

## Application of “Omics” and Systems Biology to Sarcoidosis Research

Elliott D. Crouser<sup>1</sup>, Tasha E. Fingerlin<sup>2</sup>, Ivana V. Yang<sup>3</sup>, Lisa A. Maier<sup>4</sup>, Patrick Nana-Sinkam<sup>5</sup>, Ronald G. Collman<sup>6</sup>, and Naftali Kaminski<sup>7</sup>

<sup>1</sup>Division of Pulmonary, Critical Care and Sleep Medicine, Ohio State University Wexner Medical Center, Columbus, Ohio; <sup>2</sup>Center for Genes Environment and Health, Department of Biomedical Research, National Jewish Health, Denver, Colorado; <sup>3</sup>Department of Medicine, University of Colorado School of Medicine, Aurora, Colorado; <sup>4</sup>Division of Occupational and Environmental Health Sciences, National Jewish Health, University of Colorado School of Medicine, Denver, Colorado; <sup>5</sup>Virginia Commonwealth University, Richmond, Virginia; <sup>6</sup>Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; and <sup>7</sup>Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut

ORCID ID: 0000-0003-3197-6746 (E.D.C.).

### Abstract

Sarcoidosis is a complex, polygenic disease of unknown cause with diverse clinical phenotypes, ranging from self-limited, asymptomatic disease to life-altering symptoms and early disease-related mortality. It is unlikely that a single common environmental exposure (e.g., infection, antigen) entirely explains the disease, and numerous genetic mutations are associated with the disease. As such, it is reasonable to assume, as with other phenotypically diverse diseases, that distinct genetic mechanisms and related biological biomarkers will serve to further define sarcoidosis subphenotypes, mechanisms, and possibly etiology, thus guiding personalized care. The fields of “omics” and systems biology research are

widely applied to understand polygenic and phenotypically diverse diseases, such as sarcoidosis. “Omics” refers to technologies that allow comprehensive profiling of sets of molecules in an organism. Systems biology applies advanced computational approaches to make sense of the enormous data sets that are typically generated from “omics” platforms. The primary objectives of this article are to review the available “omics” tools, assess the current status of “omics” and systems biology research in the field of sarcoidosis, and consider how this technology could be applied to advance our understanding of the mechanistic underpinnings of disease and to develop novel treatments.

**Keywords:** granuloma; genomic; proteomic; microbiome; epigenetic

(Received in original form July 18, 2017; accepted in final form October 18, 2017)

Supported by National Institutes of Health Grants R34HL123586 and R13HL126399 (E.D.C.), R01HL114587 (T.E.F.), U01-HL112712 and R01-HL113252 (R.G.C.), R01ES023826 (I.V.Y. and L.A.M.), and R01-HL127349 and U01-HL112712 (N.K.).

*Author Contributions:* All authors participated in concept, design, drafting of the manuscript, and critical review of the manuscript. All authors read and approved the manuscript.

Correspondence and requests for reprints should be addressed to Elliott D. Crouser, M.D., Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Ohio State University Wexner Medical Center, 473 West 12th Avenue, Columbus, OH 43210. E-mail: crouser.1@osu.edu

Ann Am Thorac Soc Vol 14, Supplement 6, pp S445–S451, Dec 2017

Copyright © 2017 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201707-567OT

Internet address: www.atsjournals.org

This workshop report was developed after a National Heart, Lung, and Blood Institute (NHLBI) workshop entitled “Leveraging Scientific Advancements to Understand Sarcoidosis Variability and Improve Outcomes” was held. Our working group was tasked with considering how systems biology approaches can be used to address the phenotypic variability of sarcoidosis and how the results could improve the understanding of the etiology and treatment of sarcoidosis. In this report we describe (1) the current state-of-the-art in terms of what is known about the elements of systems biology in sarcoidosis and the tools available for further

study, (2) knowledge and resource gaps that offer opportunities for development, and (3) recommendations for future research directions, with a special emphasis on studying the extreme phenotypes.

### Current State-of-the-Art of “Omics” and Systems Biology: Methodologies

#### Methodologies and Experimental Design Considerations

Systems biology is defined as a multidisciplinary holistic approach to the study of biological

systems that often utilizes unbiased computational approaches to “omics” technologies (1, 2), including microarrays (3), next-generation sequencing platforms, and other technologies that allow the study of the full molecular compendium. Many of the considerations are shared for all technologies. Initially, investigators need to determine whether the whole “omics” repertoire needs to be assayed or whether a targeted approach (4–6) is appropriate. Similarly, although the genome is identical for all cells and tissues in a given organism, other “omes” are cell-specific, posing challenges when tissues or cellular admixtures are studied. Fresh tissue sorting,

single-cell analysis, and laser capture microdissection (7) are all increasingly used, including in lung disease (8). Finally, the tissue compartment is critically important: the affected organ is preferable, although other compartments, such as bronchoalveolar lavage (BAL), sputum, exhaled breath condensate, or even blood may all be informative. In a systemic and variable disease such as sarcoidosis, obtaining multiple “omic” profiles from multiple compartments may allow better modeling of the disease process.

The ability to measure millions of data points for every single “omics” experiment makes hypothesis generation dependent on advanced computational approaches. “Machine learning” techniques are among the most widely used approaches to address this problem (9). The term “machine learning” refers to the computer science field of applying mathematical algorithms to “learn” from one source of data, usually to make predictions about a new source of data. For example, machine learning techniques can identify sets of genes whose expression distinguishes the disease of interest from other states or to predict clinical disease severity. The algorithms used in machine learning include applications of graph theory, clustering, and other data mining approaches, and can be applied in ways that incorporate existing information or ideas (supervised), that largely ignore preconceived ideas (unsupervised) or somewhere in between those two extremes (semisupervised). The relative benefits and challenges of each algorithm and approach differ depending on the application and purpose of the analysis.

### Barriers to Applying “Omics” and Systems Biology

Barriers to the application of “omics” and systems biology include the constant evolution of research technologies, the variety of platforms, the ever-expanding content of databases, and the understandable reluctance of the community to adopt the technologies. For instance, it is a challenge to directly compare genomic data generated on a gene array platform with data from next-generation sequencing. Likewise, the discovery of new molecular mechanisms and the related changes in genomic databases leads to a revised interpretation of the same “omics” data over time. Standardized research techniques, validation steps, experimental designs incorporating replication, and careful documentation of analytical approaches

are necessary. Furthermore, specific study design considerations may be required depending on the tissue analyzed and research objectives (Table 1). Because interpretation depends on advanced analytical approaches, the development of multidisciplinary teams and specific education of pulmonary faculty and trainees in systems biology approaches are needed to overcome existing barriers to adopting these innovative approaches.

## Sarcoidosis Systems Biology: Current State-of-the-Art

### Genomic Sarcoidosis Research

Genome-wide association studies (GWASs) have generated significant insights in sarcoidosis. As detailed in the accompanying article by Moller and colleagues in this issue (pp. S429–S436), various disease-modifying factors, and at least one disease-causing gene polymorphism (the butyrophilin-like 2 gene), have been identified by conventional GWAS approaches (10–14). These studies highlighted the genetic factors influencing disease phenotype and provide valuable mechanistic insights.

Despite the notable successes, the limitations of GWASs for sarcoidosis are

significant. The polygenetic nature of sarcoidosis is reflected by multiple genetic risk factors, the impact of which are often influenced by race and ethnicity. Many of the discovered loci are noncoding polymorphisms whose functions are unclear (15), and the studies have been limited to common variants. Most importantly, most sarcoidosis GWASs excluded African Americans, a population disproportionately affected with more severe disease. Thus, the future of genomic research in sarcoidosis depends on more powerful genomic technologies, careful delineation of phenotypes, and increased focus on African American subjects. The advanced research methods will allow identification of noncoding effects, including allele-specific effects relating to methylation of transcription factor-binding sites, inhibition of gene expression, or the influence of alternative splicing. The focus on detailed disease subphenotypes will allow identification of heritable traits that determine subphenotype demarcations, and the focus on African Americans will address a significant unmet need (15).

### Transcriptomic Sarcoidosis Research

“Transcriptomics” refers to profiling all the transcripts expressed in a cell, organ, or

**Table 1.** Study design considerations

Hypothesis Drives Sample Collection
<ul style="list-style-type: none"> <li>● Characteristics of subjects (human or animal) need to be carefully considered</li> <li>● Timing and frequency of sample collection depend on desired “snapshot” of biological process of interest, disease natural history, and accessibility of tissue</li> <li>● Compartment, tissue type, and cell type(s) of interest need to be considered</li> </ul>
Preparation Determines Product
<ul style="list-style-type: none"> <li>● Single cell (12) vs. multiple cells</li> <li>● Capturing “omics” vs. targeted aspects of feature</li> <li>● Feature captured, for example: <ul style="list-style-type: none"> <li>○ Total RNA</li> <li>○ RNA bound to ribosomes</li> <li>○ DNA predicted to bind to specific proteins (10, 11)</li> <li>○ Bacterial DNA or RNA (31–34)</li> <li>○ Metabolites in certain size range</li> <li>○ Capture of protein modifications (phosphorylation, glycosylation)</li> </ul> </li> </ul>
Molecular Analysis Depends on Sample Collection and Preparation
<ul style="list-style-type: none"> <li>● Sequencing DNA or RNA <ul style="list-style-type: none"> <li>○ Amplification steps (8) or single molecule (9)</li> <li>○ Depth of coverage</li> </ul> </li> <li>● Identification and quantification of proteins, metabolites (29, 30) <ul style="list-style-type: none"> <li>○ Liquid chromatography</li> <li>○ Mass spectrometry</li> <li>○ Ultraviolet detection</li> </ul> </li> </ul>

tissue including coding and noncoding RNAs, using RNA profiling techniques. At the cellular level transcriptomes reflect a summary of cellular processes at a given moment, as every stimulus, stress, or normal process is reflected by changes in RNA expression. At the level of the organ and in the context of human disease, the transcriptome represents the integration of multiple factors, including the patient's genetic background, the environmental stimuli to which the patient is exposed, the disease process, and many other variables. Thus, transcriptomics can be considered a tool for comprehensive molecular phenotyping of the patient, useful for identifying new disease subphenotypes, generating hypotheses regarding disease mechanisms and in some cases discovery of disease genetics.

The application of transcriptomics has focused thus far on discovery of novel biomarkers and disease mechanisms. Crouser and colleagues identified two matrix metalloproteinases (MMP-12, ADAMDEC1) that were highly expressed in lung tissue and BAL, and correlated with clinical disease activity (16). Other differentially expressed transcripts observed in lung tissue were subsequently shown by Koth and colleagues to be represented in peripheral blood (17), some of which, particularly the chemokine (C-X-C motif) ligand 9 gene and certain T-cell receptor-related transcripts, were predictive of chronically active and self-limited phenotypes (18). Tissue-specific gene expression profiling was shown to be useful for differentiating cardiac sarcoidosis from other common forms of myocarditis (19).

As for discovery of novel disease mechanisms, the relatedness of differentially expressed genes can be derived from a supervised "machine learning" approach, wherein gene networks are identified according to known or predicted interrelated functions or interactions (9). These networks typically represent coregulated genes, such as would be expected in response to the activation of a specific transcription factor or cell receptor. Applying this approach to sarcoidosis lung tissues identified genes conforming to networks regulated by IFN- $\gamma$  and signal transducer and activator of transcription 1 (16). Using a similar approach for microRNA expression in lung, lymph node, and peripheral blood mononuclear cell (PBMC) tissues revealed microRNAs from each tissue that were predicted to target

components of the transforming growth factor- $\beta$ /wingless-related integration site signaling pathway (20), a critical regulatory pathway of lung inflammation and fibrosis (21, 22). Although "machine learning" approaches can provide an informative and unbiased interpretation of complex transcriptomics data, subsequent experiments are required to validate and confirm the observations (23).

### Proteomic Sarcoidosis Research

Proteomics approaches provide a comprehensive analysis of all proteins present in a sample, and the most common methods are either arrays that detect proteins captured by specific probes or matrix-assisted laser desorption/ionization, which uses a laser to fragment the proteins, followed by an ionization step that allows the fragments to be identified by mass spectroscopy.

Proteomic analyses have proven useful for identifying sarcoidosis disease biomarkers. For instance, Häggmark and colleagues employed antigen microarrays to screen for autoimmunity and to identify novel disease biomarker proteins. IgG reactivity to a panel of more than 3,000 self-antigens was determined in 73 BAL samples from subjects with sarcoidosis or asthma, and from healthy control subjects. The initial findings were further validated in a larger cohort. Reactivity toward zinc finger protein 688 and mitochondrial ribosomal protein L43 was discovered with higher frequencies in patients with sarcoidosis, and L43 was particularly associated with a more severe disease phenotype (24). Silva and colleagues performed proteomic analysis (matrix-assisted laser desorption/ionization-time of flight mass spectrometry) of alveolar macrophages in subjects with sarcoidosis. Twenty-five unique proteins were identified compared with control samples, and were further shown by network analysis to be regulated by nuclear factor kappa-light-chain-enhancer of activated B cells. In particular, phospholipase D and RhoA expression differentiated patients with Löfgren's syndrome and those without Löfgren's syndrome (25).

### Metabolomics

Expanding from theoretical mechanisms implied by gene networks and related proteins, the next immediate challenge

is to determine the implications in terms of complex cell and organ functions. Metabolomics, the comprehensive study of small-molecule metabolites present in biological samples, can provide a fingerprint left by disease for the purpose of detection (i.e., a biomarker), such as was reported in the context of fibrotic pulmonary sarcoidosis (26), or could direct investigators in terms of understanding disease mechanisms to guide novel therapeutics or to monitor the progression of disease (27, 28).

### Microbiomics

Humans exist symbiotically with microbes that inhabit multiple body sites, including the skin and the gastrointestinal and respiratory tracts. The microbiome varies from one individual to the next depending on environmental and genetic variables. Significant evidence indicates that the microbiome can influence the development and function of the immune system in the context of health and disease (29). At present little is known about the role of the microbiome in the pathogenesis of sarcoidosis, but considering that microbiome alterations affect host immunity, it is possible that such changes will contribute to granuloma formation by presenting antigens and/or by modifying the function of immune cells.

Low levels of microbial DNA are present in the lungs of healthy people. These sequences largely reflect microbial communities present in the upper respiratory tract, and thus are likely derived mainly by passive microaspiration (30), which may impact immune responses (31). Garzoni and colleagues did not find any difference in the BAL microbiome profiles of patients with interstitial pneumonia and sarcoidosis (32). However, animal studies specifically implicate gut microbiota in the development of abnormal systemic inflammatory responses, including autoimmune disease (33, 34), allergy, cancer, and cardiovascular disease (35), reflecting the complexity of interactions and the need to sample relevant compartments. Genomic, and more specifically "metagenomic" (the study of genetic material from environmental sources), analysis of the intestinal

microbiome shows great promise for understanding the link between the microbial environment and inflammatory diseases (36), an approach that holds promise for advancing our understanding of sarcoidosis

### Integration of “Omics” Platforms

A major challenge is integration and interpretation of multiple “omics” and clinical data types, which are commonly classified into “horizontal” and “vertical” (37). Horizontal data integration compares a common set of “omics” data collected from various sources. The objective of the analysis is similar to traditional meta-analysis, namely, leveraging the additional power derived from multiple data sets to identify a universal pattern of change. In contrast, vertical integrative analysis deals with multiple types of “omics” data (Table 1) as well as clinical data that are measured in parallel in every subject in a cohort. It is postulated that the analysis of multiple hierarchies of “omics” data would improve disease classification (e.g., unique disease phenotypes responsive to different treatments). For example, in the context of sarcoidosis it could be valuable to integrate “omics” data horizontally across clinically similar entities (e.g., *Mycobacterium tuberculosis*) and vertically among different “omics” platforms (e.g., microbiome, genome) to identify common disease mechanisms (e.g., environmental factors, genetic variables). Although there is no definite solution at this stage of data integration, some intriguing approaches have been developed including subclassifications by principal component analysis (38), partial least squares (39), nonnegative matrix factorization (40), and an integrative phenotyping framework (41) to reduce features and identify patterns across multiple “omics” data sets, as well as novel clustering methods to identify novel disease subclasses (42).

Attempts to integrate genomic, transcriptomic, and proteomic data in sarcoidosis have been limited. Maver and colleagues reviewed the available literature and identified nine candidate molecules that are distinctly associated with sarcoidosis and that could have potential functional implications (43). Fischer and colleagues integrated

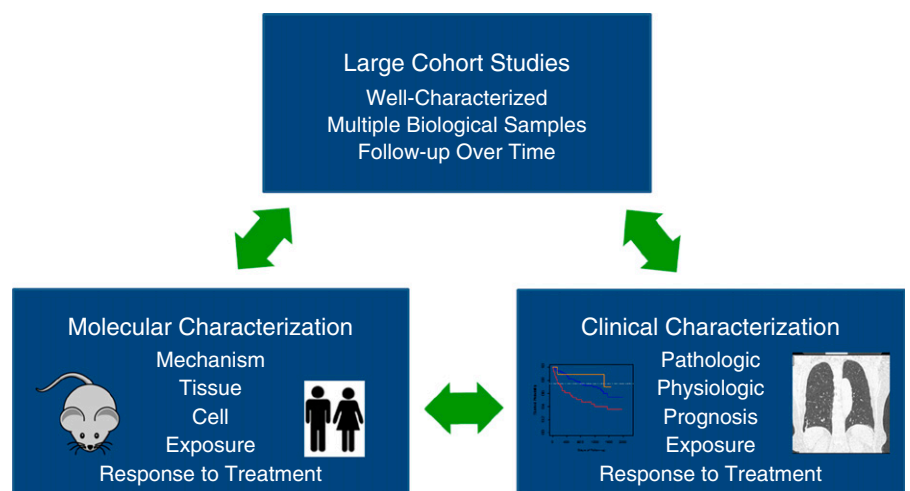
genomic and proteomic approaches to identify novel sarcoidosis risk genes. They first identified single-nucleotide risk polymorphisms in a European cohort, including novel associations with genes encoding the butyrophilin-like 2 gene promoter region, human leukocyte antigen-B\*0801, human leukocyte antigen-DPB1, and IL-23R. Related functional predictions and proteomic analysis further identified a potentially drug-targetable IL-23/T-helper cell type 17–signaling pathway (11).

It is expected that novel multiomic data sets will generate the foundations for implementing vertical integrative data analysis approaches, and in this regard the GRADS (Genomics Research in Alpha-1 Antitrypsin and Sarcoidosis) study provides a template for future sarcoidosis research. GRADS is a national, multicenter research study, funded by the NHLBI, that addresses two understudied lung diseases: alpha-1 antitrypsin deficiency and sarcoidosis (described in detail in Reference 44). GRADS planned to enroll 400 participants with a minimum of 35 in each of nine clinical phenotype subgroups (see Table 3 in Reference 44) prioritized by their clinical relevance to understanding the pathobiology and clinical heterogeneity of sarcoidosis. Participants with a confirmed

diagnosis of sarcoidosis underwent baseline and 6-month follow-up assessments, including self-administered questionnaires, computed tomography scan, pulmonary function testing, blood testing, and bronchoscopy with BAL. Fecal samples were collected for gut microbiome analysis. GRADS has generated a unique sarcoidosis resource. Blood collection included PBMCs, serum, plasma, and whole blood RNA and DNA; BAL included supernatant and cell pellet; and stool was also collected. Analyses performed through the GRADS Genomics and Informatics Center include RNA sequencing of BAL cells and PBMCs; microbiome, virome, and fungal determinations in BAL and stool; and a data-sharing portal is being created. GRADS is powerfully positioned to inform and direct studies on the pathobiology of sarcoidosis, identify diagnostic or prognostic biomarkers, and provide novel molecular phenotypes that could lead to improved personalized approaches to therapies for sarcoidosis. Results of these studies will be forthcoming.

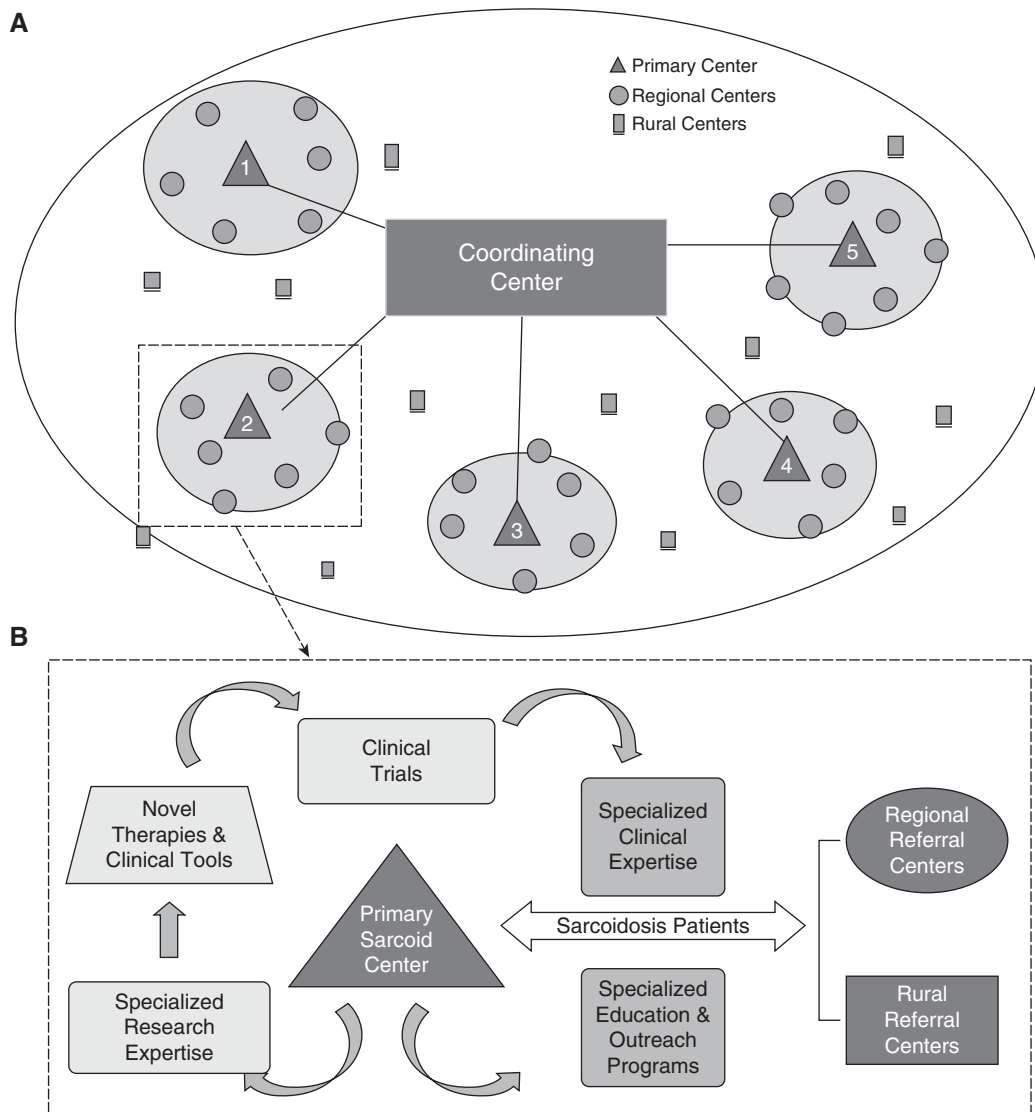
### Limitations of “Omics” and Systems Biology Research in Sarcoidosis

Sarcoidosis presents numerous challenges relating to the existence of distinct clinical phenotypes; the ability to define these



**Figure 1.** Proposed translational research paradigm. The clinical diversity of sarcoidosis mandates careful clinical phenotype characterization to inform related “omics” and systems biology research to improve the molecular characterization of the disease. Clinical validation of research observations (e.g., discovery of novel mechanisms and related treatments, biomarker discovery) requires longitudinal follow-up (e.g., large cohort studies) and sequential sampling from a large, well-characterized patient population.





**Figure 2.** Proposed international network for sarcoidosis research. (A) An international network of regional “centers of excellence” that is regulated by a coordinating center, the details of which would require further development. (B) The proposed functions of a regional center of excellence in terms of serving the clinical and educational needs of the regional sarcoidosis population, while conducting translational research (in conjunction with other national “centers”) designed to improve the access to and quality of care of patients with sarcoidosis. Sarcoid = sarcoidosis.

phenotypes easily; the lack of access to readily available and well-phenotyped tissues in a cross-sectional let alone longitudinal manner, including target tissue with active granulomatous inflammation and the ability to analyze distinct cell populations within tissues; and the lack of efforts to integrate the various “omics” platforms to strengthen the resulting systems biology-generated conclusions. The impact of the research is further influenced by the quality of the tissue samples, including variable results influenced by processing, storage techniques, and the cellular components represented in the samples.

### Challenges and Opportunities: Creating a Roadmap for “Omics” and Systems Biology Research in Sarcoidosis

Progress in sarcoidosis research is significantly impeded by the lack of relevant animal or *ex vivo* laboratory models of the disease. In lieu of such models, and given that sarcoidosis is a relatively rare disease, it is essential to optimize the scientific information that can be gained from each human participant in sarcoidosis research. Ideally, research would engage as many subjects as possible, including those with limited access to health care, many

of whom are at highest risk of severe disease complications (45, 46). Our working group identified the major goals and made recommendations aimed to address current challenges relating to rapidly recruiting biological samples from carefully phenotyped sarcoidosis cohorts, using standardized and reproducible research techniques.

### Goals

In keeping with the mission of agencies that are actively supporting sarcoidosis research, including the National Institutes of Health and Foundation for Sarcoidosis Research, our working group recommended that the

scientific priorities of emerging (e.g., GRADS) and future “omics” and systems biology research should be directed toward projects that translate readily to the clinical setting, as portrayed in Figure 1. Clinical translation is categorized as follows:

1. Biomarkers that improve the diagnosis of sarcoidosis
2. Prognostic biomarkers that identify extreme sarcoidosis phenotypes
3. Discovery of novel disease mechanisms
4. Identification of therapeutic targets
5. Inclusion of minorities and underserved populations; representing more severe clinical phenotypes (and deserving equal access to cutting edge research).

## Recommendations

### 1. Establish Clinical Centers of Excellence to Form the Foundation of Future Sarcoidosis Research

In keeping with the approach used to research other rare lung disorders (e.g., idiopathic pulmonary fibrosis [47], cystic fibrosis, and others), it was recommended that each participating site within a future sarcoidosis network conform to uniformly high standards of clinical care, maintain a registry of the sarcoidosis population and their long-term clinical data, and develop the capacity to obtain high-quality biological specimens for the purpose of scientific discovery. Sarcoidosis exemplifies health care disparities relating to race, sex, and factors influenced by lower socioeconomic status and related factors, such as diet, exercise, and environmental stress (48). Thus, the integration of underserved populations would be addressed at each research center within the sarcoidosis network through community outreach mechanisms, including patient and health care provider education, the establishment of processes to enhance health care access to underserved urban and rural communities, and by engaging with sarcoidosis patient advocacy organizations. Because sarcoidosis has pronounced geographic differences in

incidence and clinical features (10, 45), this network ideally should include sarcoidosis cohorts and centers globally, to ensure inclusion of varied disease manifestations and to enable studies to address commonality and differences across distinct geographic and clinical groups. A strategic distribution of sarcoidosis “centers of excellence” is envisioned to optimize the quality of translational research, including greater representation of severe disease phenotypes and underrepresented patient populations, while accelerating the pace of scientific discovery (Figure 2).

### 2. Develop Animal or *in Vitro* Models to Improve Research Efficiency

The development of highly relevant animal and/or *in vitro* models of pathological granuloma formation that closely resemble the features of sarcoidosis is considered to be a high priority. Such models would accelerate the rate of scientific discovery while reducing the costs of research relative to relying primarily on human tissue-based research. A thorough discussion of the barriers to creating such models is beyond the scope of this article, although the animal model is reviewed in part in an accompanying article (10).

### 3. Improve Resolution of “Omics” Data

It is important to optimize the resolution of patient-based research by more carefully dissecting disease mechanisms in specific sarcoidosis subphenotypes, particularly extreme phenotypes associated with poor outcomes. Because GRADS has addressed the requirements for bulk analysis, future studies should focus on specific cell populations, distinct tissue microenvironments, and single-cell “omics.” Such approaches will address the effects of cellular admixture and improve the mechanistic implications of the results.

### 4. Utilize Computer Models to Leverage “Omics” Data

The dynamics of complex molecular interactions identified by “omics” research can be modeled mathematically and further manipulated to simulate the effects of a

novel therapeutic intervention, such as was demonstrated in sarcoidosis by Hao and colleagues (49). Although computer models allow for the rapid “testing” of many potential therapeutic interventions, or combinations thereof, validation of the mathematical predictions ultimately requires longitudinal testing in humans. The inevitable reliance on relatively slow-moving and costly human research emphasizes the pressing need for developing relevant animal and/or *in vitro* models of disease.

## Conclusions

With the advent of powerful “omics” research platforms, it is now feasible to simultaneously assess innumerable molecular events representing multiple biological functions in the context of disease or health. The application of systems biology to “omics” data allows the recognition of biological patterns translatable into novel human disease-relevant scientific hypotheses, which are validated by more conventional laboratory and clinical research techniques. Considering the complexity and varying presentations of sarcoidosis, “omics” and systems biology techniques promise to catalyze scientific discovery to elucidate the biological mechanisms underlying disease phenotypes to identify more effective disease biomarkers and treatments. ■

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

**Acknowledgment:** The authors thank their patients; study participants who have enrolled in sarcoidosis-related research studies; the many organizations that have supported the sarcoidosis community and its research mission; Drs. Jerry Eu, George Mensah, Lora Reineck, and Antonello Punturieri, and the National Heart, Lung, and Blood Institute; and other Institutes/Centers of the National Institutes of Health for the support of this workshop. The authors also thank the remainder of the workshop participants for thoughtful discussions that led to the development of these recommendations, and Megan Marchant for assistance with editing and formatting this manuscript.

## References

- 1 Bunyavanich S, Schadt EE. Systems biology of asthma and allergic diseases: a multiscale approach. *J Allergy Clin Immunol* 2015;135:31–42.
- 2 Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010;11:476–486.
- 3 Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 2006;7:200–210.

- 4 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57–63.
- 5 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
- 6 Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–330.
- 7 Wang J, Song Y. Single cell sequencing: a distinct new field. *Clin Transl Med* 2017;6:10.
- 8 Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016;1:e90558.
- 9 Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet* 2015;16:321–332.
- 10 Moller DR, Rybicki BA, Hamzeh NY, Montgomery CG, Chen ES, Drake W, et al. Genetic, immunologic, and environmental basis of sarcoidosis. *Ann Am Thorac Soc* 2017;14(suppl6):S429–S436.
- 11 Fischer A, Ellinghaus D, Nutsua M, Hofmann S, Montgomery CG, Iannuzzi MC, et al.; GenPhenReSa Consortium. Identification of immune-relevant factors conferring sarcoidosis genetic risk. *Am J Respir Crit Care Med* 2015;192:727–736.
- 12 Grunewald J, Brynedal B, Darlington P, Nisell M, Cederlund K, Hillert J, et al. Different HLA-DRB1 allele distributions in distinct clinical subgroups of sarcoidosis patients. *Respir Res* 2010;11:25.
- 13 Levin AM, Iannuzzi MC, Montgomery CG, Trudeau S, Datta I, Adrianto I, et al. Admixture fine-mapping in African Americans implicates XAF1 as a possible sarcoidosis risk gene. *PLoS One* 2014;9:e92646.
- 14 Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 2005;37:357–364.
- 15 Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 2012;30:1095–1106.
- 16 Crouser ED, Culver DA, Knox KS, Julian MW, Shao G, Abraham S, 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.
- 17 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;184:1153–1163.
- 18 Su R, Li MM, Bhakta NR, Solberg OD, Darnell EP, Ramstein J, et al. Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. *Eur Respir J* 2014;44:985–993.
- 19 Lassner D, Kühl U, Siegismund CS, Rohde M, Elezskurtaj S, Escher F, et al. Improved diagnosis of idiopathic giant cell myocarditis and cardiac sarcoidosis by myocardial gene expression profiling. *Eur Heart J* 2014;35:2186–2195.
- 20 Crouser ED, Julian MW, Crawford M, Shao G, Yu L, Planck SR, et al. Differential expression of microRNA and predicted targets in pulmonary sarcoidosis. *Biochem Biophys Res Commun* 2012;417:886–891.
- 21 Levänen B, Wheelock AM, Eklund A, Grunewald J, Nord M. Increased pulmonary Wnt (wingless/integrated)-signaling in patients with sarcoidosis. *Respir Med* 2011;105:282–291.
- 22 Salez F, Gosset P, Copin MC, Lamblin Degros C, Tonnel AB, Wallaert B. Transforming growth factor- $\beta_1$  in sarcoidosis. *Eur Respir J* 1998;12:913–919.
- 23 Maertzdorf J, Weiner J III, Mollenkopf HJ, Bauer T, Prasse A, Müller-Quernheim J, et al.; TBnotTB Network. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci USA* 2012;109:7853–7858.
- 24 Häggmark A, Hamsten C, Wiklundh E, Lindskog C, Mattsson C, Andersson E, et al. Proteomic profiling reveals autoimmune targets in sarcoidosis. *Am J Respir Crit Care Med* 2015;191:574–583.
- 25 Silva E, Souchelnyskiy S, Kasuga K, Eklund A, Grunewald J, Wheelock AM. Quantitative intact proteomics investigations of alveolar macrophages in sarcoidosis. *Eur Respir J* 2013;41:1331–1339.
- 26 Mirsaeidi M, Banoei MM, Nienow CK, Abassi T, Hakim A, Schraufnagel D, et al. Plasma metabolomic profile in fibrosing pulmonary sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2016;33:29–38.
- 27 Devillier P, Salvator H, Naline E, Couderc LJ, Grassin-Delye S. Metabolomics in the diagnosis and pharmacotherapy of lung diseases. *Curr Pharm Des* 2017;23:2050–2059.
- 28 Geamanu A, Gupta SV, Bauerfeld C, Samavati L. Metabolomics connects aberrant bioenergetic, transmethylation, and gut microbiota in sarcoidosis. *Metabolomics* 2016;12:35.
- 29 Eloe-Fadrosh EA, Rasko DA. The human microbiome: from symbiosis to pathogenesis. *Annu Rev Med* 2013;64:145–163.
- 30 Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011;184:957–963.
- 31 Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* 2013;1:19.
- 32 Garzoni C, Brugger SD, Qi W, Wasmer S, Cusini A, Dumont P, et al. Microbial communities in the respiratory tract of patients with interstitial lung disease. *Thorax* 2013;68:1150–1156.
- 33 Scher JU, Abramson SB. The microbiome and rheumatoid arthritis. *Nat Rev Rheumatol* 2011;7:569–578.
- 34 Yurkovetskiy L, Burrows M, Khan AA, Graham L, Volchkov P, Becker L, et al. Gender bias in autoimmunity is influenced by microbiota. *Immunity* 2013;39:400–412.
- 35 Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW, Zheng SS. Application of metagenomics in the human gut microbiome. *World J Gastroenterol* 2015;21:803–814.
- 36 Martín R, Miquel S, Langella P, Bermúdez-Humarán LG. The role of metagenomics in understanding the human microbiome in health and disease. *Virulence* 2014;5:413–423.
- 37 Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res* 2012;40:3785–3799.
- 38 Lock EF, Dunson DB. Bayesian consensus clustering. *Bioinformatics* 2013;29:2610–2616.
- 39 Li W, Zhang S, Liu CC, Zhou XJ. Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. *Bioinformatics* 2012;28:2458–2466.
- 40 Zhang S, Liu CC, Li W, Shen H, Laird PW, Zhou XJ. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res* 2012;40:9379–9391.
- 41 Kim S, Herazo-Maya JD, Kang DD, Juan-Guardela BM, Tedrow J, Martinez FJ, et al. Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* 2015;16:924.
- 42 Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 2009;25:2906–2912.
- 43 Maver A, Medica I, Peterlin B. Search for sarcoidosis candidate genes by integration of data from genomic, transcriptomic and proteomic studies. *Med Sci Monit* 2009;15:SR22–SR28.
- 44 Moller DR, Koth LL, Maier LA, Morris A, Drake W, Rossman M, et al.; GRADS Sarcoidosis Study Group. Rationale and design of the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) study: sarcoidosis protocol. *Ann Am Thorac Soc* 2015;12:1561–1571.
- 45 Mirsaeidi M, Machado RF, Schraufnagel D, Sweiss NJ, Baughman RP. Racial difference in sarcoidosis mortality in the United States. *Chest* 2015;147:438–449.
- 46 Westney GE, Judson MA. Racial and ethnic disparities in sarcoidosis: from genetics to socioeconomic. *Clin Chest Med* 2006;27:453–462, vi.
- 47 Belkin A, Swigris JJ. Patient expectations and experiences in idiopathic pulmonary fibrosis: implications of patient surveys for improved care. *Expert Rev Respir Med* 2014;8:173–178.
- 48 Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. *Annu Rev Sociol* 2010;36:349–370.
- 49 Hao W, Crouser ED, Friedman A. Mathematical model of sarcoidosis. *Proc Natl Acad Sci USA* 2014;111:16065–16070.