1.a. **Full Title:** Admixture Genetic Mapping for Diabetic Retinopathy Genes in African Americans

   b. **Abbreviated Title (Length 26 characters):** Admixture Mapping Retinopathy in African Americans

2. **Writing Group:**
   Writing group members: Arti Tandon, Tien Y. Wong, Ronald Klein, Barbara EK Klein, Nick Paterson, Lucia Sobrin

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

**First author:** **Lucia Sobrin**
Address: 243 Charles Street, 12th floor, Boston, MA 02114
   Phone: 617-573-4279  Fax: 617-573-3011
   E-mail: lucia_sobrin@meei.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: **Tien Y Wong**
   Address: National University of Singapore
           16 Medical Drive
           Singapore 117597
   Phone: (65) 6772 5338  Fax: (65) 6777 7161
   E-mail: tien_yin_wong@nuhs.edu.sg

3. **Timeline:** We plan to complete the manuscript within 6 months

4. **Rationale:**

Diabetic retinopathy (DR) is a major complication of type 2 diabetes (T2D) and the leading cause of new cases of blindness among adults ages 20-74 years in the United
The frequency and severity of DR is highly heterogeneous. Duration of diabetes, glycated hemoglobin (HbA1c) levels and elevated blood pressure (BP) are the most consistently established risk factors for DR progression. However, these known risk factors explain some, but not all, of the observed heterogeneity. For example, progression of DR in some patients despite excellent control and the existence of patients who never develop any DR despite long-term hyperglycemia indicate that factors other than glycemia influence the risk of DR.

Genetic variation may explain some of the remaining heterogeneity in DR development. Heritability estimates for DR range from 18% to 27%. The heritability for the most severe form of DR, proliferative DR (PDR) has been found to be higher at 52%. Genetic association studies, including candidate gene studies and genome-wide association studies, are a potentially powerful way to identify genetic variants underlying DR, but most reported associations have not been consistently reproduced.

Admixture mapping is an alternative genetic study design that has proven to be a successful for complex traits in populations of mixed ancestry. Admixed individuals are those who inherit chromosomal segments of distinct continental ancestry. AA are one example of an admixed population; they have an average of 20% European ancestry and 80% African ancestry. Admixture mapping is useful when the disease of interest has a significant difference in prevalence in the ancestral populations. There are data to suggest that the prevalence of DR differs between African- and European-Americans. Several studies have found that AA have a higher risk of developing DR compared with Caucasians, even while controlling for the known DR risk factors. The prevalence of DR was higher for AA veterans than for Caucasian veterans in the Veterans Affairs Diabetes Trial of T2D after controlling for traditional risk factors. In other studies, among T2D patients, the odds of developing DR may be three times higher in AA compared to Caucasians while adjusting for risk factors.

The concept of whole-genome admixture mapping for DR in AA is to scan the genome in a large number of individuals with measured DR phenotypes, looking for genomic regions where the proportion of African ancestry of individuals with DR is strikingly higher or lower than that seen in individuals without DR. This is accomplished by genotyping ancestry informative markers (AIMs), polymorphisms that differ significantly between the two populations. Finding such a region would indicate the presence of at least one genetic risk variant for DR whose frequency differed between the ancestral populations.

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

Hypothesis: Admixture mapping can identify genetic loci associated with DR and PDR in AA with T2D.
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).

The Candidate-gene Association Resource (CARe)
CARe was a collaboration of large epidemiologic studies of heart disease for association analyses of genotypes and cardiovascular disease phenotypes.\(^{25}\) Three CARe cohorts have fundus photography on AA T2D participants and genome-wide genotyping done through CARe including ARIC, Jackson Heart Study (JHS), and Multi-Ethnic Study of Atherosclerosis (MESA).\(^{26-28}\)

For the purposes of this investigation, we will include all AA participants with T2D who had been genotyped as part of CARe and had graded fundus photographs. T2D is defined according to the 2003 ADA Criteria.\(^{29}\) The fundus photography protocol for each cohort has been previously described.\(^{30-32}\)

Case-Control Definitions for Diabetic Retinopathy
We will examine two DR case definitions. First we define cases as participants with ETDRS grade ≥ 14 in the eye with the higher ETDRS grade or in the only eye photographed, depending on the study’s protocol. These analyses are aimed to detect associations with the presence of any DR. Our second phenotype defines cases as participants with ETDRS grade ≥ 60 (which denotes PDR) in the eye with the higher ETDRS grade or in the only eye photographed, depending on the study’s protocol. The latter is done to reduce misclassification of patients with minimal signs of DR, which may be seen even in persons without diabetes.\(^{33-35}\) For all analyses, controls are defined as T2D participants with an ETDRS grade < 14 (no DR) and a duration of diabetes of at least 10 years.

Admixture Mapping
Genotyping
CARe participants underwent genome-wide genotyping on the Affymetrix 6.0 platform at the Broad Institute. These genotyping data sets will be each culled separately to keep only the AIMs specific to the Affymetrix panel. After genotyping, samples will be eliminated based on the following criteria: (1) samples with a <95% call rate and (2) duplicate samples defined as a >75% match in the genotypes between the two samples. The AIMs were used to estimate European/African ancestry proportion in each individual studied.

SNP quality filters
To decrease the likelihood of false-positives in our scan, we will apply a series of filters that had the goal of detecting and removing any SNPs with problematic genotyping. First, SNPS were dropped if there were atypical clustering patterns, ill-defined clusters, or low genotyping success rate (95%). We eliminate SNPs if they did not meet the requirement for Hardy-Weinberg equilibrium (P>0.01) in both ancestral West African
and European populations. We apply a filter that examined whether the observed frequency of a SNP in African Americans was statistically consistent with being a mixture of the frequencies observed in West Africans and European American samples that we used to represent the ancestral populations. Finally, we apply a filter that for each sample iteratively eliminated SNPs that were less informative (in terms of information content about ancestry until none were within 200 Kbs of each other or in detectable LD with each other in the ancestral West African or European populations.

**Estimating genome-wide ancestry**

Using the ANCESTRYMAP software, we will estimate a global percentage of African ancestry. ANCESTRYMAP uses a Markov Chain Monte Carlo approach to account for uncertainty in the unknown parameters (including SNP allele frequencies in the West African and European ancestral populations, the numbers of generations since mixture, and the average proportion of ancestry inherited from ancestral populations) that emerge from the Hidden Markov Model analysis.

**Admixture mapping**

We will use ANCESTRYMAP (genepath.med.harvard.edu/~reich) to perform admixture mapping analyses. To accumulate evidence of association in these models, we average the Bayes factors emerging from each model at each point in the genome, taking the log_{10} of this number to produce a locus-genome statistic (LGS) score. To assess whether a locus shows an unusual association of ancestry with DR or PDR, we calculate the likelihood of the data at each locus assuming a particular disease model, divided by the likelihood with no disease locus. The admixture scan was run using a range of disease risk models, which went from increased risk due to African vs. European ancestry. As previously described, we declare genome-wide significance only when the LGS score > 5. Based on simulation studies, this threshold for genome-wide significance is conservative (exceeded in < 1 in 100 repetitions of simulated data sets where there is no disease locus). A peak with a LGS score > 4 is our criterion for genome-wide suggestiveness, and further exploration.

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 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.cscce.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)? The non-diabetic retinopathy genetic analyses that are being performed as part of CHARGE Tien Wong is the PI of these efforts as well as an author on the current proposal.

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.cscce.unc.edu/ARIC/formspages/

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.cscce.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.
References