1.a. Full Title: Multivariate phenotype analysis of hyperglycemia biomarkers in the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): Glycemic marker USAT analysis

2. Writing Group:

Writing group members: Stephanie Loomis, Priya Duggal, Liz Selvin, Debashree Ray, Adrienne Tin, Anna Kottgen, Joe Coresh, Baolin Wu; others welcome

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

First author: Stephanie Loomis
Address: Johns Hopkins University
615 N. Wolfe St, W6021
Baltimore, MD 21287

Phone: 860 912-7169 Fax:
E-mail: sloomis5@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: Elizabeth Selvin
Address: Johns Hopkins University
2024 E. Monument Street, Suite 2-600
Baltimore, MD 21287

Phone: 410-955-0495 Fax: 410-955-0476
E-mail: eselvin@jhu.edu
3. **Timeline:**

Data are available. We anticipate a rapid timeline for this project and aim to have a first draft of the manuscript to co-authors in <6 months.

4. **Rationale:**

Type 2 diabetes is defined by hyperglycemia, or elevated blood glucose concentrations. Hyperglycemia is commonly characterized using glucose and/or HbA1c; however, there is growing interest in additional, nontraditional biomarkers of hyperglycemia such as fructosamine and glycated albumin \(^1\)–\(^9\). Fructosamine, which is glucose bound to total serum protein, and glycated albumin, which is glucose bound to the most prevalent serum protein, albumin, reflect average blood glucose over the previous 2-3 weeks\(^10\).

It is well-established that genetics play a role in type 2 diabetes and influence levels of glucose and HbA1c. In addition, our recent genome-wide association study (GWAS) has identified several genetic variants associated with fructosamine and glycated albumin (Loomis 2018 Diabetes, under review). Recent, large GWAS of fasting glucose and HbA1c (sample sizes of up to 133,000-160,000) has enabled identification of many variants associated with these measures (fasting glucose \((n=36)\) and HbA1c \((n=60))\(^{11,12}\). Thus, improving the power of recent fructosamine and glycated albumin GWAS may identify additional variants associated with these biomarkers. However, there are few large prospective cohort studies that have both the genetic and biomarker data available to conduct large GWAS of fructosamine and glycated albumin.

A way to achieve additional power which does not require increased sample size is through multivariate phenotype analysis\(^{13-15}\). This method models the association between variants and a combined phenotype of correlated variables (eg, multiple biomarkers of hyperglycemia). We will use the program Unified Score-based Association Test (USAT)\(^{16}\) to evaluate the association between common genetic variants and biomarkers of hyperglycemia in the ARIC Study to identify additional variants associated with fructosamine and glycated albumin.

5. **Main Hypothesis/Study Questions:**

In this study, we will evaluate if the increased power of a multivariate phenotype analysis method can identify additional genetic variants associated with fructosamine and glycated albumin using participants from the ARIC study.

**Hypothesis:** Common genetic variants are associated with fructosamine and glycated albumin and may be identified through multivariate phenotype analysis.

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 population:** GWAS data is available for approximately 10,000 (8,000 white, 2,000 black) ARIC participants.
Study design: Analysis using glycemic biomarker (fructosamine, glycated albumin) data collected at ARIC visit 2 (1990-1992). Blood for genetic data was collected at visit 1 (1987-1989), but as the DNA sequence does not change over time, it is acceptable to collect exposure (DNA) and outcome (markers of glycemia) variables at different study visits.

Inclusion/exclusion ARIC individuals with consent for genetics studies, and with GWAS data that has passed standard (ARIC approved) quality thresholds will be included. We will exclude individuals without valid fructosamine and glycated albumin data available and individuals with prevalent diabetes at visit 2, (defined by self-reported physician diagnosed diabetes or taking diabetes medication) when fructosamine and glycated albumin were measured.

Exposure variables: Hardcall genotypes from imputed dosages from GWAS data

GWAS data
DNA was extracted from blood collected at visit 1 from ARIC participants. Genotyping was done using the Affymetrix 6.0 array and imputed to 1000 Genomes Phase I reference panel for 37 million SNPs. Standard quality control measures were applied.

Outcomes: Fructosamine, glycated albumin

Fructosamine (Roche Diagnostics, Indianapolis IN, USA) and glycated albumin (Asashi Kasei GA-L, Tokyo, Japan) were measured in 2012-2013 using a Roche Modular P800 system from serum collected at visit 2 (1990-92) and stored at -70°C.17

Covariates: Age (years) at visit 2, sex, study center and significant (p<0.05) principal components.

Data analysis:

While there are several methods for multivariate phenotype analysis, many of them do not maintain high power in a variety of variant-phenotype associations and correlation between-phenotype conditions. Therefore, will implement the program USAT, which has high power in many settings16. USAT gives weights to two types of multivariate phenotype methods: MANOVA and score-based tests from marginal models, and uses the method with optimal power for each variant-phenotypes association test. We will run USAT on several combinations of phenotypes separately by ancestry: 1: fructosamine and glycated albumin; 2: fructosamine, glycated albumin, HbA1; 3: fructosamine, glycated albumin, HbA1c, fasting glucose.

Limitations:

While the aim of implementing multivariate phenotype analysis methods is to increase power, we still may be underpowered to detect some variants of small effect, particularly in the African ancestry subset (N=approx. 2000).
7.a. Will the data be used for non-CVD analysis in this manuscript? ____ 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? ____ 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?  ____ 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”?  ____ 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.cscce.unc.edu/ARIC/search.php

____ 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)?

#2496 USAT: A Unified Score-based Association Test for Multiple Phenotype- Genotype Analysis

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes  ____ No

11.b. If yes, is the proposal

A. primarily the result of an ancillary study (list number*2006.02)
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.cscce.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.csc.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.

13. Per Data Use Agreement Addendum for the Use of Linked ARIC CMS Data, approved manuscripts using linked ARIC CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __x__ No.

References:


