



Published in final edited form as:

*Curr Opin Pulm Med.* 2020 September ; 26(5): 544–553. doi:10.1097/MCP.0000000000000719.

## Recent advances in sarcoidosis genomics: epigenetics, gene expression, and gene by environment (G×E) interactions studies

Lori Garman<sup>1,\*</sup>, Courtney G. Montgomery<sup>1</sup>, Natalia V. Rivera<sup>2,3,4</sup>

<sup>1</sup>Department of Genes and Human Disease, Oklahoma Medical Research Foundation, Oklahoma City, OK

<sup>2</sup>Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

<sup>3</sup>Rheumatology Division, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

<sup>4</sup>Center of Molecular Medicine (CMM), Karolinska Institutet, Stockholm, Sweden

### Abstract

**Purpose:** We aim to review the most recent findings in genomics of sarcoidosis and highlight the gaps in the field.

**Recent Findings:** Original explorations of sarcoidosis sub-phenotypes, including cases associated with the World Trade Center and ocular sarcoidosis, have identified novel risk loci. Innovative gene-environment interaction studies utilizing modern analytical techniques have discovered risk loci associated with smoking and insecticide exposure. The application of whole-exome sequencing has identified genetic variants associated with persistent sarcoidosis and rare functional variations. A single epigenomics study has provided background knowledge of DNA methylation mechanisms in comparison to gene expression data. The application of machine learning techniques has suggested new drug repositioning for the treatment of sarcoidosis. Several gene expression studies have identified prominent inflammatory pathways enriched in the affected tissue.

**Summary:** Certainly, sarcoidosis research has recently advanced in the exploration of disease sub-phenotypes, utilizing novel analytical techniques, and including measures of clinical variation. Nevertheless, large-scale and diverse cohorts investigated with advanced sequencing methods, such as whole-genome and single-cell RNA sequencing, epigenomics, and meta-analysis coupled with cutting-edge analytic approaches, when employed, will broaden and translate genomics findings into clinical applications, and ultimately open venues for personalized medicine.

### Keywords

Genetics; genomics; DNA methylation; transcriptomics; sarcoidosis

---

\*Correspondence: Lori Garman, PhD, Department of Genes and Human Disease, Oklahoma Medical Research Foundation, 825 NE 13th, MC-104, Oklahoma City, OK 73104. lori-garman@omrf.org. Phone: +1(405)-271-1705 extension 1.

*Conflicts of Interest:* none

## Introduction

Sarcoidosis is a complex inflammatory disease characterized by noncaseating granulomas in any organ of the body, but most commonly in the lungs and lymph nodes. The incidence and clinical presentation of sarcoidosis vary across ethnic populations and are affected by, at minimum, race, sex, age, vitamin D status, and exposure to environmental triggers, such as cigarette smoking, mold, insecticides, dust, and potentially season[1]. Sarcoidosis can be divided into multiple clinical phenotypes that describe the course of the disease, including Löfgren's syndrome (LS), arguably a different disease, and non-Löfgren's sarcoidosis (non-LS).

Ancestry differences, high sibling relative risks, and linkage findings all suggest a genetic basis for sarcoidosis risk. The first large multicenter family-based study in 2001 began to address both ancestry-dependent disease risk and familial relative risk[2].

Twenty years of additional familial aggregation, familial linkage, candidate gene, and genome wide association studies (GWAS; reviewed in [3, 4]) have led to an estimated pooled prevalence proportion of familial sarcoidosis of 9.5% (CI 4.6–16.1) and an estimated heritability of 60–70%. These estimates, based on twelve study populations, suggest familial sarcoidosis is highest in French, African American, Dutch and Irish patients. Genetic variants with high penetrance have not been identified in familial sarcoidosis and the significant part of familial occurrence that is not explained by heritable factors may be explained by environmental factors [4].

Over time, it has become evident that this genetic susceptibility is complex and multifactorial. LS and non-LS differ in most risk loci, but share genetic factors localizing in the major histocompatibility complex (MHC) region[5\*\*, 6]. Within non-LS, individuals of African descent have few overlapping genetic risk factors with those of European ancestry[7–9], and genetic effects of sub-phenotypes of non-LS, like neuro-[10], ocular[11\*], or bone and joint sarcoidosis[12] appear manifestation-specific. The relative risks for first-degree familial relationships differ between European American (odds ratio: 16.6) and African American (odds ratio: 3.1) [2]. This ancestry- and phenotype-associated variation in non-LS suggests multiple genetic architectures and potentially differential exposures underlying various disease mechanisms.

It is apparent that sarcoidosis research lags behind other inflammatory disorders in the application of cutting-edge technologies and analysis techniques necessary to identify genetic, epigenetic, and transcriptomic risk factors and the endeavor to link genotypic to phenotypic data. For example, no extensive eQTL studies in sarcoidosis-relevant populations have been performed, except for a few expression quantitative trait loci (eQTL) analyses[13–15] targeted towards cell types or genes associated with sarcoidosis that explained only a small fraction of the variation in gene expression. In spite of this lag, there is a group of studies published in the last two years in which advances in sarcoidosis genetics, epigenetics, and transcriptomics have been made through the use of original cohorts, technology, and analysis methodology. In this review, we aim to summarize these studies as well as indicate the gaps in the sarcoidosis genetics literature.

## Genetics

As in most inflammatory diseases, variants located in the major histocompatibility complex (MHC) have historically been the major genetic risk factors in sarcoidosis, although these vary by ancestry as well as clinical presentation. MHC class II alleles *HLA-DRB1*\*11:01, 12:01, and 15:03 are associated with sarcoidosis in African American patients, and 15:01 and 04:01 in individuals of European descent. *HLA-DRB1*\*03:01 is associated with increased susceptibility, but also with disease resolution, in European LS patients; yet in African American patients, 03:01 is protective and 03:02 is associated with susceptibility and resolution. In addition to the MHC loci, other genes with and without known immune function have also been reported to be associated with sarcoidosis. Commonly replicated loci include *ACE*, *ADAM33*, *ANXA11*, *BTNL2*, *FUT9*, *NOD2*, *CCR2*, *IL23R*, *NOTCH4*, *C6ORF67*, *OS9*, *PRDX5*, *RAB23*, *SLC11A1*, *TGFB1*, *TLR9*, *TNFA*, and *XAF1* (descriptions of sarcoidosis-associated genes/transcripts and pathways can be found in Tables 1 and 2, respectively). Many of these are shared with other immune-mediated diseases or with sarcoidosis-associated risk factors. For example, *IL23R* is associated with Crohn's disease, ulcerative colitis, psoriasis, and ankylosing spondylitis[5\*\*].

Most of the sarcoidosis genetic associations discussed above were found with GWAS studies. still the gold standard in genetic studies of sarcoidosis, as no whole-genome sequencing studies and only a few whole-exome studies[16\*–18\*] have been performed. Sarcoidosis genetic research is also limited in that population-based cohorts from Germany, Sweden, Japan, and the United States have been historically analyzed individually, with the notable exception of two 2015 studies utilizing Immunochip data from multiple cohorts[6, 19]. Sarcoidosis is also not typically analyzed with other diseases, with a few notable exceptions,[20, 21]. Finally, gene by environment analyses are few, but have, for example, found interactions of *FUT9* with insecticide exposure[22] and *IL23R* with smoking in non-LS, but not in LS[5\*\*]. Several recent studies have begun to address these limitations with novel sequencing technologies and gene by environment analytical techniques. Finally, two multicenter studies known as Genotype–Phenotype Relationship in Sarcoidosis (GenPhenReSa)[23] and the Multi-Ethnic Sarcoidosis Consortium for Genetic Studies MESARGEN (<https://mesargen.wordpress.com>) represent international networks of population-based cohorts created for genomic studies of sarcoidosis, including genetics, transcriptomics, and epigenomics.

### Old genetics methods in a new subset of patients: WTC-associated sarcoidosis

While firefighters in the Fire Department of the City of New York have a higher incidence of sarcoidosis relative to the general population of similar sex and race (12.9 per 100,000 versus 9.4 per 100,000), the incidence of sarcoidosis doubled in this group following the World Trade Center (WTC) attacks (25 per 100,000). This cohort represents a unique research opportunity, with a known baseline rate, point of exposure, and an extensively clinically characterized study population, allowing for cohort-matching on exposure and other potentially confounding factors. The genetic differences between firefighters who responded to the collapse of the WTC in 2001 and developed sarcoidosis (n=55) have been

compared to those with similar demographics, smoking rates, and exposure to WTC-associated dust that did not develop the disease (n=100)[24\*].

In this cohort, the authors examined 51 genes involved in immune function, a portion of which had been previously associated with sarcoidosis, using a new amplicon-based enrichment method for targeted next-generation sequencing (AmpliSeq). The authors identified and classified 909 common single-nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) of  $\geq 5\%$  and 3,619 total genetic variants within their predominantly Caucasian cohort (153/155, 98.7%). Of these, 17 common variants located on chromosomes 1 and 6, were found to be associated with sarcoidosis as a whole. Not surprisingly, many of the findings were in or close to HLA genes, including *BTNL2*, *PTGS2/COX2*, and *PACERR*, a long non-coding RNA. Additionally, one of these sarcoidosis-associated SNPs, rs2066826, located in an intron of *PTGS2*, was associated with extrathoracic involvement along with six other SNPs. In a secondary analysis, the authors considered the arrival time of a firefighter at the scene as a proxy of dust exposure; interestingly, no effects of exposure were found. While the small sample size limits the study, the results suggest that the amount of dust particle exposure is not causal, but rather a trigger in genetically predisposed individuals.

### **Old genetics methods in a new subset of patients: ocular sarcoidosis**

Substantial evidence suggests the genetic effects of sub-phenotypes of non-LS, including neurosarcoidosis[10] and bone and joint sarcoidosis[12] are manifestation-specific. Earlier this year, two of the authors of this review published the first GWAS of ocular sarcoidosis in an ethnically diverse cohort of 1,271 African American sarcoidosis cases, 1,551 African American controls, 332 European American cases, and 2,046 European American controls[11\*]. While the methods employed were classic GWAS analytical techniques, the study was the first GWAS of the clinical phenotype of ocular sarcoidosis. SNPs within the novel sarcoidosis risk locus *MAG11* were associated with ocular sarcoidosis in African American individuals. These findings implicate an autoimmune mechanism underlying sarcoidosis, as genetic variants in *MAG11* have been previously reported in multiple diseases involving a loss of barrier function: spontaneous glomerulosclerosis, celiac disease, medically refractory ulcerative colitis, and the stricturing phenotype of Crohn's disease. The study also confirmed several previous HLA associations with ocular sarcoidosis.

### **New genetics analytics of gene by environment interactions: smoking**

Historically, the literature on the interaction of smoking with sarcoidosis has been conflicting, as previous studies did not take into consideration the genetic background of individuals included in the observational studies and often merged clinical phenotypes of LS and non-LS (reviewed in [5\*\*]). Rivera and colleagues examined genetic data with smoking information in LS (n=292) and non-LS (n=455) and compared these groups with healthy controls (n=2,966), all individuals from Sweden with European ancestry. The authors reported distinct sets of genetic variants in LS and non-LS (54 and 34 SNPs, respectively) that significantly interacted with smoking, and showed that risk for the disease substantially increased if an individual had a genetic predisposition and smoked. The smoking-sarcoidosis interacting variants included well-known loci previously reported associated with the

disease, such as *IL23R*, *ANXA11*, *PRDM1*, *CLEC10A*, *CCR7*, *NLRP7*, *CLIP1*, and *BCL2*. Interestingly, adjustment for the *HLA-DRB1\*03* allele revealed a large number of interacting variants, 106 in LS and 33 in non-LS, suggesting a plausible gene-gene interaction between *HLA-DRB1\*03* and other disease loci in the genome. Overall, the study found that smoking modulates the risk of sarcoidosis by 56% in LS and 62% in non-LS patients. The interaction effects with smoking were calculated using a novel model of attributable proportion or a metric interaction consistent with additive, multiplicative, and multifactorial threshold models that had been previously utilized in rheumatoid arthritis[25]. Moreover, to strengthen their findings, the authors utilized canonical pathway algorithms implemented in the MetaCore™ software to determine signaling pathways implicated in sarcoidosis pathogenesis influenced by smoking exposure. The analysis led to the identification of relevant GO processes, including positive regulation of nitrogen compound metabolic process (87.2%; 6.364e-25), response to an organic substance (82.5%; 1.268e-16), and regulation of cellular process (91.7%; 9.839e-19) – all which may be triggered by cigarette smoking.

### **New genetics analytics of gene by environment interactions: insecticide use**

Insecticide exposure has been found to be associated with sarcoidosis both independently[26] and in a previous GxE study[22]. Last year, Chen and colleagues[27\*\*] developed a novel GxE method accounting for population admixture and related individuals, as known admixed populations like African Americans are commonly and severely affected by sarcoidosis. After benchmarking the methods on simulated data, the authors applied their methodology to a previously analyzed data set of 1,877 African Americans (1,073 cases and 804 controls) collected from both family and unrelated study groups[28] to assess the effect of insecticide exposure in sarcoidosis. The most significant interacting variant with disease risk and insecticide exposure identified by the GEE-joint method was rs3129890, located 1,447 base pairs from the 3' of *HLA-DRA* on chromosome 6, which had an increased risk effect (odds ratio of 2.32 in the exposed versus 1.43 in the unexposed). This same variant was in high linkage disequilibrium ( $r^2 > 0.78$ ) with a known African-American-specific sarcoidosis risk variant rs2227139[7] that was also found to be interacting with insecticide exposure. Additionally, two novel loci with suggestive evidence of GxE interaction were also identified.

### **New genetics sequencing methods: whole-exome studies**

Whole-Exome Sequencing (WES) is a relatively new sequencing technology that provides coverage of more than 95% of the exons in the human genome, which are thought to contain roughly 85% of the genetic variants associated with human disease phenotypes. In WES, genomic DNA is fragmented, and targeted regions are captured by hybridization to probes in solution or on an array before subsequent amplification and sequencing. As exome sequences encompass only about 2% of the human genome, WES can be performed in smaller cohorts while maintaining sufficient power for identifying genetic risk factors within sequenced regions. The studies below suggest WES can provide further insights on known sarcoidosis risk variants and potentially identify novel variants.

A recent Finnish study performed WES with the goal of identifying variants associated with persistent sarcoidosis[17\*]. Of the 188 patients in the study, roughly half (n=98) had persistent activity after two years; the rest had resolved within two years of diagnosis. Thirty-six patients from each classification, cohort-matched by known Finnish HLA associations, including *HLA-DRB1*\*03:01 and haplotype *HLA-DRB1*\*04:01-*DPB1*\*04:0, were used in a discovery cohort. Nineteen variants were associated with prognosis in the discovery cohort; 5 of these, in genes *AADACL3*, *C1orf158*, *LILRB4*, and *KIR3DL1/KIR3DS1*, were replicated in the full cohort. HLA-specific subset analyses confirmed all five in those patients with known HLA associations and 3 in patients without these variations. *AADACL3* and *C1orf158* were previously reported to be associated with sarcoidosis in a WES study of German families[16\*]. In the German study, exomes of 22 sarcoidosis cases from six families were sequenced, and authors applied analysis techniques designed to address linkage and high-penetrance to identify familial-clustered, functional, and rare genetic variants. Forty functional rare variants were identified through bioinformatic approaches suggesting a link to interferon and IL-10 pathways, immune cell proliferation, bacterial defense, and vesicle-mediated transport, and enriched pathways that were most similar in inflammatory bowel disease. A similar 2019 study examined 5 French families with non-LS and found 227 disease susceptibility variants in 192 genes[18\*]. These variants were predominantly missense (223, 88.9%), but also included small numbers of splicing variants (2), missense variants, in-frame deletion/insertions (9), nonsense variants (8), and a start-stop. Functional analyses implicated the mTOR signaling, autophagy, the TCA cycle, and leptin and interferon signaling.

### Ongoing large-scale genetic efforts

In order to identify robust genetic associations of sarcoidosis, probable genetic heterogeneity, and genotype-to-phenotype connections, large cohorts diverse in ethnicity, clinical presentation, and types of data collected are necessary. Two recently established multicenter programs have generated large cohorts of Caucasian sarcoidosis patients with deeply characterized clinical phenotypes. First, the *GenPhenReSa* applied a sophisticated hierarchical clustering method to identify organ phenotype clusters in over 2,000 sarcoidosis patients with acute onset and subacute onset. Five distinct organ phenotype clusters were identified, namely 1) abdominal, 2) ocular–cardio–cutaneous–central nervous system, 3) musculoskeletal–cutaneous, 4) pulmonary–lymphonodal, and 5) extrapulmonary[23]. A similar study of 195 hospital-based cases from Greece applied the 18F-FDG PET/CT imaging modality coupled with a hierarchical clustering method to identify four cluster phenotypes: 1) thoracic nodal hilar-mediastinal, 2) thoracic nodal hilarmediastinal and lungs, 3) an extended thoracic and extra-thoracic only nodal phenotype including inguinal-abdominal-supraclavicular stations, and 4) all the above plus systemic organs and tissues such as muscles-bones-spleen and skin[29].

Both of these multicenter studies currently lack published investigations of genetics, epigenetics, or transcriptomics; however, stored physical samples or data could represent untapped potential for large-scale exploration of genetic predispositions, environmental triggers, and genotype-to-phenotype discovery. Two similar efforts focused on genetic studies in diverse populations include MESARGEN (<https://mesargen.wordpress.com>), a

developing initiative to collect samples from population-based cohorts of patients and a large, case-control whole-genome sequencing and single-cell transcriptomic effort in African and European Americans (<https://omrf.org/patient-studies/sarcoidosis/>, <https://www.nhlbiwgs.org/>).

## Epigenomics

Known genetic risk factors do not fully explain sarcoidosis susceptibility or clinical variability, implicating the presence of epigenetic processes, underlying the pathogenesis of sarcoidosis. To date, only one epigenomics study in sarcoidosis has been performed[30\*\*]. This recent study must be highlighted for its 1) comparison of multiple clinical cohorts of granulomatous diseases, 2) use of two data modalities, gene expression and DNA methylation, and 3) adjustment for clinical covariates. Genome-wide DNA methylation and mRNA expression were both assessed by array in bronchoalveolar lavage (BAL) from patients with chronic beryllium disease (CBD, n = 8), beryllium sensitization without disease (BeS, n = 8), and sarcoidosis (n = 32). Initial epigenomics analyses found extensive, genome-wide significant DNA methylation changes in CBD compared to BeS, but not in the eight sarcoidosis patients used for this portion of the study compared to BeS. When CBD-associated methylation changes were examined in sarcoidosis patients, most changes were methylated in the same direction as CBD, but with a smaller magnitude of change, compared to BeS. Variability in DNA methylation and gene expression in sarcoidosis were also increased, potentially due to increased disease heterogeneity compared to CBD. A follow-up analysis of 9 progressive and 15 remitting sarcoidosis patients found 15,215 differentially methylated CpGs, although only 801 had greater than 5% methylation change in the progressive compared to the remitting disease group. Those which were also differentially expressed in progressive compared to remitting included genes with function in immunity. Although a small sample size limits this study, the results generated crucial basic knowledge of the methylation state in BAL immune cells in sarcoidosis, supporting the notion of epigenetic processes underlying the pathobiology of the disease.

## Transcriptomics

While there are a plethora of studies of gene expression[31–46], they are generally limited by small sample sizes, heterogeneous cell mixtures, and clinically variable cohorts. In addition, except for two small RNA sequencing studies limited to monocytes[45] and regulatory T cells[46], gene expression studies of both tissue and the circulating immune system have historically been limited to PCR- or microarray-based studies, both of which have significant shortcomings in comparison to RNA sequencing. I Despite these limitations, several recent studies have generated new findings or employed new analytics to identify novel gene expression differences in sarcoidosis.

### **New analytics methods: combining cohorts and data modalities**

A recent study perfectly exemplifies the utility of bioinformatically assessing publicly available data to generate translational findings[21\*\*]. The authors first presented a novel approach identifying new diseases in which a developed drug could be clinically useful (i.e., drug repositioning). This method utilized a multilayer self-organizing maps (ml-SOM)

machine learning technique. Briefly, ml-SOM uses multiple transcriptomic data sets of disease and drug effects to cluster co-regulated and/or functionally related networks across data sets (for an excellent visual description, see figure 2 of [21\*\*]). The authors employed ml-SOM first to confirm the action of infliximab in ulcerative colitis and Crohn's disease, namely, reduction of TNF-alpha related gene signatures. With the notable assumption that if two diseases share similar differentially expressed genes with respect to controls, then the same treatment will affect the two diseases in similar ways, the authors utilized ml-SOM to evaluate the repositioning possibility of infliximab to sarcoidosis. ml-SOM predicted that genes associated with the immune/inflammatory response and TNF-alpha/interferon and elevated in sarcoidosis would be beneficially impacted by infliximab treatment. This method was additionally used for brodalumab and other inflammatory diseases.

### Old analytics methods, new findings

A set of gene expression studies published in the past year applied RT-PCR or microarray to explore previously understudied cell subsets. First, T follicular helper cells, previously not well-studied in sarcoidosis, were isolated in 3 patients and 3 controls[47\*]. Cytokine mRNA levels were analyzed by qRT-PCR; IFN  $\gamma$  and IL-4 mRNA levels were found to be reduced in patients, while PD-1 and IL-17A mRNA were upregulated. Levels of IL-21, CD40, and IL-6 mRNA in T follicular helper cells did not differ compared with healthy controls, and gene expression did not seem to correlate with cytokine levels.

Similarly, several gene expression studies published in the past year applied RT-PCR or microarray to explore new gene targets in populations of cells that have been interrogated frequently in sarcoidosis research. Huppertz and colleagues employed RT-PCR to explore a new pathway: the NLRP3 inflammasome[48\*]. BAL cells, as well as lung and skin biopsies, were taken from 19 patients and 19 controls and examined for RNA levels of NLRP3 inflammasome pathway components. In sarcoidosis alveolar macrophages, levels of miR-223, a microRNA downregulating *NLRP3*, were decreased; *NLRP3* mRNA was correspondingly increased. The authors confirmed their findings in a model of pharmacological interference: both the NLRP3 pathway inhibitor MCC950 and an anti-IL-1 $\beta$  antibody resulted in reduced granuloma formation. Another recent gene expression study compared BAL cells in sarcoidosis (n=12) and idiopathic pulmonary fibrosis (n=9) [49\*], using standard microarray techniques. They found highly differing mRNA profiles of cells in the two diseases, with enrichment of ribosome biogenesis and proteasome apparatus in sarcoidosis, and neutrophilic dysfunction in idiopathic pulmonary fibrosis.

### Summary

Recent genomics studies in sarcoidosis have made advances in both new technologies and analysis methods. Gene-environment-wide interaction studies have identified interactions between sarcoidosis-associated genetic variants, smoking[5\*\*], and insecticide use[27\*\*], yet many other environmental exposures, such as mold and asbestos, remain to be examined. WES studies have expanded the breadth and depth of genetic studies, exploring the genetics of persistent sarcoidosis[17\*], identifying rare genetic variants[16\*], and bioinformatically assessing variants for potential function[16\*, 18\*]. However, these studies should be



replicated in both larger sample sets and other ethnicities. Neither WGS or genome-wide eQTL studies have been reported; this gap should be filled in order to confirm GWAS findings of imputed genotype and to “connect the dots” between genotype and phenotype (Table 3).

The majority of recent transcriptomic studies have generally used outdated and insensitive techniques. Transcriptomic studies based on total mRNA or small RNA using RNA-sequencing technologies will therefore be essential to conduct in-silico functional studies. Similarly, epigenomics of sarcoidosis could benefit from the application of large-scale genome-wide arrays such as the Illumina EPIC array that includes 850,000 methylation markers or methylation sequencing technologies, such as whole-genome bisulfite sequencing (WGBS), methylation capture sequencing (MethylCap-Seq), methyl-CpG binding domain sequencing (MBD-Seq), and more recently nanopore sequencing, that offer extensive knowledge to understand regulation in gene expression and consequently biological processes in health and disease beyond sarcoidosis (Table 3).

Finally, limited diversity in ancestry and clinical presentation of patients, few meta-analyses of multiple sarcoidosis studies, and lack of meticulous comparisons to other diseases leave us with findings in all aspects of genomics that are not broadly applicable. Although some studies explore common risk factors, like shared loci across disease[11\*], previous polygenic risk score[20] and cross-disease comparisons[21\*\*] should be repeated with recently proposed methodologies such as PRSice[50] and LDpred[51] software for calculating genetic risk scores and large-scale data to include recent findings and a greater breadth of the disease (Table 3).

## Conclusion

The genomics studies discussed here have made advances in both new technologies and analysis methods, yet still suffer from significant limitations. Next-generation sequencing, genome-wide whole-exome sequencing, spatial transcriptomics, single-cell RNA sequencing, whole-genome sequencing, and modern informatics methods like machine learning represent huge strides forward for genomics, yet have rarely been used in sarcoidosis. To identify broadly applicable genotype-to-phenotype connections, it is imperative that sarcoidosis research utilizes both large, diverse cohorts and the innovative techniques employed in other human complex diseases.

## Financial support and sponsorship:

This work was supported by grants from the Foundation for Sarcoidosis Research (Chicago, IL), the National Institutes of Health (R01HL113326-05, P30 GM110766-01, U54GM104938-06), and the Swedish Heart-Lung Foundation (Grant No. 20170664)

## References

1. Grunewald J, Eklund A. Sex-specific manifestations of Lofgren’s syndrome. *Am J Respir Crit Care Med.* 2007;175(1):40–4. Epub 2006/10/07. doi: 10.1164/rccm.200608-1197OC. [PubMed: 17023727]

2. Rybicki BA, Iannuzzi MC, Frederick MM, et al. Familial aggregation of sarcoidosis. A Case-Control Etiologic Study of Sarcoidosis (ACCESS). *Am J Respir Crit Care Med*. 2001;164(11):2085–91. doi: 10.1164/ajrccm.164.11.2106001. [PubMed: 11739139]
3. Fingerlin TE, Hamzeh N, Maier LA. Genetics of Sarcoidosis. *Clin Chest Med*. 2015;36(4):569–84. Epub 2015/11/26. doi: 10.1016/j.ccm.2015.08.002. [PubMed: 26593134]
4. Terwiel M, van Moorsel CHM. Clinical epidemiology of familial sarcoidosis: A systematic literature review. *Respir Med*. 2019;149:36–41. Epub 2018/12/28. doi: 10.1016/j.rmed.2018.11.022. [PubMed: 30587386]
- 5\*\*. Rivera NV, Patasova K, Kullberg S, et al. A gene-environment interaction between smoking and gene polymorphisms provides a high risk of two subgroups of sarcoidosis. *Scientific reports*. 2019;9(1):18633. doi: 10.1038/s41598-019-54612-1. [PubMed: 31819081] Gene by environment interaction study of sarcoidosis and smoking in both LS and non-LS identifying distinct sets of genetic variants in LS and non-LS that significantly interacted with smoking, and showed that risk for the disease substantially increased if an individual had a genetic predisposition and smoked.
6. Rivera NV, Ronninger M, Shchetynsky K, et al. High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *Am J Respir Crit Care Med*. 2016;193(9):1008–22. doi: 10.1164/rccm.201507-1372OC. [PubMed: 26651848]
7. Adrianto I, Lin CP, Hale JJ, et al. Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PloS one*. 2012;7(8):e43907. doi: 10.1371/journal.pone.0043907. [PubMed: 22952805]
8. Hofmann S, Fischer A, Till A, et al. A genome-wide association study reveals evidence of association with sarcoidosis at 6p12.1. *The European respiratory journal*. 2011;38(5):1127–35. doi: 10.1183/09031936.00001711. [PubMed: 21540310]
9. Hofmann S, Franke A, Fischer A, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nature genetics*. 2008;40(9):1103–6. doi: 10.1038/ng.198. [PubMed: 19165924]
10. Lareau CA, Adrianto I, Levin AM, et al. Fine mapping of chromosome 15q25 implicates ZNF592 in neurosarcoidosis patients. *Ann Clin Transl Neurol*. 2015;2(10):972–7. doi: 10.1002/acn3.229. [PubMed: 26478897]
- 11\*. Garman L, Pezant N, Pastori A, et al. Genome-wide association study of ocular sarcoidosis confirms HLA associations and implicates barrier function and autoimmunity in African Americans. *Ocul Immunol Inflamm*. 2020:1–6. Epub 2020/03/07. doi: 10.1080/09273948.2019.1705985. The first genome-wide association study (GWAS) of ocular sarcoidosis in an ethnically diverse cohort
12. Bello GA, Adrianto I, Dumancas GG, et al. Role of NOD2 pathway genes in sarcoidosis cases with clinical characteristics of Blau Syndrome. *American Journal of Respiratory and Critical Care Medicine*. 2015;192(9):1133–5. doi: 10.1164/rccm.201507-1344LE. [PubMed: 26517420]
13. Wolin A, Lahtela EL, Anttila V, et al. SNP Variants in major histocompatibility complex are associated with sarcoidosis susceptibility—a joint analysis in four European populations. *Frontiers in immunology*. 2017;8:422. doi: 10.3389/fimmu.2017.00422. [PubMed: 28469621]
14. Rivera NV, Hagemann-Jensen M, Ferreira MAR, et al. Common variants of T-cells contribute differently to phenotypic variation in sarcoidosis. *Sci Rep*. 2017;7(1):5623 Epub 2017/07/19. doi: 10.1038/s41598-017-05754-7. [PubMed: 28717140]
15. Soubrier F From an ACE polymorphism to genome-wide searches for eQTL. *J Clin Invest*. 2013;123(1):111–2. Epub 2013/01/03. doi: 10.1172/JCI66618. [PubMed: 23281417]
- 16\*. Kishore A, Petersen BS, Nutsua M, et al. Whole-exome sequencing identifies rare genetic variations in German families with pulmonary sarcoidosis. *Hum Genet*. 2018;137(9):705–16. Epub 2018/07/29. doi: 10.1007/s00439-018-1915-y. [PubMed: 30054724] Whole-exome study in German families applying analysis techniques designed to address linkage and high-penetrance to identify familial-clustered, functional, and rare genetic variants.
- 17\*. Lahtela E, Kankainen M, Sinisalo J, et al. Exome sequencing identifies susceptibility loci for sarcoidosis prognosis. *Front Immunol*. 2019;10:2964 Epub 2020/01/11. doi: 10.3389/

- fimmu.2019.02964. [PubMed: 31921204] Whole-exome study in Finnish patients identifying variants associated with persistent sarcoidosis.
- 18\*. Calender A, Lim CX, Weichhart T, et al. Exome sequencing and pathogenicity-network analysis of five French families implicate mTOR signalling and autophagy in familial sarcoidosis. *Eur Respir J*. 2019;54(2). Epub 2019/04/27. doi: 10.1183/13993003.00430-2019. Whole-exome study in French families identifying rare, functional genetic variants.
  19. Fischer A, Ellinghaus D, Nutsua M, et al. Identification of Immune-Relevant Factors Conferring Sarcoidosis Genetic Risk. *Am J Respir Crit Care Med*. 2015;192(6):727–36. doi: 10.1164/rccm.201503-0418OC. [PubMed: 26051272]
  20. Lareau CA, DeWeese CF, Adrianto I, et al. Polygenic risk assessment reveals pleiotropy between sarcoidosis and inflammatory disorders in the context of genetic ancestry. *Genes and immunity*. 2017;18(2):88–94. doi: 10.1038/gene.2017.3. [PubMed: 28275240]
  - 21\*\*. Arakelyan A, Nersisyan L, Nikoghosyan M, et al. Transcriptome-guided drug repositioning. *Pharmaceutics*. 2019;11(12). Epub 2019/12/18. doi: 10.3390/pharmaceutics11120677. A novel approach identifying new diseases in which a developed drug could be clinically useful utilizing a multilayer self-organizing maps (ml-SOM) machine learning technique.
  22. Li J, Yang J, Levin AM, et al. Efficient generalized least squares method for mixed population and family-based samples in genome-wide association studies. *Genet Epidemiol*. 2014;38(5):430–8. Epub 2014/05/23. doi: 10.1002/gepi.21811. [PubMed: 24845555]
  23. Schupp JC, Freitag-Wolf S, Bargagli E, et al. Phenotypes of organ involvement in sarcoidosis. *Eur Respir J*. 2018;51(1). Epub 2018/01/27. doi: 10.1183/13993003.00991-2017.
  - 24\*. Clevlen KL, Ye K, Zeig-Owens R, et al. Genetic variants associated with FDNY WTC-related sarcoidosis. *Int J Environ Res Public Health*. 2019;16(10). doi: 10.3390/ijerph16101830. Identifies genetic differences between firefighters who responded to the collapse of the WTC in 2001 and developed sarcoidosis and those with similar demographics, smoking rates, and exposure to WTC-associated dust that did not develop the disease.
  25. Diaz-Gallo L-M, Brynedal B, Westerlind H, et al. Understanding genetic interactions and assessing the utility of the additive and multiplicative models through simulations. *bioRxiv*. 2019:706234. doi: 10.1101/706234.
  26. Rossman MD, Thompson B, Frederick M, et al. HLA and environmental interactions in sarcoidosis. *Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders*. 2008;25(2):125–32.
  - 27\*\*. Chen Y, Adrianto I, Ianuzzi MC, et al. Extended methods for gene-environment-wide interaction scans in studies of admixed individuals with varying degrees of relationships. *Genetic epidemiology*. 2019;43(4):414–26. doi: 10.1002/gepi.22196. [PubMed: 30793815] A novel GxE method accounting for population admixture and related individuals, applied to African Americans, finding novel loci associated with insecticide exposure and sarcoidosis.
  28. Rybicki BA, Levin AM, McKeigue P, et al. A genome-wide admixture scan for ancestry-linked genes predisposing to sarcoidosis in African-Americans. *Genes and immunity*. 2011;12(2):67–77. doi: 10.1038/gene.2010.56. [PubMed: 21179114]
  29. Papiris SA, Georgakopoulos A, Papaioannou AI, et al. Emerging phenotypes of sarcoidosis based on 18F-FDG PET/CT: a hierarchical cluster analysis. *Expert Rev Respir Med*. 2020;14(2):229–38. Epub 2019/10/28. doi: 10.1080/17476348.2020.1684902. [PubMed: 31647341]
  - 30\*\*. Yang IV, Konigsberg I, MacPhail K, et al. DNA methylation changes in lung immune cells are associated with granulomatous lung disease. *American journal of respiratory cell and molecular biology*. 2019;60(1):96–105. doi: 10.1165/rcmb.2018-0177OC. [PubMed: 30141971] The first and only epigenomics study in sarcoidosis, comparing of multiple clinical cohorts of granulomatous diseases, utilizing both gene expression and DNA methylation data, and including clinical covariates.
  31. Judson MA, Marchell RM, Mascelli M, et al. Molecular profiling and gene expression analysis in cutaneous sarcoidosis: the role of interleukin-12, interleukin-23, and the T-helper 17 pathway. *Journal of the American Academy of Dermatology*. 2012;66(6):901–10, 10 e1–2. doi: 10.1016/j.jaad.2011.06.017. [PubMed: 21924794]

32. Su R, Li MM, Bhakta NR, et al. Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. *Eur Respir J*. 2014;44(4):985–93. Epub 2014/08/22. doi: 10.1183/09031936.00039714. [PubMed: 25142485]
33. Rosenbaum JT, Choi D, Wilson DJ, et al. Parallel gene expression changes in sarcoidosis involving the lacrimal Gland, orbital tissue, or blood. *JAMA Ophthalmol*. 2015;133(7):770–7. Epub 2015/04/17. doi: 10.1001/jamaophthalmol.2015.0726. [PubMed: 25880323]
34. Rosenbaum JT, Pasadhika S, Crouser ED, et al. Hypothesis: Sarcoidosis is a STAT1-mediated disease. *Clinical immunology (Orlando, Fla)*. 2009;132(2):174–83. doi: 10.1016/j.clim.2009.04.010.
35. Christophi GP, Caza T, Curtiss C, et al. Gene expression profiles in granuloma tissue reveal novel diagnostic markers in sarcoidosis. *Exp Mol Pathol*. 2014;96(3):393–9. Epub 2014/04/29. doi: 10.1016/j.yexmp.2014.04.006. [PubMed: 24768588]
36. Lockstone HE, Sanderson S, Kulakova N, et al. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med*. 2010;181(12):1367–75. Epub 2010/03/03. doi: 10.1164/rccm.200912-1855OC. [PubMed: 20194811]
37. Maertzdorf J, Weiner J 3rd, Mollenkopf HJ, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci U S A*. 2012;109(20):7853–8. Epub 2012/05/02. doi: 10.1073/pnas.1121072109. [PubMed: 22547807]
38. Bloom CI, Graham CM, Berry MP, et al. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS One*. 2013;8(8):e70630 Epub 2013/08/14. doi: 10.1371/journal.pone.0070630. [PubMed: 23940611]
39. Crouser ED, Culver DA, Knox KS, et al. Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med*. 2009;179(10):929–38. doi: 10.1164/rccm.200803-490OC. [PubMed: 19218196]
40. Gharib SA, Malur A, Huizar I, et al. sarcoidosis activates diverse transcriptional programs in bronchoalveolar lavage cells. *Respir Res*. 2016;17(1):93 Epub 2016/07/28. doi: 10.1186/s12931-016-0411-y. [PubMed: 27460362]
41. Schischmanoff PO, Naccache JM, Carrere A, et al. Progressive pulmonary sarcoidosis is associated with over-expression of TYK2 and p21Waf1/Cip1. *Sarcoidosis Vasc Diffuse Lung Dis*. 2006;23(2):101–7. Epub 2007/10/17. [PubMed: 17937105]
42. Monast CS, Li K, Judson MA, Baughman RP, et al. Sarcoidosis extent relates to molecular variability. *Clin Exp Immunol*. 2017;188(3):444–54. Epub 2017/02/17. doi: 10.1111/cei.12942. [PubMed: 28205212]
43. Zhou T, Zhang W, Sweiss NJ, et al. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. *PLoS One*. 2012;7(9):e44818 Epub 2012/09/18. doi: 10.1371/journal.pone.0044818. [PubMed: 22984568]
44. Koth LL, Solberg OD, Peng JC, et al. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *American Journal of Respiratory and Critical Care Medicine*. 2011;184(10):1153–63. doi: 10.1164/rccm.201106-1143OC. [PubMed: 21852540]
45. Talreja J, Farshi P, Alazizi A, et al. RNA-sequencing identifies novel pathways in sarcoidosis monocytes. *Sci Rep*. 2017;7(1):2720 Epub 2017/06/04. doi: 10.1038/s41598-017-02941-4. [PubMed: 28577019]
46. Kachamakova-Trojanowska N, Jazwa-Kusior A, Szade K, et al. Molecular profiling of regulatory T cells in pulmonary sarcoidosis. *J Autoimmun*. 2018;94:56–69. Epub 2018/07/28. doi: 10.1016/j.jaut.2018.07.012. [PubMed: 30049532]
- 47\*. Ly NTM, Ueda-Hayakawa I, Nguyen CTH, Okamoto H. Exploring the imbalance of circulating follicular helper CD4(+) T cells in sarcoidosis patients. *J Dermatol Sci*. 2020 Epub 2020/02/18. doi: 10.1016/j.jdermsci.2020.02.002. The first gene expression study of T follicular helper cells, previously not well-studied in sarcoidosis.
- 48\*. Huppertz C, Jager B, Wieczorek G, et al. The NLRP3 inflammasome pathway is activated in sarcoidosis and involved in granuloma formation. *Eur Respir J*. 2020;55(3). Epub 2020/01/18. doi: 10.1183/13993003.00119-2019. A gene expression study targeted toward a new pathway, the

NLRP3 inflammasome, found significant upregulation and evidence of effective pharmacological targeting.

- 49\*. Paplinska-Goryca M, Goryca K, Misiukiewicz-Stepien P, et al. mRNA expression profile of bronchoalveolar lavage fluid cells from patients with idiopathic pulmonary fibrosis and sarcoidosis. *Eur J Clin Invest.* 2019;49(9):e13153 Epub 2019/06/28. doi: 10.1111/eci.13153. [PubMed: 31246273] A gene expression study identifying distinct signature in two separate diseases, sarcoidosis and idiopathic pulmonary fibrosis.
50. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.* 2015;31(9):1466–8. doi: 10.1093/bioinformatics/btu848. [PubMed: 25550326]
51. Vilhjalmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *American journal of human genetics.* 2015;97(4):576–92. doi: 10.1016/j.ajhg.2015.09.001. [PubMed: 26430803]

**Table 1:**

Descriptions of genes or transcripts found to be associated with sarcoidosis in publications of the last two years

<b>Gene</b>	<b>Protein/RNA Name</b>	<b>Highlighted citation*</b>
MAG11	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1	[11]
AADACL3	Arylacetamide deacetylase-like 3	[16, 17]
C1orf158	Uncharacterized protein C1orf158	[16, 17]
KIR3DL1/KIRDS1	Killer cell immunoglobulin-like receptor 3DL1	[17]
LILRB4	Leukocyte immunoglobulin-like receptor subfamily B member 4	[17]
BTNL2	Butyrophilin-like protein 2	[24]
PACERR	PTGS2 Antisense NFKB1 Complex-Mediated Expression Regulator RNA	[24]
PTGS2/COX2	Prostaglandin G/H synthase 2	[24]
ANXA11	Annexin 11	[25]
BCL2	Apoptosis regulator Bcl-2	[25]
CCR7	C-C chemokine receptor type 7	[25]
CLEC10A	C-type lectin domain family 10 member A	[25]
CLIP1	CAP-Gly domain-containing linker protein 1	[25]
IL23R	Interleukin-23 receptor	[25]
NLRP7	NACHT, LRR and PYD domains-containing protein 7	[25]
PRDM1	PR domain zinc finger protein 1	[25]
HLA-DRA	HLA class II histocompatibility antigen, DR alpha chain	[27]
IFNg	Interferon gamma	[47]
IL-17A	Interleukin-17A	[47]
IL4	Interleukin-4	[47]
PDCD1	Programmed cell death protein 1	[47]
miR-223	MicroRNA-223	[48]
NLRP3	NACHT, LRR and PYD domains-containing protein 3	[48]
ACE	Angiotensin-converting enzyme	
ADAM33	Disintegrin and metalloproteinase domain-containing protein 33	
C6ORF67 (TMEM30A)	Cell cycle control protein 50A	
CCR2	Complement receptor type 2	
FUT9	4-galactosyl-N-acetylglucosaminide 3-alpha-L-fucosyltransferase 9	
HLA-DRB1	HLA class II histocompatibility antigen, DRB1 beta chain	
NOD2	Nucleotide-binding oligomerization domain-containing protein 2	
NOTCH4	Neurogenic locus notch homolog protein 4	
OS9	Protein OS-9	
PRDX5	Peroxiredoxin-5, mitochondrial	
RAB23	Ras-related protein Rab-23	
SLC11A1	Natural resistance-associated macrophage protein 1	
TGFB1	Transforming growth factor beta-1 proprotein	
TLR9	Toll-like receptor 9	
TNFA	Tumor necrosis factor	

---

<b>Gene</b>	<b>Protein/RNA Name</b>	<b>Highlighted citation*</b>
XAF1	XIAP-associated factor 1	

---

\*When the citation was highlighted in this review, its reference number is given

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Pathways found to be associated with sarcoidosis in publications of the last two years

Pathway	Study type	Highlighted citation*
Antigen presentation via HLA	Multiple	[11, 25, 26]
Barrier function	GWAS	[11]
Interferon or TNFa signaling	WES, ml-SOM	[16, 18, 21]
Bacterial defense	WES	[16]
IL-10 signaling	WES	[16]
Immune cell proliferation	WES	[16]
Vesicle-mediated transport	WES	[16]
Autophagy	WES	[18]
Leptin signaling	WES	[18]
mTOR signaling	WES	[18]
TCA cycle	WES	[18]
Immune/inflammatory response	ml-SOM	[21]
(Positive) regulation of cell proliferation	GxE (smoking)	[25]
(Positive) Regulation of gamma-aminobutyric acid secretion	GxE (smoking)	[25]
Adenylate cucleus-inhibiting G-protein coupled receptor signaling pathway	GxE (smoking)	[25]
Canonical Wnt signaling pathway	GxE (smoking)	[25]
Cell motility	GxE (smoking)	[25]
Cell-cell signaling by Wnt	GxE (smoking)	[25]
Feeding behavior	GxE (smoking)	[25]
Localization of cell	GxE (smoking)	[25]
Locomotion	GxE (smoking)	[25]
Non-canonical Wnt signaling pathway	GxE (smoking)	[25]
Peptidyl-tyrosine autophosphorylation	GxE (smoking)	[25]
Positive regulation of biological process	GxE (smoking)	[25]
Positive regulation of cellular metabolic process	GxE (smoking)	[25]
Positive regulation of cellular process	GxE (smoking)	[25]
Positive regulation of macromolecule biosynthetic process	GxE (smoking)	[25]
Positive regulation of macromolecule metabolic process	GxE (smoking)	[25]
Positive regulation of neurotransmitter transport	GxE (smoking)	[25]
Positive regulation of nitrogen compound metabolic process	GxE (smoking)	[25]
Positive regulation of nucleobase-containing compound metabolic process	GxE (smoking)	[25]
Regulation of animal organ morphogenesis	GxE (smoking)	[25]
Regulation of cell migration	GxE (smoking)	[25]
Response to hormone	GxE (smoking)	[25]
Response to nitrogen compound	GxE (smoking)	[25]
Response to organic substance	GxE (smoking)	[25]
Response to organonitrogen compound	GxE (smoking)	[25]
Signaling	GxE (smoking)	[25]
Wnt-signaling pathway	GxE (smoking)	[25]



<b>Pathway</b>	<b>Study type</b>	<b>Highlighted citation*</b>
Immunity	Genome-wide methylation studies	[30]
NLRP3 inflammasome	RT-PCR	[49]
Proteasome apparatus	RT-PCR	[49]
Ribosome biogenesis	RT-PCR	[49]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

Novel genomic methodologies not yet applied to sarcoidosis

Category	Methodology	Advantage
Genetics	Whole-genome sequencing	Observed rather than imputed data, full genome coverage, better copy number variant and rare SNP determination
Epigenetics	Genome-wide methylation array	Quantitative accuracy and reproducibility
Epigenetics	Methylation sequencing (e.g., whole-genome bisulfite sequencing)	High resolution (up to single-base)
Epigenetics	single-cell epigenetics (e.g., single-cell whole-genome bisulfite sequencing)	Localization of epigenetic changes to specific cell types
Transcriptomics	single-cell RNA-sequencing	Localization of transcriptomic changes to specific cell types
Transcriptomics	spatial transcriptomics (e.g., 10x Genomic's Visium)	Localization of transcriptomic changes to specific physical locations
Transcriptomics	small RNA sequencing (e.g., miRNA-seq)	Amplification of microRNA (miRNA), small interfering RNA (siRNA), and piwi-interacting RNA (piRNA)
Analytics	Meta-analyses	Increased power and diversity of subjects
Analytics	Cross-disease comparison methods (e.g., PRSice and LDpred)	High-resolution scoring with conditioning on covariates or linkage disequilibrium