ARIC Manuscript Proposal #3517

PC Reviewed: 12/10/19 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: *NAT8* variant and N-acetyl amino acids in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters): *NAT8* in ARIC

2. Writing Group:

Writing group members: Shengyuan Luo, Bing Yu, Eric Boerwinkle, Adrienne Tin, Josef Coresh, Eugene Rhee, Zihe Zheng, Inga Steinbrenner, Anna Köttgen, Morgan Grams

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

First author: **Shengyuan Luo**
Address: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Phone: 443-287-1827 Fax: 410-955-0485
E-mail: Shengyuan.luo@jhu.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: Morgan Grams
Address: Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

Phone: 443-287-1827 Fax: 410-955-0485
E-mail: mgrams2@jhmi.edu

3. Timeline: The manuscript will be submitted within 6 months.

4. Rationale:

NAT8 is a liver- and kidney-specific gene that encodes for N-acetyltransferase 8, an enzyme suggested to play a role in detoxification processes through catalyzing the acetylation of cysteine conjugates to form mercapturic acids which can be excreted in bile and urine.[1] Previous genetic studies have identified rs13538, a common non-synonymous single nucleotide polymorphism (SNP) in *NAT8*, as well as other variants within or near *NAT8*, to be associated with estimated glomerular filtration rate (eGFR) or incident CKD.[2-5] Variants in the promotor region of *NAT8* were found to have protective effects against hypertension and kidney failure.[6]

Previously, we conducted a genome-wide association study of untargeted serum metabolites among African American Atherosclerosis Risk in Communities Study (ARIC) participants. We found that rs13538 was strongly associated with N-acetylornithine and N-acetyl-1-methylhistidine, and that higher circulating levels of these metabolites were associated with lower eGFR and higher incidence of CKD without adjusting for eGFR and the association with incident CKD became insignificant after adjusting for baseline eGFR.[7]

The associations between rs13538 and other N-acetyl amino acids were not evaluated because of the limited number of metabolites quantified in the previous study. Further, it is unknown whether *NAT8*-associated N-acetyl amino acids are also associated with ESKD. Given the expansion of the Metabolon platform and the availability of metabolomics data at Visit 5, we now propose to extend the study by addressing these additional research questions. We have performed the same analyses in the African American Study of Kidney Disease and Hypertension, the Modification of Diet in Renal Disease, and BioMe. We will conduct meta-analyses to combine results from ARIC with findings from these cohorts.

5. Main Hypothesis/Study Questions:

Hypothesis 1: The *NAT8* variant rs13538 is associated with N-acetyl amino acids in addition to those previously evaluated (i.e., N-acetylornithine and N-acetyl-1-methylhistidine) quantified using untargeted serum metabolomic profiling in ARIC participants.

Hypothesis 2: Specific N-acetyl amino acids associated with rs13538 are also associated with ESKD.

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:

Hypothesis 1: Association study in eligible participants at Visit 5, performed separately between blacks and whites.

Hypothesis 2: Longitudinal study of eligible participants followed from Visit 5 till the date of ESKD, death, or administrative censoring, whichever occurred first.

Inclusion/exclusion criteria:

Hypothesis 1: Participants age >18 at Visit 5, agreed to DNA research, and with both genotyping and untargeted serum metabolomic profiling at Visit 5 will be included. We will exclude participants who have eGFR <15 ml/min/m^2, prevalent ESKD, or missing age, sex, race, eGFR, urine ACR at Visit 5.

Hypothesis 2: Participants age >18, with serum metabolomic profiling at Visit 5 will be included. We will exclude participants who have eGFR <10 ml/min/m^2, prevalent ESKD, or missing age, sex, race, eGFR, urine ACR at Visit 5.

Outcomes:

Hypothesis 1: The cross-sectional associations between rs13538 and circulating levels of N-acetyl amino acids will be evaluated. Here, the metabolite is the outcome.

Hypothesis 2: End-stage kidney disease, ascertained using linkage to the US Renal Data System (USRDS) or based on the International Classification of Diseases, Ninth or Tenth Revision, Clinical Modification (ICD-9-CM/ICD-10-CM) codes for hospitalization or death.

Exposures:

Hypothesis 1: Rs13538 genotype.

Hypothesis 2: N-acetyl amino acids that are associated with rs13538.

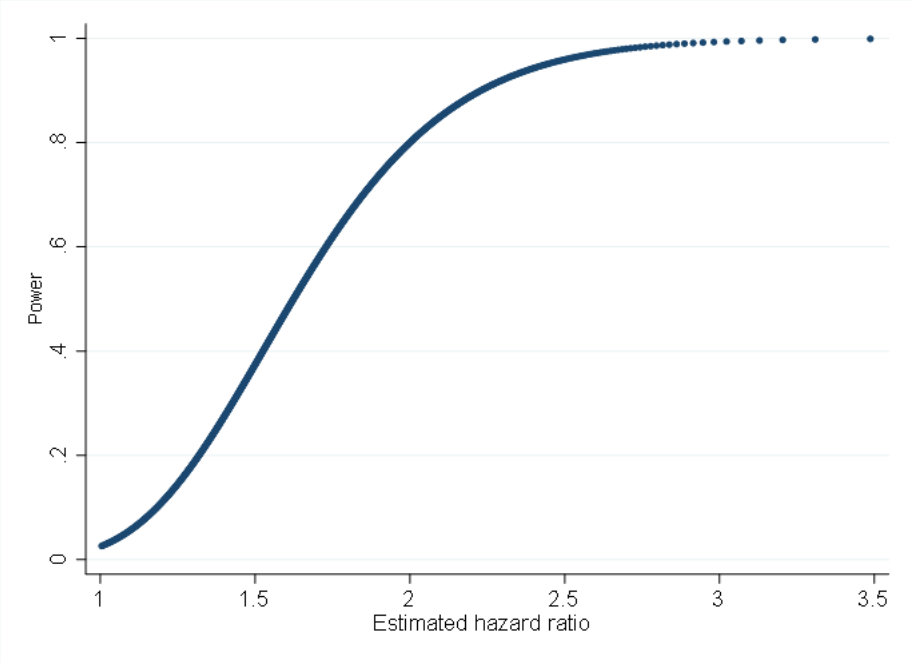
Covariables:

Covariable name	Type	Description	Time of collection
Age	Continuous	Year	Visit 5
Sex	Binary	Male/female	Visit 5
Race	Binary	Black/white	Visit 5
eGFR calculated using CKD-EPI equation	Continuous	mL/min/1.73 m^2	Visit 5
Urine albumin to creatinine ratio (ACR)	Continuous	mg albumin/gram urine creatinine	Visit 5
Smoker	Categorical	Never/Past/Current	Visit 5
Diabetes	Binary	Yes/no Fasting glucose greater than 126 mg/dL, non-fasting glucose greater than 200 mg/dL, self-reported physician diagnosis, or antidiabetic medication use	Visit 5
Hypertension	Binary	Yes/no Blood pressure of ≥140/90 mm Hg or current use of antihypertensive medication	Visit 5
Coronary artery disease	Binary	Yes/no Self-reported history of myocardial infarction or cardiac procedure	Visit 5
Body mass index	Continuous	kg/m^2	Visit 5

Analysis plan:

Hypothesis 1: Rs13538 genotype (or dosage information if imputed) will be extracted from existing genotype data. The methods of genotyping in ARIC have been previously described.[7] Details about metabolomic profiling and assessment have also been described and published previously.[7] Metabolite values will be natural log-transformed and scaled to a median of 1 prior to analyses. For the purposes of genetic associations, we will evaluate both by excluding missing values and also by imputing with the lowest detectable level. We will only include metabolites that are N-acetyl amino acids. Rs13538 will be coded using an additive genetic model. We will perform linear regressions of rs13538 on each N-acetyl amino acid, adjusting for age, sex, eGFR, urine ACR, body mass index, history of smoking, history of diabetes, history of hypertension, history of coronary artery disease. Analyses will be stratified by race. The assumption of linearity will be checked for metabolites with significant associations by categorizing metabolite levels into quartiles and quintiles and repeating analyses.

Hypothesis 2: Cox proportional hazards regression will be used to calculate the hazard ratio for ESKD. The time metric will be time since visit 5. Participants will be considered at risk from Visit 5 until ESKD, death, or administrative censoring. All Cox models will include as covariables age, sex, eGFR, urine ACR, body mass index, history of smoking, history of diabetes, history of hypertension, history of coronary artery disease. Metabolite levels below the assay detectable limit will be imputed with the lowest detected value for that metabolite in all samples. Metabolite values will be natural log-transformed and scaled to a median of 1 prior to analyses. Analyses will be stratified by race. A power calculation was performed to estimate effect size. We used the following parameters: Sample size N=6,538 (estimated number of eligible participants at Visit 5); Probability of an ESKD event Pr=0.01; Power=0.80. Using a two-sided test, the target hazard ratio is estimated to be 2.00. We plot power by estimated effect size below:



All significance levels will be Bonferroni adjusted. Statistical analyses will be performed using R (R Foundation, Vienna, Austria) or Stata/IC 14.2 (Stata Corp., College Station, TX).

Methodologic challenges:

Hypothesis 1: Measurement precision may vary by metabolite, affecting likelihood of detecting associations for different metabolites. We will take into account metabolites' coefficients of variation in blind duplicates in interpreting results.

Hypothesis 2: The total number of ESKD events is 45, which may limit our power to detect associations. We plan to meta-analyze these same associations in the African American Study of Kidney Disease and Hypertension, the Modification of Diet in Renal Disease, and BioMe cohorts to increase sample size and power.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes __X__ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ____ Yes ____ No
(The 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

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 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/mantrack/maintain/search/dtSearch.html

___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)?

#1882: Bing Yu: A longitudinal study of metabolomics and kidney function among African Americans in ARIC

#2084: Bing Yu: DNA Sequence Variation and the Human Metabolome in African Americans from 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? ___X___ Yes ____ No

11.b. If yes, is the proposal

_X_ A. primarily the result of an ancillary study (list number* _____)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

* Bing Yu's ancillary on metabolites and heart failure at ARIC visit 5.

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.

Reference

1. Veiga-da-Cunha M, Tyteca D, Stroobant V, *et al.* Molecular identification of NAT8 as the enzyme that acetylates cysteine S-conjugates to mercapturic acids. *The Journal of biological chemistry* 2010; **285:** 18888-18898.

2. Chambers JC, Zhang W, Lord GM, *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nature genetics* 2010; **42:** 373-375.

3. Boger CA, Gorski M, Li M, *et al.* Association of eGFR-Related Loci Identified by GWAS with Incident CKD and ESRD. *PLoS genetics* 2011; **7:** e1002292.

4. Kottgen A, Pattaro C, Boger CA, *et al.* New loci associated with kidney function and chronic kidney disease. *Nature genetics* 2010; **42:** 376-384.

5. Suhre K, Shin SY, Petersen AK, *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011; **477:** 54-60.

6. Juhanson P, Kepp K, Org E, *et al.* N-acetyltransferase 8, a positional candidate for blood pressure and renal regulation: resequencing, association and in silico study. *BMC medical genetics* 2008; **9:** 25.

7. Yu B, Zheng Y, Alexander D, *et al.* Genetic determinants influencing human serum metabolome among African Americans. *PLoS genetics* 2014; **10:** e1004212.