC, Iannuzzi MC. Genetic linkage analysis of sarcoidosis phenotypes: the sarcoidosis genetic analysis (SAGA) study. Genes Immun. 2007; 8:379–386. [PubMed: 17476268]

10. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. Nat Genet. 2008; 40:1103–1106. [PubMed: 19165924]

11. Rybicki BA, Major M, Popovich JJ, Maliarik MJ, Iannuzzi MC. Racial differences in sarcoidosis incidence: a 5-year study in a health maintenance organization. Am J Epidemiol. 1997; 145:234–241. [PubMed: 9012596]

12. Gundelfinger BF, Britten SA. Sarcoidosis in the United States Navy. Am Rev Respir Dis. 1961; 84(suppl):109–115. [PubMed: 13902941]

13. Sartwell PE, Edwards LB. Epidemiology of sarcoidosis in the US Navy. Am J Epidemiol. 1974; 99:250–257. [PubMed: 4818715]

14. Benatar SR. Sarcoidosis in South Africa. A comparative study in Whites, Blacks and Coloureds. S Afr Med J. 1977; 52:602–606. [PubMed: 918803]

15. Edmondstone WM, Wilson AG. Sarcoidosis in Caucasians, Blacks and Asians in London. Br J Dis Chest. 1985; 79:27–36. [PubMed: 3986110]

16. Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet. 2003; 112:387–399. [PubMed: 12579416]

17. Judson MA, Baughman RP, Thompson BW, Teirstein AS, Terrin ML, Rossman MD, et al. Two year prognosis of sarcoidosis: the ACCESS experience. Sarcoidosis Vasc Diffuse Lung Dis. 2003; 20:204–211. [PubMed: 14620163]

18. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, et al. A high-density admixture map for disease gene discovery in african americans. Am J Hum Genet. 2004; 74:1001–1013. [PubMed: 15088270]

19. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel with high ancestry information for african american admixture mapping. Am J Hum Genet. 2006; 79:640–649. [PubMed: 16960800]

20. Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol. 1998; 9:578–594. [PubMed: 9889136]

21. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A. 2006; 103:14068–14073. [PubMed: 16945910]

22. Kopp JB, Smith MW, Nelson GW, Johnson RC, Freedman BI, Bowden DW, et al. MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. Nat Genet. 2008; 40:1175–1184. [PubMed: 18794856]

23. Scherer ML, Nalls MA, Pawlikowska L, Ziv E, Mitchell GF, Huntsman S, et al. Admixture mapping of ankle-arm index: identification of a candidate locus associated with peripheral arterial disease. J Med Genet. 2009

24. Elbein SC, Das SK, Hallman DM, Hanis CL, Hasstedt SJ. Genome-wide linkage and admixture mapping of type 2 diabetes in African American families from the American Diabetes Association GENNID (Genetics of NIDDM) Study Cohort. Diabetes. 2009; 58:268–274. [PubMed: 18840782]

25. Basu A, Tang H, Arnett D, Gu CC, Mosley T, Kardia S, et al. Admixture mapping of quantitative trait loci for BMI in African Americans: evidence for loci on chromosomes 3q, 5q, and 15q. Obesity (Silver Spring). 2009; 17:1226–1231. [PubMed: 19584881]

26. Cheng CY, Kao WH, Patterson N, Tandon A, Haiman CA, Harris TB, et al. Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS Genet. 2009; 5:e1000490. [PubMed: 19461885]

27. Reich D, Patterson N, Ramesh V, De Jager PL, McDonald GJ, Tandon A, et al. Admixture mapping of an allele affecting interleukin 6 soluble receptor and interleukin 6 levels. Am J Hum Genet. 2007; 80:716–726. [PubMed: 17357077]

28. Basu A, Tang H, Lewis CE, North K, Curb JD, Quertermous T, et al. Admixture mapping of quantitative trait loci for blood lipids in African-Americans. Hum Mol Genet. 2009; 18:2091–2098. [PubMed: 19304782]

29. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. Am J Hum Genet. 2004; 74:965–978. [PubMed: 15088268]

30. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, et al. Methods for high-density admixture mapping of disease genes. Am J Hum Genet. 2004; 74:979–1000. [PubMed: 15088269]

31. Schurmann M, Reichel P, Muller-Myhsok B, Schlaak M, Muller-Quernheim J, Schwinger E. Results from a genome-wide search for predisposing genes in sarcoidosis. Am J Respir Crit Care Med. 2001; 164:840–846. [PubMed: 11549543]

32. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet. 2007; 3:e58. [PubMed: 17447842]

33. Onnie C, Fisher SA, King K, Mirza M, Roberts R, Forbes A, et al. Sequence variation, linkage disequilibrium and association with Crohn's disease on chromosome 5q31. Genes Immun. 2006; 7:359–365. [PubMed: 16724073]

34. Chang M, Li Y, Yan C, Callis-Duffin KP, Matsunami N, Garcia VE, et al. Variants in the 5q31 cytokine gene cluster are associated with psoriasis. Genes Immun. 2008; 9:176–181. [PubMed: 18075513]

35. Li Y, Chang M, Schrodi SJ, Callis-Duffin KP, Matsunami N, Civello D, et al. The 5q31 variants associated with psoriasis and Crohn's disease are distinct. Hum Mol Genet. 2008; 17:2978–2985. [PubMed: 18614543]

36. Rossman MD, Thompson B, Frederick M, Maliarik M, Iannuzzi MC, Rybicki BA, et al. HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. Am J Hum Genet. 2003; 73:720–735. [PubMed: 14508706]

37. Sato H, Grutters JC, Pantelidis P, Mizzon AN, Ahmad T, van Houte AJ, et al. HLA-DQB1*0201: A Marker for Good Prognosis in British and Dutch Patients with Sarcoidosis. Am J Respir Cell Mol Biol. 2002; 27:406–412. [PubMed: 12356573]

38. Iannuzzi MC, Maliarik MJ, Poisson LM, Rybicki BA. Sarcoidosis susceptibility and resistance HLA-DQB1 alleles in African Americans. Am J Respir Crit Care Med. 2003; 167:1225–1231. [PubMed: 12615619]

39. Voorter CE, Drent M, van den Berg-Loonen EM. Severe pulmonary sarcoidosis is strongly associated with the haplotype HLA-DQB1*0602-DRB1*150101. Hum Immunol. 2005; 66:826–835. [PubMed: 16112030]

40. Cao K, Moormann AM, Lyke KE, Masaberg C, Sumba OP, Doumbo OK, et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. Tissue Antigens. 2004; 63:293–325. [PubMed: 15009803]

41. Lazarus R, Vercelli D, Palmer LJ, Klimecki WJ, Silverman EK, Richter B, et al. Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. Immunol Rev. 2002; 190:9–25. [PubMed: 12493003]

42. Lind JM, Hutcheson-Dilks HB, Williams SM, Moore JH, Essex M, Ruiz-Pesini E, et al. Elevated male European and female African contributions to the genomes of African American individuals. Hum Genet. 2007; 120:713–722. [PubMed: 17006671]

43. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet. 1998; 63:1839–1851. [PubMed: 9837836]

44. Foreman MG, Mannino DM, Kamugisha L, Westney GE. Hospitalization for patients with sarcoidosis: 1979-2000. Sarcoidosis Vasc Diffuse Lung Dis. 2006; 23:124–129. [PubMed: 17937108]

45. Burke RR, Stone CH, Havstad S, Rybicki BA. Racial differences in sarcoidosis granuloma density. Lung. 2009; 187:1–7. [PubMed: 18716835]

46. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet. 1965; 29:51–76.

47. Rybicki BA, Kirkey KL, Major M, Maliarik MJ, Popovich J Jr. Chase GA, et al. Familial risk ratio of sarcoidosis in African-American sibs and parents. Am J Epidemiol. 2001; 153:188–193. [PubMed: 11159165]

48. Rybicki BA, Iannuzzi MC, Frederick MM, Thompson BW, Rossman MD, Bresnitz EA, et al. Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). Am J Respir Crit Care Med. 2001; 164:2085–2091. [PubMed: 11739139]

49. McDougal KE, Fallin MD, Moller DR, Song Z, Cutler DJ, Steiner LL, et al. Variation in the lymphotoxin-alpha/tumor necrosis factor locus modifies risk of erythema nodosum in sarcoidosis. J Invest Dermatol. 2009; 129:1921–1926. [PubMed: 19225544]

50. Fischer A, Valentonyte R, Nebel A, Nothnagel M, Muller-Quernheim J, Schurmann M, et al. Female-specific association of C-C chemokine receptor 5 gene polymorphisms with Lofgren's syndrome. J Mol Med. 2008; 86:553–561. [PubMed: 18311470]

51. Furuya K, Yamaguchi E, Itoh A, Hizawa N, Ohnuma N, Kojima J, et al. Deletion polymorphism in the angiotensin I converting enzyme (ACE) gene as a genetic risk factor for sarcoidosis. Thorax. 1996; 51:777–780. [PubMed: 8795663]

52. Kruit A, Ruven HJ, Grutters JC, Van Den Bosch JM. Angiotensin-converting enzyme 2 (ACE2) haplotypes are associated with pulmonary disease phenotypes in sarcoidosis patients. Sarcoidosis Vasc Diffuse Lung Dis. 2005; 22:195–203. [PubMed: 16315782]

53. Gleicher N, Barad DH. Gender as risk factor for autoimmune diseases. J Autoimmun. 2007; 28:1–6. [PubMed: 17261360]

54. Berlin M, Fogdell-Hahn A, Olerup O, Eklund A, Grunewald J. HLA-DR predicts the prognosis in Scandinavian patients with pulmonary sarcoidosis. Am J Respir Crit Care Med. 1997; 156:1601–1605. [PubMed: 9372682]

55. Grunewald J, Eklund A, Olerup O. Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. Am J Respir Crit Care Med. 2004; 169:696–702. [PubMed: 14656748]

56. Johns CJ, Schonfeld SA, Scott PP, Zachary JB, MacGregor MI. Longitudinal study of chronic sarcoidosis with low-dose maintenance corticosteroid therapy. Outcome and complications. Ann N Y Acad Sci. 1986; 465:702–712. [PubMed: 3460404]

57. Franke A, Fischer A, Nothnagel M, Becker C, Grabe N, Till A, et al. Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. Gastroenterology. 2008; 135:1207–1215. [PubMed: 18723019]

58. ACCESS Group. Design of a case control etiologic study of sarcoidosis (ACCESS). ACCESS Research Group. J Clin Epidemiol. 1999; 52:1173–1186. [PubMed: 10580780]

59. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851–861. [PubMed: 17943122]
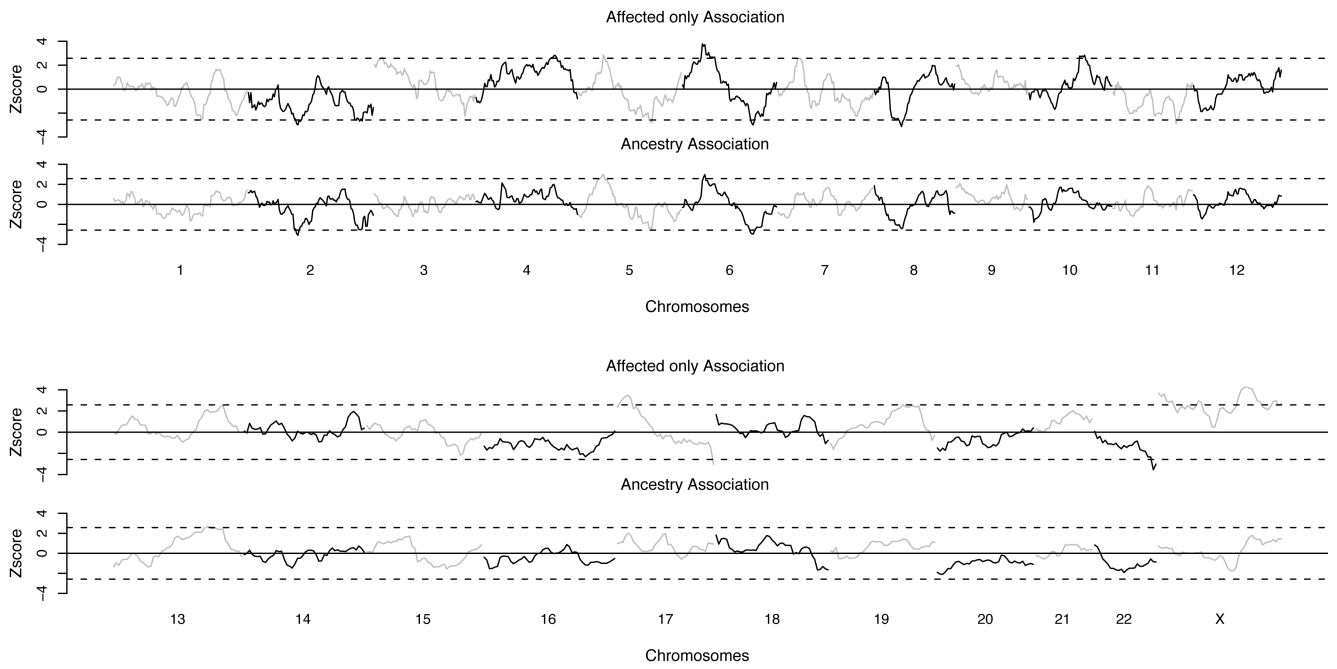
Figure 1.
Genome-wide Ancestry Associations with Sarcoidosis susceptibility based on Affected-Only and Ancestry Association Z scores.
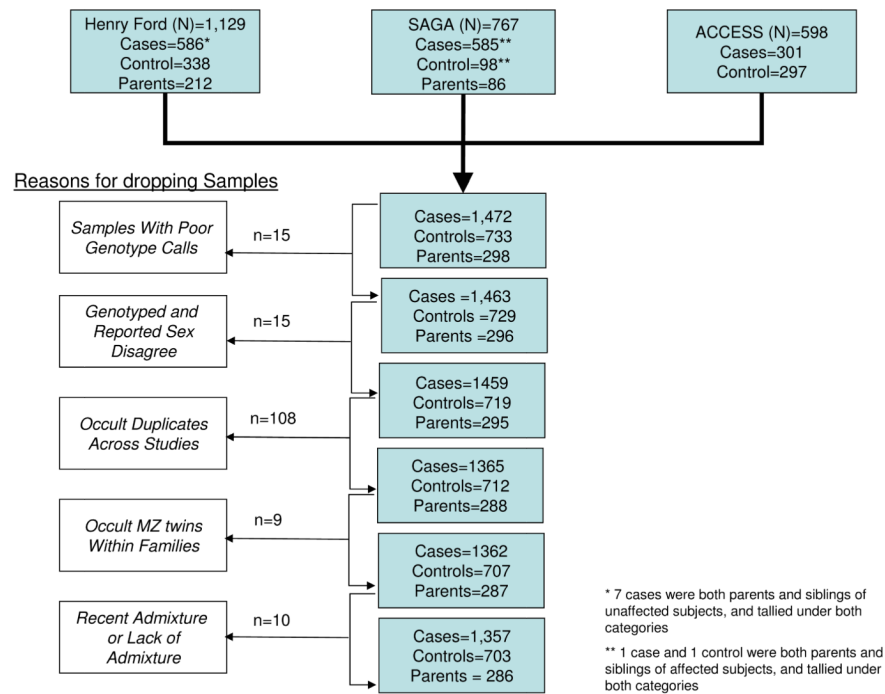
Figure 2.
Flow diagram of sample composition by study of origin and where samples dropped out to form final analysis sample.

Table 1

Markers with strongest association between sarcoidosis and genetic ancestry based on a genome-wide ADMIXMAP analysis of unrelated sarcoidosis 1,026 cases and 316 controls

GENOMIC POSITION	Marker	Affected-only			Ancestry Association		
		aRR	95%CI	P value	aOR	95%CI	P value
2p13.3-2q12.1	Rs1444543	0.67	0.52-0.88	0.004	0.69	0.54-0.89	0.003
2q35-2q36.3	Rs4674659	0.69	0.53-0.91	0.008	0.72	0.55-0.93	0.01
4q31.21-4q34.1	Rs1530044	1.53	1.14-2.05	0.005	1.27	0.94-1.71	0.1[*]
5p13.3-5p13.2	Rs35397	1.46	1.13-1.90	0.005	1.52	1.16-1.99	0.003
5q23.1-5q31.2	Rs30533	0.67	0.51-0.89	0.005	0.72	0.56-0.94	0.01
6p24.3-6p12.1	Rs11966463	1.90	1.36-2.64	0.0002	1.49	1.07-2.08	0.02
6q23.3-6q25.2	Rs276497	0.67	0.52-0.87	0.003	0.68	0.53-0.88	0.003
8p21.3-8p11.21	Rs1462906	0.65	0.5-0.85	0.002	0.73	0.57-0.94	0.02
17p13.3-17p13.1	Rs8070464	1.68	1.26-2.25	0.0005	1.37	1.01-1.85	0.04

Note: Regions were selected satisfied all of the following criteria: (1) at least one SNP with an Affected-only p-value < 0.01 (i.e. absolute Z-score > 2.576); (2) more than 3 SNPs with Affected-only p-values < 0.05; and (3) at least one SNP with an Ancestry Association p-value < 0.05.

(aRR = ancestry risk ratio; aOR = ancestry odds ratio)

[*] adjacent SNP had p value <0.05

Table 2

Markers with strongest association between sarcoidosis and genetic ancestry based on a genome-wide ADMIXMAP analysis of unrelated sarcoidosis 910 cases with a definite diagnosis and 316 controls

GENOMIC POSITION	Marker	Affected-only			Ancestry Association		
		aRR	95%CI	P value	aOR	95%CI	P value
2p13.3-2q12.1	rs1444543	0.63	0.47-0.8	0.001	0.67	0.52-0.86	0.002
2q35-2q36.3	rs4674659	0.71	0.53-0.9	0.02	0.72	0.56-0.94	0.02
4q31.21-4q34.1	rs1530044	1.45	1.06-1.97	0.02	1.20	0.89-1.63	0.2
5p13.3-5p13.2	rs35397	1.55	1.17-2.05	0.002	1.56	1.18-2.05	0.002
5q23.1-5q31.2	rs30533	0.62	0.46-0.82	0.001	0.70	0.54-0.91	0.007
6p24.3-6p12.1	rs11966463	2.01	1.42-2.84	0.00007	1.53	1.09-2.14	0.02
6q23.3-6q25.2	rs276497	0.77	0.58-1.01	0.06	0.71	0.55-0.92	0.01
8p21.3-8p11.21	rs1462906	0.63	0.47-0.84	0.001	0.71	0.55-0.92	0.01
17p13.3-17p13.1	rs8070464	1.62	1.19-2.2	0.002	1.33	0.98-1.81	0.07

Note: Regions were selected satisfied all of the following criteria: (1) at least one SNP with an Affected-only p-value < 0.01 (i.e. absolute Z-score > 2.576); (2) more than 3 SNPs with Affected-only p-values < 0.05; and (3) at least one SNP with an Ancestry Association p-value < 0.05.

(aRR = ancestry risk ratio; aOR = ancestry odds ratio)

Table 3

Markers with strongest association between sarcoidosis and genetic ancestry from a genome-wide ADMIXMAP analysis of unrelated sarcoidosis cases stratified by sex[*]

GENOMIC POSITION	Marker	All Cases (n=1026)			Males (n=252)			Females (n=774)		
		aRR	95%CI	P value	aRR	95%CI	P value	aRR	95%CI	P value
2p13.3-2q12.1	rs4852696[1]	0.67	0.51-0.87	0.003	0.79	0.46-1.36	0.403	0.62	0.46-0.84	0.002
2q35-2q36.3	rs12471078[1]	0.65	0.45-0.96	0.029	1.11	0.56-2.19	0.764	0.53	0.35-0.82	0.004
4q31.21-4q34.1	rs1530044	1.53	1.14-2.05	0.005	0.94	0.53-1.68	0.830	1.8	1.29-2.51	0.005
5p13.3-5p13.2	rs35397	1.46	1.13-1.90	0.005	1.56	0.93-2.62	0.094	1.46	1.08-1.96	0.013
5q23.1-5q31.2	rs30533	0.67	0.51-0.89	0.005	1.07	0.62-1.86	0.802	0.58	0.43-0.79	0.0005
6p24.3-6p12.1	rs11966463	1.9	1.36-2.64	0.0002	1.36	0.72-2.56	0.348	2.07	1.43-2.99	0.0001
6q23.3-6q25.2	rs276497	0.67	0.52-0.87	0.003	0.56	0.33-0.96	0.034	0.72	0.53-0.97	0.028
8p21.3-8p11.21	rs1462906	0.65	0.5-0.85	0.002	0.68	0.4-1.17	0.16	0.65	0.48-0.88	0.005
13q32.2-13q33.3[2]	rs2476230	1.34	1.03-1.75	0.033	0.84	0.49-1.43	0.524	1.60	1.18-2.17	0.003
14q32.2-14q32.33[3]	rs1131877	1.27	0.91-1.78	0.155	2.63	1.42-4.87	0.002	1.01	0.7-1.47	0.957
17p13.3-17p13.1	rs8070464	1.68	1.26-2.25	0.0005	1.84	1.05-3.25	0.035	1.67	1.2-2.32	0.003

Note: Regions selected where either significant in all cases or a sex-stratified subset based on criteria outlined in the table 2 footnote

[*]
Most significant SNP in region in females was different than most significant SNP in all cases

[1]
Significant region only for females

[2]
Significant region only for females

[3]
Significant region only for males

(aRR = ancestry risk ratio estimated from the affected-only analysis)

Table 4

Markers with strongest association between sarcoidosis chest radiographic phenotypes and genetic ancestry based on a genome-wide ADMIXMAP analysis of unrelated sarcoidosis cases

GENOMIC POSITION	Marker	All Cases (n=1026)			Radiographically Resolved Disease (n=295)			Radiographically Persistent Disease (n=525)			Scadding Stage IV Disease (n=186)		
		aRR	95%CI	P value	aRR	95%CI	P value	aRR	95%CI	P value	aRR	95%CI	P value
2p13.3-2q12.1	rs1320149	0.71	0.54-0.92	0.011	0.97	0.59-1.60	0.91	0.67	0.47-0.97	0.036	0.29	0.15-0.55	0.0001
2q35-2q36.3	rs4674659	0.69	0.55-0.91	0.008	0.76	0.46-1.25	0.272	0.62	0.42-0.90	0.013	0.66	0.35-1.24	0.195
2p24.3-2p16.1	rs1316521	0.77	0.56-1.06	0.110	0.67	0.39-1.17	0.161	0.67	0.44-1.03	0.070	0.33	0.17-0.68	0.003
4q21.21-4q24	rs10003577	1.27	0.98-1.66	0.075	1.43	0.88-2.33	0.153	1.70	1.17-2.46	0.005	1.40	0.75-2.59	0.291
4q31.21-4q34.1	rs1530044	1.53	1.14-2.05	0.005	1.52	0.9-2.58	0.117	1.45	0.97-2.17	0.068	1.66	0.87-3.17	0.128
5p13.3-5p13.2	rs40030	1.41	1.09-1.83	0.008	1.97	1.24-3.15	0.004	1.37	0.96-1.94	0.082	2.25	1.25-4.05	0.007
5q23.1-5q31.2	rs30533	0.67	0.51-0.89	0.005	0.79	0.48-1.29	0.348	0.71	0.48-1.04	0.079	0.47	0.25-0.89	0.021
6p24.3-6p12.1	rs11966463	1.90	1.36-2.64	0.0002	2.04	1.15-3.67	0.006	1.61	1.04-2.49	0.006	1.13	0.55-2.31	0.740
6q23.3-6q25.2	rs276497	0.67	0.52-0.87	0.003	0.59	0.36-0.96	0.034	0.63	0.44-0.91	0.013	0.38	0.21-0.70	0.002
8p21.3-8p11.21	rs1462906	0.65	0.5-0.85	0.002	0.92	0.56-1.5	0.723	0.68	0.47-0.98	0.037	0.79	0.43-1.46	0.446
10p12.2-10q11.23	rs7919137	0.82	0.64-1.07	0.349	0.91	0.56-1.48	0.708	0.62	0.43-0.9	0.037	0.27	0.15-0.49	2×10^{-5}
16q22.1-16q23.2	rs12444204	0.78	0.58-1.04	0.094	1.21	0.72-2.04	0.471	0.73	0.49-1.09	0.121	0.32	0.16-0.63	0.001
17p13.3-17p13.1	rs8070464	1.68	1.29-2.25	0.0005	1.66	0.97-2.82	0.062	1.68	1.12-2.51	0.011	2.60	1.36-4.97	0.004
18p11.32-18p11.31	rs568273	1.34	0.95-1.89	0.097	0.74	0.39-1.41	0.362	1.86	1.19-2.91	0.007	0.91	0.43-1.92	0.802
18q23	rs4798954	0.82	0.57-1.17	0.279	0.61	0.33-1.13	0.115	0.49	0.3-0.81	0.005	0.81	0.37-1.79	0.599
19p12-19q13.41	rs2974190	1.45	1.1-1.91	0.009	1.81	1.08-3.04	0.025	1.82	1.25-2.65	0.002	1.49	0.8-2.77	0.201

*
Note: Regions selected where either significant in all cases or a phenotype-stratified subset based on criteria outlined in the table 2 footnote

(aRR = ancestry risk ratio estimated from the affected-only anlaysis)

Table 5

Genomic regions with evidence for loci with significant heterogeneity between stage IV and radiographically resolved sarcoidosis cases based on a genome-wide ADMIXMAP analysis[*]

GENOMIC POSITION	Marker	Radiographically Resolved Disease (n=295)			Scadding Stage IV Disease[*] (n=186)			Stage 4 vs. Resolved Test of Heterogeneity	
		aRR	95%CI	P value	aRR	95%CI	P value	Z score	P value
2p13.3-2q12.1	rs260714	1.10	0.66-1.82	0.718	0.33	0.18-0.61	4×10^{-4}	−2.97	0.003
10p12.2-10q11.23	rs7919137	0.91	0.56-1.48	0.708	0.27	0.15-0.49	2×10^{-5}	−3.07	0.002
16q22.1-16q23.2	rs12444204	1.21	0.72-2.04	0.472	0.32	0.16-0.63	0.001	−3.02	0.002

[*] Note: For a region to be included in table 5, there had to be at least two consecutive markers with heterogeneity p-values < 0.01.

(aRR = ancestry risk ratio estimated from the affected-only analysis)

Table 6

Demographic and Clinical Characteristics of Sarcoidosis Cases and Controls used in Admixture Mapping Study

Factor	Full Sample		Analytic Subset of Unrelateds	
	Cases (n=1,357)	Controls (n=703)	Cases (n=1,026)	Controls (n=316)
Age [*]	36.7 ± 9.7	44.0 ± 11.2	37.4 ± 9.6	41.7 ± 11.5
Male Sex	25.4%	32.7%	24.6%	28.2%
Source of Sample:				
Henry Ford	35.9%	45.9%	44.5%	9.2%
ACCESS	18.8%	40.8%	24.9%	90.5%
SAGA	38.7%	12.4%	24.9%	0.3%
More than one study	7.0%	0.9%	5.8%	0.0%
Diagnostic Category:				
Definite	78.1%		84.22%	
Highly Probable	19.2%		13.8%	
Unknown	2.7%		1.9%	
Chest Radiographic Follow-Up Status:				
Resolved in 2 years	23.7%		23.1%	
Persistent 2 year	48.3%		46.8%	
Stage IV	14.6%		14.7%	
Unknown	28.0%		30.1%	

[*] Age at diagnosis for cases; Age at study enrollment for controls