

UCSF

UC San Francisco Previously Published Works

Title

Defining CD4 T helper and T regulatory cell endotypes of progressive and remitting pulmonary sarcoidosis (BRITE): protocol for a US-based, multicentre, longitudinal observational bronchoscopy study

Permalink

<https://escholarship.org/uc/item/9vk3j582>

Journal

BMJ Open, 11(11)

ISSN

2044-6055

Authors

Koth, Laura L
Harmacek, Laura D
White, Elizabeth K
[et al.](#)

Publication Date

2021-11-01

DOI



10.1136/bmjopen-2021-056841

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed

BMJ Open Defining CD4 T helper and T regulatory cell endotypes of progressive and remitting pulmonary sarcoidosis (BRITE): protocol for a US-based, multicentre, longitudinal observational bronchoscopy study

Laura L Koth ,¹ Laura D Harmacek,² Elizabeth K White,³ Nicholas Kostandinos Arger,¹ Linda Powers,⁴ Brenda R Werner,⁴ Roman E Magallon,² Pineet Grewal,¹ Briana Q Barks,⁵ Li Li,⁵ May Gillespie,⁵ Sarah E Collins,⁶ Jessica Cardenas,¹ Edward S Chen ,⁶ Lisa A Maier,⁵ Sonia M Leach,³ Brian P O'Connor,² Nabeel Y Hamzeh⁴

To cite: Koth LL, Harmacek LD, White EK, *et al.* Defining CD4 T helper and T regulatory cell endotypes of progressive and remitting pulmonary sarcoidosis (BRITE): protocol for a US-based, multicentre, longitudinal observational bronchoscopy study. *BMJ Open* 2021;**11**:e056841. doi:10.1136/bmjopen-2021-056841

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-056841>).

Received 31 August 2021
Accepted 27 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Nabeel Y Hamzeh;
nabeel-hamzeh@uiowa.edu

ABSTRACT

Introduction Sarcoidosis is a multiorgan granulomatous disorder thought to be triggered and influenced by gene–environment interactions. Sarcoidosis affects 45–300/100 000 individuals in the USA and has an increasing mortality rate. The greatest gap in knowledge about sarcoidosis pathobiology is a lack of understanding about the underlying immunological mechanisms driving progressive pulmonary disease. The objective of this study is to define the lung-specific and blood-specific longitudinal changes in the adaptive immune response and their relationship to progressive and non-progressive pulmonary outcomes in patients with recently diagnosed sarcoidosis.

Methods and analysis The BRonchoscopy at Initial sarcoidosis diagnosis Targeting longitudinal Endpoints study is a US-based, NIH-sponsored longitudinal blood and bronchoscopy study. Enrolment will occur over four centres with a target sample size of 80 eligible participants within 18 months of tissue diagnosis. Participants will undergo six study visits over 18 months. In addition to serial measurement of lung function, symptom surveys and chest X-rays, participants will undergo collection of blood and two bronchoscopies with bronchoalveolar lavage separated by 6 months. Freshly processed samples will be stained and flow-sorted for isolation of CD4 +T helper (Th1, Th17.0 and Th17.1) and T regulatory cell immune populations, followed by next-generation RNA sequencing. We will construct bioinformatic tools using this gene expression to define sarcoidosis endotypes that associate with progressive and non-progressive pulmonary disease outcomes and validate the tools using an independent cohort.

Ethics and dissemination The study protocol has been approved by the Institutional Review Boards at National Jewish Hospital (IRB# HS-3118), University of Iowa (IRB# 201801750), Johns Hopkins University (IRB# 00149513) and University of California, San Francisco (IRB# 17-

Strengths and limitations of this study

- This is the largest and most geographically diverse longitudinal bronchoscopy study with paired blood analysis performed to date in pulmonary sarcoidosis in the USA.
- The study design will use repeated measures of a comprehensive clinical phenotyping strategy over 18 months to ascertain pulmonary outcomes in patients with recently diagnosed disease that will be a rich clinical and biological data resource for studies in pulmonary sarcoidosis.
- The experimental design will leverage single cell analysis to define immunophenotypic changes in disease associated CD4 T cell populations over time and associate findings with progressive and non-progressive pulmonary sarcoidosis clinical phenotypes.
- The experimental plan will also measure transcriptional profiles of isolated CD4 T cell populations using RNA sequencing to enable construction of novel computational prognostic tools via expression deconvolution and association of findings with progressive and non-progressive pulmonary sarcoidosis.
- A potential methodological limitation is the reliance on flow cytometers at each centre for processing of freshly isolated samples requiring significant optimisation.

23432). All participants will be required to provide written informed consent. Findings will be disseminated via journal publications, scientific conferences, patient advocacy group online content and social media platforms.

INTRODUCTION

Sarcoidosis is a multisystem granulomatous disorder that is thought to be triggered and influenced by gene–environment interactions.¹ This condition affects at least 45–300/100 000 individuals in the USA² and has a rising mortality rate.³ Sarcoidosis affects the lungs in more than 90% of cases. The greatest gap in knowledge in sarcoidosis management is a lack of tools to discern which subjects will develop progressive disease vs stable or remitting disease. In addition, advanced therapeutics that are effective are lacking which is linked to a lack of understanding about the pathophysiology driving progressive disease.⁴ While the environmental triggers are unknown,^{5–8} the inflammatory response is characterised by a predominance of activation of CD4 +T cells, cytokine production, macrophage activation and granuloma formation.^{9–18}

The immunological determinants of clinical outcomes in sarcoidosis remain poorly understood. Sarcoidosis is traditionally characterised by enhanced production of interferon-gamma (IFN γ) by CD4 +T cells. Importantly, recent novel findings identified that a majority of IFN γ -producing cells in the bronchoalveolar lavage (BAL) from sarcoidosis patients bear a Th17 phenotype and are more properly classified as Th17.1 cells rather than Th1 cells.¹⁹ In addition, a robust expansion of IFN γ -producing Th17.1 cell was also identified in a European cohort of newly diagnosed sarcoidosis patients, and an even more striking expansion of interleukin 17-producing Th17 cells in the peripheral blood.²⁰ Mouse models have revealed functional ‘plasticity’ in Th17 cells, demonstrating their ability to ‘transdifferentiate’ to an anti-inflammatory phenotype as defined by their transcriptional profile and regulatory capacity.²¹ Other studies support that a deficit in regulatory T cell (Treg) capacity may permit disease activity in sarcoidosis.^{22–25} In other words, disease activity in sarcoidosis may depend, in part, on T cell functional plasticity, resulting from transdifferentiation of Th17 cells into pathogenic Th17.1 cells versus anti-inflammatory Tregs, as shown in vivo.²¹ As such, our study’s hypothesis states that Th17.1 and Tregs play significant yet opposing roles to determine clinical outcomes in sarcoidosis, and that the role of Th17 cells could transcriptionally reflect a multifunctional (eg, Treg) profile in patients with stable (non-progressive or improving) disease.

The BRonchoscopy at Initial sarcoidosis diagnosis Targeting longitudinal Endpoints (BRITE) study is an NIH (National Institute of Health)-sponsored, multi-centre, longitudinal study that will measure changes in the immune response and its relationship to pulmonary outcomes in sarcoidosis patients early in their disease course. This is the first study to investigate longitudinal changes in paired blood and BAL T cell phenotype. The study has three primary goals. In aim 1, we will measure changes in T cell lineage diversity (Th17.1, Th17, Th1 and Treg) that occur longitudinally as sarcoidosis subjects develop either progressive or non-progressive disease. In aim 2, we will define changes in T cell lineages during disease course through identification of genome wide

transcriptional profiles of purified Th17.1, Th17, Th1 and Treg cells using low-input RNA-sequencing (RNA-seq). In aim 3, we will construct computational tools from T cell lineage diversity and gene expression data to define and/or predict sarcoidosis endotypes using bioinformatic deconvolution and biostatistical methods. We will then extend the genomic findings from the NIH GRADS study (Genomic Research in Alpha-1 anti-trypsin Deficiency and Sarcoidosis)²⁶ by comparing our genomic T cell signatures with gene expression data from this independent cohort to predict T cell lineages and relate findings to outcome definitions used in that study. These bioinformatic tools can also be applied to other genomic datasets from US and European cohorts.

This study fills a major gap in our current knowledge by focusing on the relationship between longitudinal measures of pulmonary outcomes with T cell lineages using an integrated approach of phenotypic cell-sorting, followed by RNA transcriptional profiling, in addition to the measurement of T cell phenotypes in the blood and BAL compartments. Gene expression data obtained in this study will be deconvoluted to create a pipeline of computational tools that can be applied to other datasets to predict disease phenotypes and clinical course. These tools can also be used to assess clinical responses to emerging immunomodulatory therapies and to identify cellular signalling pathways that could be targeted with existing small molecule inhibitors or biological therapies.

METHODS AND ANALYSIS

Study setting

This study involves four academic universities with established clinical centres in sarcoidosis care and research: National Jewish Health (NJH), University of Iowa, University of California, San Francisco (UCSF) and Johns Hopkins University. All centres represent a geographic spectrum from Coast to Coast. Each site is responsible for enrolling a similar number of participants (figure 1).

Eligibility criteria

Participants will have been diagnosed within 18 months of enrolment according to criteria endorsed by the American Thoracic Society.²⁷ Additional eligibility criteria for enrolment are presented in box 1. Participants with fibrotic (chest X-ray, CXR stage IV) lung disease will be excluded as this form of lung disease was considered long-standing.

Study procedures

Study participants will undergo several procedures as depicted in table 1. A history and physical examination will be performed at each study visit to assess disease status and changes in measurements and to confirm safety of performing bronchoscopy with lavage. An anterior–posterior CXR and pulmonary function testing including forced vital capacity (FVC), forced expiratory volume in 1 s (FEV1) and a single breath diffusing capacity for carbon

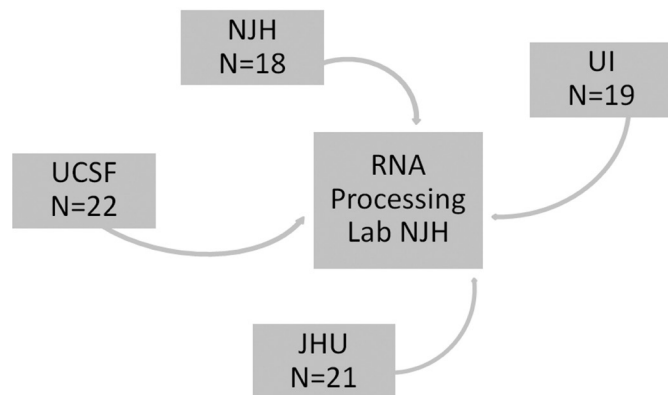


Figure 1 Organisational structure of the multisite study. Participants will be enrolled at four clinical centres, University of Iowa (UI), University of California, San Francisco (UCSF), National Jewish Health (NJH) and Johns Hopkins University (JHU). Each of these centres is responsible for collection of the clinical data and processing of biospecimens. Flow cytometric-sorted CD4 T cell populations will be shipped to the genomics facility at National Jewish Health for RNA isolation and next-generation sequencing.

monoxide (DLCO) will be performed at baseline and at 6-month intervals during follow-up. Standard bronchoscopy with lavage will be performed at baseline and 6 months later. At each 6-month study visit, participants will complete a blood draw and three patient-reported outcome measures to numerically quantify dyspnoea (University of California, San Diego shortness of Breath Questionnaire²⁸), quality of life (12-item Short Form Survey (SF-12)²⁹) and fatigue (patient-reported outcomes measurement information system³⁰).

Outcomes

The primary outcome at the completion of the study visits will be progressive pulmonary sarcoidosis. To define this outcome, we will use a composite measure of lung function and radiology measurements. For lung function, we will define progression by declines in lung function using thresholds that have precedent in the interstitial lung disease literature.^{31 32} Specifically, participants will meet the definition of progressive lung disease if there is a $\geq 10\%$ decline in FVC and/or FEV1 or a $\geq 15\%$ decline in DLCO. For chest imaging, we will define progression by increasing opacities on chest radiography as determined by the interpreting radiologist/investigator. If a participant meets either the lung function definition or the CXR definition, they will be categorised as progressive lung disease. Participants who do not meet these definitions will be categorised as non-progressive disease. We will exclude other causes that may mimic progressive disease based on clinical presentation, radiographic pattern consistent with sarcoidosis and analysis of BAL cell counts and differentials. BAL cultures will be performed based on clinical indications.

Participant timeline

Study visits and timeline are depicted in figure 2. Procedures at each visit are outlined in table 1. At enrolment

Box 1 Inclusion and exclusion criteria

Inclusion criteria

A histopathological diagnosis of sarcoidosis according to the American Thoracic Society/European Respiratory Society sarcoidosis statement with the exception of Lofgren's syndrome which are exempt from a pathological diagnosis.

Diagnosis of sarcoidosis within 18 months of enrolment.

Non-smokers (<10 pack-years and no smoking for 6 months prior to enrolment) including vaping and use of marijuana.

No history of systemic immunosuppression within 2 weeks of enrolment.

Chest X-ray Scadding stage 0, I, II, III (chest X-ray stage 0 requires lung biopsy confirming granulomatous inflammation).

Exclusion criteria

Unable to tolerate study procedures as determined by the site principal investigator.

Pregnancy.

Active smoker.

Systemic immunosuppressive therapy within 2 weeks of enrolment except for inhaled steroids.

Scadding stage IV.

Evidence of active cardiac, neurological or ophthalmic involvement with sarcoidosis and require systemic immunosuppressive therapy.

Diagnosis of beryllium sensitisation and/or disease, common variable immunodeficiency, mycobacterial and/or fungal infection or suspected hypersensitivity pneumonitis.

On anticoagulation except for aspirin.

Active bacterial or viral infection, use of antibiotics or immunisation within 4 weeks of enrolment (may reassess for participation after the 4-week period).

Known medical problems that could affect biological interpretations, including malignancy, autoimmunity, asthma and COPD (Chronic Obstructive Pulmonary Disease), chronic viral infections (hepatitis B, C, Herpes virus requiring suppressive medications, HIV (Human Immunodeficiency Virus)).

History of cancer other than presumed cured non-metastatic skin cancer.

Currently institutionalised (eg, prison, long-term care facility).

Other comorbid conditions that increase risk of complications from bronchoscopy including uncontrolled hypertension and/or diabetes, unstable coronary artery disease or decompensated heart failure, active cardiac arrhythmias.

(V1) and at 6-month intervals (V3, V5, V6), participants will be assessed with the procedures of lung function, CXR and questionnaires during the 18-month follow-up period. The V2 and V4 visits focus on biospecimen collection for bronchoscopy with lavage and venipuncture and will occur within 1 month of V1 and V3, respectively (figure 2). The second bronchoscopy will occur approximately 5–6 months after the first bronchoscopy. Participants will also undergo venipuncture for blood collection at V5 and V6.

Sample size

Our primary outcome is to determine the distribution of T helper cells and Treg cells in paired blood and BAL samples over time and relate these measurements to clinical outcome (progressive vs non-progressive disease). Based on previous cohort studies from our group,³³

Table 1 Visit procedures

Procedures	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6
Informed consent	X					
History and physical exam	X	X	X	X	X	X
Chest X-ray	X		X			X
Spirometry	X		X		X	X
Diffusing capacity	X		X		X	X
Bronchoscopy with lavage		X		X		
Blood draw		X		X	X	X
Nasal swab for COVID-19 PCR	X	X	X	X	X	X
Pregnancy test (if applicable)	X	X	X	X	X	X

approximately 50% of sarcoidosis patients met the definition for progressive disease. Therefore, we assumed 50% of the target enrolment sample (ie, 40 participants out of a total target enrolment of 80 participants) will be progressive. With $\alpha=0.00625$ (0.05/8 for four subsets, two time points), we will have >80% power to detect a 0.82 SD difference in a cell subset percentage between groups; if only 40% progress, we will have >80% power to detect a 0.85 SD difference between groups. For the gene expression analyses, we will have 80% power to detect a 0.95 SD change in expression between progressive and non-progressive participants or an R^2 of 0.19 between expression and a quantitative clinical phenotype such as FVC, assuming $\alpha=0.001$.

Data collection methods

Deidentified clinical data will be entered and stored in a central REDCap (Research Electronic Data Capture) database (<https://www.project-redcap.org>) that was created for the BRITE study and hosted by UCSF. Data instruments include demographic information, organ involvement using a modified organ assessment tool,³⁴ medical history, medications, questionnaire responses, pulmonary function measurements, complete blood counts and details related to the biospecimen collection of blood and BAL. The organ assessment tool was developed by Delphi study methodology³⁴ and is a

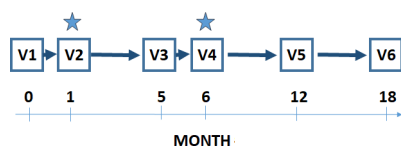


Figure 2 Study design with timeline. Participants will be assessed with the procedures of lung function, CXR and questionnaires at enrolment (visit 1 or V1) and at 6-month intervals (V3, V5, V6) during the 18-month follow-up period. The V2 and V4 visits consist of biospecimen collection for bronchoscopy with lavage and venipuncture and will occur within 1 month of V1 and V3, respectively. The second bronchoscopy will occur approximately 5–6 months after the first bronchoscopy. Participants will also undergo venipuncture for blood collection at V5 and V6. CXR, chest X-ray.

well-accepted survey to quantify organ involvement by sarcoidosis.²⁶ The questionnaires will be completed by participants on paper or through a REDCap survey. The dyspnoea and fatigue questionnaires have been validated in patients with interstitial lung disease^{28 35 36} and sarcoidosis,^{30 37} respectively. The SF-12 instrument is an extensively used quality of life assessment tool that is scored into a Mental Component Summary score and a Physical Component Summary score.²⁹ Bronchoscopy with lavage will be performed in two subsegments using up to a total of 480 mL sterile saline. Prebronchoscopy and postbronchoscopy assessment and monitoring will be performed as per each institution's standard of care. Each centre will follow recommended ATS/ERS guidelines for test performance for spirometry³⁸ and DLCO.³⁹

Biospecimen collection

A complete blood count will be collected at visits 2, 4, 5, 6 (figure 2). Serum will be collected at visits 2, 4, 6. PMBC will be collected at visits 2 and 4 and processed for flow cytometric and gene expression assays. PBMC (Peripheral Blood Mononuclear Cell) will also be collected at visits 5 and 6 and frozen. Excess PBMCs at visits 2 and 4 will also be frozen. Whole blood will be collected into DNA PAXgene tubes and frozen at visit 5 and will be collected and frozen into RNA PAXgene tubes at visits 2 and 6. All frozen samples will be stored for future use as specified in the IRB approvals at each site. Cells isolated from PBMC and BAL samples will undergo multiparameter flow cytometry with sorting of Treg, Th1, Th17.0 and Th17.1 populations. Sorted cells will be lysed with appropriate RNase inactivating buffers and kept frozen at -80°C until processed. Vials of lysed cells will be shipped to NJH for RNA isolation and RNA-seq next-generation sequencing of each sorted population.

Data management

The data analyst will perform data validation in accordance with the study-wide protocol specifications. Quarterly data check programmes will be run to identify discrepancies in entered data. Study sites will be notified about data discrepancies and query resolution will be performed and the database updated as appropriate.

Discrepancies to be flagged will include inconsistent data, missing data, range checks and deviations from the protocol. There will be a final data validation check of the study-wide dataset and the database will be locked after approval from all investigators.

Statistical analysis

Our primary outcome will be to determine the distribution of T helper cells and Treg cells at diagnosis and follow-up in paired blood and BAL samples and relate these measurements to clinical course (progressive vs non-progressive disease). The primary statistical analyses will use t-tests or rank sum tests, as appropriate, to compare the percentage of each subset (eg, Th17.1) between participants with progressive and non-progressive disease. The primary analysis will use the enrolment distribution to predict progressive vs non-progressive disease at follow-up; secondary analyses will consider the change in a subset frequency between enrolment and follow-up. Exploratory analyses will test for association between the cell subset percentages and measures of disease severity (FEV1, FEV1/FVC, DLCO).

Association between expression and clinical outcomes: To determine whether individual genes or combinations of genes are associated with clinical outcomes, a series of analyses will be computed using gene expression from BAL and PBMC samples. The main clinical outcomes are as defined in aim 1 (progressive vs non-progressive). To leverage the ability to infer more than cross-sectional correlation, primary analyses will test whether gene expression at enrolment predicts the primary clinical outcomes at follow-up. Secondary analyses will test whether significant gene expression changes between enrolment and follow-up identified above are associated with the clinical outcomes at follow-up. Primary inference regarding disease aetiology will be based on gene expression from BAL samples among all sarcoidosis cases. To identify potential gene expression markers of our primary outcome, PBMC will be used to identify genes whose expression is correlated with expression of genes in BAL that predict the clinical outcomes. Exploratory analyses will consider other time point combinations and solely use PBMC gene expression to predict clinical outcomes.

Differential expression between dichotomous groups (eg, progressive vs non-progressive): this will be identified via negative binomial regression using the software DESeq2,⁴⁰ adjusting for age and sex. We will also use CIBERSORTx to deconvolve RNA-seq data to predict both cell-type proportions⁴¹ (which we can validate against our flow analysis results) and to compute cell-type-specific gene expression for the cell types we can reliably separate from bulk. Association between continuous outcomes (eg, FVC) and read count will be tested via regression models with the clinical outcome as the independent variable and read count as a predictor, also adjusting for age and sex. We will perform these tests using read counts for all genes at enrolment for the primary analyses; the secondary analyses will use change in normalised read

counts from enrolment to follow-up for only those genes differentially expressed overall between enrolment and follow-up. After the progressive versus non-progressive disease differential expression analysis, we will use the significantly associated genes to perform hierarchical clustering to determine the ability of the combination of genes to distinguish between progressive and non-progressive cases. We will perform similar analyses as described above in terms of identifying pathways involved and overlap among the significantly associated genes for all the clinical outcomes measured in this study.

For aim 3, concordance between known and predicted cell-type proportions or cell-type-specific expression values will be determined by Pearson correlation coefficient and root mean square error to measure linear fit and estimation bias, respectively.⁴² Empirical p values are generated to test the null hypothesis that no cell types in the signature matrix are present in a given mixture using a Monte Carlo sampling strategy for randomised mixture samples.⁴² A similar strategy will be used to estimate p values for the cell-type-specific expression data.

Clinical outcomes: Similar to aim 2, we will first test for whether gene expression at specific genes or combinations of genes is associated with clinical outcomes. In this aim, we will be able to use the cell-type-specific expression data to test these associations. We will use the same approach in terms of primary analyses (expression in BAL at enrolment predicting clinical outcomes at follow-up) and secondary analyses (changes in BAL expression between enrolment and follow-up predicting clinical outcomes at follow-up). We anticipate choosing one or two cell-specific subset expression profiles rather than testing all subsets; we will use the results from aim 2 and aim 3 to determine the cell types that appear most likely to be driving any clinical associations. Exploratory analyses will evaluate all the cell subsets and PBMC expression. We will first test individual genes within each subset but will also consider using principal components analysis to reduce dimensionality in addition to other refinement of the list of genes to test based on our results in aim 2. If we test all genes, we will use a false discovery rate of ≤ 0.001 , similar to aim 2. If we have a smaller subset of genes to test in primary analyses, we will use a Bonferroni correction to determine the type I error rate. The constructed computational tool and identified endotypes from the BRITe study cohort will be tested with an independent cohort of samples (ie, the GRADS datasets).

Study timeline

Recruitment began in February 2020 with a projected completion of study visits by February 2023.

Patient and public involvement

It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

ETHICS AND DISSEMINATION

This multicentre study is being conducted in accordance with globally accepted standards of good practice, in agreement with the Declaration of Helsinki and with local regulations. The study protocol has been approved by the Institutional Review Boards at National Jewish Hospital (IRB# HS-3118), University of Iowa (IRB# 201801750), Johns Hopkins University (IRB# 00149513), and University of California, San Francisco (IRB# 17-23432). All participants will be required to provide written informed consent. All manuscripts resulting from the study will be submitted to peer-reviewed journals. Findings will be disseminated via journal publications, scientific conferences, patient advocacy group online content and social media platforms.

Author affiliations

¹Department of Medicine, University of California San Francisco, San Francisco, California, USA

²Center for Genes, Environment, & Health, National Jewish Health, Denver, Colorado, USA

³Department of Immunology and Genomic Medicine, National Jewish Health, Denver, Colorado, USA

⁴Department of Medicine, University of Iowa, Iowa City, Iowa, USA

⁵Division of Environmental and Occupational Health Sciences, National Jewish Health, Denver, Colorado, USA

⁶Division of Pulmonary and Critical Care Medicine, Baltimore, Maryland, USA

Acknowledgements This work has not been previously presented at a conference or published as a conference abstract. We would like to acknowledge Kristyn MacPhail for her contributions to the study design.

Contributors All the authors contributed to the conception and design of the study. Drafting of the manuscript was performed by NYH and LLK. Rigorous critiques were provided by LDH, EKW, NKA, LP, BRW, REM, PG, BQB, LL, MG, SEC, JC, ESC, LAM, SML and BPO'C.

Funding This study is funded by the National Heart, Lung and Blood Institute (NHLBI) of the National Institute of Health (NIH), grant number 5R01HL136681.

Competing interests The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Laura L Koth <http://orcid.org/0000-0001-9541-3622>

Edward S Chen <http://orcid.org/0000-0003-0723-8209>

REFERENCES

- Crouser ED, Maier LA, Wilson KC, *et al*. Diagnosis and detection of sarcoidosis: an official American thoracic Society clinical practice guideline. *Am J Respir Crit Care Med* 2020;201:e26–51.
- Erdal BS, Clymer BD, Yildiz VO, *et al*. Unexpectedly high prevalence of sarcoidosis in a representative U.S. metropolitan population. *Respir Med* 2012;106:893–9.
- Swigris JJ, Olson AL, Huie TJ, *et al*. Sarcoidosis-related mortality in the United States from 1988 to 2007. *Am J Respir Crit Care Med* 2011;183:1524–30.
- Hamzeh N. Sarcoidosis. *Med Clin North Am* 2011;95:1223–34.
- Maier LA. Is smoking beneficial for granulomatous lung diseases? *Am J Respir Crit Care Med* 2004;169:893–5.
- Bresnitz EA, STOLLEY PD, ISRAEL HL, *et al*. Possible risk factors for sarcoidosis. *Ann N Y Acad Sci* 1986;465:632–42.
- Gerke AK, van Beek E, Hunninghake GW. Smoking inhibits the frequency of bronchovascular bundle thickening in sarcoidosis. *Acad Radiol* 2011;18:885–91.
- Newman LS, Rose CS, Bresnitz EA, *et al*. A case control etiologic study of sarcoidosis: environmental and occupational risk factors. *Am J Respir Crit Care Med* 2004;170:1324–30.
- Gerke AK, Hunninghake G. The immunology of sarcoidosis. *Clin Chest Med* 2008;29:379–90.
- Ziegenhagen MW, Benner UK, Zissel G, *et al*. Sarcoidosis: TNF-alpha release from alveolar macrophages and serum level of sIL-2R are prognostic markers. *Am J Respir Crit Care Med* 1997;156:1586–92.
- Rothkrantz-Kos S, van Diejen-Visser MP, Mulder PGH, *et al*. Potential usefulness of inflammatory markers to monitor respiratory functional impairment in sarcoidosis. *Clin Chem* 2003;49:1510–7.
- Miyoshi S, Hamada H, Kadowaki T, *et al*. Comparative evaluation of serum markers in pulmonary sarcoidosis. *Chest* 2010;137:1391–7.
- Grutters JC, Fellrath J-M, Mulder L, *et al*. Serum soluble interleukin-2 receptor measurement in patients with sarcoidosis: a clinical evaluation. *Chest* 2003;124:186–95.
- Belli F, Capra A, Moraiti A, *et al*. Cytokines assay in peripheral blood and bronchoalveolar lavage in the diagnosis and staging of pulmonary granulomatous diseases. *Int J Immunopathol Pharmacol* 2000;13:61–7.
- Salazar A, Pintó X, Mañá J. Serum amyloid A and high-density lipoprotein cholesterol: serum markers of inflammation in sarcoidosis and other systemic disorders. *Eur J Clin Invest* 2001;31:1070–7.
- Bargagli E, Magi B, Olivieri C, *et al*. Analysis of serum amyloid A in sarcoidosis patients. *Respir Med* 2011;105:775–80.
- Grunewald J, Berlin M, Olerup O, *et al*. Lung T-helper cells expressing T-cell receptor AV2S3 associate with clinical features of pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2000;161:814–8.
- Katchar K, Eklund A, Grunewald J. Expression of Th1 markers by lung accumulated T cells in pulmonary sarcoidosis. *J Intern Med* 2003;254:564–71.
- Ramstein J, Broos CE, Simpson LJ, *et al*. IFN-γ-Producing T-helper 17.1 cells are increased in sarcoidosis and are more prevalent than T-helper type 1 cells. *Am J Respir Crit Care Med* 2016;193:1281–91.
- Kaiser Y, Lepzien R, Kullberg S, *et al*. Expanded lung T-bet+RORγT+ CD4+ T-cells in sarcoidosis patients with a favourable disease phenotype. *Eur Respir J* 2016;48:484–94.
- Gagliani N, Amezcua Vesely MC, Iseppon A, *et al*. Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* 2015;523:221–5.
- Miyara M, Amoura Z, Parizot C, *et al*. The immune paradox of sarcoidosis and regulatory T cells. *J Exp Med* 2006;203:359–70.
- Taffin C, Miyara M, Nochy D, *et al*. Foxp3+ regulatory T cells suppress early stages of granuloma formation but have little impact on sarcoidosis lesions. *Am J Pathol* 2009;174:497–508.
- Prasse A, Zissel G, Lützen N, *et al*. Inhaled vasoactive intestinal peptide exerts immunoregulatory effects in sarcoidosis. *Am J Respir Crit Care Med* 2010;182:540–8.
- Broos CE, van Nimwegen M, Kleinjan A, *et al*. Impaired survival of regulatory T cells in pulmonary sarcoidosis. *Respir Res* 2015;16:108.
- Moller DR, Koth LL, Maier LA, *et al*. Rationale and design of the genomic research in alpha-1 antitrypsin deficiency and sarcoidosis (GRADS) study. sarcoidosis protocol. *Ann Am Thorac Soc* 2015;12:1561–71.
- Hunninghake GW, Costabel U, Ando M, *et al*. ATS/ERS/WASOG statement on sarcoidosis. American thoracic Society/European respiratory Society/World association of sarcoidosis and other granulomatous disorders. *Sarcoidosis Vasc Diffuse Lung Dis* 1999;16:149–73.
- Eakin EG, Resnikoff PM, Prewitt LM, *et al*. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998;113:619–24.
- Ware J, Kosinski M, Keller SD. A 12-Item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:p. 220–33.
- Kalkanis A, Yucel RM, Judson MA. The internal consistency of pro fatigue instruments in sarcoidosis: superiority of the PFI over the Fas. *Sarcoidosis Vasc Diffuse Lung Dis* 2013;30:60–4.

- 31 Khanna D, Mittoo S, Aggarwal R, *et al.* Connective tissue disease-associated interstitial lung diseases (CTD-ILD) - report from omeract CTD-ILD Working group. *J Rheumatol* 2015;42:2168–71.
- 32 Collard HR, King TE, Bartelson BB, *et al.* Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;168:538–42.
- 33 Su R, Li MM, Bhakta NR, *et al.* Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. *Eur Respir J* 2014;44:985–93.
- 34 Judson MA, Costabel U, Drent M, *et al.* The wasog sarcoidosis organ assessment instrument: an update of a previous clinical tool. *Sarcoidosis Vasc Diffuse Lung Dis* 2014;31:19–27.
- 35 Swigris JJ, Yorke J, Sprunger DB, *et al.* Assessing dyspnea and its impact on patients with connective tissue disease-related interstitial lung disease. *Respir Med* 2010;104:1350–5.
- 36 Swigris JJ, Han M, Vij R, *et al.* The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respir Med* 2012;106:1447–55.
- 37 Judson MA, Mack M, Beaumont JL, *et al.* Validation and important differences for the sarcoidosis assessment tool. a new patient-reported outcome measure. *Am J Respir Crit Care Med* 2015;191:786–95.
- 38 Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005;26:319–38.
- 39 Graham BL, Brusasco V, Burgos F, *et al.* 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur Respir J* 2017;49:1600016.
- 40 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:550.
- 41 Newman AM, Steen CB, Liu CL, *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773–82.
- 42 Newman AM, Liu CL, Green MR, *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453–7.