1.a. Full Title: Metabolomic profiling of GSTM1 copy number variation (CNV) and APOL1 renal risk variants

b. Abbreviated Title (Length 26 characters): metabolites and CKD genes

2. Writing Group:
   Adrienne Tin, Morgan Grams, Bing Yu, Robert B. Scharpf, Josef Coresh, Dan Arking, Megan Grove, Eric Boerwinkle, and others welcome

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

First author: Adrienne Tin, PhD
Address: Department of Epidemiology
          Johns Hopkins Bloomberg School of Public Health
          615 N. Wolfe Street, W6021
          Baltimore, MD 21205
          Phone: 201-281-9577
          atin1@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, MD, PhD
   Address: 2024 East Monument St
            Room 2-638
            Baltimore, MD 21287

3. Timeline:
   Data analysis will start immediately. A manuscript is expected to be prepared within 6 months.

4. Rationale:
   Glutathione S- transferase mu 1 (GSTM1) copy number variations and the G1 and G2 renal risk variants at the apolipoprotein L1 gene (APOL1) are known common variants with significant association with chronic kidney disease (CKD) progression in African
Americans. (1, 2) GSTM1 catalyzes the conjugation of glutathione with a range of electrophiles. The details of the serum metabolites in its pathway have not been characterized. (3) Regarding the APOL1 renal risk variants, the biological pathway underlying this association with CKD progression is still under investigation. (4, 5) The APOL1 risk variants were not included in the HapMap reference panel, (6) and the GSTM1 deletion is not tagged by single nucleotide polymorphisms (SNP). Therefore the risk alleles in these two genes were not included in the previous study on genetic association of serum metabolites in the ARIC African American cohort. (7) An unbiased scan of the association of GSTM1 copy number and the APOL1 renal risk variants with serum metabolites can generate insight on the function of these two genes and the biological pathway underlying their associations with CKD progression.

The association between GSTM1 copy number and CKD progression has been shown to follow an additive genetic model. (1) The association between the APOL1 risk variant and CKD progression has been shown to follow a recessive genetic model: 2 copies of G1 or G2 alleles. However, having one copy of the risk variant has been associated with earlier age onset of dialysis. (8) In addition, the recessive genetic model for CKD progression may not hold for metabolite association for the G1 and G2 variants. For metabolite association, we will perform analyses using additive and non-additive genetic model to comprehensively evaluate the association between the APOL1 risk variants and serum metabolites levels.

5. Main Hypothesis/Study Questions:

The aim of this study is to identify metabolites with significant association with GSTM1 copy numbers or the APOL1 G1 and G2 risk variants.

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: cohort study

Inclusion criteria: Participants with data in predictors (GSTM1 copy number and APOL1 risk variants), serum metabolites, and covariates.

Outcomes: Metabolites were measured from stored fasting serum samples by Metabolon, Inc. (Durham, North Carolina) using an untargeted, gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry approach (GC-MS/LC-MS). This untargeted approach identified approximately 600-800 named and unnamed metabolites. In the present study, we will primarily focus on the ~200 metabolites with limited missing values, reasonable reliability and presented in both batches.

Predictor:
1) GSTM1 copy numbers
2) Number of \textit{APOL1} G1 and G2 risk alleles

\textbf{Genetic Model:}
Primary analysis: additive genetic model because it is more powerful for unbiased scan.

Secondary analysis: use the number of risk alleles as categorical variables to determine whether the additive genetic model is the best fit. Based on the estimate here, we will decide whether to run the dominant or recessive model.

\textbf{Other variable of interest at visit 1}: age, gender, diabetes, hypertension, eGFRcr at visit 1

\textbf{Measure of metabolites and determination of} \textit{GSTM1} \textbf{copy number}

For the determination of \textit{GSTM1} copy numbers, we will use the same methods developed in Ancillary Proposal 2015.27. Briefly, we will process the exome sequencing reads of chromosome 1 where \textit{GSTM1} is located. We will first apply quality control to remove exons with low coverage and mappability and at the extreme of GC content using the CODEX package. Then the coverage at each exon will be normalized using the median coverage of chromosome 1. The number of copies of \textit{GSTM1} will be determined by detecting break points in the distribution of the normalized coverage.

Serum metabolites of ~2000 African Americans were measured using stored samples from visit 1 by Metabolon, Inc. (Durham, NC) using an untargeted gas chromatography–mass spectrometry and liquid chromatography–mass spectrometry based protocol. Metabolites will be excluded if (1) more than 80% of the samples had values below the detection limit and (2) the reliability coefficient on the basis of a comparison of two samples from 60 individuals collected 4–6 weeks apart was less than 0.60.

\textbf{Association analysis}

For each predictor, we will perform analysis using two models

1) Model 1: Adjusted for age, sex, and genetic principal components
2) Model 2: add eGFR, diabetes and hypertension medication use

Serum metabolites will be log transformed.
Significance threshold: 0.05/number of known metabolites
Power analysis

Assuming the number of known metabolites tested is 300 (7) resulting in a significant threshold of $1.7 \times 10^{-4}$ and the number of participants with available data is 1500, the following table shows the minimum detectable change in metabolite levels per copy of risk allele with 80% power.

<table>
<thead>
<tr>
<th>Variant</th>
<th>MAF</th>
<th>Minimum detectable SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSTM1 copy number</td>
<td>0.48</td>
<td>0.17 (rsqr=0.014)</td>
</tr>
<tr>
<td>APOL1 risk alleles</td>
<td>0.13</td>
<td>0.25 (rsqr=0.014)</td>
</tr>
</tbody>
</table>

7.a. Will the data be used for non-CVD analysis in this manuscript? __X__ Yes ____ 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? __X__ 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? __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 file ICTDER03 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.csc.unc.edu/ARIC/search.php](http://www.csc.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)?

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
A. primarily the result of an ancillary study (list number: 2015.27)
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.

References