ARIC Manuscript Proposal # 2938

PC Reviewed: 2/14/17	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

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

b. Abbreviated Title (Length 26 characters): metabolomics and PF

2. Writing Group:

Writing group members: Bing Yu, Annah Wyss, Alanna Morrison, Kari North, Eric Boerwinkle, and Stephanie J. London. Other interested ARIC investigators are invited 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 St. E641, Houston, Texas 77030
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 data 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:

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, reflects lung volume and is 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, are identified to be correlated with pulmonary function (*4, 5*). Metabolomics, which characterizes small-molecular metabolites produced by a multitude of metabolic, physiologic and cellular processes, serves as

proximal reporter of early disease processes. A few metabolomic studies demonstrate that metabolites are associated with pulmonary function and lung diseases (*6-8*). Most recently, the TwinsUK cohort has conducted a metabolomics study on multiple age-related traits, including FVC and FEV1. Among 280 analyzed metabolites, C-glycosyl tryptophan (C-glyTrp), correlated strongly with age and FEV1 ($p = 1.8 \times 10^{-8}$ adjusted for age and confounders) and was replicated in an independent population (*9*). To our knowledge, no study has investigated the metabolomic association on pulmonary function among African Americans. In addition, there are no publications on longitudinal change in lung function.

This proposed research will use both a cross-sectional and longitudinal design allowing us to better characterize the relationship between the metabolome and pulmonary function in a well-characterized population-based sample of both African Americans and European Americans from the Atherosclerosis Risk in Communities (ARIC) study.

5. Main Hypothesis/Study Questions:

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

The cross-sectional hypotheses are:

1. A number of metabolites are associated with pulmonary function (FVC, FEV1 and FEV1/FVC) at visit 1.
2. A number of metabolites are associated with prevalent COPD status at visit 1.

The longitudinal hypothesis is:

1. Metabolomic factors measured at visit 1 are associated with longitudinal decline in pulmonary function beyond the traditional risk factors. The primary analysis of decline will examine visit 1 and 2 data. Although sample size is greatly reduced by death and attribution to visit 5, we will also examine change from visit 1 to 5 as a secondary analysis.
2. Metabolomic factors identified above are associated with incident COPD beyond the traditional risk factors. This will be a secondary analysis if positive findings are identified for metabolomic association with longitudinal decline in 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 and sample:
The ARIC study participants with serum metabolomic data (1,553 European Americans and 2,479 African Americans) quantified at visit 1, pulmonary function (FVC, FEV1, FEV1/FVC) at visit 1, 2 and 5, and COPD information at visit 1 and 5.

Exclusion:

- Persons with no metabolomic data;
- For cross-sectional analysis: Persons with missing outcome variables or baseline covariates at visit 1;

- For longitudinal analyses to visit 2. Persons with missing PFT data at visit 2
- For secondary longitudinal analysis to visit 5. Persons with missing PFT data at visit 5.

Outcome:

- Pulmonary function (FVC, FEV1, FEV1/FVC) and COPD information at visit 1.
- The change of pulmonary function (FVC, FEV1, FEV1/FVC) between visit 1 and 2.
- Secondary outcome – change of pulmonary function between visits 1 and 5 and incident COPD defined using pulmonary function at visit 1, 2 and 5.

Covariates:

- Age (years)
- Age^2($years^2$)
- Sex (male, female)
- Height (cm)
- $Height^2$ (cm^2)
- Smoking status (Current versus not current; Past versus all others)
- Cigarettes per day
- Cigarettes pack years
- Center (use dummy variables)
- Batch effect (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 245 named metabolites detected in both European Americans and African Americans. For the primary analysis, analytes with >25% of values below the detection limit (BDL) and/or missingness will be excluded. Metabolite levels will be analyzed as a continuous variable; where missing/BDL values will be imputed with the lowest value to those without missing data. Metabolites will be standardized (centered at its mean and scaled by its standard deviation) prior to the analysis. We will secondarily examine the metabolites with missing/BDL values between 25%-75%, and they will be analyzed as continuous variables with the values set to the square root of the limit of detection.

For each metabolite, linear models, logistic regression, generalized estimating equations (GEE) and Cox model will be used to assess its relationship to baseline pulmonary function, prevalent COPD status, pulmonary function change over visits and incident COPD respectively. The baseline value of the lung function variable will be included in the longitudinal model. African American and European American will be analyzed separately. Statistical significance for the metabolomic data will be pre-specified at $p < 2\times10^{-4}$ using Bonferroni correction for 245 metabolites. If more than one metabolite is identified, a MetScore, which sums the quartiles of each identified metabolite, will be used to test the joint effect on pulmonary function.

Cooperative Health Research in the Region Augsburg (KORA) and we have the intention to meta-analyze our summary statistics for this proposed project.

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? ____ 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
We do not plan to use the genotype data in any primary analysis. But it has become common to incorporate genotype data and we might be asked to do this in the review process.

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 #1882 Metabolomics and kidney function (Yu)

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
__X_ A. primarily the result of an ancillary study (list number* 2008.16 and 2014.20**)**
___ 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.
Agreed.

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. E. B. Schroeder *et al.*, Lung function and incident coronary heart disease: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* **158**, 1171-1181 (2003).
2. D. D. Sin, L. Wu, S. F. Man, The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest* **127**, 1952-1959 (2005).
3. H. J. Schunemann, J. Dorn, B. J. Grant, W. Winkelstein, Jr., M. Trevisan, Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* **118**, 656-664 (2000).
4. S. Ahmadi-Abhari, S. Kaptoge, R. N. Luben, N. J. Wareham, K. T. Khaw, Longitudinal association of C-reactive protein and lung function over 13 years: The EPIC-Norfolk study. *Am J Epidemiol* **179**, 48-56 (2014).
5. D. Valvi, D. M. Mannino, H. Mullerova, R. Tal-Singer, Fibrinogen, chronic obstructive pulmonary disease (COPD) and outcomes in two United States cohorts. *Int J Chron Obstruct Pulmon Dis* **7**, 173-182 (2012).
6. D. J. Adamko *et al.*, Metabolomic profiling of asthma and chronic obstructive pulmonary disease: A pilot study differentiating diseases. *J Allergy Clin Immunol* **136**, 571-580 e573 (2015).
7. S. Deja *et al.*, Metabolomics provide new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease. *J Pharm Biomed Anal* **100**, 369-380 (2014).
8. J. L. McClay *et al.*, (1)H nuclear magnetic resonance metabolomics analysis identifies novel urinary biomarkers for lung function. *J Proteome Res* **9**, 3083-3090 (2010).
9. C. Menni *et al.*, Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* **42**, 1111-1119 (2013).