ARIC Manuscript Proposal #3884

PC Reviewed: 7/13/21	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Serum metabolomic signatures of plant-based diets and incident chronic kidney disease

b. Abbreviated Title (Length 30 characters): Metabolomics of plant-based diets

2. Writing Group:
Writing group members:

Hyunju Kim, PhD, MPH
Bing Yu, PhD, MS
Morgan E. Grams, MD, PhD, MHS
Eric Boerwinkle, PhD
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: Hyunju Kim
Address: 2024 E. Monument St, Suite 2-500
Baltimore, Maryland 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 M. 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 immediately after the manuscript proposal has been approved. 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 are dietary patterns which are comprised predominantly of plant foods and are low in animal products. Recently, several prospective studies have shown that greater adherence to plant-based diets is associated with lower risk of obesity,[1] hypertension,[2] type 2 diabetes,[3] cardiovascular disease,[4,5] and all-cause mortality.[6] Some of these studies have also reported that the risks of these conditions differ by the quality of plant foods. The ARIC Study, a study of nurses and health professionals in the US, and a study of Spanish university graduates reported that those with greater adherence to healthful plant-based diets (diets high in fruits, vegetables, whole grains, nuts, legumes, tea, coffee and low in refined grains, sweets and desserts, sugar-sweetened beverages, and animal products) was associated with lower risk of cardiometabolic outcomes.[1–4] In contrast, greater adherence to less healthful plant-based diets (diets high in refined grains, sweets and desserts, sugar-sweetened beverages, and low in healthful plants foods and animal products) was associated with an elevated risk of cardiometabolic outcomes.[1–4]

In the ARIC Study, we found that those in the highest vs. lowest quintile of provegetarian diets (similar to overall plant-based diets) and healthful plant-based diets had 10-14% lower risk of incident chronic kidney disease (CKD), whereas those in the highest vs. lowest quintiles of less healthful plant-based diets had 11% higher risk of incident CKD.[7] Despite the accumulating evidence, our understanding on the mechanism through which plant-based diets are associated with CKD is limited. The mechanisms through which plant-based diets are associated with CKD are diverse, because these diets are characterized by many dietary factors (high in vegetables, fruits, legumes, and nuts; high in fiber and vitamin A, C, E; low in red and processed meats, poultry, dairy, eggs; low in dietary acid load).[2,7,8]

Metabolomics, an approach which measures low-molecular-weight metabolites in biofluids simultaneously, may be a useful tool to expand our understanding on the physiologic effects of diet.[9–11] The metabolome is influenced by several factors, including diet; thus, this approach can help identify objective biomarkers of plant-based diets and elucidate different metabolic pathways underlying associations between plant-based diets and kidney disease.

5. Main Hypothesis/Study Questions:

Aim 1: To identify metabolites associated with greater adherence to 4 different types of plant-based diets (overall plant-based diets; healthful plant-based diets; less healthful plant-based diets; provegetarian diet), as potential biomarkers of these dietary patterns

Aim 2: To evaluate whether candidate biomarkers of 4 plant-based diets are associated with incident CKD

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 metabolomics and plant-based diet scores, which were assessed at study visit 1 (1987-1989) and prospective analysis (aim 2) of diet-related metabolites and risk of incident CKD through the latest follow-up period

Eligibility criteria: Approximately 4,000 African Americans and European Americans in the ARIC Study with metabolomic profiling data from visit 1 serum specimens and data on dietary intake

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

Plant-based diet scores: At visit 1, participants' usual dietary intake was assessed by trained interviewers using a modified version of the 66-item semiquantitative Willett food frequency questionnaire.[12] 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.[7]

Table 1. Classification and calculation of plant-based diet scores [7]

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.

Metabolomics: Serum fasting metabolites were profiled by Metabolon, Inc (Durham, North Carolina) using an untargeted, ultra-performance liquid chromatography tandem mass spectrometry approach in two sub-samples/batches of the ARIC study. We will focus on named metabolites which were identified in both of the ARIC samples. As a secondary analysis, we will analyze all named metabolites identified in at least one of the samples.

Kidney Disease: A composite definition will be used to ascertain incident CKD, which was defined as meeting one of the following criteria: 1) estimated glomerular filtration rate (eGFR) <60 ml/min/1.73 m^2 with ≥25% eGFR decline at any follow-up study visit relative to baseline; 2) kidney-disease related hospitalizations or deaths according to codes from the International Classification of Diseases (ICD) Ninth and Tenth Revisions, 3) a death related to CKD stage 3+ identified through linkage to the National Death Index, and 4) end-stage renal disease identified by linkage to the US Renal Data System (USRDS) registry.[13]

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

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, separately in sample 1 and sample 2.

For the cross-sectional analysis, we will primarily focus on the 374 metabolites that were identified in both sample 1 and sample 2, using sample 1 as a discovery data set, and sample 2 as a replication data set. We will use multivariable linear regression models to assess the association between 1 unit higher in plant-based diet scores and individual metabolites. In sample 1, we will use a threshold of 3.34×10^{-5} (0.05/4 dietary patterns/374 metabolites) to account for multiple testing. In sample 2, we will focus only on statistically significant metabolites from sample 1, and adjust the Bonferroni threshold accordingly (0.05/number of significant associations in sample 1). Then, we will meta-analyze significant metabolites in sample 1 and sample 2 using fixed-effects models.[14] We will use C statistics to assess whether significant metabolites improve the prediction of those in the highest quintiles vs. lowest 4 quintiles of four plant-based diet scores. As a secondary analysis, metabolites (*n*=385) that were available only in sample 2 will be analyzed using the same analytic methods with a threshold of 3.25×10^{-5} (0.05/4 dietary patterns/385 metabolites).

For the prospective analysis, we will assess the association between statistically significant diet-related metabolites (in sample 1 and sample 2) and incident CKD using Cox proportional hazards regression models. For the prospective analysis, we will estimate hazard ratios per 1-standard deviation higher in metabolites, and additionally adjust for diabetes, hypertension, and cardiovascular disease. We will conduct mediation analyses to evaluate if metabolites are significant mediators of the association between plant-based diets and incident CKD. To assess model prediction performance of metabolites significantly associated with CKD, we will use C statistics and examine whether addition of these metabolites improve the prediction of incident CKD on top of the CKD risk prediction equation.[15]

Anticipated methodologic limitations of challenges: Given the large number of metabolites and the large number of statistical tests we will be conducting, there may be false positive findings. We plan to address this challenge by applying the most conservative method (Bonferroni adjusted *P* values).[16]

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#2034: The human metabolome is associated with dietary intake among African Americans in the Atherosclerosis Risk in Communities Study (lead author: Yan Zheng)

> The manuscript based on this proposal has been published [Zheng Z, Yu B, Alexander D, Steffen LM, Boerwinkle E. Human metabolome associates with dietary intake habits among African Americans. Am J Epidemiol 2014;179(12):1424-1433]. This proposal focused on food groups and food items, whereas our study will focus on plant-based diets. In addition, the prior study used data from only African Americans, whereas our study will use data from both African Americans and European Americans.

MS#1882: A longitudinal study of metabolomics and kidney function among African Americans in the Atherosclerosis Risk in Communities (ARIC) study (lead author: Bing Yu)

The manuscript based on this proposal has been published [Yu B, Zheng Y, Nettleton JA, Alexander D, Coresh J, Boerwinkle E. Serum metabolomic profiling and incident CKD among African Americans. Clin J Am Soc Neprhol 2014;9(8):1410-1417]. This published study used data from only African Americans, and the prior study is more comprehensive than the proposed analysis in that the authors reported on all available metabolites whereas this proposal will analyze only diet-related metabolites in association with incident CKD.

MS#3145: Serum Metabolomic Markers of Diet Quality and Kidney Disease Risk (lead author: Casey Rebholz)

The manuscript based on this proposal has been published [Kim H, Hu EA, Wong K, Yu B, Steffen LM, Seidelmann SB, Boerwinkle E, Coresh J, Rebholz CM. Serum metabolites associated with healthy diets in African Americans and European Americans. J Nutr 2021;151:40–9]. This published study used data on 4 diet quality indices (Healthy Eating Index, Alternative Healthy Eating Index, the Dietary Approaches to Stop Hypertension diet, and the alternate Mediterranean diet), but we will focus on plant-based diet scores.

Similar to MS#3145, the 3 manuscript proposals listed below are focused on distinct aspects of dietary intake, i.e., coffee, dietary acid load, and protein, and are part of the same overall body of work leveraging the metabolomics, dietary, and kidney disease data to identify diet biomarkers and disease mediators.

MS#3228: Metabolomics of coffee consumption and risk of kidney disease in the ARIC study (first author: Casey Rebholz)

MS#3458: Novel biomarkers for dietary acid load and incident CKD (ARIC author: Casey Rebholz)

MS#3718: Serum metabolomic markers of protein and kidney disease risk (first author: Casey Rebholz)

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

2014.20: Genomics, Metabolomics, and Cardiovascular Disease (PI: Eric Boerwinkle)

2008.16: Metabolomics & Heart Failure: A Novel Approach to Biomarker Discovery (PI: Jennifer Nettleton)

___ 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. Gómez-Donoso C, Martínez-González MÁ, Martínez JA, et al. A provegetarian food pattern emphasizing preference for healthy plant-derived foods reduces the risk of overweight/obesity in the SUN Cohort. *Nutrients*. 2019;11(7):1553. doi:10.3390/nu11071553

2. 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(4):842-850. doi:10.1093/jn/nxz275

3. Satija A, Bhupathiraju SN, Rimm EB, et al. Plant-based dietary patterns and incidence of type 2 diabetes in US men and women: results from three prospective cohort studies. *PLoS Med*. 2016;13(6):e1002039. doi:10.1371/journal.pmed.1002039

4. Satija A, Bhupathiraju SN, Spiegelman D, et al. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S Adults. *J Am Coll Cardiol*. 2017;70(4):411-422. doi:10.1016/j.jacc.2017.05.047

5. 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(16):e012865. doi:10.1161/JAHA.119.012865

6. Kim H, Caulfield LE, Rebholz CM. Healthy plant-based diets are associated with lower risk of all-cause mortality in US adults. *J Nutr*. 2018;148(4):624-631. doi:10.1093/jn/nxy019

7. Kim H, Caulfield LE, Garcia-Larsen V, et al. Plant-based diets and incident CKD and kidney function. *Clin J Am Soc Nephrol*. 2019;14(5):682-691.

8. Satija A, Hu FB. Plant-based diets and cardiovascular health. *Trends Cardiovasc Med*. 2018;28(7):437-441. doi:10.1016/j.tcm.2018.02.004

9. Tzoulaki I, Ebbels TMD, Valdes A, Elliott P, Ioannidis JPA. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. *Am J Epidemiol*. 2014;180(2):129-139. doi:10.1093/aje/kwu143

10. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin Chem*. 2018;64(1):82-98. doi:10.1373/clinchem.2017.272344

11. Whitfield PD, German AJ, Noble P-JM. Metabolomics: an emerging post-genomic tool for nutrition. *Br J Nutr*. 2004;92(4):549-555. doi:10.1079/BJN20041243

12. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122(1):51-65.

13. Grams ME, Rebholz CM, McMahon B, et al. Identification of incident CKD stage 3 in research studies. *Am J Kidney Dis*. 2014;64(2):214-221. doi:10.1053/j.ajkd.2014.02.021

14. Harris RJ, Deeks JJ, Altman DG, Bradburn MJ, Harbord RM, Sterne JAC. Metan: Fixed- and Random-Effects Meta-Analysis: *The Stata Journal*. Published online April 1, 2008. doi:10.1177/1536867X0800800102

15. Nelson RG, Grams ME, Ballew SH, et al. Development of Risk Prediction Equations for Incident Chronic Kidney Disease. *JAMA*. 2019;322(21):2104. doi:10.1001/jama.2019.17379

16. Curtin F, Schulz P. Multiple correlations and Bonferroni's correction. *Biol Psychiatry*. 1998;44(8):775-777. doi:10.1016/s0006-3223(98)00043-2