ARIC Manuscript Proposal #2074

PC Reviewed: 2/12/13	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title:
Incorporating genetic information in predicting end-stage renal disease (ESRD) related hospitalization

b. Abbreviated Title (Length 26 characters):
Genetics and ESRD prediction

2. Writing Group:
Adrienne Tin, Brad Astor, David Couper, Thomas Mosley, Josef Coresh, Nisa Maruthur, Linda WH Kao, and others

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

First author: Adrienne Tin
Address: 615 N. Wolfe Street, Suite W6012
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD 21205
Phone: 201-281-9577 Fax:
E-mail: atin@jhsph.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: Linda WH Kao
Address: 615 N. Wolfe Street, Suite W6012
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD 21205
Phone: 410 614-0945 Fax: 410 955 0863
E-mail: wkao@jhsph.edu

3. Timeline:
After the approval of the proposal, we will apply funding for this project from the National Kidney Foundation of Maryland Mini-Grant due in March 2013.

4. Rationale:

In comparison with European Americans (EA), African Americans (AA) carry excess burden of end stage renal disease (ESRD). The incidence of ESRD in AA has consistently been over three times higher than in EA. [1-3] This disparity has been attributed to a combination of genetic, environmental, and behavioral factors. [2, 4, 5] Recently, variants in the region of *APOL1/MYH9* on chromosome 22 were found to be associated with complex forms of kidney disease, such as focal segmental glomerulosclerosis (FSGS), HIV-associated nephropathy (HIVAN), and ESRD not attributable to diabetes in AA in case-control studies in a recessive manner. Reports of the associations between *MYH9* variants and hypertensive-attributed ESRD in AA independent of the *APOL1* G1 and G2 variants have been inconsistent. [6, 7] Therefore, this study will focus on the *APOL1* risk variants, which are associated with hypertensive-attributed ESRD with reported odds ratio of 7.3 (CI: 5.6 to 9.5)[6], and higher odds ratios were found for FSGS and HIVAN.[8] These *APOL1* variants, G1 (rs73885319 and rs60910145, two non-synonymous SNPs) and G2 (rs71785313, encoding two amino-acid deletion) are common in AA (G1 ~ 22.5% and G2 ~ 14.6%[9]) but with frequencies less than 1% in EA.[10] [9]Given the large effect size and common allele frequencies, these variants are likely to account for considerable proportion of the racial disparity in ESRD burden between AA and EA.

ESRD is a devastating condition. While prediction models for ESRD have been built,[11, 12] none currently address the racial disparity in ESRD or include genetic information. We hypothesize that given the strong association between the G1 and G2 alleles and hypertensive-attributed ESRD, incorporating genetic information will improve discrimination and calibration of ESRD prediction models. Further, we will assess whether African ancestry estimates, may capture additional genetic information for explaining the racial disparity in ESRD and for better prediction of ESRD. Increasing the accuracy in the prediction of ESRD can lead to more effective prevention of this disease and improvement of the clinical management of kidney disease.

Existing ESRD prediction models included participants with chronic kidney disease (CKD) stage 3 or higher. [11, 12] Compared to EA, prevalence of stage 3 CKD in AA has been shown to be lower[13], similar[14], or moderately higher[9] in population-based studies, while the prevalence of low eGFR (<20mL/min/1.73m^2) has consistently been higher in AA. [13, 14] This suggests more rapid disease progression among AA. Therefore, it may be useful to assess ESRD prediction model at the population level or at earlier stage of CKD before the begining of rapid progression. Predicting ESRD including asymptomatic individuals may be useful for primary prevention, while prediction among individuals with symptoms of CKD may be more useful for the clinical management.

We will conduct model development in all individuals with self-reported race of black or white using visit 1 as baseline and in the sub-population with mildly reduced eGFR (eGFR < 90mL/min/1.73m^2) and compare the importance of the types of prediction variables in the models for the overall and the reduced eGFR population. It is possible that laboratory measures may be more important in the models of the reduced eGFR population, while genetic variables may be more important in the models of the overall population.

On outcome definition, while the ARIC study does not have information on the initiation of renal replacement therapy, it has information of ESRD-related hospitalization up to 2008. We will use this as our outcome.

We will compare three types of models with different sets of predictor variables. Type I models will include demographics and genetic information which can provide risk assessment early in life. Type II models will include demographics, clinical and laboratory variables, which use information on the health status of an individual at baseline. Type III models will combine the demographics, genetics, clinical and laboratory variables.

Table 1. Population at baseline and prediction model types

	Model Type		
Population at baseline	Type I Model	Type II Model	Type III Model
Overall population: Individuals without history of ESRD and with eGFR > 15 ml/min/1.73m^2 at baseline	Demographics and genetic variables	Demographics, clinical, and laboratory variables	Demographics, genetics, clinical, and laboratory variables
Mildly reduced eGFR population: Individuals without history of ESRD and with eGFR between 15 to 90ml/min/1.73m^2			

On predictor variables, the ARIC study at visit 1 has most of the laboratory measures reported to be associated with ESRD, such as serum albumin, calcium, phosphate,[11] while these variables are not available at visit 4. At visit 4, the ARIC study has urinary albumin creatinine ratio (UACR), an important predictor of ESRD, while this variable is not available at visit 1. The primary model development will be conducted using visit 1 as baseline. The model performance will be assessed using visit 4 as baseline with the addition of UACR as a predictor.

For genetic ancestry estimates, we will use the first principal component generated from single nucleotide polymorphisms (SNPs) from the combined black and white population. The SNPs used for principal component analysis will not include the *APOL1* risk variants. Based on in-house data, the first principal component and African ancestry estimated from ancestry informative markers are highly correlated (Pearson correlation ~ 0.95).

5. Main Hypothesis/Study Questions:

1. The addition of genetic information of the *APOL1* G1 and G2 alleles and genetic ancestry information in ESRD-related hospitalization prediction models will improve model discrimination and calibration and risk stratification.
2. Replacing self-reported race with ancestry information represented by the first principal component and/or *APOL1* risk variant information in a ESRD-related hospitalization prediction model will improve model discrimination and calibration and risk stratification

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

Prospective cohort study

Inclusion criteria

All individuals in ARIC with self-reported race of white or black and data for the *APOL1* variants and candidate variables and estimated glomerular filtration rate (eGFR) > 15ml/min/1.73m^2 at baseline.

Outcome

1. Incident of ESRD-related hospitalization since baseline up to 2008 (Total events: ~200 events in AA and ~150 events in EA).

   ESRD-related hospitalization events were obtained by active surveillance and defined as:

   A. All those with ICD codes specified for stage 5 CKD, kidney transplant, dialysis or procedural code indicating dialysis (ICD 9: 585.5-585.6, ICD 10: N18.5-N18.6), except:
      a. The individual has ICD code of traumatic anuria (958.5) with the same event date;
      b. The Individuals has ICD code of acute kidney injury (AKI, 586.x and 788.9x) with same event date and without any prior CKD events as indicated by serum creatinine rise of 0.4mg/dL, eGFR < 60ml/min/1.73m^2, or surveillance ICD code for CKD.

   OR

   B. All those with AKI codes (584-584.9, 586, N17.0-N17.0) as an underlying cause of death and with a prior history of CKD as indicated by creatinine rise of 0.4mg/dL, eGFR < 60ml/min/1.73m^2 or surveillance ICD code for CKD.

Candidate variables for prediction models

The candidate variables are known risk factors of ESRD[15, 16] or significant predictor variables in ESRD prediction models[11, 12] with availability in the ARIC study. Although socioeconomic factors are known risk factors of ESRD, they are not included because these varibles are not likely to be available clinically.

Demographics variable: age, gender, and self-reported race

Clinical (including laboratory) variables at visit 1: eGFR (at 5 unit increment, based on serum creatinine calculated using the CKD-EPI equation[17]), hypertension, diabetes and current smoking status, systolic and diastolic blood pressures, HDL cholesterol, BMI, and history of coronary heart disease, serum albumin, calcium, and phosphate, uric acid, hemoglobin, and triglycerides levels.

Clinical (including laboratory) variables at visit 4: eGFR (at 5 unit increment), hypertension, diabetes, current smoking status, systolic and diastolic blood pressures, HDL cholesterol, BMI, and history of coronary heart disease, UACR, hemoglobin, and triglycerides levels.

Genetic variables: *APOL1* G1 and G2 genotypes, ancestry information represented by the first principal component estimates (PC1) using SNPs on Affymetrix 6.0 microarray.

Statistical analysis

We will use Cox proportional hazard regression for modeling. For type I models, we will compare a model including demographics variables plus self-report race with models including demographics variables plus PC1 and/or *APOL1* risk variant genotypes.

For type II models, since racial disparity exist in many candidate variables. The initial development will be conducted in the two race groups separately. We will perform univariate analysis to assess the associations between the clinical/laboratory variables and incident of ESRD-related hospitalization stratified by self-reported race. Variables associated with the outcome at p-value<0.1 will be included in model development. Then the black and white samples will be combined. Candidate variables that improved model fit in the stratified analyses will be included in the combined type II model. For candidate variables with substantial effect size differences in the stratified models, an interaction term with one of the genetic variable will be added. The interaction between diabetes and genetic variables will be tested because of the known association between the *APOL1* risk variants and hypertensive-attributed ESRD. The variables in the type I (demographics and genetic) and type II (demographics and clinical) models will be combined to form the type III model.

Model diagnostic will be perform to assess the validity of model assumptions, such as the use of scaled Schoenfeld residuals for the proportional hazard assumptions and the

martingale residuals for nonlinearity. The models will compared based on C statistics for discrimination, Hosmer-Lemeshow chi-square statistics for calibration, net reclassification improvement[18] for risk stratification.

The same model development process will be repeated in the subsample of mildly reduced eGFR.

Strengths and Limitations

The timing of the outcome was not captured precisely. Individuals might have started dialysis before their ESRD-related hospitalization events. In addition, individuals who started dialysis in clinics and did not have hospitalization events were not included as having outcomes.

Albuminuria or proteinuria, an important risk factor [15] or predictor of ESRD[11], is not available at visit 1. This may limit the performance of the models.

The biological mechanisms of the *APOL1* variants are unknown. These genetic variants may be a causal factor of one or more clinical variables, such as eGFR. Controlling for eGFR may attenuate the effect of the genetic variants.

The strengths of this study include the prospective nature of the study and having a sufficient number of events for model development and the availability of the combination of clinical, laboratory, and genetic predictors to begin to elucidate the contributions of different types of factors to racial disparity in ESRD.

7.a. Will the data be used for non-CVD analysis in this manuscript?__X Yes __ 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 ICTDER03 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)?
No previous proposal in ARIC focus specifically on predicting ESRD using genetic information.

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 at http://www.cscc.unc.edu/aric/forms/

12. 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.

References

1. Klag MJ, Whelton PK, Randall BL*, et al.* End-stage renal disease in African-American and white men. 16-year MRFIT findings. *JAMA* 1997; **277:** 1293-1298.

2. Norris KC, Agodoa LY. Unraveling the racial disparities associated with kidney disease. *Kidney Int* 2005; **68:** 914-924.

3. USRDS: USRDS 2012 Annual Data Report: Atlas of End-Stage Renal Disease in the United States. In, Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2012

4. Perneger TV, Whelton PK, Klag MJ. Race and end-stage renal disease. Socioeconomic status and access to health care as mediating factors. *Arch Intern Med* 1995; **155:** 1201-1208.

5. Powe NR, Melamed ML. Racial disparities in the optimal delivery of chronic kidney disease care. *The Medical clinics of North America* 2005; **89:** 475-488.

6. Genovese G, Friedman DJ, Ross MD*, et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 2010; **329:** 841-845.

7. Kopp JB, Freedman BI, Winkler CA*, et al.* APOL1 and MYH9 Genetic Variants Are Independently Associated with Kidney Disease Risk in African Americans. *J Am Soc Nephrol* 2011; **Abstract FR-PO1501**.

8. Kopp JB, Nelson GW, Sampath K*, et al.* APOL1 genetic variants in focal segmental glomerulosclerosis and HIV-associated nephropathy. *J Am Soc Nephrol* 2011; **22:** 2129-2137.

9. Friedman DJ, Kozlitina J, Genovese G*, et al.* Population-based risk assessment of APOL1 on renal disease. *J Am Soc Nephrol* 2011; **22:** 2098-2105.

10. O'Seaghdha CM, Parekh RS, Hwang SJ*, et al.* The MYH9/APOL1 region and chronic kidney disease in European-Americans. *Hum Mol Genet* 2011; **20:** 2450-2456.

11. Tangri N, Stevens LA, Griffith J*, et al.* A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 2011; **305:** 1553-1559.

12. Johnson ES, Thorp ML, Yang X*, et al.* Predicting renal replacement therapy and mortality in CKD. *Am J Kidney Dis* 2007; **50:** 559-565.

13. McClellan W, Warnock DG, McClure L*, et al.* Racial differences in the prevalence of chronic kidney disease among participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Cohort Study. *J Am Soc Nephrol* 2006; **17:** 1710-1715.

14. Clase CM, Garg AX, Kiberd BA. Prevalence of low glomerular filtration rate in nondiabetic Americans: Third National Health and Nutrition Examination Survey (NHANES III). *J Am Soc Nephrol* 2002; **13:** 1338-1349.

15. Hsu CY, Iribarren C, McCulloch CE*, et al.* Risk factors for end-stage renal disease: 25-year follow-up. *Arch Intern Med* 2009; **169:** 342-350.

16. Bash LD, Astor BC, Coresh J. Risk of incident ESRD: a comprehensive look at cardiovascular risk factors and 17 years of follow-up in the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Kidney Dis* 2010; **55:** 31-41.

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

18. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr.*, et al.* Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; **27:** 157-172; discussion 207-112.