UNIVERSITÀ DEGLI STUDI DI

MODENA E REGGIO EMILIA

Facoltà di Medicina e Chirurgia

Corso di Laurea Magistrale in Medicina e Chirurgia

Dipartimento di Scienze Mediche e Chirurgiche,

Materno-Infantili e dell'Adulto

Struttura Complessa di Malattie dell'Apparato Respiratorio

GENOME-WIDE ASSOCIATION FOR SARCOIDOSIS: A MULTI-ANCESTRY META-ANALYSIS FROM FOUR INTERNATIONAL DATASETS

RELATORE:

Dott.ssa **Stefania Cerri**

TESI DI LAUREA DI:

Brent Julius Marinduque

Anno accademico 2024-2025

TABLE OF CONTENTS

ABSTRACT

BACKGROUNDS

Sarcoidosis is a multisystem inflammatory disease that presents a wide range of clinical manifestations but mainly affecting the lungs and lymph nodes. The etiology is unknown, although it is assumed that multiple factors, including environmental factors and genetic risk, interplay into determining the phenotype, but the precise roles of the factors are still poorly clarified. As for genetic risk, multiple genes and loci have been associated to sarcoidosis through genome-wide association studies.

In our study we conducted a European ancestry, an African ancestry and a multi-ancestry meta-analyses of summary statistics of four GWAS for sarcoidosis. We aimed to identify novel ancestry-specific risk loci and compared our findings with available literature.

METHODS

We accessed public summary statistics from four international datasets (FinnGen freeze 12, Pan-UK Biobank Project, Million Veteran Project, and Biobank Japan) to conduct European (EUR) ancestry, African (AFR) ancestry and multi-ancestry meta-analyses, and performed risk loci plotting and annotation through FUMA GWAS. ANNOVAR and CADD were used for functional annotation. We identified novel risk loci and SNPs significantly associated to sarcoidosis and critically searched for studies in the literature that have previously identified or described a function that can be related to sarcoidosis.

RESULTS

We analyzed 9,659 cases and 1,665,804 controls from the four datasets. We found nineteen risk loci from EUR meta-analysis and two risk loci from AFR meta-analysis, including novel risk loci. We reviewed the following significantly associated to sarcoidosis SNPs and their nearest non-MHC genes with functional studies and GWAS available in literature: IL23R, PUS10, ACOXL, PLCL1, FAM117B, PPARG, ESYT2, ANXA11, CCDC88B, ATXN2,SH2B3, CCL24, HOMER2, CD19, UBASH3A, RNF215, MAT1A, RPS6KA4/MSK2, MACROD1, LAT,RP11-231C14.10,RP11-264B17.3 and SUMO2P.

CONCLUSION

In summary, we present the largest ancestry-specific GWAS meta-analysis for sarcoidosis (EUR and AFR) and the first multi-ancestry meta-analysis including EUR, AFR and East Asian ancestry (EAS) to date. A total of twenty-one loci were identified in EUR and AFR ancestry meta-analyses. We confronted significantly associated to sarcoidosis non-MHC genes and SNPs with available studies in the literature. We hope that our findings contribute to elucidate the role of genetics in sarcoidosis.

KEY WORDS: sarcoidosis, GWAS, meta-analysis, ancestry, risk loci

ABBREVIATIONS

6MWT six-minute walking test

18F-FDG-PET ^{18}F-fluorodeoxyglucose-positron emission tomography

1000G 1000 genomes

ACE angiotensin converting enzyme

AFR African

APC antigen presenting cells

ATP adenosin tryphosphate

ATS/ERS/WASOG American Thoracic Society/ European Respiratory Society/ World Association for Sarcoidosis and Other Granulomatous Disease

BAL bronchoalveolar lavage

BBJ Biobank Japan

CADD combined annotation dependent depletion

CD- cluster of differentiation

CT computer tomography

DLCO diffusing capacity of the lung for carbon monoxide

EAS East Asian

ECG electrocardiogram

EUR European

FEV1 forced expiratory volume in 1 second

FUMA functional mapping and annotation

FVC forced vital capacity

GWAS genome wide association study

HLA human leukocyte antigen

HRCT high resolution computer tomography

IFNγ interferon gamma

IL- interleukin

IPF idiopathic pulmonary fibrosis

LD linkage disequilibrium

MHC major histocompatibility complex

MRI magnetic resonance imaging

MVP Million Veteran Program

NF-κB nuclear factor kappa B

OR odds ratio

PFT pulmonary function test

PPARγ/ PPARG peroxisome proliferator-activated receptor gamma

rsID reference SNP cluster ID

RV residual volume

SNP single nucleotide polymorphism

TCR T cell receptor

TGFβ transforming growth factor beta

Th T helper

TLC total lung capacity

TNFα tumor necrosis factor alfa

Treg T regulatory

VC vital capacity

UKBB Pan-UK Biobank

UTR untranslated region

WES whole exome sequencing

WGS whole genome sequencing

SARCOIDOSIS

DEFINITION OF SARCOIDOSIS

Sarcoidosis is a multisystem inflammatory disease involving different organs but majorly affecting the lungs and intrathoracic hilar lymph nodes leading to non-necrotizing or non-caseating granulomas and adenopathy.

SHORT HISTORICAL OVERVIEW

Sarcoidosis was first described in 1869 by Jonathan Hutchinson, an English physician. A 58-year-old coal-wharf worker happened to come into his attention, showing purple tender skin plaques on both his hands and legs, but not complaining pain. He associated the lesions to gout which also affected the patient [1].

He later named the condition “Mortimer’s malady” after another patient he visited with red skin lesions on her face and forearms, not resembling lesions of other conditions like tuberculosis [2].

Up to that point physicians labelled sarcoidosis as an exclusive dermatological condition noticing cutaneous lesions that did not match any other known and frequent dermatological or systemic diseases of that time.

It was Caesar Boeck, a Norwegian dermatologist, who led to a new development of the definition of sarcoidosis. Not only he was the first to describe the granulomatous nature of

the condition and coined the term “sarkoid”, but also detected the involvement of multiple organs, including the lungs, lymph nodes and spleen [2].

Another milestone for sarcoidosis was when Christian Frederick Heerfordt, a Danish ophthalmologist, defined a syndrome presenting cutaneous lesions, uveitis, enlargement of the salivary glands, in particular the parotid, paresis of the facial nerve [1]. The syndrome was named after its discoverer.

In 1941 Sven Löfgren, a Swedish pulmonologist, observed a clinical presentation consisting of fever, bilateral hilar lymphadenopathy, polyarthritis and erythema nodosum. Again, the syndrome was named after its discoverer [2].

Collectively these observations contributed to the definition of sarcoidosis as we know it today, that is a complex multi-organ disorder with a wide range of clinical manifestations.

EPIDEMIOLOGY

Sarcoidosis is considered a rare pathology that can affect individuals of any age and ethnic group worldwide, although certain populations and age groups experience a greater number of cases compared to others. This indicates that multiple risk factors, including the difference in environment, lifestyle and occupation between countries, alongside with sex and age, play a role in the manifestation of the disease.

Regarding the age, sarcoidosis presents two peaks of incidence:

- Around the age of thirty
- Around the age of sixty. In this second peak the female population tends to have higher incidence [3].

The incidence and prevalence differ depending on the geographical region. The prevalence is higher in Western countries, most of all Canada and the Scandinavian countries with 140-160 subjects per 100,000, while Asian countries like Japan and South Korea show lower rates (1-5 cases per 100,000) [4].

The same applies to the incidence: higher rates in Sweden, Finland, North America and Australia (5-15 cases per 100,000), while in Japan and South Korea reported cases are comparatively low with 0.5-5 cases per 100.000 [4,5].

It is also important to acknowledge that there is a lack of data of the affected population in other areas such as Africa and South America. Therefore, further research in this field is needed to obtain a full picture, not only of the epidemiology of sarcoidosis but also of the risk factors that contribute to the manifestation of the disease and the prognosis.

Concerning ethnic background, different studies confirm that it is more prevalent in the Afro-American population with black females being more frequently affected [6,7]. On the contrary, Asian Americans and Hispanics are less likely to present with sarcoidosis, with females once more having higher prevalence [6,8]. This shows that both ethnicity and sex act as risk factors in determining an augmented susceptibility to sarcoidosis.

As for the mortality, an American study from 2018 that analyse its trend from 1999 to 2016 suggests that ethnicity is also involved in the prognosis of the disease as higher mortality is found in the Afro-Americans with average death rate of 15 cases per 1,000,000; in the white population the average death rate was markedly lower; in the Hispanic population the death rate was 0.8 cases per 1,000,000. It is noteworthy that in all the populations examined, females tend to have higher rates than males [9], indicating that sex is another contributing factor to the prognosis of the pathology.

ETIOLOGY AND RISK FACTORS

The etiology of sarcoidosis is currently unknown although numerous etiology hypotheses have been proposed. It is reasonable to assume that sarcoidosis is a multifactorial disease in which a wide range of factors interplay into determining an increased risk of developing the disease.

Aside from age and ethnicity, as stated in the previous section, there are additional factors, some more certain than others, that influence the susceptibility to sarcoidosis.

- Tobacco smoking: it is well known that smoke is a strong risk factor for various diseases, including cancer and respiratory diseases. First, it exposes to lung cancer due to its chemical composition which has a cancerogenic effect.

  Furthermore, it is also associated interstitial lung diseases, including IPF (idiopathic lung fibrosis) [10].

  As for sarcoidosis, the role of smoke remains unclear. A Swedish review and meta-analysis form 2024 shows that active smoking is linked to a 51% decreased risk of developing the disease [11]. However, another study examining Icelandic subjects shows a non-meaningful association between smoking and sarcoidosis [12].
- Non-infectious environmental exposure: different studies highlight an association between sarcoidosis and inorganic, work related agents. For instance, exposure to several metal dust, including beryllium silica and titanium has been connected with the disease onset [13,14].

  An interesting fact is the connection between the events of World Trade Center of September 2001 and the development in the following years of granulomas resembling those of sarcoidosis or sarcoid-like granulomatous pulmonary disease in first-response rescue workers due to dust exposure [15]. Nevertheless, it is difficult to

assess with certainty whether one inorganic agent was more strongly causative than another because of the heterogeneity of the dust composition [13].

Moreover, it is assumed that exposure to bioaerosol agents during Spring is linked to an increased manifestation [13], resembling allergic conditions.

- Infectious exposure: several studies show that there is an increased risk of developing sarcoidosis while being infected with microorganism. It has been observed that tissues from affected patients contained components linked to *Propionibacterium acnes* [14].
- Obesity: since it is a condition characterized by an underlying inflammatory state, it may favor the development of granulomas at a systemic level [5,16].

  Furthermore, due to the fact that obesity is characterized by multiple comorbidities, it can worsen the prognosis of the disease.

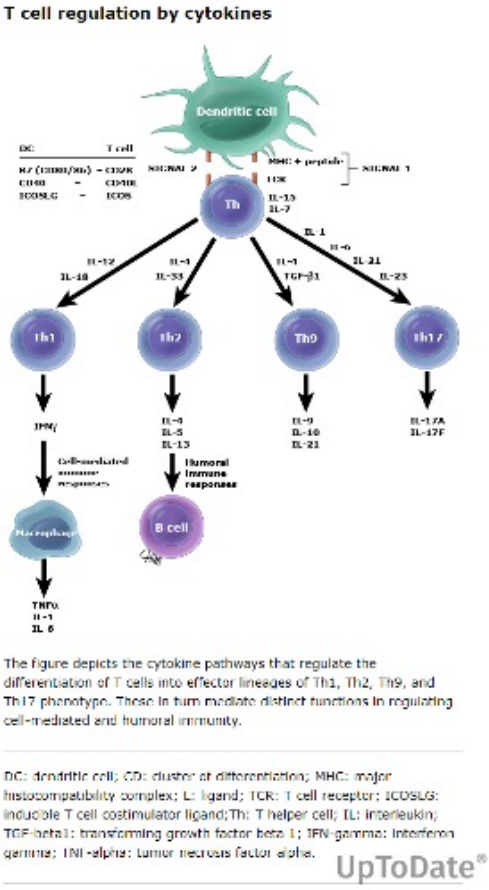
  However, a study shows that there is no correlation between body mass index (BMI) and the rates of sarcoidosis [17]. This finding questions the role of obesity as a risk factor and therefore more research is needed to elaborate on this aspect.
- Autoantigens: the influence of autoantigens in sarcoidosis have been investigated in a study by Wahlström et al. in which they isolated lung cells in the bronchoalveolar lavage fluid (BAL fluid) of HLA.DRB1*03:01 patients affected by sarcoidosis. They eluted peptides from HLA-DR and identified, through liquid chromatography-mass spectrometry, aminoacidic sequences, including vimentin and ATP synthetase. In particular, vimentin is found in fibroblasts and in asteroid bodies of sarcoid granulomas [18,19].
- Genetics: the role of genetics in sarcoidosis will be explored in a subsequent chapter.

IMMUNOPATHOGENESIS

While the exact etiology of sarcoidosis remains unclear, significant progress has been made in terms of its pathogenesis.

Per definition, sarcoidosis is a systemic disease underlying a dysregulated inflammatory response. This indicates that there is an activation of the immune system with T cells, B cell and different cytokines and chemokines being recruited to support the pathological development and the formation of sarcoid granulomas.

Figure 1. Immune response



The figure depicts the cytokine pathways that regulate the differentiation of T cells into effector lineages of Th1, Th2, Th9, and Th17 phenotype. These in turn mediate distinct functions in regulating cell-mediated and humoral immunity.

DC: dendritic cell; CD: cluster of differentiation; MHC: major histocompatibility complex; L: ligand; TCR: T cell receptor; ICOSLG: inducible T cell costimulator ligand;Th: T helper cell; IL: interleukin; TGF-beta1: transforming growth factor beta 1; IFN-gamma: interferon gamma; TNF-alpha: tumor necrosis factor alpha.

The pathway of immune cells activation is illustrated in Figure 1.

It is widely accepted that the first step of the immunopathogenesis of sarcoidosis is the phagocytosis of a yet-to-be identified antigen by antigen presenting cells (APCs) which are found in various tissues. APCs includes dendritic cells and macrophages [20].

Dendritic cells fall within the innate immune cells population and secrete cytokines and chemokines to promote the inflammatory process [21].

These cells are also located in the epithelium and interstitium of the lung where they monitor the pulmonary microenvironment [22]. Once the antigen is recognized as a result of the binding with class II MHC, dendritic cells migrate towards the lymph nodes where the MHC-antigen complex is recognized by T cell receptors, thus stimulating the differentiation of CD4+ T helper cells (Th cells) into effector Th cells through the action of cytokines they secrete. For instance, Th1 polarization is mediates by the action of interleukin 12 (IL-12) [23].

Another type of cells that contribute to the granuloma development are macrophages. In the lungs macrophages are found within the alveoli. They play a similar action as dendritic cells. After the recognition of an unknow foreign antigen they promote mobilization and transformation of monocytes, and they accumulate into forming epithelioid cells [20] and finally they fusion into multinucleated giant cells in response to various cytokines, including IL-4 and interferon gamma (IFNγ) [24].

It is worth to distinguish the two possible polarizations as well, that macrophages undergo:

- M1 polarization is the classical polarization of macrophages triggered by the main cytokines involved in sarcoidosis which are IFNγ and tumor necrosis factor alpha (TNFα). This process includes the activation of interferon-regulatory factors by nuclear factor kappa B (NF-κB) [25]. M1 macrophages supports the development of granulomas.

- M2 polarization is an alternative polarization of macrophages triggered by other agents such as peroxisome proliferator-activated receptor gamma (PPARγ) [23]. It is suggested that there is shift of M1 to M2 macrophages which may have a role in a chronic phase of sarcoidosis characterized by pulmonary fibrosis [26].

As shown in Figure 1, the second step of the immunopathogenesis of sarcoidosis is the differentiation of the activated CD4+ Th cells into Th cells effectors whose action is to trigger other components of the immune system, including B cells, thus activating the humoral response, and more macrophages. This process employs a broad range of cytokines. For example IL-1β, IL-6, IL-12and IL-23 are involved [23].

CD4+ Th cells are present in the BAL fluid and blood samples of sarcoidosis affected patients but with contrasting features. While an increased cellularity is observed in the BAL fluid, in peripheral blood samples there is a CD4+ T cell lymphopenia and a reduced CD4+/CD8+ ratio [20]. A possible explanation for these findings is that CD4+ T cells are being recruited in the affected organs, primarily the lungs, where they contribute to granulomas formation and sustain the disease process.

The main effector Th cells are Th1 and Th17.

- Th1 effector cells: the cytokines that promote Th1 polarization are IL-12 and IL-18 (Figure 1).

  The two aforementioned cytokines are mainly produced by dendritic cells and macrophages [23].

  A study from Shigehara et al. shows that increased levels of IL-12 and IL-18 led to an elevated production of IFNγ by BAL fluid cells of affected patients, providing evidence

for the role of Th1 response in sarcoidosis [27], as IFNγ is one of the key cytokines secreted by Th1 effector cells.

Th1 cells not only secrete IFNγ, but also IL-12, TNFα and chemokines [23].

IFNγ and TNFα are involved in the development and persistence of sarcoid granulomas [28].

Specifically, IFNγ stimulates the formation of multinucleated giant cells [24], a critical cellular component of these lesions.

As form chemokines, Th1 cells secrete CXCL9, CXCL10, CCL2 [23]. Chemokines are a special type of cytokines with the main function of chemotaxis, which is the process that facilitates immune cell migration towards inflammatory sites.

- Th17 effector cells: in comparison to Th1, different cytokines lead to Th17 polarization such as IL-1, IL-6, IL-21, IL-23, as illustrated in Figure 1.

  Th17 cells participate as well in granuloma pathogenesis. They secrete IFNγ and IL-17 [23]. Specifically, Th17.1, a presumed progeny of Th17 cells, are the most represented memory CD4+ T cells in affected lungs and mediastinal lymph nodes. On the contrary, Th17.1 levels are low in peripheral blood [29], reflecting a situation similar to that of Th1.

  What's more is that Rames et al. reported that corticosteroids have no effect on Th17.1 cells hinting a negative role in the chronic development of inflammatory conditions [30].

Other T cells that contribute to the pathogenesis of the disease are:

- Th2 effector cells: Figure 1 shows that il-4 and IL-33 are needed for Th2 polarization. Th2 cells produces IL-4, IL-5, IL-13 which stimulated humoral immune response from B cells.

It is known that later stages of sarcoidosis are characterized by fibrosis in affected organs. Fibrosis is associated with poor prognosis of the disease. It has been proposed that a transition from Th1 to Th2 activity might be the cause of fibrosis [31], although current evidence is insufficient for this hypothesis [23].

- Tregs: T regulatory lymphocytes or Tregs are specific T cells capable of downregulating immune response [20]. They secrete cytokines with anti-inflammatory properties, including IL-10, TGFβ (transforming growth factor beta) and IL-35 [32].

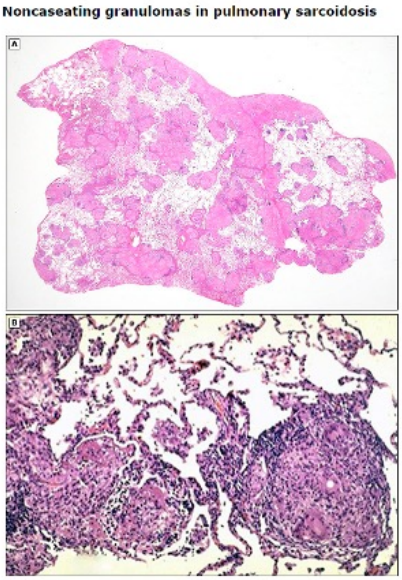
  Several studies demonstrate that sarcoidosis affected patients tend to have higher levels of Tregs but it is debated whether it is more elevated in the lungs or in peripheral blood [23].

HISTOPATHOLOGY

Sarcoid granulomas can develop in any target organ, but mostly they are found within the lungs and hilar lymph nodes. In the lungs they are located in the peri-bronchovascular interstitium and in the subpleural region, where lymphatic vessels are present.

Figure 2 in the following page depicts an example of a pulmonary non-caseating granuloma.

Figure 2. Non-caseating lung granuloma



Copyrights apply

Granulomas can be divided into a central area and a periphery. The central area is composed of epithelioid cells, multinucleate giant cells and CD4+ T cells and in rare cases it undergoes necrosis, unlike tuberculous granuloma.

Encircling the central area are CD8+ and CD4+ T cells, B cells, monocytes, mast cells and fibroblast. In the outermost layer, hyaline collagen forming lamellar structures can be observed [20].

Other structures that are usually observed in granulomas are Schauman bodies which are concretions of calcium and proteins, and asteroid bodies or stellate inclusions, They are located in the proximity of multinucleate giant cells [33].

CLINICAL MANIFESTATIONS

The clinical manifestation of sarcoidosis is a very heterogeneous as it has a systemic involvement. Since the main target organs are the lungs, the affected patient is most likely to present for a visit to a pulmonologist.

A patient can be asymptomatic, and the disease is identified accidentally for example through suspicious blood tests or radiological imaging conducted for other indications.

Symptomatic cases exhibit a wide variety of clinical features depending on the organ involvement and some of the symptoms are included into recognizable clinical syndromes.

The clinical presentation can be distinguished into two forms, depending on the progression of the disease:

- Acute sarcoidosis: the acute form is characterized by organ specific symptoms. Of note, a relevant clinical pattern known as Löfgren syndrome, consisting of erythema nodosum, bilateral hilar lymphadenopathy, fever and polyarthritis is observed. Most patients with acute manifestation tend to undergo a spontaneous resolution and relapses are not common [34].

- Chronic sarcoidosis: it is characterized by subtle onset, more progressive presentation and recurrence even after treatment [34]. Again the symptoms are variable, but chronic forms are more likely to present with uveitis and lung fibrosis for example [35].

This section provides an overview of symptomatology categorized by organ and system.

Table 1 contains a summary of the affected organs in percentage [36].

Table 1. Organ involvement in sarcoidosis within six months of diagnosis (ACCESS study)

Organ involvement in sarcoidosis within six months of diagnosis (ACCESS study)

Organ involvement	Number*	Percent
Lungs	699	95
Skin¶	117	15.9
Lymph node	112	15.2
Eye	87	11.8
Liver	85	11.5
Erythema nodosum	61	8.3
Spleen	49	6.7
Neurologic	34	4.6
Parotid/salivary	29	3.9
Bone marrow	29	3.9
Endocrine	27	3.7
ENT	22	3
Cardiac	17	2.3
Renal	5	0.7
Bone/joint	4	0.5
Muscle	3	0.4

ENT: ear, nose, and throat.

* Total n = 736.

¶ Excluding erythema nodosum.

Reproduced with permission from: Baughman, RP, Teirstein, AS, Judson, MA, et al. Clinical characteristics of patients in case control study of sarcoidosis. Am J Respir Crit Care Med 2001;164:1885. Official Journal of the American Thoracic Society. Copyright © 2001 American Thoracic Society.

UpToDate®

Most of the symptoms are highly non-specific for sarcoidosis, that can be found in many different conditions triggered by different mechanism of that of sarcoidosis; therefore, it is important to consider all the manifestations collectively to facilitate the detection of the disease.

Pulmonary sarcoidosis

In most cases symptomatic patients presents with non-specific symptoms such as cough (27-53% of cases); dyspnea (18-51% of cases) and thoracic pain (9-23% of cases) [35]. Thoracic pain is frequently located in the infrascapular, substernal or subcostal areas [36]. The

retrosternal pain and dyspnea with a perception of bronchial obstruction are mainly due to lymphadenopathy. At physical examination no particular alterations are evident. Sometimes fine crackles can be appreciated on auscultation which underlines fibrosis.

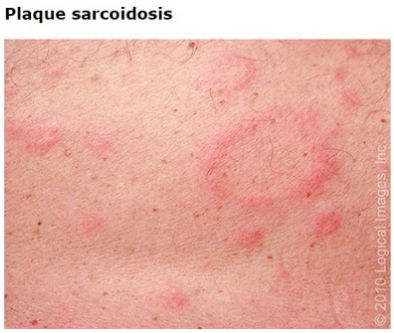
A study by Boucly et al. shows a correlation between pulmonary hypertension and sarcoidosis [37]. Pulmonary hypertension is a condition that not only causes damages to lung function, but it leads to heart failure with signs of congestion in an early phase.

It also involves the pleura with pleural effusions, pleural nodules and there is a higher risk of developing aspergillosis [34].

Cutaneous sarcoidosis

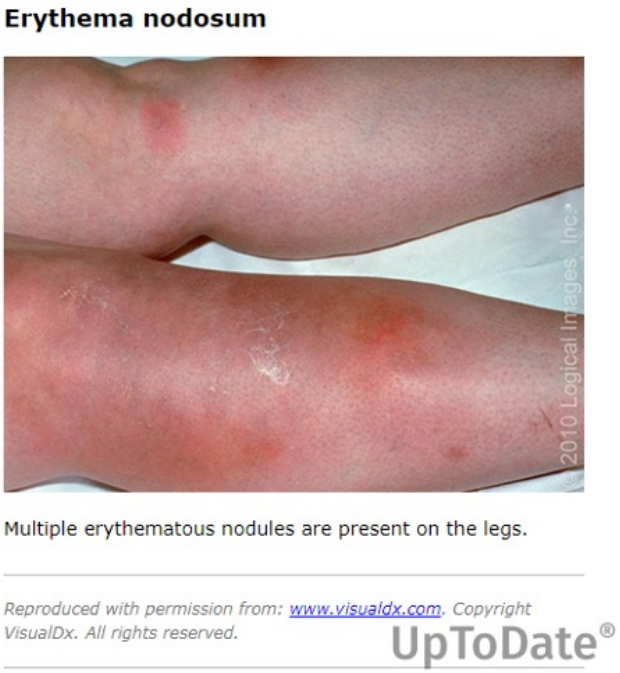
Skin lesions are the most frequent clinical manifestation after pulmonary sarcoidosis. They can be divided into sarcoidosis-specific skin lesions that histologically are granulomas, and non-specific skin lesions like papules and plaques (Figure 3) ranging in color from skin-toned to yellowish or violaceous, which are predominant in upper and lower limbs, head and neck region. Nodules of subcutaneous origin may also be found in the limbs [6].

Figure 3. Sarcoidosis plaque



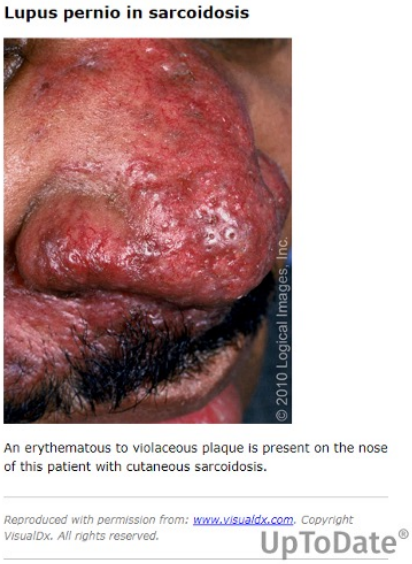
Another skin presentation of sarcoidosis is erythema nodosum (Figure 4) consisting of tender erythematous nodules. They can be found typically in the lower limbs and span diameters from 1 o 2 cm [34].

Figure 4. Erythema nodosum



Lupus pernio is a clinical subtype of sarcoidosis, consisting of facial plaques which may be erythematous or purplish, thickened and infiltrative (Figure 5). This condition might be mistaken for other plaques presenting disease for example lupus erythematosus and lupus vulgaris [38].

Figure 5. Lupus pernio



Eyes

Regarding the eyes, sarcoidosis is associated with uveitis, generally bilateral. Sarcoid uveitis is an immunomodulated non-infectious form of uveitis which is the inflammation of the uvea, a component of the vascular tunic of the eye, the other one being the choroid.

Patients complain ocular pain and redness, visual alterations such as myodesopsia or vitreous floaters. On the other hand, decline in visual acuity and blindness is infrequent [6,34].

Anterior uveitis is typically common in patients of Caucasian ethnicity. Posterior uveitis is a typical finding in the black population [6].

Joints

In comparison to lung and eye involvement, arthropathy is less common with cases ranging from 5% to 15% [6]. On the contrary, arthralgia is more likely a referred symptom by patients[34] and myopathy is rare with less than 3% of cases [35].

Generally, patients present with oligoarthritis of large joints such as the ankle articulation in 90% of cases while arthritis of small articulations are rare [6,35].

Blood

Patients may present some clinical features like fatigue, skin pallor, increased risk of infection due to hematological abnormalities. In fact, anemia, leukopenia and lymphopenia are quite common (40% of cases) [34].

Lymphatic system

Sarcoidosis is generally associated with adenopathy. Hilar and mediastinal lymph nodes are the most frequently involved, but other lymph nodes and organs of the lymphatic system may be sites of disease involvement.

Specifically, peripheral lymph nodes are affected in 40% of cases and spleen manifestations include granulomas. It is rare to note palpable spleen on physical examination [39].

Liver and Gastrointestinal tract

Liver manifestations are not common and tend to be asymptomatic. When symptoms are present, they are often non-specific in nature. Such symptoms are abdominal pain, pruritus, hepatomegaly [39] and blood test may reveal alterations in biochemical markers of cholestasis, for example increased values of alkaline phosphatase and gamma glutamyltransferase [6].

The performance of a liver biopsy reveals non-specific non-caseating granulomas in the portal and periportal areas [6,39].

In reference to the gastrointestinal tract, all organs are potentially susceptible to sarcoidosis. In particular, stomach involvement results in peptic ulcers with symptoms of epigastric pain and dyspepsia [40].

Heart

As for the heart, the inflammation targets the myocardium. Typical symptoms are heart palpitations and syncope, probably because of arrythmias, the most frequent of which being atrioventricular block [6]. Patients may also be affected by cardiomyopathy, a condition, that, if not treated, progress to heart failure [35], often the congestive form.

Kidneys

The main manifestation of renal sarcoidosis is hypercalciuria which is associated with nephrolithiasis and nephrocalcinosis [39], which are respectively the formation of calcium

stones and calcium deposition in the renal parenchyma. If untreated, affected patients may develop chronic renal failure [39], needing renal replacement therapy which is dialysis.

Neurologic sarcoidosis

Neurologic involvement of sarcoidosis can be divided into central and peripheral manifestations:

- Central manifestations include brain granulomas which are very rare and can lead to epileptic seizures. More common are the inflammation of cranial nerves, often bilateral, including the optic nerve (II cranial nerve) and facial nerve (VII cranial nerve) leading to optic neuritis and Bell's palsy respectively [6,41]. The latter is also part of a clinical syndrome known as Heerfordt syndrome, consisting of swelling of the parotid gland, fever and uveitis [35]. Further central manifestations might be myelopathy and aseptic meningitis presentations [41].
- Peripheral manifestations: affected patients may develop mononeuropathy or polyneuropathy of both sensory and motor nerves [41].

Other manifestations

Sarcoidosis may affect other organs such as salivary glands, the upper respiratory tract, including the nose and larynx [39]. There is also a curious case reported by Moretti et al. of prostatic sarcoidosis [42].

DIAGNOSIS

When there is clinical suspicion of sarcoidosis, whether its involvement is initially limited to the lungs or there are additional signs and symptoms of organ damage, besides referring the patient to the appropriate specialists (pulmonologist, ophthalmologist, etc.), different tests and radiologic investigations are needed.

Regarding laboratory tests:

- Blood tests are the first step to gain a comprehensive understanding of the correct bodily functioning. An affected patient may exhibit anemia, leukopenia and lymphopenia [34] but they are highly non-specific. As mentioned before, a diminished CD4/CD8 ratio can be noticed [20].

  It is also important to investigate liver and renal functioning as alterations can occur.

  Moreover, there are a few biochemical markers, some more helpful than others, that can be measured:

  - Inflammatory markers: C-reactive protein and erythrocyte sedimentation rate but they are also non-specific.
  - Serum ACE (angiotensin converting enzyme) levels: higher the values, greater is the specificity. About 75% of patients are positive for elevated ACE [36].
  - Other markers that have limited utility in diagnosis are serum A amyloid and IL-2 levels [36]
  - Calcium levels, for increased risk of nephrolithiasis and nephrocalcinosis [39].

It is noteworthy that these values, considered individually, are not useful, but it is crucial to associate them with clinical signs and symptoms.

- Urine analysis: hypercalciuria is common.

- Tuberculin skin test or Mantoux test, for differential diagnosis between sarcoidosis and *Mycobacterium Tuberculosis* infection [36].
- BAL fluid is obtained through bronchoscopy, with the purpose of aiding differential diagnosis, for example infections and eosinophilic lung disease. Moreover CD4/CD8 ratio is increased in comparison to the blood ratio and it accounts for above 90% specificity [36].

Concerning radiologic investigations:

- Chest radiograph is vital to identify the common manifestations of sarcoidosis being the lung and bilateral hilar adenopathy. Based on radiograph findings, four stages are identified as shown on Table 2.

Table 2. Stages of radiographic involvement in pulmonary sarcoidosis

Stages of radiographic involvement in pulmonary sarcoidosis

Stage	Radiographic findings
I	Bilateral hilar adenopathy
II	Bilateral hilar adenopathy and reticular opacities (predominantly in upper lobes)
III	Reticular opacities (predominantly in upper lobes) without hilar lymphadenopathy
IV	Reticular opacities, volume loss, conglomerated masses with traction bronchiectasis. Calcification and cavitation or cyst formation may be present.
Nodular sarcoid	Multiple, bilateral lung nodules and minimal hilar adenopathy

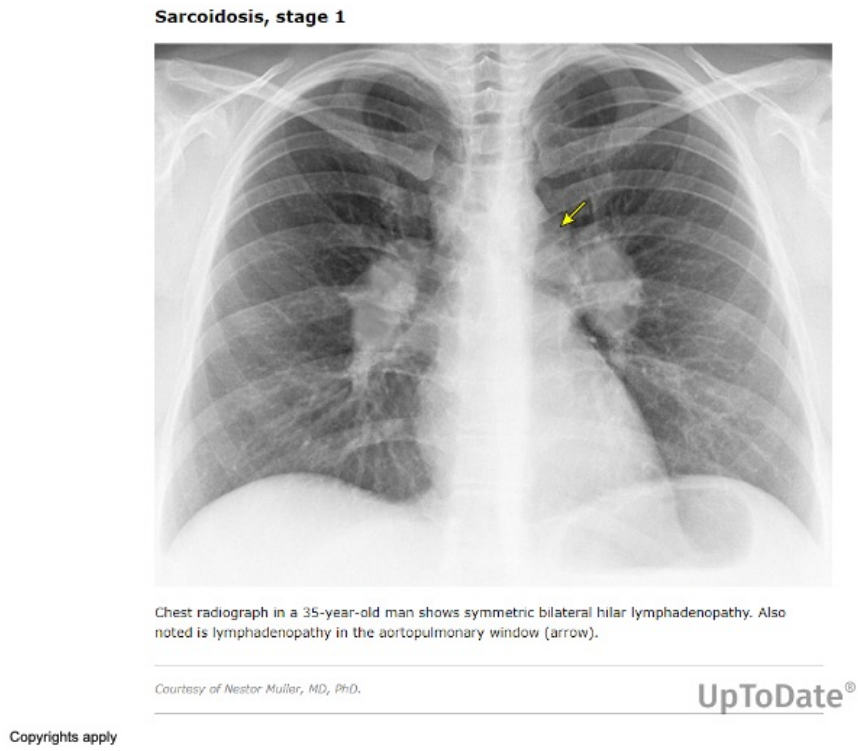
Reference:

1. Scadding JG. Prognosis of intrathoracic sarcoidosis in England: A review of 136 cases after five years' observation. Br Med J 1961; 2:1165.

UpToDate®

- Stage I (Figure 6) is characterized by bilateral hilar adenopathy (50% of cases initially presents this finding) [36].

Figure 6. Stage I (chest radiograph)

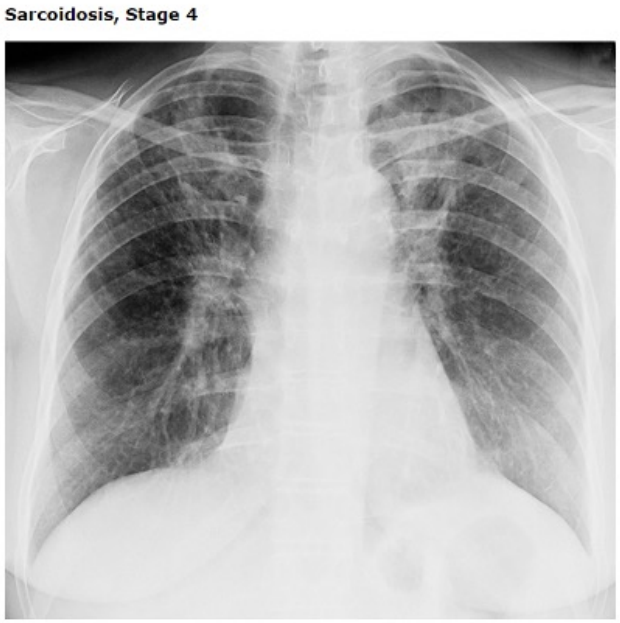


- Stage II, in addition to bilateral hilar adenopathy, pulmonary parenchymal infiltrates are appreciated [35]

- Stage III, only pulmonary parenchymal infiltrates are detected.

- Stage IV (Figure 7) is associated with a worsening of respiratory symptoms due to the presence of pulmonary fibrosis with reticular opacities. This stage of the disease may be marked by the development of cavitations or cysts [6,36].

Figure 7. Stage IV (chest radiograph)



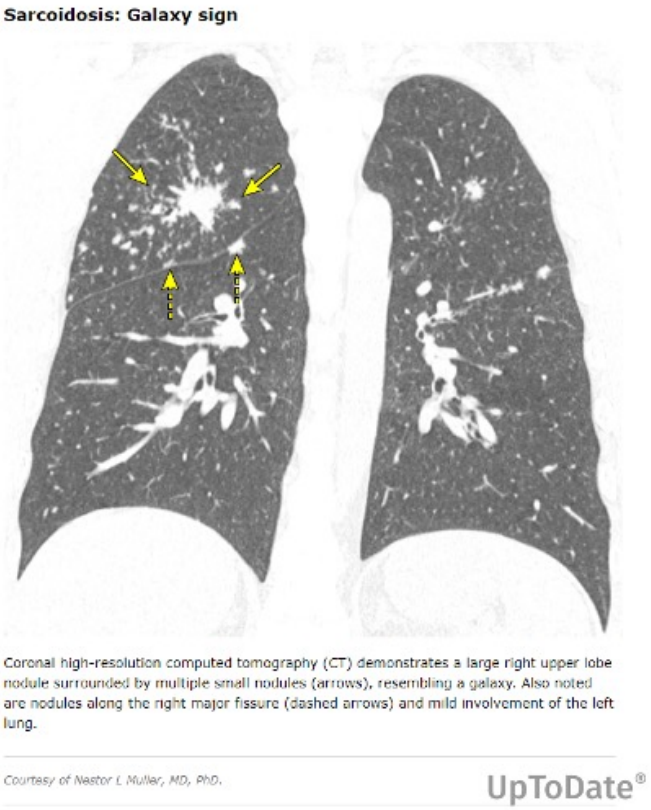
- High resolution CT scan (HRCT) allows greater image detail in comparison to chest radiograph.

  In contrast to conventional CT scan, HRCT does not require contrast agents and allows for thinner image slices.

  It assists in further investigating or clarifying symptoms that lack chest radiological correlation. HRCT can identify slightly increased size of lymph nodes in the hilar and mediastinal regions, ground glass opacifications, cysts, fibrosis, nodules, etc. [36]

  “Galaxy sign” (Figure 8), consisting of a central large nodule and peripheral micronodules or small and large nodules fused, is a pattern strongly indicative of sarcoidosis [35].

Figure 8. HRCT “galaxy sign”



Copyrights apply

- MRI or magnetic resonance imaging is not frequently used. Indications for MRI with gadolinium are heart and neurological involvement [6].
- 18F-FDG-PET (^{18}F-fluorodeoxyglucose-positron emission tomography) also assists in case of cardiac sarcoidosis although it is not helpful in terms of differential diagnosis to other malignant conditions as they, like sarcoidosis, can test positive[6,36].

Further relevant examinations are electrocardiogram (ECG), pulmonary function tests (PFTs), DLCO and 6MWT.

ECG is an important first-line test to investigate if the patient has cardiac involvement, for instance arrythmias like BAV or supraventricular tachycardia [6].

PFTs are essential to determine alterations in respiratory functions and are an indicator of the disease progression as observed in lung issue analysis comparing mild and severe form of sarcoidosis. Generally, PFTs results show a restrictive pattern [35].

Restrictive pattern is characterized by reduced inspiration flow and normal or near-normal expiration flow in the flow-volume curve; reduced lung volumes such as vital capacity (VC), total lung capacity (TLC), forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV1), normal or slightly increased Tiffenau index (FEV1/FVC x 100%), normal or decreased residual volume (RV).

Not only sarcoidosis presents a restrictive pattern but many other lung diseases (i.e. IPF), thus making it essential to associate PTFs findings with clinical examination.

DLCO or diffusing capacity of the lung for carbon monoxide is a functional test to assess if the alveolar-capillary membrane is intact, therefore allowing regular gas exchange between the alveoli and the capillaries. DLCO is reduced in restrictive disease.

As for sarcoidosis, DLCO is reduced from 15% to 20% of normal values [6].

6MWT or six-minute walking test is a diagnostic tool to evaluate if the affected patient presents progressive desaturation during an exertion.

It consists in walking back and forth along a flat corridor for 400-700 m over a period of at least 6 minutes and one of its indications is for patients with lung fibrosis [43], a possible manifestation of sarcoidosis [34].

However, an histological examination is required for definitive diagnosis: the main instruments used for pulmonary sarcoidosis are fine-needle aspiration through endoscopy with the aid of ultrasound, cryobiopsy of mediastinal lymph nodes and lung transbronchial biopsy [6,36].

GENERAL THERAPEUTIC APPROACH

The therapeutic approach for sarcoidosis is variable because it can affect different parts of the body and it also depends on the presence or absence of symptoms.

Sarcoidosis is generally characterized by spontaneous resolution and no permanent organ damage, and if the patient is asymptomatic, no therapy is needed [6].

On the other hand, if the patient is symptomatic, there are different guidelines depending on the affected system but generally, as it is an inflammatory disease, the first step is treatment with corticosteroids.

As for lung manifestations, first-line therapy is represented by oral corticosteroids [6]. The ATS/ERS/WASOG recommends prednisone, 20-40 mg daily [44].

If the patient relapses despite treatment with corticosteroids, second-line therapy is needed. Methotrexate is an immunosuppressive drug that is used not only after corticosteroids failure but also as a steroid sparing agent to avoid steroid toxicity. Other drugs included in this group are azathioprine, mycophenolate and antimalarials such as chloroquine [6,34].

Third-line therapy are reserved for patients intolerant to the previous drugs and are refractory to first and second-line treatments. TNFα inhibitors are part of this group such as infliximab and adalimumab [6].

With recent advances in the genetics of sarcoidosis, specifically the detection of genes that are associated to an increased susceptibility to it, attempts to find treatments have been made in the field of targeted therapy.

An example is gene IL23R, part of the IL-23 signaling pathway, which is targeted indirectly by Ustekinumab, an anti-IL-12 and anti-IL-23 monoclonal antibody; however the attempts to use it were unsuccessful [45].

For other manifestations of sarcoidosis, corticosteroids represents again the first-line therapy with additional specific treatment according to the system affected [6].

GENETICS OF SARCOIDOSIS

GENETIC RISK

Sarcoidosis is an inflammatory disease underlying an immune-mediated mechanism. To date, the precise etiology of sarcoidosis is still unknown. However, multiple hypotheses have been made throughout the years.

There are many risk factors that may increase the susceptibility to sarcoidosis. As mentioned earlier, sarcoidosis is a multifactorial disease, meaning that all presumed factors, including environmental and occupational, ethnic background etc. interplay into determining a major probability of manifesting the disease.

Like any other disease of multifactorial origin, genetics could play an important role in sarcoidosis.

Currently, there is no known specific gene mutation that causes disease manifestation with a mendelian inheritance like cystic fibrosis, an autosomal recessive disease caused by mutations in the CFTR gene. It is conceivable that multiple genes have a role in determining a predisposition to sarcoidosis.

As for sarcoidosis, several studies in medical literature have been conducted with the aim of finding more about its genetic susceptibility.

First of all, there is evidence of familial clustering in sarcoidosis. For example, a case control family study on Swedish subjects conducted in 2018 found that familial relative risk increases as the number of affected first-degree relatives increases, in particular, a relative risk of 4.69 when the subject had at least two first-degree relatives affected. The authors

also found that genetic heritability was 39%, meaning that other factors have a major role in establishing a stronger susceptibility to sarcoidosis [46].

Secondly, it has been noticed that monozygotic twins have a higher probability of presenting sarcoidosis than dizygotic twins by 2-4 times [34].

Furthermore, belonging to a specific ethnic group, that shares common genetic traits, increases the risk of sarcoidosis: African Americans tend to have higher cases of sarcoidosis in contrast to East Asians [4]. This suggests that there are specific risk loci that differ according to the ancestry of an individual.

Different gene regions have been linked to an elevated risk of sarcoidosis, indicating that the genetic basis of the disease susceptibility is heterogeneous.

One of the most studied gene regions in sarcoidosis is the human MHC or major histocompatibility complex.

MHC is situated within chromosome 6p21 (p stands for “petit” and it is the short arm of the chromosome, 2 is the region, 1 is the band).

An aspect of human MHC is that it is not only polygenic, meaning that it consists of different genes, but also polymorphic, which indicates that a MHC gene can have several variants or alleles in the population [47].

Human MHC has 3 classes: MHC I class, MHC II class, MHC III class.

- MHC I class and MHC II class genes are known to be encode for molecules involved in antigen presentation to immune cells: while MHC class molecules I are found on every nucleated cell and presents antigens from inside the cell to CD8+ T cells, MHC class II molecules are found on APCs and they present antigens from outside the cell to CD4+ T cells [47].

- MHC III class region contains genes that encode for proteins that are active during inflammation (for example TNFα, complement proteins, etc.) [48].

In humans, the genes from MHC I class and II class are referred to as HLA or Human Leukocyte Antigen as they were discovered through studies of antigenic differences among white blood cells [47].

Numerous genes variants from all the three classes, have been associated to an incremented predisposition to sarcoidosis [49].

Regarding HLA class I, some studies show that HLA-B7 and HLA-B8 genes are associated with sarcoidosis. HLA-B7 is particularly frequent in African Americans, whereas it is less common in the Japanese population. Moreover, there is a common haplotype (a set of alleles inherited together) of HLA-B8 and HLA-DR3 (the latter is an HLA class II gene) in the white population. This indicates that there is a linkage disequilibrium, a non-random association of alleles at different loci, between the two genes [50].

As for HLA class II, there is more evidence in literature. Alleles of HLA-DRB1 contribute to determining both an increased and decreased risk of sarcoidosis [14,49]. Specifically, HLA-DRB1*11:01 increases the risk in the black and white populations [14]; HLA-DRB1*12:01 and HLA-DRB1*15:01 also increase the risk [14,49]. On the contrary, other alleles such as HLA-DRB1*04:01 and HLA-DRB1*04:07 reduce the risk [49].

Another example is a study investigating HLA-DQA1 and its newly identified variants in Korean patients: variant HLA-DQA1*05:08 is associated with increased risk whereas variants HLA-DQA1*01:02 decreases the susceptibility [51].

Furthermore, there are variants that increase the risk of manifesting a sarcoidosis syndrome: a Swedish study shows patients with HLA-DRB1*03 have higher risk of presenting Löfgren syndrome [52].

In addition to HLA genes, there are other investigated genes that may lead to sarcoidosis susceptibility.

BTNL2 (Butyrophilin-like 2) is one most the studied genes outside HLA region. BTLN2 is adjacent to the HLA region and has a role in suppressing T cells activation [53]. In a study on a Japanese cohort, SNP (single nucleotide polymorphism) rs2076530_A in BTLN2 gene was found to be associated with sarcoidosis susceptibility. Furthermore, rs2076530_A is in linkage disequilibrium with HLA-DRB1 alleles, in particular with variant HLA-DRB1*08:03 but the authors also found that having both variants does not increase the overall risk of sarcoidosis [54].

There are also data about non-HLA genes and loci such as ANXA11, IL23R and CCDC88B[49]. These genes will be further explored in this study.

An interesting gene that failed to have consistent results in different populations is the ACE gene. The D allele of the gene is characterized by a missing non encoding sequence, and it correlates with elevated risk of sarcoidosis in African Americans and Germans, but it also correlates with low risk in Whites and the Japanese [50].

An intriguing detail about the genetics of sarcoidosis is the influence of ancestry-specific gene variants. Several studies in the literature have reported data about gene variants that are present in certain analysed populations.

A study by Levin A.M. et al. reveals that variant HLA-DRB1*03:01 is associated with increased risk in subjects of European descent, whereas in subjects of African American descent it decreases the risk. They also found that if the European alleles of HLA-DRB1*03:01 are present in African Americans, it may also increase the risk in this population [55].

Another example of ancestry specific variants is ANXA11. SNPs rs61860052 and rs4377299 of ANXA11 are found only in African Americans and appear to contribute to sarcoidosis risk with the former being protective (OR 0.62) while the latter being a risk factor (OR 1.31) [56].

GENOME WIDE ASSOCIATION STUDIES

Many discoveries have been made regarding the genetics of various disease in recent years. It is known that for most diseases the mechanisms underlying their manifestations is multifactorial. In other words, there is a wide range of factors involved, some more significant than others, but for some diseases, genetics plays an important role.

In most cases, there is no exact gene that determines disease expression, but there are a multitude of genes that may confer a susceptibility to the disease. Sarcoidosis falls within this context.

With the knowledge from the Human Genome Project, that aimed to sequence the human genome, there has been a growing interest in searching for genetic variants that may be associated with diseases [57].

In the field of genetics there are different methods to find associations. To mention a few, WES or whole exome sequencing, WGS or whole genome sequencing and GWAS or genome wide association studies. This study focuses on GWAS.

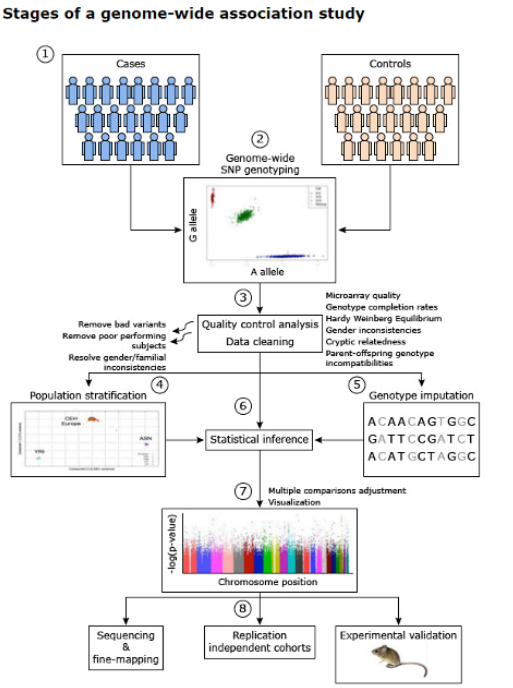
Genome wide association studies are an essential tool to explore the genetic architecture of sarcoidosis. Moreover, they provide insights about the genetic profiles that may influence disease onset, progression and severity.

GWAS is a study conducted with the aim of uncovering statistical associations between a disease and genetic variants. The genetic variants may be directly associated with the disease or they may be in linkage disequilibrium with a still unknown causative variant [58].

Most GWAS study a particular genetic variable called SNPs or single nucleotide polymorphisms. A gene is a sequence of base pairs that can vary between individuals. A SNP is a change in one of the base of the sequence that is shared by at least 1% of the population [57].

GWAS is usually a case-control study that examines a selected genetic variant that might be associated to a disease of interest and tests, for each individual, a large number of genes through microarray technology. The following steps, as shown in Figure 9, are quality control analysis (removal of bad variants, microarray quality, gender inconsistencies, etc.), population stratification, genotype imputation and statistical interference. Then, each genetic variant is analyzed to see if it's correlated to the disease by logistic regression (in case control studies) while considering other influencing factors. P-value is used to evaluate whether the result is meaningful or not. The results are displayed in a Manhattan plot, an XY graph where X axis shows chromosomes and Y axis shows P-value. Finally, further investigations such as replication studies, sequencing and fine-mapping, are performed to confirm the GWAS results [57].

Figure 9. Stages of a GWAS



It is frequent to combine the results of multiple GWAS in order to carry out a meta-analysis[58].

Meta-analyses are an effective resource as they help to find more accurate genetic associations by boosting statistical power, since the effect of genetic variants on complex diseases are often small [57].

However, GWAS are not without limitations. One key limitation is the concept of missing heritability.

Genetic heritability is the difference between estimated heritability of a particular trait or disease and the heritability of the genetic variants explained through GWAS. The hypothesized reasons for missing heritability are weak genetic effects of the SNPs, strong rare variants, environmental factors and gene interactions [58].

Other GWAS limits are potential biases from ethnicity mix, comorbidities of the examined subjects, genotyping errors due to differences in DNA samples of cases and controls, and linkage disequilibrium with a causative variant [57].

Although GWAS are not perfect studies, their contribution to clinical practice in undeniable, especially for complex diseases as sarcoidosis.

The identification of genetic variants that may confer protection or increase the risk of manifesting sarcoidosis, holds a considerable significance not only in terms of gaining deeper insight of the pathological mechanisms that lead to sarcoidosis, but also in informing prevention and therapeutic approach.

AIM OF THE STUDY

Sarcoidosis is multisystem inflammatory condition with an unknown etiology. A considerable variety of factors are suggested to participate in the disease onset, including genetic risk.

There are many genetic variants that have been implicated in susceptibility to sarcoidosis and are identified through genome wide association studies [49].

As more data about genetic variants, particularly SNPs, is being available, made possible by large population-based GWAS datasets, there has been a growing interest not only in identifying new risk loci that are potentially linked to sarcoidosis but also adding more evidence to support existing associations reported in biomedical literature.

Furthermore, sarcoidosis features differences in epidemiology patterns and genetic risk across populations [4,49].

This means that there is an ethnic variability beneath and analyzing diverse populations per genotype can offer further insight into the variability of SNP expression across ethnic groups in relation to a particular phenotype.

In this study, we conducted a comprehensive meta-analysis of the publicly available summary statistics of four genome-wide association studies for sarcoidosis. The four datasets contained data of individuals of different genetic backgrounds, specifically of European (EUR), African (AFR) and East Asian (EAS) ancestries. We included the results of EUR ancestry meta-analysis, AFR ancestry meta-analysis and a multi-ancestry meta-analysis.

This paper aimed to identify novel ancestry-specific risk loci, as well as confirming previously identified risk loci, by applying bioinformatic analyses and critically comparing our findings with the available biomedical literature.

MATERIALS AND METHODS

DATASETS AND PATIENT POPULATION

We accessed four international publicly available datasets and employed their latest available data. The website links to access the datasets are provided below.

- FinnGen freeze 12: https://www.finngen.fi/en/access_results, phenotype SARCOIDOSIS, released on November 4, 2024); the FinnGen study is a research project consisting in a large-scale genomics initiative that has analyzed over 500,000 Finnish biobank samples collected since 1969. It correlated genetic variation with health data to understand disease mechanisms and predispositions. The FinnGen study is a public-private collaboration between research organizations and biobanks within Finland and international industry partners [59];

- Pan-UK Biobank project: https://pan.ukbb.broadinstitute.org/, phenotype "D86 sarcoidosis" (last recorded update March 16, 2023). Pan-UK Biobank Project is a study across multiple ancestral groups involving a wide number of traits [60];

- Million Veteran Program (MVP): https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001672.v11.p1 (last recorded update August 22, 2023), Phecodes Phe_697.EUR.GIA (European) and Phe_687.AFR.GIA (African American). MVP is a project that investigates data of US Veterans collected through surveys, clinical information and biological samples with the aim of conducing genome sequencing analyses [61];

- Biobank Japan (BBJ): https://pheweb.jp/ , phenotype sarcoidosis. The study by Sakaue et al., published on October 2021 reports summary statistics of BBJ [62]. BBJ is project that selected subjects according to a series of diseases of interest such as lung cancer and interstitial lung diseases [63].

From each dataset we extracted ancestry-specific sample sizes for both cases and controls.

- FinnGen_SARCOIDOSIS, r.12, 5,411 cases and 492,311 controls.
- Million Veteran Program. Regarding European ancestry, 1,509 cases and 449,523 controls; as for African ancestry, 1,827 cases and 119,828 controls.
- UK Biobank (pan-UKBB). For European ancestry 639 cases and 419,892 controls. concerning African ancestry 53 cases and 6,583 controls.
- Biobank Japan for East Asian ancestry. 220 cases and 177,667 controls.

A critical point is that the datasets do not have a standardized format. FinnGenn freeze 12, MVP and UKBB datasets are updated to the GRCh38 assembly. GRCh38 stands for Genome Reference Consortium human assembly 38 [58] and it is the latest human genome sequence. The same cannot be applied for BBJ as it bases on the GRCh37 assembly.

Therefore, rsIDs (reference SNP cluster ID, a naming convention for SNPs) cross-referencing was used to replace chromosome coordinates in the BBJ dataset with coordinates from the UKBB reference dataset. R script is available in the Results chapter.

All summary statistics were individually replicated before we conducted the meta-analysis. Another critical point is that the BBJ, MVP and meta-analysis datasets were not provided with nearest gene specifications. To address the issue, we performed annotation through rsIDs cross referencing from the UKBB reference dataset.

META-ANALYSES

Regarding the conduction of the meta-analyses, R version 4.4.1 has been employed to execute the following procedures: the handling of datasets, the filtering of SNPs and the plotting of Manhattan and qq plots [64]. We used the following packages: parallel, vroom, data.table, dplyr, Biocmanager, Biostrings, SeqArray, ggmanH, SeqArray, farver, labeling, readr, knitr, kableExtra. R markdown html reports for EUR, AFR and multi-ancestry analyses are to be found in the Results chapter.

The fundamental concept behind meta-analysis is to combine the associations found from several independent studies. But not every study has the same format: they may vary in the number of the subjects recruited, study design, methodologies, etc. Therefore, it is necessary to apply appropriate weighting to account for the differences [65].

As we examined four independent datasets, we adopted METAL to perform a sample size-weighted meta-analysis of the summary statistics from EUR ancestry, AFR ancestry and multi-ancestry.

METAL approach consists in calculating a Z-score from P-value and the direction of the effect, in other words if a SNP increases or decreases disease risk. If Z score is below 0, the SNP is protective and statistically significant; on the other hand, if Z-score is positive, the SNP increases the risk of disease, and it is statistically significant; Z-score= 0 is statistically non-significant. The next step is to sum each study Z-score but according to weight which means more weight is given to large studies and the weight is proportional to the square root of the number of the participants. In our cases, the datasets of interest are unbalanced in the numbers of cases and controls, thus the application of the following formula:

$$N_{eff} = \frac{4}{\left(\frac{1}{N_{cases}} + \frac{1}{N_{ctrls}}\right)}$$

The formula is needed to balance cases and controls when different in numbers, thus calculating the effective sample size [65].

In all our analyses we considered the issue of genomic inflation. Therefore, we applied the correction for genomic inflation λ.

Furthermore, we filtered the datasets according to a common set of SNPs before conducting the meta-analyses.

We conducted three meta-analyses:

- AFR-specific meta-analysis included 1,880 cases and 126,411 controls; we analyzed 21,388,456 common SNPs.
- EUR-specific meta-analysis included 7,559 cases and 1,361,726 controls; we analyzed 10,776,775 common SNPs.
- Multi-ancestry meta-analysis included 9,659 cases and 1,665,804 controls; we analyzed 5,586,486 common SNPs. There is a similar multi-ancestry analysis that was issued with the FinnGen r.12 freeze, but it did not include data from East Asian ancestry.

RISK PLOTTING AND ANNOTATION

We generated Manhattan plots and regional plots for EUR ancestry meta-analysis, AFR ancestry meta-analysis, multi-ancestry meta-analysis, and performed annotation through the following techniques:

- Functional Mapping and Annotation of GWAS (FUMA GWAS) is an online tool that integrates data and supports visualization and functional annotation of results from GWAS [66];
- ANNOVAR is a software that allows annotation of genetic variants and their functional role [67];

We also used Combined Annotation-Dependent Depletion (CADD) for functional annotation. CADD is a tool that combines a wide range of genetic annotations into a numerical score. This allows a valuation of a genetic variant's deleteriousness and potential impact [68].

In the EUR ancestry meta-analysis, we manually selected the SNPs utilized for mapping on the basis of their CADD score. The reference value was CADD> 12.37 which means the SNP is pathologically relevant to sarcoidosis. This applies for each locus examined.

In order to calculate r^2, a coefficient for LD, and minor allele frequency estimations for risk loci plotting and annotation, Plink 1.9 (tool for data handling, summary statistics, etc.) [69,70] and GRCh38-based 1000 Genomes (1000G) Project Phase 3 reference panel (publicly available WGS data) [71,72] were used by FUMA.

It is important to emphasize that in all the analyses we excluded data from MHC region but included non-MHC loci.

RESULTS

In our study we analyzed 9,659 cases of sarcoidosis:

- 7,559 cases of European ancestry
- 1,880 cases of African ancestry
- 220 cases of East Asian ancestry

And 1,665,804 controls:

- 1,361,726 controls of European ancestry
- 126,411 controls of African ancestry
- 177,667 controls of East Asian ancestry

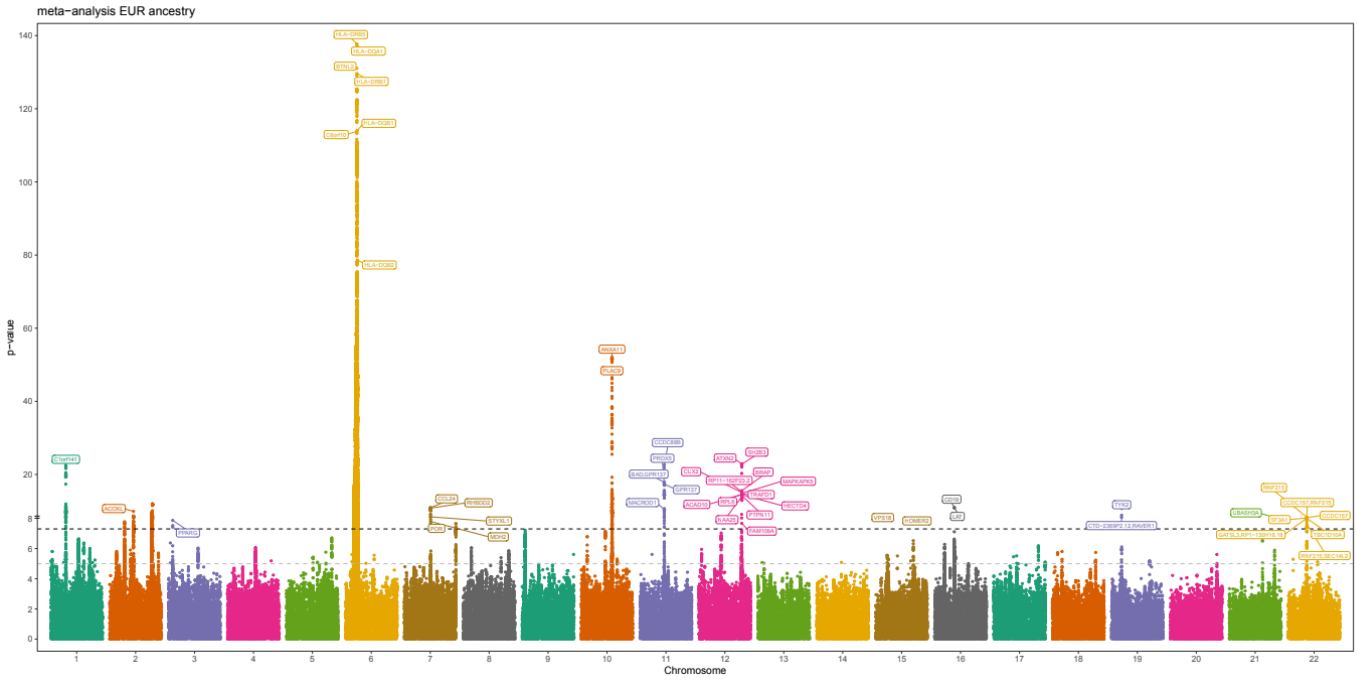
Subsequently, we proceeded to identify risk loci for each population ancestry, with visualization through Manhattan plots and regional plots and performed an assessment of genes and SNPs significantly associated with sarcoidosis.

RISK LOCI IDENTIFICATION

EUR ANCESTRY

As for the European ancestry meta-analysis, we identified nineteen risk loci (Table 3). We performed a Manhattan plot (Figure 10) and regional plots for each risk loci (Figure 11).

Figure 10. EUR ancestry Manhattan plot



Of the nineteen loci detected, around ten risk loci had not been associated to the phenotype before, with a p-value < 5×10^{-8} significance level, in any individual GWAS. This means that these risk loci represent novel genetic associations with sarcoidosis.

Here a few reported loci (see Table 3 for complete findings).

Chromosome 2

- Lead SNP 2:111610816:G:T, rs10183338, nearest gene ACOXL (p-value 1.16E-10, r^2=1)
- Lead SNP 2:198808578:A:G, rs1401090, nearest gene PLCL1 (p-value 1.98E-10, r^2=1)

Chromosome 6

- Lead SNP 6:29607101:C:T, rs3131856, nearest gene SUMO2P1 (p-value 3.20E-20, r^2=1)
- Lead SNP 6:33196397:A:C, rs34859217, nearest gene HTATSF1P (p-value 1.71E-36, r^2=1)

Chromosome 15

- Lead SNP 15:41202608:C:T, rs181956065, nearest gene RP11-540O11.1 (p-value 9.59E-09, r^2=1)
- Lead SNP 15:83515292:A:C, rs11856316, nearest gene HOMER2 (p-value 1.32E-08, r^2=1)

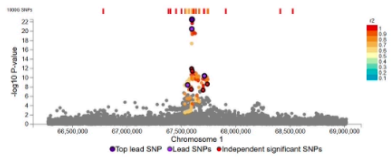
Chromosome 21

- Lead SNP 21:43855067:A:C, rs1893592, nearest gene UBASH3A (p-value 6.01E-09, r^2=1)

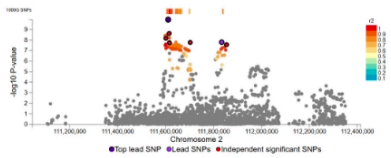
Chromosome 22

- Lead SNP 22:30776419:C:T, rs757870, nearest gene RNF215 (p-value 4.34E-09, r^2=1)

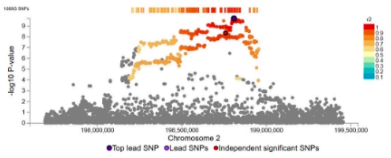
Figure 11. EUR ancestry regional plots



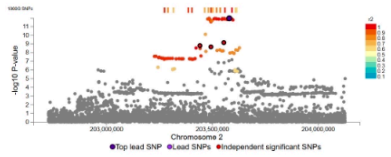
. 1:67597977:C:T



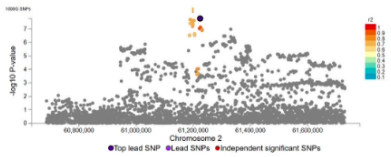
. 2:111610816:G:T



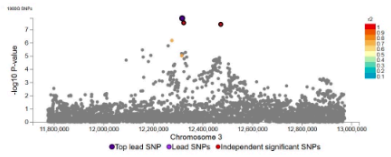
. 2:198808578:A:G



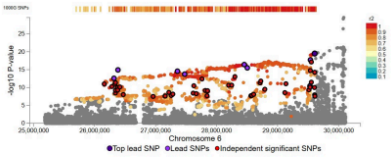
. 2:203579866:C:T



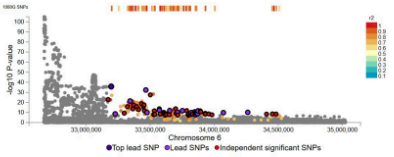
. 2:61224259:C:T



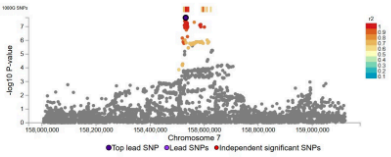
. 3:12316339:A:G



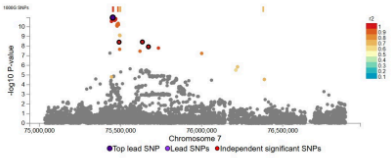
. 6:29607101:C:T



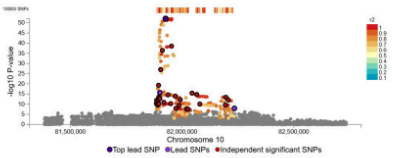
. 6:33196397:A:C



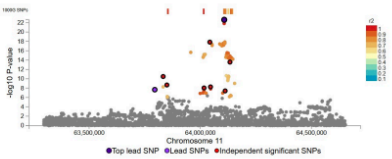
. 7:158531641:A:G



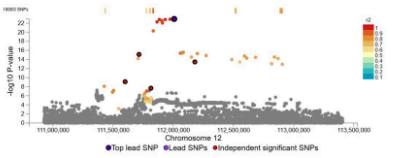
. 7:75453254:G:T



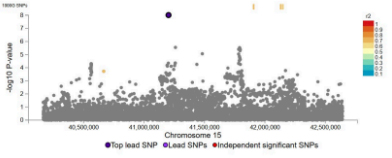
. 10:81926702:A:G



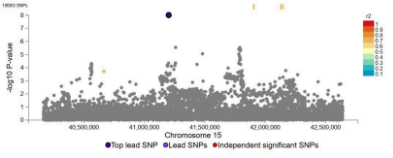
. 11:64107735:A:G



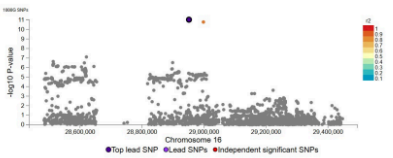
. 12:112007756:C:T



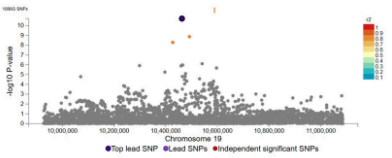
. 15:41202608:C:T



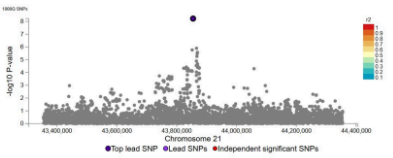
. 15:83515292:A:C



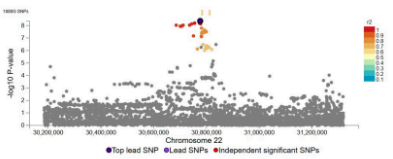
. 16:28951457:C:T



. 19:10463118:C:G



. 21:43855067:A:C



. 22:30776419:C:T

Table 3. Risk locus in EUR ancestry, mapped SNPs CADD > 12.37 and r^2 > 0.6 are reported under the lead variant (in bold)

risk locus	rsid	bp	r^2	GWAS P-value	Annotation	Nearest Gene	CADD
1:67597977:C:T	**rs2024825**	**67597977**	**1**	**2.96E-23**	**intronic**	**C1orf141**	**3.432**
	rs11209026	67705958	0.920359	1.268E-10	exonic	IL23R	26.5
	rs11208997	67560956	0.642851	0.00226	exonic	C1orf141	12.62
	rs2224501	67532526	0.619157	0.000000004463	intergenic	SLC35D1	13.75
	rs4655514	67550921	0.650973	0.001956	intergenic	C1orf141	15.59
	rs7519768	67555522	0.629107	5.933E-10	intergenic	C1orf141	12.99
	rs34388889	67563430	0.774939	NaN	intronic	C1orf141	12.53
	rs12140736	67586671	0.631274	1.427E-10	intronic	C1orf141	13.86
2:61224259:C:T	**rs10181042**	**61224259**	**1**	**1.67E-08**	**intronic**	**PUS10**	**0.43**
	rs6710043	61218280	0.787821	0.00009042	intronic	PUS10	13.93
2:111610816:G:T	**rs10183338**	**111610816**	**1**	**1.16E-10**	**intronic**	**ACOXL**	**5.723**
	rs1554005	111598958	1	3.15E-08	exonic	ACOXL	17.28
	rs6732565	111607832	0.89891	5.95E-09	intronic	ACOXL	15.15
	rs13401811	111616104	1	1.83E-08	intronic	ACOXL	21.5
	rs17525147	111619522	0.919722	4.95E-08	intronic	ACOXL	16.43
	rs9308690	111638898	0.828115	3.69E-08	intronic	ACOXL	12.69
	rs9308692	111641263	0.828115	3.57E-08	intronic	ACOXL	13.03
	rs28811027	111653937	0.834575	3.11E-08	intronic	ACOXL	12.39
	rs113122861	111657687	0.858706	NaN	intronic	ACOXL	14.45
	rs1513826	111660710	0.834575	4.00E-08	intronic	ACOXL	13.45
2:198808578:A:G	**rs1401090**	**198808578**	**1**	**1.98E-10**	**intronic**	**PLCL1**	**0.618**
	rs788018	198265526	0.739653	0.000002647	exonic	SF3B1	18.96
	rs788023	198283305	0.739653	0.000002588	exonic	SF3B1	13.29
	rs8539	198362018	0.740639	0.000001763	exonic	HSPD1	18.44
	rs1064213	198950240	0.773667	0.0000001944	exonic	PLCL1	26.7
	rs788007	198233676	0.710917	0.00000405	intergenic	NPM1P46	16.84
	rs67657812	198387401	0.698702	0.00000002455	intronic	HSPE1-MOB4:MOB4	12.46
	rs3838584	198416728	0.74175	NaN	UTR3	MOB4	15.71
	rs11336956	198540701	0.901653	NaN	UTR5	RFTN2	13.6
	rs700662	198668751	0.944576	0.00000001692	upstream	PLCL1	12.6
	rs696817	198722225	0.954087	0.000000008587	intronic	PLCL1	13.92
	rs570036157	198723904	0.954087	NaN	intronic	PLCL1	12.66
	rs4850812	198743655	0.953646	0.000000008505	intronic	PLCL1	15.42
	rs938929	198780860	0.981324	0.000000001489	intronic	PLCL1	12.86
	rs1464211	198798630	0.944517	0.00000001046	intronic	PLCL1	12.99

risk locus	rsid	bp	r^2	GWAS P-value	Annotation	Nearest Gene	CADD
	rs55696134	198821569	0.981393	4.065E-10	intronic	PLCL1	12.56
	rs61435657	198825506	0.627221	0.00006861	intronic	PLCL1	14.38
	rs5837576	198826363	0.94	NaN	intronic	PLCL1	15.33
	rs13382697	198920560	0.81266	0.0000003302	intronic	PLCL1	15.25
2:203579866:C:T	**rs191390916**	**203579866**	**1**	**1.16E-12**	**intronic**	**FAM117B**	**1.08**
	rs72925089	203359566	0.961832	5.25E-08	intronic	BMPR2	17.28
	rs140974562	203460202	0.713015	NaN	intergenic	AC009960.1	15.01
	rs933969	203499720	0.646164	NaN	upstream	FAM117B	13.85
	rs6748088	203556526	0.973392	1.39E-12	intronic	FAM117B	13.85
3:12316339:A:G	**rs2972166**	**12316339**	**1**	**1.36E-08**	**intergenic**	**PPARG**	**0.377**
6:29607101:C:T	**rs3131856**	**29607101**	**1**	**3.20E-20**	**intergenic**	**SUMO2P1**	**4481**
	rs13195402	26463575	0.877433	1.17E-13	exonic	BTN2A1	23.7
	rs13195509	26463660	0.755787	2.664E-13	exonic	BTN2A1	22.5
	rs139332558	26637724	0.889731	NaN	exonic	ZNF322	19.82
	rs7756481	27115069	0.879068	0.00000139	exonic	HIST1H2AH	16.39
	rs141138864	27223064	0.704895	NaN	exonic	PRSS16	15.96
	rs16897515	27278020	0.966338	1.82E-08	exonic	POM121L2	19.9
	rs200484	27775674	0.809201	7.57E-15	exonic	HIST1H2BL	16.23
	rs200981	27833174	0.782008	7.32E-15	exonic	HIST1H2AL	15.9
	rs200956	27839746	0.949039	3.42E-09	exonic	HIST1H3I	16.41
	rs200973	27858421	0.984932	1.88E-09	exonic	HIST1H3J	21.5
	rs34788973	27879200	0.947228	4.47E-16	exonic	OR2B2	23.2
	rs61742093	27879982	0.947228	9.87E-16	exonic	OR2B2	22.6
	rs1679709	28228342	0.844255	9.92E-11	exonic	NKAPL	22.3
	rs33932084	28268824	0.940768	8.94E-14	exonic	PGBD1	14.75
	rs2230683	28891176	0.92562	1.21E-17	exonic	TRIM27	18.69
	rs404240	29523957	0.897948	1.45E-15	exonic	UBD:GABBR1	14.02
	rs6902389	26322153	0.729201	1.05E-11	ncRNA_exonic	HIST1H3PS1	13.92
	rs6908156	26322861	0.729201	1.05E-11	upstream	HIST1H3PS1	14.27
	rs6902392	26322115	0.729201	1.05E-11	ncRNA_exonic	HIST1H3PS1	14.9
	rs9467740	26383250	0.872107	7.73E-12	upstream	BTN2A2	12.85
	rs3832422	26383523	0.872107	NaN	UTR5	BTN2A2	13.39
	rs12207181	26431285	0.847182	7.81E-10	ncRNA_exonic	BTN2A3P	12.45
	rs34104395	26478252	0.755787	2.82E-13	intergenic	BTN2A1	13.34
	rs13198716	26582035	0.890433	1.49E-13	intergenic	ABT1	12.54
	rs9393735	26582327	0.719829	4.74E-07	intergenic	ABT1	12.51
	rs13203358	26590578	0.674991	2.01E-07	intergenic	ABT1	13.55

risk locus	rsid	bp	r^2	GWAS P-value	Annotation	Nearest Gene	CADD
	rs6456742	26617075	0.785548	2.24E-08	intergenic	RP11-457M11.6	12.79
	rs72845515	26646871	0.785548	2.02E-08	intronic	ZNF322	12.63
	rs12198053	26667659	0.664209	5.50E-08	intergenic	ZNF322	13.55
	rs12192446	26995638	0.896444	3.23e-7	intergenic	LINC00240	13.3
	rs72838249	26998489	0.925126	9.65E-08	intergenic	VN1R12P	13.41
	rs72838262	27008517	0.925126	1.58E-07	intergenic	VN1R12P	13.7
	rs6934329	27158033	1	2.40E-08	intergenic	RP11-209A2.1	13.14
	rs12198077	27169307	0.868364	0.000001443	intergenic	RP11-209A2.1	12.54
	rs72843629	27175215	0.868364	9.54E-07	intergenic	RP11-209A2.1	12.52
	rs67457459	27198343	0.711219	1.45E-10	intergenic	PRSS16	13.45
	rs72843641	27203335	0.858902	8.25E-07	intergenic	PRSS16	15.71
	rs72839477	27327000	1	8.09E-15	ncRNA_exonic	ZNF204P	14.36
	rs34573979	27480526	0.986328	2.38E-13	intergenic	XXbac-BPGBPG34I8.1	15.35
	rs56405707	27640246	0.986328	1.08E-13	intergenic	RP1-15D7.1	13.34
	rs9357045	27688927	0.690251	3.36E-07	intergenic	RP1-97D16.1	19.07
	rs9348774	27688930	0.690251	4.66E-07	intergenic	RP1-97D16.1	17.07
	rs13193480	27702561	0.972983	6.80E-14	intergenic	RP1-97D16.1	15.29
	rs4713119	27712825	0.955161	8.27E-09	intergenic	RP1-97D16.1	12.66
	rs9468225	27745719	0.609622	0.000007862	intergenic	RSL24D1P1	17.65
	rs200483	27774824	0.809201	5.99e-15	upstream	HIST1H4PS1	13.69
	rs17751184	27775028	0.972983	4.02E-14	ncRNA_exonic	HIST1H4PS1	14.92
	rs375576927	27835322	1	NaN	UTR5	HIST1H1B	13.52
	rs200949	27835435	0.737187	2.81E-14	upstream	HIST1H1B	12.63
	rs45509595	27840926	0.972983	1.74E-15	UTR3	HIST1H4L	15.55
	rs184666393	27870324	1	NaN	intergenic	RNU7-26P	17.84
	rs71559054	27896799	0.947228	5.69E-16	intergenic	OR2W6P	14.09
	rs67040724	27905509	0.947228	2.84E-16	ncRNA_exonic	OR2W6P	13.54
	rs148418547	27914359	0.947228	NaN	intergenic	OR2W6P	13.54
	rs28360499	27945396	0.934796	5.44E-16	ncRNA_exonic	OR2W4P	13.56
	rs149947	27972433	0.692362	0.000009017	intergenic	IQCB2P	13.59
	rs149900	28014597	0.708989	0.00001664	ncRNA_exonic	OR2B7P	14.67
	rs35572414	28015111	0.750238	NaN	ncRNA_exonic	OR2B7P	12.49
	rs203888	28021589	0.683285	0.000008666	ncRNA_exonic	OR2B8P	15.75
	rs35155115	28021853	0.774358	0.000006234	ncRNA_exonic	OR2B8P	14.64
	rs188105	28071393	0.766448	0.000002204	intergenic	ZSCAN12P1	15.23
	rs75874576	28087832	0.902652	2.66E-12	intergenic	ZSCAN16-AS1	12.49
	rs9468300	28126840	0.66646	2.69E-10	UTR3	ZKSCAN8	13.42

risk locus	rsid	bp	r^2	GWAS P-value	Annotation	Nearest Gene	CADD
	rs1736904	28219270	0.837331	2.04E-08	intronic	ZKSCAN4	12.41
	rs146407472	28243175	1	NaN	ncRNA_intronic	RP5-874C20.3	14.64
	rs13211507	28257377	0.955051	1.58E-16	intronic	PGBD1	13.79
	rs36005309	28277346	0.955051	NaN	intergenic	PGBD1	12.91
	rs56075693	28290328	0.969715	1.24E-16	intergenic	ZSCAN31	12.71
	rs10591593	28315542	0.609775	NaN	intronic	ZSCAN31	12.96
	rs35744819	28318331	1	1.01E-16	intronic	ZSCAN31:ZKSCAN3	13.57
	rs78371185	28321502	1	NaN	intronic	ZSCAN31:ZKSCAN3	14.17
	rs71559081	28372069	0.660174	0.00004601	intergenic	ZSCAN12	12.97
	rs13209596	28396190	0.660174	0.00004634	intergenic	ZSCAN23	13.61
	rs7765989	28400295	0.703367	0.000005804	intronic	ZSCAN23	12.95
	rs11967609	28656109	0.780705	3.49E-07	intergenic	LINC00533	13.29
	rs12193659	28677144	0.899801	4.65E-12	intergenic	RPSAP2	14.4
	rs1233583	28715566	0.926851	4.83E-17	intergenic	RPSAP2	14.21
	rs9257189	28757555	0.926851	2.93E-17	intergenic	NOL5BP	14.92
	rs209181	28792477	1	3.66E-10	intergenic	XXbac-BPG308K3.5	12.38
	rs3135300	28824397	0.911716	1.64e-17	intergenic	XXbac-BPG308K3.6	13.51
	rs3135302	28829637	0.911716	1.81E-17	ncRNA_exonic	XXbac-BPG308K3.6:RPL13P	12.44
	rs3132392	28838629	0.911716	2.25e-17	intergenic	XXbac-BPG308K3.6	13.32
	rs3130746	29153155	0.92562	5.32E-17	intergenic	OR2J4P	12.86
	rs3129093	29170261	1	4.64E-09	intergenic	OR2H4P	13.4
	rs426595	29512482	0.865438	6.44E-08	intergenic	GPR53P	14.95
	rs111508444	29603512	1	1.03E-10	downstream	SUMO2P1	19
6:33196397:A:C	**rs34859217**	**33196397**	**1**	**1.71E-36**	**intergenic**	**HTATSF1P**	**3.039**
	rs9394145	33399778	0.881875	4.646E-13	exonic	SYNGAP1	17.25
	rs4713668	33690796	0.945819	1.426E-11	exonic	IP6K3	22.1
	rs114699053	33172306	1	1.38E-23	upstream	SLC39A7	13.33
	rs211448	33343859	1	4.33E-22	intergenic	LYPLA2P1	12.78
	rs116323349	33401291	0.660355	1.39e-16	intronic	SYNGAP1	16.5
	rs9366821	33444074	0.776956	8.45E-20	intergenic	ZBTB9	14.14
	rs62405954	33524820	0.874942	5.64E-29	intergenic	BAK1	12.68
	rs210122	33574635	0.980196	3.85e-12	intergenic	ITPR3	12.96
	rs9368768	33588610	0.752623	0.0000002741	UTR5	ITPR3	15.94
	rs71565395	33593577	0.904311	1.07e-9	intronic	ITPR3	13.36
	rs749847	33665001	0.932443	2.332E-11	UTR3	UQCC2	13.46
	rs9469566	33667605	0.949405	1.76e-11	intronic	UQCC2	14.21
	rs4713659	33667839	0.949405	1.948E-11	intronic	UQCC2	14.48

risk locus	rsid	bp	r^2	GWAS P-value	Annotation	Nearest Gene	CADD
	rs6457739	33673831	0.949405	1.937E-11	intronic	UQCC2	20.2
	rs58420296	33677214	0.932443	1.103E-11	intronic	UQCC2	13.5
	rs2966	33689520	0.925707	1.972E-11	UTR3	IP6K3	14.81
	rs2967	33689534	0.662091	0.004241	UTR3	IP6K3	12.69
	rs4713674	33708169	0.844356	3.958E-12	intronic	IP6K3	16.92
	rs73747318	33782048	0.796913	0.0000002631	intergenic	MLN	12.44
	rs17600945	33802263	0.88096	0.000005024	intergenic	MLN	20.9
	rs7746977	33842481	0.67781	2.555E-10	ncRNA_intronic	LINC01016	14.81
	rs11758326	34482498	0.809188	0.00000002457	intronic	PACSIN1	18.95
7:75453254:G:T	**rs7811626**	**75453254**	**1**	**1.13E-11**	**upstream**	**CCL24**	**1.552**
7:158531641:A:G	**rs7803766**	**158531641**	**1**	**2.25E-08**	**intronic**	**ESYT2**	**1.718**
	rs2305473	158536267	0.992399	0.00000005877	exonic	ESYT2	22.2
	rs2305475	158536345	0.992399	0.0000000623	exonic	ESYT2	22.7
	rs1061735	158526455	0.947305	0.000001459	UTR3	ESYT2	16.15
	rs842450	158567423	0.687358	0.000001523	intronic	ESYT2	14.26
	rs842452	158575440	0.687358	0.000001293	intronic	ESYT2	12.61
10:81926702:A:G	**rs1049550**	**81926702**	**1**	**9.08E-53**	**exonic**	**ANXA11**	**28.6**
	rs12411782	82012967	0.70449	0.000003082	exonic	AL359195.1	18.65
	rs2573360	81891731	1	4.134E-10	intronic	PLAC9	14.58
	rs2819885	81892002	0.65768	1.154E-11	intronic	PLAC9	12.69
	rs2819875	81903517	0.782327	4.26E-39	intronic	PLAC9	12.57
	rs3851053	81918582	0.635924	4.48E-28	intronic	ANXA11	14.57
	rs141872729	81927935	0.979479	NaN	intronic	ANXA11	15.34
	rs11202248	81991475	0.930316	3.77E-11	intergenic	RP11-40F6.1	13.82
	rs60492867	82022560	0.88442	NaN	intergenic	MAT1A	14.14
	rs10788579	82108610	0.962307	2.57E-14	intronic	DYDC1:DYDC2	13.51
	rs35050604	82196677	0.956171	4.65E-14	ncRNA_exonic	RP11-137H2.6	12.39
	rs10887881	82204620	0.987203	3.29E-10	intergenic	RP11-137H2.6	15.41
	rs7095954	82209232	0.84176	5.37E-08	intergenic	TSPAN14	12.88
	rs116857892	82214065	0.84176	4.91E-08	UTR5	TSPAN14	13.16
	rs12220642	82218964	0.707937	0.000001754	intronic	TSPAN14	12.7
11:64107735:A:G	**rs663743**	**64107735**	**1**	**2.38E-23**	**UTR5**	**CCDC88B**	**5934**
	rs647152	64109118	0.807175	1.64E-18	exonic	CCDC88B	13.15
	rs600377	64124940	0.682994	5.377E-11	exonic	CCDC88B	14.93
	rs521950	64127744	0.682994	3.66E-11	exonic	RPS6KA4	14.41
	rs11542299	64138805	0.876014	4.45E-15	exonic	RPS6KA4	16.99
	rs60031276	64012910	0.747794	2.52E-15	UTR5	PPP1R14B	18.33

risk locus	rsid	bp	r^2	GWAS P-value	Annotation	Nearest Gene	CADD
	rs35989122	64011614	0.992329	NaN	downstream	FKBP2	14.87
	rs11231727	64011854	0.90993	0.0000001063	downstream	PPP1R14B	12.39
	rs2510066	64052447	0.80836	4.22E-18	UTR5	GPR137	13.7
	rs542907	64124980	0.682994	1.097E-10	UTR3	CCDC88B	16.01
12:112007756:C:T	**rs653178**	**112007756**	**1**	**1.20E-23**	**intronic**	**ATXN2**	**0.312**
	rs7968960	111426615	0.79915	8.10E-09	intergenic	RP1-46F2.2	15.51
	rs10849925	111495518	0.817697	0.0000001583	intronic	CUX2	16.4
	rs4378452	111504033	0.825654	0.00000006634	intronic	CUX2	13.24
	rs7399113	111792848	0.62465	0.00001513	intergenic	CUX2	17.49
15:41202608:C:T	**rs181956065**	**41202608**	**1**	**9.59E-09**	**intergenic**	**RP11-540O11.1**	**15.11**
15:83515292:A:C	**rs11856316**	**83515292**	**1**	**1.32E-08**	**UTR3**	**HOMER2**	**2.074**
16:28951457:C:T	**rs11645302**	**28951457**	**1**	**9.45E-12**	**downstream**	**CD19**	**1.707**
19:10463118:C:G	**rs34536443**	**10463118**	**1**	**1.95E-11**	**exonic**	**TYK2**	**25.5**
	rs74956615	10427721	0.828325	5.24E-09	UTR3	CTD-2369P2.12:RAVER1	20.8
21:43855067:A:C	**rs1893592**	**43855067**	**1**	**6.01E-09**	**intronic**	**UBASH3A**	**11.16**
22:30776419:C:T	**rs757870**	**30776419**	**1**	**4.34E-09**	**intronic**	**RNF215**	**0.304**
	rs757660	30793137	0.800351	2.38E-08	exonic	RNF215:SEC14L2	19.14
	rs1860217	30685560	0.975458	9.45E-09	UTR5	GATSL3:RP1-130H16.18	15.03
	rs2240422	30700924	0.980324	1.15E-08	intronic	TBC1D10A	13.35

AFR ANCESTRY

Fewer loci were detected in the AFR ancestry meta-analysis compared to EUR ancestry meta-analysis. The two risk loci identified are shown in the Manhattan plot (Figure 12) and regional plot (Figure 13)

These are the two risk loci, located respectively in chromosomes 6 and 10:

- Lead SNP 6:26261874:A:G (MHC genes)
- Lead SNP 10:81939813:C:T (PLAC9 and ANXA11)

Figure 12. AFR ancestry Manhattan plot

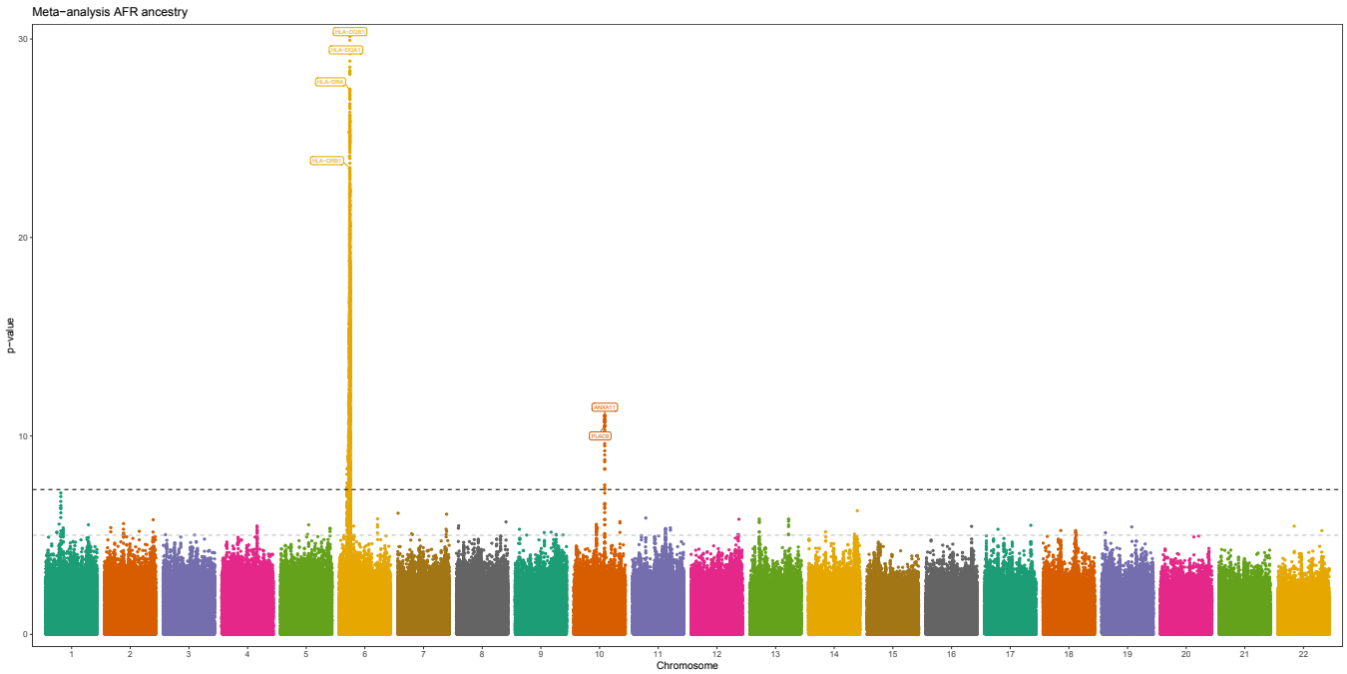
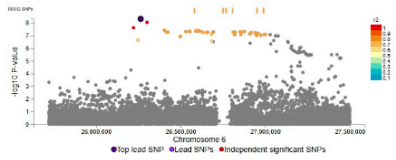
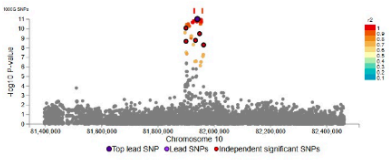


Figure 13. AFR ancestry regional plots



. 6:26261874:A:G



. 10:81939813:C:T

MULTI-ANCESTRY

The multi-ancestry meta-analysis included data of affected patients of EUR, AFR and EAS ancestry. We identified 18 risk loci, shown in the Manhattan plot (Figure 14) and regional plot (Figure 15).

Chromosome 1

- Lead SNP 1:172776938:A:G
- Lead SNP 1:67626416:C:G

Chromosome 2

- Lead SNP 2:111610816:G:T
- Lead SNP 2:198911426:A:G
- Lead SNP 2:203485086:G:T

Chromosome 3

- Lead SNP 3:112062920:C:T
- Lead SNP 3:12471871:C:T

Chromosome 6

- Lead SNP 6:29535217:A:G
- Lead SNP 6:33467086:A:T
- Lead SNP 6:34504042:G:T

Chromosome 7

- Lead SNP 7:75445194:C:T

Chromosome 10

- Lead SNP 10:64454497:C:T
- Lead SNP 10:81953371:A:G

Chromosome 11

- Lead SNP 11:64107735:A:G

Chromosome 12

- Lead SNP 12:112007756:C:T
- Lead SNP 12:58213272:A:G

Chromosome 15

- Lead SNP 15:8351592:A:C

Chromosome 16

- Lead SNP 16:28998111:A:T

Figure 14. Multi-ancestry Manhattan plot

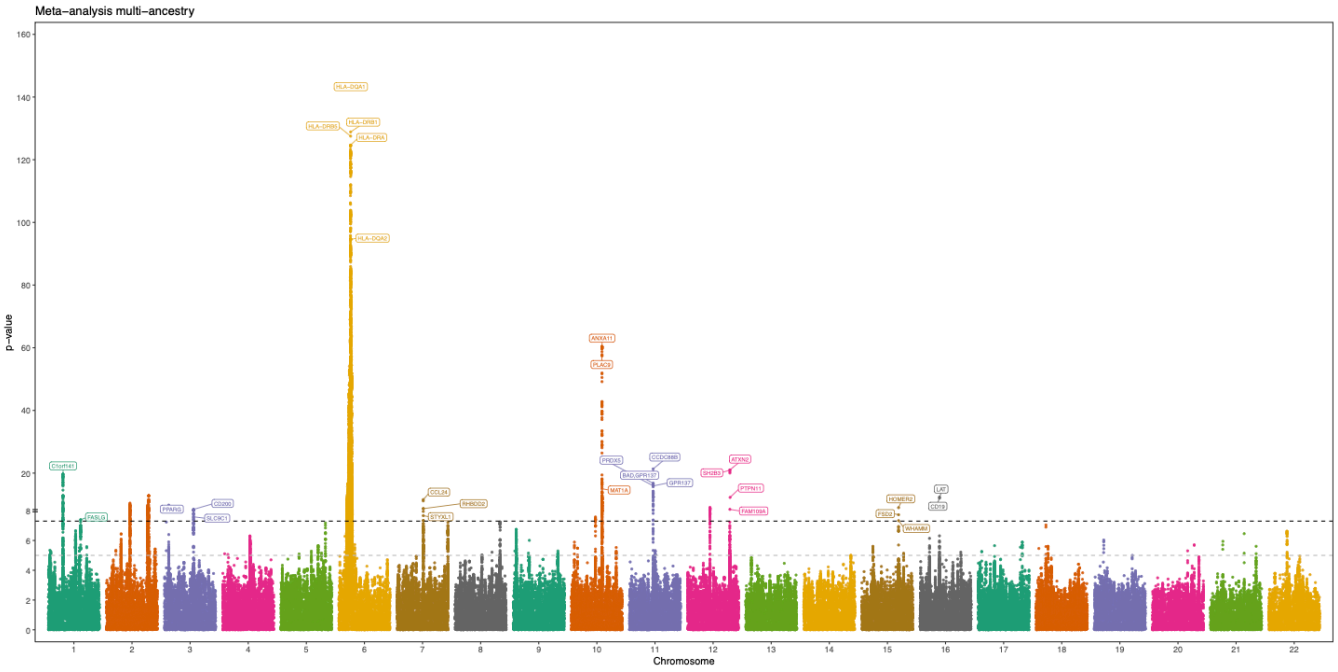
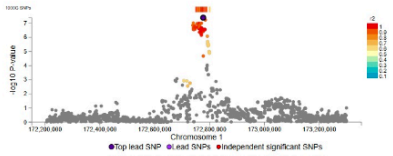
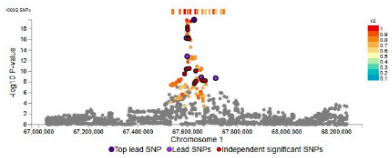


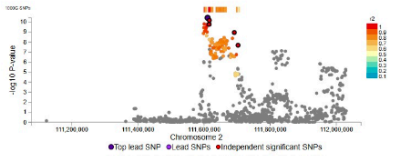
Figure 15. Multi-ancestry regional plots (continues page 55)



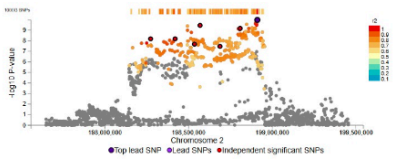
. 1:172776938:A:G



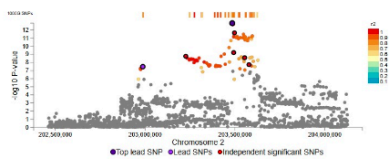
. 1:67626416:C:G



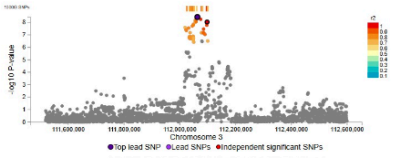
. 2:111610816:G:T



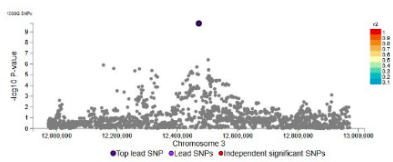
. 2:198911426:A:G



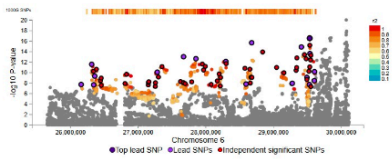
. 2:203485086:G:T



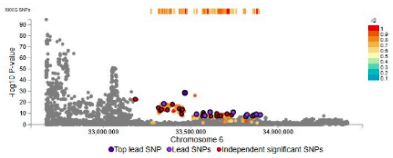
. 3:112062920:C:T



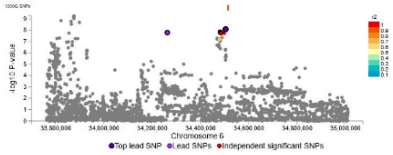
. 3:12471871:C:T



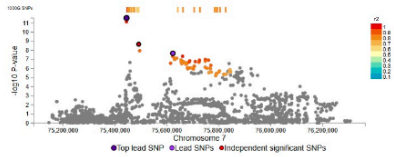
. 6:29535217:A:G



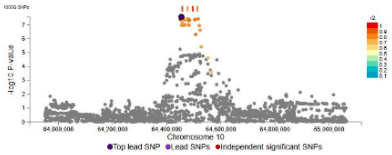
. 6:33467086:A:T



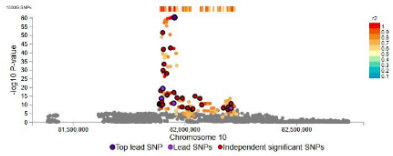
. 6:34504042:G:T



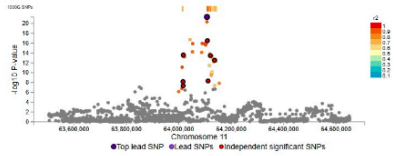
. 7:75445194:C:T



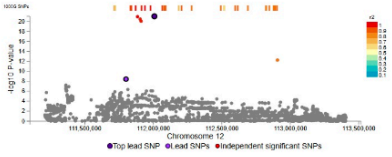
. 10:64454497:C:T



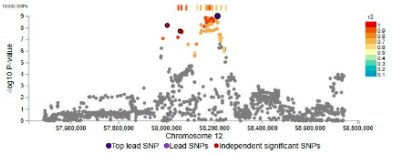
. 10:81953371:A:G



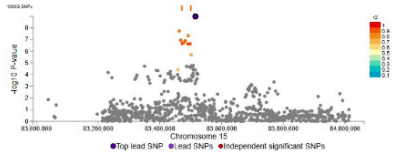
. 11:64107735:A:G



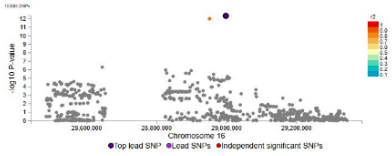
. 12:112007756:C:T



. 12:58213272:A:G



. 15:83515292:A:C



. 16:28998111:A:T

Compared to EUR ancestry meta-analysis, additional loci in chromosomes 1, 2, 3,10, 12 and 16 were detected.

Moreover, we confronted multi-ancestry results with the meta-analysis of FinnGen + UKBB + MVP datasets released by FinnGen freeze 12 and found that our results were consistent with the latter.

ASSESSMENT OF SIGNIFICANCE GENES AND SNPs

In the study we conducted a critical assessment of the non-MHC genes that were significantly associated with sarcoidosis in the EUR and multi-ancestry meta-analysis. We searched the literature for GWAS and other studies that have previously associated the non-MHC genes with sarcoidosis or that have described a function that might be related to

sarcoidosis with the aim of providing deeper insight into the putative functional roles of the genes.

We found in the literature GWAS and other studies that have identified SNPs of genes CCDC88B, ANXA11, IL23R, FAM117B, CCL24, ATXN2/SH2B3 and associated them to sarcoidosis.

Evidence from functional or experimental studies have been reported for genes PLCL1, PPARG and CD19.

For genes PUS10 and ESYT2 we found indirect evidence of their involvement in biological process containing elements implicated in sarcoidosis.

As for genes RNF215, RPS6KA4, LAT, UBASH3A, HOMER2, SUMO2P we found suggestive evidence of participation in regulated or dysregulated immune processes.

We have not found direct or indirect evidence in the existing literature regarding genes ACOXL, MAT1A and MACROD1 and their possible association to sarcoidosis or immune (dys)regulation.

Finally, to our knowledge there is no available data of the function of the following genes reported in the literature: RP11-213C14.10 and RP-11264B17.3.

Moreover, we manually selected and reported SNPs with CADD > 12.37 for each locus in EUR ancestry meta-analysis (see Table 3). In the AFR ancestry meta-analysis, only two SNPs displayed CADD score > 12.37.

DISCUSSION

In this paper, we present the most extensive ancestry-specific GWAS meta-analysis for sarcoidosis to date. Specifically, we focused on European and African ancestries. To obtain this, we aggregated publicly available summary statistics from some of the world's largest and well-established genetic biobanks, FinnGen freeze 12, UKBB, MVP and BBJ datasets, and performed a meta-analysis for each ancestry.

Moreover, we performed a multi-ancestry meta-analysis that included data of populations of European, African and East Asian ancestries together and it is the first study to do so, according to our current knowledge.

A similar meta-analysis have been published recently by the FinnGen biobank project that also accessed and included patients from the MVP and UKBB datasets [73]. What our study adds is a critical review of each identified risk locus, including both previously identified risk loci and novel risk loci we discovered in our meta-analyses. The purpose is to determine the probability of functional impact of the risk loci by comparing with existing data in the biomedical literature.

According to our research in literature, there are only a few GWAS studies focusing on sarcoidosis. We compared our findings with two GWAS [74,75]: GWAS for sarcoidosis by Fischer et al. examined 1,726 cases and 5,482 controls after quality filtering and identified 4 risk loci, of which two are also confirmed by our study. The nearest genes of the two risk loci are ATXN2/SH2B3 and FAM117B [75]; Meguro et al. GWAS examined 1,631 cases and 1,928 controls and found 3 risk loci, of which two also emerged in our study: CCL24 and C1orf141-IL23R [74].

Compared to the number of subjects included in the mentioned GWAS studies, our GWAS meta-analysis has a larger sample size which enabled to detect a greater number of risk loci. For instance, the EUR ancestry meta-analysis alone identified nineteen risk loci.

Regarding functional annotation of the SNPs and following analyses, we observed that the EUR ancestry meta-analysis generated the largest number of statistically significant risk loci compared to the other two meta-analyses. For each locus, SNPs were selected based on their CADD score > 12.37 (Table 3).

For the AFR ancestry meta-analysis, our results mostly overlap with the single summary statistics from the MVP study. The meta-analysis only produced two statistically significant risk loci on Chromosomes 6 (MHC genes) and 10 (PLAC9 and ANXA11). Once again SNPs were selected based on CADD > 12.37. Interestingly, a potential trend was observed for at least a locus in the IL23R region on chromosome 1. The paucity of results is probably due to the limited number of cases and controls of AFR ancestry available in the consulted datasets. Nevertheless, our AFR ancestry meta-analysis for sarcoidosis is the largest conducted in individuals of African ancestry to date, based on publicly available GWAS statistics. Also, data about individuals of East Asian ancestry is limited, compared to Europeans. This indicates that more attention should be devoted on GWAS studies with larger number of cases and controls of other ancestries, including AFR and EAS, in comparison to EUR.

As for the multi-ancestry meta-analysis (EUR, AFR and EAS) we noticed that some of the loci showing a statistically significant association with sarcoidosis in this meta-analysis were the same found in the EUR meta-analysis, in particular, lead SNP 1:67626416:C:G on chromosome 1; lead SNP 3:12471871:C:T on chromosome 3; lead SNP 6:34504042:G:T on chromosome 6; lead SNP 10:64454497:C.T on chromosome 10; lead SNP 12:112007756:C:T on chromosome 12 (Figure 15, regional plots).

However, there are risk loci significantly associated with sarcoidosis in the EUR ancestry meta-analysis, but lost statistical significance in the multi-ancestry meta-analysis, in particular lead SNP 7:75453254:G:T on chromosome 7; lead SNP 15:41202608:C:T on chromosome 15; lead SNP 19:10463118:C:G on chromosome 19; lead SNP 21:43855067:A:C on chromosome 21; lead SNP 22:30776419:C:T on chromosome 22 (Figures 11 and 15, regional plots).

Subsequently, we identified the nearest non-MHC genes in the risk loci. For each gene, a search in the literature has been conducted to verify whether the gene had already been linked to sarcoidosis or it was a novel finding. It was observed that at least 9 have been previously associated to sarcoidosis.

For the genes listed below, we compared available GWAS or other genetic studies in literature:

- CCDC88B encodes for protein GIPIE and has a poorly understood role as a study by Matsushita et al. shows that it suppresses apoptosis induced by endoplasmic reticulum stress in endothelial cells. GIPIE binds GRP78 and IRE1 which attenuates IRE1/ASK1/JNK signaling pathway and expression of CHOP (proteins involved in the apoptosis cascade) [76]. However, these findings were not replicated in the study by Kennedy et al. which also shows that CCDC88B contributes to the maturation and activation of T cells (it is expressed in CD3+ T cells in lymph nodes, spleen and thymus) as its absence impairs T cell function and leads to reduced cytokine production [77]. We also found evidence of CCDC88B being expressed in lung cells [78]. Fischer et al. GWAS identified SNPs, of which rs6633743 (UTR5) was also identified in our study and has a high regulatory potential. Other SNPs reported are rs647152 that may have an influence in the function of the protein, and rs671976 that likely is causative variant or in LD with causative variant [75].

- ANXA11 (Annexin A11) is a regulator of apoptosis in immune cells [79] and downregulated when immune cells are activated [80].

  ANXA11 is one of the most extensively studies genes in sarcoidosis and several GWAS have identified ANXA11 as a risk locus. In our study we identified rs1049550 (exonic), whose C allele had been linked to sarcoidosis risk [81].

  Another associated SNP is rs2784773 [80].

- IL23R is a further known gene associated with sarcoidosis. IL23R encodes for IL-23 receptor [74] which consists of IL12Rβ1 and IL23R and is found CD4+ memory T cells, in particular it stimulates INFγ production by CD45RO memory T cells [82]. IL23R is part of the IL-23/IL-17 signaling pathway, contributes to the differentiation of Th17/Th IL-17 cells and consequent production of cytokines such as IL-17 and TNFα is attributed to IL-23 binding to its receptor on CD4+ naïve T cells [83]. There is evidence of Th17/Th IL-17 cells being increased in peripheral blood and in the BAL fluid and are found within the granulomas of patients with sarcoidosis [84], suggesting a possible correlation between IL23R and Th17/ Th IL-17 cells in the pathogenesis of sarcoidosis.

  Further evidence shows that IL-23 is upregulated in cutaneous sarcoidosis [85] and a vast majority of sarcoidosis-related uveitis are genetically susceptible to IL23R [86]. Also, a murine experiment demonstrated its involvement in granuloma formation [87].

  In our study, we detected for IL23R SNPs rs11209026 (exonic) and rs117282985 that had been described by Meguro et al.: rs11209026 frequency is decreased in patients with sarcoidosis and in patients with both sarcoidosis and uveitis and rs117282985, upstream IL23R, is significantly associated with sarcoidosis in the Japanese population [74]. Another study confirms the negative association with sarcoidosis-related uveitis of rs11209206 [88].

Other reported SNPs associated with sarcoidosis are rs12069782 in the IL23R promoter [75] and rs117633859 (Japanese population), rs6664119 [74]; rs11465804 (negatively associated with sarcoidosis-related uveitis) [88].

- C1orf141, IL23R: insufficient evidence of function related to sarcoidosis for C1orf141 is available to the best of our knowledge. We identified SNP rs117282985, confirmed by Meguro et al. [74].

- FAM117B has a role in oxidative stress as it activates KEAP1/NRF2 signaling pathway [89,90]. A study by Fois et al. shows that it is associated with sarcoidosis [91].
  We identified SNP rs6748088 (intronic) which was previously associated with susceptibility for sarcoidosis in Europeans [75].

- CCL24 is involved in Th2 inflammatory response as a reduction in its expression reduces Th2 inflammatory response in favor of Th1 inflammatory response [74]. CCL24 plays a role in pulmonary fibrosis in systemic sclerosis: if CCL24 is blocked, it leads to a reduction in lung inflammation and fibrosis [92].
  Studies have demonstrated increased CCL24 levels in the BAL fluid of stage 3 sarcoidosis patients in comparison to patients with Löfgren syndrome [93] and in the aqueous humor of patients with uveitis [94].
  We identified SNP rs7811626 (upstream) but no data about it was available in the literature. We report rs472849 (intronic), associated with a reduced expression of CCL24, in particular to Löfgren syndrome [74].

- ATXN2, SH2B3: in our study we detected rs3184504 and rs653178 (intronic). Both were reported by Fischer et al., in particular, rs653178 is significantly associated with Löfgren syndrome [75].

A few identified genes of considerable interest have been implicated in sarcoidosis in functional studies:

- PLCL1 activates NLRP3 inflammatory pathway, thus regulating fibroblasts [95]. NLRP3 has been linked to Th17/Treg activation (with Th17 being a key player in sarcoidosis inflammation) [96] and its blockade modifies Th17/Treg balance [97].

  NLRP3 is a potential protein for sarcoidosis as a study by Huppertz et al. correlates NLRP3 to sarcoidosis in granuloma formation [98].

  We identified lead SNP rs1401090 (intronic) but it has not been previously associated with sarcoidosis.
- PPARG (PPARγ) is involved in IFNγ production and in patients with sarcoidosis it is reduced in the alveolar macrophages [99]. It has also a role in Th17 regulation [100]. Further evidence for PPARG contribution to sarcoidosis were observed: in PPARγ knock-out mice, sarcoidosis was induced through instillation of MWCNT [101,102].

  Another study highlights that the activation of PPARγ inhibits granuloma formation and reduces levels of CCL2 and osteopontin by decreasing NF-κB [103].

  No GWAS for PPARG was found in the literature, therefore lead SNP rs2972166 (intergenic) is a novel SNP.
- CD19 is a known antigen found on the surface of B cells which can be observed in the periphery of the sarcoidosis granuloma. A wide range of studies have shown varying patterns of its expression in affected patients: CD19+ B cells are increased in the peripheral blood [104] but are lower in BAL fluid compared to peripheral blood [105];

CD19+ B cells are increased in the BAL fluid of patients with positive endobronchial biopsy which also showed impaired pulmonary function test values, including FEV1/FVC, FEV1, FEF25 and VC, although only VC values are statistically significant[106]; in cases of severe disease, CD19+ B cells lymphopenia is observed in peripheral blood [107]; finally, in extra-pulmonary sarcoidosis CD19+ B cells are reduced [108].

As for the previous two genes, there was no available GWAS in literature, once more our finding lead SNP rs11645302 (downstream) is a novel SNP.

Given the participation of these genes in inflammation and immune regulation and responses, it is appropriate to conduct additional research to explore their potential as therapeutic targets.

For the following significant genes, there is no direct evidence of their involvement in sarcoidosis pathogenesis, but it is strongly suggested a role in immune response, specifically cytokine production and regulation of NF-κB:

- RNF215 (ring finger protein 215) negatively regulates type 1 IFN as increased expression of RNF215 blocks NF-κB p65 binding to the IFNB1 promoter [109], preventing IFNβ1 transcription [110].

  Lead SNP rs757870 (intronic) is a novel finding.

- LAT, RP11-231C14.10, RP11-264B17.3: a study reports LAT as a linker for activation of T cells [111]. To date, there is no data about the other two genes.

- RPS6KA4/MSK2 plays a role in the regulation of inflammation [112]. Increased levels of TNF, IL-6 and IL-12 are correlated with reduced levels of MSK1 and MSK2 [112,113].

- PUS10 plays a direct role in apoptosis pathways through a positive feedback mechanism involving caspase-3. Moreover, defects in PUS10 transport or interactions lead to increased cell survival [114].

  It has been shown that BAL fluid of affected patients presents lymphocytes that display a non-apoptotic morphology associated with endogenous caspase-3 activity[115].

  Lead SNP rs10181042 (intronic) is a novel finding.

- ESYT2 (extended synaptogamin 2) is a lipid transport protein involved in plasma membrane lipid homeostasis [116]. It depletes diacylglycerol from the plasma membrane causing the down regulation of T-cells activation and reduction of IL-12 levels: it is demonstrated that the absence of ESYT2 leads to high levels of diacylglycerol in the plasma membrane, thus increasing activation of T cells, TCR signaling and stimulated CD4+ T cells production of IL-12 [117]. IL-12 is a Th-1 cytokine whose expression in CD4+ and CD8+ T cells is increased in the BAL fluid of patients with sarcoidosis compared to controls [118].

  An experiment including mice with ESYT2 gene deficiency shows an altered embryogenic fibroblasts migration [119].

  Lead SNP rs7803766 (intronic) is a novel finding.

- UBASH3A (ubiquitin-associated and src-homology 3), also known as STS2 [120], is involved in reducing the activation and signaling of T cells and in the expression of cytokines such as IL-2 and IFNγ [121]. Experiments on mice show that on STS-1 and STS-2 knock out mice show an augmented response following TCR stimulation, enhanced levels of IL-2 and IFNγ compared to wild type mice [121]; on the contrary, knock-out for STS2 alone has a weaker role in TCR stimulation and IFNγ

production[122]. Furthermore, defect in UBASH3A/STS-2 and UBASH3B/STS-1 lead to an enhanced ZAP10, a transducer for TCR [120], through phosphorylation [120,121], while defect in STS-2 alone is linked with a less pronounced ZAP70 phosphorylation [122].

A study by Ge et al. reports that in type 1 diabetes It has a role in inhibiting NF-κB signaling: an overexpression of UBASH3A reduces IL-12 and NF-κB levels [123].

Intronic SNP rs1893592 was detected in our study.

- HOMER2 negatively regulates the activation of T cells, through TCR modulation and CD28 signaling, and plays a role in regulating NF-κB, NFAT and calcium signaling. HOMER triple knock-out mice produced more IL-2 compared to wild type mice, after stimulation with anti-CD3 T cells [124].

  We identified SNP rs11856316 (UTR3) as a new lead variant.

- SUMO2P: alteration in SUMOylation levels influences the differentiation of Tregs and the transcription of IL-17 in Th17 lymphocytes [125].

  Intergenic SNP rs3131856 is a novel finding.

The following genes have been significantly associated to sarcoidosis in our analyses. However, there is neither direct evidence of a functional role in sarcoidosis nor clear involvement in pathological immune response. Therefore, further research on the potential involvement of these genes in sarcoidosis, whether protective or contributes to increasing risk, is needed to complete the significance of our findings and to give insights for therapeutic approaches.

- ACOXL has a role in oxidative stress and it is implicated in the metabolism of fatty acids [126]. Lead SNP rs10183338 (intronic) emerged as novel variant in the analyses.

- MAT1A: a role in oxidative stress has been hypothesized [127].

- MACROD1 is involved in inflammation [128].

Concerning the limitations of our study, asides from those inherent to genome wide association studies, there are also limitations related to the consultation of public summary statistics to conduct a meta-analysis: first and foremost, it was not possible to access data on an individual level; secondly a more detailed exploration of genomic data, for instance validation of polygenic risk scores, could not be performed. Moreover, the datasets only provide information whether patients were affected or not by sarcoidosis, not the clinical subtypes, targeted organs or the gravity of the disease.

Despite minor differences arose from the choice of technical parameters applied in the meta-analysis and functional annotation processes, our study's main results remained stable and reliable across the repetition of analytical configurations.

CONCLUSION

Availing of publicly available summary statistics of four international datasets and projects which are FinnGen, MVP, Pan-UKBB and BBJ, we conducted the largest meta-analysis concerning European, African and East Asian ancestries to date. For EUR ancestry meta-analysis nineteen risk loci were identified, both previously associated to sarcoidosis and novel loci. As for AFR ancestry meta-analysis two risk loci were identified.

Furthermore, we conducted a multi-ancestry meta-analysis on EUR, AFR and EAS ancestries and identified eighteen risk loci. Comparing multi-ancestry and EUR meta-analyses, some loci were confirmed but others lost statistical significance in the multi-ancestry meta-analysis, meaning that they are loci seemingly private to Europeans.

Given the poor data and small sample sizes, both cases and controls of AFR and EAS ancestries, further research aimed at collecting more data on these ancestries to obtain novel significant ancestry-specific loci through meta-analysis is encouraged.

Subsequently, we reviewed the available literature for GWAS and functional studies on the identified genes. For previously identified genes, we compared our findings, particularly SNPs, and reported their direct involvement in sarcoidosis or information related to immune response.

As for the other identified genes, we observed that most participate in the regulation of immune response, that plays a crucial role in sarcoidosis pathogenesis.

We hope that our findings contribute to expanding knowledge about sarcoidosis etiology, specifically genetics as a risk factor in the development of the disease, about risk loci not only in extensively studied populations, but also in understudied populations to date. We

also aspire to provide further insights of sarcoidosis pathogenesis at a molecular level which may facilitate the development of tailored approaches for diagnosis, prognosis and treatment.

ACKNOWLEDGEMENTS

Arrivato a termine del mio percorso formativo desidero esprimere la mia sincera gratitudine a coloro che hanno reso possibile la stesura di questa tesi.

Ringrazio il Professor Enrico Clini, Direttore della Struttura Complessa di Malattie dell'Apparato Respiratorio dell'Azienda Ospedaliera-Universitaria di Modena per la sua ampia disponibilità nell'avermi accolto in reparto come tirocinante in questi anni e per aver accolto la mia richiesta di svolgere la tesi in ambito pneumologico.

Ringrazio la Dottoressa Stefania Cerri, coordinatrice del Centro Malattie Rare del Polmone (MaRP) dell'AOU di Modena, per avermi accolto in ambulatorio, introdotto alle patologie rare e per avermi permesso di partecipare al progetto che ha reso possibile l'elaborazione di questa tesi.

Ringrazio il Dottor Filippo Gozzi, che è stato il primo tutor ad avermi introdotto alla professione medica. Sono riconoscente per il suo prezioso contributo alla mia crescita professionale e per il suo supporto durante lo svolgimento del progetto.

Ringrazio il Dottor Roberto Tonelli per la sua competenza e le conoscenze che mi ha fornito durante i tirocini in reparto e per il suo prezioso consiglio di svolgere la tesi al Centro MaRP.

Un sentito ringraziamento alla Dottoressa Federica Andolfi, Specializzanda di Malattie dell'Apparato Respiratorio, per avermi affiancato per tutta la durata del progetto, per la stimolante esperienza di collaborazione, per l'entusiasmo con cui ha saputo coinvolgermi, per le conoscenze trasmesse durante il tirocinio e per la sua disponibilità.

Un ringraziamento doveroso al Dottor Andrea Ricci per aver reso possibile questo progetto, per la sua presenza e attenzione costante e per la passione per l'argomento che ha saputo trasmettermi.

Ringrazio la Prof.ssa Bianca Beghé per il suo contributo alla mia crescita professionale durante il tirocinio in ambulatorio.

Un ringraziamento va a tutti i Dirigenti e Specializzandi della Struttura Complessa di Malattie dell'Apparato Respiratorio che hanno favorito la mia maturazione personale e sviluppo professionale medico.

Ringrazio la Dottoressa Elena Gallina, bibliotecaria della Biblioteca Universitaria Medica, per i preziosi consigli offerti nella revisione della bibliografia e per la sua disponibilità.

Desidero esprimere la mia più profonda gratitudine ai miei genitori per avermi sempre sostenuto nelle mie scelte e creduto in me, ai miei parenti, nonostante la lontananza hanno sempre creduto che sarei riuscito a portare a termine questo percorso, ai miei amici da sempre per il loro supporto e per aver contribuito a rendere sereno il mio percorso, ai miei amici dell'università per le condivisioni di momenti gioiosi, supporto reciproco nei momenti di particolare sfida e per essere stati una parte integrante della mia maturazione professionale.

REFERENCES

1. Sharma OP. Definition and history of sarcoidosis. In: *Sarcoidosis (out of Print)*. European Respiratory Society; 2005:1-12. doi:10.1183/1025448x.00032001

2. Spagnolo P. Sarcoidosis: a Critical Review of History and Milestones. *Clin Rev Allergy Immunol*. 2015;49(1):1-5. doi:10.1007/s12016-015-8480-0

3. Rugarli C. *Medicina interna sistematica*. Vol <<Medicina interna sistematica>> 1. EDRA <casa editrice>; 2021.

4. Arkema EV, Cozier YC. Sarcoidosis epidemiology: recent estimates of incidence, prevalence and risk factors. *Curr Opin Pulm Med*. 2020;26(5):527-534. doi:10.1097/MCP.0000000000000715

5. Rossides M, Darlington P, Kullberg S, Arkema EV. Sarcoidosis: Epidemiology and clinical insights. *J Intern Med*. 2023;293(6):668-680. doi:10.1111/joim.13629

6. Ungprasert P, Ryu JH, Matteson EL. Clinical Manifestations, Diagnosis, and Treatment of Sarcoidosis. *Mayo Clin Proc Innov Qual Outcomes*. 2019;3(3):358-375. doi:10.1016/j.mayocpiqo.2019.04.006

7. Hena KM. Sarcoidosis Epidemiology: Race Matters. *Front Immunol*. 2020;11:537382. doi:10.3389/fimmu.2020.537382

8. Baughman RP, Field S, Costabel U, et al. Sarcoidosis in America. Analysis Based on Health Care Use. *Ann Am Thorac Soc*. 2016;13(8):1244-1252. doi:10.1513/AnnalsATS.201511-760OC

9. Kearney GD, Obi ON, Maddipati V, et al. Sarcoidosis deaths in the United States: 1999-2016. *Respir Med*. 2019;149:30-35. doi:10.1016/j.rmed.2018.11.010

10. Gandhi S, Tonelli R, Murray M, Samarelli AV, Spagnolo P. Environmental Causes of Idiopathic Pulmonary Fibrosis. *Int J Mol Sci*. 2023;24(22):16481. doi:10.3390/ijms242216481

11. Dehara M, Nguyen NV, Arkema EV. Smoking and the risk of sarcoidosis: a systematic review and meta-analysis. *Respir Med*. 2025;241:108089. doi:10.1016/j.rmed.2025.108089

12. Haraldsdóttir SÓ, Jonasson JG, Jorundsdottir KB, Hannesson HJ, Gislason T, Gudbjornsson B. Sarcoidosis in Iceland: a nationwide study of epidemiology, clinical picture and environmental exposure. *ERJ Open Res*. 2021;7(4):00550-02021. doi:10.1183/23120541.00550-2021

13. Judson MA. Environmental Risk Factors for Sarcoidosis. *Front Immunol*. 2020;11:1340. doi:10.3389/fimmu.2020.01340

14. Moller DR, Rybicki BA, Hamzeh NY, et al. Genetic, Immunologic, and Environmental Basis of Sarcoidosis. *Ann Am Thorac Soc*. 2017;14(Supplement_6):S429-S436. doi:10.1513/AnnalsATS.201707-565OT

15. Izbicki G, Chavko R, Banauch GI, et al. World Trade Center “sarcoid-like” granulomatous pulmonary disease in New York City Fire Department rescue workers. *Chest*. 2007;131(5):1414-1423. doi:10.1378/chest.06-2114

16. Guzik TJ, Skiba DS, Touyz RM, Harrison DG. The role of infiltrating immune cells in dysfunctional adipose tissue. *Cardiovasc Res*. 2017;113(9):1009-1023. doi:10.1093/cvr/cvx108

17. Judson MA, Tiwari A, Gemoets DE. The Relationship of Obesity and OSA to the Development of Sarcoidosis: A Large Retrospective Case-Control US Veterans Administration Analysis. *Chest*. 2022;162(5):1086-1092. doi:10.1016/j.chest.2022.05.008

18. Wahlström J, Dengjel J, Persson B, et al. Identification of HLA-DR-bound peptides presented by human bronchoalveolar lavage cells in sarcoidosis. *J Clin Invest*. 2007;117(11):3576-3582. doi:10.1172/JCI32401

19. Cain H, Kraus B. Immunofluorescence microscopic demonstration of vimentin filaments in asteroid bodies of sarcoidosis. A comparison with electron microscopic findings. *Virchows Arch B Cell Pathol Incl Mol Pathol*. 1983;42(2):213-226. doi:10.1007/BF02890384

20. Culver DA. Pathology and pathogenesis of sarcoidosis. In: Connor R, ed. *Uptodate*. Wolters Kluwer. Accessed May 12, 2025. www.uptodate.com

21. Cabeza-Cabrerizo M, Cardoso A, Minutti CM, Pereira da Costa M, Reis e Sousa C. Dendritic Cells Revisited. *Annu Rev Immunol*. 2021;39:131-166. doi:10.1146/annurev-immunol-061020-053707

22. Liu H, Jakubzick C, Osterburg AR, et al. Dendritic Cell Trafficking and Function in Rare Lung Diseases. *Am J Respir Cell Mol Biol*. 2017;57(4):393-402. doi:10.1165/rcmb.2017-0051PS

23. Miedema J, Cinetto F, Smed-Sörensen A, Spagnolo P. The immunopathogenesis of sarcoidosis. *J Autoimmun*. 2024;149:103247. doi:10.1016/j.jaut.2024.103247

24. Pagán AJ, Ramakrishnan L. The Formation and Function of Granulomas. *Annu Rev Immunol*. 2018;36:639-665. doi:10.1146/annurev-immunol-032712-100022

25. Arango Duque G, Descoteaux A. Macrophage Cytokines: Involvement in Immunity and Infectious Diseases. *Front Immunol*. 2014;5:491. doi:10.3389/fimmu.2014.00491

26. Shamaei M, Mortaz E, Pourabdollah M, et al. Evidence for M2 macrophages in granulomas from pulmonary sarcoidosis: A new aspect of macrophage heterogeneity. *Hum Immunol*. 2018;79(1):63-69. doi:10.1016/j.humimm.2017.10.009

27. Shigehara K, Shijubo N, Ohmichi M, et al. IL-12 and IL-18 are increased and stimulate IFN-gamma production in sarcoid lungs. *J Immunol Baltim Md 1950*. 2001;166(1):642-649. doi:10.4049/jimmunol.166.1.642

28. Miyara M, Amoura Z, Parizot C, et al. The immune paradox of sarcoidosis and regulatory T cells. *J Exp Med*. 2006;203(2):359-370. doi:10.1084/jem.20050648

29. Broos CE, Koth LL, van Nimwegen M, et al. Increased T-helper 17.1 cells in sarcoidosis mediastinal lymph nodes. *Eur Respir J*. 2018;51(3):1701124. doi:10.1183/13993003.01124-2017

30. Ramesh R, Kozhaya L, McKevitt K, et al. Pro-inflammatory human Th17 cells selectively express P-glycoprotein and are refractory to glucocorticoids. *J Exp Med*. 2014;211(1):89-104. doi:10.1084/jem.20130301

31. Kunkel SL, Lukacs NW, Strieter RM, Chensue SW. Th1 and Th2 responses regulate experimental lung granuloma development. *Sarcoidosis Vasc Diffuse Lung Dis Off J WASOG*. 1996;13(2):120-128.

32. Schmidt A, Oberle N, Krammer PH. Molecular mechanisms of treg-mediated T cell suppression. *Front Immunol*. 2012;3:51. doi:10.3389/fimmu.2012.00051

33. Robbins SL, Cotran RS. *Robbins e Cotran Le basi patologiche delle malattie*. Vol 2: Malattie degli organi e degli apparati. EDRA <casa editrice>; 2022.

34. Murray JF. *Murray and Nadel's textbook of respiratory medicine*. Vol <<Murray and Nadel's textbook of respiratory medicine>> Vol. 2. Saunders Elsevier; 2010.

35. Sève P, Pacheco Y, Durupt F, et al. Sarcoidosis: A Clinical Overview from Symptoms to Diagnosis. *Cells*. 2021;10(4):766. doi:10.3390/cells10040766

36. Talmadge EKJr, Culver DA. Clinical manifestations and diagnosis of sarcoidosis. In: Connor R, ed. *Uptodate*. Wolters Kluwer. Accessed May 16, 2025. www.uptodate.com

37. Boucly A, Weatherald J, Savale L, et al. Risk assessment, prognosis and guideline implementation in pulmonary arterial hypertension. *Eur Respir J*. 2017;50(2):1700889. doi:10.1183/13993003.00889-2017

38. Prystowksy S, Sanchez M. Cutaneous manifestations of sarcoidosis. In: Connor R, ed. *Uptodate*. Wolters Kluwer. Accessed May 17, 2025. www.uptodate.com

39. Culver DA. Overview of extrapulmonary manifestations of sarcoidosis. In: Connor R, ed. *Uptodate*. Wolters Kluwer. Accessed May 17, 2025. www.uptodate.com

40. Shah N, Mitra A. Gastrointestinal and Hepatic Sarcoidosis: A Review Article. *Clin Liver Dis*. 2021;17(4):301-307. doi:10.1002/cld.1055

41. Stern BJ. Neurologic sarcoidosis. In: Connor R, ed. *Uptodate*. Wolters Kluwer. Accessed May 17, 2025. www.uptodate.com

42. Moretti A, Bruzzi G, Andrisani D, et al. When sarcoidosis hits down: a case of prostatic sarcoidosis. *Multidiscip Respir Med*. 2024;19(1):956. doi:10.5826/mrm.2024.956

43. Kaminsky DA. Overview of pulmonary function testing in adults. In: Connor R, ed. *Uptodate*. Wolters Kluwer. Accessed May 18, 2025. www.uptodate.com

44. Hunninghake GW, Costabel U, Ando M, et al. ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis

and other Granulomatous Disorders. *Sarcoidosis Vasc Diffuse Lung Dis Off J WASOG*. 1999;16(2):149-173.

45. Judson MA, Baughman RP, Costabel U, et al. Safety and efficacy of ustekinumab or golimumab in patients with chronic sarcoidosis. *Eur Respir J*. 2014;44(5):1296-1307. doi:10.1183/09031936.00000914

46. Rossides M, Grunewald J, Eklund A, et al. Familial aggregation and heritability of sarcoidosis: a Swedish nested case-control study. *Eur Respir J*. 2018;52(2):1800385. doi:10.1183/13993003.00385-2018

47. Janeway C. *Immunobiologia di Janeway*. Piccin <casa editrice>; 2019.

48. Gs C, Av H. Genetics of susceptibility to human infectious disease. *Nat Rev Genet*. 2001;2(12). doi:10.1038/35103577

49. Liao SY, Fingerlin T, Maier L. Genetic predisposition to sarcoidosis. *J Autoimmun*. 2024;149:103122. doi:10.1016/j.jaut.2023.103122

50. Spagnolo P, du Bois RM. Genetics of sarcoidosis. *Clin Dermatol*. 2007;25(3):242-249. doi:10.1016/j.clindermatol.2007.03.001

51. Sikorová K, Moon SJ, Yoon HY, Strnad A, Song JW, Petrek M. HLA class II variants defined by next generation sequencing are associated with sarcoidosis in Korean patients. *Sci Rep*. 2022;12(1):9302. doi:10.1038/s41598-022-13199-w

52. Grunewald J, Eklund A. Löfgren's syndrome: human leukocyte antigen strongly influences the disease course. *Am J Respir Crit Care Med*. 2009;179(4):307-312. doi:10.1164/rccm.200807-1082OC

53. Valentonyte R, Hampe J, Huse K, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet*. 2005;37(4):357-364. doi:10.1038/ng1519

54. Suzuki H, Ota M, Meguro A, et al. Genetic characterization and susceptibility for sarcoidosis in Japanese patients: risk factors of BTNL2 gene polymorphisms and HLA class II alleles. *Invest Ophthalmol Vis Sci*. 2012;53(11):7109-7115. doi:10.1167/iovs.12-10491

55. Levin AM, Adrianto I, Datta I, et al. Association of HLA-DRB1 with Sarcoidosis Susceptibility and Progression in African Americans. *Am J Respir Cell Mol Biol*. 2015;53(2):206-216. doi:10.1165/rcmb.2014-0227OC

56. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun*. 2013;14(1):13-18. doi:10.1038/gene.2012.48

57. Attia J. Genetic association and GWAS studies: Principles and applications. In: Connor R, ed. *Uptodate*. Wolters Kluwer.

58. Strachan T. *Genetica molecolare umana*. Nicola Zanichelli <casa editrice>; 2021.

59. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518. doi:10.1038/s41586-022-05473-8

60. Karczewski KJ, Gupta R, Kanai M, et al. Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. Published online October 1, 2024:2024.03.13.24303864. doi:10.1101/2024.03.13.24303864

61. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016

62. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415-1424. doi:10.1038/s41588-021-00931-x

63. Nagai A, Hirata M, Kamatani Y, et al. Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol*. 2017;27(3S):S2-S8. doi:10.1016/j.je.2016.12.005

64. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2024. Accessed October 15, 2024. https://www.r-project.org/

65. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340

66. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5

67. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi:10.1093/nar/gkq603

68. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892

69. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7. doi:10.1186/s13742-015-0047-8

70. Purcell S, Chang C. PLINK 1.9 (BETA 7.6). Published online 2019. Accessed October 15, 2024. https://www.cog-genomics.org/plink/1.9

71. Byrska-Bishop M, Evani US, Zhao X, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell*. 2022;185(18):3426-3440.e19. doi:10.1016/j.cell.2022.08.004

72. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

73. FinnGen Consortium. FinnGen r12 Study-Sarcoidosis (D3_SARCOIDOSIS). FinnGen MVP+UKBB Meta-analysis Browser. 2024. Accessed May 28, 2025. https://mvp-ukbb.finngen.fi/pheno/d3sarcoidosis

74. Meguro A, Ishihara M, Petrek M, et al. Genetic control of CCL24, POR, and IL23R contributes to the pathogenesis of sarcoidosis. *Commun Biol*. 2020;3(1):465. doi:10.1038/s42003-020-01185-9

75. 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):727-736. doi:10.1164/rccm.201503-0418OC

76. Matsushita E, Asai N, Enomoto A, et al. Protective role of Gipie, a Girdin family protein, in endoplasmic reticulum stress responses in endothelial cells. *Mol Biol Cell*. 2011;22(6):736-747. doi:10.1091/mbc.E10-08-0724

77. Kennedy JM, Fodil N, Torre S, et al. CCDC88B is a novel regulator of maturation and effector functions of T cells during pathological inflammation. *J Exp Med*. 2014;211(13):2519-2535. doi:10.1084/jem.20140455

78. Fischer A, Schmid B, Ellinghaus D, et al. A novel sarcoidosis risk locus for Europeans on chromosome 11q13.1. *Am J Respir Crit Care Med*. 2012;186(9):877-885. doi:10.1164/rccm.201204-0708OC

79. Wang J, Guo C, Liu S, et al. Annexin A11 in disease. *Clin Chim Acta Int J Clin Chem*. 2014;431:164-168. doi:10.1016/j.cca.2014.01.031

80. Hofmann S, Franke A, Fischer A, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*. 2008;40(9):1103-1106. doi:10.1038/ng.198

81. Karakaya B, van der Vis JJ, Veltkamp M, Biesma DH, Grutters JC, van Moorsel CHM. ANXA11 rs1049550 Associates with Löfgren's Syndrome and Chronic Sarcoidosis Patients. *Cells*. 2022;11(9):1557. doi:10.3390/cells11091557

82. Oppmann B, Lesley R, Blom B, et al. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. *Immunity*. 2000;13(5):715-725. doi:10.1016/s1074-7613(00)00070-4

83. Iwakura Y, Ishigame H. The IL-23/IL-17 axis in inflammation. *J Clin Invest*. 2006;116(5):1218-1222. doi:10.1172/JCI28508

84. Facco M, Cabrelle A, Teramo A, et al. Sarcoidosis is a Th1/Th17 multisystem disorder. *Thorax*. 2011;66(2):144-150. doi:10.1136/thx.2010.140319

85. 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. *J Am Acad Dermatol*. 2012;66(6):901-910, 910.e1-2. doi:10.1016/j.jaad.2011.06.017

86. Zhong Z, Su G, Kijlstra A, Yang P. Activation of the interleukin-23/interleukin-17 signalling pathway in autoinflammatory and autoimmune uveitis. *Prog Retin Eye Res*. 2021;80:100866. doi:10.1016/j.preteyeres.2020.100866

87. Schreiber T, Falk-Paulsen M, Kuiper J, et al. IL23R on myeloid cells is involved in murine pulmonary granuloma formation. *Exp Lung Res*. 2021;47(7):344-353. doi:10.1080/01902148.2021.1962433

88. Kim HS, Choi D, Lim LL, et al. Association of interleukin 23 receptor gene with sarcoidosis. *Dis Markers*. 2011;31(1):17-24. doi:10.3233/DMA-2011-0796

89. Yamamoto M, Kensler TW, Motohashi H. The KEAP1-NRF2 System: a Thiol-Based Sensor-Effector Apparatus for Maintaining Redox Homeostasis. *Physiol Rev*. 2018;98(3):1169-1203. doi:10.1152/physrev.00023.2017

90. Zhou Y, Chen Y, Shi Y, et al. FAM117B promotes gastric cancer growth and drug resistance by targeting the KEAP1/NRF2 signaling pathway. *J Clin Invest*. 2023;133(3):e158705. doi:10.1172/JCI158705

91. Fois SS, Canu S, Fois AG. The Role of Oxidative Stress in Sarcoidosis. *Int J Mol Sci*. 2021;22(21):11712. doi:10.3390/ijms222111712

92. Mor A, Segal Salto M, Katav A, et al. Blockade of CCL24 with a monoclonal antibody ameliorates experimental dermal and pulmonary fibrosis. *Ann Rheum Dis*. 2019;78(9):1260-1268. doi:10.1136/annrheumdis-2019-215119

93. Arakelyan A, Kriegova E, Kubistova Z, et al. Protein levels of CC chemokine ligand (CCL)15, CCL16 and macrophage stimulating protein in patients with sarcoidosis. *Clin Exp Immunol*. 2009;155(3):457-465. doi:10.1111/j.1365-2249.2008.03832.x

94. Abu El-Asrar AM, Berghmans N, Al-Obeidan SA, et al. The CC chemokines CCL8, CCL13 and CCL20 are local inflammatory biomarkers of HLA-B27-associated uveitis. *Acta Ophthalmol (Copenh)*. 2019;97(1):e122-e128. doi:10.1111/aos.13835

95. Luo S, Li XF, Yang YL, et al. PLCL1 regulates fibroblast-like synoviocytes inflammation via NLRP3 inflammasomes in rheumatoid arthritis. *Adv Rheumatol Lond Engl*. 2022;62(1):25. doi:10.1186/s42358-022-00252-5

96. Patel D, Gaikwad S, Challagundla N, Nivsarkar M, Agrawal-Rajput R. Spleen tyrosine kinase inhibition ameliorates airway inflammation through modulation of NLRP3 inflammosome and Th17/Treg axis. *Int Immunopharmacol*. 2018;54:375-384. doi:10.1016/j.intimp.2017.11.026

97. Chen L, Hou W, Liu F, et al. Blockade of NLRP3/Caspase-1/IL-1β Regulated Th17/Treg Immune Imbalance and Attenuated the Neutrophilic Airway Inflammation in an Ovalbumin-Induced Murine Model of Asthma. *J Immunol Res*. 2022;2022:9444227. doi:10.1155/2022/9444227

98. Huppertz C, Jäger B, Wieczorek G, et al. The NLRP3 inflammasome pathway is activated in sarcoidosis and involved in granuloma formation. *Eur Respir J*. 2020;55(3):1900119. doi:10.1183/13993003.00119-2019

99. Culver DA, Barna BP, Raychaudhuri B, et al. Peroxisome proliferator-activated receptor gamma activity is deficient in alveolar macrophages in pulmonary sarcoidosis. *Am J Respir Cell Mol Biol*. 2004;30(1):1-5. doi:10.1165/rcmb.2003-0304RC

100. Sanderford V, Barna BP, Barrington RA, et al. PPARγ Deficiency in Carbon Nanotube-elicited Granulomatous Inflammation Promotes a Th17 Response to a Microbial Antigen. *J Nanomedicine Nanotechnol*. 2020;11(1):541. doi:10.35248/2157-7439.20.11.541

101. Barna BP, Culver DA, Abraham S, et al. Depressed peroxisome proliferator-activated receptor gamma (PPargamma) is indicative of severe pulmonary sarcoidosis: possible involvement of interferon gamma (IFN-gamma). *Sarcoidosis Vasc Diffuse Lung Dis Off J WASOG*. 2006;23(2):93-100.

102. Barna BP, Malur A, Thomassen MJ. Studies in a Murine Granuloma Model of Instilled Carbon Nanotubes: Relevance to Sarcoidosis. *Int J Mol Sci*. 2021;22(7):3705. doi:10.3390/ijms22073705

103. McPeek M, Malur A, Tokarz DA, Murray G, Barna BP, Thomassen MJ. PPAR-gamma pathways attenuate pulmonary granuloma formation in a carbon nanotube induced murine model of sarcoidosis. *Biochem Biophys Res Commun*. 2018;503(2):684-690. doi:10.1016/j.bbrc.2018.06.061

104. Kudryavtsev I, Serebriakova M, Starshinova A, et al. Imbalance in B cell and T Follicular Helper Cell Subsets in Pulmonary Sarcoidosis. *Sci Rep*. 2020;10(1):1059. doi:10.1038/s41598-020-57741-0

105. d'Alessandro M, Bergantini L, Cameli P, et al. Adaptive immune system in pulmonary sarcoidosis-Comparison of peripheral and alveolar biomarkers. *Clin Exp Immunol*. 2021;205(3):406-416. doi:10.1111/cei.13635

106. Kieszko R, Krawczyk P, Michnar M, Chocholska S, Milanowski J. The yield of endobronchial biopsy in pulmonary sarcoidosis: connection between spirometric impairment and lymphocyte subpopulations in bronchoalveolar lavage fluid. *Respir Int Rev Thorac Dis*. 2004;71(1):72-76. doi:10.1159/000075652

107. Sweiss NJ, Salloum R, Gandhi S, et al. Significant CD4, CD8, and CD19 lymphopenia in peripheral blood of sarcoidosis patients correlates with severe disease manifestations. *PloS One*. 2010;5(2):e9088. doi:10.1371/journal.pone.0009088

108. Zurkova M, Kolek V, Tomankova T, Kriegova E. Extrapulmonary involvement in patients with sarcoidosis and comparison of routine laboratory and clinical data to pulmonary involvement. *Biomed Pap Med Fac Univ Palacky Olomouc Czechoslov*. 2014;158(4):613-620. doi:10.5507/bp.2014.026

109. Wu Y, Chen D, Hu Y, et al. Ring Finger Protein 215 Negatively Regulates Type I IFN Production via Blocking NF-κB p65 Activation. *J Immunol Baltim Md 1950*. 2022;209(10):2012-2021. doi:10.4049/jimmunol.2200346

110. Oeckinghaus A, Ghosh S. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harb Perspect Biol*. 2009;1(4):a000034. doi:10.1101/cshperspect.a000034

111. Zhang W, Trible RP, Samelson LE. LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation. *Immunity*. 1998;9(2):239-246. doi:10.1016/s1074-7613(00)80606-8

112. Sattarifard H, Safaei A, Khazeeva E, Rastegar M, Davie JR. Mitogen- and stress-activated protein kinase (MSK1/2) regulated gene expression in normal and disease states. *Biochem Cell Biol Biochim Biol Cell*. 2023;101(3):204-219. doi:10.1139/bcb-2022-0371

113. Ananieva O, Darragh J, Johansen C, et al. The kinases MSK1 and MSK2 act as negative regulators of Toll-like receptor signaling. *Nat Immunol*. 2008;9(9):1028-1036. doi:10.1038/ni.1644

114. Jana S, Hsieh AC, Gupta R. Reciprocal amplification of caspase-3 activity by nuclear export of a putative human RNA-modifying protein, PUS10 during TRAIL-induced apoptosis. *Cell Death Dis*. 2017;8(10):e3093. doi:10.1038/cddis.2017.476

115. Stridh H, Planck A, Gigliotti D, Eklund A, Grunewald J. Apoptosis resistant bronchoalveolar lavage (BAL) fluid lymphocytes in sarcoidosis. *Thorax*. 2002;57(10):897-901. doi:10.1136/thorax.57.10.897

116. Saheki Y, De Camilli P. The Extended-Synaptotagmins. *Biochim Biophys Acta Mol Cell Res*. 2017;1864(9):1490-1493. doi:10.1016/j.bbamcr.2017.03.013

117. Benavides N, Giraudo CG. Extended-Synaptotagmin-1 and -2 control T cell signaling and function. *EMBO Rep*. 2024;25(1):286-303. doi:10.1038/s44319-023-00011-7

118. Prasse A, Georges CG, Biller H, et al. Th1 cytokine pattern in sarcoidosis is expressed by bronchoalveolar CD4+ and CD8+ T cells. *Clin Exp Immunol*. 2000;122(2):241-248. doi:10.1046/j.1365-2249.2000.01365.x

119. Herdman C, Tremblay MG, Mishra PK, Moss T. Loss of Extended Synaptotagmins ESyt2 and ESyt3 does not affect mouse development or viability, but in vitro cell migration and survival under stress are affected. *Cell Cycle Georget Tex*. 2014;13(16):2616-2625. doi:10.4161/15384101.2014.943573

120. Wang H, Concannon P, Ge Y. Roles of TULA-family proteins in T cells and autoimmune diseases. *Genes Immun*. 2025;26(1):54-62. doi:10.1038/s41435-024-00300-8

121. Carpino N, Turner S, Mekala D, et al. Regulation of ZAP-70 activation and TCR signaling by two related proteins, Sts-1 and Sts-2. *Immunity*. 2004;20(1):37-46. doi:10.1016/s1074-7613(03)00351-0

122. San Luis B, Sondgeroth B, Nassar N, Carpino N. Sts-2 is a phosphatase that negatively regulates zeta-associated protein (ZAP)-70 and T cell receptor signaling pathways. *J Biol Chem*. 2011;286(18):15943-15954. doi:10.1074/jbc.M110.177634

123. Ge Y, Paisie TK, Newman JRB, McIntyre LM, Concannon P. UBASH3A Mediates Risk for Type 1 Diabetes Through Inhibition of T-Cell Receptor-Induced NF-κB Signaling. *Diabetes*. 2017;66(7):2033-2043. doi:10.2337/db16-1023

124. Huang GN, Huso DL, Bouyain S, et al. NFAT binding and regulation of T cell activation by the cytoplasmic scaffolding Homer proteins. *Science*. 2008;319(5862):476-481. doi:10.1126/science.1151227

125. Chang HM, Yeh ETH. SUMO: From Bench to Bedside. *Physiol Rev*. 2020;100(4):1599-1619. doi:10.1152/physrev.00025.2019

126. Dong Y, Zhao Z, Simayi M, et al. Transcriptome profiles of fatty acid metabolism-related genes and immune infiltrates identify hot tumors for immunotherapy in cutaneous melanoma. *Front Genet*. 2022;13:860067. doi:10.3389/fgene.2022.860067

127. von Montfort C, Matias N, Fernandez A, et al. Mitochondrial GSH determines the toxic or therapeutic potential of superoxide scavenging in steatohepatitis. *J Hepatol*. 2012;57(4):852-859. doi:10.1016/j.jhep.2012.05.024

128. Zang L, Hong Q, Yang G, et al. MACROD1/LRP16 Enhances LPS-Stimulated Inflammatory Responses by Up-Regulating a Rac1-Dependent Pathway in Adipocytes. *Cell Physiol Biochem Int J Exp Cell Physiol Biochem Pharmacol*. 2018;51(6):2591-2603. doi:10.1159/000495931