# ARIC Manuscript Proposal Form

**ARIC Publication Admin Use Only**

Publication Committee Review Date: **08/13/24**
ARIC Manuscript Proposal Number: **# 4500**

**1.a. Full Title:** Relating DNAm protein signature to incident dementia risk in ARIC

**b. Abbreviated Title (Length 26 characters):**

**2. Writing Group [please provide a middle initial if available; EX: Adam L Williams]:**
Writing group members: Shannon M. Drouin; Gabriela Gomez; Jan Bressler: Jeannette Simino; Rebecca Gottesman; Keenan A Walker

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

| **Shannon** | **M** | **Drouin** |
|---|---|---|
| First name | Middle initial | Last name |

Address: 251 Bayview Boulevard
Room 04B316
Baltimore, MD 21224

Phone:
E-mail: shannon.drouin@nih.gov

**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: Keenan A. Walker
Address: 251 Bayview Boulevard
Room 04B316
Baltimore, MD 21224

Phone: (667) 205-2657
E-mail: keenan.walker@nih.gov

**3. Timeline**:
1-3 months: analysis of data
3-6 months: writing of manuscript

**4. Rationale**: Previously, several large-scale proteomic studies have found important associations between plasma proteins and AD-related outcomes (e.g., cognitive decline, cognitive impairment, regional brain volumetric loss, increased Aβ pathology) as well as dementia risk [1-3]. Despite these associations, reports of short-term and rapid fluctuating plasma protein levels within individuals remains a notable limitation of using plasma levels as markers of long-term protein exposure [4-6]. For example, plasma levels of interleukin-6 and C-reactive protein (CRP) have been reported to be especially prone to short-term fluctuations [7, 8]. In response to this, DNA methylation measures of proteins have been lauded as more robust and stable measures of protein exposure [6, 8]. In a recent study, a methylation measure of CRP predicted cardiovascular outcomes more strongly than measured plasma levels [9]. Similarly, a methylation-based CRP score (but not serum CRP) was found to be associated with cognitive ability and have higher re-test reliability [8].

In a parallel study, we plan to test whether protein epigenetic scores are associated with cognitive and brain volume trajectories longitudinally in the Baltimore Longitudinal Study of Aging (BLSA). Subsequently, we plan to test associations between these same protein epigenetic scores and dementia risk in the ARIC sample. Associations with dementia risk will provide additional support for our results relating DNA methylation protein signatures and cognitive/brain aging outcomes.

**5. Main Hypothesis/Study Aims**:
**Study Objective**. Determine the degree to which DNA methylation protein signatures (as compared to measured plasma protein levels) are associated with incident dementia risk.

**Aim 1**. Examine and compare the associations between DNA methylation protein signatures and measured plasma protein levels, each, with 20- and 30-year dementia.

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

**a) study design**
**b) inclusion/exclusion**
**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**

a) Study Design

Primary analyses will relate standardized (z-score) DNAm protein measures (from visit 2) to subsequent dementia risk through visit 5 (20-year risk) and visit 9 (30-year risk). After determining the association of DNAm protein measures with dementia risk, we will compare the strength of these associations to that of the association between corresponding standardized (z-score) plasma protein levels (measured with SomaScan) and dementia risk across the same time intervals. Both DNAm measures and plasma protein levels will be standardized to allow for comparisons across analyses.

b) Inclusion/Exclusion
   a. *Inclusion Criteria*:
      i. Epigenetic data available at visit 2
      ii. SomaScan plasma protein measures at visit 2
      iii. Dementia follow-up data available
   b. *Exclusion Criteria*:
      i. Non-white or non-black race, and Black participants from Minneapolis and Washington County due to small numbers
      ii. Missing essential covariate information
      iii. Missing dementia follow-up data
      iv. Diagnosis of dementia before or at visit 2

c) Outcome and Exposure Variables:

*Outcome variable*

*a.* **All-cause dementia**. We will use ARIC dementia outcomes (Level 3), as captured until visit 9.

*Exposure variable*

**b. Protein DNAm Score**. We will generate standardized epigenetic scores for a set of plasma proteins from participants with available epigenetic data collected at Visit 2, as per the methods discussed by Gadd and colleagues [6]. Secondary analyses will consider alterative published models for epigenetic score generation. Epigenetic scores have been previously developed using results of methylation-wide association studies with the measured protein as the outcome. This is comparable to polygenic risk scores for individual proteins, but using epigenetic – rather than genetic – information.

d) Summary of Data Analysis

**Aim 1**. Examine the association between standardized protein epigenetic scores with 20- and 30-year dementia risk.

**Data analysis plan**. After computing protein-specific DNAm score, we will determine the correlation between standardized DNAm scores and directly measured standardized protein level for each measured protein with a corresponding DNAm score.

For each protein DNAm score, we will use cox regression models to examine the association between visit 2 protein-specific DNA methylation scores and time to

dementia up to visit 5 (20-year dementia risk) and visit 9 (30-year dementia risk). We will methodically test multiple models to examine the effect of potential confounders. Model 1 will adjust for potentially confounding demographic variables, including age at plasma sample acquisition, sex, race-study center, education, and *APOE-ε4* status. Model 2 will additionally adjust models for other confounding factors (i.e., eGFR, BMI, hypertension, diabetes, and current smoking). As part of a sensitivity analysis, we will examine additional models that account for previously identified genetic variants that influence plasma protein level (protein quantitative trait loci [pQTLs]).

e) Limitations
Lack of sufficiently strong signal is not evidence of no association with a given protein signature, since lack of signal may be due to limited power for a host of reasons including the need to be conservative when adjusting for multiple comparisons.

**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) The current study focuses on associations between epigenetic protein signatures (epigenomics) and dementia risk . ☐ 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)?**

MP 2337A. Epigenetic Age Acceleration and Cognitive Function in African American Adults in Midlife: The Atherosclerosis Risk in Communities Study.

MP U0427. Epigenetic and integrative cross-omics analyses o cerebral white matter hyperintensities on MRI

MP 4240. Proteome-wide association study of Alzheimer's disease

MP 3327. A protomic analysis of incident dementia: The ARIC study

**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. "Proteomic longitudinal ARIC study: SOMAscan of multiple visits"

*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:


1. Dark, H.E., et al., *Proteomic Indicators of Health Predict Alzheimer's Disease Biomarker Levels and Dementia Risk.* Annals of neurology, 2024. **95**(2): p. 260-273.
2. Jiang, Y., et al., *Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging.* Alzheimer's & Dementia, 2022. **18**(1): p. 88-102.
3. Walker, K.A., et al., *Proteomics analysis of plasma from middle-aged adults identifies protein markers of dementia risk in later life.* Science translational medicine, 2023. **15**(705): p. eadf5681.
4. Conole, E.L., et al., *DNA methylation and protein markers of chronic inflammation and their associations with brain and cognitive aging.* Neurology, 2021. **97**(23): p. e2340-e2352.
5. Stevenson, A.J., et al., *Creating and validating a DNA methylation-based proxy for interleukin-6.* The Journals of Gerontology: Series A, 2021. **76**(12): p. 2284-2292.
6. Gadd, D.A., et al., *Epigenetic scores for the circulating proteome as tools for disease prediction.* Elife, 2022. **11**: p. e71802.
7. Dugué, B. and E. Leppänen, *Short-term variability in the concentration of serum interleukin-6 and its soluble receptor in subjectively healthy persons.* 1998.
8. Stevenson, A.J., et al., *Characterisation of an inflammation-related epigenetic score and its association with cognitive ability.* Clinical Epigenetics, 2020. **12**: p. 1-11.
9. Hillary, R.F., et al., *Blood-based epigenome–wide analyses of chronic low–grade inflammation across diverse population cohorts.* medRxiv, 2023: p. 2023.11. 02.23298000.


**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.**