ARIC Manuscript Proposal #3981

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

1.a. Full Title: Associations between plasma proteomics and asthma

b. Abbreviated Title (Length 26 characters): Asthma Proteomics

2. Writing Group:

Writing group members: Yura Lee, Gordon Smilnak, Mikyeong Lee, Julie White, Stephanie J. London, and Bing Yu. 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: Yura Lee
Address: 1200 Pressler Street, Suite E405
Houston, TX 77040
Phone: 737-222-1101 Fax:
E-mail: yura.lee@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: Bing Yu
Address: 1200 Pressler Street, Suite E407
Phone: 713-500-9285 Fax:
E-mail: bing.yu@uth.tmc.edu

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

4. Rationale: **copy paste**

Asthma is a chronic inflammatory disorder of the respiratory tract that impacts an estimated 262 million people in 2019 and caused 461,000 deaths worldwide[1]. Because asthma is multifactorial and clinically heterogenous, specific pathophysiologic mechanisms remain poorly understood[2, 3]. Proteomic analyses are being used to better characterize the molecular pathogenesis of asthma; however, most studies have been restricted to children, so less is known about the adult disease profile[4-6].

We aim to better understand adult asthma pathophysiology by identifying plasma proteins that are differentially expressed in adults with asthma. We also wish to understand the effects of asthma medication use on the plasma proteome. We measured plasma proteomics using the SOMAScan V4 platform in the Agricultural Lung Health Study (ALHS), a case-control study of current asthma nested within the Agricultural Health Study cohort. The ALHS analysis included 761 cases and 1,096 non-cases. We identified a number of significant associations that we would like to replicate in ARIC, which used the same proteomics platform and has a reasonably large number of individuals reporting current asthma.

5. Main Hypothesis/Study Questions:

We hypothesize that protein levels are associated with asthma status.

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 propose a cross-sectional analysis to test protein association with asthma status at ARIC visit 3. This analysis will be considered as a replication of ALHS findings or converted to a meta-analysis, if plausible. Also note that, while ARIC has data for both European and African Americans, ALHS participants are almost exclusively European in ancestry. We will perform stratified analysis to examine protein and asthma status by each race group, as asthma in adults is a heterogeneous outcome.

Exclusion criteria: Missing proteome and/or asthma data at visit 3

Exposure: SomaLogic proteome measures at visit 3

Outcome: Asthma case is defined by: YES to items PHXA8M and PHXA8N, and NO to item PHXA8L (from Visit 3 Data Book)

Covariates:

- Age = continuous (years old at time of sample collection)
- Gender = factor (male vs. female)
- Smoking Status = factor (current vs. never, former vs. never)
- Pack-Years = continuous (packs of cigarettes per day · number of years smoked)
- BMI = continuous (kg/m2)
- Study site = factor (4 sites)
- Time to Freezer = continuous (hours elapsed between sample collection and storage)
    - We collected blood on home visits to farms and mailed these back to a central lab. This created a delay that does not exist in ARIC, so we would not expect this variable to be needed
- Season = factor (fall vs. any other season)

    - Most farmers did not want their home visits in the fall harvest season, creating potential associations with both analytes and outcomes. We do not imagine that this variable would be important in ARIC
- Asthma Control Questionnaire (ACQ) Score = factor (<1.5 vs. >=1.5)
    - This covariate was present in asthma case-only analyses of medication use. An analogous variable is not available in ARIC, but we do not see this as a problem because it was not found to be an important confounder

Statistical Methods:

Model 1. Asthma status
ALHS model = SOMAmer ~ Asthma status (case vs. non-case) + Age + Gender + Smoking Status + Pack-Years + BMI + State + Hours to Freezer + Season

Suggested ARIC model = SOMAmer ~ Asthma status (case vs. non-case) + Age + Gender + Smoking Status + Pack-Years + BMI + Study Site

The below models are supplementary to Model 1 and included for completeness—they did not significantly impact ALHS results, but serve to analyze any added effect of education level and/or renal function (via cystatin C, as creatinine is not available to calculate eGFR in ALHS) on SOMAmer levels:

Model 1b. Asthma status including education level
Suggested ARIC model = SOMAmer ~ Asthma status (case vs. non-case) + Age + Gender + Smoking Status + Pack-Years + BMI + State + Study Site + Education Level

- Note that education is not included in the primary model because it is missing on about 5% of individuals in ALHS. ALHS is composed of farmers and their spouses, but education level did not seem to be a confounder for most outcomes

Model 1c. Asthma status including cystatin C (surrogate for eGFR)
Suggested ARIC model = SOMAmer ~ Asthma status (case vs. non-case) + Age + Gender + Smoking Status + Pack-Years + BMI + State + Study Site + Cystatin C

- In this model, the Cystatin C SOMAmer is reclassified from an outcome variable to a covariate

Model 1d. Asthma status including both education and cystatin C
Suggested ARIC model = SOMAmer ~ Asthma status (case vs. non-case) + Age + Gender + Smoking Status + Pack-Years + BMI + State + Study Site + Education Level + Cystatin C

Model 2.1. Use of inhaled corticosteroids over past two weeks
ALHS model = SOMAmer ~ Medication use (vs. non-use) + ACQ Score + Age + Gender + Smoking Status + Pack-Years + BMI + State + Hours to Freezer + Season

Suggested ARIC model = SOMAmer ~ Medication use (vs. non-user) + Age + Gender + Smoking Status + Pack-Years + BMI + Study Site

- Inhaled corticosteroid use can be defined as: documented use (MEDI-SCAN codes 444000, 221099, 442099024, or 442099027) over the past two weeks
- ACQ score is omitted from the proposed ARIC model due to lack of concordant variables

Model 2.2. Use of oral corticosteroids over past two weeks
ALHS model = SOMAmer ~ Medication use (vs. non-use) + ACQ Score + Age + Gender + Smoking Status + Pack-Years + BMI + State + Hours to Freezer + Season

Suggested ARIC model = SOMAmer ~ Medication use (vs. non-user) + Age + Gender + Smoking Status + Pack-Years + BMI + Study Site

- Inhaled corticosteroid use can be defined as: documented use (MEDI-SCAN code 221000) over the past two weeks
- ACQ score is omitted from the proposed ARIC model due to lack of concordant variables

Statistical significance will be determined using false discovery rate to account for the number of tests analyzed.

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)?

MS#3634 Yu et al. Proteomic Profiling and Pulmonary Function: findings from the Atherosclerosis Risk in Communities (ARIC) study

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. Diseases GBD and Injuries C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396:1204-1222.
2. Bush A. Pathophysiological Mechanisms of Asthma. *Front Pediatr*. 2019;7:68.
3. Carr TF and Bleecker E. Asthma heterogeneity and severity. *World Allergy Organ J*. 2016;9:41.
4. Pereira-Fantini PM and Tingay DG. The proteomics of lung injury in childhood: challenges and opportunities. *Clin Proteomics*. 2016;13:5.
5. Semernik O, Lebedenko A and Gunko V. Proteomic analysis of blood serum of children with bronchial asthma. *European Respiratory Journal*. 2019;54:PA4506.
6. Terracciano R, Pelaia G, Preiano M and Savino R. Asthma and COPD proteomics: current approaches and future directions. *Proteomics Clin Appl*. 2015;9:203-20.