ARIC Manuscript Proposal #2867

PC Reviewed: 10/11/16 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Metabolomic predictors of incident coronary heart disease: findings from the Atherosclerosis Risk in Communities (ARIC) study

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

2. Writing Group:

Writing group members: Bing Yu, Aaron Folsom, Christie Ballantyne, Alanna Morrison and Eric Boerwinkle.

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: Eric Boerwinkle
Address: 1200 Pressler St. W114A, Houston, Texas 77030
Phone: 713-500-9058 Fax: 713-500-9020
E-mail: Eric.Boerwinkle@uth.tmc.edu

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 six months from approval of the analysis plan.

4. Rationale:

Coronary heart disease (CHD) is the leading cause of death globally and in the United States, killing over 375,000 Americans a year (*1*). Over the last decades, a number of biomarkers have been discovered that are associated with CHD (*2-4*). However, the molecular etiology of CHD remains poorly understood. Metabolomics, which characterizes small-molecular metabolites produced by a multitude of metabolic, physiologic and cellular processes, serves as proximal reporter of early disease processes. In addition, the human metabolome is the ultimate product of gene by environment interaction which forms a bridge between the genome and human diseases, creating valuable opportunity to investigate the molecular etiology for CHD. Studies have shown that new metabolite signatures can be identified to enhance CHD risk prediction models (*5-7*).

Our preliminary results in African-Americans show that histidine is associated with incident CHD (*8*). With the availability of metabolomics data in ARIC European and African Americans, we propose to explore the longitudinal association of serum metabolome quantified by untargeted GS/MS/MS with incident CHD.

5. Main Hypothesis/Study Questions:

To identify the associations between metabolite levels and incident CHD in ARIC European and African Americans.

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 baseline (visit 1) and follow-up for incident CHD events.

Exclusion:
- Persons with no metabolomic data;
- Persons with missing outcome variables or baseline covariates;
- For incident CHD analyses, exclusion of prevalent cases

Outcome:
- Incident CHD by December 31, 2013

Likely covariates at baseline exam:
- Age (yrs)
- Gender
- Race
- Body mass index (BMI, kg/m^2)
- Systolic blood pressure (SBP) and use of hypertensive medications (yes/no)
- The presence or absence of diabetes
- Lipids levels (LDL-C, HDL-C, and triglycerides)
- Current smoking status or pack years
- Estimated Glomerular Filtration Rate ($eGFR_{CKD\text{-}EPI}$, $mL/min/1.73m^2$)

Statistical Methods:
We will primarily focus on 245 named metabolites detected in both European Americans and African Americans with missing values or values below the detection limit (BDL) ≤ 25%. 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 data between 25%-75%, and they will be analyzed as ordinal variables.

We will use Cox proportional hazards model to assess the longitudinal association between metabolite levels and incident CHD. We will adjust for traditional risk factors (including age, sex, race, BMI, prevalent diabetes, SBP, use of hypertensive medications, LDL, HDL, triglycerides, current smoking status, eGFR) and batch effect. In the secondary analyses, we will examine potential effect modification using statistical tests for interaction and by stratifying by gender and race. Statistical significance for the metabolomic data will be pre-specified at $p < 2\times10^{-4}$ using Bonferroni correction (0.05/245 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 CHD prediction of these metabolites We assume 1) a sample size at 4032, 2) among which 15.5% individuals developed incident CHD during follow up, 3) the variance of the metabolite is 1 (metabolite levels will be standardized), 4) the square of correlation between the covariate of interest is 0.3 (corresponds to r = ~0.6), 5) the alpha level at 0.0002 (Bonferroni correction: 0.05/245 named compounds), and then 80% power will be achieved if the hazard ratio sits outside the range of (0.80, 1.25). We identified a hazard ratio at 0.67 for histidine and incident CHD association in our previous work (*8*).

We will also investigate the ability of the model to predict risk using Harrell's c statistic (*9*), and calculate the net reclassification index (NRI) and the integrated discrimination index (IDI) (*10*) based on 10-year risk of CHD for four categories: 0-5% (low risk), 5-10% (low-intermediate risk), 10-20% (intermediate-high risk) and >20% (high risk) (*11*).

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? ____ Yes __X__ 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"? ____ 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 #1847 Metabolomics and heart failure (Zheng)

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. D. Mozaffarian *et al.*, Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation* **131**, e29-322 (2015).
2. R. S. Vasan, Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* **113**, 2335-2362 (2006).
3. L. E. Chambless *et al.*, Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *J Clin Epidemiol* **56**, 880-890 (2003).
4. A. R. Folsom, N. Aleksic, D. Catellier, H. S. Juneja, K. K. Wu, C-reactive protein and incident coronary heart disease in the Atherosclerosis Risk In Communities (ARIC) study. *Am Heart J* **144**, 233-238 (2002).
5. E. P. Rhee, R. E. Gerszten, Metabolomics and cardiovascular biomarker discovery. *Clinical chemistry* **58**, 139-147 (2012).

6. Z. Wang *et al.*, Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57-63 (2011).
7. A. Ganna *et al.*, Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS genetics* **10**, e1004801 (2014).
8. B. Yu *et al.*, Association of Rare Loss-Of-Function Alleles in HAL, Serum Histidine: Levels and Incident Coronary Heart Disease. *Circ Cardiovasc Genet* **8**, 351-355 (2015).
9. F. Harrell *et al.*, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15**, 361-387 (1996).
10. M. Pencina *et al.*, Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* **30**, 11-21 (2011).
11. V. Nambi *et al.*, Carotid intima-media thickness and presence or absence of plaque improves prediction of coronary heart disease risk: the ARIC (Atherosclerosis Risk In Communities) study. *J Am Coll Cardiol* **55**, 1600-1607 (2010).