ARIC Manuscript Proposal #4055

PC Reviewed: 5/17/22	**Status: _____**	**Priority: 2**
SC Reviewed: ___/___/___	**Status: _____**	**Priority: ____**

1.a. Full Title: The Proteomic Signatures of Physical Activity and their association with Type 2 Diabetes: The ARIC Study

b. Abbreviated Title (Length 41 characters): The proteomics of physical activity

2. Writing Group: Brian T. Steffen*, DJ McDonough*, James S. Pankow, Kelley Pettee Gabriel, Priya Palta, Weihong Tang, Mary R. Rooney, Mark A. Pereira. We welcome any further nominations.

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

*__Co-first authors__: Brian T. Steffen and DJ McDonough
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. 2nd St., Suite 300
Minneapolis, MN 55454
Phone: 612-626-0371
E-mail: steff293@umn.edu / mcdo0785@umn.edu

Senior author: Mark Pereira
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. 2nd St., Suite 300
Minneapolis, MN 55454
E-mail: map@umn.edu

3. Timeline: Begin analysis in May 2022
Draft of paper by July 2022

4. Rationale:

Physical activity (PA) is a well-studied lifestyle behavior associated with lower risks of adverse cardiometabolic outcomes including obesity and type 2 diabetes (T2D) [1, 2]. PA triggers dozens of biological phenomena, and both intervention studies and experimental models have identified some of the likely mediators of the metabolic health benefits of PA. Among these, increased expression and translocation of the glucose transporter GLUT4 [3], increased activity levels of multiple antioxidant enzymes [4, 5], lower levels of inflammatory cytokines associated with insulin resistance [6, 7], and lower leptin levels [7, 8] have been shown to be associated with, or directly result from, PA. These, in turn, are known to contribute to greater insulin sensitivity, lower oxidative burden, and suppression of adipogenesis [9-12]. Yet, our current understanding of PA and its benefits likely remains in its developmental stages, and many of the pathways and protein architecture through which PA affects metabolic health remain unknown.

To date, proteomics-based analyses have identified numerous proteins associated with PA. Higher antioxidant enzyme levels as well as lower inflammatory cytokines and leptin have been confirmed [13, 14], and novel associations have also been identified. Cross-sectional analyses have shown that PA levels are related to dozens of proteins including creatine kinases, endothelial adhesion molecules, tumor necrosis factor and receptors, insulin growth factor and fatty acid binding proteins [13]. Further, an intervention trial of 650 sedentary individuals showed that PA resulted in differential protein expression of 100 targets, 21 of which were found to be associated with all-cause mortality [15]. Finally, biological pathways related to PA have also been revealed by proteomics studies. Compared to inactive individuals, those who engage in aerobic PA were found to have plasma proteomic signatures corresponding to pathways involved in wound healing, apoptosis regulation, glucose-insulin signaling, and cellular stress signaling [16]. Collectively, these studies have identified numerous protein and pathway targets. However, most were limited in sample size and/or numbers of measured proteins, and it remains unclear which proteins may serve to protect against metabolic dysfunction and T2D.

The present proposal aims to extend the above proteomics findings in the ARIC cohort, which has over 4600 unique protein measurements in over 11,000 participants. Using a data-driven approach, we propose to identify proteomic signatures associated with PA—which is expected to confirm the above mechanisms and also elucidate additional plasma proteins related to PA. Identified proteins will be subjected to pathway analysis and also be tested for their associations with the risk of T2D. Proteins associated with T2D will be further interrogated using Mendelian randomization analysis to determine whether they may have causal roles in T2D development.

5. Main Hypothesis/Study Questions:

Can we identify unique plasma proteomic signatures associated with PA? Are any of these proteins causally related to lower risk of T2D?

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).

Design: cross-sectional. Proteomic data were measured in plasma samples using the modified aptamer-based platform SOMAscan v 4.0. This platform measured levels of 4,870 proteins from Visit 3 plasma samples. The substantial protein coverage of SOMAscan v 4.0 in over 11,000 ARIC participants with available data provide the opportunity to identify proteomic signatures associated with PA. Visit 3 was selected since it is the first visit with PA self-report variables. Visit 5 data will be used in replication analysis to confirm proteins identified from Visit 3. Following the primary analysis, secondary analyses will be conducted in which identified proteins will be interrogated for their: (1) cross-sectional associations with T2D at Visit 3; and (2) prospective associations with T2D using Visit 3 as the baseline. Finally, for proteins found to be related to T2D, Mendelian randomization analysis will be conducted to determine whether these proteins may be causally related to T2D development.

Inclusion/Exclusion: We will exclude participants who do not have SOMAscan data, failed proteomic quality control, self-reported PA, or estimated glomerular filtration rate (eGFR) data at Visit 3. In the prospective secondary analysis, those with T2D at baseline will be excluded.

Exposure: The sport and exercise index of PA will serve as the exposure variable. Change in PA was initially considered as an exposure of interest; however, the measurement errors would likely be compounded using this approach and skew results toward the null, presuming a non-differential classification bias.

Outcomes: Natural log2-transformed SomaScan protein levels will serve as the primary outcome variables, and ARIC analytic recommendations for the SomaScan data will be followed. The proteins were assessed using a Slow Off-rate Modified Aptamer (SOMAmer)-based capture array (SomaLogic, Inc, Boulder, Colorado). Non-human proteins and proteins with unacceptable QC will be removed [e.g., have large CVs (e.g. > 20%), poor reproducibility between the blind duplicate pairs, or show non-specific binding].

In a secondary analysis, we will examine whether proteins identified in the primary analysis are associated with T2D as defined by physician diagnosis, glucose-lowering medication use, an HbA_{1c} level ≥6.5%, or a fasting glucose level ≥126 mg/dL.

Data analysis: In the primary analysis, multiple linear regression will be conducted. PA will serve as the independent variable and protein levels measured at Visit 3 will serve as the dependent variables. Statistical adjustments will be made for age at Visit 3, sex, race-field center,

education status, smoking status, self-reported alcohol intake, and eGFR. PA will be examined as both a continuous and categorical variable (i.e., top vs bottom decile). Bonferroni correction for multiple comparisons will be applied—4,955 proteins stipulate a significance threshold of $p<1.0x10^{-5}$. As a replication step, proteins associated with PA at Visit 3 will be examined at Visit 5.

Secondary analyses will examine whether proteins identified and replicated in the primary analysis are associated with risk of T2D. Logistic regression will be conducted to estimate odds of T2D with levels of identified proteins at Visit 3. Relative risk regression will additionally be conducted in which identified proteins will serve as the exposures and incident T2D over subsequent follow up years will serve as the outcome. Statistical adjustments will be made for age, sex, field center-race, cigarette smoking status, and estimated glomerular filtration rate at Visit 3. BMI is a likely mediator of the association between PA and T2D and will be included in a separate model to examine this possibility. In addition, a stratified analysis will be conducted across strata of BMI category to examine whether it may modify PA-protein associations.

To aid in the interpretation of the results of the primary analysis, we will use Ingenuity Pathway Analysis to identify the most relevant signaling and metabolic pathways, molecular networks, and biological functions associated with PA-related proteins.

Finally, to examine whether the proteins may be causally related to T2D, two-sample Mendelian randomization will be conducted using the MR-base platform. Instrument variables or protein quantitative trait loci will be identified by performing genome-wide association analyses for proteins found to be related to prevalent and incident T2D with adjustment for age, sex, field center, and ten principal components of ancestry. Summary statistics for these instrument variables will be uploaded to MR-base. The T2D outcome summary statistics for instrument variables will be provided by a previous GWAS [17], available through the MR-base platform. Where only single instrument variables are available for protein exposures, Wald ratios will be generated. Where multiple instrument variables are available, random effects inverse variance weighted estimates will be generated and MR-Egger will test for horizontal pleiotropy as a sensitivity analysis.

Limitations: Self-reported PA is well-documented for measurement error. [18]. For the primary outcome, aptamer-based protein measurements have been found to have inconsistent correlations with their corresponding protein when levels are determined by mass-spectrometry or immuno-based assay methods. Caution will be exercised in interpreting results; where available, aptamer-based measurements will be compared to other assays to evaluate their accuracy.

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? ____ 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 file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? ____ Yes __x__ 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:

___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)?

No investigators have so far proposed to examine proteomic signatures associated with PA. However, proposed research by Dr. Mary Rooney is examining the proteomics of T2D (#3803), and there will likely be some overlap in our findings. Dr. Rooney is a co-author here, and we will collectively determine which proteomic findings are unique to PA and incident diabetes for publication.

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___)

___ 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 at https://www2.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

1. Aune D, Norat T, Leitzmann M, et al. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol. 2015;30(7):529-42. Epub 20150620. doi: 10.1007/s10654-015-0056-z. PubMed PMID: 26092138.
2. Brellenthin AG, Lee DC, Bennie JA, et al. Resistance exercise, alone and in combination with aerobic exercise, and obesity in Dallas, Texas, US: A prospective cohort study. PLoS Med. 2021;18(6):e1003687. Epub 20210623. doi: 10.1371/journal.pmed.1003687. PubMed PMID: 34161329; PMCID: PMC8266085.
3. Hansen PA, Nolte LA, Chen MM, et al. Increased GLUT-4 translocation mediates enhanced insulin sensitivity of muscle glucose transport after exercise. J Appl Physiol (1985). 1998;85(4):1218-22. doi: 10.1152/jappl.1998.85.4.1218. PubMed PMID: 9760308.
4. Berzosa C, Cebrián I, Fuentes-Broto L, et al. Acute exercise increases plasma total antioxidant status and antioxidant enzyme activities in untrained men. J Biomed Biotechnol. 2011;2011:540458. Epub 20110309. doi: 10.1155/2011/540458. PubMed PMID: 21436993; PMCID: PMC3062968.
5. Wiecek M, Szymura J, Maciejczyk M, et al. Anaerobic Exercise-Induced Activation of Antioxidant Enzymes in the Blood of Women and Men. Front Physiol. 2018;9:1006. Epub 20180727. doi: 10.3389/fphys.2018.01006. PubMed PMID: 30140236; PMCID: PMC6094974.
6. Enarsson M, Feldreich T, Byberg L, et al. Association between Cardiorespiratory Fitness and Circulating Proteins in 50-Year-Old Swedish Men and Women: a Cross-Sectional Study. Sports Med Open. 2021;7(1):52. Epub 20210726. doi: 10.1186/s40798-021-00343-5. PubMed PMID: 34312731; PMCID: PMC8313632.
7. Sirico F, Bianco A, D'Alicandro G, et al. Effects of Physical Exercise on Adiponectin, Leptin, and Inflammatory Markers in Childhood Obesity: Systematic Review and Meta-Analysis. Child Obes. 2018;14(4):207-17. Epub 20180515. doi: 10.1089/chi.2017.0269. PubMed PMID: 29762052; PMCID: PMC5994661.
8. Fedewa MV, Hathaway ED, Ward-Ritacco CL, et al. The Effect of Chronic Exercise Training on Leptin: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Sports Med. 2018;48(6):1437-50. doi: 10.1007/s40279-018-0897-1. PubMed PMID: 29582381.
9. Ryan BJ, Schleh MW, Ahn C, et al. Moderate-Intensity Exercise and High-Intensity Interval Training Affect Insulin Sensitivity Similarly in Obese Adults. J Clin Endocrinol Metab. 2020;105(8):e2941-59. doi: 10.1210/clinem/dgaa345. PubMed PMID: 32492705; PMCID: PMC7347288.
10. Teixeira-Lemos E, Nunes S, Teixeira F, et al. Regular physical exercise training assists in preventing type 2 diabetes development: focus on its antioxidant and anti-inflammatory properties. Cardiovasc Diabetol. 2011;10:12. Epub 20110128. doi: 10.1186/1475-2840-10-12. PubMed PMID: 21276212; PMCID: PMC3041659.
11. Freeman A, Cellura D, Minnion M, et al. Exercise Training Induces a Shift in Extracellular Redox Status with Alterations in the Pulmonary and Systemic Redox Landscape in Asthma. Antioxidants (Basel). 2021;10(12). Epub 20211130. doi: 10.3390/antiox10121926. PubMed PMID: 34943027; PMCID: PMC8750917.
12. Allerton TD, Savoie JJ, Fitch MD, et al. Exercise reduced the formation of new adipocytes in the adipose tissue of mice in vivo. PLoS One. 2021;16(1):e0244804. Epub 20210120. doi: 10.1371/journal.pone.0244804. PubMed PMID: 33471817; PMCID: PMC7817033.
13. Stattin K, Lind L, Elmståhl S, et al. Physical activity is associated with a large number of cardiovascular-specific proteins: Cross-sectional analyses in two independent cohorts. Eur J Prev Cardiol. 2019;26(17):1865-73. Epub 20190814. doi: 10.1177/2047487319868033. PubMed PMID: 31409108.
14. Corlin L, Liu C, Lin H, et al. Proteomic Signatures of Lifestyle Risk Factors for Cardiovascular Disease: A Cross-Sectional Analysis of the Plasma Proteome in the Framingham Heart Study. J Am Heart Assoc. 2021;10(1):e018020. Epub 20201229. doi: 10.1161/jaha.120.018020. PubMed PMID: 33372532; PMCID: PMC7955453.
15. Robbins JM, Peterson B, Schranner D, et al. Human plasma proteomic profiles indicative of cardiorespiratory fitness. Nat Metab. 2021;3(6):786-97. Epub 20210527. doi: 10.1038/s42255-021-00400-z. PubMed PMID: 34045743.
16. Santos-Parker JR, Santos-Parker KS, McQueen MB, et al. Habitual aerobic exercise and circulating proteomic patterns in healthy adults: relation to indicators of healthspan. J Appl Physiol (1985). 2018;125(5):1646-59. Epub 20180920. doi: 10.1152/japplphysiol.00458.2018. PubMed PMID: 30236049; PMCID: PMC6295489.
17. Xue A, Wu Y, Zhu Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018;9(1):2941. Epub 20180727. doi: 10.1038/s41467-018-04951-w. PubMed PMID: 30054458; PMCID: PMC6063971.

18. Lichtman, S. W., Pisarska, K., Berman, E. R., Pestone, M., Dowling, H., Offenbacher, E., ... & Heymsfield, S. B. (1992). Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. New England Journal of Medicine, 327(27), 1893-1898.