1.a. Full Title: Joint Meta-Analysis of Genome-Wide Interactions for Type 2 Diabetes in African Americans

b. Abbreviated Title (Length 26 characters): African American T2D Interaction Genetics

2. Writing Group:
   Writing group members:
   Lead Author: Jacob M Keaton (WFU)
   Senior Author: Donald W. Bowden (WFU)
   Coauthors and cohort: Maggie C-Y Ng (WFU), Nichole Allred (WFU), Suzette J Bielinski (Mayo Clinic), Laura J Rasmussen-Torvik (Northwestern University), David Couper (UNC), James S Pankow (University of Minnesota), Eric Boerwinkle (University of Texas Health Science Center at Houston)
   Multiple other authors from multiple cohorts: MESA, 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. __JK__ [please confirm with your initials electronically or in writing]

First author: Jacob M Keaton
Address:
Center for Genomics and Personalized Medicine Research
Wake Forest University School of Medicine
Medical Center Blvd.
Winston-Salem, NC 27157

Phone: 336-713-7529           Fax: 336-713-7566
E-mail: jkeaton@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: James Pankow
Address: Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

Phone:               Fax:                E-mail: panko001@umn.edu
3. Timeline:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition of data</td>
<td>August 2014</td>
</tr>
<tr>
<td>Data cleaning/QC</td>
<td>August-October 2014</td>
</tr>
<tr>
<td>Analysis</td>
<td>October-November 2014</td>
</tr>
<tr>
<td>Manuscript writing begins</td>
<td>December 2014 – January 2015</td>
</tr>
</tbody>
</table>

4. Rationale:

**Background.** Type 2 diabetes (T2D) is one of this country’s most troubling health dilemmas with more than 25 million affected individuals in the United States alone. T2D is the seventh leading cause of death in the U.S., and the risk of death for an affected individual is about twice that of an age-matched unaffected individual. Affected individuals are at high risk for several comorbidities including cardiovascular disease, hypertension, kidney disease, and nervous system damage. T2D has a significant impact on the U.S. economy and healthcare system with an estimated annual cost for treatment and lost productivity exceeding $240 billion. Ethnic disparities in T2D prevalence are well documented, with one of the largest observable differences occurring between individuals of European and African descent. The most recent data indicates that T2D prevalence among African Americans (12.6%) is much higher than among European Americans (7.1%). Several studies have shown that the higher risk in African Americans persists even after adjustment for known environmental and socioeconomic risk factors such as body mass index (BMI), physical activity, education level, and income. Genetic risk factors that may underlie the disparity in diabetes prevalence are not well understood.

Although common variants examined in genome-wide association studies (GWAS) have identified ~70 loci associated with T2D risk, these variants explain only a fraction of T2D heritability. With these limitations, many avenues of investigation have not been effectively pursued. A portion of the missing heritability may be explained by genetic epistasis, a phenomenon that occurs when the effect of a genetic risk factor is modified by other factors in an individual’s genetic background. GWAS have largely ignored epistatic contributions to T2D risk due to the heavy computational and multiple testing burden of exhaustive analytical approaches. For example, testing all two-locus interactions using a regression-based approach in a panel of 500,000 SNPs would require the evaluation of several billion models. To further illustrate this point, a recent genome-wide scan for two-locus interactions in the WTCCC T2D GWAS data did not reveal any significant epistatic signals at a Bonferroni-corrected p-value threshold of $2.14 \times 10^{-11}$ after adjusting for the main effects of the most strongly associated T2D locus, TCF7L2. However, much of this burden may be alleviated through hypothesis-driven approaches.

T2D is characterized by two major metabolic defects: impaired insulin secretion and insulin resistance. Impaired insulin secretion arises from pancreatic beta cell dysfunction, and insulin resistance in hepatic, skeletal muscle, and other peripheral tissues leads to decreased plasma glucose uptake. Due to the underlying bimodal pathophysiology, T2D is a particularly well-suited disease model for hypothesis-driven investigation of epistatic interactions.
interactions. Genetic insults to both insulin secretion and insulin sensitivity metabolic pathways may jointly increase an individual’s genetic risk of T2D. Considering the low degree of linkage disequilibrium in admixed populations in conjunction with higher prevalence rates, African Americans may be an ideal population for the study of genetic interactions that contribute to T2D risk.

**Rationale.** A singular observation of the T2D associations observed to date has been the high representation of loci with influence on insulin secretion. Thus interaction analysis of these variants with variants across the genome suggests the potential to identify insulin resistance genes. In preliminary analysis, SNPs in the MTNR1B gene (encoding melatonin receptor 1B) are powerfully associated with acute insulin response (AIR), a measure of first phase insulin secretion. The top AIR association in MTNR1B was with the intronic SNP rs10830963 (p=1.2x10^{-5}). Initial analyses have shown promising evidence of interaction with genes related to circadian regulation (melatonin receptor is implicated in circadian rhythms). These results and additional results from other genes implicated in insulin secretion suggest expansion of the analysis to additional cohorts, including ARIC.

5. **Main Hypothesis/Study Questions:**

*Variants associated with measures of insulin secretion and insulin sensitivity in African Americans interact to confer T2D genetic risk.*

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).**

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

- Type 2 diabetes: diagnosed type 2 diabetes after age 25 with at least one of the following: fasting glucose ≥126 mg/dL, 2-hr oral glucose tolerance test glucose ≥200 mg/dL, random glucose ≥200 mg/dL, use of oral hypoglycemic agents and/or insulin, or physician diagnosed diabetes using all visits
- Controls: non-diabetic, normal glucose tolerance over age 25 with fasting glucose <100 mg/dL and 2-hr oral glucose tolerance test glucose <140 mg/dL (if available) without reported use of diabetes medications using all visits

**Genotyping data:** Directly genotyped or 1000 Genomes Phase 1 imputed genotype data from ARIC African Americans in the CARe dataset. Our group already has the GWAS data through a previous CARe study. We do not require genotype data from the coordinating center.

**Main Outcome Variables:** Type 2 diabetes mellitus

**Secondary Outcome Variables:** None
Covariates:
Association analysis: Age, gender, principal components
Primary interaction analysis: Age, gender, principal components, BMI
Secondary interaction analysis: Age, gender, principal components

**Brief Methods/Analysis Plan**

Cases will be derived primarily from the Wake Forest African American type 2 diabetes GWAS collection supplemented with cases from the collaborating CARe cohorts. The total number of cases will number approximately 2700 (900 from WFU, the remainder from collaborating cohorts). Controls (non-diabetic, normal glucose tolerance controls) will be derived from GWAS data from the WFU collection (approximately 900) and additional samples from CARe for a total of 4100 controls. Standard GWAS QC will be performed (much if not all has already been performed in support of CARe manuscripts). Association analysis will be simple logistic regression with adjustment for age, gender, and principle components. SNPs 1) associated with T2D and measures of glucose homeostasis in prior African American and European American GWAS and 2) exhibiting nominal association with T2D in the WFU and collaborating CARe cohorts will be tested for interaction. Interaction analysis will be multiple logistic regression including marginal effects of a test SNP and an interacting SNP, interaction between these two SNPs, and adjustment for age, gender, principal components, and BMI. Secondary interaction analysis will explore the influence of BMI. The primary analysis tools will be SNPTEST (current v2), ProbABEL (current v0.4.3), and METAL (current v2011-03-25). SNPTEST computes a series of estimates and tests appropriate for GWAS and smaller SNP sets based on logistic regression models. SNPTEST v2 generates summary statistics for each SNP (MAF, genotype counts, info), tests each SNP for departures from Hardy-Weinberg equilibrium, and genotypic association is computed under five models: additive, dominant, recessive, general, and heterozygote. SNPTEST v2 incorporates covariate adjustment and can handle allelic dosages for imputed SNPs. ProbABEL v0.4.3 tests for association of marginal and interaction effects and computes a robust variance-covariance matrix of parameter estimates. Joint effects (i.e. marginal effect of the interacting SNP combined with the interaction effect) will be analyzed in the context of meta-analysis using the joint meta-analysis (JMA) method as described by Manning et al. The JMA method summarizes regression coefficients for the SNP and interaction terms from a logistic regression model while combining cohorts by fixed-effects inverse variance weighting. JMA will be implemented in METAL using a patch provided by the authors of this method.

After analysis interpretation of the primary data, replication will take place in a second set of cases and controls defined in the same way and composed of Wake Forest cases and controls and samples from other studies broadly comparable to CARe.

**Brief Discussion of Power for Interaction Study**

Figure 1 summarizes the power estimates for the combined interaction analysis of GWAS data (estimated 2725 cases, 4167 controls) to detect association at different significance thresholds ($\alpha=5 \times 10^{-2}$, $\alpha=5 \times 10^{-4}$, $\alpha=5 \times 10^{-6}$, $\alpha=5 \times 10^{-8}$). These thresholds were chosen to reflect power accounting for multiple testing burdens of 1, 100, 10000,
and 1 million tests, respectively. This model assumes additive effects, MAF=35%, and OR=1.1 for both SNPs and a population risk of 12.6%. Power calculations were performed in Quanto (http://hydra.usc.edu/gxe/). These results show that we have good to excellent power to detect effect sizes consistent with other complex genetic traits.

7.a. Will the data be used for non-CVD analysis in this manuscript? **Yes**

   **Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”?**  **Yes**

   **X**  **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? **Yes**

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”?  **Yes**

   **X**  **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**  **Yes**

   **X**  **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.