ARIC Manuscript Proposal #2034

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

1.a. Full Title:
The Human Metabolome is Associated with Dietary Intake among African Americans in the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): Metabolomics, dietary intake

2. Writing Group:
Writing group members: Yan Zheng; Bing Yu; Danny Alexander; Lyn M. Steffen; Jennifer A. Nettleton; and Eric Boerwinkle

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

First author: Yan Zheng, MPH
Address: Division of Epidemiology, Human Genetics & Environmental Sciences
University of Texas Health Science Center
1200 Herman Pressler, suite E-405
Houston, TX 77030
Phone: (713)500-9823
E-mail: Yan.Zheng@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, Ph.D
Address: Division of Epidemiology, Human Genetics & Environmental Sciences
University of Texas Health Science Center
1200 Herman Pressler, suite E-447
Houston, TX 77030
Phone: (713)500-9816 Fax: 713-500-0900
E-mail: Eric.Boerwinkle@uth.tmc.edu

3. Timeline:
We expect that the manuscript will be prepared within six months from approval of the analysis plan.

4. Rationale:
An unhealthy diet is considered one of the factors that have contributed to a rapid increase in the incidence of cardiovascular diseases. To date, the framework in nutrition

science including the role of diet has been largely reductionist in nature (e.g., by focusing on deviations from the norm, one nutrient at a time).(1) However, emerging technologies including metabolomics now allow investigation of the complexity of links between diet, lifestyle, and genetics, since such links can be eventually reflected by levels of the metabolome.(2) Furthermore, metabolomics offers potential in the area of nutritional assessment because it measures a range of small molecules present in a biological system that can characterize the physiologic state of the human body.(3, 4) Therefore, metabolomics could play roles in identification of novel biomarkers of dietary intake and in dietary assessment. In a well-characterized, population-based sample of African Americans from the Atherosclerosis Risk in Communities (ARIC) study, we propose to explore the cross-sectional associations of multiple named metabolites quantified by untargeted GS/MS/MS with dietary intakes measured by a semi-quantitative food frequency questionnaire at baseline.

5. Main Hypothesis:
Metabolomic factors are associated with self-reported dietary intakes in African Americans at baseline independent of age, sex, body mass index, overall energy intakes per day and kidney function (as measured by estimated Glomerular Filtration Rate, eGFR).

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 nested study consists of ARIC African Americans with serum metabolomic data quantified (metabolomics profiling was completed in a random sample of African Americans from the Jackson, MS, field center) and usual dietary intake information at baseline (visit 1) measured by food-frequency questionnaire (FFQ).

Exposures:

Food subgroups, major food groups, and derived dietary patterns (see below in *Dietary assessment*)

Outcome:

The standardized levels of the measured named metabolites

Covariates:

- Age (yrs)
- Sex
- Body mass index (kg/m^2)
- Energy intakes per day (kcal/d)
- Kidney function, measured as eGFR ($mL/min/1.73m^2$, calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation(5))

Metabolomics data:

Based on both practical and theoretical considerations, we have placed each of the 357 measured named metabolites into **three** groups based on the percentage of individuals having values below the detection limit (bdl) for that metabolite.

- Group 1 contains named metabolites (n=309) with < 50% of the individuals having values bdl. The metabolites are to be treated as continuous variables during data analysis with the value bdl replaced by the lowest measured value in this sample for that metabolite.
- Group 2 contains named metabolites (n=29) that have a moderate amount of data bdl (50- 80% of the individuals). For this group, we consider values bdl 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. These categories form an ordinal variable and it is to be used during data analysis.
- Group 3 contains named metabolites (n=19) that have >80% the data bdl. For this group, we consider those values bdl as category 1 and the measured (non-missing) values as category 2. An ordinal variable will be used during data analysis.

Dietary assessment:

At baseline, usual dietary intake was assessed by a 66-item interviewer-administered semiquantitative FFQ, which was a modified version of the 61-item instrument developed by Willett et al.(6) Food and beverages from the FFQ were categorized into 29 food subgroups, which were used to derive dietary patterns via principal components analyses (PCA). The subgroups were further collapsed into 5 major food groups: meat, dairy, fruits and vegetables, refined grains, and whole grains.(7) The 29 single food subgroups, PCA-derived dietary patterns and 5 major food groups will be the exposure variables for this proposed manuscript.

Statistical Methods:

For each metabolite, linear regression will be conducted to estimate its relations with each of single food subgroups, each of PCA-derived dietary patterns, and each of major food groups, independent of above covariates at baseline.

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.(8, 9)

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.

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.
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.
MS#1882 Yu B, et al. A longitudinal Study of Metabolomics and Kidney Function 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.

References:

1. Jones DP, Park Y, Ziegler TR. Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr* 2012;32:183-202.
2. Mercuro G, Bassareo PP, Deidda M, et al. Metabolomics: a new era in cardiology? *J Cardiovasc Med (Hagerstown)* 2011;12(11):800-5.
3. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet* 2008;4(11):e1000282.
4. Suhre K, Shin SY, Petersen AK, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011;477(7362):54-60.
5. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150(9):604-12.
6. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122(1):51-65.
7. Lutsey PL, Steffen LM, Stevens J. Dietary intake and the development of the metabolic syndrome: the Atherosclerosis Risk in Communities study. *Circulation* 2008;117(6):754-61.
8. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med* 1997;16(22):2529-42.
9. Blakesley RE, Mazumdar S, Dew MA, et al. Comparisons of methods for multiple hypothesis testing in neuropsychological research. *Neuropsychology* 2009;23(2):255-64.