ARIC Manuscript Proposal #3533

PC Reviewed: 12/10/19	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Proteomics and kidney disease in a community based population

b. Abbreviated Title (Length 26 characters): Proteomics and CKD

2. Writing Group:

Writing group members: Morgan E. Grams, Teresa K. Chen, Casey M. Rebholz, Bing Yu, Jingsha Chen, Christie Ballantyne, Eric Boerwinkle, Dan Arking, Pamela Lutsey, Tom Mosley, Josef Coresh, *others welcome* (order TBD).

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

First author: Morgan Grams
Address: 2024 E. Monument Street, Suite 2-638
Baltimore, Maryland 2187
Phone: 443-287-1827 Fax: 410-955-0485
E-mail: mgrams2@jhmi.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: Josef Coresh
Address: 2024 East Monument Street, Suite 2-600
Baltimore, Maryland 21287
Phone: 410-955-0495 Fax: 410-955-0485
E-mail: coresh@jhu.edu

3. Timeline: Analyses will begin once the manuscript proposal has been approved. We anticipate that the manuscript will be written and submitted to the ARIC Publications Committee within one year of the manuscript proposal being approved.

4. Rationale: Chronic kidney disease (CKD) disproportionately affects older adults. The prevalence of CKD, defined by estimated glomerular filtration rate (eGFR) <60 ml/min/1.73 m^2 or urine albumin-to-creatinine ratio (ACR) >30 mg/g ,[1] was nearly 50% in persons older than 70 years in the National Health and Nutrition Examination Survey (1999-2004).[2] CKD is associated with myriad morbidity and mortality, including end-stage renal disease (ESRD),[3,4] cardiovascular disease,[5] heart failure, and all-cause and cardiovascular mortality.[6] Still, the

progression of CKD is quite variable; identifying factors associated with poor prognosis could help accurately risk stratify as well as identify potential targets for intervention.

The kidney affects blood concentrations of proteins through diverse processes of filtration, secretion, production, and degradation. Preliminary analyses suggest that protein levels are closely correlated with eGFR.[7] Protein levels from blood integrate both genetic and environmental exposures[105–107] and may reflect processes that mediate adverse outcomes associated with CKD.

5. Main Hypothesis/Study Questions:

Our overarching hypothesis is that proteomic pathways will provide insight on the pathogenesis of CKD-associated outcomes.

Aim 1: To identify proteins and pathways associated with eGFR and albuminuria in ARIC participants.

Aim 2: To determine whether levels of proteins are associated with incident CKD, eGFR decline, and ESRD in ARIC participants.

Aim 3: To determine whether levels of proteins are associated with ESRD and mortality in ARIC participants with eGFR <60 ml/min/1.73 m^2.

Aim 4: To perform a genome-wide association study (GWAS) detecting genetic markers of levels of proteins implicated in CKD-associated outcomes separately in African-American and white ARIC participants. We will also explore the risk associations of identified genotypes with CKD-related outcomes in external datasets using Mendelian randomization methods.

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: We will conduct analyses of the ARIC cohort, treating Visit 3 (1993-1995) as the baseline visit. In sensitivity analyses, we will also consider visit 5 as the baseline visit.

Study Population: The study population will consist of white and African-American ARIC participants with proteomics data from Visits 3 and/or 5.

Exposure: For Aim 1, the primary exposure will be baseline eGFR based on creatinine and cystatin and secondary exposure will be albuminuria based on urine albumin-to-creatinine ratio (ACR). For Aims 2 and 3, the exposure of interest will be protein levels at Visit 3 and, in sensitivity analysis, at Visit 5. For Aim 4, the exposures will be genetic markers in the genome.

Outcomes: For Aim 1, the outcome will be proteomics data previously measured from Visits 3 and Visit 5 samples using the SomaScan Platform (SomaLogic, Inc). For Aim 2, the outcomes will be: 1) incident CKD; 2) eGFR decline; and 3) incident ESRD. Creatinine-based and/or cystatin-based Chronic Kidney Disease Epidemiology (CKD-EPI) Collaboration equations will

be used to estimate GFR.[8,9] Consistent with prior ARIC publications, incident CKD will be defined as having any one of the following: 1) an eGFR <60 ml/min/1.73 m^2 at follow-up (accompanied by a ≥25% eGFR decline relative to baseline; 2) CKD-related hospitalization or death based on the International Classification of Diseases (ICD) 9 or 10 codes; or 3) ESRD as identified by the US Renal Data System (USRDS) registry.[10-12] For exploratory analyses of eGFR decline, at the time of ESRD onset, eGFR will be imputed as 15 ml/min/1.73 m^2. For Aim 3, mortality will be determined from active surveillance techniques, including linkage to the National Death Index.[13] For Aim 4, outcomes will be the levels of identified proteins of interest at Visit 3 and in secondary analysis we will explore protein levels at Visit 5.

Statistical Analysis: We will use descriptive statistics, including means, medians, and proportions to compare baseline characteristics by CKD status at Visit 3. Formal testing will be performed using student's t-test or Wilcoxon rank-sum test for continuous variables and chi-squared for categorical variables. We anticipate that the distributions of proteins will be skewed: we plan to transform (e.g., log base-2) to achieve a more normal distribution. For Aim 1 (n~11,500), linear regression models will be used to study the associations of eGFR (and in visit 5, log-transformed urine ACR) with proteins in the proteome. Bonferroni correction will be used to account for multiple comparisons. Model 1 will be unadjusted; Model 2 will adjust for age, sex, and race-center; Model 3 will further adjust for baseline cholesterol, HDLc, diabetes, systolic blood pressure, anti-hypertension medication at Visit 3. For Aim 2 (n~11,500), Cox proportional hazards models will be constructed to study the associations of proteins at Visit 3 with: 1) incident CKD; and 2) incident ESRD. We will adjust for the same covariates as in Aim 1 but also add in eGFR to Model 3. To examine the association of proteins with subsequent eGFR decline, we will fit linear mixed-effects models with random intercepts and random slopes, adjusting for the same covariates as above except eGFR, which will be the independent variable. We will perform these analyses overall and stratified by race. For Aim 3, we will perform similar analyses but evaluating the associations between protein level and ESRD and death among participants with baseline eGFR <60 ml/min/1.73 m^2. For Aim 4, we will perform a GWAS to identify genetic markers that are associated with higher levels of proteins of interest separately within African-American and white ARIC participants.

Limitations: We acknowledge that our proposed study has a few limitations. First, we are only using data from Visit 3 onwards. Still, duration of follow-up for the current proposed study is still long (up to 17 years). Second, the SOMA is an aptamer based platform and the accuracy of protein identification is not always known. Third, power is limited for Mendelian randomization analyses, which requires us to evaluate associations in other, larger datasets such as CKD-Gen and the UK Biobank.

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? __X__ 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? __X__ Yes ____ 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"? __X__ 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/search.php

___**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)?

Manuscript #: TBD, APOL1 risk variants and proteomics in a community based population
Description: This manuscript proposal will evaluate the associations between proteins and *APOL1* risk status.

Manuscript #: 3324
Whole Genome Sequence and Proteomics for Gene Discovery in the Atherosclerosis Risk in Communities (ARIC) Study
Description: This group will evaluate genetic determinants of the proteome using whole genome sequencing data.

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*__ 2017.27____)
_ _ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __2011.03_(Selvin for funding on visit 6 labs, Matsushita for funding of visit 3 labs))

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __**X**__ No.

References

1. Kidney Disease: Improving Global O. Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int.Suppl.* 2013;3(Journal Article):1-150.

2. Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *JAMA.* 2007;298(17):2038-2047.

3. Astor BC, Matsushita K, Gansevoort RT, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with mortality and end-stage renal disease. A collaborative meta-analysis of kidney disease population cohorts. *Kidney Int.* 2011;79(12):1331-1340.

4. Gansevoort RT, Matsushita K, van der Velde M, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. *Kidney Int.* 2011;80(1):93-104.

5. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *NEJM.* 2004;351(13):1296-1305.

6. Matsushita K, van der Velde M, Astor BC, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. *Lancet.* 2010;375(9731):2073-2081.

7. Coresh JI, L.A.; Sang, Y.; Chen, J.; Shafi, T.; Post, W.S.; Shlipak, M.G.; Ford, L.; Goodman, K.; Perichon, R.; Greene, T.; Levey, A.S. Metabolomic profiling to improve GFR estimation: A proof of concept. *Nephrol Dial Transplant.* 2018;In press.

8. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* May 5 2009;150(9):604-612.

9. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med.* Jul 5 2012;367(1):20-29.

10. Rebholz CM, Selvin E, Liang M, et al. Plasma galectin-3 levels are associated with the risk of incident chronic kidney disease. *Kidney Int.* Jan 2018;93(1):252-259.

11. Sumida K, Kwak L, Grams ME, et al. Lung Function and Incident Kidney Disease: The Atherosclerosis Risk in Communities (ARIC) Study. *Am J Kidney Dis.* Nov 2017;70(5):675-685.

12. Tin A, Scharpf R, Estrella MM, et al. The Loss of GSTM1 Associates with Kidney Failure and Heart Failure. *J Am Soc Nephrol.* Nov 2017;28(11):3345-3352.

13. Grams ME, Rebholz CM, Chen Y, et al. Race, APOL1 Risk, and eGFR Decline in the General Population. *J Am Soc Nephrol.* Sep 2016;27(9):2842-2850.