1.a. Full Title: Transcription factor 7-like 2 (TCF7L2) gene and retinopathy

b. Abbreviated Title (Length 26 characters): TCF7L2 and retinopathy

2. Writing Group:
   Writing group members: Yu Yan, Ronald Klein, Barbara Klein, Eric Boerwinkle, Kathryn Rose, Suzanne West, Cynthia Girman, Frederick Brancati, James Pankow, Christy Ballantyne, Anna Kottgen, Kari North.

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

First author: Yu Yan
Address: Department of Epidemiology
         School of Public Health
         University of North Carolina at Chapel Hill

         E-mail: yuyan@email.unc.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author): Kari North
Address: Department of Epidemiology and Carolina Center for Genome Sciences
         University of North Carolina Chapel Hill
         Bank Of America Center
         137 E. Franklin St., Suite 306
         CB #8050
         Chapel Hill, NC 27514
         Phone: (919) 966-2148
         Fax: (919) 966-9800
         E-mail: kari_north@unc.edu

3. Timeline: The 5 SNPs (rs12255372, rs7903146, rs7901695, rs11196205, rs7895340) within the TCF7L2 gene have been typed by Dr. Boerwinkle’s laboratory on the ARIC cohort. Approval of this manuscript by the ARIC publications Committee will then enable work on this manuscript. Once started, this work will lead to a manuscript within one year.
4. Rationale:

Type 2 diabetes mellitus (T2DM) affects up to 150 million people worldwide and is one of the leading causes of morbidity and mortality in middle-aged persons. Diabetic retinopathy, one of the common complications of T2DM, is a leading cause of blindness in people 20 to 74 years of age. Duration and degree of diabetes are major predictors for the development of diabetic retinopathy. Retinopathy is found in people with prediabetes which suggests that microvascular disease may contribute to the development of T2DM. Studies showed that microvascular abnormalities such as arteriolar narrowing and impaired microvascular blood flow in the skin and skeletal muscles have been noted in persons with T2DM and in persons at high risk of developing diabetes, such as those with prediabetes and first-degree relatives of persons with diabetes. Previous ARIC studies suggested that the retinal arteriolar narrowing is independently associated with risk of diabetes and that retinopathy predicts subsequent risk of clinical diabetes in individuals with a family history of diabetes, supporting a microvascular role in the development of diabetes. Therefore, early identification of individuals with increased risk for retinopathy among diabetics and non-diabetics is important for effective intervention.

Retinal microvascular signs (e.g., retinopathy, arteriolar narrowing, arterio-venous nicking) are potential markers of systemic arteriolar disease. Previous ARIC studies have demonstrated that narrower retinal arteriolar diameters are related to elevated blood pressure, incident T2DM and incident hypertension. Larger venular calibers have been further shown to predict the progression of retinopathy, independent of severity of retinopathy, suggesting that wider venular diameter may also be a marker of diabetes severity.

T2DM is a multifactorial, heterogeneous group of disorders. Apart from conventional risk factors like obesity, smoking, and physical inactivity, genetics plays an important role in its complex etiology. There is now compelling evidence that common variants in the genes influence the susceptibility to T2DM. There is also evidence that genetic factors may influence the calibers of retinal blood vessels, although the exact genetic determinants are unknown. In the Beaver Dam Eye Study, the between siblings correlation (95% confidence interval) for arteriolar and venular diameters were 0.23 (0.16, 0.31) and 0.20 (0.12, 0.28), respectively, whereas the spousal correlations were 0.03 and 0.04, respectively (P>0.05).

Following reports of linkage of T2DM to chromosome 10q, Grant et al. reported an association of the intronic variant DG10S478 of the transcription factor 7-like 2 gene with T2DM. Heterozygous (38% of the population) and homozygous (7% of the population) carriers of at risk alleles had prevalence ratios of 1.45 (1.26-1.67) and 2.41 (1.94-3.00), respectively. A population attributable risk of 21% was estimated in Caucasian populations of Iceland, Denmark and the U.S. This observation has been replicated in multiple studies and variants within the TCF7L2 gene have been compellingly associated with T2DM across different populations. Preliminary results from the ARIC study confirm that genetic variants in the TCF7L2 gene confer risk for
T2DM (manuscript proposal #1141). However, very few studies of diabetes related microvascular diseases have been conducted and their findings are not consistent. One study in a French population reported no evidence of an association with rs7903146 and prevalent T2DM complications such as severe retinopathy or nephropathy\(^2\). The InCHIANTI study found that T carrier diabetics of rs7903146 were more likely to have nephropathy (OR: 3.15; 95% CI: 1.27-7.81), and retinopathy (7.15; 0.87-58.51)\(^3\). The effect estimates for nephropathy and retinopathy from the InCHIANTI study were imprecise due to a very small number of cases. Previous studies of TCF7L2 gene variants were limited by sample size (especially in African-Americans), and also limited by not accounting for gene-by-environment interaction. Moreover, previous studies on TCF7L2 gene variants and retinopathy did not stratify the analyses by T2DM status. To our knowledge, no study has reported effect estimates for associations with microvascular diseases in a community-based study of African-Americans.

This proposed study will address these major gaps in the literature. The availability of data on the TCF7L2 gene on the ARIC cohort in a large, bi-racial sample will enable analyses focused on retinal microvascular diseases. The ARIC study rephotographed 2000 selected participants based on the distribution of carotid IMT 10 years after initial retinal photography, which provides an ideal opportunity to prospectively study the association between the TCF7L2 gene and incidence, progression and regression of retinal microvascular diseases besides the cross-sectional association at visit 3. Improved risk estimates as well as an expanded understanding of the impact of the TCF7L2 variants on the natural history and stages of T2DM and retinopathy will be obtained. The detailed phenotypic characterization available on cohort members will permit adjustment for a range of potential confounders and stratification on important modifiers, including T2DM status, gender, physical activity, center, tobacco exposure, family history of diabetes, hypertension, BMI, LDL cholesterol and HDL cholesterol. The proposed study will potentially contribute significant knowledge about the etiology of T2DM and retinopathy, and may aid in the development of screening strategies and treatment regimes utilizing genetic information.

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

(1) Will genetic variants (SNPs/haplotypes) of the TCF7L2 gene be associated with variations in the prevalence of retinal microvascular abnormalities?

(2) Will genetic variants (SNPs/haplotypes) of the TCF7L2 gene be associated with the incidence, progression and regression of retinal vascular abnormalities, and longitudinal changes in retinal vessel calibers?

(3) If associations are present,
   a. Will this association vary among diabetics, non-diabetics and participants with impaired fasting glucose (IFG)?
   b. Will this vary among African Americans and Caucasians?
   c. Will associations vary across different signs of retinal microvascular disease?
d. To which extent are associations influenced by adjustment for or stratification on T2DM, gender, center, family history of diabetes, physical activity, body mass index, smoking, hypertension, hyperglycemia, high LDL cholesterol and low HDL cholesterol?

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:** A cross-sectional study will be conducted among all ARIC participants at the third examination (1993-1995); a longitudinal study will be conducted among ARIC participants who attended the third (1993-1995) and fifth examinations (2004-2005).

**Inclusion/Exclusion Criteria:** Participants who did not attend the third examination will be excluded from analysis. Of the 12,642 participants examined at the third examination, individuals reporting race other than African American or Caucasian and African American residents living in Minneapolis and Maryland will be excluded as well. Participants who did not consent to genotyping will be excluded (use of DNA data distributed by the Coordinating Center with confirmation by using the variable RES_DNA = “No use/storage DNA” in the file ICTDER02). Retinal photographs will be used to define retinal phenotypes so participants who did not undergo the retinal examination or had ungradable fundus photographs at visit 3 will be excluded from the analyses. Participants with retinal microvascular diseases at visit 3 will be excluded from the incidence analysis. Individuals lacking information on diabetic status will be excluded from the stratified analysis by diabetic status. The missing number of individuals excluded due to missing genotype information will depend on the each SNP.

**Outcome:** The primary outcomes are retinal abnormalities which include five markers (visit 3,5): retinopathy, focal retinal arteriolar narrowing, arterio-venous (A/V) nicking, retinal arteriolar caliber (CRAE) and venular caliber (CRVE).

Retinopathy will be defined if any characteristic lesion as defined by the Early Treatment Diabetic Retinopathy Study severity scale was present: microaneurysms, retinal hemorrhages (blot or flame shaped), soft exudates (cotton-wool spots), macular edema, hard exudates, intraretinal microvascular abnormalities. A retinopathy severity score will be assigned based on the presence of various lesions: level 10, none; level 20, minimal nonproliferative retinopathy (microaneurysms only or blot hemorrhages only); level 35, early nonproliferative retinopathy (microaneurysms and one or more of the following: venous loops, soft exudate or hard exudate, and questionable intraretinal microvascular abnormalities or venous beading); levels 43 to 47, moderate to severe nonproliferative retinopathy (microaneurysms and one or more of the following: intraretinal microvascular abnormalities, venous beading, hemorrhages, and microaneurysms exceeding those in standard photograph); level 60+, proliferative retinopathy.

Focal arteriolar narrowing was assessed in each of four quadrants of the eye. Focal narrowing will be considered definite if an arteriole estimated to be 50µm in diameter or
greater had a constricted area of two thirds or less the width of the proximal and distal vessel segments. The maximum grade from the four quadrants defines two endpoints: focal narrowing on the disc or within 0.5 disc diameter of its margin and outside the disc. Both will be categorized as definite versus questionable or absent. A/V nicking will also be categorized as definite versus questionable or absent.

To detect and quantify generalized retinal arteriolar narrowing, the diameters of all arterioles and venules coursing through a specified area surrounding the optic disc were measured with an image processor and summarized as the central retinal artery equivalent (CRAE) and the central retinal venular equivalent (CRVE), representing average calibers of retinal arterioles and venules, respectively.

For the cross-sectional approach, the primary outcomes will be defined as the presence of retinal abnormalities; for the longitudinal approach, the primary outcomes will be defined as the incidence, progression and regression of retinal vascular abnormalities over 10 years. The incidence of retinal microvascular diseases will be defined as the presence of retinal signs at visit 5 in persons without these retinal signs at visit 3, progression will be defined as an increase in severity of retinal signs, and regression as disappearance of these retinal signs between visit 3 and 5.

Other Variables of Interest (time of collection):
(1) Five SNPs (rs12255372, rs7903146, rs7901695, rs11196205, rs7895340) within the TCF7L2 gene. Priority will be given to SNP rs7903146 as this SNP is the best candidate in the region for a functional effect and is most predictive across populations, including in the ARIC study.
(2) Baseline cardiovascular risk factors (visit 3): hypertension, systolic blood pressure (visit 1,2,3), diastolic blood pressure (visit 1,2,3), anti-hypertension medication use, diabetes (visit 1,2,3), glucose (visit 1,2,3), diabetic medication use, duration of diabetes, family history of diabetes, cigarette smoking status (ever/never, current/past, pack-years), HDL cholesterol, LDL cholesterol, triglycerides, total cholesterol, body mass index, waist to hip ratio, physical activity.

Diabetes at each visit (visit 1-3) will be defined as present if fasting plasma glucose levels of at least 7.0 mmol/L (126 mg/dL) (variable fast0802), nonfasting glucose levels of at least 11.1 mmol/L (200 mg/dL) (variable glucos01), current use of medications prescribed to treat diabetes (eg, insulin or sulfonylureas) (variable msra08f), or a positive response to the question "Has a doctor ever told you that you had diabetes (sugar in the blood)?" (variable HOM10E). IFG at each visit (visit 1-3) will be defined as fasting plasma glucose levels fell between 6.1 and 6.9 mmol/l, nonfasting glucose levels less than 11.1 mmol/L, no diabetic medications and a negative response to the question "Has a doctor ever told you that you had diabetes (sugar in the blood)?".
(3) Other covariates: age, gender, race, center, alcohol consumption, HBA1C.
(4) Variables to correct measurement error (visit 3): the repeat measurements of retinal vascular variables such as arteriolar and venular diameters, CRAE, and CRVE from the Individual Variability Study (n = 206), and the Grader Variability Study (n = 495).

Data Analysis

Data Quality Analyses: Tests of Hardy-Weinberg equilibrium will be performed for each SNP in the cohort using the control samples. Significant deviations from Hardy-
Weinberg will be assessed using a chi-square test, by comparing the observed distribution of genotypes to the Hardy-Weinberg 'expected' distribution, with degrees of freedom equal to the number of alleles (n) – 1.

The frequency distribution of all the variables in this proposed study will be examined in exploratory plots to examine their general characteristics. The two-sample t test and the \( \chi^2 \) test will be employed to detect differences in continuous and categorical characteristics between 'cases' and 'controls' at visit 3, and included (gradable) individuals and excluded (not-gradable) individuals at visit 3. Additionally, allele frequencies will be calculated and a two-sample test for binominal proportions will be used to assess differences in allele frequencies between 'cases' and 'controls'. The pattern for missing data for the exposure and the outcome of interest will be categorized into Missing Completely at Random, Missing at Random, or Not Missing at Random, which leads to the corresponding appropriate analysis of complete case analysis, data imputation, or maximum likelihood method. A similar strategy will apply when assessing the missing data due to no retinal examination or ungradable fundus photographs.

**Association Analyses:** All association analyses will be first examined within each ethnic (African American or Caucasian) group and then stratified by T2DM status.

For the association of SNP/haplotypes of the \( TCF7L2 \) gene with prevalent retinopathy, focal retinal arteriolar narrowing, and A/V nicking, prevalence odds ratios will be calculated from logistic regression models. Multivariable relationships between characteristics and severity of retinopathy will be examined with ordinal logistic regression in which severity of retinopathy will be categorized into four groups: level 10, level 20, level 35, and level 43 or worse. Analysis of variance (ANOVA) will be used to compare the mean CRAE and CRVE for SNP/haplotypes of the \( TCF7L2 \) gene; multiple linear regression will be used to evaluate the association of SNP/haplotype of the \( TCF7L2 \) gene with continuous CRAE and CRVE. Models will be constructed to account for potential confounders. For the assessment of confounders, a change-in-estimate approach will be used with a criterion of 0.10.

Due to the nonrandom nature of the sample that was rephotographed at ARIC visit 5 and selected to obtain 60% of baselines participants with high carotid IMT (>85 percentile) and 40% of remaining participants with low carotid IMT(<85 percentile), we will perform analyses stratified by carotid IMT category, and perform weighted analysis, adjusting for sampling fractions. For the association of SNP/haplotypes of the \( TCF7L2 \) gene with the incidence, progression and regression of retinal vascular abnormalities, hazard odds ratios will be calculated from proportional hazards regression models. The entry-age-adjusted age-scale model will be used to model the time-to-event\(^{44}\). Plots of the log (-log) survival curves and the Cox test will be utilized to assess violations of proportional hazard assumptions. Multiple linear regression will be used to assess longitudinal changes in retinal vessel calibers (Visit 5 vs. Visit 3).

For each SNP, a general model (no mode of inheritance assumption) using a 2-degree-of-freedom F-test will be utilized to analyze the gene-disease association in the cross sectional analysis. If this result is found to be statistically significant, or if there is an \textit{a priori} hypothesis based on the literature, further testing of the SNP effects assuming genetic modes of inheritance will then be performed. For dominant or recessive genetic transmission models, a single indicator variable (e.g. taking the value 1 if an individual
has genotype X0 or 00 and 0 otherwise) will be used. A variable taking on the values -1 for genotype XX, 0 for genotype X0, and 1 for genotype 00 will be used to test for additive genetic effects.

Haplotype Analysis: The association between multiple genotypes within TCF7L2 gene and retinal abnormalities will be examined using haplotype analysis methods. We will infer haplotypes using the method developed by Stephens, Smith, and Donnelly for population-based data (SSD algorithm)\(^45\). The PHASE software, available from the Oxford Mathematical Genetics Group, with suggested modifications from Lin et al. (2002)\(^46\) will be used to perform the haplotype analyses for this study. Haplotypes can then be incorporated into the analyses described in the current proposal in place of genotypes, weighted by their posterior probability. Several strategies can be used for testing the effects of haplotypes. Rare haplotypes may be collapsed into more common haplotype groups, which may help to remove the statistical noise that may be introduced by error in haplotype estimation. Presence or absence of individual haplotypes is most easily tested, with systematic modification of sites within the haplotype being added to the model as alternate “alleles.”

Assessment of Modification: The effect measure modification of the association between SNP/haplotypes of the TCF7L2 gene and retinal phenotypes will be assessed in logistic regression and linear regression models stratified by gender (male, female), hypertension (absence, presence), family history of diabetes (absence, presence), fasting glucose (<6.1, 6.1–6.9 mmol/l), HbA1c (<6.1, ≥6.1%), ever smoking (absence, presence) and body mass index (<28, ≥28 kg/m\(^2\)). The interaction term will be fit into the models and the likelihood ratio test with alpha = 0.15 will be utilized to assess effect measure modifications on the multiplicative scale. Additive interaction will be also assessed using the interaction contrast ratios (ICR)\(^47\). For logistic regression analysis, ICR = OR_AB – OR_A – OR_B + 1, where OR_AB represents the joint effect of exposure of interest and the SNP/haplotype and OR_A and OR_B represent the main effects of exposure and the SNP/haplotype, respectively. Departures from zero suggest additive interaction. The OR and variance covariance matrix is used to calculate values for ICR and the confidence intervals\(^48, 49\). For all tests, a two sided hypothesis will be adopted and a \(P\) value < 0.15 will be regarded as statistically significant. The effect measure modification of the association of SNP/haplotypes of the TCF7L2 gene with each continuous phenotype will be assessed using a product interaction term.

Methodologic Limitations or Challenges: An important limitation of this proposed study is the selection bias which may limit the generalizability of our study findings. One source of selection bias is the significant number of participants with ungradable fundus photographs which may introduce selection bias into this study. A previous ARIC study has suggested that persons without gradable photographs were more likely to be older and African American\(^50\). Also persons with more severe diabetes and more severe retinopathy are less likely to survive to the age at which the cohort was photographed. To evaluate the selection bias due to ungradable fundus photographs, we will impute values and conduct a sensitivity analysis if the missing is random. Another source of selection bias is the nonrandom nature of the sample at visit 5. The sample was selected based on the distribution of carotid IMT which may lead to selection bias, however previous ARIC
study did not demonstrate carotid IMT is associated with the retinal arteriolar abnormalities, suggesting that the selection on the distribution of carotid IMT may not bias our study\textsuperscript{51}. Due to the variability in caliber measurements it may be even harder to demonstrate relations with change in CRAE or CRVE in this selected sample. We will perform stratified analyses and weighted analysis accounting for sampling fractions.

Only one single 45-degree fundus photograph, from a randomly chosen eye, was taken through nonpharmacologically dilated pupils and graded. In comparison with other studies with multiple photographs through pharmacologically dilated pupils of both eyes, the technique used in ARIC study underestimates the prevalence of diabetic retinopathy\textsuperscript{52, 53}, which may attenuate associations if extant.

7.a. Will the data be used for non-CVD analysis in this manuscript?  
\_\_\_ Yes  
\_X\_ No

b. If Yes, is the author aware that the file ICTDER02 must be used to exclude persons with a value RES\_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES\_D\_NA = “CVD Research” would be used?  
\_\_\_ Yes  
\_\_\_ No
(This file ICTDER02 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 ICTDER02 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.csec.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)?

Ms. #1141 titled "Transcription factor 7-like 2 (TCF7L2) gene and type 2 diabetes" (first author: Yu Yan; corresponding/senior author: Gerardo Heiss)

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/

12. 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.
References


