ARIC Manuscript Proposal # 1882

PC Reviewed: 12/13/11 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title:
A longitudinal Study of Metabolomics and Kidney Function among African Americans in the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): Metabolome & Kidney Function

2. Writing Group:
Writing group members: Bing Yu, Yan Zheng, Jennifer A. Nettleton, Danny Alexander, Josef Coresh, 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: Division of Epidemiology
University of Texas School of Public Health
1200 Hermann Pressler
Houston, TX, 77225
Phone: 713-500-9892 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: Human Genetics Center
University of Texas School of Public Health
1200 Hermann Pressler
Houston, TX, 77225
Phone: 713-500-9816 Fax: 713-500-0900
E-mail: Eric.Boerwinkle@uth.tmc.edu

3. Timeline:
Analysis is to start as soon as approval is obtained. Manuscript is to be prepared as soon as analysis is available. We expect that the manuscript will be prepared within nine months from approval of the analysis plan.

4. Rationale:
Kidney function is commonly characterized by estimated Glomerular Filtration Rate (eGFR). Declining eGFR over time is an early harbinger of future chronic kidney disease

(CKD). CKD is defined as presence of kidney damage or reduced function for at least 3 months. ARIC has data on albuminuria for kidney damage and eGFR for kidney function, as well as hospitalization codes which can help when participants are sicker and more likely to miss visits. [A comparison of different definitions of CKD in ARIC have been described previously (1).] Kidney failure is treated with dialysis or transplantation. Kidney function is influenced by both genetic and environmental factors (2). The human metabolome is a reflection of the interaction between genes and the environment, and studies relating kidney function with metabolomic profiles may enhance our understanding of the physiology underlying development and progression of CKD (3). Establishing a clear temporal relationship among these factors is challenging; the metabolome may influence kidney function and kidney function may, in turn, influence the metabolome. In fact, it is likely that both scenarios are true.

Multiple metabolomic derangements can negatively affect kidney function. For example vasoactive molecules such as nitric oxide have important activities in the kidney. We recognize that not all metabolites can be detected in a broad metabolomics screen but this study will provide a first rigorous assessment of these associations.

The kidney filters all metabolites from the blood. Molecules of <10,000 Da molecular weight are freely filtered by the kidney and subsequently metabolized (reabsorbed, catabolized and/or secreted). Therefore an important first step will be to understand the cross-sectional relationship of estimated kidney function to the measured metabolome. For example, uric acid is elevated in CKD and we have studied the role of genes in influencing kidney transport of urate (4, 5). Uric acid was measured in multiple ARIC visits (initially visits 1 and 2 and recently visit 4 and 5) providing a useful positive control.

This proposed research will use both a cross-sectional and longitudinal design allowing us to better characterize the temporal relationship between the metabolome and kidney function among African Americans in ARIC.

5. Main Hypothesis/Study Questions:

The cross-sectional hypothesis is:

1. Estimated kidney function will be associated with a number of metabolites detectable in the current metabolomic profile.

The longitudinal hypotheses are:

1. Metabolomic factors can be identified that are associated with kidney function (as estimated by eGFR at visit 1) beyond traditional CKD risk factors in African Americans free of prevalent CKD.
2. Metabolomic factors can be identified that are associated with longitudinal decline in kidney function (characterized by repeated measures of eGFR) beyond the traditional CKD risk factors.
3. Metabolomic factors can be identified to predict that are prospectively associated with incident CKD (defined as eGFR < 60ml/minute/1.73m^2 and ≥ 25% drop in

follow-up visit) or hospitalization with a kidney related code (including dialysis or transplantation).

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

This is a longitudinal study that consists of African American ARIC participants at the baseline (visit 1), visit 2 and visit 4 examinations, where appropriate additional follow-up will be incorporated. Hospitalization codes including CKD ICD codes allow for detection of dialysis or severe CKD after visit 4. Participants in the ARIC CARMRI visit have serum creatinine measured as well.

Exclusion:

- Non-African Americans
- Prevalent CKD (defined as eGFR < 60ml/minute/1.73m^2 at visit 1) will be excluded from longitudinal analyses
- Missing outcome or covariates information

Outcome:

- Baseline eGFR
- eGFR at each follow-up examination based on serum creatinine. At visit 4 additional filtration markers (Cystatin C, Beta trace protein, Beta-2 microglobulin) were measured and can be used in secondary analyses.
- Incident CKD (defined as eGFR < 60ml/minute/1.73m^2 and ≥ 25% drop in a follow-up visit)

eGFR will be calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation: $eGFR_{CKD\text{-}EPI} = 141\times$ (minimum of Standard Serum Creatinine [mg/dL]/κ or 1)$^{\alpha}$ × (maximum of Standard Serum Creatinine [mg/dL]/κ or 1)$^{-1.209}$ × 0.993^{age} × (1.018 if female) × (1.159 if black), where κ is 0.7 if female and 0.9 if male; α is -0.329 if female and -0.411 if male (6).

Covariates:

Susceptibility factors

- Age
- Sex
- Body mass index (kg/m^2)
- Education level

Initiation factors

- Current smoking status
- Prevalent hypertension status
- Prevalent diabetes status
- Serum glucose (by Clinical Chemistry Laboratory with calibration) (mmol/L)
- Prevalent coronary heart disease status

- Lipids: Serum HDL cholesterol (mmol/L) is associated with CKD incidence. We will also examine other lipid measurements and adjust for lipid-lowering medication use.

Metabolomics data:

Based on both practical and theoretical considerations, we have placed each measured metabolite into three groups by reliability coefficient (RC; from either the medium-term reliability study or the blind duplicate study) and missing percentage (MS#1847 Zheng Y, et al.).

- Group 1 contains metabolites (n=187) that are reliably measured (RC $\geq$0.60) and have missing values in fewer than 50% of the sample. The metabolites are to be treated as continuous variables during data analysis with the missingness of metabolites are replaced by the lowest measured value.
- Group 2 contains metabolites (n=17) that are reliably measured (RC $\geq$0.60) but have a moderate amount of missing data (values missing in 50- 80% of the sample). For this group, we consider missing values as category 1. For the measured (non-missing) values, we consider values below the median as category 2 and values above the median as category 3. An ordinal variable is to be used during data analysis.
- Group 3 contains metabolites (n=398) that have >80% missing data or RC <0.6; this group is not included in data analysis.

Statistical Methods:

For each metabolite, linear model will be conducted to estimate the relation of baseline kidney function (cross-sectional hypothesis); and then linear models, generalized estimating equations (GEE) and Cox proportional hazard model will be performed to assess its relationship to longitudinal change of kidney function and incident CKD, respectively, in several models. Minimally adjusted models will adjust for demographics (age, sex; race is limited to African-Americans). Next we will further adjust for hypertension and diabetes measures which are the strongest risk factors for CKD. Finally we will add additional risk factors. The potential combinations of baseline risk factors will be investigated by inspecting how beta-coefficient changes for the metabolite as we add individual risk factor in the extended model for each metabolite. Statistical significance for the metabolomic data will be pre-specified with an experiment-wise α=0.05 (2 tailed) and a modified Bonferroni procedure will be used to consider the correlations among metabolites (7, 8).

References:

1. Bash LD, Coresh J, Kottgen A, Parekh RS, Fulop T, Wang Y, et al. Defining incident chronic kidney disease in the research setting: The ARIC Study. American journal of epidemiology 2009;170(4):414-24.
2. Keller BJ, Martini S, Sedor JR, Kretzler M. A systems view of genetics in chronic kidney disease. Kidney Int. 2011 Oct 19.
3. Weiss RH, Kim K. Metabolomics in the study of kidney diseases. Nat Rev Nephrol. 2011 Oct 25.
4. Dehghan A, Kottgen A, Yang Q, Hwang SJ, Kao WL, Rivadeneira F, et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet 2008;372(9654):1953-61.
5. Woodward OM, Kottgen A, Coresh J, Boerwinkle E, Guggino WB, Kottgen M. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proc Natl Acad Sci U S A 2009;106(25):10338-42.

6. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150:604-612.
7. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustments methods in clinical trials. Statistics in Medicine. 1997; 16:2529-2542.
8. Blakesley RE, Mazumdar S, Dew MA, et al. Comparisons of methods for multiple hypothesis testing in neuropsychological research. Neuropsychology. 2009 Mar; 23(2):255-64.

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 ICTDER03 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

8.c. If yes, is the author aware that the participants with RES_DNA = 'not for profit' restriction must be excluded if the data are used by a for profit group?
___ 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

Yes. There is no overlap between this proposal and current proposals/published manuscripts. This proposal builds off of the metabolomic HF proposal submitted by Zheng and Nettleton (MS #1847) both of whom are authors here.

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)?

Nettleton J, Follis J L, Alonso A, et al. "Metabolomic predictors of incident heart failure: a case-control study nested within the Atherosclerosis Risk in Communities (ARIC) Study". Poster section presented at American Heart Association(AHA) Epidemiology Council meeting in Atlanta, GA; March 2011.

MS#1847 Zheng Y, et al. Role of the Human Metabolome in Incident Heart Failure Etiology among African Americans in the Atherosclerosis Risk in Communities (ARIC) Study (under review)

MS#1853 Yu B, et al. Genome-Wide Association Study of Heart Failure-related Human Metabolite Profiles among African Americans in 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? Yes

11.b. If yes, is the proposal

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

___ **B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)*** 2008.16 "Metabolomics & Heart Failure: A Novel Approach to Biomarker Discovery"**)**

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

12. 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.

Yes, the lead author is aware that manuscript preparation is expected to be completed in 1-3 years, and if this expectation is not met, the manuscript proposal will expire.