ARIC Manuscript Proposal #3818

(Submitted 11/4/2020)

PC Reviewed: 4/13/21	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Omic data and adverse kidney outcomes

b. Abbreviated Title (Length 26 characters): Omics and adverse outcomes in CKD

2. Writing Group:
Morgan Grams, Josef Coresh, Linda Zhou, Adi Surapaneni, Pascal Schlosser, Eric Boerwinkle, Anna Kottgen, Bing Yu, *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: MG
Address:

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: Morgan Grams
Address: 2024 East Monument Street, Suite 2-638
Baltimore, Maryland 21287
Phone: 443-287-1827
E-mail: mgrams2@jhmi.edu

3. Timeline: Analysis will begin upon receipt of the data and approval of the manuscript proposal. Manuscript will be written and submitted for ARIC Publications Committee review within one year of manuscript proposal approval.

4. Rationale:
Selection methods that relate omic data to clinical outcomes may not identify weak but cumulative effects of correlated features. Methods for determining statistical significance such as Bonferroni correction are overly conservative in this setting. Further, integrating difference sources of proteomics and metabolomics may help identify common molecular pathways that underlie disease progression. We propose to apply Netboost, a dimension reduction technique, to aggregate information across high-dimensional omic data and use these modules to relate the summary meaures to clinical outcomes.[1]

The disease of interest is chronic kidney disease, a common and morbid disease that is associated with risk of cardiovascular disease, end-stage kidney disease, and mortality, as well as outcomes such as decreased quality of life, bone disease, and decreased cognitive function.[2,3] Chronic kidney disease is highly related to both the metabolome and the proteome, components of which may precipitate disease progression.[4,5] Particularly in older populations, chronic kidney disease can afflict over 40% of the population.[6] Identifying molecular pathways that are up- or down-regulated in people with adverse outcomes may produce proteins or metabolites that may targeted in disease prevention efforts.

5. Main Hypothesis/Study Questions:

Aim 1: To identify clusters of metabolites, and to relate them to adverse kidney outcomes.
Aim 2: To identify clusters of proteins, and to relate them to adverse kidney outcomes.
Aim 3: To integrate metabolites and proteins into clusters, and to relate them to adverse kidney outcomes.

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 a prospective analysis of the ARIC cohort, using study visit 5 with follow-up through December 31, 2018 (or the most recent surveillance year).

Study Population: The study population will include all members of the ARIC cohort with available SOMAScan data, consent to participate in cardiovascular research, and available metabolomics data. All analyses will exclude those with ESKD at the study visit.

Exposure: The exposure for each aim will be eigen-modules: for aim 1, eigen-metabolites, for aim 2, eigen-proteins, and for aim 3, a mixture of metabolites and proteins. Because Netboost requires non missing data, we will focus on metabolites missing <80% of the time, and evaluate how robust clustering is to different methods of missing value imputation. A priori, we plan to impute missing data with the lowest value, as done in previous studies. Both metabolites and proteins have been measured using plasma from ARIC visit 5.

Outcomes:

The main outcomes will be slope of eGFR decline, CKD progression, and incident ESKD. Incident ESKD will be defined as the initiation of renal replacement therapy (either dialysis or transplant) and cases will be defined through linkage of the ARIC study with the United States Renal Data System (USRDS) registry. CKD progression will be defined as ESKD, or 50% decline in GFR using subsequent visit data. We will also assess cross-sectional relationships with eGFRcrcys and log-ACR (urine albumin-to-creatinine ratio). We will estimate GFR based on creatinine and cystatin using the CKD-Epi equation.

Statistical Analysis:

We will cluster omic measures using Netboost or a related dimension-reduction technique such as weighted correlated network analysis. We will evaluate the association between eigenmodules and outcomes using the appropriate statistical models: linear mixed models with random intercept and slope for eGFR decline; Cox proportional hazards model for CKD progression and incident ESKD. We will test associations in unadjusted models, demographic- and study-center adjusted models, and fully-adjusted models, which include known risk factors for adverse kidney outcomes (demographics, study center, cardiovascular disease, smoking, eGFRcrcys, urine albumin-to-creatinin, hypertension, diabetes). For modules that are significantly associated with kidney outcomes, we will evaluate genetic associations with eigenmodules separately in whites and Blacks for potential Mendelian randomization studies and seek replication at other visits or using other cohorts as available.

Limitations:

SOMAscan provides aptamer levels, which may not perfectly correlate with protein levels.

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? __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)?

ARIC Manuscript Proposal #3533: Proteomics and kidney disease in a community based population
ARIC Manuscript Proposal #1182: A longitudinal Study of Metabolomics and Kidney Function among African Americans in the Atherosclerosis Risk in Communities (ARIC) Study

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, 2013.21)

___ 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 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.

1. Schlosser P, Knaus J, Schmutz M, et al. Netboost: Boosting-supported Network Analysis Improves High-Dimensional Omics Prediction in Acute Myeloid Leukemia and Huntington's Disease. *IEEE/ACM Trans Comput Biol Bioinform.* 2020;PP.
2. Gorodetskaya I, Zenios S, McCulloch CE, et al. Health-related quality of life and estimates of utility in chronic kidney disease. *Kidney Int.* 2005;68(6):2801-2808.
3. Chronic Kidney Disease Prognosis C, Matsushita K, van der Velde M, 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.
4. Tin A, Yu B, Ma J, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *J Appl Lab Med.* 2019;4(1):30-39.
5. Grams ME, Shafi T, Rhee EP. Metabolomics Research in Chronic Kidney Disease. *J Am Soc Nephrol.* 2018;29(6):1588-1590.
6. Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *Jama.* 2007;298(17):2038-2047.