ARIC Manuscript Proposal #4112

PC Reviewed: 9/13/22 **Status: _____** **Priority: 2**

1.a. Full Title: Plasma proteome-wide association study of motoric cognitive risk

b. Abbreviated Title (Length 26 characters):
Proteomics of motoric cognitive risk

2. Writing Group:
Writing group members:
Gabriela T. Gomez, Sanish Sathyan, Jingsha Chen, Myriam Fornage, Pascal Schlosser, Zhongsheng Peng, Priya Palta, Kevin Sullivan, Adrienne Tin, B. Gwen Windham, Rebecca F. Gottesman, Josef Coresh, Nir Barzilai, Sofiya Milman, Joe Verghese, Keenan A. Walker (senior author), others welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal.

First author: Gabriela T. Gomez

Address: NIA (National Institute on Aging)
BRC BG RM 04B311
251 Bayview BLVD
Baltimore MD 21224
Phone: 667-205-2657

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: Keenan Walker, PhD.

Address: NIA (National Institute on Aging)
BRC BG RM 04B311
251 Bayview BLVD
Baltimore MD 21224
Phone: 667-205-2657

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

4. Rationale:

Motoric cognitive risk (MCR), a syndrome defined by slow gait speed and subjective cognitive complaints, has been associated with increased risk for cognitive impairment and subsequent dementia.[1–3] While MCR may overlap with other predementia syndromes, such as mild cognitive impairment (MCI), MCR and MCI can and often do occur independently of one another.[4]

Our workgroup recently demonstrated in the Atherosclerosis Risk in Communities (ARIC) visit 5 cohort that MCR has predictive value for incident dementia that is unique from that of MCI[5]; we also found that, compared to a reduction in brain volume associated with MCI, MCR-associated regional atrophy was more dominant in frontal and parietal brain regions,[5] a pattern consistent with other neuroimaging studies.[6,7] In addition, we demonstrated that the severity of white matter hyperintensities associated with MCR was greater than that associated with MCI. Also linked to frontal lacunar infarcts,[8] MCR may reflect subcortical cerebrovascular disease. Elucidating the pathogenesis of MCR, compared with other predementia syndromes like MCI, will further clarify the shared and distinct pathophysiological events preceding MCR and MCI.

Limited genetic studies characterizing MCR exist. One report from the Health and Retirement study found that obesity related polygenic scores (PGS) (i.e., PGS for greater BMI and waist circumference) increased risk for MCR.[9] Another study of Ashkenazi Jewish adults aged 65 and older in the LonGenity study reported increased risk for incident MCR associated with the interleukin-10 gene.[10] Genomic studies have been performed for slow gait speed[11,12] and cognitive decline[13–15]; however, a proteomic analysis of MCR has not been reported. A deeper understanding of peripheral biological changes that promote development of the diverse clinical presentations of predementia syndromes, including MCR, may be achievable using a data-driven analysis of the plasma proteome.

The goal of the current study is to use SomaScan Multiplexed Proteomic technology[16,17] to examine the relationship between the plasma level of a set of proteins and the MCR syndrome within the ARIC cohort. In this proteome-wide association study of MCR, we first aim to identify proteins measured at ARIC Visit 5 (designated “late-life”) that are associated with concurrent MCR. We will compare the plasma proteins associated with MCR to those associated with MCI to determine whether these two pre-dementia syndromes have distinct proteomic signatures. The LonGenity study is also conducting a discovery analysis of MCR-associated proteins; in collaboration with LonGenity, we will conduct an external replication analysis of select plasma proteins identified by this group. Identified MCR-associated proteins within ARIC will next be examined during the mid- to late life transition (ARIC Visit 2) to determine whether they continue to demonstrate an association with the MCR syndrome approximately two decades earlier. MCR-associated proteins in ARIC will also be validated in external samples. Using available GWAS summary statistics of vascular dementia and Alzheimer’s disease,[18–20] we will conduct bidirectional two-sample Mendelian randomization to investigate potential causal relationships between the proteins of interest and MCR. Finally, systems level analyses will be applied to identify protein networks represented by the candidate proteins to better understand the peripheral biological processes and regulatory mechanisms underlying MCR pathogenesis.

5.Main Hypothesis/Study Questions:

Objective 1. Identify plasma proteins cross-sectionally associated with MCR in late life.

Hypothesis. A number of proteins measured during late-life (Visit 5) will be associated with MCR, compared to normal cognition, after correction for multiple comparisons.

Hypothesis. Expression of protein networks, defined using Netboost, will be associated with MCR, compared to normal cognition.

Objective 2. Compare plasma proteins and protein networks cross-sectionally associated with MCR to plasma proteins and protein networks associated with MCI in late life.

Hypothesis. The set of MCR-associated proteins will overlap partially with MCI-associated proteins.

Hypothesis. The proteins associated with MCR, but not MCI, will be enriched for vascular and cardiometabolic pathways.

Objective 3. Midlife and external replication of ARIC MCR-associated proteins and networks.

Hypothesis. A subset of ARIC-identified MCR-associated discovery proteins will also be associated with late-life MCR when these proteins are examined in blood collected during midlife (Visit 2).

Hypothesis. A subset of Visit 5 MCR-associated proteins and protein networks identified in ARIC will also be associated with MCR when measured within select available external cohorts (e.g., the LonGenity cohort).

Objective 4. External replication of LonGenity discovery analysis of MCR-associated proteins
Hypothesis. A subset of MCR-associated proteins identified in LonGenity will also be associated with MCR when measured within the ARIC cohort.

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).

Participants

Inclusion criteria:

- Participants who have SOMAscan protein measurements available from blood collected at Visit 5 (discovery analysis) and Visit 2 (for midlife replication analysis).
- Visit 5 4m walk and Subjective Memory Form data available at Visit 5

Exclusion Criteria:

- Non-white or non-black race
- Non-white participants in Washington County and Minnesota
- Missing proteomic data
- Missing covariate information
- Missing information needed to classify cognitive status (i.e., normal/MCI/dementia

classification) at Visit 5.

Exposure Variables

Proteomic measurement: Using plasma collected at Visit 2 (1990-92) and Visit 5 (2011-13), proteins were measured using a Slow Off-rate Modified Aptamer (SOMAmer)-based capture array (SomaLogic, Inc, Boulder, Colorado). Using chemically modified nucleotides, this process transforms protein signals to a nucleotide signal quantifiable using relative florescence on microarrays. Previous work indicates a median intra- and inter-run coefficient of variation of approximately 5% and intra-class correlation coefficients of ~0.9.[21–24] Individual protein levels (N=4,877 after QC) will be examined as the primary exposure. Proteins will be log2 transformed and outliers >5 SD away from the mean winsorized.

Proteins will also be grouped into networks/modules using the Bioconductor R package Netboost v1.0.0[25,26]. The first principal component score for the rank matrix of each protein module, referred to as the module eigenprotein (ME), will be calculated to quantify person-specific protein module expression. Module membership values will be used to define hub proteins for each module. The ME values will be examined as exposure variables and related to MCR in separate multivariable logistic regression models. Modules will be further characterized using Ingenuity Pathway Analysis to gain an improved understanding of the biological pathways and upstream regulators (e.g., transcription factors) associated with expression of MCR-relevant protein modules.

Primary outcome variables:

Motoric Cognitive Risk (MCR). MCR diagnosis is defined as (1) slow gait and (2) the presence of cognitive (memory) complaints, among non-demented (See *Dementia*) individuals[7].

1. *Walking speed* was measured at visit 5 as the time needed to walk 4 m at a usual pace. Slow walking speed will be defined as a time one standard deviation (within lowest 16th percentile) below the mean, stratified by age (+/-75 years) and sex, in the total V5 sample. We will conduct a sensitivity analysis using the cutoff defined in the Cardiovascular Health Study (CHS).
2. *Subjective Memory Complaints* were assessed at Visit 5 using the Subjective Memory Form (SMF). Participants will be categorized as having a subjective memory complaint if they respond “often” or “very often” to either of the following questions: 1) “In the past month, how often did you misplace or lose things around the house?” 2) In the past month, how often did you have trouble remembering conversations that occurred just a few days earlier?” A continuous subjective memory rating score will also be derived.

Mild Cognitive Impairment (MCI). MCI was defined as at least one domain score worse than -1.5 Z-score, a CDR sum of boxes between >0.5 and ≤3, an FAQ ≤5, and a decline on the serial ARIC cognitive battery below the 10th percentile on one test or below the 20th percentile on two tests[12]. MCR and MCI are not mutually exclusive; participants may meet criteria for neither, one, or both syndromes.

Cognitively Normal. Participants who did not meet criteria for MCI or MCR at ARIC visit 5, as defined above.

Dementia: Dementia will be defined at visit 5 based on the following criteria: at least two cognitive domain scores worse than 1.5 SD below the normative mean; a CDR sum of boxes greater than 3 or an FAQ >5; and a decline on the cognitive battery. For those who attended visit 6, date of dementia onset was assessed using information captured in annual follow-up interviews using the Six Item Screener (SIS) and the Ascertain Dementia 8-item Informant Questionnaire (AD8). For participants who did not attend visit 6, the SIS, AD8, hospital discharge codes, and death certificates were used to define dementia diagnosis and date of onset. Follow-up ended on December 31, 2017.

Other Variables

Visit 1 demographic variables, including race (black/white), sex (male/female), education (less than high school/high school, general education diploma [GED], or vocational school/college, graduate or professional school), *APOE* ε4 status, and center will be extracted. A combined-race-center variable will be created as follows: White-Washington County, White-Forsyth County, Black-Forsyth County, White-Minneapolis, or Black-Jackson. Additionally, participant age and laboratory and physiologic data, including systolic and diastolic blood pressures, total/high density lipoprotein cholesterol, body mass index (BMI, kg/m^2), and measures of kidney function will be extracted from the visit concurrent with plasma proteomic measurement (i.e., Visit 2 and Visit 5). Cardiovascular risk factors and disease information (i.e., diabetes, hypertension, coronary heart disease, and cigarette use) as well as medication information (i.e. treatment with antihypertensive drugs) will also be extracted from Visit 2 and Visit 5.

Data Analysis

Identification of plasma proteins associated with MCR. Multivariable logistic regression models will be used to examine the association between the relative level of each protein and the binary MCR variable. Analyses will be first adjusted for age at sample acquisition, sex, education, and race-center (model 1). Second, analyses will be adjusted for the aforementioned demographic variables as well as kidney function (model 2). Model 3 will adjust for model 2 covariates in addition to cardiovascular risk factors, including body mass index (BMI), hypertension, diabetes, and smoking status. FDR corrected $P<0.05$ will be used to identify candidate proteins. Sensitivity analyses will be conducted further adjusting for (a) prevalent stroke and (b) depression at the time of plasma collection.

Additionally, analyses will be repeated using Netboost-defined protein network expression as the exposure variable. Netboost will be used to identify networks of correlated proteins within the set of 4,877 proteins using the full set of participants with available SomaScan proteins at Visits 2 and 5. After modules are identified, module expression values (module eigenproteins [MEs]) will be calculated from the first principal component of each module for each participant. These values represent measures of network expression that can be related to participant traits and outcomes.

Comparison of plasma proteins and protein networks associated with MCR and MCI. The prevalence of MCR and MCI as well as the overlap of these syndromes will be characterized. Models 1-3, as outlined above, will be used to relate the full set of plasma proteins to MCI. Netboost will again be used to identify protein networks associated with MCI and compare these to networks associated with MCR.

Midlife Replication Analysis. Using the candidate proteins identified by the primary analyses (Visit 5) that pass FDR corrections, the same models outlined above will be replicated at midlife (Visit 2).

Ingenuity pathway analysis. To understand the regulatory mechanisms underlying the association between the candidate proteins and both MCR and MCI, we will use Ingenuity Pathway Analysis (IPA), a bioinformatics platform providing interpretations of omics data using the Ingenuity Knowledge Base (IPA, QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). We will use the top candidate proteins associated with (a) MCR and (b) MCI. The threshold for identifying *top* proteins will vary based on our results. We will aim to included 100-500 proteins in the *enrichment set*. We will perform IPA canonical pathway, upstream regulator, and mechanistic pathway analyses to construct biological interpretations of the pattern of candidate proteins identified to be associated with MCR and MCI.

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)?
MP 3327r Walker K et al. A proteomic analysis of incident dementia: The ARIC Study
MP 3574 Gomez G et al. The Association of Motoric Cognitive Risk with Neuroimaging Characteristics and Incident Dementia: The ARIC Study
MP 3808 Gomez et al. Proteomic analysis of cerebral white matter hyperintensities and small vessel disease

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes ____ No

11.b. If yes, is the proposal

__X_ A. primarily the result of an ancillary study (list number* _____)
ARIC Neurocognitive Study 2008.06
"Proteomic longitudinal ARIC study: SOMAscan of multiple visits"

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s) * 2013.10)

*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.
Understood

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.
Understood

13. Per Data Use Agreement Addendum, approved manuscripts using 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 pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __X__ No.
Page Break

References

1. Verghese, J., Wang, C., Lipton, R. B. & Holtzer, R. Motoric Cognitive Risk Syndrome and the Risk of Dementia. *J. Gerontol. A. Biol. Sci. Med. Sci.* **68**, 412–418 (2013).
2. Verghese, J. *et al.* Motoric cognitive risk syndrome. *Neurology* **83**, 718–726 (2014).
3. Verghese, J. *et al.* Motoric Cognitive Risk Syndrome and predictors of transition to dementia: Multi-center Study. *Alzheimers Dement. J. Alzheimers Assoc.* **15**, 870–877 (2019).
4. Verghese, J. *et al.* Motoric cognitive risk syndrome: multicountry prevalence and dementia risk. *Neurology* **83**, 718–726 (2014).
5. Gomez, G. T. *et al.* The association of motoric cognitive risk with incident dementia and neuroimaging characteristics: The Atherosclerosis Risk in Communities Study. *Alzheimers Dement. J. Alzheimers Assoc.* **18**, 434–444 (2022).
6. Beauchet, O., Allali, G., Annweiler, C. & Verghese, J. Association of Motoric Cognitive Risk Syndrome With Brain Volumes: Results From the GAIT Study. *J. Gerontol. A. Biol. Sci. Med. Sci.* **71**, 1081–1088 (2016).
7. Sekhon, H. *et al.* Motoric cognitive risk syndrome, incident cognitive impairment and morphological brain abnormalities: Systematic review and meta-analysis. *Maturitas* **123**, 45–54 (2019).
8. Wang, N. *et al.* Cerebral Small Vessel Disease and Motoric Cognitive Risk Syndrome: Results from the Kerala-Einstein Study. *J. Alzheimers Dis. JAD* **50**, 699–707 (2016).
9. Sathyan, S., Wang, T., Ayers, E. & Verghese, J. Genetic basis of motoric cognitive risk syndrome in the Health and Retirement Study. *Neurology* **92**, e1427–e1434 (2019).

10. Sathyan, S. *et al.* Association of anti-inflammatory cytokine IL10 polymorphisms with motoric cognitive risk syndrome in an Ashkenazi Jewish population. *Neurobiol. Aging* **58**, 238.e1-238.e8 (2017).
11. Ben-Avraham, D. *et al.* The complex genetics of gait speed: genome-wide meta-analysis approach. *Aging* **9**, 209–246 (2017).
12. Santanasto, A. J. *et al.* Identification of a Novel Locus for Gait Speed Decline With Aging: The Long Life Family Study. *J. Gerontol. A. Biol. Sci. Med. Sci.* **76**, e307–e313 (2021).
13. Sherva, R. *et al.* Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **10**, 45–52 (2014).
14. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
15. Hu, X. *et al.* Genome-wide association study identifies multiple novel loci associated with disease progression in subjects with mild cognitive impairment. *Transl. Psychiatry* **1**, e54 (2011).
16. Ganz, P. *et al.* Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA - J. Am. Med. Assoc.* **315**, 2532–2541 (2016).
17. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004 (2010).
18. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).

19. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).
20. Schrijvers, E. M. C. *et al.* Genome-wide association study of vascular dementia. *Stroke* **43**, 315–319 (2012).
21. Sattlecker, M. *et al.* Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimers Dement. J. Alzheimers Assoc.* **10**, 724–734 (2014).
22. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS One* **5**, e15004 (2010).
23. Kiddle, S. J. *et al.* Candidate blood proteome markers of Alzheimer's disease onset and progression: a systematic review and replication study. *J. Alzheimers Dis. JAD* **38**, 515–531 (2014).
24. Ganz, P. *et al.* Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. *JAMA* **315**, 2532–2541 (2016).
25. Schlosser, P. *et al.* Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nat. Genet.* **52**, 167–176 (2020).
26. Schlosser, P. *et al.* Netboost: Boosting-supported Network Analysis Improves High-Dimensional Omics Prediction in Acute Myeloid Leukemia and Huntington's Disease. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **PP**, (2020).

Appendix

SMF Coding

Dichotomous measure
If SMF1 is Often or Very Often (3 or 4) (losing items question)
Or
If SMF3 is Often or Very Often (3 or 4) (problems remembering conversations)
Then MemProb_SR = 1
Otherwise MemProb_SR = 0.
MemProb_SR is missing only if both SMF1 and SMF3 are missing