ARIC Manuscript Proposal #3634 (Amended)

PC Reviewed: 12/14/21	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Proteomic Profiling and Pulmonary Function: findings from the Atherosclerosis Risk in Communities (ARIC) study

b. Abbreviated Title (Length 26 characters): Proteomics and PFTs

2. Writing Group:

Writing group members: Bing Yu, Christine Ladd-Acosta, Alanna Morrison, Kari North, Amil Shah, Christie Ballantyne, Eric Boerwinkle and Stephanie J. London (order TBD). Other interested ARIC investigators are welcome to participate.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __BY___ **[please confirm with your initials electronically or in writing]**

First author: Bing Yu
Address: 1200 Pressler Street, Suite E407
Houston, TX 77040
Phone: 713-500-9285 Fax: 713-500-0900
E-mail: bing.yu@uth.tmc.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Stephanie London
Address: NIEHS, PO Box 12233, MD A3-05, RTP NC 27709
Phone: 919-541-5772 Fax:
E-mail: london2@niehs.nih.gov

3. Timeline: The proteome data at visit 2, 3 and visit 5 are available, and analysis is to start as soon as approval is obtained. VWe expect that the manuscript will be prepared within a year from approval of the analysis plan.

4. Rationale:

Pulmonary function, which reflects respiratory health, is a long-term predictor of morbidity and mortality [1-3]. Forced vital capacity (FVC), forced expiratory volume in the first second (FEV1) and its ratio to FVC (FEV1/FVC) are spirometric measures of pulmonary function, which reflect lung volume and are used to diagnose and monitor lung diseases, such as chronic obstructive pulmonary disease (COPD). Impairment of pulmonary function is multifactorial. Several blood

biomarkers, including C-reactive protein (CRP) and fibrinogen have been associated with pulmonary function [4, 5].

Proteomics, which characterizes a set of proteins produced in a biological context, is a novel approach to advance our understanding on the pathology and pathogenesis of various diseases. Several studies have implicated a critical role for specific matrix metalloproteinases (MMPs) proteins in COPD development induced by smoking, and its severe complication, emphysema.[6] A recent proteomic study has reported that protein markers, metalloproteinase inhibitor 3 (TIMP3) and matrix metalloproteinase 28 (MMP-28), are associated with COPD. [7] To date, no large studies have employed targeted proteomics to identify molecular pathways related to pulmonary function and its decline.

This proposed research will use both a cross-sectional and longitudinal design to evaluate the association between the proteome with pulmonary function, its decline and COPD in ARIC African and European Americans to identify novel protein biomarkers to improve the understanding of COPD pathophysiology.

5. Main Hypothesis/Study Questions:

To identify the associations between protein levels and pulmonary function in ARIC European and African Americans.

The cross-sectional hypotheses is:

1. A number of proteins are associated with pulmonary function (FVC, FEV1 and FEV1/FVC).

The longitudinal hypotheses are:

1. Proteins are associated with longitudinal decline in pulmonary function beyond the traditional risk factors.
2. Proteins identified above are associated with incident COPD, incident cardiovascular diseases and mortality beyond the traditional risk factors. This will be a secondary analysis if positive findings are identified for proteomic association with pulmonary function.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study design: We will relate proteomes with pulmonary function (FVC, FEV1 and FEV1/FVC) at visit 2, and pulmonary function decline measures from visit 2 to visit 5. We will start the analysis by relating proteome with pulmonary function measured at visit 2. We will repeat our findings using visit 5 data.

Exclusion criteria: Missing proteome and/or pulmonary function data at visit 2, 3 or 5

Outcome:

- Pulmonary function (FVC, FEV1, FEV1/FVC) and COPD information at visit 2 and 5.
- The change of pulmonary function (FVC, FEV1, FEV1/FVC) between visit 2 and 5.

- Secondary outcome: incident COPD defined using pulmonary function at visit 2 and 5. Incident cardiovascular events including coronary heart disease, stroke, and heart failure identified, as well as mortality, through ARIC follow up from visit 2 until 2019.

Covariates:

- Age (years)
- Age^2($years^2$)
- Sex (male, female)
- Race
- Height (cm)
- $Height^2$ (cm^2)
- Smoking status (Current versus not current; Past versus all others)
- Smoking pack-years
- Center (use dummy variables)
- Estimated glomerular filtration rate (eGFR, mls/min/1.73m2)
- Weight (kg; include in FVC model only)

Statistical Methods:

We will primarily focus on the 4877 proteins that pass quality control. First, we will relate each protein value to pulmonary measures at visit 2 (n = 11000 estimated) and visit 5 (n = 3781) respectively using linear regression. In a previous paper where we used weighting methods to account for attrition at visit 5 for PFTs and we found that overall there were not major differences compared to findings from unweighted analyses.[8] If visit 2 protein data is not ready to use when the analyses start, we propose to relate visit 2 pulmonary measures to visit 3 proteins (i.e. those stable over time [9]) under the assumption that visit 3 proteins would be a good reflection of visit 2. The analysis will account for covariates listed above. We will secondly relate visit 2 to visit 5 change in pulmonary measures (n=3549 estimated) with visit 3 proteins accounting for the same covariates. Lastly, for either proteins identified to be associated with pulmonary function cross-sectionally or its decline, we will explore their associations with prevalent and incident COPD. The Benjamini–Hochberg procedure (BH step-up procedure) will be used to control false discovery rate (FDR) at level of 0.05 to correct for multiple testing. Sex- and race- stratified analysis will be conducted to evaluate potential effect modification.

In addition, we will explore variable selection methods, such as the elastic net regularized regression method, which uses cross-validation to select model predictors and estimate regression parameters,[10] to identify a parsimonious set of proteins that relate to pulmonary function measures or COPD. To avoid overfitting, we will apply the one standard error rule for model selection. For proteins that are associated with the outcomes, we will identify protein quantitative trait loci (pQTLs) via an ARIC ongoing project (MS# 3324, Yu) and/or public database and examine if these pQTLs are associated with pulmonary function traits. We will also evaluate the degree to which circulating protein levels mediate the effect of SNPs on pulmonary traits using two approaches. 1) Examine whether these are protein products of identified GWAS signals for pulmonary function and perform analyses adjusted for the associated SNPs; and 2) adjust proteomic analyses for genetic risk scores (GRS) generated using public GWAS summary and ARIC TOPMed imputed genotypes for each pulmonary trait. Finally, we will consider the use of

Mendelian Randomization (MR) to assess causality of protein associations, when the key assumptions of MR hold.

For the identified proteins, we will apply Cox proportional hazard regressions to estimate their effect on incident cardiovascular events and mortality adjusting traditional clinical risk factors. Statistical significance will be defined using Bonferronni Correction or FDR-BH procedure.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes ___X_ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? __X__ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __X__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

___**X**___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes ____ No

11.b. If yes, is the proposal
___ A. primarily the result of an ancillary study (list number*__ AS2017.27__)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __**X**__ No.

References

1. Schroeder EB, Welch VL, Couper D, Nieto FJ, Liao D, Rosamond WD and Heiss G. Lung function and incident coronary heart disease: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2003;158:1171-81.
2. Sin DD, Wu L and Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest*. 2005;127:1952-9.
3. Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr. and Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest*. 2000;118:656-64.
4. Ahmadi-Abhari S, Kaptoge S, Luben RN, Wareham NJ and Khaw KT. Longitudinal association of C-reactive protein and lung function over 13 years: The EPIC-Norfolk study. *Am J Epidemiol*. 2014;179:48-56.
5. Valvi D, Mannino DM, Mullerova H and Tal-Singer R. Fibrinogen, chronic obstructive pulmonary disease (COPD) and outcomes in two United States cohorts. *Int J Chron Obstruct Pulmon Dis*. 2012;7:173-82.
6. Gharib SA, Manicone AM and Parks WC. Matrix metalloproteinases in emphysema. *Matrix Biol*. 2018;73:34-51.
7. Ahrman E, Hallgren O, Malmstrom L, Hedstrom U, Malmstrom A, Bjermer L, Zhou XH, Westergren-Thorsson G and Malmstrom J. Quantitative proteomic characterization of the lung extracellular matrix in chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *J Proteomics*. 2018;189:23-33.
8. Mirabelli MC, Preisser JS, Loehr LR, Agarwal SK, Barr RG, Couper DJ, Hankinson JL, Hyun N, Folsom AR and London SJ. Lung function decline over 25 years of follow-up among black and white adults in the ARIC study cohort. *Respir Med*. 2016;113:57-64.
9. Tin A, Yu B, Ma J, Masushita K, Daya N, Hoogeveen RC, Ballantyne CM, Couper D, Rebholz CM, Grams ME, Alonso A, Mosley T, Heiss G, Ganz P, Selvin E, Boerwinkle E and Coresh J. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *J Appl Lab Med*. 2019;4:30-39.
10. Zou H HT. Regularization and variable selecton via the elastic net. *J Royal Statist Soc: Series B (Statistical Methodology)*. 2005;67:301-320.