ARIC Manuscript Proposal Form

ARIC Publication Admin Use Only

Publication Committee Review Date: **01/14/25**
ARIC Manuscript Proposal Number: **#4593**

1.a. Full Title Proteomics of prediabetes in the ARIC Study

b. Abbreviated Title (Length 26 characters): Proteomics of prediabetes

2. Writing Group [please provide a middle initial if available; EX: Adam L Williams]:
Writing group members: Mary R. Rooney, Jingsha Chen, Keenan A. Walker, Justin B. Echouffo Tcheugui, Christie M. Ballantyne, Eric Boerwinkle, Chiadi E. Ndumele, James S. Pankow, Morgan E. Grams, Elizabeth Selvin, Josef Coresh (others welcome; replication and specialized analyzes will lead to adding individuals who make substantive contributions)

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

First author [please provide a middle initial; EX: Adam L Williams]:

Mary	**R**	**Rooney**
First name	Middle initial	Last name

Address: 2024 E. Monument St., Suite #2-600 ; Baltimore, MD 21287

Phone: 630-995-5772
E-mail: mroone12@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 (The ARIC author should be involved enough in ARIC to be able to point the lead author to appropriate ancillary study PIs and to be able to search ARIC manuscript proposals if the lead author doesn't have the access needed to do such a search).

Name: Josef Coresh
Address: TRB 227 E. 30th Street 7th, 707; New York, NY 10003

Phone: 646-501-2503
E-mail: josef.coresh@nyulangone.org

3. Timeline: We aim to complete the manuscript within 1 year from the time of approval.

4. Rationale:

Prediabetes is a state of intermediate hyperglycemia and indicates an elevated diabetes risk. However, the majority of people with prediabetes do not actually progress to diabetes within ~5-10 years[1-3] (including work led within ARIC at middle-age[2] and older age[3]). Prediabetes is based solely on binary elevations in one or more glycemic biomarkers and do not account for variation in disease pathophysiology within these groups.

There are underlying biologic mechanisms leading to diabetes beyond hyperglycemia alone. Improvements in proteomic measurements provide an opportunity to explore biologic pathways implicated in the progression from prediabetes to diabetes. A growing number of studies have identified known and novel proteomic biomarkers associated with diabetes incidence[4-910,11]. In ARIC, for example, we identified 26 proteins that improved prediction of ~20-year diabetes incidence beyond traditional risk factors (including baseline FG, HbA1c).[10] Few studies have characterized the proteomic signatures of prediabetes and identified proteins associated with short-term progression to diabetes.[12]

The underlying biology related to prediabetes remission, i.e., reverting to normoglycemia, is not well-understood. Prediabetes remission may relate to weight loss or intra-individual variability near prediabetes thresholds is another explanation. Proteomics could improve our understanding of the mechanisms underlying short-term prediabetes remission. Interestingly, prespecified secondary analysis of the SELECT trial showed that semaglutide (vs placebo) was associated with higher reversion to normoglycemia independent of weight loss.[13] These novel glucose-lowering medications also appear to target inflammatory pathways.

Using SomaScan data with ~5000 proteins measured in ARIC participants, we will examine the proteomic signatures associated with prediabetes and identify proteins associated with short-term progression from prediabetes to diabetes and regression to normoglycemia.

5. Main Hypothesis/Study Aims:

What are the proteomic predictors of prediabetes and short-term progression to diabetes and regression to normoglycemia?

6. Design and analysis - please address the following aspects:

a) inclusion/exclusion
b) study design
c) outcome and other variables of interest with specific reference to the time of their collection
d) summary of data analysis
e) Any anticipated methodologic limitations or challenges if present

Design: Cross-sectional and prospective analysis in persons without prevalent diabetes (visit 2). Prospective analyses will focus on persons with prevalent prediabetes (visit 2 FG 100-125 mg/dL; HbA1c measured at visit 2 but not visit 3).

Exclusions:

- Missing SOMAscan data at visit 2 or failed QC
- Missing covariate data

- Non-fasting blood draw
- Prevalent diabetes at baseline (visit 2)

Exposures: SomaScan proteins with ANML (~5,000 proteins measured using SomaScan's aptamer-based proteomics platform). We will be guided by the QC document following the current recommended data cleaning, focus on log protein levels, and use flag 2 to exclude unreliable data (flag2=0 for N = 4955 aptamers excluding flag 1 cat1-3: Fc mouse/Contaminants, or variance<0.01 or SomaScan QC CV>50% at v2, v3, or v5).

Covariates: age, sex, race, center, eGFR, family history of diabetes, body mass index, current smoking, total cholesterol, HDL-cholesterol, systolic blood pressure, hypertension medication use, physical activity (sport index)

Outcomes: We will examine and compare cross-sectional proteomic associations with different ADA definitions of prediabetes (e.g. visit 2 FG 100-125 mg/dL vs visit 2 HbA1c 5.7-6.4%). In prospective analyses, our outcome will be progression to diabetes (visit 3 diabetes diagnosis, medication, or FG >=126 mg/dL) or regression to normoglycemia (visit 3 FG <100 mg/dL).

Statistical Analysis

We will $\log_2$-transform protein abundances. When comparisons across proteins is desirable, we will analyze protein levels on a standardized scale (mean=0, SD=1). We will use logistic regression to examine proteomic associations with prevalent prediabetes (based on FG 100-125 mg/dL and separately based on HbA1c 5.7-6.4%). We will compare proteins that are associated with both definitions vs proteins that are associated with only one definition. In prospective analyses, we will estimate 3-year rates of 1) progression from prediabetes to diabetes and of 2) regression to normoglycemia. We will use logistic regression models to examine progression from prediabetes to diabetes or regression from prediabetes to normoglycemia. We will examine correlations of the top protein hits with baseline FG, HbA1c, and BMI (modeled continuously). We will use a set of hierarchical models, shown below.

- M0: unadjusted
- M1: age, sex, eGFR
- M2: M1 + race-center, BMI, current smoking, total cholesterol, HDL-cholesterol, systolic blood pressure, hypertension medication use, physical activity
- we will further adjust for hyperglycemia to discover novel markers:

M3 (prospective analysis only): M2 + baseline fasting glucose

To determine statistical significance threshold, we will apply the Bonferroni correction in the most conservative analysis. FDR (q<0.05) will be used to examine a broader list of proteins for pathway analyses (using ENRICHR software).

Prediction: We will examine 3-year risk of diabetes and 3-year risk of regression to normoglycemia. Using the Bonferroni significant hits, we will consider a comparison of M3 above (full risk factor model including measures of hyperglycemia) as the comparison model to be improved (testing change in AUC).

Limitations: HbA1c measurement is not available at visit 3 for the progression analysis. The lack of a separate cohort for external validation is another important potential limitation – we will seek external replication in collaboration with Peter Ganz and others working with SomaLogic.

f) **Will the author need Limited data to complete the proposed manuscript? ☒ Yes, Limited data is needed (Provide a brief (2-3 sentences) justification for requesting PHI data)** This proposal requires the use of proteomics/SomaLogic data, which falls under the limited data category. Other limited data (including dates, CMS, genomic, geocoded) are not required. **☐ No, De-identified data will be sufficient.**

*Please note, Limited dataset access is strict and rarely provided. Limited data includes identifiable information such as dates (birthdays, visit dates, etc.). CMS, Genomic, Geocoded, *Proteomics/Somalogic*, and other -omic data all fall under the limited data category. De-identified data does not include dates. All dates are date adjusted to "Days since Visit 1".

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? (Non-ARIC analysis means that the authors are not regarded as ARIC investigators and the "ARIC author" is essentially just a facilitator rather than an integral part of the writing group.) ☐ Yes ☒ 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 is distributed to ARIC PIs annually, 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 current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"? ☐ Yes ☐ No

9. The lead author or the "sponsoring" ARIC 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 website at: https://aric.cscc.unc.edu/aric9/proposalsearch [ARIC Website□Publications□Proposal Search]

☒ 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)?
Rooney MS #3803 proteomics of diabetes

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or does it use current [or ongoing] ancillary study data (this includes ACHIEVE)? ☒ Yes ☐ No → Skip to question 12

11.b. If yes to 11.a., is the proposal

☒ **A. primarily the result of an ancillary study**

☐ **B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables)**

11.c. If yes to 11.a., list number* 2017.27

*ancillary studies are listed by number
https://aric.cscc.unc.edu/aric9/researchers/ancillary_studies/approved_ancillary_studies [ARIC Website□Ancillary Studies□Approved Ancillary Studies]

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 https://aric.cscc.unc.edu/aric9/publications/policies_forms_and_guidelines [ARIC Website□Publications□Publication Policies, Forms, and Guidelines]. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References: ____

To view publications materials, click "Log in" at the top right of the ARIC website. Click "Forgot Password" if you are experiencing issues with logging in.

1. van Herpt TTW, Ligthart S, Leening MJG, et al. Lifetime risk to progress from pre-diabetes to type 2 diabetes among women and men: comparison between American Diabetes Association and World Health Organization diagnostic criteria. *BMJ open diabetes research & care.* 2020;8(2).
2. Wallace AS, Rooney MR, Fang M, Echouffo-Tcheugui JB, Grams M, Selvin E. Natural History of Prediabetes and Long-term Risk of Clinical Outcomes in Middle-aged Adults: The Atherosclerosis Risk in Communities (ARIC) Study. *Diabetes Care.* 2022.

3. Rooney MR, Rawlings AM, Pankow JS, et al. Risk of Progression to Diabetes Among Older Adults With Prediabetes. *JAMA internal medicine.* 2021;181(4):511-519.
4. Nowak C, Sundström J, Gustafsson S, et al. Protein Biomarkers for Insulin Resistance and Type 2 Diabetes Risk in Two Large Community Cohorts. *Diabetes.* 2016;65(1):276-284.
5. Gudmundsdottir V, Zaghlool SB, Emilsson V, et al. Circulating Protein Signatures and Causal Candidates for Type 2 Diabetes. *Diabetes.* 2020;69(8):1843-1853.
6. Beijer K, Nowak C, Sundström J, Ärnlöv J, Fall T, Lind L. In search of causal pathways in diabetes: a study using proteomics and genotyping data from a cross-sectional study. *Diabetologia.* 2019;62(11):1998-2006.
7. Molvin J, Pareek M, Jujic A, et al. Using a Targeted Proteomics Chip to Explore Pathophysiological Pathways for Incident Diabetes- The Malmö Preventive Project. *Scientific reports.* 2019;9(1):272.
8. Elhadad MA, Jonasson C, Huth C, et al. Deciphering the Plasma Proteome of Type 2 Diabetes. *Diabetes.* 2020;69(12):2766-2778.
9. Gou W, Yue L, Tang X-y, et al. Circulating Proteome and Progression of Type 2 Diabetes. *The Journal of Clinical Endocrinology & Metabolism.* 2022;107(6):1616-1625.
10. Rooney MR, Chen J, Echouffo-Tcheugui JB, et al. Proteomic Predictors of Incident Diabetes: Results From the Atherosclerosis Risk in Communities (ARIC) Study. *Diabetes Care.* 2023;46(4):733-741.
11. Yao P, Iona A, Pozarickij A, et al. Proteomic Analyses in Diverse Populations Improved Risk Prediction and Identified New Drug Targets for Type 2 Diabetes. *Diabetes Care.* 2024;47(6):1012-1019.
12. Carrasco-Zanini J, Pietzner M, Lindbohm JV, et al. Proteomic signatures for identification of impaired glucose tolerance. *Nature Medicine.* 2022.
13. Kahn SE, Deanfield JE, Jeppesen OK, et al. Effect of Semaglutide on Regression and Progression of Glycemia in People With Overweight or Obesity but Without Diabetes in the SELECT Trial. *Diabetes Care.* 2024;47(8):1350-1359.