AWARD NUMBER: **W81XWH-22-1-0629**

TITLE: Unraveling Molecular Mechanisms Underlying Chronic Obstructive Pulmonary Disease Heterogeneity Using Combinatorial Barcoding and Single-Nuclear RNA Sequencing

PRINCIPAL INVESTIGATOR: Maor Sauler

CONTRACTING ORGANIZATION: Yale University, New Haven, CT

REPORT DATE: July 2023

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 is unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
July 2023

2. REPORT TYPE
Annual

3. DATES COVERED
01Jul2022-30Jun2023

4. TITLE
Unraveling Molecular Mechanisms Underlying Chronic Obstructive Pulmonary Disease Heterogeneity Using Combinatorial Barcoding and Single-Nuclear RNA Sequencing

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-22-1-0629

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
PR

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

Maor Sauler

E-Mail: maor.sauler@yale.edu

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

Yale School of Medicine
300 Cedar Street
New Haven, CT 06520

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

Chronic Obstructive Pulmonary Disease (COPD) is the 4th leading cause of death due to chronic illness in the United States, and is particularly common amongst current and former United States military personnel. COPD is a heterogeneous disorder characterized by diverse clinical and pathologic manifestations that vary amongst susceptible individuals. However, the profound complexity of COPD heterogeneity has made it challenging to link clinical features with cellular mechanisms and develop targeted therapies. Our goal is to generate comprehensive single-cell transcriptional profiles of lung tissue samples from the full spectrum of disease. Then, we will use systems-based computational approaches to identify cell-specific molecular phenotypes in COPD (Aim 1), novel endotypes and their associated biomarkers (Aim 2), and cell-targeted therapies (Aim 3). Our hypothesis for Aim 1 was that Cell-specific transcriptional phenotypes underlie progression of COPD-related disease traits. We have performed single nuclear RNAseq of COPD and disease-free lung tissue samples and are analyzing the results currently.

15. SUBJECT TERMS
None listed.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USMRDC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b PHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	15	

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

TABLE OF CONTENTS

1. INTRODUCTION:

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of death due to chronic illness in the United States. COPD is defined by mostly irreversible airflow obstruction and is commonly caused by chronic exposure to aerosolized toxins such as cigarette smoke (CS). However, COPD is a heterogenous syndrome. The clinical and pathologic manifestations of COPD vary amongst susceptible individuals, and the extent to which specific disease mechanisms manifest within a given cell type over time or between patients is not well-known. This lack of understanding of COPD heterogeneity represents a major gap in knowledge limiting the development of targeted therapies. Our goal is to identify cell-specific mechanisms underlying the heterogeneous manifestations of COPD using scalable single-nuclear RNA sequencing (snRNAseq) methods and translate those findings into precision-based therapies.

2. KEYWORDS:

COPD, emphysema, single-cell RNA sequencing, single-nuclear RNA sequencing, airway disease, lung, smoking.

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Below I have included Tasks and milestones that fall within Year 1 of this DOD annual report.

Aim 1. Define cellular phenotypes and trajectories associated with clinicopathologic features of COPD. **Hypothesis 1:** Cell-specific molecular phenotypes underlie progression of COPD-related disease traits.		
Major Task 1: Regulatory and Registration	Months	
• Acquire IACUC approval	1-2	Sauler
Major Task 2: Single-Nuclear sequencing of COPD lung tissue	Months	
• Perform single-nuclear sequencing on lung tissue samples	1-6	Sauler, McDonough, Kaminski
• Ensure appropriate quality control and replication of poor samples if/when needed	6-8	Sauler, McDonough, Kaminski
Major Task 3: Identification of aberrant cellular phenotypes	Months	
• Identify cellular clusters associated with disease features of COPD	8-18	Sauler, McDonough, Kaminski, Yan

- Utilize trajectory inference to identify aberrant cellular phenotypes	8-18	Sauler, McDonough, Kaminski, Yan
- Identify Differentially expressed genes amongst specific cell-types in COPD	8-18	Sauler, McDonough, Kaminski, Yan
Milestone 1: Successful single-nuclear sequencing	Completed	
Milestone 2: Identification of aberrant cell populations	Started (Major Task 3)	

What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

Major Task 1: We have successfully acquired IACUC approval. Protocol Number

Major Task 2: We made use of the curated biospecimens from the Lung Tissue Resource Consortium (LTRC), a biobank established by the National Institute of Health (NIH) to collect and distribute lung tissue, blood samples, clinical data, and radiographic studies from participants with chronic lung disease. The data dictionary of the LTRC includes 1670 variables including lung function testing, pathologic characterization of biopsied tissue, results of questionnaires that assessed disease exacerbations, associated medical illnesses, smoking, environmental, and occupational exposures, and quality of life. There are also lobe specific-radiographic assessments of emphysema, airway wall thickness, and vascular volume, which we can then link back to the lobe from where the tissue biopsy.

We performed snRNAseq on snap frozen lung tissue samples from 142 subjects, with the demographic features described in (**Fig. 1A**). We also generated principal components of the clinical features and identified 10 clusters of patients. While these clusters are unstable, we did this to ensure adequate representation across the full spectrum of the disease (**Fig. 2A**). We have already evaluated the quality of our single-cell sequencing, and we have identified a few key important factors suggesting good quality data.

	n	% Female	Age	Pack-Years	FEV_1 % predicted
Never smokers	16	62%	60 ±12	0	81±24
Smokers w/ preserved function	18	50%	66 ±11	46 ±30	98±13
GOLD I/II	31	58%	69 ±10	55 ± 22	70±25
GOLD III	27	37%	67 ± 6	45 ± 36	38±7
GOLD IV	50	48%	60 ± 7	51 ± 29	20±5

Figure 1 A) Basic demographic features of subjects undergoing single-nuclear sequencing. B) Clustering of all LTRC subjects (n = 142). Columns represent clinical features. Rows represent subjects. Subjects are grouped into clusters. Numbers along y-axis represent the number of subjects per each cluster.

1) Samples cluster by cell type, not by subject (with a few exceptions) and all major cell types in the distal lung are identified (**Fig. 2A, B**).
2) Sequencing depth is similar across samples (**Fig 2C**). There are a few samples that had lower sequencing depth but were still quality samples and we have sent these for resequencing. In addition to read depth, other quality metrics have been assessed including percent.mitochondria, splice fraction, and the existence of doublets. For example, the removal of doublet clusters generates a UMAP projection of the data with clear cell populations (**Fig 2D**). A few samples were poor quality, and they have been removed. But we have maintained representation across disease severity and clusters identified in **Fig 1B**. We have generated a cDNA library based on these samples have been re-sequenced. We are now actively integrating these data into our current data set to finalize our dataset.

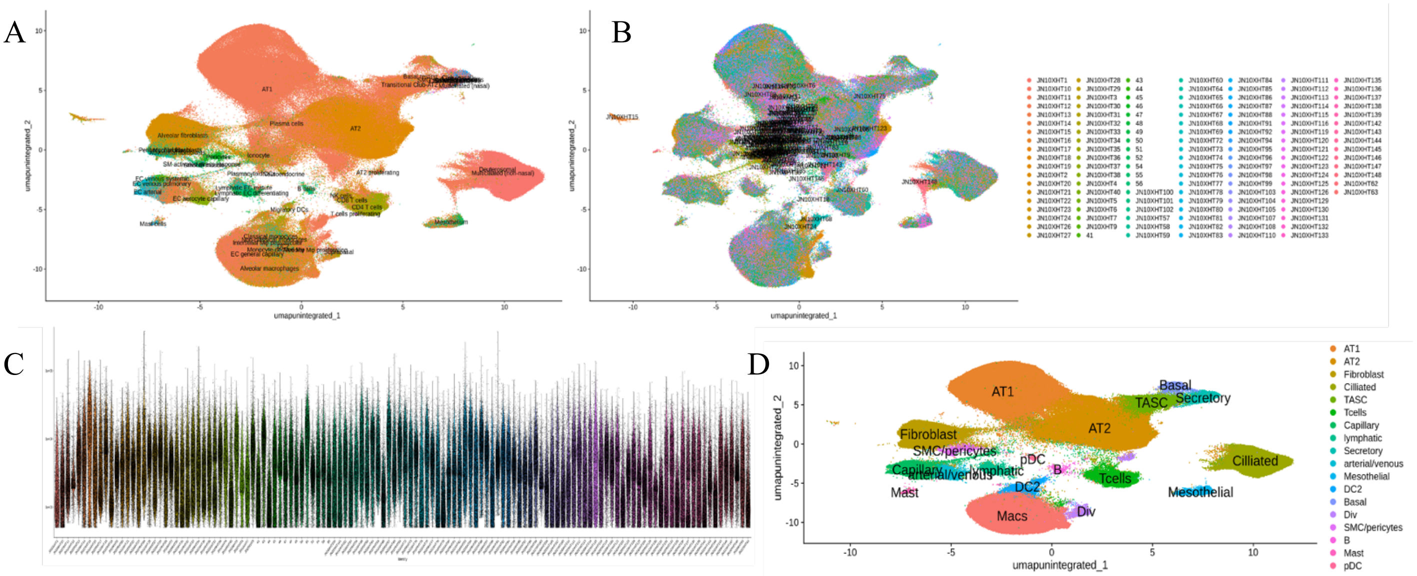


Figure 2 A) Preliminary UMAP of all samples sequenced, labeled by cell type. B) UMAP samples labeled by subject. What is critical is that the data clusters by cell type (A) rather than subject (B). C) Read Depth (n_Count) across all subjects. D) UMAP of samples separated by cell type.

We have focused considerable effort towards analyzing quality metrics and searching for potential problems with the data. This is important; while the generation of figures is easy with single-cell sequencing data, identifying challenges that may affect downstream analyses and ensuring that the data is of the upmost quality will be critical to the success of the project. An important finding is that we can identify major cell populations within subjects using integration (and have tested multiple approaches, including ScVii, Harmony, and Seurat RPCA integration methods). For example, we can identify endothelial and fibroblast cell populations across subjects (**Figures 3A-D**). These cell populations emosntrate canonical marker genes (**Figures 3E-F**). We have also benchmarked different background removal methods for our analysis (CellBender and SoupX). What this has allowed us to do is assess how technical variation in the data may

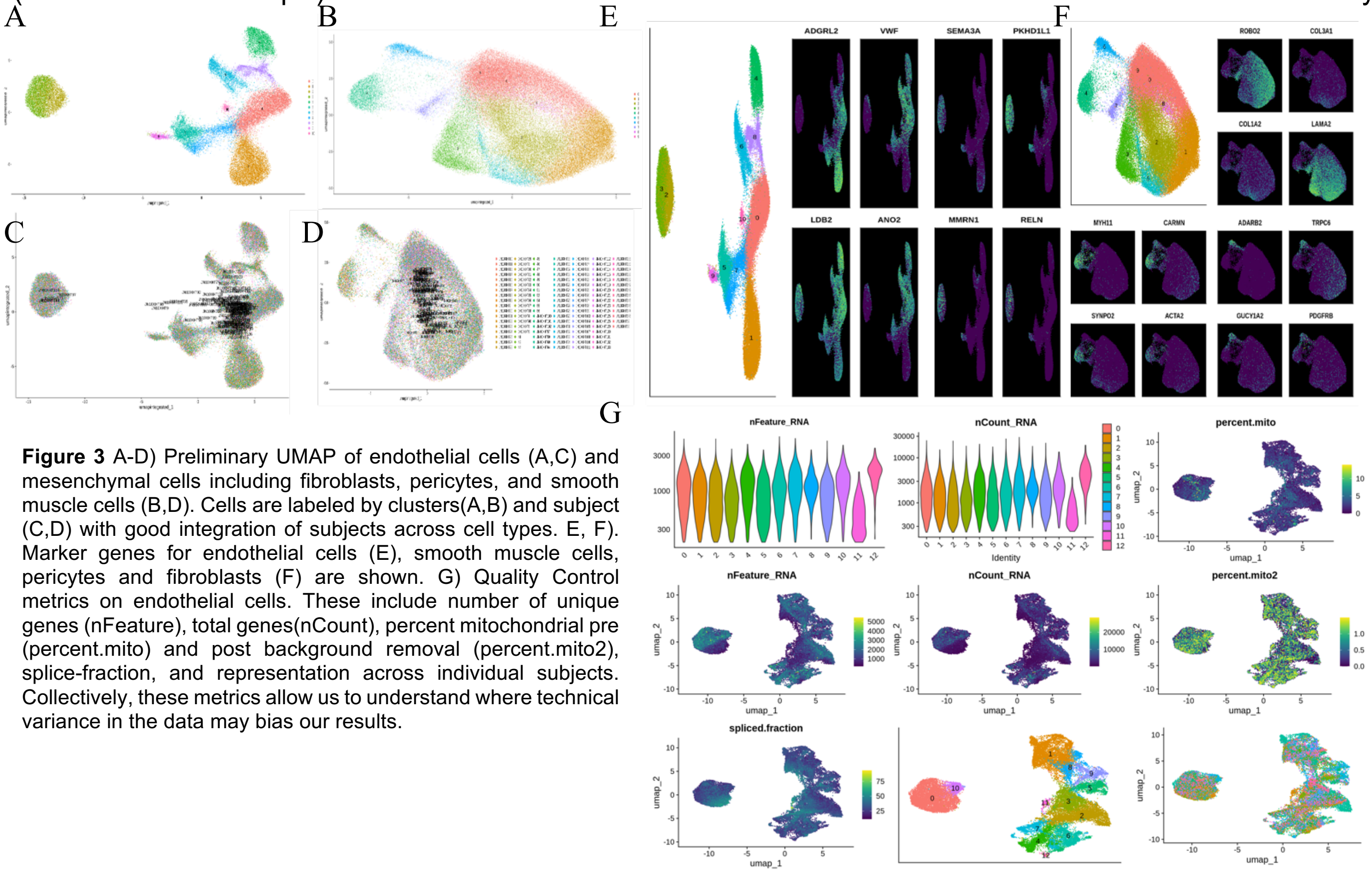
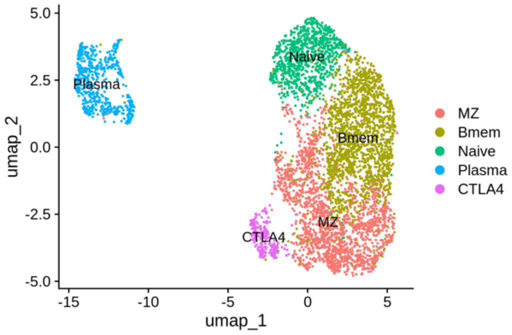


Figure 3 A-D) Preliminary UMAP of endothelial cells (A,C) and mesenchymal cells including fibroblasts, pericytes, and smooth muscle cells (B,D). Cells are labeled by clusters(A,B) and subject (C,D) with good integration of subjects across cell types. E, F). Marker genes for endothelial cells (E), smooth muscle cells, pericytes and fibroblasts (F) are shown. G) Quality Control metrics on endothelial cells. These include number of unique genes (nFeature), total genes(nCount), percent mitochondrial pre (percent.mito) and post background removal (percent.mito2), splice-fraction, and representation across individual subjects. Collectively, these metrics allow us to understand where technical variance in the data may bias our results.

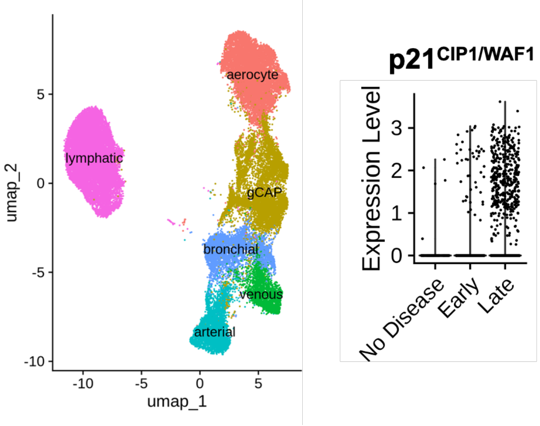
bias our results, and therefore, identify strategies to mitigate these biases. Sample metrics for endothelial cells are shown (**Figure 3G**). All these now allow us to move forward with the final analysis.

Major Task 3:
Preliminary Data: We have started analyzing data to ensure we have an analytic pipeline for successful analysis, as the number of samples for resequencing and for new library prep are small compared to the total number of samples. This mainly involves identifying major cell populations and potential cellular states.

B cells: We have identified a unique population of B-cells with germinal center markers enriched in COPD. We are beginning to explore the potential role of these B-cells as they related to various disease stages and subtypes.



Endothelial Cells: We have already begun analysis on endothelial cells. As previously performed, we have been able to identify all major cell populations of endothelial cells in the lung, and we have preliminary data that there are increased senescence markers in endothelial cells of subjects with late disease.



What opportunities for training and professional development has the project provided?

Nothing to Report

How were the results disseminated to communities of interest?

Preliminary data were presented at the American Thoracic Society Meeting, in May of 2023

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

By next reporting period, we expect to

a) Finish analysis of single-nuclear sequencing data
b) Begin analysis of mouse exposed to cigarette smoke
c) Begin to identify potential biomarkers of disease and potential therapeutic targets.

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:

Changes in approach and reasons for change

Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them

We tried using PARSEseq to analyze our data. However, the results from PARSE did not give us quality data. Therefore, we used 10x 3' sequencing which provided us the results above.

Changes that had a significant impact on expenditures

Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nothing to Report

Significant changes in use or care of vertebrate animals

Nothing to Report

Significant changes in use of biohazards and/or select agents

Nothing to Report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Nothing to report

Books or other non-periodical, one-time publications.

Nothing to report

Other publications, conference papers and presentations.

Preliminary data was reported at the American Thoracic Society meeting, 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**

We anticipate that the final Seurat object will be a database for other investigators to interrogate when it is completed.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: *Reginal Brewster*
Project Role: *post-graduate student*
Researcher Identifier *(e.g. ORCID ID):* *P300268*
Nearest person month worked: *4*

Contribution to Project: *Mr. Brewster helped organize all the clinical data for this project*

Name: *Jessica Nouws*
Project Role: *Associate Research Scientist*
Researcher Identifier *(e.g. ORCID ID):* *P265399*
Nearest person month worked: *3*

Contribution to Project: *Dr. Nouws has performed all the sequencing experiments for this project*

Project Role: *Maor Sauler*
Researcher Identifier *(e.g. ORCID ID):* *P234318*
Nearest person month worked: *Associate Professor*

Contribution to Project: *Dr. Sauler has designed all experiments, organized projects, and lead all analyses*

Project Role: *Xiting Yan*
Researcher Identifier *(e.g. ORCID ID):* *P6503*
Nearest person month worked: *Associate Professor*

Contribution to Project: *Dr. Yan is leading al the biostatistical analyses related to this project*

Project Role: *Jiayi Zhao*
Researcher Identifier *(e.g. ORCID ID):* *N/A*
Nearest person month worked: *stueent*

Contribution to Project: *Ms. Zhao is starting to perform some of the single-cell sequencing analyses related to this project*

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

Nothing to Report

What other organizations were involved as partners?

Nothing to Report

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:

QUAD CHARTS:

9. APPENDICES: