1.a. Full Title: Multi-Ethnic Genome-wide Association Study of Diabetic Retinopathy using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control

b. Abbreviated Title (Length 26 characters): Multi-ethnic GWAS Diabetic Retinopathy

2. Writing Group:

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

First author: Samuela Pollack
Address: 243 Charles Street
          Boston, MA 02114
Phone: 617-573-4279          Fax: 617-573-3011
E-mail: spollack@jimmy.harvard.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: Lucia Sobrin
Address: 243 Charles Street, Boston, MA, 02114
Phone: 6175734279          Fax: 6175733011
E-mail: lucia_sobrin@meei.harvard.edu

3. Timeline: We plan to complete and submit the manuscript within 5 months.
4. Rationale:

Diabetic retinopathy (DR) is the leading cause of blindness in Americans between 20-74 years of age. Established risk factors (RFs) for DR include longer duration of diabetes (DoD) and poor glycemic control. Genetic factors are also believed to play an important role, with heritability estimates of up to 27% for overall DR and 52% for proliferative diabetic retinopathy (PDR). Several candidate gene and genome-wide association studies (GWAS) have now been conducted. While polymorphisms in several genes have been reported to be associated with DR, few associations have been convincingly replicated across studies.

There are several reasons why genetic association studies have yet to yield consistent findings. The genetic effects are likely to be modest and require large sample sizes to be identified. Data sets from diverse populations have not yet been combined because of the underlying technical challenges. Previous studies have compared diabetic individuals with DR to diabetic individuals without DR, without optimally accounting for covariates such as the DoD and glycemic control. Because DoD and glycemic control each strongly affect the risk of DR, individuals who progress to DR a short time after diabetes diagnosis and with tighter blood sugar control are likely to carry higher genetic risks than individuals who progress to DR a long time after diabetes diagnosis and with worse blood sugar control. Conversely, patients who do not develop DR after having diabetes for several decades with poor blood sugar control might harbor protective genetic variants. These differences can provide a means to increase statistical power as explained below.

Standard approaches that adjust for covariates all fail to capture the increase in statistical power that is available when case-control studies include data on covariates that affect genetic risk. To address incorporation of covariates to optimize statistical power, a method based on the classical technique of liability threshold (LT) modeling was developed which outperforms standard approaches by a wide margin in simulation and in empirical case-control studies of T2D, prostate cancer, breast cancer, rheumatoid arthritis, macular degeneration, and end stage renal disease.

The idea behind LT modeling is that there exists an underlying unobserved quantitative trait, which is function of covariates, a genetic effect being evaluated, and statistical noise from a standard normal distribution. An individual is a case if the quantitative trait is above some threshold, and a control otherwise. Only case-control status is observed, and from this information the posterior mean value of the quantitative trait after adjusting for covariates can be computed for each individual. Standard association tests for quantitative traits can then be applied to these posterior mean values. Application of these methods to DR could substantially increase the ability to discover genes. We expect that subjects with DR and short DoD and/or good glycemic control harbor greater genetic risk than those with DR and long DoD and poor glycemic control; LT modeling will leverage this understanding of covariates’ effects to maximize this study’s power.

Therefore, the purpose of this study is to identify genetic variants associated to DR by (1) assembling a large sample size through inclusion of patients of multiple ethnicities in discovery and replication and (2) incorporating the covariates of DoD and glycemic control via LT modeling to increase power.
5. **Main Hypothesis/Study Questions:** A well-powered, multi-ethnic genome-wide association study of diabetic retinopathy can identify genetic loci associated with diabetic retinopathy.

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).**

*Discovery Sample Description*

The discovery sample will consist of a consortium of eleven studies of diabetic retinopathy in participants of European or African American ancestry. Inclusion criteria for this study were (1) T2D, (2) Caucasian or African American ethnicity and (3) availability of both DR phenotypes and genome-wide genotypes. Diabetes was defined using American Diabetes Association criteria and T2D was defined as onset of diabetes after age 30 years. Table 1 summarizes the diabetic retinopathy phenotyping protocols and covariate measures available for each study included in the discovery sample. Additional details of the phenotyping protocols have been previously described for each of these studies. With the exception of the Wake Forest University (WFU) cohort, all of the discovery cohort samples used photography with ETDRS grading for DR phenotyping. The WFU cohort based DR diagnosis on questionnaire items of previous laser treatment for DR. Therefore, WFU could only be included in the Phenotype 2 definition (see below) analyses as it was impossible to determine the degree of retinopathy in patients without laser treatment. RISE/RIDE did not include any patients without retinopathy and therefore could only be incorporated in the Phenotype 2 analyses as well.

*Case/Control Definitions*

We will examine four different phenotypes (Table 2). There are several reasons for examining four different phenotypes. First, because previous GWAS for DR have examined varying case definitions, it will allow us to examine and replicate previous findings across a range of DR definitions from different studies. For replication of our own findings, it will allow the greatest opportunity for well-powered replication given that existing GWAS results exist for different phenotype definitions in different studies. Phenotypes 1 and 2 have the advantages of incorporating the largest sample sizes into the analyses, thus maximizing power. They will help distinguish if there are genetic variants that increase risk at the earliest stage of retinopathy development (Phenotype 1) vs. advanced retinopathy (Phenotype 2). Phenotype 3, although able to only take advantage of a subset of the sample available, will answer whether there are genetic variants that influence retinopathy at its intermediate stage. Because many of the cohorts were population-based studies, they do not have large numbers of patients with advanced DR. Phenotype 3 would preserve sample size while trying to examine a more rigorous case definition than Phenotype 1. Phenotype 4 is an extreme-of-phenotype analysis with the expectation that the decrease in power from smaller sample size would be compensated by enrichment for genetic variants in participants at the extremes of DR development.

*Covariate Definitions*

The glycemic control measures available for each cohort were either Hemoglobin A1c (HbA1c), fasting blood sugar (FBS), or both. DoD was obtained by patient self-report for all participants.
Table 1. Studies to be included in the discovery sample

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Diabetes Type</th>
<th># of Eyes/# of Fields/ Size of Fields Photographed</th>
<th>Diabetes Duration</th>
<th>Glycemic Control Measure</th>
<th>Genotyping Platform</th>
<th>SNPs genotyped</th>
<th>SNPs after QC</th>
<th>SNPs after imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAPDR</td>
<td>AA</td>
<td>2</td>
<td>2/7/30 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>883851</td>
<td>579834</td>
<td>12194970</td>
</tr>
<tr>
<td>AGES*</td>
<td>Caucasian</td>
<td>2</td>
<td>2/2/45 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Illum 370</td>
<td>353202</td>
<td>304677</td>
<td>7811063</td>
</tr>
<tr>
<td>ARIC</td>
<td>AA</td>
<td>2</td>
<td>1/1/45 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>603146</td>
<td>505092</td>
<td>12308412</td>
</tr>
<tr>
<td></td>
<td>Caucasian</td>
<td>2</td>
<td>1/1/45 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>603146</td>
<td>505092</td>
<td>6665985</td>
</tr>
<tr>
<td>AUST</td>
<td>Caucasian</td>
<td>2</td>
<td>NA‡</td>
<td>Y</td>
<td>HbA1C</td>
<td>Illum 670</td>
<td>605258</td>
<td>603947</td>
<td>6933006</td>
</tr>
<tr>
<td>BMES</td>
<td>Caucasian</td>
<td>2</td>
<td>2/5/30 deg.</td>
<td>Y</td>
<td>FBS</td>
<td>Illumina Express</td>
<td>544799</td>
<td>525303</td>
<td>6886208</td>
</tr>
<tr>
<td>CHS</td>
<td>AA</td>
<td>2</td>
<td>1/1/45 deg.</td>
<td>Y</td>
<td>FBS</td>
<td>Illum Express</td>
<td>886457</td>
<td>844504</td>
<td>8501510</td>
</tr>
<tr>
<td></td>
<td>Caucasian</td>
<td>2</td>
<td>1/1/45 deg.</td>
<td>Y</td>
<td>FBS</td>
<td>Illum 670</td>
<td>886457</td>
<td>844504</td>
<td>2168325</td>
</tr>
<tr>
<td>FIND-Eye*</td>
<td>AA</td>
<td>2</td>
<td>2/2/45 deg.†</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>Unknown</td>
<td>798920</td>
<td>19463075</td>
</tr>
<tr>
<td></td>
<td>Caucasian</td>
<td>2</td>
<td>2/2/45 deg.†</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>Unknown</td>
<td>690150</td>
<td>8813873</td>
</tr>
<tr>
<td>JHS</td>
<td>AA</td>
<td>2</td>
<td>2/7/30 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>633267</td>
<td>580336</td>
<td>12441123</td>
</tr>
<tr>
<td>MESA</td>
<td>AA</td>
<td>2</td>
<td>2/2/45 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>735518</td>
<td>671745</td>
<td>12733230</td>
</tr>
<tr>
<td></td>
<td>Caucasian</td>
<td>2</td>
<td>2/2/45 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Affy 6.0</td>
<td>735518</td>
<td>671745</td>
<td>6727876</td>
</tr>
<tr>
<td>RISE/RIDE</td>
<td>Caucasian</td>
<td>2</td>
<td>2/7/30 deg.</td>
<td>Y</td>
<td>HbA1C</td>
<td>Illum 1M</td>
<td>2362143</td>
<td>§</td>
<td>§</td>
</tr>
<tr>
<td>WFU</td>
<td>AA</td>
<td>2</td>
<td>NA‡</td>
<td>Y</td>
<td>HbA1C</td>
<td>Illum 6.0</td>
<td>868155</td>
<td>817755</td>
<td>7044061</td>
</tr>
</tbody>
</table>

**Crls= Controls, AAPDR = African American Proliferative Diabetic Retinopathy Study, AGES = Age, Gene/Environment Susceptibility Study, ARIC = Atherosclerosis Risk In Communities Study, AUST= Australian Genetics of Diabetic Retinopathy Study, BMES = Blue Mountains Eye Study, CHS= Cardiovascular Health Study, FIND-Eye = Family Study of Nephropathy and Diabetes-Eye, JHS = Jackson Heart Study, MESA = Multi-Ethnic Study of Atherosclerosis, RIDE/RISE= Ranibizumab Injection in Subjects with Clinically Significant Macular Edema with Center Involvement Secondary to Diabetes, WFU=Wake Forest University, AA=African American, Illum=Illumina, Affy=Affymetrix, NA=not available, Y=information on diabetes duration is available, HbA1C=hemoglobin A1C, FBS=fasting blood sugar, deg.= degrees, SNPs= single nucleotide polymorphisms, QC=quality control

* Cohorts without access to raw genotype information
† Not all FIND-Eye subjects had photographs but all participants had harmonization of exam and clinical data to an ETDRS score.
‡ The AUST study used examination by an ophthalmologist to ascertain diabetic retinopathy. The WFU study used a questionnaire to ascertain diabetic retinopathy.
§ The RISE and RIDE studies were imputed separately, RISE has 1174886 SNPs after QC and 7417483 SNPs after imputation. RIDE had 1607163 SNPs after QC and 7888767 SNPs after imputation.

Table 2. Four phenotype definitions and the number of samples available for discovery for each phenotype

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Control Definition ETDRS Score</th>
<th>African American</th>
<th>Caucasian</th>
<th>Case Definition ETDRS Score</th>
<th>African American</th>
<th>Caucasian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype 1</td>
<td>&lt; 14</td>
<td>941</td>
<td>1970</td>
<td>≥ 14</td>
<td>911</td>
<td>1079</td>
</tr>
<tr>
<td>Phenotype 2</td>
<td>&lt; 60</td>
<td>1514</td>
<td>2848</td>
<td>≥ 60</td>
<td>1097</td>
<td>398</td>
</tr>
<tr>
<td>Phenotype 3</td>
<td>&lt; 14</td>
<td>941</td>
<td>1970</td>
<td>≥ 30</td>
<td>768</td>
<td>644</td>
</tr>
<tr>
<td>Phenotype 4</td>
<td>&lt; 14</td>
<td>941</td>
<td>1970</td>
<td>≥ 60</td>
<td>1097</td>
<td>398</td>
</tr>
</tbody>
</table>
**Genotyping**

The genotyping platform and the number of SNPs genotyped for each cohort is summarized in Table 1.

**Initial Quality Control and Data Filtering**

The following data filtering and quality-control measures will be employed. First, for each study (e.g. WFU) separately, we will remove any SNP with minor allele frequency (MAF) <0.1% and with >1% missing data, and that does not have a human genome build 37 physical position and rs id. Next, for each cohort separately, we will infer the continental (EUR, ASI, AFR) ancestry proportions of each individual using SNPweights. Caucasian individuals will be defined based on >90% European ancestry. African American individuals were defined based on >50% African ancestry and <10% East Asian ancestry. If there are individuals with African ancestry between 10% and 50% and less than 10% East Asian ancestry, and these will be removed for the primary analyses.

From this point forward, all data filtering and quality control steps will continue to be performed for each study separately and in addition will be performed separately for African Americans and Caucasians as defined above if that study had both African Americans and Caucasians. A Hardy-Weinberg filter will be applied for each SNP. We will compare the actual number of heterozygotes to the expected number of heterozygotes based on MAF. For Caucasians, we will remove the SNP if the actual number of heterozygotes is greater than 5 standard deviations from expected. In African Americans, we will only remove the SNP if there is an excess of heterozygotes (greater than 5 standard deviations from expected). This is because African Americans are a highly structured population and a reduction in heterozygotes can be expected to occur even with no quality control issues.

We will remove one of each pair of duplicate samples. Two of the studies recruited patients from overlapping geographic areas and there was a higher chance for duplicate samples between them. The Jackson Heart Study (JHS) overlapped in recruitment area with the Atherosclerosis Risk in Communities (ARIC) Study. For these studies, we will merge the cohorts prior to the duplicate check, perform the duplicate check on the reduced set of merged SNPs, remove the duplicate sample, and then restore the full set of cohort-specific SNPs to each cohort before proceeding with the analyses below.

**1000 Genomes Imputation**

Next, imputation will be done with the 2.3.0 version of impute2. We will use the cosmopolitan reference sample based on 1,092 1000 Genomes samples. After imputation, any SNP with low imputation accuracy (INFO <0.6), MAF<1%, or < 5 copies in imputed data will be removed from that cohort. SNPs that are retained in only a subset of cohorts were included in the meta-analysis. Association analysis will be run on rounded dosages, retaining highly uncertain genotypes (e.g. P<0.9) so long as the SNP has imputation $r^2 \geq 0.6$.

**Related individuals**

After imputation, we will remove one of each pair of related individuals separately for each cohort, except for JHS and ARIC where overlap was expected as described above. We will remove one of each pair of individuals with genome-wide relatedness >0.10 with the smartrel program in EIGENSOFT, see [http://www.hsph.harvard.edu/alkes-price/software](http://www.hsph.harvard.edu/alkes-price/software).
**Association Analyses**

The quality control steps and analyses will be performed for all studies centrally at Harvard Medical School except for the Family Investigation of Nephropathy and Diabetes (FIND) and Age/Gene Susceptibility Study (AGES). For these two studies, the consents and permissions do not allow for sharing of raw genotypes so these cohorts performed the above analyses at the primary study site and shared their data for meta-analysis. Because there is also overlap between WFU and FIND with regards to recruitment area, the FIND study will identify duplicate individuals and those will be removed from the WFU study for the analyses.

We will analyze Caucasians and African American samples separately. We will run principal components analysis using typed SNPs only separately for African Americans and Caucasians for each study and for each phenotype definition using EIGENSOFT. We will use 5 principal components (PCs) for the Caucasian analysis. Separately, we will also use 5 PCs for the African American analysis. We will used both the Armitage Trend Test\(^\text{18}\) with PCs and the LTSCORE test with PCs.\(^\text{38}\) Both tests were implemented in version 2.0 of LTSOFT software (http://www.hsph.harvard.edu/alkes-price/software).

**Liability Threshold Modeling**

The LTSCORE test has been previously described.\(^\text{26}\) Let \(x\) be a phenotypic covariate with mean \(\bar{x}\), and let \(\phi(x) = \alpha(x - \bar{x}) + \beta + \varepsilon\) be an underlying quantitative phenotype such that individuals with \(\phi(x) \geq 0\) are disease cases. We assume \(\varepsilon \sim \mathcal{N}(0,1)\) and model disease prevalence using the affine term \(\beta\), precluding the need for a nonzero threshold. For example, for DR with \(x=\text{DoD}\), we fit \(\alpha=0.03\) and \(\beta=-0.5\), based on higher DR prevalence with higher values of DoD (e.g. 22%, 47% and 57% for DoD 0-9 yrs, 10-19 yrs and 20-29 yrs; see below).

Instead of conducting an association test using case-control phenotype \(z\), we use the posterior mean estimate \(E(\varepsilon \mid z, x) = \int_{\varepsilon_{\text{min}}}^{\varepsilon_{\text{max}}} \varepsilon \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\varepsilon^2}{2} \right) d\varepsilon / \int_{\varepsilon_{\text{min}}}^{\varepsilon_{\text{max}}} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\varepsilon^2}{2} \right) d\varepsilon\) of the underlying quantitative phenotype \(\varepsilon\), where \((\varepsilon_{\text{min}},\varepsilon_{\text{max}}) = (\alpha(x - \bar{x}) - \beta, \infty)\) for \(z=1\) (cases) and \((\varepsilon_{\text{min}},\varepsilon_{\text{max}}) = (-\infty, -\alpha(x - \bar{x}) - \beta)\) for \(z=0\) (controls). We define our LT statistic as the number of samples times the squared correlation between genotype \(g\) and posterior mean \(E(\varepsilon \mid z, x)\) across samples, generalizing the standard Armitage trend test.\(^\text{18}\) This statistic is a \(\chi^2(1\ \text{dof})\) statistic and is equivalent to a Score test, which is commonly used in genetic association studies.\(^\text{39}\) The appropriate \(\chi^2(1\ \text{dof})\) distribution is observed in simulations and in permuted real data.\(^\text{40}\)

The LTSCORE test makes use of information on two clinical covariates: DoD and glycemic control. Glycemic control could be captured by either HbA\(_1c\) or FBS. We will use phenotype and covariate information to compute the posterior mean \(E(\varepsilon \mid z, x)\) of each sample, based on the above LT model parameters. Covariate files specifying liability-scale effect sizes of DoD and HbA\(_1c\) or DoD and FBS covariates in the population will be created separately for African Americans and Caucasians. If HbA\(_1c\) data is available, that will be used preferentially over FBS as it is more precise measure of glycemic control. We will estimate liability-scale effect sizes of DoD, HbA\(_1c\) and FBS using cohorts that ascertained/enrolled diabetic participants without regard to DR status. For African Americans, these cohorts are JHS, ARIC, Cardiovascular Health Study (CHS) and Multi-Ethnic Study of Atherosclerosis (MESA). For Caucasians, these cohorts are ARIC, CHS, MESA, and Blue Mountains Eye Study (BMES). The numbers of African American participants who contribute information on DoD, FBS and HbA\(_1c\) to calculate liability-scale effect sizes are 702, 988 and 866, respectively. The numbers of Caucasian
participants who contributed information on DoD, FBS and HbA$_{1c}$ to calculate liability-scale effect sizes are 917, 1389 and 1013, respectively.

**Computing association statistics**
The p value threshold for genome-wide significance will be determined according to recent empirically determined thresholds for different ancestral populations which take into account the large burden of multiple testing in GWAS. This study recommended that the empirical threshold should be:

- P < 3.24 $\times 10^{-8}$ for SNPs ascertained in populations of African ancestry
- P < 5.0 $\times 10^{-8}$ for SNPs ascertained in populations of European ancestry
- P < 3.24 $\times 10^{-8}$ for SNPs ascertained in transethnic meta-analyses

We will further correct these thresholds for the additional multiple testing due to examination of 4 phenotypes, each with and without incorporation of covariates, for a total of 8 additional tests to correct for. This will yield the following P value thresholds for our study:

- P < 3.75 $\times 10^{-9}$ for SNPs ascertained in populations of African ancestry
- P < 6.25 $\times 10^{-9}$ for SNPs ascertained in populations of European ancestry
- P < 3.75 $\times 10^{-9}$ for SNPs ascertained in transethnic meta-analyses

We will compute association statistics for each study separately and then meta-analyze across cohorts. This has the advantage of being less susceptible to artifact, but the disadvantage that it is not able to make use of data from cohorts with only DR cases for the primary analysis definition. Meta-analysis will be performed using inverse-variance weighting, either accounting only for effective sample size (defined as 4/[1/Ncase + 1/Ncontrol]) or accounting for both effective sample size and allele frequency. Meta-analysis statistics will be computed only for SNPs with at least half of the total available effective sample size. For any given analysis, only studies with at least 10 cases and at least 10 controls (for a given ethnicity) will be included. This is done because such low numbers of cases or controls may lead to test statistics that do not follow the theoretical null distribution and because they do not add any effective power to the analyses.

**Sensitivity analysis**
We will perform the following sensitivity analyses of the meta-analysis for the primary phenotype definition removing cohorts that do not have bilateral grading for diabetic retinopathy (ARIC, CHS) or do not have fundus photograph grading for retinopathy (FIND).

**Replication Meta-Analysis**
Twenty replication cohorts will provide summary statistics on SNPs with a P value < 1 $\times 10^{-5}$. Replication cohorts included patients of European, Asian and Hispanic ancestry. Details of the phenotyping and genotyping protocols of the studies have been previously described. Replication will be in silico with existing genotyping. The cohorts will choose the index SNP if it is available in their dataset. The replication cohorts will chose the case-control definition from our discovery analyses that most closely matches their existing case-control definitions. If possible, the replication cohorts will re-code their case-control definitions to match all of the phenotypes in the discovery analysis. However, this is not always feasible and therefore, every cohort will not provide information on every phenotype definition. The replication cohorts will incorporate DoD and glycemic control using traditional adjustment in logistic regression.
Replication meta-analysis will be performed using inverse-variance weighting, either accounting only for effective sample size (defined as 4/[1/Ncase + 1/Ncontrol]) or accounting for both effective sample size and allele frequency. Replication meta-analysis will be performed across all ethnicities together as well as individually by each ethnicity (Caucasians, Hispanics, Asians). Unfortunately, no African or African-American cohorts are available for replication meta-analysis.

Protein-Protein Interaction Analysis of Top GWAS Loci

To try to identify specific genes involved in DR within the loci with highest statistical evidence for association to the DR phenotype, we will apply the Disease Association Protein-Protein Link Evaluator (DAPPLE) to our discovery GWAS results. Using this analysis pipeline we will examine the top 1000 independent loci from the discovery GWAS. The loci will be defined based on linkage disequilibrium. Within these loci, we will identify direct and indirect protein-binding networks based on established protein-protein interaction database and test whether cross-locus connections are present. Permutation methods will be used to score individual proteins for likelihood of harboring causal variants and the P values generated will be corrected for the number of possible candidates within each locus. This methodology has been used successfully with GWAS from other diseases to identify protein networks with biological relevance.

Type 2 Diabetes and Associated Glycemic Traits Loci

To further determine whether participants with retinopathy might be enriched for SNPs which have been associated with type 2 diabetes overall and/or with associated glycemic traits, we will compute a correlation between Phenotype 1 DR case status and the sum of the beta*genotype (for quantitative traits) or logOR*genotype (for T2D) of the trait-associated SNPs for each cohort and each trait. This will be performed only using the Caucasian cohorts because the majority of the trait SNPs and their effect sizes were originally assessed in Caucasians. The SNPs will be obtained from the largest powered meta-analyses for these traits: T2D, fasting glucose, fasting insulin, 2-hour glucose, pro-insulin, and HbA1c. For each trait, we will compute a correlation (R) meta-analyzed across the cohorts as well as a Z score (Z) for that correlation being nonzero as follows:

\[ R = \frac{\text{sum-over-cohorts}(N_{\text{indiv}} \times \text{correlation})}{\text{sum-over-cohorts}(N_{\text{indiv}})} \]

\[ Z = \sqrt{\text{sum-over-cohorts}(N_{\text{indiv}})} \times R \]

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 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 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.csc.c.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)?
Admixture Genetic Mapping for Diabetic Retinopathy Genes in African Americans (already published)

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

___X__ A. primarily the result of an ancillary study (list number* _2011.08_) ___ 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.csc.c.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.csc.c.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


