ARIC Manuscript Proposal # 2952

PC Reviewed: 3/14/17   **Status:** _____   **Priority: 2**
SC Reviewed: _____   **Status:** _____   **Priority:** ____

1.a. **Full Title**: A GWAS Meta-Analysis of Type 2 Diabetes-attributed End-Stage Kidney disease in African Americans

 b. Abbreviated Title (Length 26 characters): AA T2D-ESKD GWAS

2.   Writing Group:
     Writing group members:
Lead Author: Meijian Guan (WFU)
Senior Author: Maggie CY Ng (WFU)
ARIC Coauthors: Nora Franceschini (UNC), Joe Coresh (JHU), Adrienne Tin (JHU), Anna Kottgen (JHU)
Wake Forest Coauthors: Nichole Allred, Barry I. Freedman, Carl Langefeld, Jacob Keaton, Latchezar Dimitrov, Donald Bowden
FIND Coauthors: John Sedor, Rulan Parekh
MESA Coauthors: Jerome I. Rotter (LA BioMed), Stephen S. Rich (U Virginia), Ida Chen (LA BioMed)
JHS Coauthors: James G Wilson (U Mississippi)
CARDIA Coauthors: Laura J Rasmussen-Torvik (NorthWestern U)

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

Note that all co-authors are not yet assigned due to the multiple cohort study design. ARIC authors have approved.

 First author: Meijian Guan
 Address: Center for Genomics and Personalized Medicine Research
           Wake Forest School of Medicine
           Medical Center Blvd
           Winston-Salem, NC 27157

           Phone: 336-716-9358   Fax: 336-713-7566
           E-mail: mguan@wakehealth.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: **Nora Franceschini**
Address:
137 E. Franklin St., Suite 306
Campus Box 8050
Chapel Hill 27514
USA

Phone: 919-966-1305 Fax:
E-mail: noraf@unc.edu

3. **Timeline**: Data cleaning/QC Feb-Mar 2017
GWAS analysis Mar-April 2017
Manuscript writing begins May 2017

4. **Rationale**:

Background. Type 2 diabetes mellitus (T2D)-attributed end-stage kidney disease (ESKD), i.e. T2D-ESKD, is the most common etiology of chronic renal failure in the United States. Diabetic kidney disease (DKD), including ESKD and advanced chronic kidney disease (CKD), currently accounts for more than 40% of incident Medicare supported dialysis cases(USRDS, 2016). In 2014, 120,688 new ESKD cases were diagnosed and 97.4% of them received dialysis while only 2.6% of them started renal replacement therapy(USRDS, 2016). The mortality rates for ESKD, dialysis, and transplant patients, were 136, 166, and 30, per 1,000 patient per years, respectively in 2014(USRDS, 2016). In addition, ESKD has been a financial burden in the US – the annual expenditure for ESKD borne solely by Medicare exceeded $32.8 billion in 2014, 3.3% higher than in 2013 and accounting for 7.2% of all Medicare paid claims costs(USRDS, 2016). Actual costs are higher as this excludes non-Medicare costs and pre-dialysis care for CKD. Incident ESKD rates, predominantly related to diabetes, remained relatively flat in recent years despite medical advances in the treatment of hypertension and hyperglycemia.

African Americans have a disproportionately high risk of developing DKD. ESKD incidence rates per million people in the U.S. were 1,003 for African Americans, 301 (Caucasians) and 499 (Native Americans) in 2014(USRDS, 2016). One explanation put forth to explain these disparities is that racial or ethnic minorities have lower economic status and consequently more limited access to medical care. While this may explain some of the differences, several studies have controlled for socio-economic status and still observed significant differences between rates of renal complications when comparing Caucasians and African American patients(Spray et al. 1995; Freedman et al. 1995; Lei et al. 1998; Song et al. 2009). It remains unclear why only some of the many African American diabetic patients will ultimately develop kidney failure.

Many studies in type 1 and type 2 diabetes performed in multiple ethnic groups, have come to the conclusion that DKD has a significant genetic component (Maeda 2004; Pezzolesi et al. 2009; McDonough et al. 2011; Sandholm et al. 2012; Iyengar et al. 2015). A key observation is that the risk of developing kidney disease is not strongly related to measures of glycemic control or duration of diabetes. Subclinical measures of

nephropathy including urine albumin creatinine ratio (ACR) and microalbuminuria are relatively poor predictors of subsequent ESKD or falling glomerular filtration rate (GFR). In fact, urine ACR and GFR appear to be under independent genetic control (Langefeld et al. 2004; Placha et al. 2005). Thus the focus of genetic studies in our research group has been in people on dialysis, i.e. ESKD or with advanced DKD (defined as urine albumin/creatinine ratio (ACR) ≥300 mg/g and estimated glomerular filtration rate (eGFR) <30 ml/min/1.73m^2).

Rationale. We published the first genome wide association study (GWAS) of African American T2D-ESKD (McDonough et al. 2011). While the results of this study were of considerable interest, the clear message was that discovery of T2D-ESKD loci will require larger sample size. This represents a challenge since the absolute number of samples with our study criteria for inclusion as cases (e.g. on dialysis or advanced DKD) is relatively modest and derives primarily from our recruiting at Wake Forest School of Medicine over the last 20 years. In order to improve study power, in past few years, we have recruited and genotyped additional 1962 T2D-ESKD cases, 1766 non-diabetic non-nephropathy controls, 885 T2D non-nephropathy patients, as well as 2233 non-diabetic-ESKD samples using Affymetrix Axiom Biobank Genotyping Array (Axiom), Illumina Multi-Ethnic Genotyping Array (MEGA). As suggested by many studies, including external genotype data can remarkably increase statistical power and decrease the financial burden (Zhuang et al. 2010; Ho and Lange 2010; Mukherjee et al. 2011). Herein, we propose to include GWAS data from ARIC African Americans derived from the CARe study, primarily as **external control**, in creating as large as possible GWAS dataset. Along with genotype data from WFU, FIND and other CARe cohorts (JHS, CARDIA and MESA), we will carry out a GWAS meta-analysis to combine a total of 3591 T2D-ESKD cases, 2198 T2D non-nephropathy diabetic patients, and 6237 healthy controls. In addition, 2233 non-diabetic-ESKD patients will be used to exam the contribution of T2D-ESKD loci to ESKD that attributes to non-diabetic causes.

This manuscript proposal draws heavily from our prior participation in the CARe Diabetes phenotyping group where Dr. Ng is lead author of several papers and was primary analyst and Dr. Bowden was the group "convener". As part of this effort Dr. Ng has the primary African American genotype data from CARe including ARIC and has worked extensively with them. The CARe data include phenotype data for diabetes and kidney function for analysis in this study.

5.   Main Hypothesis/Study Questions:
Polymorphisms associated with African American T2D-ESKD can be identified by combined analysis of phenotypically informative African American GWAS samples

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. Discovery stage will be a meta-analysis combining association results from three genotyping platforms, specifically, Affymetrix Genome-wide Human SNP

array 6.0 (Affy6.0) (WFU, CARe cohorts and FIND), Affymetrix Axiom array (WFU), and Illumina MEGA array (WFU). Imputed GWAS data will be used for analysis within each platform. WFU, CARe cohorts and FIND genotypes will be combined directly rather than by meta-analysis because many cohorts primarily contribute control samples. A discrimination stage will be followed to test T2D-ESKD SNPs for association with T2D and removed. The third stage is to test T2D-ESKD associated SNPs in an additional 2233 African American non-diabetic-ESKD patients, to evaluate the contribution of T2D-ESKD SNPs to all-cause ESKD. We have also been actively worked with other collaborators, such as BioUV and BioMe cohorts, to recruit additional AA samples for replication of the top hits from the first stage GWAS.

In order to account for the confounding effect of *APOL1*, the same serial of analyses will be repeated (*APOL1*-negative model) by removing *APOL1* renal-risk-variant carriers and those missing *APOL1* genotypes.

Subjects: All ARIC cohort members who provided DNA samples and were included the GWAS analysis performed under the auspices of the CARe project.

- T2D-ESKD: T2D diagnosed $\geq$5 years prior to the onset of ESKD (on renal replacement therapy (RRT) OR urine ACR>300 mg/g at any visits (4 or 5) OR eGFR <30 ml/min/1.73 m^2 at any visits (1, 2, 4, or 5). T2D will be defined by 1) fasting glucose <126mg/dL, self-reported physician diagnosis of diabetes or on diabetes medications at ARIC visits 1 to 5, or 2) self-report diagnosis of diabetes or on diabetes medications at annual follow-up phone interview. RRT will be defined by linkage to USRDS up to 2013.
- T2D non-nephropathy: T2D diagnosed prior to or concurrent with eGFR $\geq$60 ml/min/1.73 m^2 in all available data (visits 1, 2, 4, 5) and UACR <30 mg/g in all available data (visits 4 and 5) and no ESRD based on linkage to USRDS up to 2013.
- Controls: non-diabetic and non-nephropathy (urine ACR <30 mg/g in all available data [visits 4 and 5], eGFR $\geq$60 ml/min/1.73m^2 in all available data [visits 1, 2, 4, 5] and no ESRD based on linkage to USRDS up to 2013
- Non-diabetic-ESKD: lacked diabetes or diabetes developed after initial renal replacement therapy. Diabetes and ESKD are defined as in the first bulletin.

Main Outcome Variables: T2D-ESKD

Secondary Outcome Variables: None

Covariates: Primary analysis: Age, sex, eigenvector

Analysis Plan and Methods:
Single variant analysis will be performed within each platform (Affy6.0, Axiom, and MEGA). Cohorts genotyped on Affy6.0 chip (WFU, CARe and FIND) will be directly combined. The number of diabetes affected subjects in the CARe cohorts is relatively modest (10-20%) and of these subjects, only a very small number of the samples have phenotypic data consistent with a stringent diagnosis of advanced DKD. Meta-analysis approach may lead to higher false positive rate and lower study power. Genotypes from

all platforms will be imputed to a combined cosmopolitan reference haplotype panel from the African Genome Variation Project (AGVP)(Gurdasani et al. 2015) and the 1000 Genomes Project phase 3 (1000 Genomes Project Consortium 2010). Another potential issue is the potential differences in population substructure within and across study cohorts. We will apply logistic mixed models implemented in the program GMMAT (Chen et al. 2016) for association analysis. This method is able to control for population structure and cryptic relatedness through incorporating a genetic relationship matrix (GRM) estimated from a set of high-quality autosomal SNPs as a random effect. Principal components analysis will also be performed and the first eigenvector correlate with ancestry will be adjusted in the model. A meta-analysis will follow to combine summary statistics from the three platforms using METAL (Willer et al. 2010). We will also implement a discrimination analysis by testing association between T2D non-nephropathy and control to identify T2D-ESKD signals that are driven by T2D only. Putative T2D-ESKD SNPs that pass discrimination stage will be tested for association with non-diabetic-ESKD in additional 2233 samples with ESKD attributed to chronic glomerular disease (e.g. FSGS), HIV-associated nephropathy, hypertension or unknown cause. Moreover, results can be cross referenced with additional external collaborators, such as BioUV and BioMe.

Two sets of variants in the *APOL1* gene are strong predictors of non-diabetic kidney disease in AAs. A secondary analysis will be performed to explore the influence of *APOL1* risk alleles by removing *APOL1*-risk-variant carriers and those missing *APOL1* genotypes. Individuals are considered *APOL1* renal-risk-variant carriers if they possessed two G1 alleles (rs60910145 G allele, rs73885319 G allele), two G2 alleles (6 base pair in-frame deletion), or were compound heterozygotes (one G1 and one G2 allele) (Genovese et al. 2010).

Finally, bioinformatic characterization will be conducted to uncover functional relevance underlying the identified associations. An expression quantitative trait loci (eQTL) query will be applied to identify the potential roles of T2D-ESKD SNPs in gene expression using database GTEx (http://www.gtexportal.org/home/). Annotation tools, such as ANNOVAR and VEP, will be used to indicate their potential impacts on protein coding, contribution to gene regulatory, predicted deleterious score, conservative score and related clinical outcomes. Furthermore, a bioinformatic tool, Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) (Pers et al. 2015), will be used to identify related pathways and enriched cell type/tissue of T2D-ESKD associated SNPs.

Brief Discussion of Power for Association Study
   Table 1 summarizes the power estimates for the discovery analysis of GWAS data (3591 cases, 6237 controls) to detect nominal association with $\alpha=1\times10^{-6}$ (Skol et al. 2006). This less stringent α was chosen since the initial aim is designed to identify SNPs for the subsequent replication study. These results show that we have good power (>70%) to detect most categories of SNPs stratified by effect sizes (1.2 to 1.4) and effect allele frequencies (0.1 to 0.5).

Table 1. Power analysis of discovery analysis of 3591 cases and 6237 controls (%)						
Prevalence=0.01	α=1e-6	Effect allele frequency				
		0.1	0.2	0.3	0.4	0.5

Odds Ratio	1.2	12	50	72	79	79
	1.25	39	86	96	98	97
	1.3	71	98	100	100	100
	1.35	91	100	100	100	100
	1.4	100	100	100	100	100

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? ___ Yes _X_ No
(This file ICTDER03 has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

Our presumption is that the CARe dataset is made up of individuals who meet this restriction.

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)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ___ Yes _X_ 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/

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.

Reference

Chen H, Wang C, Conomos MP, et al (2016) Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. Am J Hum Genet 98:653–666. doi: 10.1016/j.ajhg.2016.02.012

Freedman BI, Tuttle AB, Spray BJ (1995) Familial predisposition to nephropathy in African-Americans with non-insulin-dependent diabetes mellitus. Am J Kidney Dis 25:710–713. doi: 10.1016/0272-6386(95)90546-4

Genovese G, Friedman DJ, Ross MD, et al (2010) Association of Trypanolytic ApoL1 Variants with Kidney Disease in African Americans. Science 329:841–845. doi: 10.1126/science.1193032

Ho LA, Lange EM (2010) Using Public Control Genotype Data to Increase Power and Decrease Cost of Case-Control Genetic Association Studies. Hum Genet 128:597–608. doi: 10.1007/s00439-010-0880-x

Iyengar SK, Sedor JR, Freedman BI, et al (2015) Genome-Wide Association and Trans-ethnic Meta-Analysis for Advanced Diabetic Kidney Disease: Family Investigation of Nephropathy and Diabetes (FIND). PLoS Genet 11:e1005352. doi: 10.1371/journal.pgen.1005352

Langefeld CD, Beck SR, Bowden DW, et al (2004) Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. Am J Kidney Dis 43:796–800. doi: 10.1053/j.ajkd.2003.12.043

Lei HH, Perneger TV, Klag MJ, et al (1998) Familial aggregation of renal disease in a population-based case-control study. J Am Soc Nephrol 9:1270–1276.

Maeda S (2004) Genome-wide search for susceptibility gene to DKD by gene-based SNP. Diabetes Res Clin Pract 66:S45–S47. doi: 10.1016/j.diabres.2003.09.017

McDonough CW, Palmer ND, Hicks PJ, et al (2011) A GENOME WIDE ASSOCIATION STUDY FOR DKD GENES IN AFRICAN AMERICANS. Kidney Int 79:563–572. doi: 10.1038/ki.2010.467

Mukherjee S, Simon J, Bayuga S, et al (2011) Including additional controls from public databases improves the power of a genome-wide association study. Hum Hered 72:21–34. doi: 10.1159/000330149

Ng, Barry Freedman, Donald Bowden (2011) Genetic ancestry and population structure of geographically separated African American populations.

Pers TH, Karjalainen JM, Chan Y, et al (2015) Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 6:5890. doi: 10.1038/ncomms6890

Pezzolesi MG, Poznik GD, Mychaleckyj JC, et al (2009) Genome-wide association scan for DKD susceptibility genes in type 1 diabetes. Diabetes 58:1403–1410. doi: 10.2337/db08-1514

Placha G, Canani LH, Warram JH, Krolewski AS (2005) Evidence for different susceptibility genes for proteinuria and ESRD in type 2 diabetes. Adv Chronic Kidney Dis 12:155–169. doi: 10.1053/j.ackd.2005.02.002

Sandholm N, Salem RM, McKnight AJ, et al (2012) New Susceptibility Loci Associated with Kidney Disease in Type 1 Diabetes. PLoS Genet. doi: 10.1371/journal.pgen.1002921

Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38:209–213. doi: 10.1038/ng1706

Song E-Y, McClellan WM, McClellan A, et al (2009) Effect of Community Characteristics on Familial Clustering of End-Stage Renal Disease. Am J Nephrol 30:499–504. doi: 10.1159/000243716

Spray BJ, Atassi NG, Tuttle AB, Freedman BI (1995) Familial risk, age at onset, and cause of end-stage renal disease in white Americans. J Am Soc Nephrol JASN 5:1806–1810.

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191. doi: 10.1093/bioinformatics/btq340

Zhuang JJ, Zondervan K, Nyberg F, et al (2010) Optimizing the Power of Genome-Wide Association Studies by Using Publicly Available Reference Samples to Expand the Control Group. Genet Epidemiol 34:319–326. doi: 10.1002/gepi.20482

United States Renal Data System. 2016 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2016.