

ARIC Manuscript Proposal #3835 (Amended)

PC Reviewed: 4/12/22
SC Reviewed: _____

Status: _____
Status: _____

Priority: 2
Priority: _____

1.a. Full Title: Validation of proteomic measurements across platforms in the ARIC Study

b. Abbreviated Title (Length 26 characters): V5 lab validation

2. Writing Group:

Writing group members: Mary Rooney, Jingsha Chen, Olive Tang, Christie Ballantyne, Ron Hoogeveen, Adrienne Tin, Kunihiro Matsushita, Junichi Ishigami, David Couper, Weihong Tang, Morgan Grams, Elizabeth Selvin, Josef Coresh (order TBD; others welcome – including representatives of the Olink ancillary study)

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

First author: Mary Rooney

Address: 2024 E. Monument St, Suite 2-600

Baltimore, MD 21286

E-mail: mroone12@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: **Elizabeth Selvin**

Address: 2024 E. Monument St, Suite 2-600

Baltimore, MD 21286

E-mail: eselvin@jhu.edu

3. Timeline: Data are available. MS 3835 was recently approved by ARIC P&P. As an extension of this work, we anticipate completion an additional manuscript within 1 year of the approval of this amended proposal.

4. Rationale:

The advent of multiplex technology to measure the thousands of proteins at a time has helped fuel a boom in proteomics research over the past decade. There are different approaches to measure the human proteome; two approaches include an aptamer-based approach and immunoassay approach. Soma and Olink both provide the relative concentrations of proteins, using different methods (Soma aptamer-based¹ and Olink immunoassay). The extent to which protein measurements correlate across platforms is relatively unknown and is an important consideration for researchers designing studies. Likewise, characterizing the comparability across different platforms for assessing proteomics are crucial for synthesizing findings across (or within) studies that utilize different platforms to measure the human proteome.

One of the larger validation studies was recently published by Raffield et al and reported correlations between immunoassays and SOMAScan based on data from multiple cohorts. Their findings indicated a wide range in the correlations across platforms for the proteins tested. For Soma vs Olink, for example, a prior comparison of Soma (1.1k version) vs Olink of 425 proteins in 48 myocardial infarction patients yielded a median Spearman correlation coefficient of 0.36 with Spearman's correlation coefficients ranging from -0.58 to 0.93. Moreover, 13% of the proteins had good correlation ($r \geq 0.7$) and 42% with poor correlation ($r < 0.3$).² However, most prior studies that have performed validation studies of proteomic platforms have been based in smaller sample sizes and/or using on older panels which measured fewer proteins.²⁻⁵ Pietzner et al published findings this year on 871 overlapping proteins measured on the aptamer-based (SomaScan Version 4) and immunoassay (Olink) using plasma from 485 participants. Their findings highlight many similarities in protein measurements across highly multiplexed proteomic platforms, however, substantial differences for certain proteins have also been found.

Using data from ARIC visit 5, we intend to assess 1) the comparability of Soma (N~5000 proteins) and Olink (N~460 proteins) untargeted platforms, and 2) the comparability of 14 targeted Soma protein measurements against standard immunoassays (58 cases [58 controls]). Additionally, using data from ARIC visits 2, 3, and 5, we will assess the comparability of Soma proteins versus standard immunoassays in the entire ARIC population (V2: N~14000; V3: N~13000; V5: N~6000).

5. Main Hypothesis/Study Questions:

Our aims are:

- 1) To assess the comparability of Olink's immunoassay method (untargeted) against SomaLogic's aptamer-based method (untargeted).
- 2) To assess the comparability of SomaLogic's aptamer-based assays (targeted) against standard immunoassay methods for 14 analytes in a case-control study.
 - a. GDF-15, ST2, Osteopontin, IL-6, MMP-1, TIMP-1, MMP-3, MMP-7, MCP-1, IL-10, VCAM-1, ICAM-1, IL-18 and TNF- α
- 3) To assess the comparability of SomaLogic's aptamer-based assays (targeted) against standard immunoassay methods used in the broader ARIC population.
 - b. Visit 2 immunoassays: ALT, albumin, cystatin C, B2M, TSH, FGF23, CRP, troponin-T, PTH, NTproBNP
 - c. Visit 3: cystatin C, fibrinogen (n~900), coagulation factor VII (n~900)
 - d. Visit 5: ALT, albumin, cystatin C, B2M, TSH, FGF23, CRP, troponin-T, troponin-I, NTproBNP

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).

Untargeted Soma vs Untargeted Olink (Aim 1)

Aim 1 Study Design

The SomaScan and Olink platforms were used to analyze previously stored plasma samples obtained at ARIC visit 5. We provide a brief description of the platforms below and the

participants included. We expect that >300 proteins will overlap between the Soma and Olink panels measured in ARIC.

Soma: The SomaScan Version 4 platform uses multiplexed modified DNA-based aptamer technology.¹ assay. Using previously stored plasma from the n>5000 ARIC visit 5 participants, the relative concentrations of ~5000 proteins were measured at Soma Logic.

Olink: The Olink platform (measured as part of ancillary study #2018.09) uses a multiplex immunoassay approach based on proximity extension assay (PEA) technology. Five Olink panels were performed in ARIC: CVD II, CVD III, inflammation, organ damage, and cardio-metabolic. For Olink, stored plasma samples for 500 participants (250 cases, 250 controls) were tested in as part of an ancillary study. The relative concentrations of 460 proteins were measured within each sample.

Eligible cases were those without prevalent heart failure at visit 5 but developed HF subsequently. Eligible controls were those without prevalent heart failure at visit 5 but did not develop heart failure over the same time period and matched based on cohort, age, and sex. Incident heart failure was defined using the FHS heart failure definition and/or as patients who were admitted to hospital at least once with a heart failure diagnosis without a prior diagnosis of heart failure.

Aim 1 Statistical Analysis

Olink and Soma protein values will both be presented on the log scale. We will report Spearman's correlation coefficients (with p-values) for each protein found in both Olink and Soma. We will summarize the correlations using histograms and using summary statistics (mean, median, percentiles). We will also generate scatterplots for each protein pair (Olink [y-axis] vs Soma [x-axis]) with linear regression and loess regression overlaid. We will also examine summary statistics for each $\log(\text{protein}_{\text{olink}})$ and $\log(\text{protein}_{\text{soma}})$ using mean, SD, and SD/mean. Additionally, after excluding outliers (>3 standardized residual), we will obtain Spearman's correlation and p-value for each protein pair and report the intercept, slope from linear regression. We will then graph a scatter plot of the correlation coefficients for each protein pair with vs without excluding outliers.

When the assays are poorly correlated, we will conduct analyses to try to determine which assay is more biologically relevant. We can look at differences in the correlation with kidney function (eGFR is related to approximately one third of plasma proteins). We can also look at disease (e.g. Heart Failure) associations as long as it is synergistic and does not interfere with the parent ancillary study.

Visit 5 Targeted Soma vs Immunoassays in 110 participants (Aim 2)

Aim 2 Study Design

We will use data from a pilot study conducted using stored plasma samples of ARIC visit 5 participants (who had 7 or 8 vials remaining). Participants with prevalent coronary heart disease (CHD), stroke, or heart failure at visit 5 were not included in this pilot study. Participants with incident CHD, stroke, or heart failure after visit 5 could be cases and controls included those without incident CHD, stroke, or heart failure, and did not die within 5 years of visit 5. The pilot study included 58 cases, 58 controls, and 26 Baylor QC pools. Cases were balanced by

age (< or ≥ median age of 73), sex (M/F), race (B/W) and eGFR (≥60 or <60). Controls were frequency matched to the age (+/- 10 years), sex, race and eGFR groupings of cases.

Aim 2 Statistical Analysis

We will assess the comparability of 14 analytes measured on Soma vs standard immunoassay methods: GDF-15, ST2, Osteopontin, IL-6, MMP-1, TIMP-1, MMP-3, MMP-7, MCP-1, IL-10, VCAM-1, ICAM-1, IL-18 and TNF-α (Baylor will provide information about assay conduct details). We will report Spearman’s correlation coefficients (and corresponding p-values). We will generate scatterplots for each of the 14 analyte pairs (Soma vs standard lab methods). We will summarize the correlations using histograms and using summary statistics (mean, median, percentiles). We will also generate scatterplots for each protein pair (Olink [y-axis] vs Soma [x-axis]) with linear regression and loess regression overlaid. We will also examine summary statistics for each log(analyte_{immunoassay}) and log(protein_{soma}) using mean, SD, and SD/mean. Additionally, after excluding outliers (>3 standardized residual), we will obtain Spearman’s correlation and p-value for each protein pair and report the intercept, slope from linear regression. We will then graph a scatter plot of the correlation coefficients for each protein pair with vs without excluding outliers.

Visits 2, 3, and 5 Targeted Soma vs Immunoassay in all ARIC participants (Aim 3)

Aim 3 Study Design

The SomaScan platform was used to analyze previously stored plasma samples obtained at ARIC visits 2 (N~14000), 3 (N~13000), and 5 (n~6000). We will assess the comparability of standard clinical immunoassays against SomaScan measured proteins across ARIC visits. Analytes that will be considered are summarized in the table below.

Visit	Standard immunoassays to compare to Soma
Visit 2	ALT, albumin, cystatin C, B2M, TSH, FGF23, CRP, troponin-T, PTH, NTproBNP
Visit 3	cystatin C, fibrinogen (n~900), coagulation factor VII (n~900), PTH, hs-TnT, NT-proBNP, albumin, FGF23
Visit 5	ALT, albumin, cystatin C, B2M, TSH, FGF23, CRP, troponin-T, troponin-I, NTproBNP, galectin-3, SHBG

Aim 3 Statistical Analysis

We will assess the comparability of analytes (see Table above) measured on Soma vs standard immunoassay methods. We will exclude SOMAmers that were flagged. We will report correlation coefficients (and corresponding p-values) and generate scatterplots for each of the analyte pairs (Soma vs standard lab methods).

Using Cox regression, we will evaluate associations of Soma proteins (modeled per 1 SD) and standard immunoassays (modeled per 1 SD) with all-cause mortality and with outcomes of well-established relevance to specific biomarkers (e.g. troponin-T with incident CHD, cystatin-C with incident CKD) through 2019. We will use the likelihood ratio tests to compare models that included biomarkers measured using both the immunoassay and SOMAscan measurement and compare to models with each biomarker method individually. We will use seemingly unrelated regression to compare the strength in association of the immunoassay vs SOMAscan measurements with outcomes.

We will explore factors associated with discordance across assays. We will obtain residuals from the linear regression of Y=standard immunoassay (interpretation of residual will be on the same scale as the standard immunoassay) on X=Soma. We will examine whether (A) the residual and (B) the absolute value of the residual are associated with mortality, and consider demographics, clinical characteristics (e.g. eGFR, diabetes, BMI), or certain relevant genotypes as potential factors related to the assay discordance.

For these analyses, we will use the newest data available from SomaLogic (adaptive normalization using maximum likelihood [ANML] that includes normalization to an external reference population [SMP]). We will also consider assessing the comparability of SomaLogic data without external reference normalization to the immunoassays.

Anticipated Limitations: The comparison of assays provides information about agreement but validity requires knowing the biologically relevant aspect of the protein. We know that immunoassays often disagree with each other so disagreement with an omic platform does not definitively imply which is superior. First, long-term storage of the plasma specimen is a potential concern due to possible protein degradation. However, we have previously demonstrated in ARIC the high reliability of protein markers over time using the Soma platform.⁷ Second, for both Olink and Soma, relative (not absolute concentrations) abundance of the proteins are examined and may not be on the same scale.

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? ___ Yes ___X___ 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 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.csc.unc.edu/aric/mantrack/maintain/search/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)?

- Tin (Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay)

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

11.b. If yes, is the proposal

A. primarily the result of an ancillary study

(list number* 2017.27 2018.09)

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.csc.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.csc.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. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS one*. 2010;5(12):e15004.
2. Raffield LM, Dang H, Pratte KA, et al. Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics*. 2020;20(12):e1900278.
3. Kukova LZ, Mansour SG, Coca SG, et al. Comparison of Urine and Plasma Biomarker Concentrations Measured by Aptamer-Based versus Immunoassay Methods in Cardiac Surgery Patients. *The Journal of Applied Laboratory Medicine*. 2019;4(3):331-342.
4. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558(7708):73-79.
5. Petrer A, von Toerne C, Behler J, et al. Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. *J Proteome Res*. 2021;20(1):751-762.
6. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nature Communications*. 2021;12(1):6822.
7. Tin A, Yu B, Ma J, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *J Appl Lab Med*. 2019;4(1):30-39.