February 6th, 2017

Dear ARIC Publications Committee Members,

Attached please find a copy of our revised manuscript proposal #2788, “GSTM1 Copy Number Variation (CNV) and Cardiovascular Disease Risk”.

This manuscript proposal was previously approved for the study of incident CHD, stroke, and heart failure. We would like to add incident hypertension because of emerging evidence. The revisions are marked with track changes.

Sincerely,

Adrienne Tin

Adrienne Tin, PhD, MS
Assistant Scientist
Department of Epidemiology, W6017
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD 21205
Phone: 443-287-4740
1.a. Full Title: **GSTMI Copy Number Variation (CNV) and Cardiovascular Disease Risk**

b. Abbreviated Title (Length 26 characters): GSTM1 and CVD

2. Writing Group:
   Adrienne Tin, Morgan Grams, 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 map

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

4. Rationale:

Glutathione S-transferase mu 1 (**GSTMI**) catalyzes the conjugation of glutathione with a range of electrophiles. **GSTMI** null genotype is common across populations with a frequency of ~50% in individuals of European ancestry and ~27% in individuals of
African ancestry. The association between having 0 copies of \textit{GSTM1} and the risk of coronary heart disease or coronary arterial disease has been inconsistent. Prior studies on \textit{GSTM1} and cardiovascular disease risk were mostly case-control studies with relatively small sample size. In the ARIC study, Li et al. used a case-cohort design and showed a significant association between the \textit{GSTM1} null genotype and coronary heart disease among ever-smokers. The \textit{GSTM1} null genotype is not tagged by single nucleotide polymorphisms (SNPs) in genome-wide arrays or reference panels. Therefore, large scale studies on the association of the \textit{GSTM1} null genotype with CVD have been limited.

With the availability of large scale exome sequencing, \textit{GSTM1} copy number can be determined using sequencing reads as described in the ARIC ancillary study #2015. Taking advantage of the rich phenotype and genetic data in the ARIC study, we will investigate the association of \textit{GSTM1} copy number with incident coronary heart disease, stroke, and heart failure.

In addition, in a mouse model of renal vascular injury, \textit{Gstm1} deletion has been shown to promote vascular smooth muscle cell migration and proliferation in the renal vasculature. This suggests a potential association between \textit{GSTM1} deletion and increased blood pressure, a risk factor of cardiovascular disease (CVD), and provides a mechanistic link between \textit{GSTM1} deletion and CVD.

5. Main Hypothesis/Study Questions:

Having 0 copy of \textit{GSTM1} will be associated with higher risk for CVD compared with those with 1 or 2 copies. 

Hypothesis 2: Having 0 copy of \textit{GSTM1} will be associated with higher risk for incident hypertension compared with those with 1 or 2 copies.

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: prospective cohort study

Inclusion criteria: Participants with exome sequencing data in Freeze 5 and with data on CVD outcomes, and relevant covariates. In the analysis of each CVD outcome, participants with prevalent disease will be excluded. For example, in the analysis of incident CHD, participants with prevalent CHD will be excluded.

Outcomes: incident events up to the end of 2013

Incident CHD includes definite or probable myocardial infarction, definite coronary death, or coronary revascularization procedure.
Incident stroke included definite or probable cases defined as sudden or rapid onset of neurologic symptoms that lasted for 24 hours or led to death in the absence of another cause.

Incident heart failure is defined as the first heart failure hospitalization or presence of heart failure code on death certificate.

**Predictor:** *GSTM1* copy numbers estimated using exome sequencing reads

Other variables of interest at visit 1: age, gender, race, diabetes, hypertension, eGFR, BMI, total cholesterol, triglycerides, prevalent coronary heart disease, heart failure, and stroke.

**Data analysis:**

For the determination of *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 *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 *GSTM1* will be determined by detecting break points in the distribution of the normalized coverage.

The association between *GSTM1* copy number and CVD outcomes will be analyzed in European and African Americans separately to avoid confounding by population. The association will be evaluated using Cox regression controlling for age, sex, and known risk factors of CVD. We test for the interaction between *GSTM1* copy number and smoking status on CVD outcomes within each ancestry groups and perform meta-analysis to estimate the association combining participants of European and African ancestry.

**Power analysis:**

We are assuming an additive genetic model, a sample size of 5000 in participants of European ancestry and 2000 in participants of African ancestry and calculate the minimal hazard ratio (HR) per copy of *GSTM1* that is detectable with an alpha of 0.05 for a 2-sided test and 80% power.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>European ancestry</th>
<th>African ancestry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~ number of cases</td>
<td>Minimal detectable HR</td>
</tr>
<tr>
<td>Incident CHD</td>
<td>809</td>
<td>1.19</td>
</tr>
<tr>
<td>Incident stroke</td>
<td>375</td>
<td>1.29</td>
</tr>
<tr>
<td>Incident heart failure</td>
<td>875</td>
<td>1.18</td>
</tr>
<tr>
<td><strong>Incident hypertension</strong></td>
<td><strong>2045</strong></td>
<td><strong>1.12</strong></td>
</tr>
</tbody>
</table>

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?   

___ 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

___ 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)?  

#1949 Validation of inter-visit kidney events  
#1929 Genome-wide DNA methylation profiling in peripheral blood: quality control and association with demographic characteristics

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: 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.csc.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


ARIC Manuscript Proposal #2788

PC Reviewed: 7/12/16  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: _____

1.a. Full Title: GSTMI Copy Number Variation (CNV) and Cardiovascular Disease Risk

b. Abbreviated Title (Length 26 characters): GSTM1 and CVD

2. Writing Group:
   Adrienne Tin, Morgan Grams, 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 map

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) catalyzes the conjugation of glutathione with a range of electrophiles. GSTM1 null genotype is common across populations with a frequency of ~50% in individuals of European ancestry and ~27% in individuals of
African ancestry. The association between having 0 copies of GSTM1 and the risk of coronary heart disease or coronary arterial disease has been inconsistent. Prior studies on GSTM1 and cardiovascular disease risk were mostly case-control studies with relatively small sample size. In the ARIC study, Li et al. used a case-cohort design and showed a significant association between the GSTM1 null genotype and coronary heart disease among ever-smokers. The GSTM1 null genotype is not tagged by single nucleotide polymorphisms (SNPs) in genome-wide arrays or reference panels. Therefore, large scale studies on the association of the GSTM1 null genotype with CVD have been limited.

With the availability of large scale exome sequencing, GSTM1 copy number can be determined using sequencing reads as described in the ARIC ancillary study #2015.27. Taking advantage of the rich phenotype and genetic data in the ARIC study, we will investigate the association of GSTM1 copy number with incident coronary heart disease, stroke, and heart failure.

5. Main Hypothesis/Study Questions:

Having 0 copy of GSTM1 will be associated with higher risk for CVD compared with those with 1 or 2 copies.

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: prospective cohort study

Inclusion criteria: Participants with exome sequencing data in Freeze 5 and with data on CVD outcomes, and relevant covariates. In the analysis of each CVD outcome, participants with prevalent disease will be excluded. For example, in the analysis of incident CHD, participants with prevalent CHD will be excluded.

Outcomes: incident events up to the end of 2013
Incident CHD includes definite or probable myocardial infarction, definite coronary death, or coronary revascularization procedure.
Incident stroke included definite or probable cases defined as sudden or rapid onset of neurologic symptoms that lasted for 24 hours or led to death in the absence of another cause.
Incident heart failure is defined as the first heart failure hospitalization or presence of heart failure code on death certificate.

Predictor: GSTM1 copy numbers estimated using exome sequencing reads
Other variables of interest at visit 1: age, gender, race, diabetes, hypertension, eGFR, BMI, total cholesterol, triglycerides, prevalent coronary heart disease, heart failure, and stroke.

Data analysis:

For the determination of 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 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 GSTM1 will be determined by detecting break points in the distribution of the normalized coverage.

The association between GSTM1 copy number and CVD outcomes will be analyzed in European and African Americans separately to avoid confounding by population. The association will be evaluated using Cox regression controlling for age, sex, and known risk factors of CVD. We test for the interaction between GSTM1 copy number and smoking status on CVD outcomes within each ancestry groups and perform meta-analysis to estimate the association combining participants of European and African ancestry.

Power analysis:
We are assuming an additive genetic model, a sample size of 5000 in participants of European ancestry and 2000 in participants of African ancestry and calculate the minimal hazard ratio (HR) per copy of GSTM1 that is detectable with an alpha of 0.05 for a 2-sided test and 80% power.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>European ancestry</th>
<th>African ancestry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~ number of cases</td>
<td>Minimal detectable HR</td>
</tr>
<tr>
<td>Incident CHD</td>
<td>809</td>
<td>1.19</td>
</tr>
<tr>
<td>Incident stroke</td>
<td>375</td>
<td>1.29</td>
</tr>
<tr>
<td>Incident heart failure</td>
<td>875</td>
<td>1.18</td>
</tr>
</tbody>
</table>

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? ___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
   _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)?

    #1949 Validation of inter-visit kidney events
    #1929 Genome-wide DNA methylation profiling in peripheral blood: quality control and
    association with demographic characteristics

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: 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.csc.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
     shows you which journals automatically upload articles to PubMed central.
References


