ARIC ManuScript Proposal #1975

PC Reviewed: 8/14/12  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: _____

1.a. **Full Title**: A GWAS Meta Analysis of African American Diabetic Nephropathy

b. **Abbreviated Title (Length 26 characters)**: AfAm Diabetic Nephropathy

2. **Writing Group**:
   Writing group members:
   Lead Author: Maggie C-Y Ng (WFU)
   Senior Author: Donald W. Bowden (WFU/FIND)
   ARIC Coauthors: Nora Franceschini (UNC), Suzette J. Bielinski (Mayo Clinic), Laura J. Rasmussen-Torvik (Northwestern)
   Coauthors: Nichole Allred, Barry I. Freedman Carl Langefeld from Wake Forest & FIND
   Jerome I. Rotter (Cedars-Sinai), Stephen S. Rich (University of Virginia) MESA
   Jason Bonomo, Jessica Bailey from Wake Forest
   Holly Kramer (Loyola), Joe Mychaleckyj (University of Virginia), Ida Chen (Cedars-Sinai) MESA
   James G Wilson (Univ Mississippi) Jackson Heart
   David Siscovick (U Washington) CARDIA
   Multiple other authors from multiple cohorts: Family Investigation of Nephropathy (FIND), CARDIA, Jackson Heart Study (we are submitting manuscript proposals to all contributors to the Candidate Gene Association Resource (CARe) study).

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _MCYN___ [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**: Maggie C-Y Ng
**Address**: Center for Human Genomics
  Wake Forest University School of Medicine
  Medical Center Blvd
  Winston-Salem, NC 27104

Phone: 336-713-7500  Fax: 336-713-7566
E-mail: mng@wakehealth.edu
3. **Timeline:**
   - Acquisition of data: August-November 2012
   - Data cleaning/QC: Dec-Jan 2013
   - GWAS analysis: Feb-Mar 2013
   - Replication analysis: April-July 2013
   - Manuscript writing begins: August 2013

4. **Rationale:**

   **Background.** Type 2 diabetes mellitus (T2D)-associated end-stage renal disease (ESRD), i.e. T2D-ESRD, is the most common etiology of chronic renal failure in the United States. Diabetic nephropathy (ESRD and advanced chronic kidney disease (CKD)) currently accounts for 44.3% of incident Medicare supported dialysis cases and patients with diabetic ESRD suffer nearly 50% mortality after 2 years of treatment with dialysis. In 2009, more than 113,600 Americans were diagnosed with ESRD and 570,000 received renal replacement therapy (USRDS 2011). The annual expenditure for ESRD borne solely by Medicare exceeded $29 billion in 2009, 3.1% higher than in 2008 and accounting for ≈6.5% of all Medicare expenditures. Actual costs are higher as this excludes non-Medicare costs and pre-dialysis care for CKD. Incident ESRD 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 diabetic nephropathy. ESRD incidence rates per million people in the U.S. were 1,010 for African Americans, 280 (Caucasians) and 489 (Native Americans) in 2006 (USRDS 2009). 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 (Freedman 1995; Spray 1995; Lei 1998; Song 2009). It remains unclear why only some of the many African American diabetic patients will ultimately develop renal failure.

   Many studies in type 1 and type 2 diabetes performed in multiple ethnic groups, have come to the conclusion that diabetic nephropathy has a significant genetic component (e.g., Seaquist 1989; Borch-Johnsen 1992; DCCT 1997; Petit 1990; Freedman 1995, 1997, 2005). A key observation is that the risk of developing renal disease is not strongly
related to measures of glycemic control or duration of diabetes. Subclinical measures of nephropathy including urine albumin creatine ratio (ACR) and microalbuminuria are relatively poor predictors of subsequent ESRD or falling glomerular filtration rate (GFR). In fact, urine ACR and GFR appear to be under independent genetic control (Placha 2005; Langefeld 2004). Thus the focus of genetic studies in our research group has been in people on dialysis, i.e. ESRD or with diabetic nephropathy (defined as urine albumin/creatinine ratio (ACR) ≥30 mg/g and estimated glomerular filtration rate (eGFR) <60 ml/min/1.73m²).

Rationale. We published the first genome wide association study (GWAS) of African American diabetic nephropathy (McDonough 2011). While the results of this study were of considerable interest, the clear message was that discovery of T2D-ESRD genes will require larger samples. This represents a challenge since the absolute number of samples with our study criteria for inclusion as cases (e.g. on dialysis; ESRD) is relatively modest and derives primarily from our recruiting at Wake Forest over the last 20 years. We propose to include GWAS data from ARIC African Americans as one element in creating as large as possible discovery GWAS dataset. As part of R01 DK53591 we propose to carry out a GWAS meta-analysis of data from >2000 African American T2D-ESRD cases and 8000 non-diabetic and non-nephropathy controls, followed by a similarly powered second, replication stage.

This manuscript proposal draws heavily from our 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 and imputed African American genotype data from CARe including ARIC and has worked extensively with them. The CARe data include phenotype data not only for diabetes diagnosis, but kidney health which can be used for analysis in this study.

5. Main Hypothesis/Study Questions:
Polymorphisms associated with African American diabetic nephropathy 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 issues. This analysis will depart from the conventional meta analysis approaches since the primary data from multiple studies will be combined for analysis. This is because 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 diabetic nephropathy (in ARIC approximately 100 on preliminary evaluation out of 1204 in the CARe ARIC dataset this is likely to be reduced on more rigorous examination). There are several issues that arise with such a study design. First are potential differences in population structure between different study samples. While at a first level this can be accounted for by use of principal components adjustment, we have also carried out a detailed analysis of this issue in CARe African Americans (Ng et al. 2011; Ng et al in preparation) which demonstrates African American samples from the different CARe cohorts (and by inference all African
Americans) are analytically indistinguishable from each other. A second issue is that combining datasets will introduce serious problems through differentially missing data from study to study. We account for this possibility by using analysis approaches that account for and measure differential missing data (e.g. implemented in analysis pipeline for SNPGWA, www.phs.wfubmc.edu). Finally one might ask why not get the data from dbGAP? Although not widely appreciated, only the primary genotyping data is in dbGAP. The imputed genotypes are not in dbGAP, thus this request to ARIC.

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

- Diabetic nephropathy: diagnosed type 2 diabetes on renal replacement therapy OR urine ACR>300 mg/g OR eGFR <60 ml/min/1.73 m² [Stage 3 CKD (GFR 30-59), Stage 4 CKD (GFR 15-29) and Stage 5 CKD (<15 = on or near dialysis].
- Controls: non-diabetic, non-nephropathy (urine ACR<30 mg/g, eGFR >60 ml/min/m²)

Genotyping data: Directly genotyped or HapMap 2.0 imputed genotype data from ARIC African Americans. Most conveniently those in the CARe dataset. If the ARIC coordinating center has additional data, we would be happy to hear about it.

Main Outcome Variables: Diabetic nephropathy (defined above)

Secondary Outcome Variables: None

Covariates: Primary analysis: Age, sex, principal components, APOL1/MYH9 genotype
Secondary analysis: Age, sex, principal components,
Tertiary analysis: diabetes status, BMI, hypertension,

Analysis Plan and Methods:
Cases will be derived primarily from the Wake Forest African American Diabetic Nephropathy Study and the Family Investigation of Diabetic Nephropathy (the latter a case only sample) supplemented with cases from the collaborating CARe cohorts. The total number of cases will number approximately 2200 (1675 cases from WFU and FIND, the rest from collaborating cohorts). Controls (non-diabetic, non-nephropathy controls) will be derived from GWAS data from the WFU collection (approximately 1025) and additional samples from CARe for a total of 8000 controls. Standard GWAS QC will be performed. Kinship coefficients will be evaluated to reduce relatedness. Analysis will be simple logistic regression with adjustment for age, sex, principle components, and APOL1/MYH9 genotype. (APOL1 coding variants are powerful contributors to non-diabetic ESRD and it is essential to control for their impact). Secondary analysis will explore the influence of APOL1/MYH9. Tertiary analysis will explore the influence of additional covariates as listed above. The primary analysis tool will be SNPGWA (current v4.0). SNPGWA computes a series of estimates and tests appropriate for GWAS and smaller SNP sets based on logistic regression models. SNPGWA v4.0 tests each SNP for departures from Hardy-Weinberg equilibrium and genotypic association is computed: two degrees of freedom overall test for 2x3 tables, dominant model, additive model (Cochran-Armitage trend test), recessive model, and
lack-of-fit to an additive model. SNPGWA 4.0 incorporates covariate adjustment and covariate adjusted out moving window (2 SNP, 3 SNP) haplotype analysis. The program also has machine learning modules suited for SNP-SNP interactions. Routinely we carry out two or more analyses with increasing covariate adjustment. A sensitivity analysis is computed to determine what the most appropriate principal components are to adjust. We note that the most appropriate PCs might not be just the first $k$ PCs and search deeper into the list of PCs to minimize the inflation factor, while attempting to thoughtfully minimize the total number of PCs. Once the PCs are determined, subsequent analyses include additional covariates, e.g. age and gender.

After analysis interpretation of the primary data, replication will take place in a second set of cases and controls defined in the same way made up of Wake Forest cases and controls and samples from other studies broadly comparable to CARe. As is likely clear, this analysis does not distinguish clearly between diabetic nephropathy loci and diabetes loci. In practice this does not present a problem. We are part of the MEDIA African American type 2 diabetes genetics consortium consisting of over 27,000 DNA samples for T2D analysis. Any result can quickly be cross referenced with MEDIA. Second we have a substantial collection of African American DNAs which have T2D, but not nephropathy, and nephropathy without T2D. Differential analysis in these groups was a critical part of the study design in McDonough et al (2011).

**Brief Discussion of Power for Association Study**

Table 1 summarizes the power estimates for the combined analysis of GWAS data (estimated 2274 GWAS cases, 8,000 controls) to detect nominal association ($\alpha=1\times10^{-5}$) based on Skol (2006). This less stringent $\alpha$ was chosen since the initial aim is designed to identify SNPs for the subsequent replication study. These results show that we have good to excellent power to detect effect sizes consistent with other complex genetic traits across a range of MAF. Table 2 summarizes the combined power for the combined GWAS analysis and replication to detect loci with $\alpha=5.0\times10^{-8}$.

<table>
<thead>
<tr>
<th>Type I Error</th>
<th>Power $q=0.9$</th>
<th>$q=0.8$</th>
<th>$q=0.7$</th>
<th>$q=0.6$</th>
<th>$q=0.5$</th>
<th>$q=0.4$</th>
<th>$q=0.3$</th>
<th>$q=0.2$</th>
<th>$q=0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.00E-05$</td>
<td>1.47</td>
<td>1.29</td>
<td>1.23</td>
<td>1.21</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.21</td>
<td>1.32</td>
</tr>
<tr>
<td>$5.00E-08$</td>
<td>1.38</td>
<td>1.23</td>
<td>1.19</td>
<td>1.17</td>
<td>1.17</td>
<td>1.16</td>
<td>1.17</td>
<td>1.19</td>
<td>1.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type I Error</th>
<th>Power $q=0.9$</th>
<th>$q=0.8$</th>
<th>$q=0.7$</th>
<th>$q=0.6$</th>
<th>$q=0.5$</th>
<th>$q=0.4$</th>
<th>$q=0.3$</th>
<th>$q=0.2$</th>
<th>$q=0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.00E-05$</td>
<td>1.61</td>
<td>1.36</td>
<td>1.28</td>
<td>1.25</td>
<td>1.24</td>
<td>1.24</td>
<td>1.25</td>
<td>1.28</td>
<td>1.37</td>
</tr>
<tr>
<td>$5.00E-08$</td>
<td>1.42</td>
<td>1.26</td>
<td>1.21</td>
<td>1.18</td>
<td>1.18</td>
<td>1.17</td>
<td>1.18</td>
<td>1.20</td>
<td>1.27</td>
</tr>
</tbody>
</table>

---

**Table 1. Power analysis: combined GWAS analysis of 2274 T2D-ESRD cases and 8000 controls ($q=$ minor allele frequency)**

**Table 2. Power analysis: combined GWAS analysis of 4744 T2D-ESRD cases and 16000 controls ($q=$ minor allele frequency)**
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/](http://publicaccess.nih.gov/) are posted in [http://www.cscc.unc.edu/aric/index.php](http://www.cscc.unc.edu/aric/index.php), under Publications, Policies & Forms. [http://publicaccess.nih.gov/submit_process_journals.htm](http://publicaccess.nih.gov/submit_process_journals.htm) shows you which journals automatically upload articles to Pubmed central.

References


