ARIC Manuscript Proposal # 3021

1.a. Full Title: Exome sequencing analysis of nontraditional markers of hyperglycemia in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters): Exome seq. of glucose markers

2.   Writing Group:

Writing group members: Stephanie Loomis, Priya Duggal, Liz Selvin, Adrienne Tin, Anna Kottgen, Joe Coresh, Eric Boerwinkle, James Pankow; 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 elevated blood glucose levels, or hyperglycemia. There are multiple ways to evaluate glucose levels: fasting glucose and hemoglobin A1c (HbA1c) are traditional glycemia biomarkers that are commonly used clinically[1], and fructosamine, glycated albumin and 1,5-AG are more recently proposed nontraditional glycemia biomarkers. [2,3] While each of these biomarkers aim to capture blood glucose levels, they vary in their molecular structure, timespan and limitations. Fasting glucose is a direct measure of serum glucose after an 8 hour fast, representing instantaneous blood glucose levels, but has high intra-individual variability and is affected by factors such as acute illness, recent physical activity and time of day.[4] HbA1c is formed as glucose binds to hemoglobin molecules within erythrocytes, and represents average blood glucose over the erythrocyte lifespan, 2-3 months.[2] Factors that impact erythrocyte turnover, such as hemolytic anemia or severe kidney disease, as well as rare hemoglobin variants, alter HbA1c levels in a manner not related to blood glucose levels.[5,6] Fructosamine is glucose bound to total serum protein; glycated albumin is glucose bound to serum albumin, and is similar to fructosamine, as the majority of serum protein is comprised of albumin. Both represent average blood glucose over the previous 2-3 weeks. Fructosamine and glycated albumin levels can be affected by changes in serum protein and serum albumin metabolism, respectively.[7] 1,5-AG is a molecule structurally similar to glucose that competes with glucose for reabsorption in the kidney at high concentrations of glucose and competes with glucose for enternal uptake among persons without diagnosed diabetes (Loomis et al, manuscript in preparation). It represents glycemic excursions over the previous 1-2 weeks.[8]

    Studying the genetics of these biomarkers is important for two main reasons: it can add to the understanding of type 2 diabetes pathophysiology, and it can help identify the limitations of the individual biomarkers in accurately reflecting blood glucose levels. We can identify genetic variants associated with multiple biomarkers, which gives strong evidence of the variants' roles in regulation of blood glucose concentrations, and we can also take advantage of the biomarkers inherent differences to uncover various aspects of diabetes biology that might be captured by one biomarker but missed by another. In addition, because none of these glycemia biomarkers are perfect indicators of hyperglycemia, their levels may be influenced by nonglycemic (non diabetes-related) factors.[9,10] Thus, we may identify significant variants in or near genes with known functions related to the nonglycemic portion of the biomarkers (eg, variants that affect serum albumin for glycated albumin), indicating limitations of the measured biomarker levels to accurately reflect glucose. These genetic limitations are important to understand for the clinical utility of these markers; if nonglycemic genetic variants strongly impact glycemic biomarker levels, this may need to be taken into account in the interpretation of these biomarkers as measures of glycemic control.

    The genetics of fasting glucose and HbA1c have been well characterized.[11] The majority of HbA1c genetic variants that have been discovered have been in or near genes involved in red blood cell turnover, highlighting HbA1c limitations in reflecting blood glucose levels.[10,9] Multiple genetic variants associated with fasting glucose are not associated with HbA1c.[12] This

could partially be due to sample size (fasting glucose GWAS are larger than HbA1c GWAS) but could also reflect diabetes-relevant genetics that are not captured by HbA1c alone.

The genetics of fructosamine, glycated albumin and 1,5-AG have been less well studied. With the exception of a GWAS and whole exome sequencing analysis of 1,5-AG as part of large biomarker panels[13,14], two in-progress manuscripts of GWAS for fructosamine/glycated albumin and 1,5-AG in ARIC constitute the literature on fructosamine, glycated albumin and 1,5-AG genetics. The fructosamine and glycated albumin GWAS has identified a likely biomarker-specific association, and a variant associated with other glycemic biomarkers. The GWAS of 1,5-AG has identified seven significant variants representing a novel, potentially diabetes-related pathway not captured by other glucose biomarkers. While these GWAS capture common genetic variants, the genetic architecture likely also includes rare coding variants, which GWAS do not capture. We will examine exome sequencing data to study rare coding variants that will add to the picture of fructosamine, glycated albumin and 1,5-AG genetics, leading to a better understanding of their respective limitations as glucose biomarkers and the understanding of diabetes pathophysiology.

5. Main Hypothesis/Study Questions:

In this study, we will identify and characterize the associations between low frequency and rare (MAF<0.05), exonic variants with fructosamine, glycated albumin and 1,5-AG in participants from the ARIC study

Hypothesis: There are rare variants that contribute to the genetic architecture of fructosamine, glycated albumin and 1,5-AG. These rare variants reflect genetic control of both glycemic and nonglycemic properties of nontraditional glycemic biomarkers.

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: Whole exome sequencing data is available for approximately 10,000 individuals in the ARIC study (approximately 2,000 blacks, 8,000 whites).

Study design: Analysis using glycemic biomarker (fructosamine, glycated albumin and 1,5-AG) 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 exome sequencing data that has passed standard (ARIC approved) quality thresholds will be included. We will exclude individuals without valid fructosamine, glycated albumin and 1,5-AG data available and individuals with prevalent diabetes at visit 2, (defined by self-reported physician diagnosed diabetes or taking diabetes medication) when fructosamine, glycated albumin and 1,5-AG were measured.

Exposure variables: Genetic variants sequenced across the exome.

Exome sequencing data
DNA was extracted from blood collected at visit 1 from ARIC participants. All sequencing was done as part of the CHARGE consortium at the Baylor College of Medicine Human Genome Sequencing Center (HGSC). Samples were pooled and sequenced using paired-end sequencing, run on the Illumina HiSeq 2000 or 2500 platform (San Diego, CA), and exome capture performed with VCRome 2.1 (NimbleGen, Inc., Madison, WI). Sequence alignment and variant calling were done with the Mercury pipeline in the DNAnexus, and reference sequence mapping was done using the Burrows-Wheeler alignment tool with the Genome Reference Consortium Human Build 37 reference sequence. Aligned reads were then recalibrated using the Genome ANalysis ToolKit (GATK) and allele calling and VCF creation was done with the Atlas2 suite (Atlas-SNP and Atlas-Indel).

Exome Sequencing Quality Control
Standard quality control exclusion measures were implemented to ensure accurate, reliable results. Single nucleotide variants (SNVs) were excluded if they met any of the following criteria: posterior probability<0.95, variant read count <3, variant read ratio <0.25 or >0.75, strand bias >99% in single direction, total convergence <10 fold for SNVs (<30x for indels), outside exon capture regions, monomorphic variant, missing rate >20%, mappability score <0.8, mean depth coverage >500 fold, Hardy Weinberg Equilibrium $p<5x10^{-6}$ in ancestry-specific groups. Samples were excluded if they had >20% missing data or beyond 6 standard deviations from the mean read depth, singleton count, heterozygote to homozygote ratio, or transition to transversion (Ti/Tv) ratio. After quality control, 2,556,859 SNVs and 76,133 indels remained, and 7,810 European Ancestry individuals and 3,180 African Ancestry individuals remained.

Outcomes: Fructosamine, glycated albumin, 1,5-AG

Fructosamine (Roche Diagnostics, Indianapolis IN, USA), glycated albumin (Asahi Kasei GA-L, Tokyo, Japan) and 1,5-AG (GlycoMark assay implemented on the Roche ModP, Wiston-Salem, NC) were measured in 2012-2013 using a Roche Modular P800 system from serum collected at visit 2 (1990-92) and stored at -70ºC.[15]

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

Data analysis:

Single variant analysis
Variants have been annotated with functional categories using ANNOVAR and dbNSFP v2.0 ref genome GRCh37 and NCBI RefSeq. We will perform single variant analyses separately by ethnic group and for each glycemic biomarker. Analyses will be linear regression controlling for age, sex, study center and relevant principal components, and run using the comprehensive SeqMeta R package. To prioritize potentially causal and rare variants, we will exclude common variants (MAF>5%) and restrict analysis to variants predicted to be deleterious, namely missense, nonsense, frameshift and splice-site variants.

Gene-based analysis
To augment power for situations where multiple rare variants affect association with a phenotype, methods for aggregating sequencing data into units (eg genes or pathways) have been developed. Each method has unique strengths and limitations, so we will evaluate results from three different methods: a score test with a MAF threshold of <0.05, a burden test and the Sequence Kernel Association Test (SKAT, a kernel based, variance components tests). Burden tests collapse variants into a score which is then evaluated in a regression framework for association with the phenotype, and have greater power when variants are associate with the phenotype in the same direction, but lose power when the associations are in different directions. Variance components tests evaluate the variance of aggregated genetic effects with the phenotype. These tests have greater power when fewer variants are causal or affect risk in both directions. We will run the score test, burden and SKAT, implemented using the SeqMeta package described above. We will perform analyses separately by ethnic group and for each glycemic biomarker.

Comparing rare exonic variants between traditional and nontraditional biomarkers of glycemia
For both single variant and gene based analyses, I will compare results from nontraditional glycemic biomarkers with type 2 diabetes variants previously identified using traditional glycemic markers (fasting glucose and HbA1c). Overlap between the genetic variants associated with traditional and nontraditional biomarkers of glycemia will likely indicate rare, potentially causal variants that are relevant to type 2 diabetes pathogenesis by excluding variants that show an association with a single biomarker due to measurement error rather than a true association with hyperglycemia. We will also evaluate individuals who develop diabetes after visit 2 for enrichment of any significant variants we find associated with fructosamine, glycated albumin and 1,5-AG.

Limitations:

Exome sequencing focuses on the coding region of the genome, and cannot capture variants in the noncoding intronic and intergenic regions. While this will be important for future studies to investigate, there is still a great deal of potential to discover important variants associated with fructosamine, glycated albumin and 1,5-AG. In addition, we are limited in our power to detect associations of rare alleles and more modest effect sizes.

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 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.cscc.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)?

#2387 Comparative genetics of fructosamine, glycated albumin, and 1,5- anhydroglucitol in the Atherosclerosis Risk in Communities Study

#2114 – "Prognostic utility of fructosamine and glycated albumin for incident diabetes and microvascular complications"

#2113 – "The associations of fructosamine and glycated albumin with vascular outcomes"
#2112 – "The prognostic value of 1,4-anhydroglucitol"
#1309 – "Genome-wide admixture mapping analyses of cardiovascular and related metabolic traits"
        -these were included in the GWAS manuscript proposal, necessary to include here too?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes   __x__ 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.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.

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:

1.   Marathe PH, Gao HX, Close KL. American Diabetes Association Standards of Medical Care in Diabetes 2017. *J Diabetes*. 2017;40 supplem:1-142. http://www.ncbi.nlm.nih.gov/pubmed/24357209.
2.   Goldstein DE, Little RR, Lorenz RA, et al. Tests of glycemia in diabetes. *Diabetes Care*. 2004;27(7):1761-1773. http://www.ncbi.nlm.nih.gov/pubmed/15220264.
3.   True MW. Circulating biomarkers of glycemia in diabetes management and implications for personalized medicine. *J diabetes Sci Technol*. 2009;3(4):743-747.
4.   Parrinello CM, Sharrett AR, Maruthur NM, et al. Racial differences in and prognostic value of biomarkers of hyperglycemia. *Diabetes Care*. 2016;39(4):589-595. doi:10.2337/dc15-1360.
5.   Sacks DB. A1C versus glucose testing: A comparison. *Diabetes Care*. 2011;34(2):518-523. doi:10.2337/dc10-1546.
6.   Sacks DB, John WG. Interpretation of hemoglobin A1c values. *JAMA*. 2014;311(22):2271-2272. doi:10.1001/jama.2014.6342.
7.   Armbruster DA. Fructosamine: structure, analysis, and clinical usefulness. *Clin Chem*. 1987;33(12):2153-2163. http://www.ncbi.nlm.nih.gov/pubmed/3319287.
8.   Buse JB, Freeman JLR, Edelman S V, Jovanovic L, McGill JB. Serum 1,5-anhydroglucitol (GlycoMark): a short-term glycemic marker. *Diabetes Technol Ther*. 2003;5(3):355-363. doi:10.1089/152091503765691839.
9.   Soranzo N, Sanna S, Wheeler E, et al. Common Variants at 10 Genomic Loci Influence Hemoglobin A1C Levels via Glycemic and Nonglycemic Pathways. *Diabetes*. 2010;59(12):3229-3239. doi:10.2337/db10-0502.
10.   Chen P, Takeuchi F, Lee JY, et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. *Diabetes*. 2014;63(7):2551-2562. doi:10.2337/db13-1815.
11.   Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. *Hum Mol Genet*. 2015;24(July):ddv264. doi:10.1093/hmg/ddv264.
12.   Burdett T (EBI), Hall PN (NHGRI), Hastings E (EBI), Hindorff LA (NHGRI), Junkins HA (NHGRI), Klemm AK (NHGRI), MacArthur J (EBI), Manolio TA (NHGRI), Morales J (EBI) PH (EBI) and WD (EBI). The NHGRI-EBI Catalog of published genome-wide association studies. www.ebi.ac.uk/gwas. Accessed January 9, 2016.

13. Shin S-Y, Fauman EB, Petersen A-K, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543-550. doi:10.1038/ng.2982.
14. Long T, Hicks M, Yu H-C, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet*. 2017;49(4):568-578. doi:10.1038/ng.3809.
15. Selvin E, Rawlings AM, Grams M, et al. Fructosamine and glycated albumin for risk stratification and prediction of incident diabetes and microvascular complications: A prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) study. *Lancet Diabetes Endocrinol*. 2014;2(4):279-288. doi:10.1016/S2213-8587(13)70199-2.
16. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics*. 2016;32(9):1423-1426. doi:10.1093/bioinformatics/btw079.