Fine mapping of chromosome 15q25 implicates ZNF592 in neurosarcoidosis patients

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1002/acn3.229</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:23474121">http://nrs.harvard.edu/urn-3:HUL.InstRepos:23474121</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Fine mapping of chromosome 15q25 implicates ZNF592 in neurosarcoidosis patients

Caleb A. Lareau1,2, Indra Adrianto1, Albert M. Levin3,4, Michael C. Iannuzzi5, Benjamin A. Rybicki3 & Courtney G. Montgomery1

1Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma
2Departments of Biostatistics, Harvard University, Cambridge, Massachusetts
3Department of Public Health Services, Henry Ford Health System, Detroit, Michigan
4Center for Bioinformatics, Henry Ford Health System, Detroit, Michigan
5Department of Medicine, Upstate Medical University, Syracuse, New York

Abstract

Neurosarcoïdosis is a clinical subtype of sarcoidosis characterized by the presence of granulomas in the nervous system. Here, we report a highly significant association with a variant (rs75652600, \( P = 3.12 \times 10^{-9} \), odds ratios = 4.34) within a zinc finger gene, ZNF592, from an imputation-based fine-mapping study of the chromosomal region 15q25 in African-Americans with neurosarcoidosis. We validate the association with ZNF592, a gene previously shown to cause cerebellar ataxia, in a cohort of European-Americans with neurosarcoidosis by uncovering low-frequency variants with a similar risk effect size (chr15:85309284, \( P = 0.0021 \), odds ratios = 5.36).

Introduction

Sarcoidosis is a multisystem disorder characterized by granulomatous inflammation in one or more affected organs.1 Though the etiology of sarcoidosis remains elusive, studies have implicated both environmental and genetic components, including a recent common-variant study that identified a suggestive association on

---

"The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health."
chromosome 15q25. While up to 90% of patients diagnosed with sarcoidosis have lung involvement, prevalence of nervous system involvement, known as neurosarcoidosis, has been reported as high as 10%. Specific studies that examine the phenotypic variance leading to a diagnosis of neurosarcoidosis could elucidate the genetic and environmental architecture leading to sarcoidosis in general.

Though sarcoidosis impacts individuals of all races, ages, and genders, African-Americans (AA) are more commonly and severely affected in the United States than European Americans (EA). To date, no study has identified the genetic or environmental etiology of neurosarcoidosis. Using a subset of AA neurosarcoidosis cases from a recent genome-wide association (GWA) study, we aimed to identify genetic variation associated with risk of neurosarcoidosis and validate these associations in an EA cohort. Rather than performing a genome-wide analysis that could lead to inflated results due to small samples, we focused our analysis to a particular region of chromosome 15 to determine possible association with neurosarcoidosis. We hypothesized that this region could be in association with neurosarcoidosis from two previous results. First, a linkage mapping association was uncovered near this same region using the neurosarcoidosis and lymph subphenotype. Additionally, a suggestive association signal on this region of chromosome 15 from a recent sarcoidosis GWA common variant study.

### Subjects and Methods

Our AA sample originally comprised 1487 cases and 1684 controls taken from four different studies. Sample quality control measures removed duplicate patients, individuals that were extreme population outliers, and samples with a low genotyping call rate (<90%) among other criteria previously described. The final AA sample was comprised of 1273 sarcoidosis cases and 1645 health controls. Of the 1273 cases, 83 were sub classified with neurosarcoidosis. This combined dataset comprised A Case Control Etiologic Study of Sarcoidosis (ACCESS), Sarcoidosis Genetic Analysis (SAGA), Henry Ford Health System in Detroit Michigan, healthy controls taken from the Oklahoma Medical Research Foundation (OMRF) Lupus Family Registry and Repository, and HapMap controls from Yoruba in Ibadan, Nigeria; ASW, African ancestry in Southwest USA. For the EA validation sample, 32 of the 442 cases that passed identical quality control measures were classified neurosarcoidosis cases, which were compared against 2284 EA controls. Full details of our sample collection have previously been reported.

Genotyping was performed at OMRF using the Illumina HumanOmni1-Quad array for over 1.1 million variants across the genome. The standard quality control methods, including a 1% variant minor allele frequency (MAF) and 95% call rate, were applied to the common-variant dataset as described in Adrianto et al. To determine the effect of low-frequency variants driving the signal in 15q25, we performed targeted resequencing of this region using the Illumina HiSeq 2000 platform with Illumina Pipeline software version 1.7 using standard procedures. Genomic DNA (3–5 μg) from 187 Sarcoidosis cases (12 neurosarcoidosis) and 293 healthy controls of AA decent was prepared for sequencing using an Illumina Paired-End Genomic DNA Sample Prep Kit. No additional EA samples were resequenced. We compared sequence-based variant calls with single-nucleotide polymorphisms (SNPs) previously genotyped on the Illumina HumanOmni1-Quad array platform and found >99% concordance between platforms. Imputation was performed spanning the hg19 chromosomal coordinates chr15:85,275,210–85,494,027 with the results of targeted sequencing experiment as a reference panel along with the 1000 Genomes Project Phase I integrated variant set. The result of the imputation was 1783 variants that passed quality control (information measure >0.5; average maximum posterior genotype call probability >0.9) measures, totaling 4703 variants evaluated in the analyses of 15q25.

Single-marker association analyses were performed using the EMMAX® and PLINK® logistic regression. As EMMAX implements a variance component approach to the linear mixed model, the software simultaneously adjusts for both pairwise genetic relatedness between individuals and corrects for population stratification. As EMMAX does not compute odds ratios (OR) or confidence intervals, PLINK logistic regression was employed.

Table 1. Summary of African-American samples after quality control.

<table>
<thead>
<tr>
<th>Study</th>
<th>Sarcoïdosis cases</th>
<th>Neurosarcoidosis cases</th>
<th>Healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS</td>
<td>222</td>
<td>25</td>
<td>251</td>
</tr>
<tr>
<td>SAGA</td>
<td>566</td>
<td>37</td>
<td>482</td>
</tr>
<tr>
<td>Henry Ford</td>
<td>485</td>
<td>21</td>
<td>482</td>
</tr>
<tr>
<td>OMRF</td>
<td>0</td>
<td>0</td>
<td>557</td>
</tr>
<tr>
<td>HapMap YRI-ASW</td>
<td>0</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td>Total</td>
<td>1273</td>
<td>83</td>
<td>1645</td>
</tr>
</tbody>
</table>

A summary of the AA sample composition for the sarcoidosis cases, neurosarcoidosis cases, and healthy controls. ACCESS, A Case Control Etiologic Study of Sarcoidosis; SAGA, Sarcoidosis Genetic Analysis; OMRF, Oklahoma Medical Research Foundation; YRI, Yoruba in Ibadan, Nigeria; ASW, African ancestry in Southwest USA.
to further analyze single-marker associations. The Local ancestry in AdMixed Populations (LAMP) program was then utilized to estimate local ancestry defined as the probability of carrying zero, one, or two copies of West African (or European) ancestral allele at each SNP for each individual. Ancestral allele frequency estimates for Illumina Omni-Quad SNPs were obtained from the HapMap Yoruba and CEPH European Utah catalog database. Variants strongly associated with the neurosarcoïdosis phenotype ($P < 1 \times 10^{-8}$) were examined for functional activity using a standard query into the RegulomeDB webserver.

Results

In our previous genome-wide common variant analysis for AA sarcoidosis cases and controls, we observed suggestive association in the 15q25 region spanning two genes, ZNF592 and ALPK3, for all AA sarcoidosis cases and controls (Fig. 1A). Figure 1B shows the results of association between only neurosarcoidosis cases and healthy controls. Although the sample size dropped considerably from our first association analysis (1273 cases/1645 controls) to our second association analysis (83 neurosarcoïdosis cases/1645 controls), we observed a substantial increase in statistical significance. While no variant in the first analysis had a $P$-value less than $1 \times 10^{-8}$, several of the variants in the second analysis surpassed genome-wide significance ($P < 5 \times 10^{-8}$ from EMMAX) with the most significant SNP at rs75652600 ($P = 3.12 \times 10^{-8}$; OR = 4.34, 95% CI = 2.30–8.20; MAF = 4.6%). We note that rs75652600 was imputed with strong linkage disequilibrium (LD) with rs62019469 ($r^2 = 0.98$; $P = 9.05 \times 10^{-8}$; OR = 3.43, 95% CI = 2.12–5.54; MAF = 4.9%), which was present in the original genotyping. When assessing potential functionality from our highly significant variants using RegulomeDB, we determined that a genotyped variant with a strong association signal in ZNF592, rs3748376 ($P = 5.5 \times 10^{-7}$; OR = 3.64, 95% CI = 2.12–6.25; $r^2 = 0.75$ with rs75652600), was likely to affect binding between USF1 and ZNF592 in the K562 cell line evidenced by a transcription factor-binding site and a DNase hypersensitivity peak. RegulomeDB revealed no other functionality for highly significant variants.

As the AA genome is admixed for African and European descent, genetic association signals could result from an association with local ancestry of common alleles instead of disease. Thus, we adjusted for not only global but also local ancestry to protect against this source of false positives. As shown in Figure 1C, our result indicates that this signal in 15q25 is not due to ancestry, but the result of a statistical association between the variants in this region and neurosarcoïdosis as rs75652600 ($P = 3.99 \times 10^{-7}$; OR = 4.27, 95% CI = 2.44–7.49) was still strongly significant, particularly noting the minimal change in OR.

While we were able to establish a strong association between neurological involvement and 15q25, two genes, ALPK3 and ZNF592, each contained variants strongly associated with sarcoidosis. Many SNPs in the ALPK3-ZNF592 150 kb region, including rs75652600, are in strong LD with each other. This implies that while the association signal persists across both genes, it is likely only one effect in this region. To verify the presence of a single association signal in this region, we used multivariate logistic regression adjusting for the most significant SNP, rs75652600, within ZNF592 and found the significance of the remaining association signals within this region significantly diminished (minimum $P = 0.0036$, chr15:85277267:D, Fig. 1D). This is not true when conditioning on SNPs within ALPK3. Thus, we suggest that the broad association effect is a product of a single association in ZNF592 and strong LD in the region. Though we had specifically hypothesized that the association with this region could be better characterized with the neurosarcoïdosis subphenotype, the same association models were performed for 11 other subphenotypes (lung, eye, thoracic, cardiac, liver, spleen, renal, skin, parotid/salivary gland, muscle, and bone marrow), and none of these secondary analyses surpassed genome-wide significance as in the neurosarcoïdosis sample.

After characterizing the association in the AA cohort, we examined variants in ZNF592 within the EA neurosarcoïdosis samples. Performing an identical case/control analysis with a smaller subset of cases (32 EA neurosarcoïdosis cases/2284 EA controls), we identified two variants with a similar risk effect size, chr15:85301273 (P = 0.018; OR = 3.80; 95% CI = 1.158–12.45; MAF = 1.4%) and chr15:85309284 ($P = 0.0021$; OR = 5.36; 95% CI = 1.616–12.45; MAF = 1.0%). Notably, these two implicated variants represented independent effects ($r^2 = 0.0057$) of one another. Common variants implicated in the AA analysis were not significant in the EA analysis (rs62019469 $P = 0.304$; rs75652600 not imputed in EA sample), highlighting the importance of the fine-mapping analysis performed in this study. Our results suggest a multi-ethnic association between the ZNF592 gene and risk of neurosarcoïdosis. Table 2 summarizes the results of the variant associations in both ethnicities.

Discussion

Though our AA discovery sample represented a subset of a larger genetic association, the association analysis using the 83 neurosarcoïdosis cases and 1645 controls had 96.7% power to explain 10% of the heritability for a
marker at the genome-wide significance level. From our neurosarcoidosis case/control analysis, we observed variants in ZNF592, notably rs75652600, that achieved genome-wide significance. Zinc finger protein 592 plays a role in cerebral development and is expressed throughout the central nervous system at relatively high levels. Notably, ZNF592 mutations have been shown to cause a cerebral ataxia.

To further substantiate the role of ZNF592, we examined genetic factors that interact with this gene. Malovannaya et al. showed that ZNF592 functions in a multiple-protein complex with ZMYND8. Notably, ZMYND8 showed...
significant chromatin modification patterns in CD4+ T cells, suggesting a role for ZMYND8 in T-cell immune response. Additionally, as the results of our RegulomeDB query uncovered a potential interaction between USF1 and ZNF592, we note that the selective interaction between USF1 and STAT1 mediates the activation of CIITA (a MHC Class II transactivator) by IFNγ, suggesting that USF1 has a defined role in MHC Class II activity. Both CD4+ T Cells and MHC Class II activity have been implicated in sarcoidosis, and these interactions support a role for ZNF592 in the existing immune-based genetic understanding of the systemic disease. Finally, a strong interaction between ZNF592 and SMAD9 was originally uncovered in a study to determine additional proteins involved in SMAD signaling. SMAD9, like many proteins in this pathway, plays an important role in TGF-β signaling. As TGF-β has been associated with the severity of pulmonary sarcoidosis, we hypothesize that the interaction between ZNF592 and SMAD9 could be significant in granuloma formation.

Though sarcoidosis is thought to aggregate in families, the risk to relatives of neurosarcoidosis patients is largely unknown. From our 83 neurosarcoidosis cases in the AA sample, only one sib pair and one parent pair was observed. Noting the familial risk ratio (λ) of sarcoidosis has been reported as 2.49 for parents and sibs combined, an accurate measure as the degree of familial risk for neurosarcoidosis must be examined in larger samples.

Our identification of a genetic effect driven by common variant association in one ethnicity but low-frequency variants in another is unusual but has been observed in other disease states. This effect could be explained due in part to the difference in allele frequencies of the common variant association. The genotyped SNP, rs62019469, has a MAF in the EA sample at 28.4% while only 4.9% in the AA sample. Though the conditional analysis verified the AA signal was not driven by particular ethnic differences in the samples, the imbalance of the proportion of risk alleles between these two ethnicities likely differentiates the disparity in association. We note that a similar risk variant architecture has been observed in interferon induced with helicase C domain 1 (IFIH1) in systemic lupus erythematosus (SLE). Similar to our association of ZNF592 in neurosarcoidosis, common and low frequency variants from IFIH1 were differently associated with the EA and AA samples in SLE, suggesting a differential role in protection through an evolutionary mechanism.

While our approach compared neurosarcoidosis cases against healthy controls, we note an association analysis was performed comparing neurosarcoidosis cases against sarcoidosis cases that were not diagnosed with nervous system involvement. This analysis yield no strongly associated variants in 15q25 as the prevalence of undiagnosed neurosarcoidosis patients likely confounded the signal. Though the prevalence of neurological symptoms was 6.5% in our sample, previous studies have shown that the true prevalence may be as high as 14% after assessing sarcoidosis patients postmortem. Uncovering the genetic basis of neurosarcoidosis will aid in the understanding of this subtype of sarcoidosis. As our novel association of ZNF592 with neurosarcoidosis explains only a fraction of the phenotypic variance of the systemic disease, additional studies to augment this association are needed to sufficiently characterize the genetic etiology of neurosarcoidosis.

**Acknowledgments**

We recognize the NHLBI-funded ACCESS and SAGA research groups for the original data collection efforts that contributed the majority of the patient samples to this study. Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS), Heart, Lung and Blood Institute (NHLBI), and National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health under award numbers: P20GM103456 and P30GM110766 to I. A., R01-HL54306 and U01-HL060263 to M. C. I., R56-AI072727 and R01-HL092576 to B. A. R., and

---

**Table 2. Summary of variants in ZNF592 associated with neurosarcoidosis.**

<table>
<thead>
<tr>
<th>Variant</th>
<th>AA P-value</th>
<th>OR</th>
<th>MAF (%)</th>
<th>EA P-value</th>
<th>OR</th>
<th>MAF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs62019469</td>
<td>9.05 × 10⁻⁸</td>
<td>3.43</td>
<td>4.9</td>
<td>0.30</td>
<td>1.42</td>
<td>28.4</td>
</tr>
<tr>
<td>rs75652600</td>
<td>3.12 × 10⁻⁸</td>
<td>4.34</td>
<td>4.6</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>chr15:85301273</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.018</td>
<td>3.8</td>
<td>1.4</td>
</tr>
<tr>
<td>chr15:85302984</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.0021</td>
<td>5.36</td>
<td>1.0</td>
</tr>
</tbody>
</table>

A summary of the P-values, OR, and MAF for the key variants identified in this study. The genotyped SNP, rs62019469, was called in both the EA and AA samples while rs75652600 was imputed only in the AA sample, chr15:85301273 and chr15:85302984 were called in the EA sample after targeted resequencing. AA, African-Americans; EA, European-Americans; OR, odds ratios; MAF, minor allele frequency.
RC2HL101499 and R01HL113326 to C. G. M. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

All authors contributed substantively to this work. C. A. L., I. A., A. M. L., M. C. I., B. A. R., and C. G. M. were involved in study conception and design. C. A. L., I. A., and A. M. L. were involved in the data organization and statistical analysis. M. C. I., B. A. R., and C. G. M. coordinated the data collection and processing. C. A. L., I. A., and C. G. M. drafted the manuscript and prepared the figures. All authors were involved in the reviewing and editing of the manuscript.

Conflict of Interest

None declared.

References