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: August 2022

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
August 2022	Annual	01Aug2021-31Jul2022

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
Dr. Jonathan Kropski, MD

E-Mail: nbanovich@tgen.org, 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** USAMRDC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	8	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

	Page
1. Introduction	4
2. Keywords	4
3. Accomplishments	4
4. Impact	6
5. Changes/Problems	7
6. Products	7
7. Participants & Other Collaborating Organizations	7
8. Special Reporting Requirements	8
9. Appendices	8

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 object is to collect and process non-IPF ILD lungs from 75 patients. At the end of our last reporting period, we had collected 42 lung samples with completed scRNA-seq on 30 of the samples. In this reporting period we collected an additional 42 samples, for a total of 84. At present we have completed scRNA-seq on 50 of these samples and expect to complete scRNA-seq data generation and analysis during the requested NCE. 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. In addition, while not directly supported by DOD funds but with planned utilization of the data for integrated analyses as a part of

these studies, we have now also generated scRNA-seq libraries and/or performed sequencing from 67 IPF lungs and 55 control lungs.

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

Aim 2: Our data analysis is performed iteratively, as batches of library sequencing are completed. We have begun analyzing the scRNA-seq data generated to identify cell-type-specific gene expression changes associated with PF. To this end, we have integrated this dataset with a second dataset, generated by us under the support of the NIH/NHLBI, which consists of lung tissue from 55 controls (declined donors) and 67 IPF patients. Together, this is the most comprehensive dataset of scRNA-seq data from PF. Using a negative binomial regression framework, we have identified hundred of differentially expressed (DE) genes in the majority of cell-types. We have made comparisons between control lung and all disease, as well as between control lung and individual diagnosis – identifying both conserved and specific DE programs.

Additionally, we have remained significant contributors large scale atlasing efforts, specifically through the Human Cell Atlas Lung Network. These efforts are essential to generating robust and well harmonized cell-type annotations across the field. In this reporting period our work in this area resulted in a manuscript deposited on bioRxiv (Sikkema et al. 2022), which we anticipate will be published soon. This work carried out a meta-analysis of over 2.2 million cells from 444 individuals representing both healthy individuals and those with chronic lung disease. This work is the most complete single cell genomics based characterization of lung tissue to date. 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. Importantly, this work carried out a comprehensive analysis of processing pipelines (**Figure 2D** of attached manuscript), and demonstrated how variation in analysis pipelines can lead to mislabeling or under labeling of cell-types (**Figure 3C and E** of attached manuscript).

Two additional manuscripts are currently in development related to subsets of data generated through this project. We continue to collaborate with Dr. Frank McCormack (U. Cincinnati) on analysis of samples with coal-worker's pneumononiosis (now n=7). These data served as preliminary data for a grant application that was submitted in 2022.

Additional analysis is ongoing examining post-COVID PF samples (now n=4) at different timepoints after infection. We anticipate submitting this as a manuscript during the coming year.

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

Aim 3: At present, statistical power for endotype analyses is still limited, and we plan to begin these analyses once data generation is completed – within the next 3 months. We have continued structured analyses of the imaging and pathology. There has been continued progress generating and evaluating different organoid model systems. In particular, we have demonstrated that regionally distinct fibroblast populations isolated from ILD lungs maintain distinct cellular profiles and

phenotypes during ex-vivo culture (Figure 2).

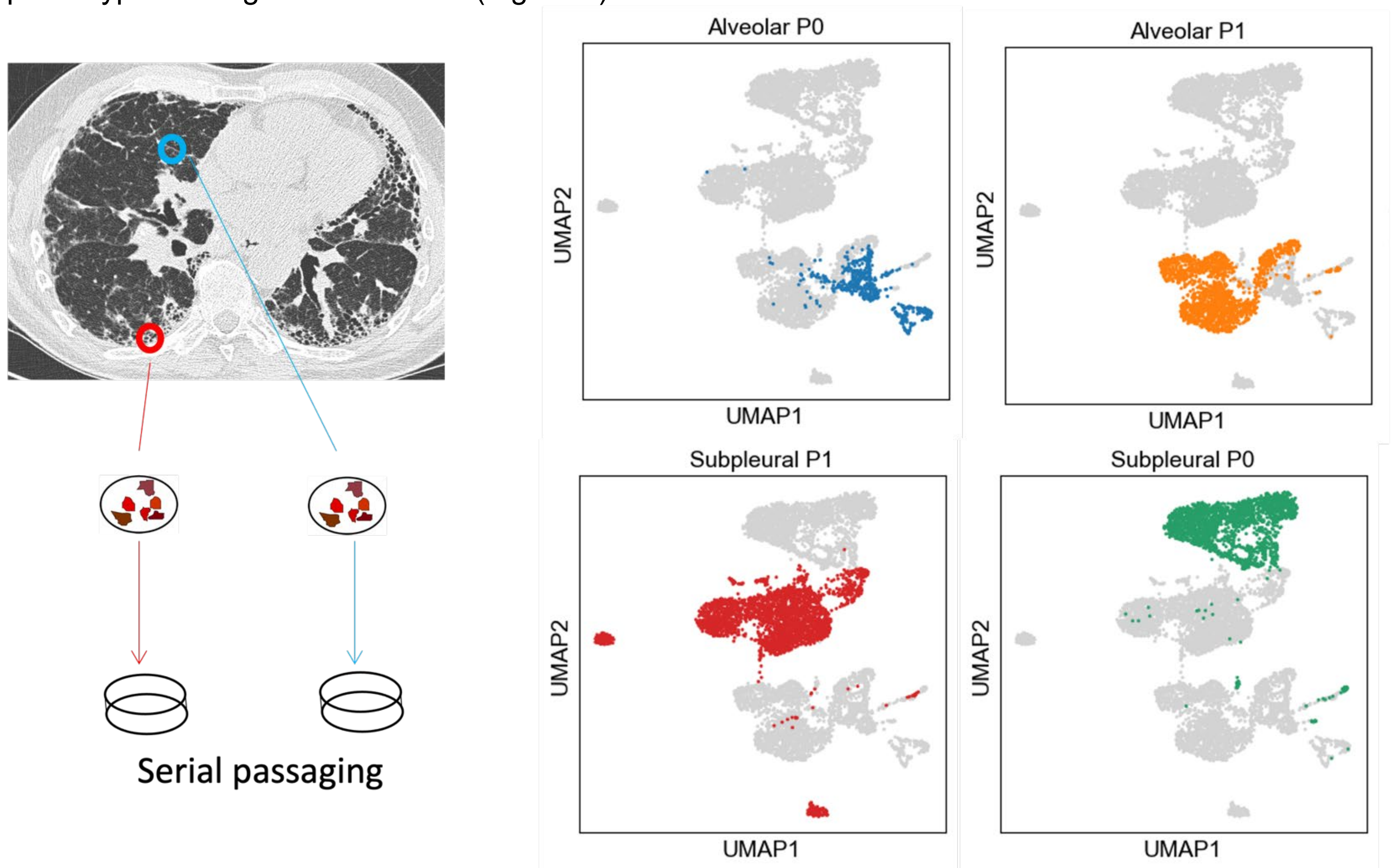

Figure 2. Subpleural and alveolar fibroblasts were isolated and cultured ex-vivo. scRNA-seq was performed at passage 0 or after passage 1. UMAP embeddings depict different samples/timepoints.

In addition, we have made considerable progress using feeder-free organoid models. We have successfully cultured freshly isolated AT2 cells using serum-free, feeder-free organoid media, and experiments are ongoing comparing efficiency and identity of organoids generated from distinct distal lung progenitor populations from ILD lungs.

What opportunities for training and professional development has the project provided? Dr. Heini Natri received new training around computational analyses to perform the work carried out here. Dr. Jason Gokey (Instructor 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.

How were the results disseminated to communities of interest? In addition to the bioRxiv manuscript noted above, this work was presented at the American Thoracic Society in May of 2022, the 2022 Scleroderma Workshop, and the 2022 FASEB Lung Epithelium in Health and Disease meeting.

What do you plan to do during the next reporting period to accomplish the goals? We will continue to work towards completion of the goals as outlined in the SOW.

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 requested and were granted a one year NCE to complete the work outlined in the SOW.

6. Products
Publications, conference papers, and presentations:
Journal publications.
Sikkema L et al. An integrated cell atlas of the human lung in health and disease. bioRxiv. March 11, 2022. https://doi.org/10.1101/2022.03.10.483747

Books or other non-periodical, one-time publications.
Nothing to Report.
Other publications, conference papers, and presentations.
American Thoracic Society 2022
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?
Name: Nicholas Banovich, PhD
Project Role: P.I.
Nearest person month worked: 3
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: 9
Contribution to the project: Dr. Natri has taken over the primary analysis role on this project. She worked on data processing and analysis.
Name: Lance Peter, MSc
Project Role: Research Associate
Nearest person month worked: 8
Contribution to the project: Mr. Peter is responsible for processing primary lung tissue from patients with ILD, including 10X genomics library preparation and next generation sequencing. Mr. Peter is also responsible for sample management and organization.

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, PhD
New: Chan Zuckerberg Initiative: Immune Cell Atlas of Environmental and Ancestral Diversity in Indonesia
Ended: None

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
Nothing to report