Early View

Research letter

The Landscape of Transcriptomics and Proteomics Studies in Sarcoidosis

Maneesh Bhargava, Shu-Yi Liao, Elliott D. Crouser, Lisa A. Maier, Sonia M. Leach

Please cite this article as: Bhargava M, Liao S-Y, Crouser ED, *et al.* The Landscape of Transcriptomics and Proteomics Studies in Sarcoidosis. *ERJ Open Res* 2021; in press (<https://doi.org/10.1183/23120541.00621-2021>).

This manuscript has recently been accepted for publication in the *ERJ Open Research*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJOR online.

Copyright ©The authors 2021. 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

The Landscape of Transcriptomics and Proteomics Studies in Sarcoidosis
Maneesh Bhargava^{1*}, Shu-Yi Liao^{2,3,4*}, Elliott D. Crouser⁵, Lisa A. Maier^{2,3,4}, Sonia M. Leach⁶

¹Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Minnesota Medical School

²Division of Environmental and Occupational Health Sciences, Department of Medicine, National Jewish Health, Denver, CO, United States

³Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, United States

⁴Department of Environmental and Occupational Health Colorado School of Public Health, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, United States

⁵Division of Pulmonary and Critical Care Medicine, The Ohio State University, Columbus, OH

⁶Center for Genes, Environmental and Health, National Jewish Health, Denver, CO

*These authors contributed equally to this work.

Corresponding Author: Maneesh Bhargava MD, PhD
University of Minnesota
420 Delaware St SE, Minneapolis, MN
Phone: (612) 626 9338
Fax: (612)625 2174
E-mail: bharg005@umn.edu

To the Editor:

Sarcoidosis is a systemic disease, and the gene/protein expression patterns may be different among different tissues, particularly based on the presence or absence of granulomas and also based on subphenotypes with progressive or non-progressive 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 emerging is unknown. Furthermore, the transcriptional impact on the proteome is required to validate molecular pathways driving heterogeneity in sarcoidosis. The purpose of this study is 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 (DEG, via microarray, bulk or single-cell RNA-seq) or differentially abundant proteins (DAP, via unbiased proteome-wide screens) through PubMed and Gene Expression Omnibus (GEO). These studies compared cases with controls or compared progressive (P) vs. non-progressive (NP) pulmonary sarcoidosis phenotypes in peripheral blood/blood mononuclear cells (PB/PBMC), bronchoalveolar lavage (BAL) cells or fluid, lung tissue, micro dissected lung granulomas, lymph node, or an established *in vitro* granuloma model (stimulated PBMCs). The definition for P vs. NP pulmonary sarcoidosis varies among studies, but overall, progressive sarcoidosis refers to worsening chest images, need for systemic anti-inflammatory therapy and/or worsening lung function/symptoms. When DEG/DAP lists were not already available as supplementary information in the publication or did not originate from our labs, 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 summarized in a visualization across studies using the R heatmap package.

Using 15 datasets¹⁻¹¹ with 10 datasets comparing sarcoidosis to control and 5 comparing sarcoidosis phenotypes (P vs. NP, Figure 1F), de novo pathway analyses revealed 549 canonical pathways mapping to DEG or DAP. Of these, 229 pathways were present in at least four of 15 datasets, several of which have previously been investigated (>3 PubMed publications, Figure 1A), but some not well studied thus far (≤ 3 PubMed publications, Figure 1B), such as integrin signaling, IL-8 signaling, and neuroinflammation signaling. The overlapping 229 pathways were present in sarcoidosis lung/lymph node compared to controls (Figure 1E). 191 (83%) of these pathways were also identified in PB (Figure 1C). In contrast, 38 (10%) canonical pathways were only detected in lung/lymph node or BAL but not in PB (Figure 1D), including mTOR signaling, phagosome formation, phagosome maturation, JAK-Stat, or Rho-Rac kinase signaling. Among the 229 pathways distinguishing sarcoidosis from controls, 177 (77%) were also detected in the P vs. NP comparison. There were no pathways specific to the P vs. NP comparison that were not present in the sarcoidosis vs. control comparison. The pathways enriched in the *in vitro* granuloma model were all represented in lung/lymph node/BAL, and 88% were represented in PB.

Our integrative analysis identified many canonical pathways apparent in multiple tissue compartments and also 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, Th1, and Th2 activation, etc., as noted above and in Figure 1. Some of the canonical pathways were identified only in BAL/lung/lymph nodes, reflecting compartment-specific granulomatous inflammatory mechanisms. Other pathways are present in blood and BAL/lung/lymph node, 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 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 or for further characterization by classical molecular techniques.

Our analyses revealed new and novel pathways in sarcoidosis, highlighting the power of "unbiased discovery." For example, neuroinflammation signaling has not been previously implicated in sarcoidosis. However, the "neuroinflammatory response" was identified in all compartments and might implicate neurotoxic cytokines/interleukins¹². Likewise, the prolactin signaling 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 signaling, might participate in granuloma formation, promoting cell adhesion and regulating immune cell survival, polarity, proliferation in pulmonary sarcoidosis¹⁴. Our novel discovery of enhanced caveolae-mediated endocytic signaling elements implies altered environmental antigen processing that could promote granulomatous inflammation. Similarly, Aryl hydrocarbon (AhR) receptor signaling 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 + Treg/Th17s and Th22 cell differentiation, balancing effector, and regulatory cells¹⁵. AhR signaling is implicated in other granulomatous diseases such as Crohn's disease. Enhanced antigen processing (e.g., caveolae-mediated, 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 signaling pathway, promoting granulocytic inflammation

and phagocytosis, is increased in sarcoidosis vs. controls and distinguished P vs NP based on DEG/DAP. Supporting this finding, higher IL-8 levels are reported in chronic sarcoidosis.

Our analysis was restricted to datasets publicly available in a format allowing identification of differentially expressed features and thus may have excluded some important studies without this format. In addition, as we only have 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 P vs. NP varied among studies; this may have limited our detection of phenotypic pathways. We summarized enriched pathways across studies rather than comparing DEG/DAP lists directly or attempting to analyze the studies as a single dataset to minimize 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 lung/lymph node/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/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 examine sarcoidosis vs. controls and do 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.

References

1. Bhargava M, Viken KJ, Barkes B, Griffin TJ, Gillespie M, Jagtap PD, Sajulga R, Peterson EJ, Dincer HE, Li L, Restrepo CI, O'Connor BP, Fingerlin TE, Perlman DM, Maier LA. Novel protein pathways in development and progression of pulmonary sarcoidosis. *Sci Rep.* 2020;10(1):13282. Epub 2020/08/09. doi: 10.1038/s41598-020-69281-8. PubMed PMID: 32764642; PMCID: PMC7413390.
2. Casanova NG, Gonzalez-Garay ML, Sun B, Bime C, Sun X, Knox KS, Crouser ED, Sammani N, Gonzales T, Natt B, Chaudhary S, Lussier Y, Garcia JGN. Differential transcriptomics in sarcoidosis lung and lymph node granulomas with comparisons to pathogen-specific granulomas. *Respir Res.* 2020;21(1):321. Epub 2020/12/06. doi: 10.1186/s12931-020-01537-3. PubMed PMID: 33276795; PMCID: PMC7716494.
3. Crouser ED, Culver DA, Knox KS, Julian MW, Shao G, Abraham S, Liyanarachchi S, Macre JE, Wewers MD, Gavrilin MA, Ross P, Abbas A, Eng C. Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med.* 2009;179(10):929-38. Epub 2009/02/17. doi: 10.1164/rccm.200803-490OC. PubMed PMID: 19218196; PMCID: PMC2684019.
4. Garman L, Pelikan RC, Rasmussen A, Lareau CA, Savoy KA, Deshmukh US, Bagavant H, Levin AM, Daouk S, Drake WP, Montgomery CG. Single Cell Transcriptomics Implicate Novel Monocyte and T Cell Immune Dysregulation in Sarcoidosis. *Front Immunol.* 2020;11:567342. Epub 2020/12/29. doi: 10.3389/fimmu.2020.567342. PubMed PMID: 33363531; PMCID: PMC7753017.
5. Koth LL, Solberg OD, Peng JC, Bhakta NR, Nguyen CP, Woodruff PG. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med.* 2011;184(10):1153-63. Epub 2011/08/20. doi: 10.1164/rccm.201106-1143OC. PubMed PMID: 21852540; PMCID: PMC3262024.
6. Liao SY, Atif SM, Mould K, Konigsberg IR, Fu R, Davidson E, Li L, Fontenot AP, Maier LA, Yang IV. Single-cell RNA sequencing identifies macrophage transcriptional heterogeneities in granulomatous diseases. *Eur Respir J.* 2021;57(6). Epub 2021/02/20. doi: 10.1183/13993003.03794-2020. PubMed PMID: 33602861; PMCID: PMC8247459.
7. Locke LW, Crouser ED, White P, Julian MW, Caceres EG, Papp AC, Le VT, Sadee W, Schlesinger LS. IL-13-regulated Macrophage Polarization during Granuloma Formation in an In Vitro Human Sarcoidosis Model. *Am J Respir Cell Mol Biol.* 2019;60(1):84-95. Epub 2018/08/23. doi: 10.1165/rcmb.2018-0053OC. PubMed PMID: 30134122; PMCID: PMC6348723.
8. Lockstone HE, Sanderson S, Kulakova N, Baban D, Leonard A, Kok WL, McGowan S, McMichael AJ, Ho LP. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med.* 2010;181(12):1367-75. Epub 2010/03/03. doi: 10.1164/rccm.200912-1855OC. PubMed PMID: 20194811.
9. Rosenbaum JT, Pasadhika S, Crouser ED, Choi D, Harrington CA, Lewis JA, Austin CR, Diebel TN, Vance EE, Braziel RM, Smith JR, Planck SR. Hypothesis: sarcoidosis is a STAT1-mediated disease. *Clin Immunol.* 2009;132(2):174-83. Epub 2009/05/26. doi: 10.1016/j.clim.2009.04.010. PubMed PMID: 19464956; PMCID: PMC2733945.
10. Vukmirovic M, Yan X, Gibson KF, Gulati M, Schupp JC, Deluliis G, Adams TS, Hu B, Mihaljiniec A, Woolard TN, Lynn H, Emeagwali N, Herzog EL, Chen ES, Morris A, Leader JK, Zhang Y, Garcia JGN, Maier LA, Collman RG, Drake WP, Becich MJ, Hochheiser H, Wisniewski SR, Benos PV, Moller DR, Prasse A, Koth LL, Kaminski N, Investigators G. Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. *Eur Respir J.* 2021. Epub 2021/06/05. doi: 10.1183/13993003.02950-2020. PubMed PMID: 34083402.
11. Zhou T, Zhang W, Sweiss NJ, Chen ES, Moller DR, Knox KS, Ma SF, Wade MS, Noth I, Machado RF, Garcia JG. Peripheral blood gene expression as a novel genomic biomarker in

- complicated sarcoidosis. PLoS One. 2012;7(9):e44818. Epub 2012/09/18. doi: 10.1371/journal.pone.0044818. PubMed PMID: 22984568; PMCID: PMC3440319.
12. Becher B, Spath S, Goverman J. Cytokine networks in neuroinflammation. *Nat Rev Immunol.* 2017;17(1):49-59. Epub 2016/12/06. doi: 10.1038/nri.2016.123. PubMed PMID: 27916979.
13. Ramos-Martinez E, Ramos-Martinez I, Molina-Salinas G, Zepeda-Ruiz WA, Cerbon M. The role of prolactin in central nervous system inflammation. *Rev Neurosci.* 2021;32(3):323-40. Epub 2021/03/05. doi: 10.1515/revneuro-2020-0082. PubMed PMID: 33661585.
14. Teoh CM, Tan SS, Tran T. Integrins as Therapeutic Targets for Respiratory Diseases. *Curr Mol Med.* 2015;15(8):714-34. Epub 2015/09/24. doi: 10.2174/1566524015666150921105339. PubMed PMID: 26391549; PMCID: PMC5427774.
15. Gutierrez-Vazquez C, Quintana FJ. Regulation of the Immune Response by the Aryl Hydrocarbon Receptor. *Immunity.* 2018;48(1):19-33. Epub 2018/01/19. doi: 10.1016/j.immuni.2017.12.012. PubMed PMID: 29343438; PMCID: PMC5777317.

Conflict of Interest: None of the authors have any relevant financial conflicts to disclose.

Figure 1: Summary of Existing Transcriptional and Proteomic Studies in Sarcoidosis: Ten datasets comparing sarcoidosis with healthy controls and five comparing progressive to non-progressive sarcoidosis were analyzed. Differentially expressed genes (DEG) and differentially abundant proteins (DAP) were used for 'core analysis' with Ingenuity Pathway Analysis to identify over-represented canonical pathways for each dataset controlling for FDR at <0.05 . IPA 'comparison analysis' tool was used to determine the common canonical pathways across various datasets studied. We focus on pathways observed in at least four datasets. A) highlights the canonical pathways that were well established based on prior studies (>3 PubMed publications); B) shows the pathways that have limited studies published to date in sarcoidosis (≤ 3 PubMed publications); C) represents some of the pathways common to lung/lymph node tissue, bronchoalveolar lavage, and blood/peripheral blood mononuclear cells and the *in vitro* granuloma model; D) includes 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) shows the numbers of overlapping pathways among different tissues presented as a Venn Diagram; and F) provides an overview of the referenced datasets. *FDR: false discovery rate; BALC: bronchoalveolar lavage cells; BALF: bronchoalveolar lavage fluid; Sarc: sarcoidosis; P vs NP: progressive versus non-progressive sarcoidosis*

E. Venn Diagram

F. Overview of Datasets

Tissue	Author	Platform	Reference	Data Set	Samples	DEG/DAP FDR=0.1	DEG/DAP p=0.05
PBMC	Garman 2020	scRNA-seq	PMID:33363531	GSE132338	30 Sarc/5 Ctrl	196	1295
PBMC	Rosenbaum 2009	mRNA array	PMID:19464956	GSE18781	12 Sarc/12 Ctrl	5512	6353
PBMC	Zhou 2012	mRNA array	PMID:22984568	GSE37912	17 Sarc/35 Ctrl	1	821
Blood	Koth 2011	mRNA array	PMID:21852540	GSE19314	38 Sarc/20 Ctrl	7263	7904
BAL Cells	Bhargava 2020	protein LC MS/MS	PMID:32764642	Expanded Suppl Table S1 from author	4 Sarc/4 Ctrl	685	890
BAL Fluid	Bhargava 2020	protein LC MS/MS	PMID:32764642	Expanded Suppl Table S3 from author	10 Sarc/7 Ctrl	363	406
Lung	Crouser 2009	mRNA array	PMID:19218196	GSE16538	6 Sarc/6 Ctrl	3260	5551
Lung	Casanova 2020	RNA-seq	PMID:33276795	GSE157671	6 Sarc/3 Ctrl	937	1000
Lymph	Casanova 2020	RNA-seq	PMID:33276795	GSE157671	12 Sarc/3 Ctrl	441	678
Granuloma Model	Locke 2019	RNA-seq	PMID:30134122	Table from author	6 Sarc/5 Ctrl	1801 DEG at FDR=0.05	
PBMC	Garman 2020	scRNA-seq	PMID:33363531	GSE132338	10 P/20 NP	34	1191
BAL Cells	GRADS 2020	RNA-seq	PMID: 34083402	P vs NP subset**	22 P/35 NP	13	816
BAL Cells	Liao 2021	scRNA-seq	PMID:33602861	Table from author	2 P/2 NP	214	763
BAL Fluid	Bhargava 2020	protein LC MS/MS	PMID:32764642	Table from author	5 P/5 NP	186	278
Lung	Lockstone 2010	mRNA array	PMID:20194811	Suppl Table 1	7 P/8 NP	334 DEG at p=0.01	