AWARD NUMBER: W81XWH-19-1-0416

TITLE: Integrated Molecular Pathogenesis of Pulmonary Fibrosis

PRINCIPAL INVESTIGATOR: Dr. Nicholas Banovich, PhD

CONTRACTING ORGANIZATION: Translational Genomics Research Institute, Phoenix, AZ

REPORT DATE: November 2023

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
November 2023	Final	01Aug2019-31Jul2023

4. TITLE AND SUBTITLE

Integrated Molecular Pathogenesis of Pulmonary Fibrosis

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-19-1-0416

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Nicholas Banovich, PhD, E-Mail: nbanovich@tgen.org
Dr. Jonathan Kropski, MD, Email: jon.kropski@vumc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Translational Genomics Research Institute, 445 N 5th St, Phoenix AZ, 85004

Vanderbilt University Medical Center,
1211 Medical Center Dr., Nashville, TN
37232

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Pulmonary fibrosis (PF) is a heterogeneous clinical syndrome that represents the end-stage of chronic interstitial lung diseases. Dozens of different occupational, environmental, immune and genetic risk factors have been associated with PF, and through the past several decades, risk factor exposures have been the driving force in the diagnostic classification of PF, thus in the current paradigm, there are dozens of different "diagnoses" of pulmonary fibrosis. This emphasis on distinction has focused much attention on the most "common" form of this syndrome (Idiopathic Pulmonary Fibrosis, IPF), which comprises only 20% of PF patients. Today there are 2 modestly effective FDA-approved treatments for IPF; however, for the 80% of PF patients with other diagnoses, there are no known effective treatments. The current paradigm emphasizing diagnostic distinction has limited progress in understanding how different risk factors lead to a common end-stage lung pathology. In order to rapidly accelerate progress towards better treatments for all PF patients, a radical departure from this approach is needed. We believe any subdividing of PF should be driven by demonstrated relevant differences in disease biology; to this end, it has become clear that a more nuanced understanding of "upstream" disease mechanism of disease initiation and propagation, as well as the convergent "downstream" mechanisms of lung fibrosis is critical. By leveraging the inherent heterogeneity of disease state in the PF lung, we will employ innovative single-cell genomic approaches – in particular single cell RNA-seq (scRNA-seq) and culture models to recreate the molecular natural history of disease, determine the convergent mediators and pathways that drive PF pathogenesis and identify mechanistically-driven disease endotypes.

15. SUBJECT TERMS
Single cell RNA-seq, pulmonary fibrosis

16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
Unclassified	Unclassified	Unclassified	Unclassified	10	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. Introduction

Pulmonary fibrosis (PF) is a heterogeneous clinical syndrome that represents the end-stage of chronic interstitial lung diseases. Dozens of different occupational, environmental, immune and genetic risk factors have been associated with PF, and through the past several decades, risk factor exposures have been the driving force in the diagnostic classification of PF, thus in the current paradigm, there are dozens of different "diagnoses" of pulmonary fibrosis. This emphasis on distinction has focused much attention on the most "common" form of this syndrome (Idiopathic Pulmonary Fibrosis, IPF), which comprises only 20% of PF patients. Today there are 2 modestly effective FDA-approved treatments for IPF; however, for the 80% of PF patients with other diagnoses, there are no known effective treatments. The current paradigm emphasizing diagnostic distinction has limited progress in understanding how different risk factors lead to a common end-stage lung pathology. In order to rapidly accelerate progress towards better treatments for all PF patients, a radical departure from this approach is needed. We believe any subdividing of PF should be driven by demonstrated relevant differences in disease biology; to this end, it has become clear that a more nuanced understanding of "upstream" disease mechanism of disease initiation and propagation, as well as the convergent "downstream" mechanisms of lung fibrosis is critical. By leveraging the inherent heterogeneity of disease state in the PF lung, we will employ innovative single-cell genomic approaches – in particular single cell RNA-seq (scRNA-seq) and culture models to recreate the molecular natural history of disease, determine the convergent mediators and pathways that drive PF pathogenesis and identify mechanistically-driven disease endotypes.

2. Keywords

Pulmonary fibrosis
Interstitial lung disease
Genomics
Single cell RNA sequencing (scRNA-seq)

3. Accomplishments

What were the major goals of the project?

This project consists of three specific aims:

Aim 1. Profile the clinical, cellular and molecular landscape of PF lungs.

Aim 2: Determine the conserved cell-type specific gene expression programs driving PF pathogenesis

Aim 3. Determine the mechanisms underlying molecular endotypes of pulmonary fibrosis.

What was accomplished under these goals?

Aim 1: Our objective was to collect and process non-IPF ILD lungs from 75 patients. During the course of the award, we successfully collected lung tissue from 75 non-IPF ILD lungs and performed scRNA-seq. Our lung samples include a diversity of diagnoses including unclassifiable ILD, CTD-ILD, chronic hypersensitivity pneumonitis, coal worker pneumoconiosis (CWP), IPAF, NSIP, sarcoidosis, and post-ARDS pulmonary fibrosis. From the majority of these samples were able to collect samples from both a highly fibrotic and a less affected region of the lung; in the remaining cases, disease was homogenous and samples from multiple regions were pooled for sequencing. The data quality has remained excellent, nearly all samples passing our quality control thresholds for cells captured, genes per-cell, and mitochondrial content. The proposed research plan included joint analysis of data from these 75 lungs with control and IPF lung samples collected through an independent project in our group (the control and IPF data generation was not supported by DOD funds), and during the course of this project,

lung tissue samples were collected from 80 control and 56 IPF lungs for scRNA-seq. As outlined below, these data have been jointly analyzed as planned in support of the research aims of the project.

Tasks by PI:
Sample acquisition – Banovich and Kropski
Single cell library generation – Banovich and Kropski
Sequencing – Banovich

Aim 2: As anticipated, data analysis was performed iteratively in support of our specific aims, and new research questions focusing on subsets of the larger dataset emerged during the course of the project. For the primary analysis, we have generated an integrated dataset including >250 unique libraries and performed compositional analyses and cell-type specific differential expression testing as planned across clinical diagnoses. Key findings included 1) the observation that KRT5-/KRT17+ “aberrant basaloid cells” can be found with different abundance and distinct “activation” states in different forms of ILD, 2) SCGB3A2+/SFTPB+ distal airway secretory cells (variably termed RASC, TASC, or TR-BSC by different groups) constitute a larger proportion of the distal lung epithelium in most forms of ILD compared to controls, 3) CTHRC1+ activated fibroblasts shared a relatively similar phenotype across forms of fibrotic ILD, 4) COL15A1+ “systemic-venous-like” endothelial cells are found not only in IPF lungs, but also other forms of ILD, and 5) a diversity of macrophage phenotypes can be found in PF lungs, and these differ across diagnoses. (**Figures 1-3**).

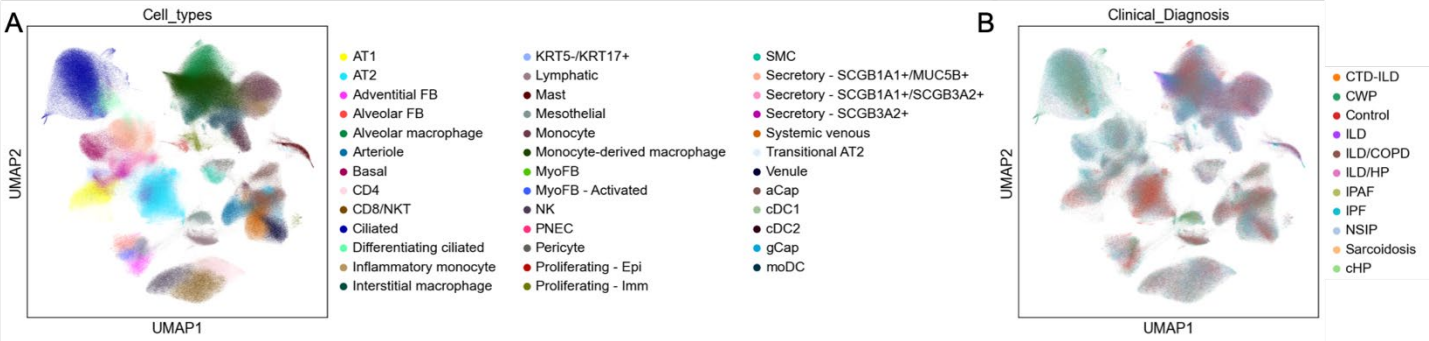

Figure 1. UMAP embedding of scRNA-seq of >1.1 million cells from >250 libraries prepared from subjects with non-IPF forms of ILD (jointly analyzed with IPF and control lung data generated through an independent project). UMAP is annotated by A) Cell-type and B) Clinical diagnosis.

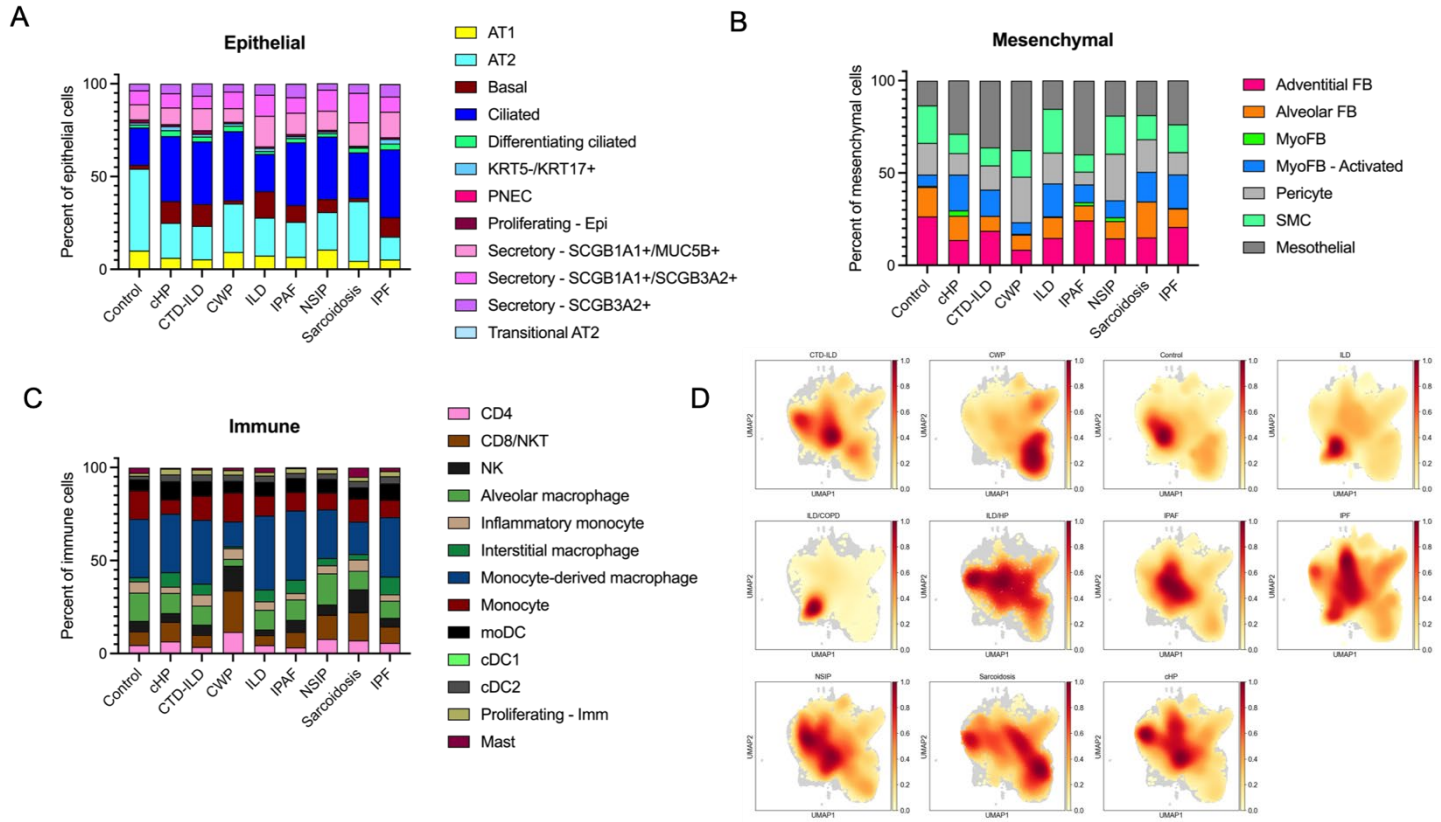

Figure 2. A-C) Quantification of proportions of cell-types within a given lineage across ILD diagnoses. D) Density embedding of macrophage phenotypes across ILD diagnoses.

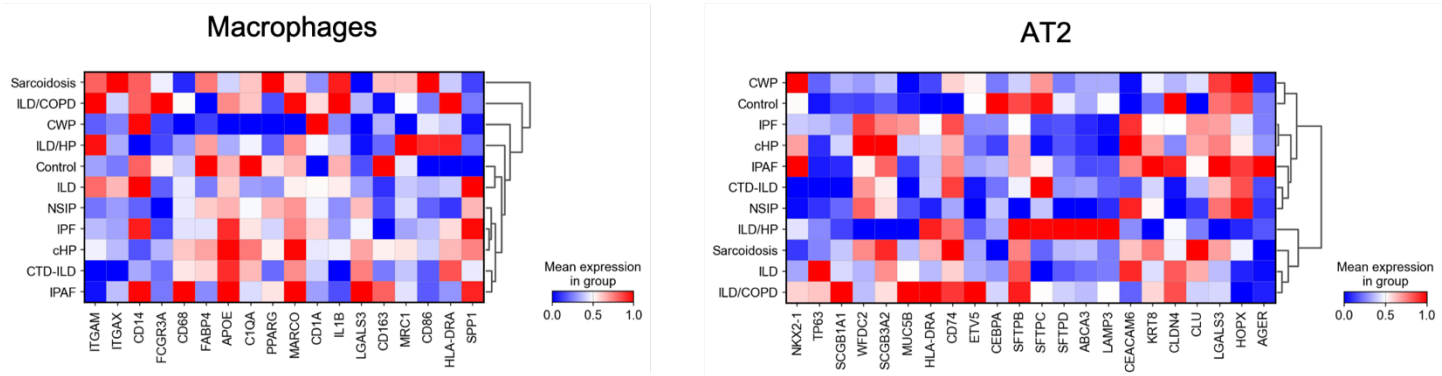

Figure 3. Heatmap depicting expression of selected differentially expressed markers in macrophages and AT2 cells across diagnosis. Dendrograms depict hierarchical clustering based on marker expression.

A second major objective of this aim was to perform single-cell expression-quantitative trait mapping of the human lung in a disease-state-informed manner. These results have been released as a preprint on bioRxiv and are currently in revision at Nature Genetics. This study, which is first reported comprehensive solid-tissue based single-cell eQTL of any organ, had a number of novel and exciting findings which are described in detail in the manuscript. Briefly, we found that eQTL can be subdivided into multiple subtypes, including those with universal, multi-state, or cell-type specific gene regulatory effects. Cell-type specific sc-eQTL tend to be located further from transcription start sites (often in enhancer regions) and exhibit larger effect sizes. We also found that eQTL sharing is greatest among cells within a common lineage (although this is not universal). Finally, we identified a subset of sc-eQTL which unexpectly exhibit significant differences in their regulatory effects in control compared to disease tissue (i.e. different direction of effect or statistically different magnitude of effect). Intriguingly, these disease-interacting sc-eQTL are enriched within promoters of stress-induced transcription factors,

implicating a previously unrecognized role of genetic variation in mediating cellular responses to exogenous stimuli.

Finally, we anticipated that generation of this rich and robust dataset would facilitate new collaborations could be analyzed in support of additional projects as this study progressed. To this end, we developed new collaborations through which we became major contributors to numerous ongoing lung single-cell atlasing projects through the Human Cell Atlas Lung Biological Network. During the COVID-19 pandemic, we led a focused analysis of these data with the objective of elucidating the cell-type specific, disease-state informed regulations of SARS-CoV-2 entry factors; this manuscript was published in *Nature Communications*. More recently, data generated through this project contributed to the recently released Human Lung Cell Atlas v1.0, including more than 2.2 millions cells from 444 individuals; work from our group represented nearly 25% of the samples included. From a methodological perspective this manuscript compared approaches to data integration, normalization, and batch correction leading to a set of best practices for the field. Furthermore, this work demonstrated how variation in analysis pipelines can lead to mislabeling or underlabeling of cell-types. These results were published in Nature Medicine in 2023.

To summarize, the aims and objectives of this aim were completed and exceeded during the course of this project.

Tasks by PI:
Data integration: Banovich and Kropski
Cell-type annotation: Kropski
Differential expression analysis: Banovich
HCA analysis: Banovich and Kropski

Aim 3: We have completed diagnostic adjudication and disease endotype analyses have been performed jointly with clinical-diagnosis based analyses. The results of the unbiased analyses are not entirely straightforward, as in some cases consensus clustering based on lineage-specific PCA yields distinct sample groupings when seeded using different lineages, while compositional changes are sensitive to differences in regional sampling. As a result, we have focused primarily on our hypothesis-driven approaches. In a collaboration with Dr. Franck McCormack (U. Cincinnati) that emerged during the course of this project, we observed that in lungs from subjects with one subtype of ILD (coal-worker's pneumoconiosis), a subset of lung macrophages adopt an osteoclast-like phenotype which promotes fibrosis (this manuscript is currently under review at *Science Advances*); this phenotype is also observed in a subset of samples from other forms of ILD. Another key observation was that a subset of IPF and other ILD samples share evidence of hypoxia-inducible factor activation in the distal lung epithelium, particularly in KRT5-/KRT17+ cells. In a series of experiments utilizing ex-vivo/organoid models, we found that persistent activation of HIF2 specifically is a hallmark of certain ILD lungs, and HIF2 drives polarization of mucous secretory cells and aberrant basaloid cells in the human lung epithelium, highlighting HIF2 as a potential novel therapeutic target for PF. These studies, led by Dr. Scott McCall, a postdoctoral fellow in Dr. Kropski's group, have been reported as a preprint on *bioRxiv* and are under review at *Science Translational Medicine*. To summarize, the primary objectives of this aim have been completed as anticipated.

What opportunities for training and professional development has the project provided? Dr. Heini Natri (Dr. Banovich's group) received new training around computational analyses to perform the work carried out here. Dr. Nichelle Winters (a postdoctoral fellow in Dr. Kropski's group) gained experience analyzing genomic data and led the COVID-19 focused Nature Communications manuscript along with Dr. Lin Bui from Dr. Banovich's group. Dr. Jason Gokey (junior faculty member in Dr. Kropski's group) gained experience contributing to the interpretation of genomic data. Dr. Scott McCall

(postdoctoral fellow in Dr. Kropski's group, supported by a T32 grant) gained experience establishing organoid models and led the HIF2-focused work.

How were the results disseminated to communities of interest? In addition to the manuscripts this work was presented at numerous international meetings including the American Thoracic Society in May 2022 and 2023, the 2022 Scleroderma Workshop, and the 2022 FASEB Lung Epithelium in Health and Disease meeting, 2023 Gordon Lung Injury and Repair Conference, and the American Society of Human Genetics meetings.

What do you plan to do during the next reporting period to accomplish the goals? N/A

4. Impact
What was the impact on the development of the principal discipline(s) of the project?
Nothing to Report.
What was the impact on other disciplines?
Nothing to Report.
What was the impact on technology transfer?
Nothing to Report.
What was the impact on society beyond science and technology?
Nothing to Report.

5. Changes/Problems
We completed a 1 year NCE

6. Products
Publications, conference papers, and presentations:
Journal publications.
Habermann AC[†], Gutierrez AJ[†], Bui LT[†], Yahn SL, Winters NI, Calvi CL, Peter L, Chung MI, Taylor CJ, Jetter C, Raju L, Roberson J, Ding G, Wood L, Sucre JMS, Richmond BW, Serezani AP, McDonnell WJ, Mallal SB, Bacchetta MJ, Loyd JE, Shaver CM, Ware LB, Bremner R, Walia R, Blackwell TS, **Banovich NE*, Kropski JA***. Single cell RNA-sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Science Advances* 2020. 06:eab1972. PMCID: PMC9439444.

Bui LT*, Winters NI*, Chung MI, Joseph C, Guiterrez AJ, Habermann AC, Adams TS, Schupp JC, Poli S, Peter LM, Taylor CJ, Blackburn JB, Richmond BW, Nicholson AG, Rass D, Wallace WA, Jenkins RG, Kaminski N, **Kropski JA**, Banovich NE**** and the Human Cell Atlas Lung Biological Network. Single-cell RNA-sequencing reveals dysregulation of molecular programs associated with SARS-CoV-2 severity and outcomes in patients with chronic lung disease. *Nature Communications* 2021. 12:4314. PMCID: PMC8280215.

Sikkema L, Strobl DC, Zappia L, Madissoon E, Markov NS, Zaragosi LE, Ansari M, Arguel MJ, Apperloo L, Bevavin C, Merg M, Chichelnitskiy E, Chung MI, Collin A, Gay ACA, Hooshiar B, Kashani BH, Jain M, Kapellos T, Kole ™, Mayr CH, Von Ppane M, Peter L, Ramirez-Suastegue C, Schniering J, Taylor CJ, Walzthoeni T, Xu C, Bui LT, de Donno C, Dony L, Guo M, Gutierrez AJ, Heumos L, Huang N, Ibarra IL, Jackson ND, Murthy PKL, Lotfollahi M, Tabib T, Talavera-Lopez C, Travaglini K, Wilbrey-Clark A, Worlock KB, Yoshia M, Desai T, Eickelberg O, Falk C, Kaminski N, Krasnow MA, Lafyatis R, Nikoli MZ, Powell JE, Rajagopal J, Rozenblatt-Rosen O, Seibold MA, Sheppard D, Shepherd DP, Teichmann SA, Tsankov AM, Whitsett J, Xu Y, **Banovich NE**, Barbry P, Duong TE, Meyer KB, **Kropski JA**, Pe'er D, Schiller HB, Tata PR, Schultze JL, Misharin AV, Nawijn MC, Luecken MD, Theis FJ. An integrated cell atlas of the human lung in health and disease. *Nature Medicine* 2023.29:1563. PMCID – PMC10287567.

Natri HM, Azodi CBD, Peter L, Taylor CJ, Chugh S, Kendle R, Chung MI, Flaherty DK, Matlock BK, Calvi CL, Blackwell TS, Ware LB, Bacchetta M, Wali R, Shaver CMS, **Kropski JA***, McCarthy DJ*, **Banovich NE***. Cell-type specific and disease-associated eQTL in the human lung. *bioRxiv* 2023.03.17.533161. [in revision - *Nature Genetics*]

McCall AS, Tanjore H, Burman A, Sherrill T, Chapman M, Calvi CL, Camarata J, Hunt RP, Nichols D, **Banovich NE**, Lawson WE, Gokey JJ, **Kropski JA***, Blackwell TS*. Hypoxia-inducible factor-2 (HIF2) regulates alveolar regeneration after repetitive injury. *bioRxiv* 2023. doi.org/10.1101/2023.09.17.557477. [Under review - *Science Translational Medicine*]

Hasegawa Y, Franks J, Tanaka Y, Uehara Y, Read D, Williams C, Srivastan S, Pitstick LB, Nikolaidis N, Shaver CM, **Kropski JA**, Ware LB, Taylor CJ, **Banovich NE**, We H, Gardner J, Osterbug A, Yu J, Kropas E, Teitelbaum S, Wikenheiser-Brokamp K, Trapnell C, McCormack FX. Pulmonary osteoclast-like cells in silica-induced pulmonary fibrosis. 2023 [Under review - *Science Advances*]

Books or other non-periodical, one-time publications.
Nothing to Report.
Other publications, conference papers, and presentations.
American Thoracic Society 2022, 2023
Website(s) or other Internet site(s)
Nothing to Report.
Technologies or techniques
Nothing to Report.
Inventions, patent applications, and/or licenses
Nothing to Report.
Other Products
Nothing to Report.

7. Participants & Other Collaborating Organizations

What individuals have worked on the project? (Level of Effort reporting for NCE period)

Name: Nicholas Banovich, PhD
Project Role: P.I.
Nearest Person month worked: 2.4 (NCE)
Contribution to the project: Dr. Banovich, along with co-PI Dr. Kropski, oversees the overall aims of this project, in particular the genomic and computational analyses. He participates in data interpretation, presentation, and publication.

Name: Heini Natri, PhD
Project Role: Computational Scientist
Nearest Person month worked: 12 (NCE)
Contribution to the project: Data processing and analysis.

Name: Lance Peter, MSc
Project Role: Research Associate
Nearest person month worked: 8 (Year 1-3)
Contribution to the project: Sample processing and library preparation.

Name: Brandon Fischer
Project Role: Research Associate
Nearest Person month worked: 3 (NCE)
Contribution to Project: Sample processing and library preperation

Name: Evan Mee

Project Role: Bioinformatician
Nearest Person month worked: 3 (NCE)
Contribution to Project: Data processing.

Name: Annika Vanna
Project Role: Post-doc Fellow
Nearest Person month worked: 2
Contribution to Project: Data processing and analysis.

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nicholas Banovich
New: NIH/NHGRI- Integrated analysis of multi-omic QTLs at single cell resolution
New: Marcus Foundation - CAR-T cell priming of an endogenous immune response to overcome antigen escape
New: NIH - Integrated Molecular and Cellular Drivers of Alveologenesis

What other organizations were involved as partners?

Organization name: Norton Thoracic Institute at Dignity Health
Location of Organization: Phoenix Arizona
Partner's contribution to the project: Collaboration

- The Norton Thoracic Institute collects biopsies from explant lungs to be used in this study. Rajat Walia, MD, serves as the primary collaborator.

8. Special Reporting Requirements

Collaborative Award

9. Appendices

N/A