ARIC Manuscript Proposal # MP4104 (Amended)

PC Reviewed: 02/13/24	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Associations between plasma proteome, leukocyte telomere length, polygenic risk scores for leukocyte telomere length, and abdominal aortic aneurysm in ARIC

b. Abbreviated Title (Length 26 characters): Proteomics for LTL PRS and AAA

2. Writing Group:

Writing group members:

Aixin Li, Ingrid Jacobson, Brian Steffen, Jiaqi Xie, Nathan Pankratz, Dan Arking, Thomas Mosley, Weihua Guan, Sanaz Sedaghat, James S. Pankow, Pamela Lutsey, Weihong Tang

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

First author: **Aixin Li**
Address: 1300 South 2nd St, WBOB 300
Minneapolis, MN 55454
Phone: 3142035049 Fax:
E-mail: li002478@umn.edu

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: **Weihong Tang**
Address: 1300 South 2nd St, WBOB 300
Minneapolis, MN 55454

Phone: 612-626-9140 Fax:
E-mail: tang0097@umn.edu

3. Timeline: Data analysis begins immediately with an anticipated draft around Spring 2024

4. Rationale:

Telomeres are repeating DNA sequences located at the end of chromosomes that protect chromosomal regions from recombination, degradation, and the loss of coding DNA during

replication. [1] With each cell division, telomeres get shorter in their nucleotide repeats throughout an individual's lifespan. Leukocyte telomere length (LTL), due to the accessibility of leukocytes, is commonly used to study telomere length dynamics. [2] Numerous studies have associated LTL with aging, [3] age-related diseases, [4,5], and certain cancers. However, the underlying biological mechanisms of telomere length in age-related diseases are not fully elucidated. While a few studies have linked candidate plasma proteins with LTL, [6,7] to the best of our knowledge, no large-scale unbiased proteomics study has been conducted for LTL. To address this gap, our study conducted a large-scale proteomic analysis to characterize the proteomic signatures of LTL in the Atherosclerosis Risk in Communities (ARIC) cohort. To extend the comprehensive understanding of underlying mechanisms associated with LTL, we also estimated individuals' genetic predisposition for leukocyte telomere length by calculating polygenic risk scores (PRS) [8] based on published GWAS data. [9] A proteomics analysis will also be conducted for LTL PRS in the ARIC cohort to analyze the protein signals associated with genetically determined LTL.

We propose to derive a PRS for LTL using the published GWAS summary statistics [9] and ARIC GWAS data, followed by associating the LTL PRS with ~5,000 SOMAscan proteins measured in ARIC visits 2. In addition, we will conduct a prospective analysis to associate the LTL at visit 2 with the SOMAscan proteins measured at visit 3. We will compare the proteins associated with LTL PRS and those associated with the actual LTL measure, which may reveal new information on genetic vs non-genetic pathways involving LTL and the plasma proteome. Notably, as part of a funded R21 study (AS2020.25, PI: Weihong Tang), we are completing the estimate of LTL in ARIC based on the whole genome sequencing (WGS) data made available through NIH TOPMed[10] / CCDG[11] programs. The ARIC WGS were measured in visit 2 or 3 samples (>80% samples from visit 2). The analyses of LTL and LTL PRS with SOMAscan proteins proposed here are supported by the R21 study.

Abdominal aortic aneurysm (AAA), an important age-related disease and characterized by a localized enlargement of the abdominal aorta, [12] can lead to life-threatening rupture if left untreated. [13] While animal and Mendelian randomization studies suggest a causal role of LTL in AAA development, [14,15] the longitudinal relationship between LTL and AAA has not been well characterized in large prospective studies with long follow-up. Moreover, the underlying biological mechanisms mediating the association of LTL with AAA are poorly understood. Aiming to enhance our understanding of the relationship between LTL and AAA, we also propose a prospective cohort study between LTL and AAA in ARIC, followed by a mediation analysis to investigate LTL-related proteins that might mediate the pathway from LTL to AAA.

ARIC SomaLogic protein data is of high quality and excellent reliability using aptamer technologies. [16,17] Investigating LTL and LTL PRS with the ~5000 proteins by SOMAscan will provide an unprecedented opportunity for the discovery of proteomics signatures for LTL and that residing in the genetic pathways for LTL. This comprehensive research will not only benefit the understanding of the etiology of AAA but also the research on many other chronic diseases associated with LTL, including cardiovascular disease and cancers. [18-21]

5. Main Hypothesis/Study Questions:

Main hypothesis: We hypothesize that proteomic signatures are associated with LTL and genetically determined LTL as measured by LTL PRS, with significant overlap between LTL and genetically determined LTL. We also hypothesize that LTL is prospectively associated with

the incidence of AAA and that protein signals associated with LTL or LTL PRS mediate the pathway from LTL to AAA.

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).

Outcome: 1) The protein outcome variables will be each of the SOMAscan v 4.0 proteins (natural log2-transformed) measured at visits 2 and 3 after exclusion of proteins with large CV (>20%) or poor reproducibility between blind duplicate pairs. We will follow the ARIC analytic recommendations for the analysis of SOMAscan data. 2) Incident AAA has been ascertained in ARIC and was determined based on ICD codes for hospitalizations and deaths in ARIC up to 2019.

Exposure: 1) LTL PRS will be calculated based on the published GWAS summary statistics [9] and ARIC TOPMed genetic imputation data. 2) The estimation of phenotypic LTL will be based on TOPMed/CCDG whole genome sequencing data at ARIC visit 2 as part of the funded R21 study.

Mediators: Proteins associated with LTL or LTL PRS in the analysis of LTL with AAA.

Inclusion/Exclusion: We will conduct proteomics analyses among ARIC white and black participants. We will exclude those missing SOMAscan protein data at visits 2 or 3, LTL, or the corresponding genetic variants for calculation of PRS, failed quality control for SNPs, SOMAscan proteins, or LTL. In the causal mediation analysis, LTL will be measured at Visits 1, 2, or 3, with protein mediators will be assessed at Visit 3. The outcome will be the incidence of AAA cases after Visit 3.

1) Analysis of LTL PRS with proteins: LTL PRS will be computed as the sum of the number of LTL-increasing alleles, weighted by their measured effects from the published GWAS summary statistics.[9] We will conduct race-specific generalized linear regression of LTL PRS (as a continuous exposure variable) on the levels of ~5,000 human proteins (as the outcome) measured in ARIC visit 2 and visit 3, adjusting for age, sex, center, eGFR, and first ten principal components for genetic ancestry. We will exclude proteins with a large coefficient of variations (CVs) (>20%) or poor reproducibility between blind duplicate pairs and use appropriate statistical procedures (exclusion or winsorization) to handle extreme or unrealistic outliers and non-normal distribution. The results will be compared between white and black participants, with the analysis of white participants as a discovery and that of black participants as a replication. Statistical significance will be determined after Bonferroni correction for the number of proteins tested in each race group. We will treat the analysis of visit 3 proteins as the primary analysis (also for comparison with the LTL analysis) and that of visit 2 proteins as an internal replication.

2) Analysis of direct associations between LTL at visit 2 and proteins at visit 3: We will follow a similar approach to that described in 1) above, with additional adjustment for potential

confounders related to LTL and LTL measurement, such as smoking, and TOPMed/CCDG status for the source of WGS data.

3) Analysis of LTL with AAA: We will associate LTL with AAA using Cox proportional hazards regression. In the primary analysis, we will treat LTL as a continuous variable and adjust for the following potential confounders: age, gender, race, field center, BMI, diabetes, hypertension, smoking status, estimated glomerular filtration rate (eGFR), and the sample visit for LTL estimation. The covariates will be those measured at the visits corresponding to the WGS samples for LTL estimation. Additionally, we will analyze LTL in quartiles and conduct a trend test, in which the rank-ordered LTL quantile will be analyzed as a numerical variable (e.g. 1 to 4). This analysis will adjust for the same covariates as in the primary analysis.

4) Causal Mediation Analysis: To assess if the LTL- or LTL PRS-associated proteins mediate the association between LTL and AAA, we will conduct a mediation analysis. Given the limitations of standard approaches, such as the difference-of-coefficients or product-of-coefficients methods in linear models without interactions, [22] we will utilize a counterfactual framework for causal mediation. [23] This approach involves a clear causal structure, enabling the simultaneous estimation of causal estimates and allowing for the inclusion of interaction terms. [24] Mediation analyses will be performed based on a causal counterfactual framework with survival data under a regression-based approach to estimate direct effects (LTL and AAA association, not mediated by proteins), indirect effects (mediation due to proteins), and total effects (direct and indirect effects). [24] We will assess the mediating effect of each individual protein as well as a joint mediation effect by including all relevant proteins in one model.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes __x__ 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 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 current derived consent file ICTDER05 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/aricproposals/dtSearch.html

___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)?

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* _ AS2017.27_ AS2020.25(R21AG072530))

___ 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.

Yes.

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.

1. Herrmann M, Pusceddu I, Marz W, Herrmann W. Telomere biology and age-related diseases. *Clin Chem Lab Med*. Jul 26 2018;56(8):1210-1222. doi:10.1515/cclm-2017-0870
2. Factor-Litvak P, Susser E, Kezios K, et al. Leukocyte Telomere Length in Newborns: Implications for the Role of Telomeres in Human Disease. *Pediatrics*. Apr 2016;137(4)doi:10.1542/peds.2015-3927
3. Vaiserman A, Krasnienkov D. Telomere Length as a Marker of Biological Age: State-of-the-Art, Open Issues, and Future Perspectives. Review. *Frontiers in Genetics*. 2021-January-21 2021;11doi:10.3389/fgene.2020.630186
4. Weischer M, Bojesen SE, Cawthon RM, Freiberg JJ, Tybjærg-Hansen A, Nordestgaard BG. Short telomere length, myocardial infarction, ischemic heart disease, and early death. *Arterioscler Thromb Vasc Biol*. Mar 2012;32(3):822-9. doi:10.1161/atvbaha.111.237271
5. Fyhrquist F, Saijonmaa O, Strandberg T. The roles of senescence and telomere shortening in cardiovascular disease. *Nat Rev Cardiol*. May 2013;10(5):274-83. doi:10.1038/nrcardio.2013.30
6. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet*. Aug 2005;6(8):611-22. doi:10.1038/nrg1656

7. Aviv A, Anderson JJ, Shay JW. Mutations, Cancer and the Telomere Length Paradox. *Trends in Cancer*. 2017;3(4):253-258. doi:10.1016/j.trecan.2017.02.005
8. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*. 2020/09/01 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1
9. Codd V, Wang Q, Allara E, et al. Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet*. Oct 2021;53(10):1425-1433. doi:10.1038/s41588-021-00944-6
10. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Cold Spring Harbor Laboratory; 2019.
11. NIH NHGR. Centers for common disease genomics. https://www.genome.gov/Funded-Programs-Projects/NHGRI-Genome-Sequencing-Program/Centers-for-Common-Disease-Genomics
12. Walden R, Tomlinson B. Cardiovascular Disease. In: Benzie IFF, Wachtel-Galor S, eds. *Herbal Medicine: Biomolecular and Clinical Aspects*. CRC Press/Taylor & Francis Copyright © 2011 by Taylor and Francis Group, LLC.; 2011.
13. Sakalihasan N, Michel JB, Katsargyris A, et al. Abdominal aortic aneurysms. *Nat Rev Dis Primers*. Oct 18 2018;4(1):34. doi:10.1038/s41572-018-0030-7
14. Liao Q, He J, Tian FF, Bi FF, Huang K. A causal relationship between leukocyte telomere length and multiple sclerosis: A Mendelian randomization study. *Front Immunol*. 2022;13:922922. doi:10.3389/fimmu.2022.922922
15. Findeisen HM, Gizard F, Zhao Y, et al. Telomerase Deficiency in Bone Marrow–Derived Cells Attenuates Angiotensin II–Induced Abdominal Aortic Aneurysm Formation. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2011;31(2):253-260. doi:doi:10.1161/ATVBAHA.110.218545
16. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. Dec 7 2010;5(12):e15004. doi:10.1371/journal.pone.0015004 [doi]
17. Smith JG, Gerszten RE. Emerging Affinity-Based Proteomic Technologies for Large-Scale Plasma Profiling in Cardiovascular Disease. *Circulation*. Apr 25 2017;135(17):1651-1664. doi:CIRCULATIONAHA.116.025446 [pii]
10.1161/CIRCULATIONAHA.116.025446 [doi]
18. Van Eldik LJ, Carrillo MC, Cole PE, et al. The roles of inflammation and immune mechanisms in Alzheimer's disease. *Alzheimers Dement (N Y)*. Jun 2016;2(2):99-109. doi:10.1016/j.trci.2016.05.001
19. Kinney JW, Bemiller SM, Murtishaw AS, Leisgang AM, Salazar AM, Lamb BT. Inflammation as a central mechanism in Alzheimer's disease. *Alzheimers Dement (N Y)*. 2018;4:575-590. doi:10.1016/j.trci.2018.06.014
20. Li C, Stoma S, Lotta LA, et al. Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. *Am J Hum Genet*. Mar 5 2020;106(3):389-404. doi:10.1016/j.ajhg.2020.02.006
21. Telomeres Mendelian Randomization C, Haycock PC, Burgess S, et al. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol*. May 1 2017;3(5):636-651. doi:10.1001/jamaoncol.2016.5945

22. Lee H, Cashin AG, Lamb SE, et al. A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies: The AGReMA Statement. *JAMA*. 2021;326(11):1045-1056. doi:10.1001/jama.2021.14075
23. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *International journal of epidemiology*. 2013;42(5):1511-1519.
24. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass)*. 2011;22(4):582.