ARIC Manuscript Proposal #4196

PC Reviewed: 02/14/23 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title:

Protein signatures of plant-based diets and incident cardiovascular disease

b. Abbreviated Title (Length 26 characters): Proteomics of plant-based diets

2. Writing Group [please provide a middle name if available; EX: Adam Lee Williams]:
Writing group members:

Hyunju Kim, PhD, MPH
Maura E. Walker, PhD
Alexis C. Wood, PhD
Morgan E. Grams, MD, PhD, MHS
Bing Yu, PhD, MS
Eric Boerwinkle, PhD
Vasan S. Ramachandran, MD, PhD
Peter Ganz, MD
Wendy S. Post, MD, MS
Josef Coresh, MD, PhD, MHS
Casey M. Rebholz, PhD, MS, MPH, MNSP

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

First author [please provide a middle name if available; EX: Adam Lee Williams]:
Address:

Hyunju Kim
2024 E Monument St, Suite 2-500
Baltimore MD 21287

Phone: Fax:
E-mail: hkim25@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: Casey Rebholz
Address: 2024 E. Monument St, Suite 2-500
Baltimore, Maryland 21287

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

3. Timeline:

Analyses will begin in April 2023. We anticipate that a first draft of the manuscript will be available within approximately one year of manuscript proposal approval.

4. Rationale:

Plant-based diets, which are comprised predominantly of plant foods and are low in animal products, are associated with a lower risk of cardiovascular disease (CVD).[1,2] In a community-based cohort (ARIC Study), we used four published plant-based diet indices to investigate the associations with CVD risk factors (hypertension, chronic kidney disease (CKD)), and incident CVD. Greater adherence to an overall plant-based diet and a vegetarian diet was associated with a lower risk of these clinical outcomes.[2–4]Healthful plant-based diets (high in healthful plant foods; low in unhealthful plant foods and animal products) were associated with a lower risk of CVD mortality, hypertension, and CKD.[2–4] In contrast, unhealthful plant-based diets (high in unhealthful plant foods; low in healthful plant foods and animal products) were associated with an elevated risk of hypertension and CKD.[2–4] Potential mechanisms by which red and processed meat influence CVD risk (e.g., formation of trimethylamine-N-oxide) have emerged.[5] However, the molecular mechanism relevant for an overall diet that is high in plant-based diets and low in animal products and CVD is poorly understood.

Recently, technological advancement in high throughput proteomics has allowed simultaneous identification and quantification of thousands of proteins.[6] Large-scale proteomics has the potential to offer insights on biological processes that regulate food metabolism,[7] given that dietary intake directly and indirectly influences the proteome. Vitamins A and C has plasma carrier proteins (vitamin A for retinol binding protein 4; vitamin C for glucose transporters and solute carriers).[8,9] Vitamin C is also an antioxidant that reduces oxidative stress, indirectly impacting associated protein networks.[10] Supplementation of other nutrients that are rich in plant-based diets (e.g., vitamin E and folate) has been shown to alter the abundance of plasma proteome.[11,12] Based on these nutrients, protein markers of plant-based diets may implicate inflammation and other pathways. However, no study examined the associations between the plasma proteome of plant-based diet indices and linked them to incident cardiovascular disease.

5. Main Hypothesis/Study Questions:

Aim 1: To identify plasma proteins associated with greater adherence to 4 different types of plant-based diets (overall plant-based diets; healthful plant-based diets; unhealthful plant-based diets; and provegetarian diets)

Aim 2: To evaluate the prospective associations between plant-based diet-related proteins and incident CVD

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).

Study Design: Cross-sectional analysis (aim 1) of large-scale proteomics (SomaLogic) and plant-based diet indices, which were assessed at study visit 3 (1993-1995) and prospective analysis (aim 2) of diet-related proteins and risk of incident CVD through the latest follow-up period

Eligibility criteria: Approximately 11,000 ARIC participants with proteomic profiling data from visit 3 plasma specimens and data on dietary intake

Exposures & Outcomes: For aim 1, the exposure will be 4 plant-based diet indices (overall plant-based diet index (PDI); healthful plant-based diet index (hPDI); unhealthful plant-based diet index (uPDI); provegetarian diet index), and the outcome will be proteins. For aim 2, the exposure will be diet-related proteins (i.e., the subset of proteins that are statistically significantly associated with plant-based diets in aim 1) and the outcome will be incident CVD.

Plant-based diet indices: At visit 3, participants' usual dietary intake was assessed by trained interviewers using a modified version of the 66-item semiquantitative Willett food frequency questionnaire.[13] Participants reported the frequency with which they consumed foods and beverages of a specific serving size in the previous year. We will calculate the PDI, hPDI, uPDI, and provegetarian diet index using the responses from the food frequency questionnaire. Details on calculation of these scores are provided in our previous study and described in the table below (**Table 1**).[3]

Table 1. Classification and calculation of plant-based diet scores [3]

Table 1. Classification of food items into food groups for creation of the plant-based diet scores[a]

Food Groups	Food Items	Overall Plant-Based Diet Index	Healthy Plant-Based Diet Index	Less Healthy Plant-Based Diet Index	Provegetarian Diet Index
Plant foods[b]					
Whole grains	Cooked cereals such as oatmeal, grits, cream of wheat, dark or whole grain bread	Positive	Positive	Reverse	Positive[c]
Fruits	Apples, pears, oranges, peaches, apricots, plums, bananas, other fruits	Positive	Positive	Reverse	Positive
Vegetables	Broccoli, cabbage, cauliflower, brussels sprouts, carrots, corn, spinach, collards, or other greens, dark yellow, winter squash such as acorn, butternut, sweet potatoes, tomatoes	Positive	Positive	Reverse	Positive
Nuts	Peanut butter, nuts	Positive	Positive	Reverse	Positive
Legumes	String beans, green beans, baked beans or lentils (pinto, blackeye, baked beans), peas or lima beans	Positive	Positive	Reverse	Positive
Tea and coffee	Coffee, tea (iced or hot)	Positive	Positive	Reverse	Not scored
Refined grains	Biscuits, cornbread, cold breakfast cereal, white bread, pasta, rice	Positive	Reverse	Positive	Positive[c]
Potatoes	Potato or corn chips, French-fried potatoes, mashed potatoes	Positive	Reverse	Positive	Positive
Fruit juices	Orange juice, grapefruit juice	Positive	Reverse	Positive	Not scored
Sugar-sweetened and artificially sweetened beverages	Low-calorie soft drinks (any diet Coke, diet Pepsi), Regular soft drinks (Coke, Pepsi, 7-Up, ginger ale), fruit-flavored punch or noncarbonated beverages	Positive	Reverse	Positive	Not scored
Sweets and desserts	Chocolate bars or pieces, candy without chocolate, pie homemade from scratch, pie (ready-made or from mix), donuts, cake or brownie, cookies, Danish pastry, sweet roll, coffee cake, croissant	Positive	Reverse	Positive	Not scored
Animal foods					
Animal fat	Butter added to food or bread, butter used for cooking	Reverse	Reverse	Reverse	Reverse
Dairy	Skim or low-fat milk, whole milk, yogurt, ice cream, cottage cheese or ricotta cheese, other cheese	Reverse	Reverse	Reverse	Reverse
Eggs	Eggs	Reverse	Reverse	Reverse	Reverse
Fish or seafood	Canned tuna; dark meat fish, such as salmon, mackerel, swordfish, sardines, bluefish; other fish, such as cod, perch, catfish, shrimp, lobster, scallops	Reverse	Reverse	Reverse	Reverse
Meat	Chicken or turkey without skin, chicken or turkey with skin, hamburgers, hot dogs, processed meats (sausage, salami, bologna), bacon, beef, pork, or lamb as a sandwich or mixed dish, beef, pork, or lamb as a main dish, steak, roast, ham, liver	Reverse	Reverse	Reverse	Reverse
Miscellaneous animal foods	Home-fried food, such as any meats, poultry, fish, shrimp	Reverse	Reverse	Reverse	Not scored

[a]Food categorization scheme is similar to previous publications (7,8,22). Positive indicates that higher consumption of this food group received higher scores. Reverse indicates that higher consumption of this food group received lower scores.
[b]In the overall, healthy, and less healthy plant-based diet indices, whole grains, fruits, vegetables, nuts, legumes, tea, and coffee were considered "healthy plant foods." Refined grains, potatoes, fruit juices, sugar-sweetened and artificially sweetened beverages, and sweets and desserts were considered "less healthy plant foods." The provegetarian diet index did not differentiate plant foods as healthy or less healthy.
[c]In the provegetarian diet index, consumption of whole grains and refined grains was aggregated as the "grains" food group.

Proteomics: Fasting plasma samples were shipped to SomaLogic (SomaLogic, Boulder, CO) for proteomic profiling. We will focus on approximately 5,000 proteins which were identified in ARIC participants (visit 3).

Cardiovascular Disease: 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. 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. Information from these sources were used to adjudicate CVD outcomes by a physician review panel.

Other variables of interest: For aim 1, we will adjust for sociodemographic characteristics (age, sex, race-center, education), health behaviors (physical activity, smoking, alcohol intake, margarine intake), body mass index, eGFR, and total energy intake. For aim 2, we will additionally adjust for CVD risk factors, including diabetes and hypertension status.

Statistical Analysis:

We will examine the characteristics of the study population according to quintiles of plant-based diet scores using proportions for categorical variables and means (standard deviations (SD)) for continuous variables.

For Aim 1, we will use multivariable linear regression models to assess the association between 1-unit higher in plant-based diet indices and 5,000 individual proteins, adjusting for sociodemographic factors, health behaviors, and clinical factors (details above). To account for multiple testing, we will apply the false discovery rate at 0.05. As a secondary analysis, we will explore machine learning models. A protein signature of each of the 4 plant-based diet indices will be identified using elastic net models for optimal selection of proteins to represent plant-based diets. A tenfold cross-validation approach, randomly splitting 90% of the ARIC study data as a training set and 10% as a validation set, will be used to select strength of penalty (lambda value) for the elastic net model.

For Aim 2, we will use Cox proportional hazards regression models to evaluate whether individual plant-based diet-related proteins and protein signatures for the 4 plant-based diet indices (from elastic net models) are prospectively associated with incident cardiovascular disease, additionally adjusting for diabetes and hypertension status. Statistical significance will be assessed as FDR <0.05 for individual diet-related proteins and at P <0.05 for the protein signature.

Anticipated methodologic limitations of challenges: The present proposal will conduct a large number of statistical tests for approximately 5,000 proteins. There may be false positive findings, but we will apply the false discovery rate threshold of 0.05, and will replicate the findings in independent populations (Framingham Heart Study, Multi-Ethnic Study of Atherosclerosis).

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.cscc.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)?

MS#3884 Serum metabolomic signatures of plant-based diets and incident chronic kidney disease

The manuscript based on this proposal has been published [Kim H, Yu B, Li X, Wong KE, Boerwinkle E, Seidelmann SB, Levey AS, Rhee EP, Coresh J, Rebholz CM. Serum metabolomic signatures of plant-based diets and incident chronic kidney disease. Am J Clin Nutr 2022;116:151–64]. The prior proposal used untargeted metabolomics data, whereas our study will use proteomics data.

MS#3755 Protein biomarkers of the DASH diet and hypertension risk in the ARIC study

The manuscript based on this proposal is in press [Du S, Chen J, Kim H, Walker ME, Lichtenstein AH, Chatterjee N, Ganz P, Ramachandran VS, Coresh J, Rebholz CM. Protein biomarkers of healthy dietary patterns: results from the Atherosclerosis Risk in Communities (ARIC) Study. J Nutr 2022 (Epub ahead of print: https://doi.org/10.1016/j.tjnut.2022.11.008)]. This prior proposal focused on the DASH diet and hypertension risk, whereas our study is focused on a different dietary exposure (plant-based diets) and outcome (cardiovascular disease).

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: 2022.07)

___ 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.

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. Satija A, Bhupathiraju SN, Spiegelman D, Chiuve SE, Manson JE, Willett W, Rexrode KM, Rimm EB, Hu FB. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S Adults. *J Am Coll Cardiol*. 2017;70:411–422.

2. Kim H, Caulfield LE, Garcia-Larsen V, Steffen LM, Coresh J, Rebholz CM. Plant-Based Diets Are Associated With a Lower Risk of Incident Cardiovascular Disease, Cardiovascular Disease Mortality, and All-Cause Mortality in a General Population of Middle-Aged Adults. *J Am Heart Assoc*. 2019;8:e012865.

3. Kim H, Caulfield LE, Garcia-Larsen V, Steffen LM, Grams ME, Coresh J, Rebholz CM. Plant-based diets and incident CKD and kidney function. *Clin J Am Soc Nephrol*. 2019;14:682–91.

4. Kim H, Rebholz CM, Garcia-Larsen V, Steffen LM, Coresh J, Caulfield LE. Operational differences in plant-based diet indices affect the ability to detect associations with incident hypertension in middle-aged US adults. *J Nutr*. 2019;150:842–50.

5. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, Didonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WHW, Bushman FD, Lusis AJ, Hazen SL. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Medicine*. 2013;19:576–585.

6. Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, Jonasson C, Sarzynski MA, Shipley MJ, Alexander L, Ash J, Bauer T, Chadwick J, Datta G, DeLisle

RK, Hagar Y, Hinterberg M, Ostroff R, Weiss S, Ganz P, Wareham NJ. Plasma protein patterns as comprehensive indicators of health. *Nat Med*. 2019;25:1851–1857.

7. Wang J, Li D, Dangott LJ, Wu G. Proteomics and Its Role in Nutrition Research. *The Journal of Nutrition*. 2006;136:1759–1762.

8. Quadro L, Hamberger L, Colantuoni V, Gottesman ME, Blaner WS. Understanding the physiological role of retinol-binding protein in vitamin A metabolism using transgenic and knockout mouse models. *Mol Aspects Med*. 2003;24:421–430.

9. Padayatty SJ, Levine M. Vitamin C physiology: the known and the unknown and Goldilocks. *Oral Dis*. 2016;22:463–493.

10. Kojo S. Vitamin C: basic metabolism and its function as an index of oxidative stress. *Curr Med Chem*. 2004;11:1041–1064.

11. Aldred S, Sozzi T, Mudway I, Grant MM, Neubert H, Kelly FJ, Griffiths HR. Alpha tocopherol supplementation elevates plasma apolipoprotein A1 isoforms in normal healthy subjects. *Proteomics*. 2006;6:1695–1703.

12. Duthie SJ, Horgan G, de Roos B, Rucklidge G, Reid M, Duncan G, Pirie L, Basten GP, Powers HJ. Blood Folate Status and Expression of Proteins Involved in Immune Function, Inflammation, and Coagulation: Biochemical and Proteomic Changes in the Plasma of Humans in Response to Long-Term Synthetic Folic Acid Supplementation. *J Proteome Res*. 2010;9:1941–1950.

13. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122:51–65.