ARIC Manuscript Proposal Form

ARIC Publication Admin Use Only

Publication Committee Review Date: **06/11/24**
ARIC Manuscript Proposal Number: **# 4474**

1.a. Full Title: The plasma proteome of whole grain intake and dietary fiber intake: results from the Cardiovascular Health Study and Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters):proteomics of whole grain and fiber

2. Writing Group [please provide a middle initial if available; EX: Adam L Williams]:
Writing group members: Hyunju Kim, Jennifer Brody, Tom Austin, Kerri Wiggins, Rozenn N. Lemaitre, Amanda Fretts, Denise C. Hasson, Nona Sotoodehnia, James Floyd, Casey M. Rebholz, Bruce M. Psaty

Others welcome

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

First author [please provide a middle initial; EX: Adam L Williams]:
Hyunju | | **Kim**
First name | Middle initial | Last name

Address: 3980 15th Avenue NE
Rm 886, Seattle WA 98105

Phone: 206-685-7902
E-mail: hyunjuk1@uw.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (The ARIC author should be involved enough in ARIC to be able to point the lead author to appropriate ancillary study PIs and to be able to search ARIC manuscript proposals if the lead author doesn't have the access needed to do such a search).

Name: Dr. Casey M. Rebholz
Address: 2024 E. Monument St, Suite 2-500
Baltimore, Maryland 21287

Phone: 410-502-2539
E-mail: crebhol1@jhu.edu

3. Timeline: Analyses will begin when the proposal is approved. We anticipate that a first draft of the manuscript will be available within approximately one year of manuscript proposal approval.

4. Rationale:

The American Heart Association recommends whole grain intake instead of refined grains, and diets rich in fiber to improve cardiovascular health (1). Greater intake of whole grains and dietary fiber is consistently associated with a lower risk of cardiovascular disease (CVD). In a meta-analysis of 45 studies, per 90g higher intake of whole grain was associated with a 19%, 12%, and 22% lower risk of coronary heart disease, stroke, and cardiovascular disease (CVD) (2). Total dietary fiber intake was inversely associated with a lower risk of cardiovascular disease in several meta-analyses (3–5). In the Cardiovascular Health Study (CHS), a community-based cohort of older adults in the US, each 5g higher in total dietary fiber and cereal fiber was inversely associated with incident CVD (6). Additionally, total dietary fiber and cereal fiber intake was negatively correlated with circulating levels of inflammatory biomarkers (C-reactive protein, interleukin-1 receptor antagonist), suggesting that inflammation is an important mechanism through which dietary fiber is associated with a lower risk of incident CVD (6). There may be diverse pathways that help explain the associations between whole grains, dietary fibers, and incident CVD. However, our understanding of the underlying pathways remains incomplete. In addition, dietary biomarkers, which are useful for accurately assessing the intake of whole grain and fiber, are limited.

Large-scale proteomics can be helpful for identifying objective biomarkers of whole grain and dietary fiber, and can offer important insights through which whole grain intake and fiber intake is associated with a reduction in CVD risk. Plasma proteins play a role in many biological processes, by carrying nutrients, acting as an enzyme to metabolize nutrients, mediating cellular responses, and activating a cascade of reactions to support immune responses (7). These functions are related to metabolism, absorption, and digestion of foods. Further, previous studies reported that nutrients that are abundant in whole grains (vitamin A, E, selenium) are known to be strongly associated with circulating levels of their carrier proteins (e.g., retinol and retinol binding protein 4; alpha-tocopherol and apolipoprotein C III; selenium and selenoprotein P isoform 1, respectively) (8,9).

The present study aims to build upon the previous study in CHS which focused on 8 inflammatory proteins, measured using immunoassays (6). Our objectives are to 1) identify plasma protein biomarkers of whole grain and dietary fiber intake in CHS, 2) prospectively link whole grain- and fiber-related proteins to CVD risk in CHS, and 3) replicate these findings in the ARIC Study.

5. Main Hypothesis/Study Aims:

Hypothesis 1: Several plasma proteins (such as inflammatory biomarkers) associated with a) whole grain intake and b) dietary fiber intake in CHS will replicate in the ARIC Study

Hypothesis 2: Several whole grain- and dietary fiber- related proteins prospectively associated with incident CVD will replicate in the ARIC Study

6. Design and analysis - please address the following aspects:

a) Inclusion/exclusion: For both aims, ARIC visit 3 participants with data on whole grain intake, fiber intake, and large-scale proteomics (approximately 5,000 proteins) will be eligible. For aim 1, we will exclude participants who are missing proteomics data, missing whole grain intake or dietary fiber intake data, those with implausible dietary intake (men <700 or 4500 kcal and women <500 or >3500 kcal), those who are missing more than 10 items in the food frequency questionnaire (FFQ), and missing covariates (sociodemographic characteristics, lifestyle factors, and clinical characteristics described below). For aim 2, in addition to the exclusion criteria outlined in aim 1, we will exclude participants with prevalent CVD before or at visit 3, or those who are missing follow-up time for incident CVD.

b) Study design: Cross-sectional analysis (aim 1) of whole grain intake, fiber intake, and large-scale proteomics (SomaLogic), which are available at ARIC study visit 3 (1993-1995) and prospective analysis (aim 2) of diet-related proteins and incident CVD through the latest follow-up period.

c) Outcome and other variables of interest with specific reference to the time of their collection:

Aim 1:
Exposure: Whole grain intake, derived from the FFQ, and dietary fiber intake, calculated from the FFQ. Whole grain intake in the ARIC study will be defined as cooked cereal such as oatmeal, grits, cream of wheat, dark or whole grain bread. Whole grain intake and dietary fiber intake will be energy-adjusted using the residual approach (10).

Outcome: Large-scale proteomics. Given that the ARIC study is used as an external replication population, we will restrict the proteins to the ones that were significantly associated with whole grain intake or fiber intake in CHS at false discovery rate (FDR) <0.05.

Aim 2:
Exposure: Proteins significantly associated with whole grain intake or dietary fiber intake in CHS at FDR <0.05.

Outcome: Incident CVD. Incident CVD will be defined as a composite of fatal or nonfatal coronary heart disease, definite or probable myocardial infarction, definite or probable stroke, and heart failure in the ARIC Study. Incident CVD was ascertained via

annual telephone calls, active surveillance of local hospital discharge lists, health department death certificate files, and linkage to the National Death Index. We will repeat the analysis by CVD subtypes (coronary heart disease, stroke, and heart failure).

Other variables of interest: Covariates include sociodemographic characteristics (age, sex, race-center, education), lifestyle factors (physical activity, smoking, alcohol intake), body mass index, estimated glomerular filtration rate (Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) 2021 race-free equation), and total energy intake (11). To identify whole grain-specific and dietary fiber-specific biomarkers, we will consider additionally adjusting for food groups that may be correlated with whole grain intake and fiber intake, such as % of energy from protein, % of energy from saturated fat, and ratio of polyunsaturated fatty acids and saturated fatty acids.

d) Summary of data analysis:

In the ARIC Study, for aim 1, four multivariable linear regression models which sequentially adjust for important confounders will be conducted, using whole grain intake as the independent variable and $\log_2$transformed proteins as the response variable. Model 1 will adjust for age (continuous), sex (men, women), race-center, and total energy intake (continuous). Model 2 will adjust for covariates in model 1 plus education (<12 years, 12 years, >12 years), physical activity (continuous), smoking status (never, former, and current smoker), and alcohol intake (continuous). Model 3 will adjust for covariates in model 2 and body mass index (continuous), eGFR (continuous), hypertension (yes, no), and diabetes (yes, no). Model 4 will adjust for covariates in model 3 and several dietary factors (% of energy from protein, % of energy from saturated fat, and ratio of polyunsaturated fatty acids and saturated fatty acids). We will repeat the analyses with dietary fiber intake as the independent variable).

For aim 2, the same set of models will be adjusted as for aim 1. We will use Cox proportional hazard regression models, using proteins as the independent variable and incident CVD as the response variable. Days of follow-up time will be used as the time metric.

In the ARIC Study, we will consider proteins to be replicated if they have the same direction of associations as the CHS, and if FDR is <0.05.

e) Any anticipated methodologic limitations or challenges if present

This study proposes to replicate the associations between whole grain intake and dietary fiber intake and a large number of proteins from the CHS. Therefore, the proposed study carries high multiple testing burden. It is possible that no diet-related proteins are associated with incident CVD.

f) Will the author need Limited data to complete the proposed manuscript?

☒ **Yes, Limited data is needed (Provide a brief (2-3 sentences) justification for requesting PHI data)** . ☐ **No, De-identified data will be sufficient.***Please note, Limited dataset access is strict and rarely provided. Limited data includes identifiable information such as dates (birthdays, visit dates, etc.). CMS, Genomic, Geocoded, Proteomics/Somalogic, and other -omic data all fall under the limited data category. De-identified data does not include dates. All dates are date adjusted to "Days since Visit 1".

This proposal aims to replicate the associations between dietary exposures (whole grain intake, dietary fiber intake), and large-scale proteomics in CHS. Proteomics/Somalogic data sets are needed for the replication analyses.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? (Non-ARIC analysis means that the authors are not regarded as ARIC investigators and the "ARIC author" is essentially just a facilitator rather than an integral part of the writing group.) ☐ **Yes** ☒ **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 is distributed to ARIC PIs annually, 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** ☒ **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 or the "sponsoring" ARIC 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 website at: https://aric.cscc.unc.edu/aric9/proposalsearch [ARIC Website□Publications□Proposal Search]

☒ 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)?

MS #4196 Plasma proteins associated with four different types of plant-based diets

This manuscript proposal aimed to identify proteomic signatures of plant-based diets and incident CVD. However, our proposal is focused on specific dietary intake (whole grain intake and dietary fiber intake). The lead author of this manuscript proposal is the same as the lead author of the present manuscript proposal.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or does it use current [or ongoing] ancillary study data (this includes ACHIEVE)? ☒ Yes ☐ No → Skip to question 12

11.b. If yes to 11.a., is the proposal

☒ **A. primarily the result of an ancillary study**

☐ **B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables)**

11.c. If yes to 11.a., list number*

2019.26 – Discovery, Replication, and Validation of Biomarkers of the DASH Diet and Hypertension (PI: Rebholz)

2017.27 – Proteomic longitudinal ARIC study: SOMAscan of multiple visits (PI: Coresh))

*ancillary studies are listed by number
https://aric.cscc.unc.edu/aric9/researchers/ancillary_studies/approved_ancillary_studies [ARIC Website□Ancillary Studies□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 https://aric.cscc.unc.edu/aric9/publications/policies_forms_and_guidelines [ARIC Website□Publications□Publication Policies, Forms, and Guidelines].
http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References:

1. Lichtenstein AH, Appel LJ, Vadiveloo M, Hu FB, Kris-Etherton PM, Rebholz CM, Sacks FM, Thorndike AN, Van Horn L, Wylie-Rosett J, et al. 2021 Dietary Guidance to Improve Cardiovascular Health: A Scientific Statement From the American Heart Association. Circulation American Heart Association; 2021;144:e472–87.

2. Aune D, Keum N, Giovannucci E, Fadnes LT, Boffetta P, Greenwood DC, Tonstad S, Vatten LJ, Riboli E, Norat T. Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. BMJ 2016;353:i2716.

3. Threapleton DE, Greenwood DC, Evans CEL, Cleghorn CL, Nykjaer C, Woodhead C, Cade JE, Gale CP, Burley VJ. Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis. BMJ British Medical Journal Publishing Group; 2013;347:f6879.

4. Veronese N, Solmi M, Caruso MG, Giannelli G, Osella AR, Evangelou E, Maggi S, Fontana L, Stubbs B, Tzoulaki I. Dietary fiber and health outcomes: an umbrella review of systematic reviews and meta-analyses. Am J Clin Nutr 2018;107:436–44.

5. Ramezani F, Pourghazi F, Eslami M, Gholami M, Mohammadian Khonsari N, Ejtahed H-S, Larijani B, Qorbani M. Dietary fiber intake and all-cause and cause-specific mortality: An updated systematic review and meta-analysis of prospective cohort studies. Clin Nutr 2024;43:65–83.

6. Shivakoti R, Biggs ML, Djoussé L, Durda PJ, Kizer JR, Psaty B, Reiner AP, Tracy RP, Siscovick D, Mukamal KJ. Intake and Sources of Dietary Fiber, Inflammation, and Cardiovascular Disease in Older US Adults. JAMA Netw Open 2022;5:e225012.

7. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002;1:845–67.

8. Cole RN, Ruczinski I, Schulze K, Christian P, Herbrich S, Wu L, Devine LR, O'Meally RN, Shrestha S, Boronina TN, et al. The plasma proteome identifies expected and novel proteins correlated with micronutrient status in undernourished Nepalese children. J Nutr 2013;143:1540–8.

9. West KP, Cole RN, Shrestha S, Schulze KJ, Lee SE, Betz J, Nonyane BA, Wu LS-F, Yager JD, Groopman JD, et al. A plasma α-tocopherome can be identified from proteins associated with vitamin E status in school-aged children of Nepal. J Nutr Oxford Academic; 2015;145:2646–56.

10. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 1997;65:1220S-1228S.

11. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, Crews DC, Doria A, Estrella MM, Froissart M, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. N Engl J Med 2021;385:1737–49.

To view publications materials, click "Log in" at the top right of the ARIC website. Click "Forgot Password" if you are experiencing issues with logging in.