ARIC Manuscript Proposal # 3057

PC Reviewed: 10/3/17	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Repeatability and Longitudinal Variability of the Plasma Proteome

b. Abbreviated Title (Length 26 characters): **change in proteome**

2. Writing Group:
Adrienne Tin, Bing Yu, Kuni Matsushita, Ron C. Hoogeveen, Christie M. Ballantyne, David Couper, Alvaro Alonso, Tom Mosley, Gerardo Heiss, Jianzhong Ma, Peter Ganz, Liz Selvin, Eric Boerwinkle, Josef Coresh, and others welcome

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

F**irst author**: **Adrienne Tin, PhD**
Address: Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, W6021
Baltimore, MD 21205
Phone: 201-281-9577
atin1@jhu.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: **Josef Coresh, MD, PhD**
Address: Johns Hopkins Bloomberg School of Public Health
2024 E. Monument Street, Room 2-635
Suite 2-600
Baltimore, Maryland 21287

3. Timeline:
Data analysis will start immediately. A manuscript is expected to be prepared within 6 months.

4. Rationale:

Protein biomarkers in peripheral blood have been essential for diagnosis and prediction of disease in clinical practice (1). The protein biomarkers with known disease association or

currently used in clinical practice represent a small subset of the proteome. Studying the proteome can reveal biological pathways underlying diseases and discover novel protein biomarkers for diseases diagnosis and prognosis (2-4). Studies on the repeatability and longitudinal variability of the proteome are limited. Multiple technologies have proposed for proteomic profiling. The aptamer array has been shown to have wide dynamic range and good reproducibility (1). The ARIC study has conducted a pilot ancillary study using the SOMAscan assay, an aptamer platform, which provides relative quantification of 4,001 proteins (5). The pilot study included 42 individuals who participated in visits 2, 5 and the repeated visit 5. The samples from the repeated visit 5 were split to serve as blind duplicates for quality control purpose. The participants were selected based on stratified sampling in seven groups with six participants in each groups. The stratification factors were race, sex, and CKD defined as serum creatinine-based eGFR < 60 mL/min/1.73m^2 at visit 5. Proteomic profiling was performed using ethylenediaminetetraacetic acid (EDTA) preserved plasma.

5. Main Hypothesis/Study Questions:

The goals of this study are to assess the repeatability, and short (~six weeks) and long (~20 years) term variability of protein levels as measured by the SOMAscan assay.

Hypotheses:

1) Greater than 1000 proteins will have excellent reproducibility (split sample CV<10%). Furthermore, the majority of proteins that were also previously measured in ARIC will show good agreement (correlation >0.8).
2) The short and long term variability will be the dominant components of the overall variance for most proteins.

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).

Inclusion criteria: All participants in the repeatability and longitudinal variability study.

Outcomes: Protein levels measured by SOMAscan and passed the quality control check of SomaLogic. We will investigate the appropriate transformation of the relative florescence units (RFU) of protein levels, such as log and inverse normal transformation.

Predictor: ARIC ID, visit, and split sample status to be used for variance component estimate.

Other variables: age, sex, and CKD.

Data analysis:

Repeatability study. Previously reported blind duplicate analysis of the SOMAscan used the commercial platform, which included ~1,300 proteins.(3) The repeatability of many protein measures in the research platform in this pilot study is unknown. To assess the repeatability of the protein levels, we will conduct the following analyses:

A) Use of coefficient of variation, Spearman correlation, and principal component analysis of the split samples to assess repeatability. Since the underlying distribution of the protein levels can vary from quite symmetric to very skewed, the coefficient of variation (CV) of the split samples will be calculated by three methods: 1) mean of the pairwise CV (=SD/mean); 2) CVanova; and CV described by Bland and Altman (6, 7). We will also conduct similar analyses for short term and long term variability, which incorporate biologic changes. Our primary analysis will not exclude outliers but secondary analyses will quantify the proportion of outliers and their impact on the results.
B) Assess the normalization methods that are appropriate for longitudinal studies.

We will describe (graphically and using percentiles) the distribution of coefficients of variation and correlations. We will also categorized coefficients of variation at <10%, 10-20% and >20% and test the hypothesis that the scan will yield >1,000 proteins with a CV of <10% providing a wealth of reliable information.

Variability study. We will use mixed effect model with restricted maximum likelihood to estimate experiment-wise, the short and long term with-in-person variability for each protein using the following models.
Model 1 includes the split sample observations from the repeated visit 5 and will estimate the experiment-wise variability as random effect.
Model 2 includes the observations from visit 5 and the average of the pairwise observations from the repeated visit 5 and will estimate short variability as random effect.
Model 3 include the observations from visits 2 and 5 and will estimate the long term variability as random effect. We will evaluate the association between protein levels with age across the 4 measures adjusting for the sample selection stratifying factors (sex, race, and CKD) use generalized estimating equation (GEE) to estimate the population average association. The association can inform the extent of the variability relating to chronological age.

Significance threshold: We will account for the number of independent tests.

Limitations: The protein levels were quantified on a relative scale, and the sample size is limited.

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? _X__ 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)?

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* 2017.14**)**
__ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 2015.12_____

*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.

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

1. Smith JG, Gerszten RE. Emerging Affinity-Based Proteomic Technologies for Large-Scale Plasma Profiling in Cardiovascular Disease. Circulation. 2017 Apr 25;135(17):1651-64. PubMed PMID: 28438806. Pubmed Central PMCID: 5555416. Epub 2017/04/26. eng.
2. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, et al. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. JAMA. 2016 Jun 21;315(23):2532-41. PubMed PMID: 27327800. Epub 2016/06/22. eng.
3. Ngo D, Sinha S, Shen D, Kuhn EW, Keyes MJ, Shi X, et al. Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. Circulation. 2016 Jul 26;134(4):270-85. PubMed PMID: 27444932. Pubmed Central PMCID: 4963294. Epub 2016/07/23. eng.
4. Menni C, Kiddle SJ, Mangino M, Vinuela A, Psatha M, Steves C, et al. Circulating Proteomic Signatures of Chronological Age. J Gerontol A Biol Sci Med Sci. 2015 Jul;70(7):809-16. PubMed PMID: 25123647. Pubmed Central PMCID: 4469006. Epub 2014/08/16. eng.
5. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One. 2010 Dec 07;5(12):e15004. PubMed PMID: 21165148. Pubmed Central PMCID: 3000457. Epub 2010/12/18. eng.
6. Roraas T, Stove B, Petersen PH, Sandberg S. Biological Variation: The Effect of Different Distributions on Estimated Within-Person Variation and Reference Change Values. Clin Chem. 2016 May;62(5):725-36. PubMed PMID: 27001492. Epub 2016/03/24. eng.
7. Bland JM, Altman DG. Measurement error proportional to the mean. BMJ. 1996 Jul 13;313(7049):106. PubMed PMID: 8688716. Pubmed Central PMCID: 2351517. Epub 1996/07/13. eng.