ARIC Manuscript Proposal #4217

PC Reviewed: 3/14/23	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Urine metabolites and kidney function in an older community-based population

b. Abbreviated Title (Length 26 characters):

2. Writing Group:

Writing group members: Wan-Jin Yeo, Adi Surapaneni, Josef Coresh, Jingsha Chen, Pascal Schlosser, Casey Rebholz, Bing Yu, Peggy Sekula, Anna Köttgen, Morgan Grams

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

First author: Wan-Jin Yeo
Address: 227 East 30th Street, #826H
New York, NY 10016
Phone: 484.809.5965
E-mail: wan-jin.yeo@nyulangone.org

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: 227 East 30th Street, #825
New York, NY 10016
Phone: 646.501.2732
E-mail: morgan.grams@nyulangone.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:

Estimated glomerular filtration rate (eGFR) is a metric that estimates the filtering capability of kidneys, and is calculated using the concentration of serum creatinine[4], cystatin C[5], or both[6].

However, eGFR may be inaccurate since serum creatinine and cystatin C levels may be influenced by factors that are at least partially independent of kidney function. For example, creatinine levels are influenced by muscle mass[7,8], and cystatin C levels can be influenced by factors such as age, mass, and drug use[9,10]. Moreover, there may be a latency in creatinine concentration changes during early stages of chronic kidney disease (CKD)[11], which renders eGFR to be insensitive to detecting CKD onset. Due to similar reasons, other metrics of kidney function such as albumin-to-creatinine ratio (ACR) or protein-to-creatinine (PCR) may thus be inaccurate as well[12].

Finding strong associations between certain urine metabolites and eGFR or ACR cross-sectionally allows for the identification of candidates that may be able to quantify kidney function more accurately or provide insight on different aspects of kidney function. This allows for the possibility of improving eGFR estimates by incorporating the identified metabolites that are freely-filtered into current estimation calculations. The identified metabolites may also highlight potential pathophysiological mechanisms related to CKD progression. In previous studies, associations between eGFR and serum metabolites have been found[1,2,3,13]. Identifying metabolites highly associated with eGFR that are common in both serum and urine metabolomics ("paired metabolites") may further inform us of the underlying mechanisms in CKD progression. Associations between eGFR and the fractional excretion (FE) of metabolites, which can be calculated for each paired metabolite, may also lend to this aim[18].

Since ARIC contains data for multiple visits of the same population, longitudinal analyses to find associations between urine metabolites and changes in eGFR or ACR can also be done. The ARIC data available extends to visit 8, which is representative of an older population. CKD is known to be more prevalent in older populations[16,17], thus we may identify metabolites associated with CKD progression. Previous studies have found associations between urine metabolites and eGFR/eGFR decline/adverse kidney outcomes[14,15]. However, these studies utilize data solely from individuals of European ancestry. The ARIC data available to us is enriched with both European American and African American individuals, which may be more representative of a general population.

5. Main Hypothesis/Study Questions:

We aim to find new associations and replicate known associations between urine metabolites and eGFR/ACR via a cross-sectional analysis of ARIC visit 5. A paired study design to find cross-sectional associations between serum metabolites and eGFR/ACR will also be performed for ARIC visit 5. For face validity, the relationship between urine/serum metabolites and age and sex will be examined.

With the paired urine and serum metabolomics studies, will identify metabolites found both in urine and serum that are associated with kidney function, kidney function related genetic variants, and kidney-related medications ("paired metabolites"). For each paired metabolite, its FE will be calculated and ranked, and the cross-sectional association with eGFR/ACR will be determined.

We also aim to find associations between baseline urine metabolite levels/serum metabolite levels/FEs and decline in kidney function via longitudinal analysis across ARIC visits 5, 6 and 7. Specifically, to investigate associations between the predictors and CKD progression, we will

look into a decline of ≥ 40% in eGFR levels or the doubling of ACR levels as well as incidence of ESKD.

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 and Population:

Cross-sectional analyses to determine associations between urine/serum metabolites/FE and kidney function will be performed on the dataset of ARIC participants with urine metabolomics at visit 5. Longitudinal analyses to determine associations between urine/serum metabolites/FE and decline in kidney function will be performed on ARIC visits 5, 6 and 7.

Exposure:

For cross-sectional study:
eGFR, $\log_2$(ACR), FE, age, sex, race, study center (other clinical covariates may be added in a subsequent, fully adjusted model, including hypertension, diabetes, cardiovascular disease, cholesterol, serum albumin if available)

For longitudinal study:
$\log_2$(urine metabolite levels), eGFR, $\log_2$(ACR), FE, age, sex, race, study center (other clinical covariates may be added in a subsequent, fully adjusted model, including hypertension, diabetes, cardiovascular disease, cholesterol, serum albumin if available)

Outcomes:

For cross-sectional study:
$\log_2$(urine metabolite levels) or $\log_2$(serum metabolite levels)

For longitudinal study:
Decline in kidney function or ESKD

Statistical Analysis:

Data preprocessing
First, samples with >50% missingness of known non-xenobiotic metabolites will be excluded from the dataset. Then, non-xenobiotic metabolites (both known and unknown) with missingness of >80% will be excluded. For each metabolite, any missing entries will be imputed with the minimum value measured for that metabolite. Probability quotient normalization will then be performed for all samples to account for varying levels of dilution. Since the distribution of each metabolite is expected to be skewed, a $\log_2$ will be performed. Finally, metabolites with a variance of ≤ 0.01 will be excluded, and outliers that are beyond 5 standard deviations from the

mean for each of the remaining metabolites will be capped at the maximum/minimum measured value.

FE values for paired metabolites will be determined with the standard FE equation

$$\text{FE} = \frac{(\text{urine metabolite})/(\text{urine creatinine})}{(\text{serum metabolite})/(\text{serum creatinine})} \times 100$$

Analysis steps

For the cross-sectional analysis, an estimate of the association between each urine/serum metabolite/FE and eGFR/$\log_2$(ACR)/age/sex will be obtained via multiple linear regression models as follows:

For associations between metabolites and eGFR:

$$\log_2(\text{metabolite}) \text{ or FE} \sim \text{eGFR} + \text{age} + \text{sex} + \text{race} + \text{study center} + \text{additional covariates}$$

For association between metabolite and ACR:

$$\log_2(\text{metabolite}) \text{ or FE} \sim \log_2(\text{ACR}) + \text{age} + \text{sex} + \text{race} + \text{study center} + \text{additional covariates}$$

For association between metabolite and age/sex:

$$\log_2(\text{metabolite}) \text{ or FE} \sim \text{eGFR} + \log_2(\text{ACR}) + \text{age} + \text{sex} + \text{race} + \text{study center} + \text{additional covariates}$$

For the longitudinal analysis, an estimate of the association between each metabolite and CKD progression will also be obtained via Cox proportional hazards regression models as follows:

CKD progression defined with eGFR decline:

$$\text{A 40\% decline in eGFR from baseline} \sim \log_2(\text{metabolite}) + \text{eGFR} + \log_2(\text{ACR}) + \text{FE} + \text{age} + \text{sex} + \text{race} + \text{study center} + \text{additional covariates}$$

CKD progression defined with ACR decline:

$$\text{Doubled ACR level from baseline} \sim \log_2(\text{metabolite}) + \text{eGFR} + \log_2(\text{ACR}) + \text{FE} + \text{age} + \text{sex} + \text{race} + \text{study center} + \text{additional covariates}$$

We will also evaluate associations that account for the competing risk of death or loss to follow-up using competing risk regression with the method of Fine and Gray.

The statistical significance of each association will be assessed against a Bonferroni-corrected *p*-value (i.e. $p = 0.05/\#\text{ metabolites} < 0.05$).

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" ? __x__ 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? __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 current derived consent file ICTDER05 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/aricproposals/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)?

Manuscript # 4005: Eicosanoids and kidney outcomes in a community-based population

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

___ A. primarily the result of an ancillary study (list number* _____)

___ 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.

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. Peng, Hongquan, et al. "Identification of Metabolite Markers Associated with Kidney Function." *Journal of Immunology Research* 2022 (2022).
2. Sekula, Peggy, et al. "A metabolome-wide association study of kidney function and disease in the general population." *Journal of the American Society of Nephrology: JASN* 27.4 (2016): 1175.
3. Titan, S. M., et al. "Metabolites related to eGFR: Evaluation of candidate molecules for GFR estimation using untargeted metabolomics." *Clinica Chimica Acta* 489 (2019): 242-248.
4. Levey, Andrew S. "CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration): a new equation to estimate glomerular filtration rate." *Ann Intern Med* 150 (2009): 604-612.
5. Stevens, Lesley A., et al. "Estimating GFR using serum cystatin C alone and in combination with serum creatinine: a pooled analysis of 3,418 individuals with CKD." *American journal of kidney diseases* 51.3 (2008): 395-406.
6. Inker, Lesley A., et al. "Estimating glomerular filtration rate from serum creatinine and cystatin C." *New England Journal of Medicine* 367.1 (2012): 20-29.
7. Baxmann, Alessandra Calábria, et al. "Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C." *Clinical journal of the American Society of Nephrology: CJASN* 3.2 (2008): 348.
8. Thongprayoon, Charat, Wisit Cheungpasitporn, and Kianoush Kashani. "Serum creatinine level, a surrogate of muscle mass, predicts mortality in critically ill patients." *Journal of thoracic disease* 8.5 (2016): E305.
9. Knight, Eric L., et al. "Factors influencing serum cystatin C levels other than renal function and the impact on renal function measurement." *Kidney international* 65.4 (2004): 1416-1421.
10. Stevens, Lesley A., et al. "Factors other than glomerular filtration rate affect serum cystatin C levels." *Kidney international* 75.6 (2009): 652-660.
11. Levey, Andrew S., Lesley A. Inker, and Josef Coresh. "GFR estimation: from physiology to public health." *American Journal of Kidney Diseases* 63.5 (2014): 820-834.
12. Ellam, Timothy J. "Albumin: creatinine ratio–a flawed measure? The merits of estimated albuminuria reporting." *Nephron Clinical Practice* 118.4 (2011): c324-c330.
13. Coresh, Josef, et al. "Metabolomic profiling to improve glomerular filtration rate estimation: a proof-of-concept study." *Nephrology Dialysis Transplantation* 34.5 (2019): 825-833.
14. Steinbrenner, Inga, et al. "Urine metabolite levels, adverse kidney outcomes, and mortality in CKD patients: a metabolome-wide association study." *American Journal of Kidney Diseases* 78.5 (2021): 669-677.
15. Dekker, Shosha EI, et al. "Urinary metabolites associate with the rate of kidney function decline in patients with autosomal dominant polycystic kidney disease." *PLoS One* 15.5 (2020): e0233213.
16. Tonelli, Marcello, and Miguel Riella. "Chronic kidney disease and the aging population." *Brazilian Journal of Nephrology* 36 (2014): 1-5.
17. Coresh, Josef, et al. "Prevalence of chronic kidney disease in the United States." *Jama* 298.17 (2007): 2038-2047.

18. Gohda, Tomohito, et al. "Fractional excretion of tumor necrosis factor receptor 1 and 2 in patients with type 2 diabetes and normal renal function." *Journal of diabetes investigation* 12.3 (2021): 382-389.