ARIC Manuscript Proposal #3999

PC Reviewed: 1/11/22 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Protein patterns as predictors of dementia and neuropathology risk.

b. Abbreviated Title (Length 26 characters): Aptamers and dementia risk.

2. Writing Group:
Writing group members:
SomaLogic: Clare Paterson, Kelsey Loupy, Rachel Ostroff, Stephen Williams, Leigh Alexander, Amy Zhang
Collaborators: potentially collaborators from Whitehall II study.
ARIC Collaborators: Josef Coresh, *others welcome*

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

First author: Clare Paterson
Address: 2995 Wilderness Place
Boulder CO 80031

Phone: 720-214-7487 Fax:
E-mail: cpaterson@somalogic.com

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: Josef Coresh
Address: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health and Welch Center for Prevention, Epidemiology, and Clinical Research,
2024 E. Monument St., Suite 2-600,
Baltimore, MD 21287

Phone: 410-955-0495 Fax:
E-mail: coresh@jhu.edu

3. Timeline: We anticipate that the manuscript will be drafted and submitted to the ARIC Publications Committee within 6 months of the manuscript proposal being approved.

4. Rationale:

The US Preventive Services Task Force (USPSTF) determined that there is inconclusive evidence for or against recommending regular screening for dementia in asymptomatic adults, although the need to identify cognitive impairment early on is acknowledged (US Preventive Services Task Force, 2020). While the greatest risk factor for dementia is age, with the incidence of dementia doubling every 5.9 years after the age of 60 (Sosa-Ortiz et al., 2012), genetic, demographic, and health history factors can alter an individual's future dementia risk. In many dementia patients by the time cognitive decline is detectable and a clinical diagnosis is established, brain pathology is already evident (Dubois et al., 2016).

Given the increasing global impact of dementia (for example Alzheimer's disease is the 6th leading cause of death currently affecting over 6 million adults, a number expected to double in the next 30 years (Alzheimer's Association Report, 2021)) development of precision medicine prognostic tools that can stratify an individual's future risk for dementia or dementia-related brain pathology may enable both physicians and patient's clinical decision-making. Moreover, these tools may also be of utility in identifying eligible individuals for dementia-related therapeutics that have recently received or are currently seeking FDA fast track approvals (Mullard, 2021).

To this end, we have developed an aptamer-based plasma protein-only model that predicts risk of a dementia diagnosis within 20 years from blood draw. This model was developed in dementia-free participants from ARIC Visit 3 using level 3 adjudication criteria for dementia diagnosis. Current research is focused on applying this model to additional datasets, including ARIC visit 5 without known dementia at time of blood draw, as well as comparing protein-only model performance to that of known genetic risk factors (e.g., ApoE). Additionally, identifying whether this incident dementia risk-prediction model can also predict dementia-related brain pathology (e.g. beta-amyloid accumulation) is of key interest.

Within the next year, we are interested in developing a similar protein-based predictive risk test using blood samples from ARIC Visit 5 (older population and within a shorter time frame) as well as models that predict future and current-state dementia-related brain pathology based on brain imaging and biomarker data that are available from ARIC Visits 5 and 6.

5. Main Hypothesis/Study Questions:

We hypothesize:

A. A plasma proteomic risk model for 20-year prediction of dementia diagnosis from middle-age blood draw will be at least equivalent to the maximum published performance of genetic risk prediction using the APOE2 and APOE4 gene variants (AUC= 0.678) (Escott-Price et al., 2015), the variant assessed in the only FDA-approved direct-to-consumer test available for Alzheimer's disease prediction through 23andMe® (U.S. Food and Drug Administration, 2017).
B. The mid-life proteomic dementia risk model will perform better (based on AUC) to predict risk of future dementia diagnosis than a model based on APOE genotype only.

C. The mid-life proteomic dementia risk model will also be successful at predicting 5-year risk of dementia diagnosis (AUC $\geq$ 0.678) when applied to an older population (using blood samples from ARIC Visit 5).

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).

The proposed manuscript will describe the development and validation of the mid-life proteomic dementia risk model, which is currently in validation phase. We are currently testing this model in new populations, including elderly ARIC participants (Visit 5) with no history of dementia.

Plasma proteins were measured using the SomaScan assay, version 4, which measures approximately 5000 human proteins (Williams et al., 2019). ARIC Visit 3 was the source of model training, verification, and validation. We are additionally assessing the performance of the model built on ARIC visit 3 to apply to samples from ARIC visit 5, predicting risk for dementia within 5 years.

The objective of the mid-life dementia risk model is to identify individuals who are at an increased risk of developing dementia within 20 years from blood draw. The intended use population for this test is dementia-free middle-aged adults (49-73 years old) who are at average-risk of developing, but without a current diagnosis of, dementia. The endpoint for this model is a binary yes/no outcome of dementia diagnosis based on level 3 adjudication (a combination of cognitive assessment tests, telephone screening, informant ratings, hospital records, and death record review). A goal of future dementia-related models would be to distinguish patients who may be eligible, in the future or at the time of blood draw, for amyloid beta-directed Alzheimer's disease treatments, several of which are now FDA-approved or seeking fast-track approval (Mullard, 2021).

The ARIC Visit 3 dataset includes individuals with 11,277 total samples passing SomaScan QC metrics and with 1937 (17.7%) dementia diagnoses observed within 20 years of visit 3 blood draw. The data were split into 70% for training, 15% for verification, and 15% for validation.

The final model is an AFT survival model with a Weibull distribution. This model has 25 aptamers as features. The output of this model will be risk probability predictions for a dementia diagnosis within 20-years.

The prespecified model performance criteria was based on AUC, requiring that AUC must be better than the performance of the APOE4 genotype at predicting lifetime dementia risk.

The performance metrics have been met through verification, as described in the table below.

Dataset	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	C-Index (95% CI)
Training (70% ARIC V3)	0.732 (0.712, 0.748)	0.612 (0.577, 0.646)	0.729 (0.717, 0.741)	0.696 (0.679, 0.709)

Verification (15% ARIC V3)	0.733 (0.691, 0.772)	0.605 (0.541, 0.676)	0.743 (0.713, 0.768)	0.690 (0.658, 0.716)

We are currently assessing the performance of this model 1) compared to the prediction of 20-year dementia risk based on APOE genotype and 2) predicting 5-year risk of dementia, using a subset of the same participants at ARIC Visit 5 without current dementia. Additionally, we seek to determine the ability of the mid-life dementia risk proteomic model in predicting the risk of developing neuropathology (e.g., beta-amyloid accumulation) as assessed at Visits 5 and 6 in a subset of study participants.

If successful, this publication will describe the derivation and validation of the mid-life dementia risk model in a diverse population.

Additionally, in the near future we seek to collaborate with the ARIC investigators to develop protein-based models that predict risk of future and current states of dementia-related neurological pathology, correlating risk scores with brain imaging and biomarker data available from ARIC Visits 5 and 6, as well as a protein model which predicts short-term risk of a dementia diagnosis from ARIC visit 5 samples. To accomplish this, we request discussions with the relevant ARIC investigators and access to the additional clinical data for brain imaging and biomarkers.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? Yes

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
(The 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

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"? No

Commented [CP1]: @ARIC team, we do not have this field in our data set, can we please have confirmation that the samples sent to SomaLogic for proteomic analyses consented to DNA data use. Thank you

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/mantrack/maintain/search/dtSearch.html

I don't have access to the above link. I have searched public databases and I believe these are the most relevant ARIC publication:

1. Keenan et al. Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. Nature Aging 2021 (1)473-489.

This study focused on development of a protein model of 5-year dementia risk including genetic, clinical and demographic co-factors from ARIC visit 5 (elderly population), and studied the underlying biology of dementia-related proteins. The current proposal will focus on a protein-only model of 20-year dementia risk from visit 3 (middle age), therefore the overlap is limited. This work will be referenced in our proposed manuscript.

2. Lindbohm et al. Plasma proteins, cognitive decline, and 20-year risk of dementia in the Whitehall II and Atherosclerosis Risk in Communities studies. Alzheimer's Dementia 2021 1-13

This study focused on identifying overlapping proteins associated with cognitive decline and 20-year dementia risk in the Whitehall II cohort, and replicated proteomic findings in ARIC visit 3. While the timepoint (visit 3) and endpoint (dementia diagnosis) overlap with the current proposal the main focus of study was the identification of biomarkers rather than proteomic model development and also integrated cognitive decline which we have not considered in our analayis. Therefore, the overlap to the current proposal is limited. This work will be referenced in our proposed manuscript.

Overall, the biggest areas of non-overlap between the current proposal and published findings from ARIC so far, are that the current proposal approaches are completely agnostic of underlying biology. The focus is to develop and validate protein-only models selected by machine learning approaches for the purpose of developing precision medicine tools.

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)?

I am not aware of related manuscript proposals.

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

Yes

11.b. If yes, is the proposal

_X A. primarily the result of an ancillary study (list number* 2017.27**)**

___ 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 https://sites.cscc.unc.edu/aric/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 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.

References

Alzheimer's Association Report. 2021 Alzheimer's disease facts and figures. *Alzheimer's Dement.* 2021; 17(3): 327–406.

Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimers Dement*. 2016; 12(3): 292–323.

Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain.* 2015; 138(12): 3673–3684.

Mullard, A. More Alzheimer's drugs head for FDA review: what scientists are watching. *Nature* 2021; 599: 544-545 . https://doi.org/10.1038/d41586-021-03410-9

Sosa-Ortiz AL, Acosta-Castillo I, Prince MJ. Epidemiology of dementias and Alzheimer's disease. *Arch Med Res*. 2012; 43(8): 600–608.

U.S. Food and Drug Administration. "FDA allows marketing of first direct-to-consumer tests that provide genetic risk information for certain conditions" [press release]. April 2017. https://www.fda.gov/news-events/press-announcements/fda-allows-marketing-first-direct-consumer-tests-provide-genetic-risk-information-certain-conditions

US Preventive Services Task Force. Screening for Cognitive Impairment in Older Adults: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2020; 323(8): 757–763.

Williams S.A., Kivimaki M, Langenberg C, et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* 2019; (12):1851–1857. https://doi.org/10.1038/s41591-019-0665-2