MOLECULAR STRATEGIES TO DISTINGUISH KEY SUBPHENOTYPES IN SARCOIDOSIS

by

Nancy González Casanova

A Dissertation Submitted to the Faculty of the

COLLEGE OF MEDICINE

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN CLINICAL TRANSLATIONAL

SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2020

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Nancy G. Casanova, titled: Molecular strategies to distinguish key su phenotypes in sarcoidosis

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Dr. Joe G.N. Garcia Date: 12/12/19
[Committee Chair Name]

Dr. Jorge Gomez Date: 12/12/19
[Committee Member Name]

Dr. Christian Bime Date: 12/12/2019
[Committee Member Name]

Xiaoguang Sun Xiaoguang Sun Date: 12/12/2019
[Committee Member Name]

Dr. Melanie Culver Date: 12/12/2019
[Committee Member Name]

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dr. Joe G.N. Garcia Date: 12/12/19
[Committee Chair Name]
Dissertation Committee Chair

ACKNOWLEDGEMENTS

To my mother, Nancy Sandoval, my husband Daniel Casanova and my children, Daniel, Pablo and Nancy for all their support and encouragement.

To my mentor Dr. Joe G. N. Garcia, for his continued inspiration, support and guidance. I would like to express my gratitude to my committee members Gomez, Dr. Bime, Dr. Sun and Dr. Culver for their inspiration and suggestions; they were instrumental in the development of this dissertation.

Finally, I would like to thank to all my colleges at the Garcia Lab members for their continued support and help.

Thank you all for your contributions that made me a better scientist.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AA	African Americans
ACCESS	A Case-Control Etiology Study of Sarcoidosis
BAL	Bronchioalveolar lavage
Cocci	Coccidioidomycosis
DE	Differentia Expression
DEGs	Differentially Expressed Genes
DNA	Deoxyribonucleic acid
EA	European Americans
eQTL	Expression quantitative loci
FC	Fold Change
FFPE	Formalin fixed paraffin-embedded
GAIN	Genetic Association Information Network
GRADS	Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis
GTEx	Genotype-Tissue Expression
GWAS	Genome Wide Association Studies
HLA	Human Leucocyte Antigen
IPF	Idiopathic Pulmonary Fibrosis
LD	Linkage disequilibrium
MHC	Major Histocompatibility Antigen
PBMC	Peripheral Mononuclear Cells
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
SNP	Single Nucleotide Polymorphism
TB	Tuberculosis

ABSTRACT

Sarcoidosis is a multisystemic disease of unknown etiology and unpredictable course, characterized by histopathological conglomerates of inflammatory cells defined as granulomas. These lesions however are non-pathognomonic, and in the absence of an identifiable etiologic agent, there are not specific diagnostic test for sarcoidosis. Despite the variable course of sarcoidosis, the lungs are affected in 90 percent of the cases. Approximately 25-30% of sarcoidosis patients progress to a complicated phenotype with progressive disease, leading to pulmonary fibrosis and organ dysfunction with increased mortality. These cases are in desperate need for biomarkers, conventional sarcoidosis biomarkers have proven to be insufficiently sensitive for implementation in routine clinical care.

In this dissertation, I focused on the use of alternate strategies for biomarkers development utilizing genomic base approaches based on high-throughput molecular assays to characterize genotype, gene expression, and epigenetics that define sarcoidosis subphenotypes. Our results demonstrated that the integration of expression quantitative loci (eQTL) studies increase the power of Genome-wide association studies (GWAS). We identified SNPs that were associated to complicated sarcoidosis in African Americans (AA) and in European Americans (EA), and then we validated these SNPs by Massarray. Furthermore, at the transcript level, we identified the Peripheral Mononuclear Cells (PBMCs) responses to TNF-α exposure, a cytokine involved in the initiation of granulomas and progression of fibrosis in sarcoidosis and identified a differential dysregulation in pathways unique to complicated sarcoidosis. At the transcriptome level, we profiled microdisected granulomas from lung and lymph nodes, and identified a hub of genes that were dysregulated only in sarcoidosis in both compartments. Additionally, we compared the genomic profile of these granulomas in Sarcoidosis vs Tuberculosis (TB) and Coccidioidomycosis. We corroborated that some genes previously suggested as potential sarcoidosis markers were also present in fungal or mycobacterium granulomas, pointing to a common mechanistic origin. We also demonstrated a strong similarity at the transcriptional level between Sarcoidosis and TB. The contribution of the epigenetic mechanisms to the clinical presentation of sarcoidosis was assessed through DNA methylation analysis, complicated sarcoidosis reveled a hypo-methylated pattern in genes within HLA complex while the miRNA analysis derived a molecular

signature consisting of 17 protein-coding genes, potentially regulated by 8 miRNAs dysregulated in complicated sarcoidosis.

Molecular Strategies to Distinguish Key Subphenotypes in Sarcoidosis

1. INTRODUCTION

1.1 Sarcoidosis definition and Historical Context.

Sarcoidosis is a multisystemic disease of unknown origin and unpredictable course, characterized by the formation of granulomatous lesions, especially in the lungs, liver, skin, and lymph nodes, with a heterogeneous set of clinical manifestations (1). The clinical course in Sarcoidosis is unpredictable and is influenced by granulomatous inflammation that most often affects the lungs and intrathoracic lymph nodes but can involve any organ.

From a historical perspective, J. Hutchinson initially described sarcoidosis as a skin disease in coal-warf workers in 1869. A Norwegian clinician, Caesar Boeck, coined the term “sarkoid” because the lesion resembled to a benign sarcoma. He later described multiple cases of the “benign military lupois” with involvement of the lungs, conjunctiva, bone lymph nodes, spleen and nasal mucous membrane (2). S. Löfgren, a Swedish physician, was the first link the triad of biopsy specimens, the erythema nodosum and bilateral hilar adenopathy as part of the acute presentations of sarcoidosis, known today as Löfgren’s syndrome. In was not until the 1940’s J.G. Scadding promoted the obtaining of sarcoidosis biopsy, which become the most valuable test for sarcoidosis diagnosis until the upcoming of bronchoscopy. With time, it became clear that sarcoidosis occurs throughout the world, affecting individuals of both genders and all races, although its prevalence varies widely across specific ethnic and racial groups(3).

1.2 Sarcoidosis Etiological Factors

The exact etiology of Sarcoidosis remains to be elucidated. Sarcoidosis granulomas are histopathological conglomerates of epithelioid cells, giant cells, CD4+ T cells in the center, and CD8 + T lymphocytes and B lymphocytes at their periphery(4). Activated alveolar macrophages produce cytokines, chemokines, and various inflammatory mediators, including TNF-ά, IL-6, IL-1, IL-12, IL-16, IL-8, MIP-1, RANTES (CCL5), MCP-1 (CCL3), and IP-10. With recruitment of additional monocytes and lymphocytes to the lung, macrophage differentiation into epithelioid and multinucleated giant cells, a granuloma is formed.

These lesions, however are non-pathognomonic, and in the absence of an identifiable etiologic agent, are not diagnostic of sarcoidosis or any other specific disease. Diseases to be

excluded include mycobacterial, fungal, and parasitic infections, chronic beryllium disease and other pneumoconiosis, hypersensitivity pneumonitis, and Wegener's granulomatosis. Pathologists have tried to characterize the granuloma origin based of the histological appearance, prototypical examples of necrotizing granulomas are seen with mycobacterial infections and non-necrotizing granulomas with sarcoidosis. However, broad differential diagnoses exist within each category(5), and it is not unusual that the granulomas of sarcoidosis may exhibit focal necrosis of minimal extent. It is has been established that in cases with granulomas that exhibit a greater degree of necrosis an infectious or other non-sarcoidosis related etiology should be strongly suspected(6). While granulomatous inflammation may or may not have an infectious etiology, independent molecular and immunologic investigations for multiple pathogens as potential etiologic agents including pathogenic mycobacteria are require in sarcoidosis (1) due to the histologic similarity between tuberculosis and sarcoidosis (7). Human metagenomics has provided insight into the complex composition of the microbiome of several body sites, and this information has allowed investigators to draw hypothesis about the relationship between specific microbiomes and health (*8-12*).

The Kveim–Siltzbach reaction, a local delayed granulomatous response to the intradermal inoculation of sarcoidosis splenic or lymph node tissue, indicates transmissibility of sarcoidosis, however the identification of specific antigens in the Kveim reagent that induce a sarcoid-like reaction has not been accomplished yet (*13*). Multiple researchers have attempted with limited success to identify **microbial lineages in sarcoidosis**. Propionibacteria and mycobacteria have been identified as targets of the sarcoidosis BAL-derived adaptive immune response (4-7). Mycobacteriophages have been isolated from sarcoidosis granulomas, as well as other human specimens with granulomatous inflammation. Despite the evidence of mycobacterial infection in some tissue biopsies and peripheral mononuclear cells (PBMCs), IgG anti MKat positive ESAT-6 and KatG peptides (*14, 15*), there is insufficient evidence to prove mycobacteria as the etiological agent in sarcoidosis. Propionibacterium acnes, a normal commensal of the skin flora is another candidate organism proposed in sarcoidosis, Propionibacterium DNA is present in lymph nodes of Japanese and European patients with sarcoidosis in higher proportions than DNA from mycobacteria spp (*16, 17*), this association reported among Japanese subjects, (*18, 19*) that has not been replicated in American

sarcoidosis subjects. Interestingly, some of these antigens cause granulomatous inflammation in animals (*20, 21*).

The Genomic Research Alpha 1 Antitrypsin deficiency and Sarcoidosis (GRADS) network, from which the mentor (site PI) and the candidate were participants, aimed to analyze the patterns in the lung microbiome are characteristic of sarcoidosis phenotypes and are reflected in changes in systemic inflammatory responses as measured by peripheral blood changes in gene transcription(*22*). Investigators have identified Th-1 immune responses to mycobacterial virulence factors, ESAT-6, katG, Ag85A, and superoxide dismutase A in sarcoidosis in peripheral blood mononuclear cells (PBMC), compared to controls as well as responses to these same virulence factors in bronchoalveolar lavage (BAL) (*23, 24*).

Other studies suggest that an active infection is not part of sarcoidosis pathobiology, but that sarcoidosis results from a host response to tissue antigens associated with an **aberrant innate response** involving serum amyloid A (SAA). (*25*) In this model, progressive accumulation of SAA aggregates within sarcoidosis granulomas and results in feed-forward amplification of local Th1 responses to tissue antigens, including those from remnant microbial antigens from prior infections associated with immune control ((*26*). Other potential pathogens such as viruses have been implicated but its correlation has been inconclusive. A Case Control Etiologic Study of Sarcoidosis (ACCESS) multicenter research group, a multicenter designed to determine the etiology of sarcoidosis also tried unsuccessfully to identify wall deficient forms of mycobacteria from blood in sarcoidosis patients, however, it provided evidence that certain environmental and occupational exposures modestly increase the risk for developing sarcoidosis (*27*). In this regard, **environmental and occupational studies** hypothesized that sarcoidosis may be caused by a dysregulated immune response to nanoparticles derived from common metals and minerals in the environment. To support this, recently, a study reported an increased risk of developing sarcoidosis in construction works exposed to silica(*28*). Also, of interest is the histopathological and immunological similarity between beryllium exposure and sarcoidosis. Following the collapse of the World Trade Center on September 2011, multiple reports associated the dust exposure and the development of sarcoidosis or "sarcoid-like" granulomatous pulmonary disease in first responders(*29, 30*) Recent geoepidemiological and occupational research have evaluated these combined effects, and identified certain risk occupations (*31*) and emphasized the

genetic individual composition or ethnicity-driven variations. This additive factor influences the frequency, epidemiology, clinical expression and outcome of sarcoidosis.

Many studies (*32-35*) indicate **genetic factors** as a contributory effect to the development of sarcoidosis involving an aberrant immune response triggered by an unidentified antigen in genetically susceptible individuals. To support the evidence of a dysfunctional genetic mechanism, there have been reported familiar clustering of sarcoidosis; studies have demonstrated an increased incidence of sarcoidosis in monozygotic twins (*36*). Furthermore, ethnicity heavily influences the clinical phenotype by modifying the age at diagnosis and the rates of thoracic and extra thoracic involvements. The highest incidence rates are uniformly reported in the United Stated for Black/African-American people, independently of the geographical location, with rates between 2- and 10-fold higher than those reported in White people living in the same geographical area (*37*). Significant racial and gender disparities exist in sarcoidosis development and prognosis in the U.S., with highest incidence in African American women(*38*). Furthermore, the interactions between changes in systemic inflammation and individual genomic composition that may influence the heterogeneity of disease phenotypes are not well understood

1.3 Sarcoidosis Clinical Presentation and Staging.

Despite significant progress in the understanding of the genetic predisposition and role of immunity, it is still a challenge to explain the clinical presentation of sarcoidosis. The diagnosis of sarcoidosis is further hindered by the lack of any reliable and specific diagnostic test, as there are no laboratory or imaging findings that enable the definitive diagnosis of sarcoidosis Usually the diagnosis is established when characteristic clinical-radiological features are supported by compatible histopathology of epithelioid cell granulomas, following exclusion of known causes of granulomatous inflammation (*3, 13*).

Non-specific constitutional symptoms such as fever, fatigue, malaise, and weight loss may occur in about one-third of patients with sarcoidosis. Most patients with pulmonary sarcoidosis recover spontaneously, whereas a significant number of patients exhibit progressive disease leading to varying degrees of pulmonary fibrosis. Standard clinical assessment, imaging, and pulmonary function tests (PFTs) do not allow prediction of disease course and response to therapy. However, a correlation with the mode of the onset and the extent of the disease has been observed. Most patients are middle-aged adults and frequently present with bilateral hilar

lymphadenopathy, pulmonary infiltration and ocular and skin lesions. Although not very common, Löfgren syndrome an acute well characterized presentation of sarcoidosis, is differentiated by the presence of erythema nodosum, an acute nodular panniculitis that affects the lower legs, BHL, and polyarthralgia. It is associated with spontaneous regression or usually heralds a self-limiting course, whereas an insidious onset, especially with multiple extrapulmonary lesions, may be followed by relentless, progressive fibrosis of the lungs and other organs(*13*). Ocular and skin lesions may become sight threatening or disfiguring thereby requiring aggressive treatment. Cardiac and neurologic involvement may cause morbidity and death.

Despite the variable course of sarcoidosis the lung is affected in 90% of cases but also may affect other organs (*39*). A classification based on intrathoracic radiographic changes, the Scadding Classification has been proposed (*13*) with the majority (50-60%) of sarcoidosis patients categorized at stage I with hilar and mediastinal lymphadenopathy, but without lung involvement **(Table 1)** . Stage II is defined as hilar lymphadenopathy with lung infiltrates. Frequently, at this stage patients will experience spontaneous remission whereas remission will occur in only 10-20% of individuals with Stage III which is characterized by the presence of diffuse interstitial infiltrates, without lymphadenopathy. Patients with Stage IV disease rarely have remissions after the development of pulmonary fibrosis and honeycombing (Stage IV). **(Table 1).**

TABLE 1. Chest Radiographic Stages of Sarcoidosis

Stage	Finding
0	Normal chest radiograph
I	Bilateral hilar lymphadenopathy (BHL)
II	BHL plus pulmonary infiltrations
III	Pulmonary infiltrations (without BHL)
IV	Pulmonary fibrosis

Greater than 50% of sarcoidosis patients will experience remission within 3 years after diagnosis, with over 66% of patients experiencing remission within 10 years(*40*). While certain disease phenotypes and chest radiograph stages portend a good prognosis, a large

proportion of patients would benefit from technological advances in biomarker development that would allow for identification of individuals at risk for increased severity of disease and the potential initiation of immunological therapies

1.4 Simple vs Complicated Sarcoidosis

In the absence of predictive biomarkers, clinical stratification of sarcoidosis remains a significant challenge. There are two major classifications of sarcoidosis severity based on disease evolution simple vs complicated. Patients with simple or uncomplicated sarcoidosis, typically asymptomatic stage I or II, tend to either experience spontaneous resolution or recover completely in up to 70- 90% of cases (*41*). In certain cases, the initial presentation of sarcoidosis is cutaneous and self-limited, exhibiting papules, ulcers, ichthyosis to erythema nodosum, a non-granulomatous panniculitis that usually occurs in acute sarcoidosis, and lupus pernio, a disfiguring red-to-purple plaques and nodules on the nose and cheeks. Epidemiological studies suggest that when cutaneous sarcoidosis associated to systemic symptoms require systemic work out to monitor closely the progression to multiorgan involvement (*42*). In general sarcoidosis morbidity tends to be more severe in black patients, whereas, white patients are more likely to have spontaneous resolution (*43*).

Approximately 25-30% of sarcoidosis patients exhibit a complicated sarcoidosis phenotype **(Table 2)** with progressive disease leading to varying degrees of pulmonary fibrosis and organ dysfunction (heart, neurological, etc.) with increased mortality(*44*). **Table 2 depicts** criteria our research group has utilized to categorize sarcoidosis subphenotypes. Our research group proposed the following criteria to stratify complicated sarcoidosis, it consolidates imaging (radiographic and CT scan) classification, pulmonary function test (Forced Vital Capacity) and involvement to vital organs(*45*) **Table 2.**

TABLE 2. Sarcoidosis clinically defined phenotype
Complicated Sarcoidosis
Lung involvement: Documented parenchymal lung disease by CT scan, radiographic stages III and IV and/or PFT with FVC <50%. Organ involvement: Cardiac or neurological involvement in addition to the lungs.
Uncomplicated Sarcoidosis
Skin or pulmonary involvement limited to the mediastinal lymphadenopathy documented by CT or radiographic stages I and II.

This table was originally published in Scientific Reports by T. Zhou and N. Casanova (2017) (*46*).

The clinical manifestations and disease course of sarcoidosis are both variable and largely unpredictable; from acute presentations with systemic symptoms that spontaneously remit without requiring immunosuppression, to relapsing and remitting disease that can lead to significant organ dysfunction over time, to chronic disease requiring long-term immunosuppression. To date there are not curative or effective treatments that reverse the course of the disease or that can modify the disease course. Therapeutic approaches such as corticosteroids, remain the first line of treatment, and are designed to suppress inflammation. While steroids can affect short-and intermediate-term improvement of sarcoidosis, particularly in early stage disease, there is little evidence these agents favorably alter long-term outcomes of patients with more advanced disease.

To avoid the side effects of long-term corticosteroids, new disease-modifiers such anti-TNF agents such as infliximab have been beneficial in limited number of patients, and are associated with considerable toxicities (*47, 48*). In addition, TNF-α blockers increase the risk of granulomatous and intracellular infections (*49, 50*).

1.5 Challenges in Mechanistic Investigation in Sarcoidosis: Absence of Preclinical Models.

In contrast to other interstitial lung diseases, such as pulmonary fibrosis, a well-accepted preclinical model of sarcoidosis is absent. Sarcoidosis' lack of a preclinical model has limited the study of potential etiological and histological features, the reproducibility and use of new

therapeutic targets and biomarkers. Therefore, the research to address these unmet needs is markedly hindered. Multiple attempts have derived in inconclusive models, as one example is, Crohn's disease, another granulomatous noninfectious disease considered as a possible model for basic research on sarcoidosis. However, there are no animal models that completely simulate the disease process of Crohn's disease(*13*). Another proposed model considers multiwall carbon nanotube (MWCNT)-based murine model for granulomatous inflammation researchers have used this model to compare inflammatory transcriptome signal similarities to BAL sarcoidosis in humans (*51*).

Recently, the Foundation of Sarcoidosis Research (FSR) sponsored projects for the development of novel experimental models for sarcoidosis. One of them based on the formation of granuloma-like structures in peripheral blood mononuclear (PBMCs) when these cells are challenged with immunogenic tuberculosis antigens, such as purified protein derivative (PPD). Other models include, murine model using Type I IFN activation to potentiate granuloma formation following administration of trehalose-6,6´-dimycolate (TDM), a lipid found in the cell wall of Mycobacterium tuberculosis. In addition to in silico multi-scale computational model of sarcoidosis using transcriptomics and flow cytometry and a model to study lung microenvironment in fibrotic pulmonary sarcoidosis employing culture scaffolds from decellularized lung tissues(*52*).

Further research is required to elucidate the destructive tissue inflammatory response that occurs with the granuloma formation, to identify which factors are responsible for accumulation of destructive T cells and monocytes in organs and how best to target therapies that will block the chronic and multisystemic response that occurs in sarcoidosis.

2. PRIOR BIOMARKER STUDIES IN SARCOIDOSIS.

Portions of this section were originally published in Clinics in Chest Med. Casanova N, Zhou T, Knox KS, Garcia JGN. Identifying Novel Biomarkers in Sarcoidosis Using Genome-Based Approaches. Clin Chest Med. 2015 Dec;36(4):621-630.

2.1 Traditional Blood Biomarkers in Sarcoidosis. The need to develop useful biomarkers in the diagnosis and prognosis of subjects with sarcoidosis has long been recognized as sarcoidosis is a diagnosis of exclusion and may mimic multiple other rheumatologic illnesses(*53, 54*). Furthermore, a significant percentage of patients with sarcoidosis (~ a third) develop complications with granulomatous involvement of vital organs

with progressive disease. Thus, monitoring subclinical disease activity and likelihood of progression or remission in a longitudinal fashion remains an important but challenging goal. Recently, we review traditional biomarkers in sarcoidosis as well as biomarkers emerging from technology-driven strategies (*55*). Despite limitations, the use of biomarkers to support diagnosis and predict disease activity remains a focal point in routine clinical care. Multiple methodologies have been applied to detect biomarkers in serum, lung tissue, bronchoalveolar lavage fluid (BALF) and exhaled breath condensate (EBC) using enzyme-linked immunosorbent assays (ELISA), proteomic analysis and mass spectrometry (*56*). Traditionally, clinical biomarkers measured in sarcoidosis were limited to soluble factors measured in blood, BAL or cerebral spinal fluid. Data remains inconsistent regarding the validity of EBC biomarkers (*57*). As technology improves, biomarker “panels” generated from array data will continue to emerge.

Several potential sarcoidosis biomarkers have been associated with **monocyte-macrophage activation** including **Angiotensin-Converting Enzyme (ACE),** the most commonly used biomarker in sarcoidosis. ACE is derived from activated macrophages in granulomatous pulmonary remodeling (*58*), and is useful in supporting a diagnosis and monitoring disease activity in some patients (*58, 59*). ACE levels are increased in approximately two-thirds of patients with sarcoidosis with elevated levels reported in neurosarcoidosis (*60*). However, elevated ACE levels are not specific for sarcoidosis, and are observed in granulomatous diseases such as tuberculosis, fungal infections, and Gaucher disease (*58, 61, 62*). Moreover, there is no evidence that ACE levels significantly differ between active and inactive sarcoidosis (*59*) and do not reliably correlate with the severity of disease (*63*).Serum concentrations at diagnosis were noted to be significantly lower in acute sarcoidosis and in Löfgren's syndrome (*64*). In addition, ACE concentration and activity is influenced by genetic polymorphisms with enzymatic activity significantly higher in individuals with the DD genotype than in individuals with the II genotype (*62, 65, 66*). The utility of ACE as a biomarker in diagnosis of sarcoidosis remains limited, although future studies utilizing conformational fingerprinting of ACE may yield better specificity [73]. While ACE is the prototypic sarcoid biomarker, similar conclusions are made regarding other soluble biomarkers outlined below.

Serum amyloid A (SAA) is an acute phase protein produced in the liver and upregulated by monocyte and macrophage-derived cytokines (*67*). SAA levels were significantly higher in sarcoidosis patients than in healthy controls (*56, 68, 69*) and significantly higher in sarcoidosis

patients with active disease (*59*). However, like ACE, this biomarker suffers from low specificity although SAA levels may be more useful during follow-up as SAA levels are less sensitive to immunosuppressive drugs, such as corticosteroids. Cytokines such as **tumor necrosis factor (TNF-α, TNF-β)** play a major role in granuloma formation and are released in greater quantities from alveolar macrophages obtained from sarcoidosis patients(*70*).

Lysozyme is an enzyme produced by macrophages to degrade bacteria and is elevated in numerous inflammatory conditions, including sarcoidosis. Lysozyme historically has been associated with extrapulmonary sarcoidosis, particularly uveitis (*71*) and in a recent study was more sensitive than ACE(*72*). **Serum chitotriosidase** concentrations were significantly higher in advanced (stage 3) sarcoidosis compared to healthy controls, directly correlated with ACE levels and were highest in those with persistent disease on therapy [26].

Among the potential sarcoidosis biomarker associated with **lymphocyte origin is Interleukin-2 receptor (sIL-2R) whose** levels shed from lymphocytes are increased in active sarcoidosis and, similar to ACE levels, may predict response to therapy (*62*). Elevated levels have been correlated with parenchymal infiltration and lung function (*73, 74*). Similarly, persistently elevated sIL-2R may suggest extrapulmonary manifestations of sarcoidosis (*59, 70, 75*). **IL-17**, an interleukin important in mucosal immunity and autoimmunity, is increased in patients with ocular sarcoidosis (*76*) and in the BAL fluid of patients with Lofgren's syndrome. Biomarkers associated with **fibrosis and the extracellular matrix** include **Tenascin-C,** an extracellular matrix molecule expressed during wound healing in various tissues, is increased in granulomatous sarcoidosis [29] and in BALF in patients with parenchymal infiltration on chest radiographs [30]. T**ransforming growth factor TGF-β1** is associated with tissue healing and recruiting of fibroblasts and myofibroblasts to the matrix, with over expression of TGF-β1 fibrosis can occur. Significantly higher levels of TGF-β1 and ACE were reported in sarcoidosis patients (*58*).

Measurement of biomarkers to diagnose and predict remitting or progressive disease remains promissory and relevant in the management of sarcoidosis. Despite a rich history and intense study, however, no single soluble biomarker has proven to be sufficiently sensitive and specific to be recommended for widespread clinical use.

TABLE 3. CONVENTIONAL SARCOIDOSIS BIOMARKERS AND THEIR CLINICAL ASSOCIATION.

Biomarker	Origin and clinical association	Reference
ACE	Monocyte-macrophage origin. Acute stage, levels influenced by polymorphisms.	A. Salazar et al (2010), A.D. Vorselaars et al (2015); B. Baudin et al (2005);A. Pietinalho et al (1999)
sIL-2R	Lymphocyte associated. Disease severity, extra-pulmonary organ involvement.	A.D. Vorselaars (2015); J. C. Grutters et a (2003)
SAA	Monocyte-macrophage origin. Higher level in tissue and serum in sarcoidosis.	A. Salazar et al (2010),
Alpha 1-antitrypsin (BALF)	Cytokine associated. Down-regulated only in patients without LS. Associated with spontaneous resolution.	R.C.Young et al (1973)
Protocadherin-2 precursor (BALF)	Cell adhesion. Up-regulated in sarcoidosis across all studied phenotypes.	E. Kiegova (2006)
Chitotriosidase	Monocyte-macrophage origin. Disease progression.	E. Bargagli et al (2008)
Tenascin-C (BALF)	Fibrosis and ECM associated. Levels correlated with infiltrates on chest radiographs in sarcoidosis.	L. E. Crowley et al (2011)
IL-17RC	Lymphocyte associated. Elevated expression in retinal tissues.	W. Wu et al et al (2014)
TGF-β1	Fibrosis and ECM related. Associated with pulmonary fibrosis.	H. Ahmadzai et al (2013)

Abbreviations: ACE, angiotensin converting enzyme; sIL-2R, soluble Interleukin-2 receptor; SAA, Serum amyloid; LS, Lofgren's syndrome; BALF, bronchioalveolar lavage fluid; ECM, extracellular matrix **Adapted from Casanova et al, *Clin Chest Med* 36, 621-630 (2015).**

2.2 Alternate Strategies for Biomarker Development in Sarcoidosis

Genomic-based biomarkers are different types of biomarkers originating from various "omics" technologies that combines genomics, proteomics and metabolomics(*77*). Recent establishment of high-throughput molecular assay technologies, such as single-nucleotide

polymorphism (SNP) arrays, gene expression microarrays, and protein arrays, has allowed discovery of potential new disease biomarkers. These high-throughput approaches have been applied to diseases of unknown cause like sarcoidosis to help understand disease pathogenesis (*78-80*) including identification of disease-associated candidate genes and dysregulated pathways with diagnostic and prognostic applications (*81-84*). Current imaging and clinical tools fail to identify individuals at risk for complicated sarcoidosis.

Identification of punctual variations in genomic DNA, has identified single nucleotide polymorphisms (SNPs) that are disease (sarcoidosis vs controls) or trait specific (complicated vs uncomplicated) and that may be associated to an increased racial disparity or poor prognosis.

Similarly, transcriptomic markers based on RNA expression profiles in peripheral blood are the most frequently studied disease-related transcriptomes. Genomic expression profiling provides the opportunity to identify those genetic variants that contribute to complicated sarcoidosis. They are informative in multiple disease classes including cancer(*85, 86*), autoimmune disease(*87, 88*), psychiatric disease(*89*), cerebro-vascular disease(*90*), heart failure (*91*), transplant rejection after cardiac transplantation(*92*) and sarcoidosis (*93*).

Microarray analysis of peripheral blood mononuclear gene expression (PBMC) in two independent cohorts with Idiopathic Pulmonary Fibrosis (IPF) identified 2595 genes associated with outcome distinguished 2 major clusters of IPF patients with a significant difference in survival. Zhou et al., reported a 31-gene signature comprised of T cell signaling pathway genes associated with sarcoidosis using genome-wide peripheral blood gene expression 20-gene sarcoidosis biomarker signature distinguished sarcoidosis (n = 39) from healthy controls and served as a molecular signature for complicated sarcoidosis (*94*). Interestingly, the CTLA4 pathway was most significantly associated with survival in the derivation and replication cohorts and SmartChip qRT-PCR confirmed that a decrease in PBMC expression of CD28, ITK, ICOS or LCK was predictive of decreased survival in patients with IPF compared to those expressing high levels. To determine whether we could find a similar signature in sarcoidosis we downloaded the published dataset (*93*) (GSE1907) and demonstrated that T-cell genes distinguished sarcoidosis peripheral blood from control and that this difference faded 6 months later. Our analysis demonstrated an increase in CD4, CXCR4 and CCR3, and a decrease in expression of CTLA4 markers

(CD28, LCK, ICOS) in the blood of patients with active disease. Interestingly – while CCR3 and CCXCR4 were reduced 6 months later, the decrease in CTLA4 genes persisted.

3. PRIOR GENETIC STUDIES IN SARCOIDOSIS SUSCEPTIBILITY AND SEVERITY

3.1 **Family Genetic Studies in Sarcoidosis (ACCESS).** Today's overwhelming evidence indicates that sarcoidosis susceptibility and severity is influenced by genetic variations in a population-specific manner (*33, 95*). Previous familial clustering cases in African American patients and Swedish strongly support a genetic basis for the heterogeneous phenotype with an increased incidence in monozygotic twins (*36, 96*). One of the first studies addressing this association was ACCESS (A Case-Control Etiology Study of Sarcoidosis) the major goal of ACCESS was to generate hypotheses about the etiology of sarcoidosis. Here a significant elevated risk of sarcoidosis was observed among first-and second-degree relatives compared with controls in both African Americans and whites (36). Despite the familial aggregation and association to genetic polymorphisms another family-based study in African Americans revealed no association between sarcoidosis susceptibility and polymorphisms in the angiotensin converting enzyme, vitamin D receptor, and tumor necrosis factor-alpha genes (*97*). HLA class II was among the genetic sarcoidosis- susceptible regions evaluated in ACCESS, HLA DRB1 alleles were significantly differently between cases and controls, DPB1 was associated with sarcoidosis in African Americans suggesting that some of the difference in susceptibility to sarcoidosis between whites and African Americans depends on the differential expression of the HLA alleles (*98*). Another study derived from ACCESS identified the polymorphisms of the IL-1β, tumor necrosis factor (TNF)-α, and the immunoglobulin genes Km and Gm in AA, indicating how race heavily influences the clinical phenotype by modifying the age at diagnosis and the rates of thoracic and extra thoracic involvements(*99*).

3.2 **Genome-Wide Association Studies.** Genome wide association studies (GWAS) have been utilized to study the genomic predisposition in sarcoidosis and have led to the identification genomic variants associated with sarcoidosis susceptibility or specific traits. The first genome wide association study (GWAS) in sarcoidosis was conducted by Hofmann et al (*100*), using Affymetrix Genome-Wide Human SNP Array 5.0 in a

predominant German population. This study identified a nonsynonymous single nucleotide polymorphism (SNP), rs1049550, residing in the first of four annexin core domains within the gene ANXA11 (annexin A11), which was associated with sarcoidosis susceptibility (*100*), a finding confirmed in an independent case-control study in Czech patients (*101*) and in African and European Americans (*32, 102*). In a separate GWAS conducted by Hofmann et al, a nonsynonymous SNP, rs1040461 in gene RAB23 (RAB23, member RAS oncogene family), was associated with sarcoidosis in German population, this association was replicated in African Americans but not in European Americans (32). More recently, a sarcoidosis GWAS in both African and European Americans, using Illumina HumanOmni1-Quad array (32), identified a novel sarcoidosis-associated SNP, rs715299, within the intro region of the gene NOTCH4 (notch 4) in African Americans but not in European Americans (32). The locus of this SNP is close to several MHC Class II genes known to be associated with sarcoidosis (*103, 104*) and within the region of high linkage disequilibrium (*105*). Using stepwise conditional association analyses, it was confirmed that the observed signal within NOTCH4 is independent of the SNPs within the MHC Class II genes (32). Recently, the use of Whole Exome Sequencing (WES), identified 37 genes that could be putatively linked to a pediatric form of sarcoidosis in three familial cases, functional analysis of these genes were implicated with IFN-γ and anti-inflammatory cytokine IL-10, leukocyte proliferation, bacterial defense, and vesicle-mediated transport (*106*).

3.3 Epigenetic genome-wide studies (miRNA and DNA Methylation).

Epigenetic regulation controls the expression of genome specific regions allowing gene transcription to be enhanced or repressed through a heritable mechanism. Epigenetic processes, including DNA methylation and histone modification, influence gene expression at the level of transcription. Various DNA methyl-binding proteins (MBPs) exist (MBD1, MBD2, MBD3, MBD4 and MECP2), they share a related DNA-binding domain and react to the methylation status of the DNA in promoter regions influencing deacetylation, another epigenetic mechanism. Specific MBPs, such as MBD1 are transcriptional repressors that participate in the recruitment of co-repressors and histone deacetylases, others, like MECP2, probably the best characterized of the MBD family, binds methylated CpG via its MBD domain and exerts repressive effects on transcription over distances of several hundred base pairs (*107*). **DNA methylation studies in sarcoidosis** are limited. One study aimed to study

SHOX2 methylation pattern from bronchial aspirates in 523 suspected lung cancer patients, which putatively allowed a distinction of malignancy from infectious disease, scleroderma a and sarcoidosis. Hypermethylation of SHOX2 was only present in lung cancer tissue (*108*). Other study recently published, analyzed the DNA methylation changes in chronic beryllium disease and sarcoidosis, in progressive versus remitting sarcoidosis, identified 15,215 CpGs sites with significant methylation changes, but only 801 of them have greater than 5% methylation change, demonstrating that DNA methylation may be of utility in tracking progression changes but may suffer from statistical power issues (*109*).

Epigenetic regulation also includes RNA-based mechanisms, through the regulation of gene expression. The majority of RNA transcripts, do not code for protein, microRNAs (miRNAs) are part of these non-coding RNAs they act as posttranscriptional regulators. MiRNAs base-pair with target mRNAs to inhibit translation or direct mRNA degradation via the RISC complex. The first study profiling miRNA in sarcoidosis was conducted in 2012, using genome-wide microarray in lung tissue and PBMCs, identifying a distinct pattern of differentially expressed-miRNAs (DE-miRNAs) in both tissues, in both cases targeting transforming growth factor (TGFβ)-regulated pathways (*110*). Recently another study derived from ACCESS specimens studied the expression of miRNAs from peripheral blood mononuclear cells in sarcoidosis vs controls, identifying eight DE-miRNAs in sarcoidosis and two miRNAs (hsa-miR-150 and hsa-miR-342) significantly associated with reduced lymphocytes and airflow limitations (*111*). Another study profiled the miRNA of T regulator cells in pulmonary sarcoidosis, identifying miR-155 and miR-223 as disease related miRNAs associated to alterations in TLR-2 and NF-κB signaling pathways (*112*). Recently we profiled microRNA and protein-coding gene expression in peripheral blood mononuclear cells from 70 subjects with Idiopathic Pulmonary Fibrosis (IPF). We linked the microRNA/gene expression level with the quantitative phenotypic variation in IPF, including diffusing capacity of the lung for carbon monoxide and the forced vital capacity percent predicted. We intersected these predicted mRNA targets against the microarray-observed 712 differentially expressed protein-coding genes revealing a total of 22 miRNA-mRNA pair that consisted of 21 unique protein-coding genes yielding a 21-gene signature Each signature performed well in a validation cohort comprised of IPF patients aggregated from distinct patient populations recruited from different sites

3.4 Expression Quantitative Loci (eQTL) studies. Genome-wide association studies (GWAS), have demonstrated that most genetic variants reside in non-coding regions and therefore involved in gene regulation, complicating the understanding of the role that such variants play in the disease pathogenesis. The study of these variants that influence the gene expression are called expression quantitative trail loci (eQTLs) that explains a fraction of the genetic variance of a gene expression phenotype(*113*). These regulatory variants are identified as **cis or trans** depending on the physical distance they are located from the gene they regulate, 1 megabase or > 5 megabases, respectively. **Figure 1.**

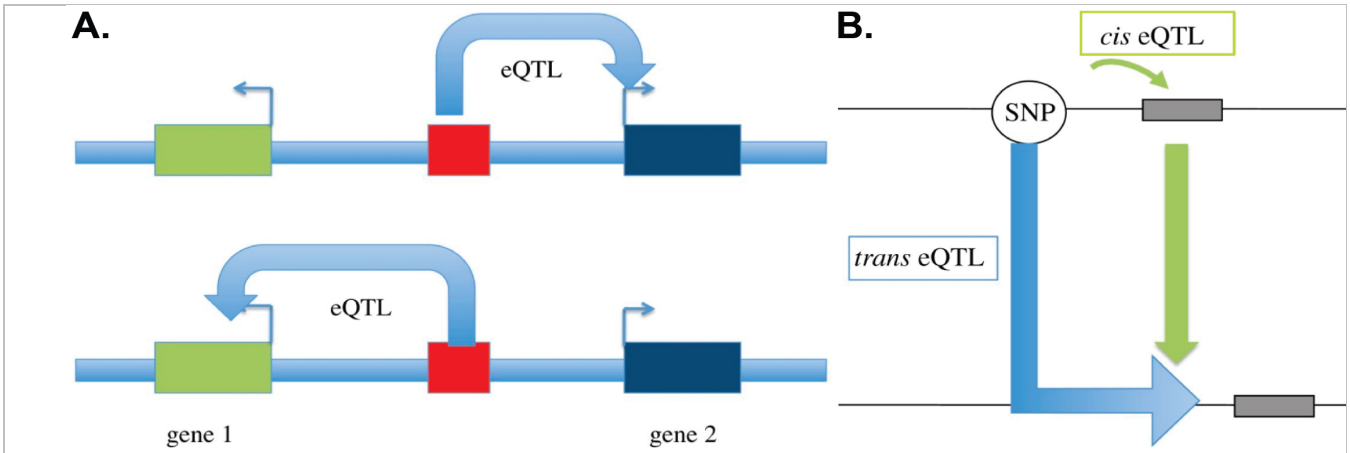


Figure 1. eQTL regulation of gene expression. **A)** Regulatory variants can be eQTL for gene 2 in (a) tissue and for gene 1 (b) in a different tissue. **B)** Network relationships based on cis and trans eQTLs. Reproduced with permission (Dermitzakis E., Nica et al Philos Trans R Soc Lond B Biol Sci. 2013 Jun 19; 368(1620).

Identification of gene networks that are perturbed by susceptibility loci and that, in turn lead to disease beyond the susceptibility genes provides a more integral approach to study complex disease traits (*114*). Consequently, new approaches to increase the power of GWAS and to understand the functional effects of most GWAS-implicated variants involve the integration of expression quantitative loci (eQTL) studies. To provide insight into how genes are expressed differently across multiple human tissues and how they are regulated, the NIH established the Genotype-Tissue Expression (GTEx) program, the eQTLs identified in post mortem donors, are accessible to the scientific community through the GTEx database. This publicly available database has generated data from 54 non-

diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq. (Casanova et al. 2020 manuscript in press) (*115*).

STUDIES PERFORMED BY THE CANDIDATE

4. CURRENT GENE EXPRESSION STRATEGIES IN SARCOIDOSIS PHENOTYPING

4.1 Genome-wide gene expression in PBMCs

We conducted an exploration to identify genetic variants that confer susceptibility to developing sarcoidosis or the development of complicated sarcoidosis. We analyzed the genetic risk predisposing to complicated versus uncomplicated sarcoidosis in two different racial groups, African Americans (AA) and European Americans (EA) (n=274 and n=239, respectively) in an Affymetrix 6.0 GWAS platform. After quality control assessment our data set included 209 AA and 193 EA. We identified the top GWAS race-specific, risk variants/genes potentially involved in sarcoid pathogenesis and stratified them by race and complicated status then we created a list of the top SNPs associated to JAK/STAT pathway (**Table 4).** However, we noticed our GWAS results were underpowered to identify true significant variants.

Table 4. Race-specific, risk variants associated to JAK/STAT pathways

Top SNP in AA						Top SNP EA					
SNP	Chr	Location	P-value	Gene_loc	Gene	SNP	Chr	Location	P-value	Gene_loc	Gene
rs11127238	2	30066492	0.0001124	intron	ALK	rs1872568	11	69418138	0.0001356	upstream	CCND1
rs12650052	4	87138738	0.0001979	intron	MAPK10	rs9382081	6	52094091	0.0001924	downstream	IL17A
rs10170924	2	213164561	0.0005845	intron	ERBB4	rs2069835	7	22767871	0.0006381	intron	IL6
rs4142620	6	92173726	0.0006123	upstream	MIR4643	rs4845144	1	207016648	0.0007858	downstream	IL19
rs8002655	13	83419471	0.000752	downstream	SPRY2	rs2465665	2	201478264	0.0008094	intron	AOX1
rs13005140	2	212362352	0.0007955	intron	ERBB4	rs6742109	2	201466732	0.0008094	intron	AOX1
rs311638	8	90603149	0.000803	upstream	RIPK2	rs939269	8	27181456	0.0009378	intron	PTK2B
rs606565	8	90615793	0.0008083	upstream	RIPK2	rs16848890	2	213461841	0.001082	intron	ERBB4
rs12620169	2	212362328	0.001047	intron	ERBB4	rs1316370	8	27181065	0.001204	intron	PTK2B
rs9546142	13	83048094	0.001059	upstream	SPRY2	rs11689233	2	213317794	0.001402	intron	ERBB4

Therefore, we incorporated eQTL data to our analysis. We intersected the 906,600 probes from Affymetrix 6.0 GWAS corresponding to 541,309 rs id dbSNPs with eQTLS from GTEx Analysis V7, that included the following tissues; Lung, brain, hearth, whole blood, skin, spleen, artery and pancreas. Using a logistic model, we run four different

comparisons according to race (EA vs AA) and complicated status (complicated vs non-complicated) and identified the top 50 SNPs from each tissue independently. Each SNP pair was evaluated for Linkage Disequilibrium (LD) or the non-random association of marker alleles and can arise from marker proximity or from selection bias. 621 unique SNPs were identified in total, 285 where in linkage equilibrium while 336 where in LD. According to GTEx annotations, the 621 SNPs affect 730 genes. Next, we conducted pathway enrichment using these 730 genes through Consensus Pathways (http://consensuspathdb.org) and identified an important enrichment with genes part of the human leukocyte antigen (HLA) system or complex **Table 5**, this system encodes the major histocompatibility complex (MHC) proteins in humans. These cell-surface proteins are responsible for the regulation of the immune system in humans located in chromosome 6p21.

Table 5. Top pathways derived from 730 eQTL derived SNPs.

pathway name	set size	candidates contained	p-value	q-value	pathway source
Allograft Rejection	80	13 (16.2%)	2.08e-07	0.000211	Wikipathways
Allograft rejection - Homo sapiens (human)	38	9 (24.3%)	4.52e-07	0.000229	KEGG
Graft-versus-host disease - Homo sapiens (human)	41	9 (22.0%)	1.16e-06	0.000376	KEGG
Type I diabetes mellitus - Homo sapiens (human)	43	9 (20.9%)	1.78e-06	0.000376	KEGG
Staphylococcus aureus infection - Homo sapiens (human)	56	10 (18.2%)	1.86e-06	0.000376	KEGG
Translocation of ZAP-70 to Immunological synapse	24	7 (29.2%)	2.35e-06	0.000397	Reactome
Viral myocarditis - Homo sapiens (human)	59	10 (17.2%)	3.08e-06	0.000446	KEGG
MHC class II antigen presentation	60	10 (16.7%)	4.24e-06	0.000537	Reactome
Phosphorylation of CD3 and TCR zeta chains	27	7 (25.9%)	5.62e-06	0.000634	Reactome
PD-1 signaling	28	7 (25.0%)	7.32e-06	0.000743	Reactome
Autoimmune thyroid disease - Homo sapiens (human)	53	9 (17.3%)	9.35e-06	0.000862	KEGG
Asthma - Homo sapiens (human)	31	7 (23.3%)	1.2e-05	0.00102	KEGG
Antigen processing and presentation - Homo sapiens (human)	77	10 (13.0%)	4.1e-05	0.0032	KEGG
Retrograde transport at the Trans-Golgi-Network	49	8 (16.3%)	4.59e-05	0.00332	Reactome
Generation of second messenger molecules	38	7 (18.4%)	6.18e-05	0.00418	Reactome
Phagosome - Homo sapiens (human)	154	14 (9.2%)	7.06e-05	0.00448	KEGG
Metabolism of nucleotides	100	11 (11.0%)	8.12e-05	0.00484	Reactome
Rheumatoid arthritis - Homo sapiens (human)	90	10 (11.2%)	0.000143	0.00804	KEGG
Systemic lupus erythematosus - Homo sapiens (human)	133	12 (9.1%)	0.000245	0.0127	KEGG

4.2 Exvivo Genome-wide gene expression in PBMCs

Recent advances in genome-wide expression profiling techniques, such as gene expression microarray and RNA sequencing, provide opportunities to discover novel disease mechanisms. These high-throughput approaches have been applied to diseases of unknown cause to help understand disease pathogenesis (*78-80*) including identification of disease-associated candidate genes with diagnostic and prognostic applications (*81-84*).

Prior studies have shown that transcription patterns measured in blood may be an indirect and powerful tool to study inflammatory signals in the diseased organs of sarcoidosis patients(*116*). Furthermore, mRNA species are measurable in both plasma and serum and could represent another area of investigation into disease biomarkers. Taken together, blood components such as PBMCs are easily obtained and processed and serve as non-invasive ways to learn tremendous insights about the biology of sarcoidosis and other diseases with systemic manifestations.

Idiopathic Pulmonary Fibrosis (IPF) and Sarcoidosis are both interstitial lung diseases that share some similarities both are of unknown etiology; sarcoidosis pulmonary involvement at the end stages also develops pulmonary fibrosis to IPF, both have poor prognosis leaving lung transplantation as the only alternative. Recently, we evaluated the **PBMC expression profile in IPF after the CO exposure** during a 12-week treatment. These studies identified a clear CO signature dominated by significant dysregulation of genes in the oxidative phosphorylation pathway. Significant differential expression was observed when comparing transcriptome from week 12 to week 24 with gene dysregulation response that was abolished after the treatment termination. These findings suggest this signature may serve as a potential genomic biomarker for CO exposure and perhaps for titration of dosage to allow precision testing of therapies in future low dose CO therapeutic studies in IPF (*117*).

The **transcriptome for PBMCs in sarcoidosis** has been studied previously using microarray technologies and PCR. One of the earliest studies conducted in sarcoidosis measured the presence of IFN-gamma transcripts, concluding that mRNA for IFN-γ is significantly associated with stage I disease(*118*). Preliminary and published data by the candidate‘s mentor investigative team identified via genome expression, using Affymetrix Human Exon 1.0 ST Array, a differentially expressed genes that allowed defining uncomplicated from complicated sarcoidosis. The initial analysis set comprised of 1233 genes

differentially expressed between AA or EA complicated sarcoidosis cases vs. healthy controls was utilized. This analysis generated a 20-gene transcriptomic signature in peripheral blood mononuclear cells (PBMCs) that distinguishes sarcoidosis from not only healthy individuals as well as uncomplicated sarcoidosis from complicated patients with progressive lung disease, cardiac or neurologic sarcoidosis involvement (*94*). Two genes within the unbiased 20-gene signature, *HBEGF* (heparin-binding EGF-like growth factor) and *SAP30* (Sin3A-associated protein, 30kDa), were strongly up-regulated in complicated sarcoidosis whereas the remaining 18 signature genes were down-regulated in complicated sarcoidosis **(Figure 2).** The non-targeted 20-gene signature distinguished all sarcoidosis patients from healthy controls with an accuracy of 86.0% (sensitivity=88.2% and specificity=83.3%).

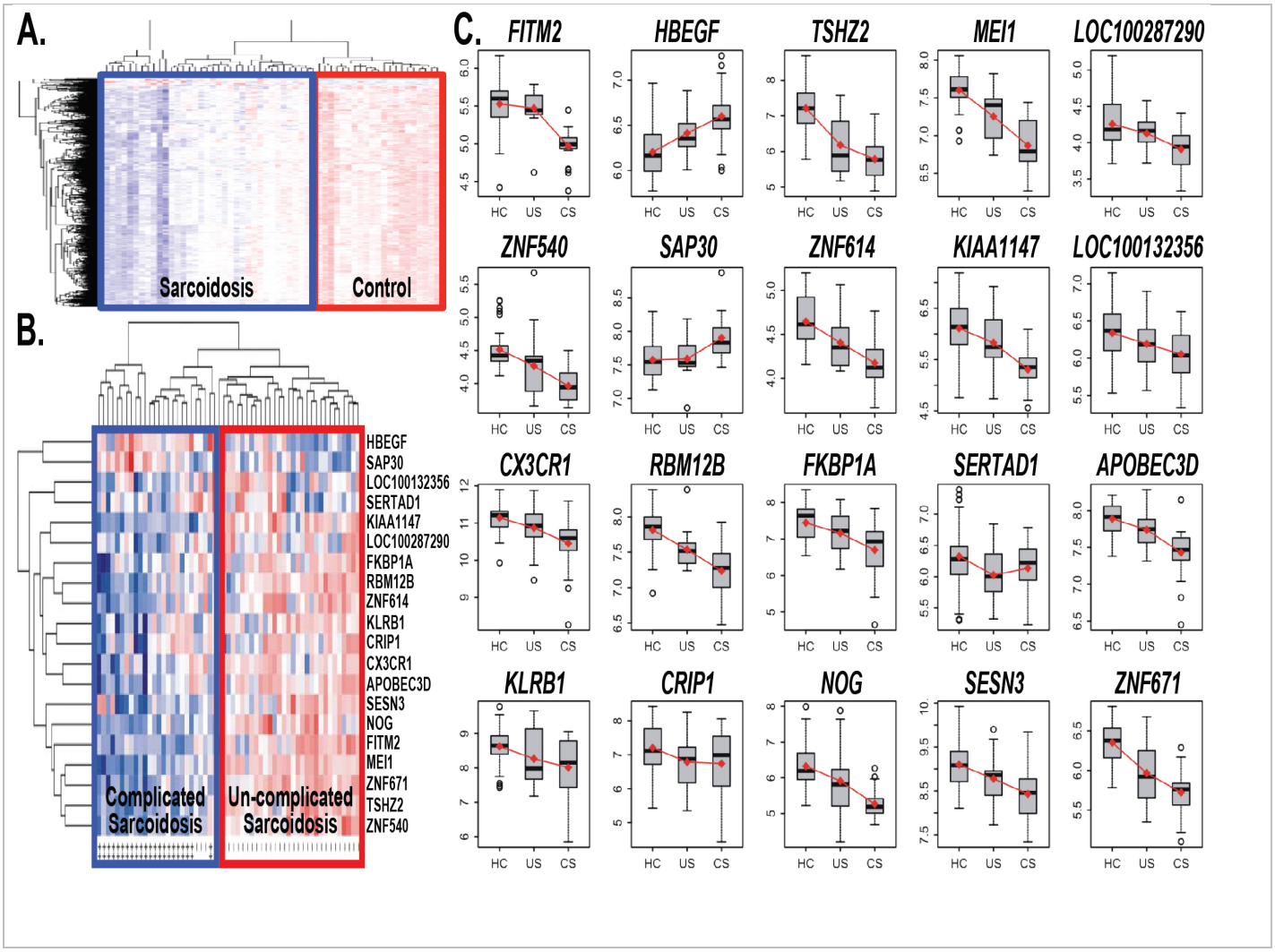


Figure 2. Hierarchical clustering discriminates patients with Sarcoidosis and healthy controls and differentiated complicated and non-complicated Sarcoidosis. Panel A. DGE Cluster 1: sarcoidosis (n=31), cluster 2: healthy controls (n=18). Blue represents low gene expression, red is high gene expression. **Panel B**. Heatmap of patients with complicated sarcoidosis and uncomplicated sarcoidosis. **Panel C**. Boxplot of expression of the 20 signature genes. The dark grey points and lines indicate the geometric mean of expression in each category. HC: healthy controls; US: uncomplicated sarcoidosis; and CS: complicated sarcoidosis. Y-axis: log2-transformed expression values. Reproduced with permission (Zhou T. et al, PLoS doi: 10.1371/journal.pone.0044818).

We evaluated the performance of our gene signatures in two different independent sarcoidosis blood gene expression datasets. One dataset (GEO - GSE19314) from University of California, San Francisco (UCSF) this validated peripheral blood molecular gene signature appears to be a valuable biomarker in identifying cases with sarcoidosis and predicting risk for complicated sarcoidosis. This gene signature was superior in prediction accuracy in each of the AA and EA populations when compared to a second signature (UCSF) comprised of genes within the T cell receptor–innate immunity pathway that includes genes previously associated with sarcoidosis. It also distinguished sarcoidosis patients from idiopathic pulmonary fibrosis (IPF) (*94*).

Gene Set Enrichment Analysis of lung samples with fibrotic pulmonary sarcoidosis showed that the gene up-regulated in lung samples obtained from patients with progressive-fibrotic sarcoidosis comprised predominantly the genes involved in host defense and immune responses. Genes overexpressed in patients exhibiting progressive-fibrotic sarcoidosis are also significantly enriched for genes up-regulated in hypersensitivity pneumonitis, another granulomatous lung diseases but different than those in idiopathic pulmonary fibrosis (*119*). Recently, we published a PBMC microarray (Affymetrix GeneChip Human Exon 1.0) derived gene signature in **sarcoidosis patients with pulmonary hypertension**, a common comorbidity associated to progressive lung involvement. The discriminative accuracy of this 18-gene signature was 100% in separating sarcoidosis patients with PH from those without it. The total number of DEG between sarcoidosis patient with and without Ph were 275. Among the 18 genes, *HIST1H4C* (encoding histone cluster 1, H4c), *CACHD1* (encoding cache domain containing 1), *STOX1* (encoding storkhead box 1), and *NRCAM* (encoding neuronal cell adhesion molecule) were strongly downregulated in

PH patients, whereas the remaining 14 genes were upregulated in PH. In summary, despite the small number of patients included in the study, we have identified a molecular gene signature as a potential novel molecular biomarker in the diagnosis of PH in sarcoidosis (*120*).

The exact mechanism for the modulation of inflammatory-related gene expression remains unknown. **Tumor necrosis factor** (TNF-α , TNF-β) is synthesized by activated macrophages and T cells and it has a key role in macrophage activation, phagosome activation, differentiation of monocytes into macrophages, recruitment of neutrophils and macrophages, granuloma formation, and maintenance of granuloma integrity (*1, 121*). Two TNFA polymorphisms have been identified in cardiac sarcoidosis, no significant difference was noted between the patients with cardiac involvement and those without, concerning the -1.031T/C and -238G/A (*122*). Similarly, *TNFA2* (*TNF_307 A*) polymorphism was associated significantly with Lofgren's syndrome and cardiac sarcoidosis (*123*).

Increasing evidence supports the influence of cytokines such as tumor necrosis factor alpha (TNF-α) in the initiation, maintenance of granulomas and the progression of fibrosis. Furthermore, the efficacy of biologic TNF antagonists (etanercept, infliximab, and adalimumab) as therapeutic strategies has been validated in refractory sarcoidosis. However, despite its effectiveness, paradoxical sarcoidosis-like reactions have been reported during anti-TNF therapy and limit these approach in particular at-risk patients(*124*). Recently, the candidate studied the transcriptional effect of TNF-α in an Ex vivo model on PBMCs for sarcoidosis patients. 282 TNF-α responsive differentially expressed genes (DEGs) were identified in uncomplicated sarcoidosis and 550 TNF-α responsive genes in complicated sarcoidosis. Sixty genes were common to both uncomplicated and complicated sarcoidosis groups. **Figure 3**.

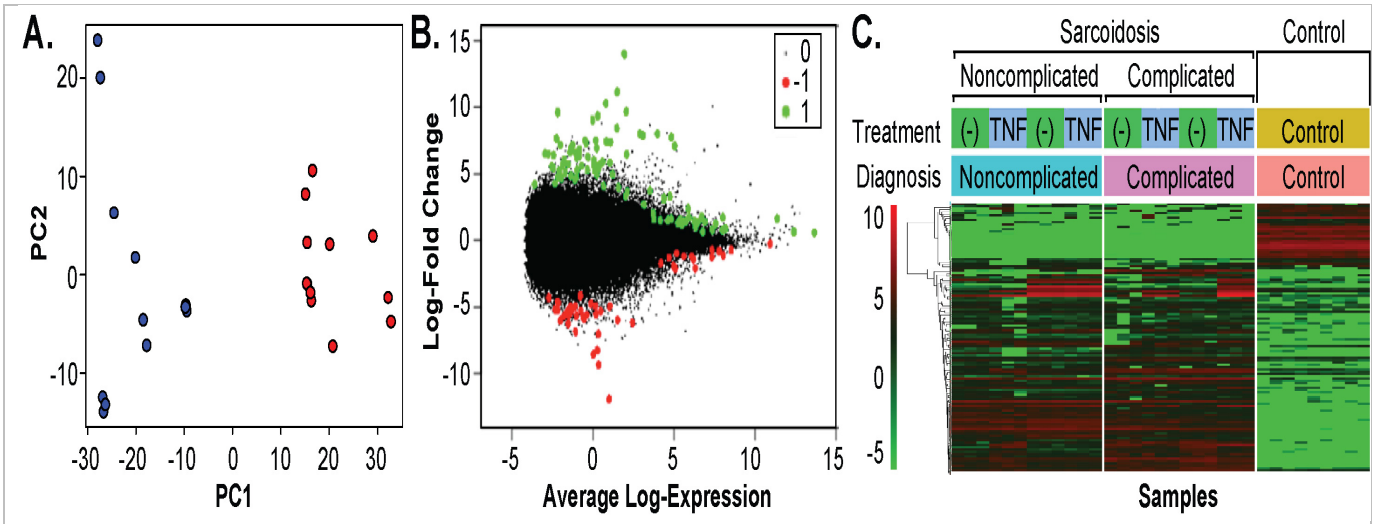


Figure 3. TNF-α responsive differentially expressed genes (DEGs) in sarcoidosis. (A) Principal component analysis on transcripts characterizing the TNF-α response (Red -TNF stimulated, Blue-untreated). X axis: first component with eigenvalue; Y-axis: second component with eigen values. **(B)** Mean-Difference Plot of TNF-α Expression in Complicated and Non-complicated cases. **(C)** Heatmap of top 100 DEGs in PBMCs of complicated and non-complicated sarcoidosis with and without TNF-stimulation and untreated healthy controls. These 100 genes show significantly enrich for the JAK-STAT pathway, VEGF, cytokine signaling, MAPK activity and the Immune system.

Pathway analysis identified the TNF-α related transcripts within pathways significantly related to the Immune system (95 DEGs; p <10-E10,); NF-kappa B signaling (18 DEGs; p<10-9); Cytokine-cytokine receptor interaction (22 DEGs; p <10-9). Other pathways significantly enriched included the, Jak-STAT ; TNF-receptor, Apoptosis, and Toll-Like Receptors (p-value <0.01).These analyses also showed a prominent MAPK, NF-kB activation and interferon response to TNF-α stimulation that was unique to those PBMCs from complicated sarcoidosis. **(Table 6).**

Table 6. Top pathways associated to the most significant dysregulated genes in response to TNF α stimulation in PBMCs in complicated and uncomplicated Sarcoidosis

Pathways associated to TNF-α in Complicated Sarcoidosis

pathway name	set size	candidates contained	p-value	q-value	pathway source
NF-kappa B signaling pathway - Homo sapiens (human)	95	12 (12.8%)	1.6e-08	1.16e-05	KEGG
Thymic Stromal LymphoPoietin (TSLP) Signaling Pathway	47	8 (17.0%)	4.42e-07	0.000118	Wikipathways
Innate Immune System	1309	43 (3.3%)	4.9e-07	0.000118	Reactome
RANKL-RANK (Receptor activator of NFKB (ligand)) Signaling Pathway	55	8 (14.5%)	1.54e-06	0.00028	Wikipathways
CLEC7A (Dectin-1) signaling	42	7 (16.7%)	2.76e-06	0.0004	Reactome
Selenium Micronutrient Network	83	9 (10.8%)	4.23e-06	0.000472	Wikipathways
Osteoclast differentiation - Homo sapiens (human)	132	11 (8.4%)	4.56e-06	0.000472	KEGG
C-type lectin receptors (CLRs)	86	9 (10.5%)	5.69e-06	0.000515	Reactome
Immune System	1950	53 (2.7%)	6.56e-06	0.000528	Reactome
TNFR2 non-canonical NF-kB pathway	52	7 (13.5%)	1.2e-05	0.000751	Reactome
Photodynamic therapy-induced NF-kB survival signaling	35	6 (17.1%)	1.23e-05	0.000751	Wikipathways

Pathways associated to TNF-α in Non-complicated Sarcoidosis

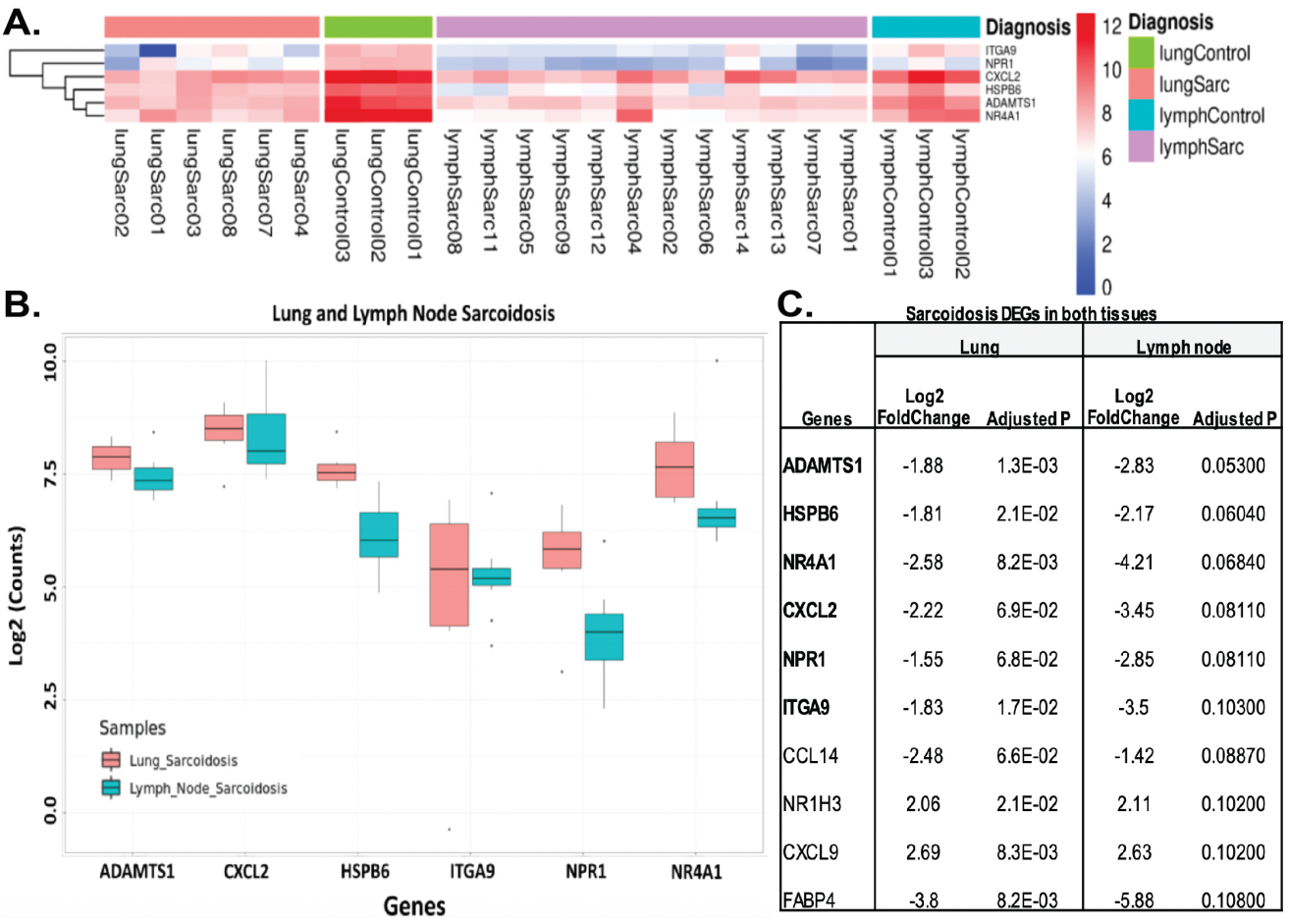
pathway name	set size	candidates contained	p-value	q-value	pathway source
EGFR1	458	24 (5.3%)	4.89e-06	0.00467	NetPath
mRNA Processing	127	11 (8.7%)	2.6e-05	0.0124	Wikipathways
Integrin-mediated Cell Adhesion	101	9 (8.9%)	0.000114	0.0339	Wikipathways
Neutrophil degranulation	497	22 (4.5%)	0.000142	0.0339	Reactome
Immune System	1950	57 (2.9%)	0.000247	0.0472	Reactome
Prednisolone Action Pathway	2	2 (100.0%)	0.000349	0.0476	SMPDB
Prednisolone Metabolism Pathway	2	2 (100.0%)	0.000349	0.0476	SMPDB
Extracellular matrix organization	295	15 (5.1%)	0.000432	0.0509	Reactome
VEGFA-VEGFR2 Signaling Pathway	236	13 (5.5%)	0.000507	0.0509	Wikipathways
Integrin	124	9 (7.3%)	0.000533	0.0509	INOH

In summary, while the number of patients in this pilot group is limited, this study suggests the utility of transcriptomic-paired analysis to identify genotype-specific pathways; transcript level responses to TNF-α exposure in sarcoidosis; and the capacity to identify pathways related to disease severity.

4.3 Selected Gene expression in sarcoidosis granulomatous tissue

Sarcoidosis is an extremely heterogeneous disease that affects multiple organs, it lacks clinical markers capable to differentiate from genomic markers at different multi-organ involvement. Previously published transcriptomic studies from blood, lung and lymph node, lacrimal gland, orbital tissue and skin (*125-127*) have identified other sarcoidosis candidate genes, many of them that are tissue specific. Thus, assessment of a multisystemic disease as sarcoidosis at different compartmental levels is relevant to identify common dysregulated transcripts at different stages and compartments of the disease.

Our goal in this study was to conduct an evaluation on the differentially expressed genes (DEG) in lung and lymph node, the most commonly affected organs in sarcoidosis. The HTG EdgeSeq Oncology-biomarker panel consisted of 2,535 probes. We analyzed 2,430 probes that uniquely mapped to single genes excluding all the probes with multiple hits or partial hits. Transcripts with a fold change (FC) >2 and a p value <0.01 were deemed differentially expressed. We identified 250 significantly dysregulated genes between healthy controls and diseased tissues. A total of 138 DEGs were present in sarcoidosis granulomas from lung and lymph node with a predominant down regulation most marked in lung, 93% of the transcripts were downregulated vs 56 % in the lung. 87 of these 138 DEGs were only dysregulated in sarcoidosis Granulomas from lymph node and lung had 10 overlapping dysregulated genes; ADAMTS1, CXCL2, HSPB6, ITGA9, NPR1, NR4A1, CCL14, CXCL9, FABP4 and NR1H3**,** while only the six first genes were exclusively dysregulated in sarcoidosis but not in Tb or cocci. (**Figure 4)**. ADAMST1, CXCL2, CXCL9, FABP4 were validated in independent microarray sets that were tissue specific (lung and lymph node) (Manuscript under revision).



Genes	Lung		Lymph node	
	Log2 FoldChange	Adjusted P	Log2 FoldChange	Adjusted P
ADAMTS1	-1.88	1.3E-03	-2.83	0.05300
HSPB6	-1.81	2.1E-02	-2.17	0.06040
NR4A1	-2.58	8.2E-03	-4.21	0.06840
CXCL2	-2.22	6.9E-02	-3.45	0.08110
NPR1	-1.55	6.8E-02	-2.85	0.08110
ITGA9	-1.83	1.7E-02	-3.5	0.10300
CCL14	-2.48	6.6E-02	-1.42	0.08870
NR1H3	2.06	2.1E-02	2.11	0.10200
CXCL9	2.69	8.3E-03	2.63	0.10200
FABP4	-3.8	8.2E-03	-5.88	0.10800

Figure 4. Sarcoidosis dysregulated genes present in lung and lymph node. A) Heatmap representing the expression of the six DEGs only present in granulomas from sarcoidosis lung and lymph node. **B)** Box plot of the six transcripts dysregulated in lung and lymph nodes from sarcoidosis. **C)** Total transcripts dysregulated in lung and lymph node. Top six genes in bolded, where present exclusively in sarcoidosis the last four were also significantly dysregulated in TB.

Despite the similarities present at the histological level, the granuloma from sarcoidosis revealed significant divergence at the genomic level at different compartments. Sarcoidosis transcriptome derived from lymph nodes in sarcoidosis delineated a clear immunological response, involving leucocyte migration and neutrophil chemotaxis; while a structural regenerative response was observed at the lung level characterized by cell migration and angiogenesis. The later changes associated to a progressive lung involvement.

Interestingly, the genes shared between comparisons among the three granulomatous diseases all have the same orientation regarding the Fold Change, with the exception of two, indicating a common patho-genomic mechanism between the granulomas.

4.4 Sarcoidosis vs other granulomatous diseases genomic profile

Recently, the candidate profiled the gene expression associated with sarcoidosis and compare to Coccidioidomycosis (Cocci) and Tuberculosis (TB), to assist in the development genomic-based biomarkers specific for sarcoidosis (manuscript under revision). Microdisected sections of formalin fixed paraffin-embedded (FFPE) granulomas were obtained from the pathology department for RNA sequencing, utilizing an immune-oncopanel (detailed in the methods section). In this study, the candidate analyzed the expression profile in granulomas in two of the most commonly affected tissues in sarcoidosis and compare it with the transcriptome from granulomas from TB and cocci, two common granulomatous diseases in the Southwest of the United States. We unveiled the transcriptome profile, in cells conforming the granuloma at the time of the diagnosis in treatment naïve specimens to unravel the molecular mechanisms that occur at the compartment level in sarcoidosis and related to diseases that share a common histopathological hallmark. We captured transcriptomic changes occurring within the granuloma of these granulomatous diseases by using next generation sequencing expression not previously reported before. This method, considered as far more precise for measurement of transcripts (20), requires low amount of RNA to obtain expression signal in paraffin embedded tissues. The immune-based transcriptomic profiles of 3,000 genes, of micro-dissected granulomas from sarcoidosis was analyzed and compared these findings to granulomas from cocci and TB, our comparisons were conducted taking in consideration the tissue origin on each granulomatous disease . A total of 138 DEGs were present in sarcoidosis granulomas from lung and lymph node with a predominant down regulation most marked in lung, 93% of the transcripts were downregulated vs 56 % in the lung (Figure 5).

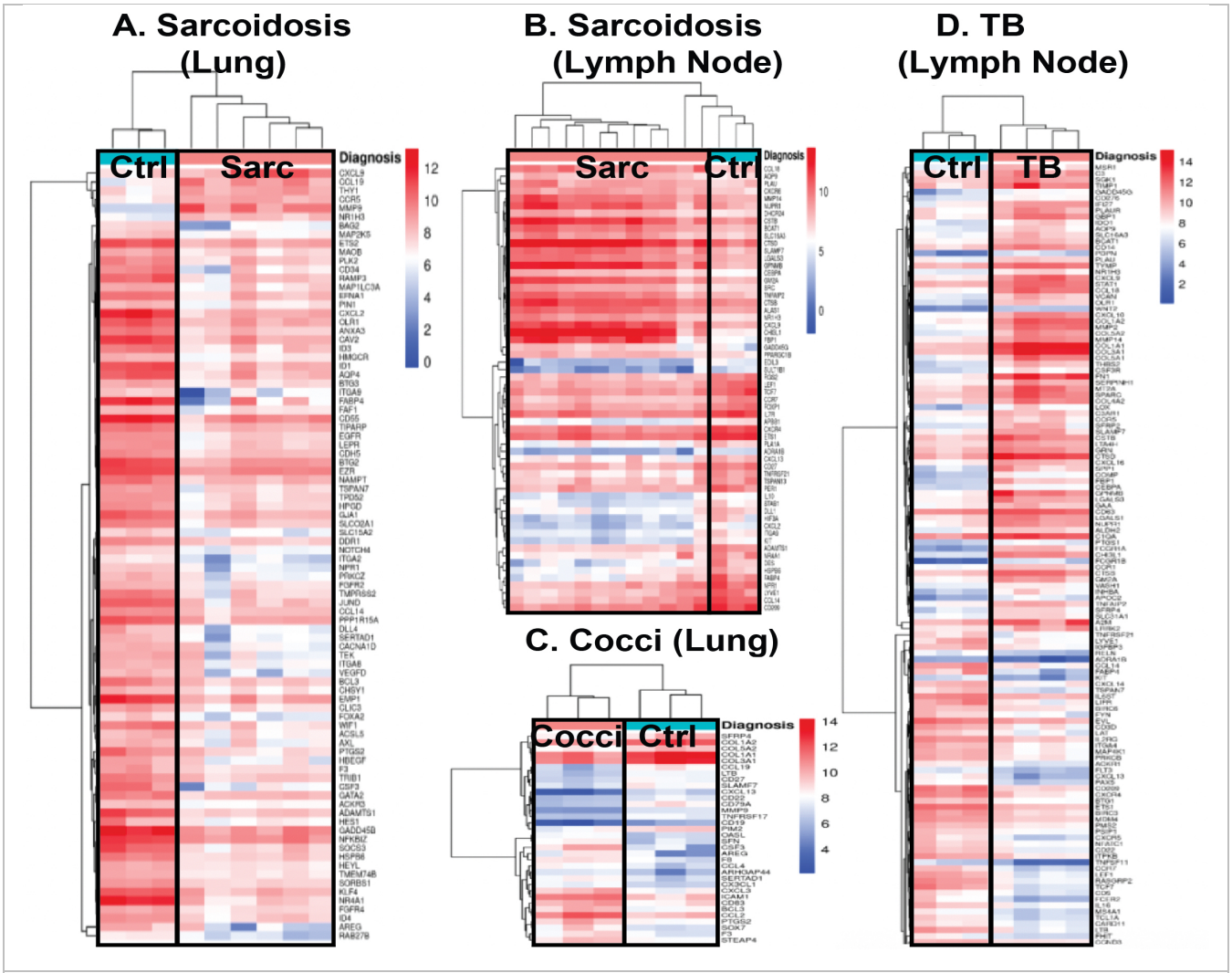


Figure 5. Heatmap representing the expression profiles of the DEG from granulomatous tissue against healthy controls. A) Sarcoidosis lung., B) Sarcoidosis lymph node., C) Cocci., D) Tuberculosis lymph node.

Interestingly, the genes shared between comparisons among the three granulomatous diseases all have the same orientation regarding the FC, with the exception of two, indicating a common patho-genomic mechanism between the granulomas. Granulomas from lung Cocci shared a proportionally a smaller number of genes indicating a lower level of similarity between Cocci and Sarcoidosis than TB and Sarcoidosis. The number of DE genes in each comparison using the same thresholds varied significantly, the number of dysregulated genes did not correlate with the number of samples per comparison, or the size of the microdissection, suggesting that the granulomas are qualitatively different depending on the disease and tissue of origin.

We confirmed that transcriptome similitudes and divergences that originate from same histopathological formation in three different diseases. This study corroborated that some genes previously suggested as potential sarcoidosis markers were also present in fungal or mycobacterium granulomas pointing to a common mechanistic origin. We demonstrated that granulomas from sarcoidosis share stronger similarity at the transcriptional and pathway level with TB than with Cocci.

The transcriptome analysis of the granulomas in the most frequently affected tissues, lung and lymph nodes, fills a crucial gap in our knowledge on gene expression in sarcoidosis. We were able to corroborate the expression of previously genomic markers postulated as potential biomarkers for sarcoidosis and identify the existence on commonly dysregulated genes that are exclusively present in sarcoidosis in both lymph and lung tissue (Manuscript under revision).

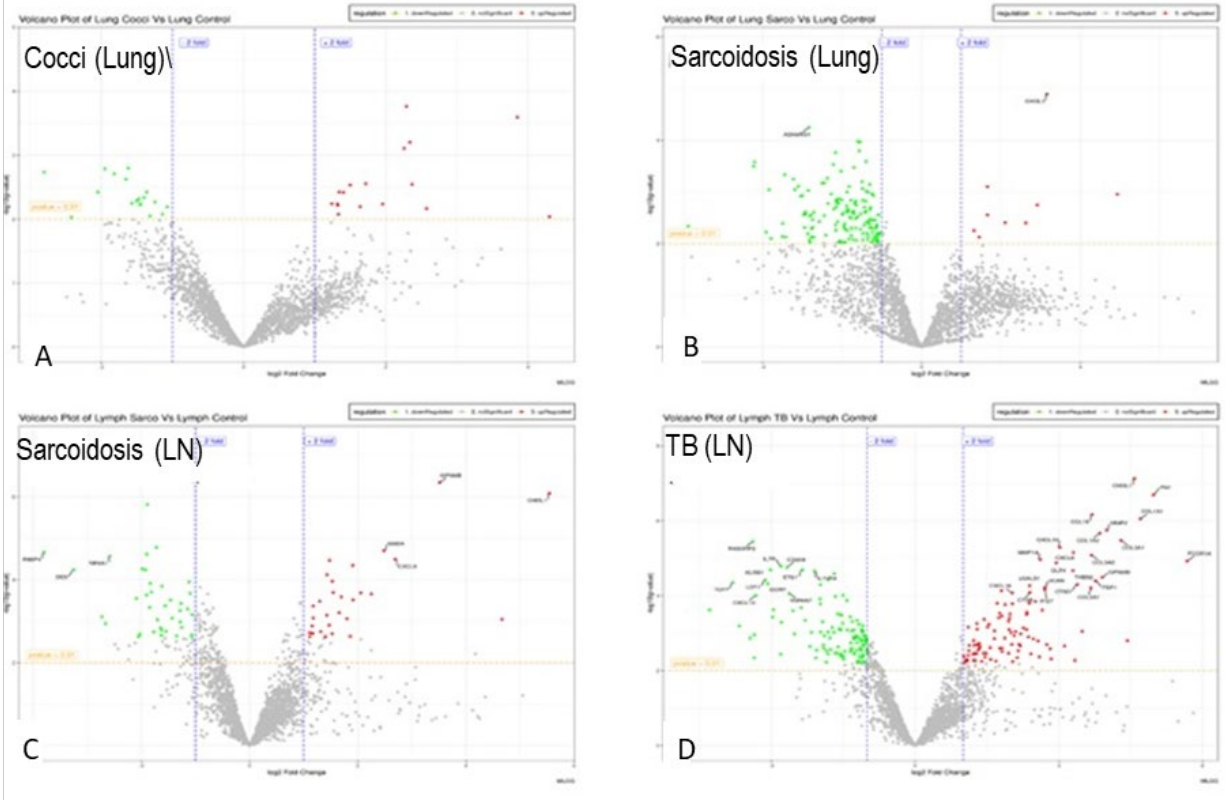


Figure 6. Results of the differential expression analysis. Volcano plot of overall gene-based differential expression. Comparisons were disease and tissue specific vs healthy controls. The x-axis corresponds to the log(base2) of the fold change difference , y-axis corresponds to the negative log(base10) of the p-values. **A)** Lung cocci, **B)** Lung Sarcoidosis**, C)** Lymph Sarcoidosis **D)** Lymph TB.

4.5 Functional Genomics (sarcoidosis dysregulated pathways).

Pathway analysis was performed to understand the pathways that were modulated by the DEG in each disease by compartment. We identified that most of the enriched transcripts in Sarcoidosis encoded proteins associated with protein binding, regulation of biological processes, and cellular processes. Pathway enrichment comparison between lung and lymph nodes in sarcoidosis delineated a clear immunological response, involving leucocyte migration and neutrophil chemotaxis; while a structural regenerative response was observed at the lung level characterized by cell migration and angiogenesis. We also identified common biological pathways that were present in both tissues in sarcoidosis **Figure 7**. Furthermore, our analysis identified common dysregulated pathways in the three disease categories (p <0.006)

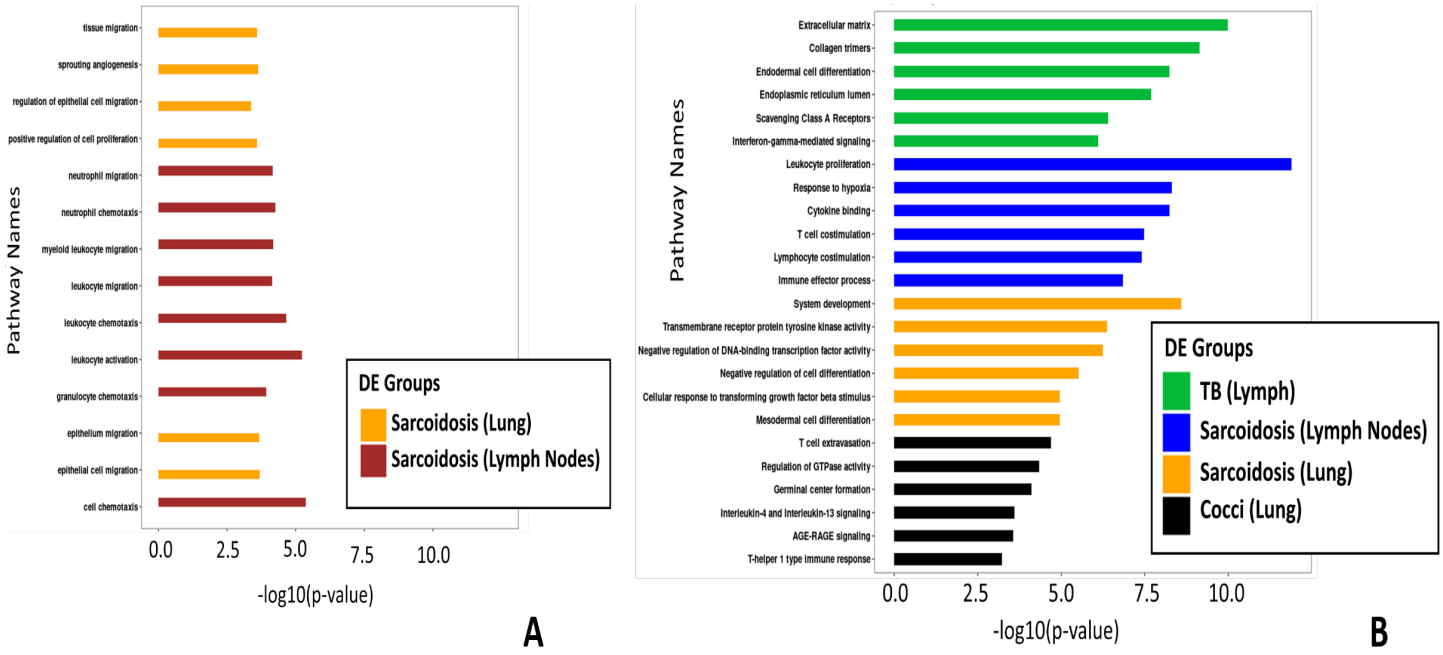


Figure 7. Top over representations of DEGs against the pathway databases and the gene ontology. The x- axis corresponds to the negative log (base 10) of the p-values. Consensus Pathways. A. Sarco**idosis** DEGs in lung and Lymph nodes. B. DEG in TB (green), Sarcoidosis lymph node (blue), Sarcoidosis lung (yellow) Cocci (black).

Sarcoidosis response was characterized by a predominant leucocyte activation and cell chemotaxis, TB response indicated a marked Interferon gamma-mediated response with collagen and extracellular matrix involvement, Cocci response was minor compared to other diseases at the pathway level, (<5 -log10 p value).

5. CURRENT GENETIC AND EPIGENETIC STRATEGIES IN SARCOIDOSIS PHENOTYPING

5.1 Genome-wide DNA Methylation studies

DNA methylation is one of the most important mechanisms involved in gene and microRNA expression regulation. In this study, the candidate aimed to explore the potential role of DNA methylation in the heterogeneous clinical spectrum of sarcoidosis. Despite most sarcoidosis, patients enter remission, complicated sarcoidosis with progressive lung fibrotic disease occurs in 20% of the cases. Complications of fibrosis pulmonary include pulmonary hypertension from capillary obliteration. Lung transplantation is the last resource in these cases which carries and high mortality. What triggers the evolution from contained granulomatous inflammation to pulmonary fibrosis is still unknown. Identifying epigenetic variants associated to DNA methylation that modulates the gene expression of genes that contribute to complicated sarcoidosis is relevant to predict complicate sarcoidosis clinical course. Cytosine methylation in CpG dinucleotides is an important mechanism involved in the regulation of multiple biological processes and pathological status.

Using the microarray-based array by Illumina we assessed the methylation profile of 47 sarcoidosis patients, 25 of them with progressive pulmonary involvement defined as pulmonary fibrosis (radiographic scadding staging IV or Ct scan) and 20 with self-limited pulmonary sarcoidosis(radiographic stages 0-IIor CT scan). DNA was extracted from PBMCs using standard methods, followed by Illumina EPIC methylation 850K array CHIP. The BeadChip are mapped to gene regions based on their functional genome distribution and CpG island regions based on CpG content. We then calculated the Beta values, the ratio of the methylated probe intensity (sum of the methylated and methylated probe intensities), values were normalized and adjusted using combat. Then we compared the methylation between sarcoidosis with progressive lung fibrosis and the ones with the non complicated sarcoidosis (Figure 7).

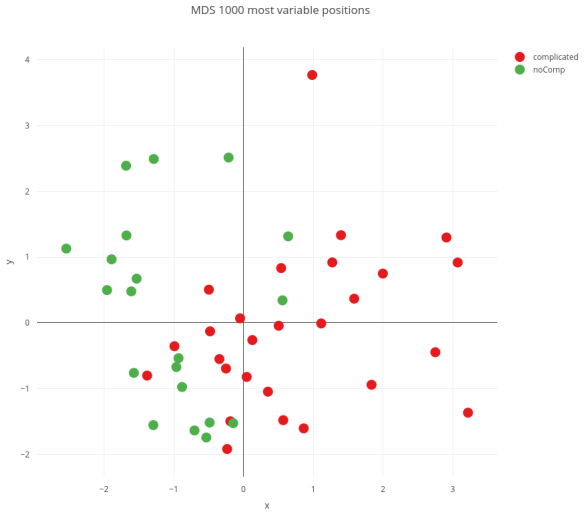


Figure 8. MSD plot. Methylated Site Display (MSD) in 27 complicated sarcoidosis defined as pulmonary fibrosis (radiographic stage IV)(Red) and 20 non complicated sarcoidosis.

Our results indicate complicated sarcoidosis presents an important reduction in methylation in HLA associated genes; HLA-DRB1, HLA-DQA1, HLA-DRB6, HLA-DPA1, and KLRC4-KLRK1 (β <0.18) and increased methylation in these genes; RADIL, DPPA3, MYT1L, HLA-DRB1, and PAQR4 (β > 0.12). Overrepresentation analysis indicated that the most significant pathways dysregulated associated to the hypo-methylated genes were MHC class II antigen presentation, Phosphorylation of CD3 and TCR, PD-1 signaling and TCR signaling.

5.2 MiRNA Studies

Subsequently, our research group explored MiRNA epigenetic regulatory mechanism in sarcoidosis. This study was published in Scientific Reports. Zhou T, Casanova N, Pouladi N, Wang T, Lussier Y, Knox KS, Garcia JGN. Identification of Jak-STAT signaling involvement in sarcoidosis severity via a novel microRNA-regulated peripheral blood mononuclear cell gene signature. Sci Rep. 2017 Jun 26;7(1):4237.

In this study, we paired the MiRNA regulation and the utility of a protein-coding gene to identify an expression signature in sarcoidosis. We initially identified a 8-miRNA signature (7 upregulated and 1 downregulated miRNAs in complicated sarcoidosis) that potentially relate to sarcoidosis severity, including miR-23a, miR-23b, miR-30c, miR-185, and miR-223, now recognized to be related to pulmonary hypertension(*128-132*) and lung cancer(*133-136*). To determine the differentially expressed protein-coding genes in sarcoidosis PBMCs, we analyzed the gene expression pattern from 35 healthy controls, 17 patients with uncomplicated sarcoidosis, and 22 patients with complicated sarcoidosis (Gene Expression Omnibus [GEO][12] accession: GSE37912). Spearman's rank correlation test was conducted between gene expression and sarcoidosis severity. In total, 1,559 genes showed significant correlation (adjusted $P < 0.0005$), among which 340 genes showed positive correlation between gene expression and severity while the expression of 1,219 genes was negatively correlated with severity. Then we identified the 17-gene signature, in which the genes were potentially regulated by a 8-miRNA signature. We next searched the enriched Gene Ontology (GO) biological process terms (*137*) for the 17-gene signature and found that the 17-gene signature is significantly associated with "JAK-STAT cascade". In agreement with prior reported studies, the 17-gene signature was enriched in Jak-STAT signaling pathway (*126, 138, 139*). Jak-STAT signaling pathway is an intracellular cascade initiated in response to cytokine signaling. Several genes in Jak-STAT pathway have already been associated with sarcoidosis, such as *IL15* (*140*), *IL23R* (*141*), and *STAT1* (*142*)derived a molecular signature consisting of 17 protein-coding genes, which are potentially regulated by deregulated miRNA in sarcoidosis (**Figure 9).**

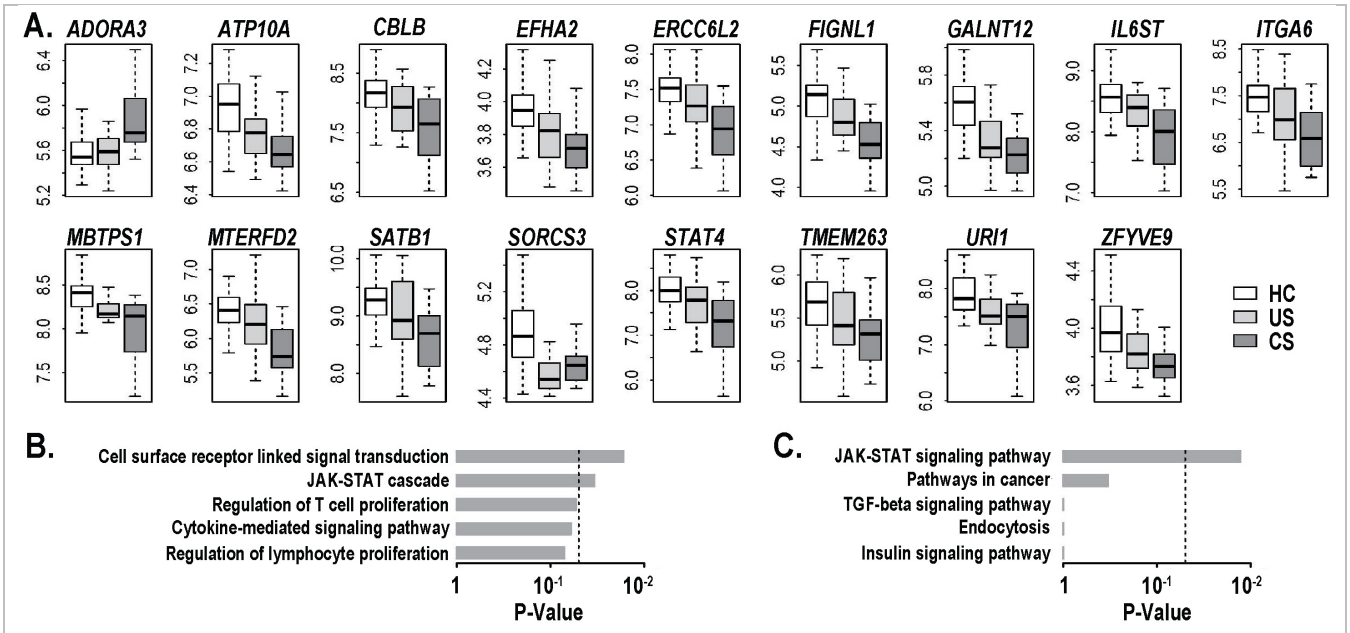


Figure 9. The pathway analysis of the MIRPSS 17 protein-coding gene revealed by Jak-Stat signaling pathway. Panel A. The 17 protein-coding genes were differentially expressed with the severity of sarcoidosis. Y-axis indicates the gene expression level. HC: healthy controls; US: uncomplicated sarcoidosis; CS: complicated sarcoidosis. **Panel B.** The top five GO biological process terms associated **with the 17-gene signature. Panel C.** The top five KEGG pathway terms associated with the 17-gene signature. The P-values were calculated by Fisher's exact test. The dash line denotes the significance level of 0.05.

Despite multiple efforts to identify concordant DE-miRNA in sarcoidosis, it is recognized that patterns of miRNA expression differ between BAL cells, lung tissue, lymph nodes and PBMCs suggesting organ specific regulation.

We assigned a severity score (S_{gene}) to each patient based on the expression of the 17-gene signature, which is a linear combination of the protein-coding gene expression values weighted by direction of differential expression (see Methods for details). Patients with more severe sarcoidosis are expected to have a higher S_{gene}. Not surprisingly, we found a significant positive correlation (Spearman's rank correlation test: $\rho = 0.615$ and P = 5.6×10-9). We next validated the predictive power of the 17-gene signature in two independent blood gene expression datasets. One dataset (GEO accession: GSE19314) (*126*) is from University of California, San Francisco that contains 20 healthy controls and 40 sarcoidosis patients (UCSF cohort). The other dataset (GEO accession: GSE18781) (*142*) is from Oregon Health Sciences University and includes 25 healthy controls and 12 sarcoidosis patients.

In agreement with prior reported studies, the 17-gene signature was enriched in Jak-STAT signaling pathway (*126, 138, 139*) . In our 17-gene signature, *CBLB* and *IL6ST*

are known to be involved in T-cell receptor signaling and cytokine-cytokine receptor interaction, respectively, according to the definition in KEGG database. However, we failed to identify a single gene within our 17-gene signature that overlapped with the previously published sarcoidosis blood gene signatures by Koth et al (*126*) and Zhou et al (*138*), respectively.

In summary, we derived a molecular signature consisting of 17 protein-coding genes, which are potentially regulated by deregulated miRNA in sarcoidosis. This signature can be independently used as potential novel molecular markers for differentiating patients with sarcoidosis, especially for distinguishing the patients with risk of complicated sarcoidosis.

5.3 GWAS-eQTL Polymorphism Studies with Massarray Validation

Genome-wide association studies (GWAS) require large sample size to have enough power. Furthermore, the majority of SNPs detected are in noncoding regions complicating the understanding of the role that such variant play in the disease mechanism. New approaches to increase the power of GWAS involve the integration of expression quantitative loci (eQTL) studies, a powerful method to identify regions of the genome that co-segregate with a given trait.

In this work, the candidate and her research team took advantage of multiple catalogs of eQTLdata, and regulatory annotations to identify SNPs that co-localize with the genetic information from GWAS Sarcoidosis studies performed by the Garcia Lab in EA and AA populations. Our analysis, enriched for 60 intragenic and 190 intergenic coding SNPs that colocalize with previously identified eQTLs, ten of them with lung eQTLs and eleven have whole blood eQTLs. 175 SNPs had significant FDR values for cis westra eQTLs (<1E-03). We identified SNPs with eQTLs for genes associated to JAK/STAT pathway STAT1, IL12B and PSMC1, a pathway we previously reported involved in complicated Sarcoidosis. Furthermore, we identified two specific variants in complicated sarcoidosis related to lung fibrosis pathway, rs8124478 and rs2546893. (rs7300241and rs758167) from the same gene associated to complicated sarcoidosis in both AA and EA.

Table 7. Top intersected SNPS with eQTLs by race and complicated status and top signaling pathways.

	eQTL	Chr	p-val	eQTL-pval	Gene symbols	Tissues	Pathway function
AA Case vs control	rs502771	6	1.27E-04	1.52E-10	C4A; C4B; CLIC1; CYP21A1P; CYP21A2; HLA-DMA; HLA-DQA1; HLA-DQB1; HLA-DQB1-AS1; HLA-DQB2; HLA-DRB1; HLA-DRB5; HLA-DRB9; NOTCH4; PSMB9	Lung, Heart, Brain & Blood	Immunological
	rs7746922	6	4.29E-04	2.15E-10	C2; C4A; C4B; CYP21A1P; HLA-DMA; HLA-DQA1; HLA-DRB5; MICB; NOTCH4; PSMB9; RNF5	Lung, Heart, Brain & Blood	Immunological
	rs17787966	11	5.83E-05	2.10E-08	ACP2; ARHGAP1; C11orf49; C1QTNF4; DDB2; LRP4; LRP4-AS1; PACSIN3	Lung, Heart & Blood	Vesicle-mediated transport
EA Case vs Control	rs3094228	6	1.71E-05	3.35E-10	C4A; C4B; CCHCR1; CYP21A1P; CYP21A2; HCG22; HLA-C; MICB; NOTCH4; PPP1R18; PSORS1C1; VARS2	Lung, Heart, Brain & Blood	Immunological
	rs9268362	6	1.38E-04	7.99E-11	FKBPL; HLA-DQA2; HLA-DRB1; HLA-DRB6; NOTCH4	Lung, Heart, Brain & Blood	Immunological
	rs2844627	6	6.31E-05	3.31E-12	C4A; C4B; C6orf15; CCHCR1; CYP21A1P; HCG20; HCG22; HLA-B; HLA-C; MICB; NOTCH4; POU5F1; PPP1R18; PSORS1C2; RNF5; VARS2; WASF5P	Heart, Brain & Blood	Immunological
AA Complicated vs Non-complicated	rs7787110	7	9.37E-04	2.08E-68	AK3P3; GPNMB; KLHL7; KLHL7-AS1; NUPL2	Lung, Heart, Brain & Blood	Biological oxidations
	rs10903129	1	8.29E-04	4.62E-21	RHCE; RHD; SDHDP6; TMEM50A; TMEM57	Lung, Heart, Brain & Blood	Biological oxidations
	rs113413	22	9.24E-04	9.80E-62	DDT; DDTL; GSTT2; GSTT2B; MIF; MIF-AS1	Lung, Heart, Brain & Blood	Biological oxidations
EA Complicated vs Non-complicated	rs9889755	17	4.86E-05	1.27E-27	ATAD5; CRLF3; LRRC37BP1; SUZ12P1; TEFM	Lung, Heart, Brain & Blood	Lymphocyte activation
	rs1052053	1	2.23E-04	1.52E-11	GLMP; PAQR6; SEMA4A; SLC25A44; SMG5	Lung & Blood	Lymphocyte activation
	rs11912715	22	2.26E-04	1.52E-20	MMP11; SLC2A11; SMARCB1; VPREB3	Lung, Heart, Brain & Blood	Immunoglobulin production

Next, we conducted a validation utilizing MassARRAY, a DNA analysis platform that uses mass spectrometry to measure the amount of genetic target material and/or variations. A total of 976 DNA samples were included, 336 and 625 African American and whites. Data was generated for 103 of 113 SNPs, with 54 SNPs called over 99.5% success rate. Our results validated 13 polymorphisms associated to risk and severity in sarcoidosis in AA, EA and one SNP, rs7219 associated to risk for complicated sarcoidosis in EA. **Table 7.**

Table 8. GWAS eQTL variants validated associated to sarcoidosis susceptibility and severity.

SNP	Group risk association	Genes Affected
rs7248735	AA	IL27RA
rs10044736	AA	CTNNA1 SIL1
rs11677881	AA	DDX11L2 FAM138B PGM5P4 PGM5P4-AS1 RABL2A RPL23AP7 WASH2P
rs1476792	AA	PTPRCAP
rs17787966	AA	ACP2 ARHGAP1 C11orf49 C1QTNF4 DDB2 LRP4 LRP4-AS1 PACSIN3
rs2054517	AA	GLIPR1 GLIPR1L2 KRR1
rs2596509	AA	ATP6V1G2 CCHCR1 HLA-B HLA-S MICA PSORS1C1 PSORS1C2 STK19B VARS2 ZBTB12
rs443532	AA	PIK3R5
rs502771	AA	C4A C4B CLIC1 CYP21A1P CYP21A2 HLA-DMA HLA-DQA1 HLA-DQB1 HLA-DQB1-AS1 HLA-DQB2 HLA-DRB1 HLA-DRB5 HLA-DRB9 NOTCH4 PSMB9
rs6779819	EA	C3orf18 DOCK3 HEMK1 LINC02019 MAPKAPK3 TEX264
rs1442533	EA	CNTLN
rs7549445	EA	JAK1 RP11-182I10.3
rs7219	EA Complicated	NUP85 MRPS7 GRB2

Additionally we compiled literature published SNPs consisting in 15 SNPs and 16 genes, GWAS catalog 41 SNPs and 45 Genes and included our variants validated Massarray SNPs. We annotated SNPs and map them to specific regions in the genome; our results identified a proportionally high number of genes in chromosome 6p21.**Figure 10.**

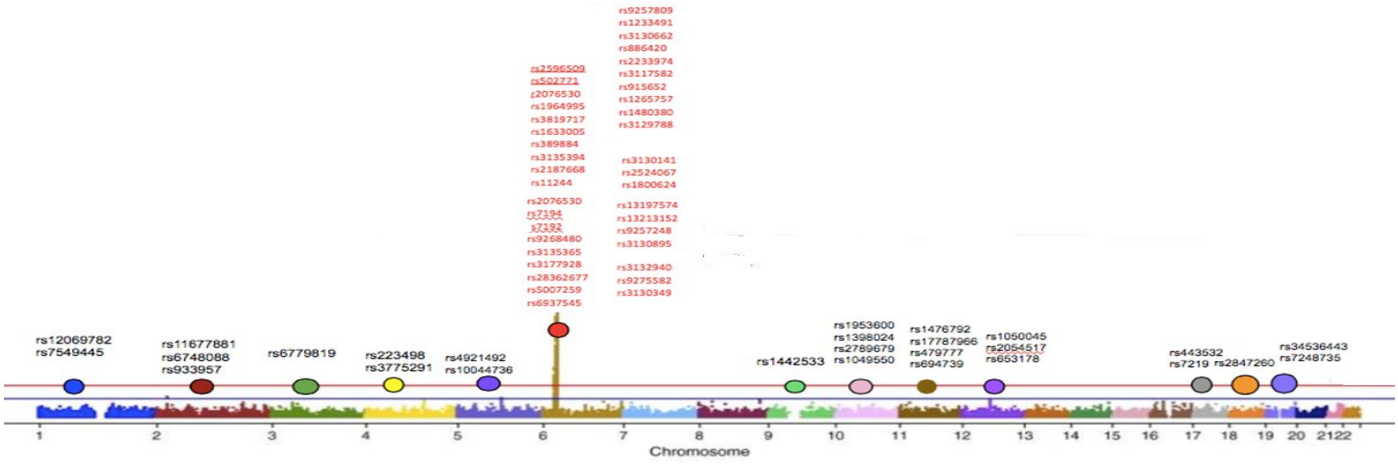


Figure 10. Genomic location of SNPs associated with sarcoidosis.

We confirmed several SNPs associated with complicated sarcoidosis majority associated to immunological pathways associated to HLA complex-genes. By integrating functional datasets and eQTLs we can prioritize the results and identify risk associated SNPs that predict complicated sarcoidosis.

6. MATERIALS AND METHODS

Human Subjects' Protections and Risk. All study personnel was certified in the ethical conduct of human biomedical, genetics research, and HIPPA information security. Data recorded and collected by the investigators at the clinical centers was coded and given a unique ID code. This information along with consents will remain in a locked filing cabinet within the Principal Investigator's office. Data was stored in computer files within a locked office and will require a password to access via an encrypted, password-protected web based RedCap system at University of Arizona. Training of clinical center staff occurred prior to the enrollment of the first patient. Since the system is expected to be completed online, an electronic copy of the website submission will be available at the end of the encounter and will serve as the source document kept at the clinical site. Source documents were saved electronically according to Good Clinical Practice. All genomic data was be behind a University of Arizona firewall and protected as required to minimize risk of breach of confidentiality present in all genetic/genomic studies. We also included the use of codes to identify subjects and securing all genetic samples and clinical data at locked sites and genomic data will be stored in secure servers.

De identified samples from about 1200 sarcoidosis patients from the Garcia laboratory's **biorepository**, all met ERS/ATS criteria for a diagnosis of sarcoidosis, were available for this research. Samples were part of the biorepository from Dr. Garcia, obtained from material transfer agreements and additional recruiting studies IRB approved # 1312168664 and1312166668. Sarcoidosis phenotypic data included clinical data to evaluate the organ involvement among complicated and uncomplicated cases. Groups were stratified in complicated and uncomplicated categories based on organ involvement type and acute and not acute presentation and number of organs affected. Complicated sarcoidosis defined as Lung involvement with parenchymal lung disease confirmed by CT scan, X-R Scadding stages III and IV and/or PFT with FVC <50%. Organ involvement: Cardiac or neurological involvement in addition to the lungs. In addition, clinical Course (remitting, chronic, and uncertain) was determined.

Ex vivo Genome-wide gene expression in PBMC

The experiment consisted of stimulation of peripheral blood mononuclear cells (PBMC)with TNF-α (20ng/ul for 6 hours). RNA from stimulated and unstimulated PBMCs was extracted following Qiagen's miRNA easy protocol.SMARTer Stranded Total RNA-seq kit-Pico was used for cDNA library preparation. Samples were sequenced using High-throughput RNA sequencing (RNA-seq) on Illumina HiSeq 3000 instrument. For this pilot study, four sarcoidosis patients (two complicated and two on-complicated cases) with three replicas from each one were submitted to RNAseq. Fastq files were aligned using the HISAT software. The transcripts were reconstructed and quantified using the StirngTie software and the latest version of the GTF annotations from ensemble (GRCh38.87). The differential expression analysis (DEA) was performed independently by using two programs Ballgown and cuffdiff. We consider a genes to be dysregulated if the p-values of the DEA was < 0.0001. The differentially expressed genes (DEGs) were submitted to gene enrichment using the integrative interaction network server "ConsensusPathDB-human" provided by the Max Planck Institute for Molecular Genetics, it provides access to most of the available pathway databases on a single site.

Gene Expression in Granulomatous Tissue

We analyzed gene expression profiles of microdisected granulomatous tissue from 30 subjects and 6 healthy individuals. Tissue specimens were collected from clinically indicated biopsies for diagnosis purposes. De-identified specimens were obtained from the Tissue Acquisition Shared Resource at the University of Arizona and approved by the Human subjects Protection Program IRB # 1509097312A001. Formalin fixed paraffin-embedded (FFPE) microdissections from lung and lymph node granulomatous tissue from Healthy (6) Sarcoidosis (22), Coccidioidomycosis (4) and Tuberculosis (4) assayed using Next Generation Sequencing-based gene expression by HTG EdgeSeq Oncology-biomarker panel assay (HTG Molecular Diagnostics, Inc.). The panel conveyed 2558 probes, 15 housekeeper genes, four negative process controls and four positive process controls for quantitative analysis of targeted mRNAs. Multiple genes part of pathways previously identified as relevant in sarcoidosis such as the interferon, MAP kinase, NFKB, the JAK/STAT (*46, 126, 138*) and other immune-oncologic related pathways were covered by this panel (*143*).

Gene expression profiling platform. Representative tissue sections from granulomatous tissue were hematoxylin and eosin- stained. Microdissection areas were selected by an expert pathologist, approximately 1.5 mm² of a 5µm tick FFPE tissue section from granulomas were microdissected, about 5-10 sections per subject were lysed and followed by and RNA extraction-free chemistry method using the HTG®'s assay kit. Next, for library preparation the HTG Edge processor for nuclease protection steps was used, followed by PCR tagging, library amplification, quantitation and normalization. Sequencing was performed on an Illumina MiSeq using 150-cycle V3 kits. Expression data was generated using FASTQ files. For the **bioinformatic analysis**, we generated the transcriptome index using the salmon package (*144*) and a fasta file containing the sequence of the HTG EdgeSeq 2567 probes. Quantification was performed for each sample using the salmon quant command(*144*), using the index file and the corresponding fastq file. The raw counts were loaded into R using the tximport package(*145*). The Differential Expression (DE) was calculated using Limma. Functional **analysis**. A list with statistically significant genes for each comparison was submitted to the human ConsensusPathDB (*146*) website using the over representation gene set analysis against the pathway databases and

the gene ontology categories using a minimum overlap with input list of 2 and a p-value cutoff of 0.01.

GWAS-eQTL Polymorphism Studies with Massarray Validation

We analyzed the genetic risk predisposing to complicated sarcoidosis in African Americans (AA), European Americans and healthy controls from the Genetic Association Information Network (GAIN) (627 AA and 579 EA). Prioritized SNPs on each category were annotated and intersected with eQTLdata sets from GTEx 7.0. Massarray validation. For SNP Genotyping validation we utilized the MassARRAY®system, which uses the iPLEX®assay with primer extension chemistry to allow efficient multiplexing of upwards of 30SNPs. PCR products are crystallized on the matrix surface of the SpectroCHIP®. A total of 976 DNA samples were included for MassARRAY, 336 and 625 African American and whites respectively, about 30% of them were classified as complicated sarcoidosis. Data was generated for 103 of 113 SNPs, with 54 SNPs called over 99.5% success rate.

DNA methylation.

Genomic DNA was extracted from PBMC collected at subject enrollment. Using the Illumina 850K EPIC array, the DNA was hybridized to the EPIC BeadChip using Zymo Bisulfite Conversion for DNA methylation. DNA methylation assays were performed at the Genomics and Microarray Core at the University of Colorado Denver. Briefly, DNA (250–750 ng) was treated with sodium bisulphite using the EZ DNA methylation kit (Zymo Research, CA, USA). DNA methylation was quantified using the Illumina Infinium and HumanMethylationEPIC (EPIC) BeadChip (Illumina, CA, USA). Raw IDAT files were processed with Illumina's GenomeStudio software V2011.1 and background normalized using negative control probes to generate methylation β-values which were used for all downstream analyses. All analyses were conducted in R statistical software. Limma Bioconductor package was used to parse data and to perform differential methylation analysis. In total, we included 47 sarcoidosis patients, 25 of them with progressive pulmonary involvement defined as pulmonary fibrosis (radiographic scadding staging IV or CT scan) and 20 with self-limited pulmonary sarcoidosis (radiographic stages 0-IIor CT scan). We removed any technically unreliable probes. We calculated methylation levels from raw data using B values.

7. DISCUSSION AND CONCLUSION

Conventional biomarkers have been be insufficiently sensitive for implementation in routine clinical care. In my dissertation, I focused on the utilization utilizing genomic base approaches based on high-throughput molecular assays to characterize genotype, gene expression, and epigenetics that define sarcoidosis subphenotypes.

Our results demonstrated that the integration of eQTL data from GTex studies increase the power of underpowered GWAS studies. This innovative method allowed the identification of **SNPs associated to complicated sarcoidosis in AA and in EA**, and then we validated some of these SNPs by Massarray. Following a Masarray validation, we replicated associations for several previously reported sarcoidosis-susceptibility risk loci in our AA collection, including MHC Class II region genes. It is known that the MHC Class II region plays a major role in immune-mediated disorders, including associations to celiac disease, insulin-dependent diabetes mellitus, rheumatoid arthritis, multiple sclerosis, and systemic lupus erythematosus.

Our analysis also replicated associations HLA-DRA, HLA-DRB5, HLA-DRB1, BTNL2, and ANXA11 in both our AA and EA datasets. Similarly, BTNL2, and ANXA11 have been suggested to play a role in T-cell activation (*100, 147*). Following MAsrray validation confirmed sarcoidosis association with two HLA related SNPS, rs2596509 (HLA-B, PSORS1C1) and rs502771 (HLA-DQB1, HLA DQA1); other non-HLA related variants were also validated, among them rs72248735 (RLN3/ IL2RA). NOTCH4 gene, previously identified as sarcoidosis-associated locus in AA (32), reached genome-wide significance in our AA samples (rs715299). We identified significant association between sarcoidosis and a previously underreported locus (JAK1) in our EA cohort, JAK1 is a membrane protein member of a class of protein-tyrosine kinases (PTK) that phosphorylates STAT proteins (signal transducers and activators of transcription) and plays a key role in interferon-alpha/beta and interferon-gamma signal transduction. JAK/STAT pathway was reported as dysregulated in our first miRNA **epigenetic** study (*148*). In this study, we identified 46 microRNAs and 1,559 genes that were differentially expressed across sarcoidosis and 19 microRNA-mRNA regulatory pairs that derived a 17 unique protein-coding gene signature whose pathway analysis revealed JAK/STAT signaling pathway as the most significantly represented pathway. A severity score based on the expression of the 17-gene signature showed an increasing trend in complicated sarcoidosis useful in

diagnosis and severity assessment of sarcoidosis. **Additional relevant epigenetics correlations with methylation** data were identified in the GWA eQTL variants. For example, HLA-DRB1 had a reduction of 50% methylation in complicated sarcoidosis; similarly, HLA-DR QA1 observed a reduction of 31% of methylation in complicated cases. The integration of epigenetics to the genomic variants corroborates the presence of specific regions in the genome relevant for risk and severity of sarcoidosis.

The **PBMCs transcriptomic responses to TNF-α** exposure in complicated and uncomplicated sarcoidosis was analyzed. We identified the presence of a differential dysregulation in genes associated to metabolic pathways. As expected, immune system (p value1.77e-12 for 80 genes) and Cytokine-cytokine receptor interaction (p-value of 8.2e09) were the top dysregulated pathways. Our analyses also showed a prominent MAPK activation and interferon response to TNF-α stimulation that was unique to those PBMCs drawn from complicated sarcoidosis patients. This mitogen-activated protein kinase (MAPK) cascade is a critical pathway described in human cancer cell survival, specifically involved in cellular proliferation (*149*). It allows for the periodic environmental adaptation necessary for activation and regulation for cell survival, events that are critical for the granuloma proliferation. We also showed for the first time, activation of interferon by PBMCs in response to TNF-α in complicated sarcoidosis. Our study therefore, not only provides transcriptional changes in PBMCs but also reveals a role for TNF-α signaling in sarcoidosis severity. We recognized that one of the limitations of this study is the limited number of patients included, four in total. Despite this limitation, this study highlights the capacity to differentiate response to the same cytokine at different clinical stages.

We captured **transcriptomic profile within the granulomas of sarcoidosis** in the two most commonly affected compartments using next generation sequencing expression not previously reported before. Results revealed significant divergence at the genomic level between these compartments. Granulomas from lymph node and lung had 10 only overlapping dysregulated genes; ADAMTS1, CXCL2, HSPB6, ITGA9, NPR1, NR4A1, CCL14, CXCL9, FABP4 and NR1H3**,** while only the six first genes were exclusively dysregulated in sarcoidosis (not in Tb or cocci). These hub genes ADAMST1, CXCL2, CXCL9, and FABP4 were validated in independent microarray sets. At the pathway level, transcriptome derived from lymph nodes delineated a clear immunological response, involving leucocyte migration and neutrophil chemotaxis; while a structural

regenerative response was observed at the lung level characterized by cell migration and angiogenesis. These later changes associated to a progressive lung involvement that follows the initial hilar lymphadenopathy involvement. Additionally, we compared the **genomic profile of granulomas in Sarcoidosis vs TB and Cocci.** This study corroborated that some genes previously suggested as potential sarcoidosis markers were also present in fungal or mycobacterium granulomas. For example, we identified that CXCL9, a chemoattractant for lymphocytes, previously proposed as a marker of severity in sarcoidosis (26)is also over-expressed in TB. Previous microarray studies identified that CXCL9 is also dysregulated in Beryllium disease (27), another granulomatous disease. Similarly, we identifiedaNOTCH4 downregulation in our panel in lung sarcoidosis (p value 0.0008). NOTCH4 was previously reported as a sarcoidosis-associated locus in genome-wide associated study in African Americans (1), its signaling is initiated after activation of toll-like receptor in macrophages. CXCL13, another chemokine promotor of B cell migration by regulation of Ca influx, it was previously proposed as a biomarker for uveitis (28, 29), a condition affecting one third of sarcoidosis patients even before the systemic disease is manifested (30). We identified that CXCL13, CD27 and SLAMF7, were the only transcripts dysregulated in Cocci, TB and sarcoidosis (lymph nodes). CXCL13 is a chemokine with key role in B cell migration (24), while SLAMF7 is a regulator of T lymphocyte development and function such as lytic activity and a modulator of B cell activation and memory (25). Furthermore, CD27 regulates B-cell activation and immunoglobulin synthesis through to the activation of NF-kappa B and MAPK8/JNK. This finding, points to a diverse molecular mechanism present in granulomatous diseases that result in a similar histopathological finding. Three of the downregulated genes, SERTAD1, HBEGF and KLRB1, profiled by this oncopanel, were also previously identified as part of our 20gene signature generated by microarray in peripheral blood from PBMCs(26). However, we found that KLRB1, however, is not exclusively downregulated in sarcoidosis, but also present in TB. Similarly, SERTAD1 was also significantly downregulated in Cocci. Another gene part of the same gene signature was the Heparin-binding epidermal growth factor (HBEGF), involved in epithelialization, wound contraction and angiogenesis. It also has a pro-inflammatory role in skin and lung lesions such as alveolar regeneration after pneumectomies (24), pneumonitis and commonly found in early stages of systemic sclerosis (25).We

confirmed that HBEGF was down-regulated in granulomas from lung (FC-2.36 p0.0016) exclusively in sarcoidosis. Furthermore, our research group previously conducted epigenetic studies in sarcoidosis identifying IL6STand STAB1as part of a MiRNA-derived gene signature for complicated sarcoidosis (5). This time we identified that IL6STwas significantly dysregulated only in TB granuloma(FC -1.42 and p 0.0008), whileSTAB1 gene was significantly downregulated in lymph nodes in sarcoidosis (FC -1.64 p 0.0014). STAB1 codes for a transmembrane receptor that is expressed in endothelial cells and lymph nodes, with functions in angiogenesis, lymphocyte homing and cell adhesion. ADMATS1 seems to be an important regulator of the pathogenic mechanism associated to the granuloma formation in sarcoidosis. Initially identified as DEG by this oncopanel as exclusively dysregulated in both compartments, then it was validated by microarray analysis in both tissues in sarcoidosis. This gene encodes a disinterring and metalloproteinase with thrombospondin motif associated with various inflammatory processes. Leveraging the power derived from multiple microarray data sets, were able to conduct a validation of the expression level of the DE panel-derived genes. We validated 70 of the 241 DEG derived from the biomarker panel were also dysregulated in independent cohorts of sarcoidosis, tuberculosis and valley fever. Despite the selection of granuloma microdisected areas for oncopanel sequencing, we were able to detect that tissue origin imprints an effect at the expression level in both platforms. Clinical implications Identifying transcription profiles from cells histologically derived microdissections of granulomas, in diverse diseased tissue can help as a diagnostic tool to identify affected individuals and indicate genes, pathways and cells affected, improving the understanding of enigmatic diseases like sarcoidosis and enabling the design of new therapeutic interventions. This relevant translational information was derived from newly diagnosed and treatment naïve patients. In summary, this study corroborated that some genes previously suggested as potential sarcoidosis markers were also present in fungal or mycobacterium granulomas pointing to a common mechanistic origin. We demonstrated that granulomas from sarcoidosis share stronger similarity at the transcriptional and pathway level with TB than with Cocci.

In summary, the studies presented as part of this dissertation summarized the utility of the use of alternate strategies for biomarkers development utilizing genomic base approaches based on high-throughput molecular assays. The candidate identified novel

variants associated to sarcoidosis sub-phenotypes, analyzed and compared the gene expression at the tissue level, and highlighted some of the epigenetics modifications associated to sarcoidosis subphenotypes. From the epigenetics analysis, she also identified dysregulated novel miRNAs associated to complicated sarcoidosis and altered methylation patterns in loci, HLA among others associated to sarcoidosis by GWAS.

REFERENCES

1. R. P. Baughman, M. Iannuzzi, Tumour necrosis factor in sarcoidosis and its potential for targeted therapy. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* **17**, 425-431 (2003).
2. E. Hem, [Multiple benign sarcoid of the skin--100 years since Caesar Boeck's pioneering article]. *Tidsskr Nor Laegeforen* **119**, 4567-4569 (1999).
3. P. Spagnolo, Sarcoidosis: a Critical Review of History and Milestones. *Clin Rev Allergy Immunol* **49**, 1-5 (2015).
4. Y. Ma, A. Gal, M. N. Koss, The pathology of pulmonary sarcoidosis: update. *Semin Diagn Pathol* **24**, 150-161 (2007).
5. K. K. Shah, B. S. Pritt, M. P. Alexander, Histopathologic review of granulomatous inflammation. *J Clin Tuberc Other Mycobact Dis* **7**, 1-12 (2017).
6. Y. Rosen, Pathology of sarcoidosis. *Semin Respir Crit Care Med* **28**, 36-52 (2007).
7. S. Saidha, E. S. Sotirchos, C. Eckstein, Etiology of sarcoidosis: does infection play a role? *Yale J Biol Med* **85**, 133-141 (2012).
8. I. Kroes, P. W. Lepp, D. A. Relman, Bacterial diversity within the human subgingival crevice. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 14547-14552 (1999).
9. B. J. Paster *et al.*, Bacterial diversity in human subgingival plaque. *J Bacteriol* **183**, 3770-3783 (2001).
10. Z. Gao, C. H. Tseng, Z. Pei, M. J. Blaser, Molecular analysis of human forearm superficial skin bacterial biota. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2927-2932 (2007).
11. E. M. Bik *et al.*, Molecular analysis of the bacterial microbiota in the human stomach. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 732-737 (2006).
12. S. R. Gill *et al.*, Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359 (2006).
13. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* **160**, 736-755 (1999).
14. W. P. Drake *et al.*, Cellular recognition of Mycobacterium tuberculosis ESAT-6 and KatG peptides in systemic sarcoidosis. *Infect Immun* **75**, 527-530 (2007).

15. Z. Song *et al.*, Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. *J Exp Med* **201**, 755-767 (2005).
16. Y. Eishi *et al.*, Quantitative analysis of mycobacterial and propionibacterial DNA in lymph nodes of Japanese and European patients with sarcoidosis. *J Clin Microbiol* **40**, 198-204 (2002).
17. K. Yamamoto *et al.*, Catalase expression of Propionibacterium acnes may contribute to intracellular persistence of the bacterium in sinus macrophages of lymph nodes affected by sarcoidosis. *Immunol Res* **67**, 182-193 (2019).
18. Y. Ebe *et al.*, Proliferative response of peripheral blood mononuclear cells and levels of antibody to recombinant protein from Propionibacterium acnes DNA expression library in Japanese patients with sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* **17**, 256-265 (2000).
19. I. Ishige *et al.*, Propionibacterium acnes is the most common bacterium commensal in peripheral lung tissue and mediastinal lymph nodes from subjects without sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* **22**, 33-42 (2005).
20. C. M. Swaisgood *et al.*, Development of a sarcoidosis murine lung granuloma model using Mycobacterium superoxide dismutase A peptide. *Am J Respir Cell Mol Biol* **44**, 166-174.
21. J. Minami *et al.*, Pulmonary granulomas caused experimentally in mice by a recombinant trigger-factor protein of Propionibacterium acnes. *J Med Dent Sci* **50**, 265-274 (2003).
22. D. R. Moller *et al.*, Rationale and Design of the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) Study. Sarcoidosis Protocol. *Ann Am Thorac Soc* **12**, 1561-1571 (2015).
23. K. A. Oswald-Richter *et al.*, Cellular responses to mycobacterial antigens are present in bronchoalveolar lavage fluid used in the diagnosis of sarcoidosis. *Infect Immun* **77**, 3740-3748 (2009).
24. K. A. Oswald-Richter *et al.*, Multiple mycobacterial antigens are targets of the adaptive immune response in pulmonary sarcoidosis. *Respir Res* **11**, 161.
25. E. Chen, e. al, Serum Amyloid A Regulates Granulomatous Inflammation in Sarcoidosis through Toll-like Receptor-2. *Am J Respir Crit Care Med* **181**, 360-373 (2010).
26. E. Chen, D. R. Moller, Sarcoidosis—Scientific Progress and Clinical Challenges. *Nat Rev Rheumatol* **7**, 457-467 (2011).
27. S. T. Brown *et al.*, Recovery of cell wall-deficient organisms from blood does not distinguish between patients with sarcoidosis and control subjects. *Chest* **123**, 413-417 (2003).
28. E. Jonsson, B. Jarvholm, M. Andersson, Silica dust and sarcoidosis in Swedish construction workers. *Occup Med (Lond)*, (2019).
29. L. E. Crowley *et al.*, "Sarcoid like" granulomatous pulmonary disease in World Trade Center disaster responders. *Am J Ind Med* **54**, 175-184 (2011).
30. G. Izbicki *et al.*, World Trade Center "sarcoid-like" granulomatous pulmonary disease in New York City Fire Department rescue workers. *Chest* **131**, 1414-1423 (2007).
31. M. Ramos-Casals, B. Kostov, P. Brito-Zeron, A. Siso-Almirall, R. P. Baughman, How the Frequency and Phenotype of Sarcoidosis is Driven by Environmental Determinants. *Lung* **197**, 427-436 (2019).
32. I. Adrianto *et al.*, Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PLoS One* **7**, e43907 (2012).
33. T. E. Fingerlin, N. Hamzeh, L. A. Maier, Genetics of Sarcoidosis. *Clin Chest Med* **36**, 569-584 (2015).
34. A. Fischer *et al.*, Genetics of sarcoidosis. *Semin Respir Crit Care Med* **35**, 296-306 (2014).
35. M. C. Iannuzzi, B. A. Rybicki, Genetics of sarcoidosis: candidate genes and genome scans. *Proc Am Thorac Soc* **4**, 108-116 (2007).
36. B. A. Rybicki *et al.*, Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). *Am J Respir Crit Care Med* **164**, 2085-2091 (2001).

37. P. Brito-Zeron, B. Kostov, D. Superville, R. P. Baughman, M. Ramos-Casals, Geoepidemiological big data approach to sarcoidosis: geographical and ethnic determinants. *Clin Exp Rheumatol*, (2019).
38. M. H. Tukey *et al.*, Mortality among African American women with sarcoidosis: data from the Black Women's Health Study. *Sarcoidosis Vasc Diffuse Lung Dis* **30**, 128-133 (2013).
39. M. C. Iannuzzi, B. A. Rybicki, A. S. Teirstein, Sarcoidosis. *N Engl J Med* **357**, 2153-2165 (2007).
40. H. Nunes, D. Bouvry, P. Soler, D. Valeyre, Sarcoidosis. *Orphanet J Rare Dis* **2**, 46 (2007).
41. N. Soto-Gomez, J. I. Peters, A. M. Nambiar, Diagnosis and Management of Sarcoidosis. *Am Fam Physician* **93**, 840-848 (2016).
42. M. H. Noe, M. Rosenbach, Cutaneous sarcoidosis. *Curr Opin Pulm Med* **23**, 482-486 (2017).
43. R. P. Baughman *et al.*, Defining the clinical outcome status (COS) in sarcoidosis: results of WASOG Task Force. *Sarcoidosis Vasc Diffuse Lung Dis* **28**, 56-64 (2011).
44. M. C. Iannuzzi, J. R. Fontana, Sarcoidosis: clinical presentation, immunopathogenesis, and therapeutics. *JAMA* **305**, 391-399.
45. G. C. Nancy, Z. Tong, N. Jorge, S. K. Kenneth, G. N. G. Joe, in *A108. MICRO RNAS, RNA SEQ, LNCRNA: BIOLOGY AND FUNCTION*. (American Thoracic Society, 2016), pp. A2809-A2809.
46. T. Zhou *et al.,* U. o. N. University of Arizona, Ed. (Scientific Reports, 2017).
47. R. P. Baughman, U. Costabel, R. M. du Bois, Treatment of sarcoidosis. *Clin Chest Med* **29**, 533-548, ix-x (2008).
48. A. S. Morgenthau, M. C. Iannuzzi, Recent advances in sarcoidosis. *Chest* **139**, 174-182.
49. J. Harris, J. C. Hope, J. Keane, Tumor necrosis factor blockers influence macrophage responses to Mycobacterium tuberculosis. *The Journal of infectious diseases* **198**, 1842-1850 (2008).
50. O. Y. Saliu, C. Sofer, D. S. Stein, S. K. Schwander, R. S. Wallis, Tumor-necrosis-factor blockers: differential effects on mycobacterial immunity. *The Journal of infectious diseases* **194**, 486-492 (2006).
51. A. Mohan *et al.*, Transcriptional survey of alveolar macrophages in a murine model of chronic granulomatous inflammation reveals common themes with human sarcoidosis. *Am J Physiol Lung Cell Mol Physiol* **314**, L617-L625 (2018).
52. I. S. Jacob Sellares, Elliot D. Crouser, Marina A. Freudenberg, Mridu Gulati, Simon Hart, Erika Herzog, Martin Kolb, Thomas Weichhart, Wonder P. Drake, Ginger Spitzer, Noopur Singh, Daniel A. Culver New advances in the development of sarcoidosis models: a synopsis of a symposium sponsored by the Foundation for Sarcoidosis Research. *SARCOIDOSIS VASCULITIS AND DIFFUSE LUNG DISEASES* **35**, 2-4 (2018).
53. N. J. Sweiss *et al.*, Rheumatologic manifestations of sarcoidosis. *Semin Respir Crit Care Med* **31**, 463-473 (2010).
54. M. Drent, J. P. Cremers, T. L. Jansen, Pulmonology meets rheumatology in sarcoidosis: a review on the therapeutic approach. *Curr Opin Rheumatol* **26**, 276-284 (2014).
55. N. Casanova, T. Zhou, K. S. Knox, J. G. Garcia, Identifying Novel Biomarkers in Sarcoidosis Using Genome-Based Approaches. *Clin Chest Med* **36**, 621-630 (2015).
56. J. A. Bons *et al.*, in *Respir Med*. (England, 2007), vol. 101, pp. 1687-1695.
57. A. Rozy *et al.*, Inflammatory markers in the exhaled breath condensate of patients with pulmonary sarcoidosis. *J Physiol Pharmacol* **57 Suppl 4**, 335-340 (2006).
58. H. Ahmadzai *et al.*, Measurement of neopterin, TGF-beta(1) and ACE in the exhaled breath condensate of patients with sarcoidosis. *Journal of Breath Research* **7**, (2013).
59. S. Gungor *et al.*, Conventional markers in determination of activity of sarcoidosis. *Int Immunopharmacol* **25**, 174-179 (2015).
60. A. J. Tahmoush *et al.*, CSF-ACE activity in probable CNS neurosarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* **19**, 191-197 (2002).

61. B. Beneteau-Burnat, B. Baudin, Angiotensin-converting enzyme: clinical applications and laboratory investigations on serum and other biological fluids. *Critical reviews in clinical laboratory sciences* **28**, 337-356 (1991).
62. R. Baughman, *Sarcoidosis*. C. Lenfant, Ed., Lung Biology in Health Disease (Taylor and Francis Group, New York, NY, 2006), vol. 210.
63. B. Baudin, [Angiotensin I-converting enzyme (ACE) for sarcoidosis diagnosis]. *Pathologie-biologie* **53**, 183-188 (2005).
64. P. Alia, J. Mana, O. Capdevila, A. Alvarez, M. A. Navarro, in *Scand J Clin Lab Invest*. (Norway, 2005), vol. 65, pp. 691-697.
65. A. Floe, H. J. Hoffmann, P. H. Nissen, H. J. Moller, O. Hilberg, in *Dan Med J*. (Denmark, 2014), vol. 61, pp. A4815.
66. A. Pietinalho, K. Furuya, E. Yamaguchi, Y. Kawakami, O. Selroos, The angiotensin-converting enzyme DD gene is associated with poor prognosis in Finnish sarcoidosis patients. *Eur Respir J* **13**, 723-726 (1999).
67. A. Salazar, X. Pinto, J. Mana, Serum amyloid A and high-density lipoprotein cholesterol: serum markers of inflammation in sarcoidosis and other systemic disorders. *European journal of clinical investigation* **31**, 1070-1077 (2001).
68. E. Bargagli *et al.*, Human Chitotriosidase: a Sensitive Biomarker of Sarcoidosis. *Journal of Clinical Immunology* **33**, 264-270 (2013).
69. Y. Zhang *et al.*, in *Respir Res*. (England, 2013), vol. 14, pp. 18.
70. E. Bargagli, A. Mazzi, P. Rottoli, Markers of Inflammation in Sarcoidosis: Blood, Urine, BAL, Sputum, and Exhaled Gas. *Clinics in Chest Medicine* **29**, 445-458 (2008).
71. C. P. Herbort, N. A. Rao, M. Mochizuki, International criteria for the diagnosis of ocular sarcoidosis: results of the first International Workshop On Ocular Sarcoidosis (IWOS). *Ocular immunology and inflammation* **17**, 160-169 (2009).
72. C. Febvay *et al.*, Clinical features and diagnostic evaluation of 83 biopsy-proven sarcoid uveitis cases. *The British journal of ophthalmology*, (2015).
73. S. Miyoshi *et al.*, COmparative evaluation of serum markers in pulmonary sarcoidosis. *Chest* **137**, 1391-1397 (2010).
74. A. D. Vorselaars *et al.*, ACE and sIL-2R correlate with lung function improvement in sarcoidosis during methotrexate therapy. *Respir Med* **109**, 279-285 (2015).
75. J. C. Grutters *et al.*, Serum soluble interleukin-2 receptor measurement in patients with sarcoidosis: a clinical evaluation. *Chest* **124**, 186-195 (2003).
76. W. Wu *et al.*, in *J Transl Med*. (England, 2014), vol. 12, pp. 152.
77. G. Novelli, C. Ciccacci, P. Borgiani, M. Papaluca Amati, E. Abadie, Genetic tests and genomic biomarkers: regulation, qualification and validation. *Clin Cases Miner Bone Metab* **5**, 149-154 (2008).
78. S. E. Calvano *et al.*, A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032-1037 (2005).
79. M. S. Rolph, M. Sisavanh, S. M. Liu, C. R. Mackay, Clues to asthma pathogenesis from microarray expression studies. *Pharmacol Ther* **109**, 284-294 (2006).
80. M. Selman, A. Pardo, N. Kaminski, Idiopathic pulmonary fibrosis: aberrant recapitulation of developmental programs? *PLoS Med* **5**, e62 (2008).
81. J. D. Herazo-Maya *et al.*, Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci Transl Med* **5**, 205ra136 (2013).
82. T. Zhou, T. Wang, J. G. Garcia, Expression of nicotinamide phosphoribosyltransferase-influenced genes predicts recurrence-free survival in lung and breast cancers. *Sci Rep* **4**, 6107 (2014).
83. A. A. Desai *et al.*, A novel molecular signature for elevated tricuspid regurgitation velocity in sickle cell disease. *Am J Respir Crit Care Med* **186**, 359-368 (2012).

84. S. P. Pitroda *et al.*, Tumor endothelial inflammation predicts clinical outcome in diverse human cancers. *PLoS One* **7**, e46104 (2012).
85. M. K. Showe *et al.*, Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer research* **69**, 9202-9210 (2009).
86. M. J. Baine *et al.*, Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. *PLoS One* **6**, e17014 (2011).
87. E. C. Baechler *et al.*, Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A* **100**, 2610-2615 (2003).
88. A. Achiron, M. Gurevich, N. Friedman, N. Kaminski, M. Mandel, Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. *Ann Neurol* **55**, 410-417 (2004).
89. R. H. Segman *et al.*, Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors. *Molecular psychiatry* **10**, 500-513, 425 (2005).
90. D. F. Moore *et al.*, Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: a pilot investigation. *Circulation* **111**, 212-221 (2005).
91. J. E. Kontaraki, F. I. Parthenakis, E. G. Nyktari, A. P. Patrianakos, P. E. Vardas, Myocardial gene expression alterations in peripheral blood mononuclear cells of patients with idiopathic dilated cardiomyopathy. *Eur J Heart Fail* **12**, 541-548 (2010).
92. M. X. Pham *et al.*, Gene-Expression Profiling for Rejection Surveillance after Cardiac Transplantation. *N Engl J Med*.
93. M. P. Bihl *et al.*, Progressive pulmonary sarcoidosis--a fibroproliferative process potentially triggered by EGR-1 and IL-6. *Sarcoidosis Vasc Diffuse Lung Dis* **23**, 38-50 (2006).
94. T. Zhou *et al.*, in *PLoS One*. (United States, 2012), vol. 7, pp. e44818.
95. W. S. Loke, C. Herbert, P. S. Thomas, Sarcoidosis: Immunopathogenesis and Immunological Markers. *Int J Chronic Dis* **2013**, 928601 (2013).
96. C. Kneitz *et al.*, [Sarcoidosis in monozygotic twins]. *Dtsch Med Wochenschr* **120**, 867-873 (1995).
97. B. A. Rybicki, M. J. Maliarik, L. M. Poisson, M. C. Iannuzzi, Sarcoidosis and granuloma genes: a family-based study in African-Americans. *Eur Respir J* **24**, 251-257 (2004).
98. A. M. Levin *et al.*, Association of HLA-DRB1 with Sarcoidosis Susceptibility and Progression in African Americans. *Am J Respir Cell Mol Biol* **53**, 206-216 (2015).
99. J. P. Pandey, M. Frederick, TNF-alpha, IL1-beta, and immunoglobulin (GM and KM) gene polymorphisms in sarcoidosis. *Hum Immunol* **63**, 485-491 (2002).
100. S. Hofmann *et al.*, Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* **40**, 1103-1106 (2008).
101. F. Mrazek *et al.*, Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* **12**, 490-494 (2011).
102. A. M. Levin *et al.*, Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* **14**, 13-18 (2013).
103. M. D. Rossman *et al.*, HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am J Hum Genet* **73**, 720-735 (2003).
104. J. Grunewald, A. Eklund, O. Olerup, Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. *Am J Respir Crit Care Med* **169**, 696-702 (2004).
105. M. M. Miretti *et al.*, A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* **76**, 634-646 (2005).
106. A. Kishore *et al.*, Whole-exome sequencing identifies rare genetic variations in German families with pulmonary sarcoidosis. *Hum Genet* **137**, 705-716 (2018).

107. E. R. Gibney, C. M. Nolan, Epigenetics and gene expression. *Heredity (Edinb)* **105**, 4-13 (2010).
108. B. Schmidt *et al.*, SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer* **10**, 600 (2010).
109. I. V. Yang *et al.*, DNA Methylation Changes in Lung Immune Cells Are Associated with Granulomatous Lung Disease. *Am J Respir Cell Mol Biol* **60**, 96-105 (2019).
110. E. D. Crouser *et al.*, Differential expression of microRNA and predicted targets in pulmonary sarcoidosis. *Biochem Biophys Res Commun* **417**, 886-891 (2012).
111. C. Ascoli *et al.*, A Circulating Micro-RNA Signature Serves as a Diagnostic and Prognostic Indicator in Sarcoidosis. *Am J Respir Cell Mol Biol*, (2017).
112. N. Kachamakova-Trojanowska *et al.*, Molecular profiling of regulatory T cells in pulmonary sarcoidosis. *J Autoimmun* **94**, 56-69 (2018).
113. A. C. Nica, E. T. Dermitzakis, Expression quantitative trait loci: present and future. *Philos T R Soc B* **368**, (2013).
114. Y. Chen *et al.*, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429-435 (2008).
115. N. Casanova, Zhou,T., Gonzalez-Garay,M., Lussier, Y., Sweiss N., Ma, SF., Noth, I., Knox, K., Garcia, J.G.N., MicroRNA and Protein-Coding Gene Expression Analysis in Idiopathic Pulmonary Fibrosis Yields Novel Biomarker Signatures Predicting Survival. *Translational Research*, (2020).
116. L. Koth *et al.*, Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med* **184**, 1153-1163 (2011).
117. N. Casanova *et al.*, Low Dose Carbon Monoxide Exposure in Idiopathic Pulmonary Fibrosis Produces a CO Signature Comprised of Oxidative Phosphorylation Genes. *Sci Rep* **9**, 14802 (2019).
118. C. Swider *et al.*, Presence of mRNA for interferon-gamma (IFN-gamma) in blood mononuclear cells is associated with an active stage I sarcoidosis. *Clin Exp Immunol* **100**, 401-405 (1995).
119. H. E. Lockstone *et al.*, Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med* **181**, 1367-1375 (2010).
120. S. Singla *et al.*, Expression Profiling Elucidates a Molecular Signature for Pulmonary Hypertension in Sarcoidosis. *Pulmonary Circulation* **NA**, (2016).
121. S. Koo, F. M. Marty, L. R. Baden, Infectious complications associated with immunomodulating biologic agents. *Infectious disease clinics of North America* **24**, 285-306 (2010).
122. E. Gialafos *et al.*, Relationship between tumor necrosis factor-alpha (TNFA) gene polymorphisms and cardiac sarcoidosis. *In Vivo* **28**, 1125-1129 (2014).
123. J. C. Grutters *et al.*, Increased frequency of the uncommon tumor necrosis factor -857T allele in British and Dutch patients with sarcoidosis. *Am J Respir Crit Care Med* **165**, 1119-1124 (2002).
124. K. T. Amber, R. Bloom, U. Mrowietz, M. Hertl, TNF-α: a treatment target or cause of sarcoidosis? *J Eur Acad Dermatol Venereol* **29**, 2104-2111 (2015).
125. J. T. Rosenbaum *et al.*, Parallel Gene Expression Changes in Sarcoidosis Involving the Lacrimal Gland, Orbital Tissue, or BloodParallel Gene Expression Changes in SarcoidosisParallel Gene Expression Changes in Sarcoidosis. *JAMA Ophthalmology* **133**, 770-777 (2015).
126. L. L. Koth *et al.*, Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med* **184**, 1153-1163 (2011).
127. E. D. Crouser *et al.*, Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med* **179**, 929-938 (2009).
128. C. Wei *et al.*, Circulating miRNAs as potential marker for pulmonary hypertension. *PLoS One* **8**, e64396 (2013).
129. I. Sarrion *et al.*, Role of circulating miRNAs as biomarkers in idiopathic pulmonary arterial hypertension: possible relevance of miR-23a. *Oxid Med Cell Longev* **2015**, 792846 (2015).

130. Y. Xing *et al.*, MicroRNA-30c contributes to the development of hypoxia pulmonary hypertension by inhibiting platelet-derived growth factor receptor beta expression. *Int J Biochem Cell Biol* **64**, 155-166 (2015).
131. A. E. Hale, K. White, S. Y. Chan, Hypoxamirs in Pulmonary Hypertension: Breathing New Life into Pulmonary Vascular Research. *Cardiovasc Diagn Ther* **2**, 200-212 (2012).
132. J. Meloche *et al.*, MiR-223 Reverses Experimental Pulmonary Arterial Hypertension. *Am J Physiol Cell Physiol*, ajpcell 00149 02015 (2015).
133. M. Cao *et al.*, MiR-23a regulates TGF-beta-induced epithelial-mesenchymal transition by targeting E-cadherin in lung cancer cells. *Int J Oncol* **41**, 869-875 (2012).
134. K. Zhong, K. Chen, L. Han, B. Li, MicroRNA-30b/c inhibits non-small cell lung cancer cell proliferation by targeting Rab18. *BMC Cancer* **14**, 703 (2014).
135. Y. Yang *et al.*, The role of microRNA in human lung squamous cell carcinoma. *Cancer Genet Cytogenet* **200**, 127-133 (2010).
136. H. Liang *et al.*, MicroRNA-223 delivered by platelet-derived microvesicles promotes lung cancer cell invasion via targeting tumor suppressor EPB41L3. *Mol Cancer* **14**, 58 (2015).
137. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
138. T. Zhou *et al.*, Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. *PLoS One* **7**, e44818 (2012).
139. J. T. Rosenbaum *et al.*, Hypothesis: sarcoidosis is a STAT1-mediated disease. *Clin Immunol* **132**, 174-183 (2009).
140. S. Muro *et al.*, Expression of IL-15 in inflammatory pulmonary diseases. *J Allergy Clin Immunol* **108**, 970-975 (2001).
141. H. S. Kim *et al.*, Association of interleukin 23 receptor gene with sarcoidosis. *Dis Markers* **31**, 17-24 (2011).
142. S. M. Sharma *et al.*, Insights in to the pathogenesis of axial spondyloarthropathy based on gene expression profiles. *Arthritis Res Ther* **11**, R168 (2009).
143. H. M. diagnostics. (2019).
144. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
145. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
146. A. Kamburov, U. Stelzl, H. Lehrach, R. Herwig, The ConsensusPathDB interaction database: 2013 update. *Nucleic acids research* **41**, D793-800 (2013).
147. R. Valentonyte *et al.*, Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* **37**, 357-364 (2005).
148. T. Zhou, Casanova N., Puladi, N., Wang, T., Lussier, Y., Kox, S.K., Garcia, J.G.N, in *Scientific Reports*. (2017).
149. M. Katz, I. Amit, Y. Yarden, Regulation of MAPKs by growth factors and receptor tyrosine kinases. *Biochim Biophys Acta* **1773**, 1161-1176 (2007).