The landscape of transcriptomic and proteomic studies in sarcoidosis

Copyright ©The authors 2022

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

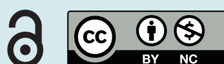
Received: 3 Nov 2021
Accepted: 16 Nov 2021

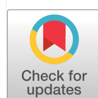
To the Editor:

Sarcoidosis is a systemic disease with gene/protein expression patterns that may be different among different tissues, based on the presence or absence of granulomas, and on subphenotypes with progressive or nonprogressive disease manifestations. There is a growing body of data evaluating global transcriptomic changes across multiple tissue compartments in sarcoidosis. However, whether similar biological pathways are involved is unknown. Furthermore, an understanding of the transcriptional impact on the proteome is required to validate molecular pathways driving heterogeneity in sarcoidosis. The purpose of this study was to compare biological inferences from published datasets and explore the compartment specificity of these responses in sarcoidosis. Common pathways identified across datasets or tissue types may serve as convenient biomarkers and could lead to the discovery of novel therapeutic targets.

We identified published sarcoidosis studies of differentially expressed genes (DEGs) (*via* microarray, bulk or single-cell RNA sequencing) or differentially abundant proteins (DAPs) (*via* unbiased proteome-wide screens) through PubMed and Gene Expression Omnibus (GEO). These studies compared cases with controls or compared progressive *versus* nonprogressive pulmonary sarcoidosis phenotypes in peripheral blood or peripheral blood mononuclear cells (PBMCs), bronchoalveolar lavage (BAL) cells or fluid, lung tissue, microdissected lung granulomas, lymph nodes, or an established *in vitro* granuloma model (stimulated PBMCs). The definition of progressive *versus* nonprogressive pulmonary sarcoidosis varied among studies but overall, progressive sarcoidosis referred to worsening chest imaging, need for systemic anti-inflammatory therapy and/or worsening lung function or symptoms. When DEG/DAP lists were not already available as supplementary information in the publication or did not originate from our laboratories, we downloaded the expression data as processed by the original authors from GEO and used the R package DEseq2 for RNA sequencing data or limma for microarray data to compute a DEG list, controlling for Benjamini and Hochberg's false discovery rate (FDR) <0.1 and compared to results found when using uncorrected p-value <0.05. Significantly enriched pathways found for each DEG/DAP independently using Ingenuity Pathway Analysis (QIAGEN Inc.) at FDR <0.05 were summarised in a visualisation across studies using the R pheatmap package.

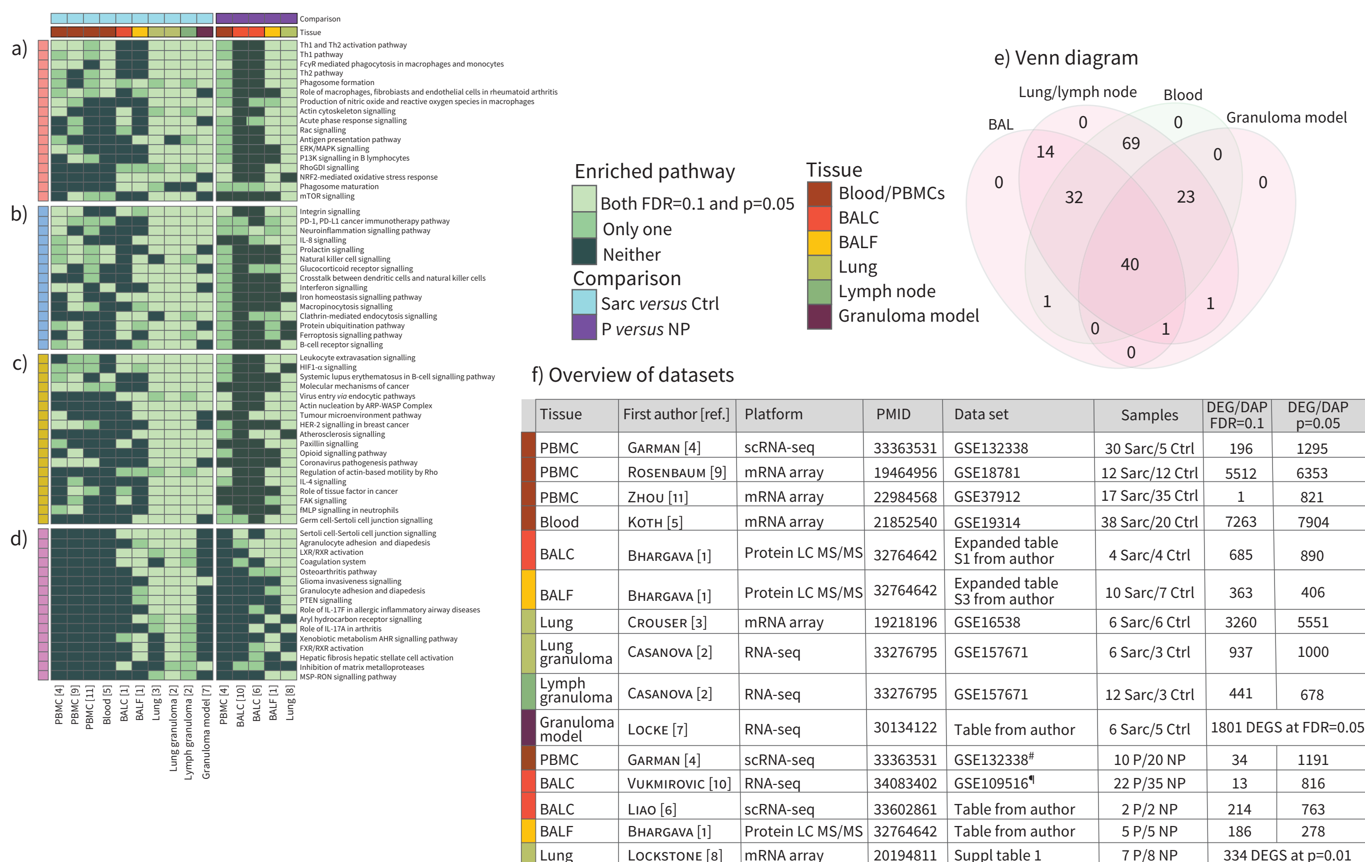
Using 15 datasets [1–11] with 10 datasets comparing sarcoidosis to control and five comparing sarcoidosis phenotypes (progressive *versus* nonprogressive) (figure 1f), *de novo* pathway analyses revealed 549 canonical pathways mapping to DEGs or DAPs. Of these, 181 pathways were present in more than four of 15 datasets, several of which have previously been investigated (more than three PubMed publications) (figure 1a) but some not well studied thus far (three or fewer PubMed publications) (figure 1b), such as integrin signalling, interleukin (IL)-8 signalling and neuroinflammation signalling. All but one of the 181 overlapping pathways were present in sarcoidosis lungs or lymph nodes compared to controls (figure 1e). 164 (91%) of these pathways were also identified in peripheral blood (figure 1c). In contrast, 16 (9%) canonical pathways were only detected in lungs or lymph nodes or BAL but not in peripheral blood (figure 1d), including mechanistic target of rapamycin signalling, phagosome formation, phagosome maturation, Janus kinase–signal tranducer and activator of transcription or Rho–Rac kinase signalling. Among the 181 pathways distinguishing sarcoidosis from controls, 161 (89%) were also detected in the progressive *versus* nonprogressive comparison. There were no pathways specific to the progressive *versus* nonprogressive comparison that were not present in the sarcoidosis *versus* control comparison. The pathways enriched in the *in vitro* granuloma model were all represented in lungs, lymph nodes or BAL, and 97% were represented in peripheral blood.





Shareable abstract (@ERSpublications)
Multiple overlapping pathways are identified in tissue, BAL cells, PBMCs and a sarcoidosis *in vitro* granuloma model. Inferences from omic studies are constrained by small sample sizes. Studies comparing differences between sarcoidosis phenotypes are needed. https://bit.ly/30NaHz4

Cite this article as: Bhargava M, Liao S-Y, Crouser ED, *et al*. The landscape of transcriptomic and proteomic studies in sarcoidosis. *ERJ Open Res* 2022; 8: 00621-2021 [DOI: 10.1183/23120541.00621-2021].



Tissue	First author [ref.]	Platform	PMID	Data set	Samples	DEG/DAP FDR=0.1	DEG/DAP p=0.05
PBMC	Garman [4]	scRNA-seq	33363531	GSE132338	30 Sarc/5 Ctrl	196	1295
PBMC	Rosenbaum [9]	mRNA array	19464956	GSE18781	12 Sarc/12 Ctrl	5512	6353
PBMC	Zhou [11]	mRNA array	22984568	GSE37912	17 Sarc/35 Ctrl	1	821
Blood	Koth [5]	mRNA array	21852540	GSE19314	38 Sarc/20 Ctrl	7263	7904
BALC	Bhargava [1]	Protein LC MS/MS	32764642	Expanded table S1 from author	4 Sarc/4 Ctrl	685	890
BALF	Bhargava [1]	Protein LC MS/MS	32764642	Expanded table S3 from author	10 Sarc/7 Ctrl	363	406
Lung	Crouser [3]	mRNA array	19218196	GSE16538	6 Sarc/6 Ctrl	3260	5551
Lung granuloma	Casanova [2]	RNA-seq	33276795	GSE157671	6 Sarc/3 Ctrl	937	1000
Lymph granuloma	Casanova [2]	RNA-seq	33276795	GSE157671	12 Sarc/3 Ctrl	441	678
Granuloma model	Locke [7]	RNA-seq	30134122	Table from author	6 Sarc/5 Ctrl	1801 DEGS at FDR=0.05	
PBMC	Garman [4]	scRNA-seq	33363531	GSE132338#	10 P/20 NP	34	1191
BALC	Vukmirovic [10]	RNA-seq	34083402	GSE109516¶	22 P/35 NP	13	816
BALC	Liao [6]	scRNA-seq	33602861	Table from author	2 P/2 NP	214	763
BALF	Bhargava [1]	Protein LC MS/MS	32764642	Table from author	5 P/5 NP	186	278
Lung	Lockstone [8]	mRNA array	20194811	Suppl table 1	7 P/8 NP	334 DEGS at p=0.01	

FIGURE 1 Summary of existing transcriptional and proteomic studies in sarcoidosis: 10 datasets comparing sarcoidosis (Sarc) with healthy controls (Ctrl) and five comparing progressive (P) to nonprogressive (NP) sarcoidosis were analysed. Differentially expressed genes (DEGs) and differentially abundant proteins (DAPs) were used for "core analysis" with Ingenuity Pathway Analysis (IPA) to identify over-represented canonical pathways for each dataset controlling for a false-discovery rate (FDR) at <0.05. IPA "comparison analysis" tool was used to determine the common canonical pathways across various datasets studied. We focused on pathways observed in at least five datasets, showing the top 50 overall and 16 not found in blood. a) The canonical pathways that were well established based on prior studies (more than three PubMed publications); b) pathways that have limited studies published to date in sarcoidosis (three or fewer PubMed publications); c) some of the pathways common to lung/lymph node tissue, bronchoalveolar lavage (BAL) and blood or peripheral blood mononuclear cells (PBMCs) and the *in vitro* granuloma model; d) pathways only detected in the BAL/lung/lymph node but not in peripheral blood cells and to a lesser degree in the *in vitro* granuloma model; e) the numbers of overlapping pathways among different tissues presented as a Venn diagram; and f) an overview of the referenced datasets. #: P (active) *versus* NP (non-active); ¶: untreated P (chronic Stage II/III) *versus* NP (non-acute or remitting Stage 0/I); Th: helper T-cell; FcγR: IgG receptor; ERK: extracellular signal-regulated kinase; MAPK: mitogen-activated protein kinase; PI3K: phosphoinositide 3-kinase; GDI: GDP dissociation inhibitor; mTOR: mechanistic target of rapamycin; PD-L1: PD-1 ligand; IL: interleukin; HIF: hypoxia-inducible factor; HER: human epidermal growth factor receptor; FAK: focal adhesion kinase; fMLP: *N*-formyl-methionyl-leucyl-phenylalanine; LXR: liver X receptor; RXR: retinoid X receptor; PTEN: phosphatase and tensin homologue; AHR: aryl hydrocarbon receptor; FXR: farnesoid X receptor; BALC: bronchoalveolar lavage cells; BALF: bronchoalveolar lavage fluid; sc: single-cell; RNA-seq: RNA sequencing; LC MS/MS: liquid chromatography–tandem mass spectrometry; PMID: PubMed identifier.

Our integrative analysis identified many canonical pathways apparent in multiple tissue compartments and those unique to a specific compartment. Most studies compared sarcoidosis to controls, whereas mechanisms of progression are best studied when comparing sarcoidosis phenotypes. Many pathways observed in our analysis were well established in sarcoidosis pathogenesis, such as antigen presentation, helper T-cell (Th)1 and Th2 activation, *etc.*, as noted above and in figure 1. Some of the canonical pathways were identified only in BAL, lungs or lymph nodes, reflecting compartment-specific granulomatous inflammatory mechanisms. Other pathways were present in blood and BAL, lungs or lymph nodes, reflecting systemic immune alterations. The well-established *in vitro* granuloma model demonstrates substantial overlap with diseased lung and lymph node tissues, reflective of *in vivo* granulomatous inflammation. For the protein dataset, the DAP map included fewer canonical pathways, likely due to a smaller number of proteins identified by contemporary technologies. However, altered protein pathways could be high-priority targets for testing as therapeutic targets or for further characterisation by classical molecular techniques.

Our analyses revealed new and novel pathways in sarcoidosis, highlighting the power of "unbiased discovery." For example, neuroinflammatory response has not been previously well studied in sarcoidosis. However, neuroinflammation signalling was identified in all compartments and might implicate neurotoxic cytokines or ILs [12]. Likewise, the prolactin signalling pathway, which regulates pro- and anti-inflammatory cytokines in the context of neural damage [13], was detected in all tissues. It is interesting to speculate that neuroinflammatory pathways might contribute to neurocognitive manifestations commonly observed in pulmonary sarcoidosis. Other pathways identified, such as integrin signalling, might participate in granuloma formation, promoting cell adhesion, and regulating immune cell survival, polarity, proliferation in pulmonary sarcoidosis [14]. Our novel discovery of enhanced caveolae-mediated endocytic signalling elements implies altered environmental antigen processing that could promote granulomatous inflammation. Similarly, aryl hydrocarbon receptor (AHR) signalling is emerging as an important immunoregulator in response to endogenous and exogenous ligands; AHR regulates T-cell responses at multiple levels, including T-cell fate, induction of $CD4^+$ regulatory T-cells and Th17s, and Th22 cell differentiation, balancing effector and regulatory cells [15]. AHR signalling is implicated in other granulomatous diseases such as Crohn's disease. Enhanced antigen processing (*e.g.* caveolae-mediated and phagolysosomal) and related T-cell responses were expected, and likely drive a strong adaptive immune response; our analysis also provides evidence for aberrant innate immune responses that likely contribute to sarcoidosis development and progression. The IL-8 signalling pathway, promoting granulocytic inflammation and phagocytosis, is increased in sarcoidosis *versus* controls and distinguished progressive from nonprogressive based on DEGs and DAPs. Supporting this finding, higher IL-8 levels are reported in chronic sarcoidosis.

Our analysis was restricted to datasets that were publicly available in a format allowing identification of differentially expressed features and thus may have excluded some important studies not available in this format. In addition, as we only had summary statistics available instead of individual data with detailed information, we cannot account for differences in race and ethnicity. While the definition for sarcoidosis was uniform among studies, pulmonary phenotypes and the definition of progressive *versus* nonprogressive varied among studies; this may have limited our detection of phenotypic pathways. We summarised enriched pathways across studies rather than comparing DEG and DAP lists directly or attempting to analyse the studies as a single dataset to minimise biases due to variable sample numbers, the number of features assayed by each platform or other sources of variance across studies. Although differences in sample size implicitly bias the size of the resulting DEG/DAP list, the pathway enrichment statistic accounts for input list size. However, emphasis should be placed on detection of an enriched pathway across multiple studies, rather than the lack of detection, since no detection in a given study could be attributed to lack of power, not lack of biological relevance. Our analysis highlights the paucity of comprehensive studies integrating omics and systems biology in sarcoidosis. Pathways involved in granulomatous inflammation can be identified in lungs, lymph nodes, BAL cells or fluid, and/or blood. Evaluation of each compartment may answer specific questions; blood or BAL may be useful to develop biomarkers, while lung and lymph nodes may be superior for investigating disease biology. We have demonstrated that findings from one compartment, such as blood, can be seen in respiratory tract specimens to define pathways shared and distinct across compartments. The *in vitro* granuloma model demonstrates pathways that overlap with both lung and blood, providing a highly relevant and convenient tool to improve knowledge in sarcoidosis. As most studies examined sarcoidosis *versus* controls and did not consider pulmonary phenotypes, the latter is an urgent unmet research need as pulmonary sarcoidosis results in significant sarcoidosis-related mortality; carefully designed studies designed to narrow existing knowledge gaps in mechanisms driving sarcoidosis heterogeneity are needed.

Maneesh Bhargava[1,7], Shu-Yi Liao[2,3,4,7], Elliott D. Crouser[5], Lisa A. Maier[2,3,4] and Sonia M. Leach[6]

[1]Division of Pulmonary, Critical Care and Sleep Medicine, Dept of Medicine, University of Minnesota Medical School, Minneapolis, MN, USA. [2]Division of Environmental and Occupational Health Sciences, Dept of Medicine, National Jewish Health, Denver, CO, USA. [3]Division of Pulmonary Sciences and Critical Care Medicine, Dept of Medicine, University of Colorado Denver, Aurora, CO, USA. [4]Dept of Environmental and Occupational Health, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA. [5]Division of Pulmonary and Critical Care Medicine, The Ohio State University, Columbus, OH, USA. [6]Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA. [7]These authors contributed equally.

Corresponding author: Maneesh Bhargava (bharg005@umn.edu)

Provenance: Submitted article, peer reviewed.

Conflict of interest: None declared.

Support statement: This study was supported by a Foundation for Sarcoidosis Research grant to E.D. Crouser and S-Y. Liao; and National Institutes of Health grants HL077466 to E.D. Crouser, HL112695 and HL112707 to L.A. Maier, HL140357 to L.A. Maier and S-Y. Liao, and HL153613 to M. Bhargava and L.A. Maier. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Bhargava M, Viken KJ, Barkes B, *et al.* Novel protein pathways in development and progression of pulmonary sarcoidosis. *Sci Rep* 2020; 10: 13282.
2 Casanova NG, Gonzalez-Garay ML, Sun B, *et al.* Differential transcriptomics in sarcoidosis lung and lymph node granulomas with comparisons to pathogen-specific granulomas. *Respir Res* 2020; 21: 321.
3 Crouser ED, Culver DA, Knox KS, *et al.* Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2009; 179: 929–938.
4 Garman L, Pelikan RC, Rasmussen A, *et al.* Single cell transcriptomics implicate novel monocyte and T cell immune dysregulation in sarcoidosis. *Front Immunol* 2020; 11: 567342.
5 Koth LL, Solberg OD, Peng JC, *et al.* Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med* 2011; 184: 1153–1163.
6 Liao SY, Atif SM, Mould K, *et al.* Single-cell RNA sequencing identifies macrophage transcriptional heterogeneities in granulomatous diseases. *Eur Respir J* 2021; 57: 2003794.
7 Locke LW, Crouser ED, White P, *et al.* IL-13-regulated macrophage polarization during granuloma formation in an *in vitro* human sarcoidosis model. *Am J Respir Cell Mol Biol* 2019; 60: 84–95.
8 Lockstone HE, Sanderson S, Kulakova N, *et al.* Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2010; 181: 1367–1375.
9 Rosenbaum JT, Pasadhika S, Crouser ED, *et al.* Hypothesis: sarcoidosis is a STAT1-mediated disease. *Clin Immunol* 2009; 132: 174–183.
10 Vukmirovic M, Yan X, Gibson KF, *et al.* Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. *Eur Respir J* 2021; 58: 2002950.
11 Zhou T, Zhang W, Sweiss NJ, *et al.* Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. *PLoS One* 2012; 7: e44818.
12 Becher B, Spath S, Goverman J. Cytokine networks in neuroinflammation. *Nat Rev Immunol* 2017; 17: 49–59.
13 Ramos-Martinez E, Ramos-Martinez I, Molina-Salinas G, *et al.* The role of prolactin in central nervous system inflammation. *Rev Neurosci* 2021; 32: 323–340.
14 Teoh CM, Tan SS, Tran T. Integrins as therapeutic targets for respiratory diseases. *Curr Mol Med* 2015; 15: 714–734.
15 Gutierrez-Vazquez C, Quintana FJ. Regulation of the immune response by the Aryl hydrocarbon receptor. *Immunity* 2018; 48: 19–33.