ARIC Manuscript Proposal Form

ARIC Publication Admin Use Only

Publication Committee Review Date: **12/10/24**
ARIC Manuscript Proposal Number: **# 4583**

1.a. Full Title: Epigenome-wide association study of blood proteome in CKD in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters): **Epigenome-wide association study of blood proteome in CKD**

2. Writing Group [please provide a middle initial if available; EX: Adam L Williams]:
Writing group members: Yang Li, Aditya Surapaneni, Pascal Schlosser, Josef Coresh, Morgan Grams

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

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

Yang | | Li
First name | Middle initial | Last name

Address: 227 E. 30th Street, #827E, New York, NY 10016

Phone:347-261-6030
E-mail: yang.li@nyulangone.org

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:Morgan Grams
Address: 227 E. 30th Street, #825, New York, NY 10016

Phone: 917-232-7960
E-mail: morgan.grams@nyulangone.org

3. Timeline: two years

4. Rationale:

The blood proteome consists of thousands of circulating proteins that serve as potential predictors of a wide range of diseases. DNA methylation, a key epigenetic mechanism regulating gene expression, plays a critical role in modulating dynamic changes in protein levels, thereby influencing disease development and prognosis. Moreover, DNA methylation is highly impacted by environmental exposures, both in long- and short-term.

Previously, we utilized the SomaScan V4 platform to profile ~5,300 plasma proteins in around 5,200 participants and the Illumina methylation EPIC V1 array to analyze 850,000 CpG sites in peripheral blood cell DNA from 2,150 participants diagnosed with chronic kidney disease (CKD) within the ARIC study. Building on this data (~2,000 participants have both proteomic and methylomic data), we propose a systematical epigenome-wide association study (EWAS) of the blood proteome to identify associations between DNA methylation and circulating proteins, investigate their connections to environmental and lifestyle factors, and evaluate their roles in the development of various diseases and progression of CKD.

5. Main Hypothesis/Study Aims:

We hypothesize that circulating protein levels are associated with DNA methylation at CpG sites in peripheral blood cell DNA, which are influenced by environmental and lifestyle factors. Furthermore, we propose that these environmental exposures contribute to the development of diseases such as diabetes and cardiovascular disease (CVD) and the progression of CKD such as 40% decline in eGFR and End-stage kidney disease (ESKD) through the DNA methylation-circulating protein axis.

Specific aims:

1. to perform epigenome-wide association study (EWAS) to systematically assess associations between DNA methylation at ~850,000 CpG sites and ~5,300 circulating proteins, providing insights into potential epigenetic mechanisms regulating circulation proteins among people with CKD.

2. to investigate environmental and lifestyle influences on DNA methylation associated with circulating proteins, revealing the interplay among environmental exposure, DNA methylation, and circulating proteins among people with CKD.

3. to elucidate the contribution of the identified DNA methylation-protein associations to cardiometabolic and cardiorenal phenotypes, as well as the risk of developing diabetes and CVD and CKD progression, aiming to uncover biomarkers and potential pathways for disease prevention and intervention among people with CKD.

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

a) inclusion/exclusion

Inclusion:

Participants with both blood proteomic data and methylomic data in the ARIC study who were selected into the V5 epigenetic study (funding Coresh/Grams)
Available baseline characteristics, including demographics (e.g. age, gender), environmental exposures (e.g. drinking, smoking), lifestyle factors (e.g. physical activity, diet, sedentary behavior, sleeping pattern).
Available follow-up clinical outcomes at visit 5-10, including cardiometabolic and cardiorenal phenotypes (e.g. glycemic measures, lipid profiles, blood pressure, kidney functions) as well as incidence of ESKD, diabetes, and CVD.

Exclusion:

Missing proteomic and methylomic data

b) study design

1. Cross-sectional study to conduct epigenome-wide association study of circulating proteins in patients with CKD (e.g., eGFR <60 ml/mn/1.73 m2 or ACR >30 mg/g) (visit 5)
2. Retrospective study to assess associations between DNA methylation at visit 5 in people with CKD and antecedent environmental exposures (visit 1 and visit 2)
3. Longitudinal study to examine associations between DNA methylation/circulating proteins at visit 5 among people with CKD and clinical outcomes (visit 6-10)

c) outcome and other variables of interest with specific reference to the time of their collection

1. Primary outcomes are ~5000 circulating proteins
2. Secondary outcomes are fasting glucose, SBP, DBP, TG, TC, LDL, HDL, eGFR, uACR, diabetes, CVD events (including myocardial infarction, stroke, heart failure, and other major adverse cardiovascular events), and CKD progression (including 40% decline in eGFR and ESKD)

d) summary of data analysis

1. Summary statistics of circulating proteins and DNA methylation
2. Descriptive statistics of baseline demographics, environmental factors, and follow-up clinical outcomes
3. Associations between DNA methylation and circulating proteins
4. Associations between environmental factors and identified DNA methylation.
5. Association between identified DNA methylation and clinical outcomes.
6. Mediating effect of identified DNA methylation in the relation between environmental factors and circulating proteins, and effect of corresponding circulation proteins in the relation between DNA methylation and clinical outcomes.

e) Any anticipated methodologic limitations or challenges if present

1. Interpretation of associations between DNA methylation and circulating proteins, particularly when CpG sites are not located in the promoters of the corresponding protein-coding genes.
2. Interpretation of the regulation network between multiple identified CpG sites and circulation proteins.
3. Assess pathways and mediating effect size of DNA methylation and circulating proteins in the relationship between environmental factors and clinical outcomes.

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) . ☐ 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)?

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* __

*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.