ARIC Manuscript Proposal # 3087

PC Reviewed: 12/12/17   Status: _____   Priority: 2
SC Reviewed: _________   Status: _____   Priority: _____

1.a. Full Title:
Associations of the Human Metabolome and Blood Pressure in the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters):
Metabolomics and Blood Pressure

2. Writing Group:
Writing group members:
Zhe Wang, Alanna Morrison, Myriam Fornage, Eric Boerwinkle, and Bing Yu

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

First author: Zhe Wang
Address: Human Genetics Center
Division of Epidemiology, Human Genetics and Environmental Science
University of Texas Health Science Center
1200 Pressler Street, Suite E-435
Houston, TX 77030
Phone: 713-500-9827    Fax: 713-500-0900
E-mail: zhe.wang@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, PhD
Dean and M. David Low Chair in Public Health
Kozmetsky Family Chair in Human Genetics
Professor, Department of Epidemiology and Human Genetics
Associate Director, Human Genome Sequencing Center at BCM
Address: 1200 Pressler Street, Suite W114A
Houston, TX 77030
Phone: 713-500-9058    Fax: 713-500-9020
E-mail: Eric.Boerwinkle@uth.tmc.edu

3. Timeline:
The metabolomics data and blood pressure data are already available. No new data collection is proposed. Once this proposal is approved, analysis will start immediately. The manuscript is to be prepared as soon as analysis is finished (~8-12 months).
4. **Rationale:**

High blood pressure (BP) (i.e. hypertension) is an important worldwide public-health challenge[1] as it is a leading risk factor for cardiovascular diseases [2] and overall mortality [3, 4]. African-Americans (AAs) have higher prevalence of hypertension than European-Americans (EAs) in the U.S., which may be in part due to a greater frequency of obesity [5]. The entire ensemble of small-molecule metabolites presented in a biologic sample is commonly referred to as the human metabolome. These small-molecule metabolites may reveal pathologic or etiologic pathways between as yet unknown genetic and environmental exposures and complex diseases [6]. Animal studies using hypertensive rats have pointed to metabolites, such as succinate and free fatty acids and their role in BP regulation [7, 8]. Recent advances in metabolomics enables large-scale human studies, which showed promise in identifying metabolites that are associated with BP levels or incident hypertension [9, 10]. To date, no study has compared and contrasted the metabolomics association with BP in European and African Americans and examined its association with BP changes over time. Linking human metabolomic data to blood pressure levels and its longitudinal changes may identify novel biomarkers to advance our knowledge of the underlying physiopathology of hypertension.

Therefore, we propose to evaluate associations of serum metabolite levels with cross-sectional and longitudinal BP levels in 1553 EAs and 2479 AAs of the Atherosclerosis Risk in Communities (ARIC) study

5. **Main Hypothesis/Study Questions:**

   1. Metabolite levels are cross-sectionally associated with baseline systolic/diastolic blood pressure (SBP and DBP) level independent of traditional hypertension risk factors, for instance BMI and smoking.
   2. Metabolite levels are associated with longitudinally change of SBP and DBP levels over ten years (visit 1-4) independent of traditional hypertension risk factors, for instance BMI and smoking.

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:**
   1. This is a study including both cross-sectional and longitudinal design using both EAs and AAs of the ARIC study.
   2. Serum metabolome were measured at 2010 and 2014 using baseline ARIC examination blood samples. BP measurements are available for visit 1-4.

   **Exclusion:**
   1. We will exclude individuals without metabolomics data or with missing values in BP and covariates.

   **Variables:**
   
   *Outcome variables:*
1) BP (mmHg, including SBP and DBP) at ARIC visit 1-4. BP will be analyzed as
continuous variables and a dichotomous variable, hypertension status yes/no (defined as
SBP≥140mmHg or DBP≥90mmHg or taking antihypertensive medication during past 2
weeks) at ARIC visit 1. Natural log transformation will be applied to correct the
skewness of BP distribution when analyzing continuous BP measurements.

*Exposure variables:* named metabolites primarily.
Metabolites will be excluded if:
1) Only detected in one batch;
2) More than 40% of the samples have missing values or values below the detection limit
(BDL) within each batch;
3) The Pearson correlation coefficient (r2) between 2010 and 2014 measurements on the
same stored sample is less than 0.30.
4) Poor medium-term variability: intraclass correlation coefficients (ICC) <0.4 [11, 12].
Metabolite levels will be analyzed as continuous variables, where missing/BDL values will
be imputed using random forest imputation based on the remaining observed measurements
[13-15]. After applying such exclusion criteria, 196 named metabolites will be
standardized to have mean 0 and standard deviation 1.

*Considering Covariates:*
1) Study center from baseline examination.
2) Risk factors based on Framingham Hypertension risk score [16]: age and sex from
baseline examination, body mass index (BMI) and cigarette smoking (current smoker,
yes/no) from visit 1-4.
3) Estimate glomerular filtration rate (eGFR) from visit 1,2 and 4 (visit 3 not available).
4) Taking any hypertension medication (yes/no) from visit 1-4.

Data analysis:
Statistical analysis will be conducted within each race group respectively. Basic models will
adjust for minimal demographic characteristics (age and sex). To test the association
between metabolites and longitudinal BP in dependent of traditional risk BP risk factors, we
will conduct fully adjusted models adjusting for age, sex, study center, BMI, smoking
status, hypertension medication and eGFR. Age, sex, BMI, smoking status and hypertension
medication can be possible effect modifiers in the relation between metabolites and BP.
Therefore, the interactions between metabolites and each of the five aforementioned
covariates will be evaluated. Using a backward selection procedure, only the statistically
relevant (p< 0.01, considering multiple comparisons) ones will be kept in the models. All
statistical analysis will be conducted using R (Version 3.4.2).

**Hypothesis 1:** For baseline SBP and DBP, linear regression will be conducted for each
metabolite levels. Logistic regression model will be used for estimating the association
between prevalent hypertension and each metabolite levels at baseline.

**Hypothesis 2:** To explore the association between baseline metabolite levels and the
longitudinal changes in SBP, DBP and hypertension status (visit 1-4), we will use
generalized estimation equations (GEE) to fit repeated measure linear (SBP and DBP) and
logistic (hypertension status) regressions. In these models, the dependent variables (SBP, DBP and hypertension status), as well as time dependent covariates such as smoking, BMI, hypertension medication and eGFR (visit 1, 2 and 4) will be assessed at baseline and follow-up 2-4 visits. Baseline age, sex and study center will be used in the models. After testing a few different structures in preliminary data, an autoregressive (AR-1) structure will be specified for the working correlation structure. R package “geepack” will be used [17] for longitudinal analysis.

To exclude the possible effect of hypertension medication, we will also conduct the same analyses in individuals without any hypertensive medication as a sensitivity analysis.

A Bonferroni corrected p-value will be used as the significant threshold to correct for multiple comparisons. A significance level of $2.55 \times 10^{-4}$ ($0.05/(196$ metabolites)) will be considered for testing the metabolome and BP associations.

Follow-up analysis: for any statistically significant metabolites we identified, we will review literatures for pathways these metabolites may be involved and how these related to BP etc. For example, when a prior path information is available and one or more of the identified metabolites are involved, I will investigate whether there is an enrichment of findings in one or more pathways. For metabolites that we identified though models involving interaction terms, we will further stratify by the potential effect modifiers to investigate the association within each strata.

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
There is no overlap between this proposal and current active proposals/published manuscripts. There are no metabolomics manuscript proposals examining the longitudinal change in BP levels.

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)?

MP1918 Zheng Y, Yu B et al. Associations of the Human Metabolome with Blood Pressure, Prevalent and Incident Hypertension among African Americans in the Atherosclerosis Risk in Communities (ARIC) study

MP1918 investigates the metabolomics of incident hypertension. Our proposal’s primary focus is the metabolomics of BP and BP changes over time. We have coordinated this proposal with Dr 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* __ AS#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.cscce.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.
   Agree.

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.cscce.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.
   Yes, the lead author is aware of the policy.

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.
Reference: