ARIC Manuscript Proposal #2808

1.a. Full Title: A genome-wide association study for severe diabetic retinopathy in Scotland

b. Abbreviated Title (Length 26 characters): GWAS severe retinopathy

2. Writing Group:

Writing group members: Weihua Meng,1,* Kaanan Shah,2 Samuela Pallock,3 Iiro Toppila,4 Harry L Hebert,1 Wellcome Trust Case Control Consortium 2 (WTCCC2), Surrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT) study group, Mark I. McCarthy,5 Leif Groop,6 Emma Ahlqvist,6 Valeriya Lyssenko,6 Elisabet Agardh,6 Mark Daniell7, Georgia Kaidonis8, Jamie E Craig8, Paul Mitchell,9 Gerald Liew,9 Annette Kifley,9 Jiejin Wang,9 Mark W. Christiansen,10 Richard A. Jensen,10 Alan Penman,11 Heather A. Hancock,11 Ching J. Chen,11 Jane Z. Kuo,12 Xiaohui Li,13 Yui-der I. Chen,13 Jerome I. Rotter,13 Ronald Klein,14 Barbara Klein,14 Tien Y. Wong,15 Andrew D. Morris,13 Alexander S.F. Doney,14 Helen M. Colhoun,15 Michael A. Grassi,16 Alkes L. Price,3 Kathryn P. Burdon17, Niina Sandholm,4 Lucia Sobrin,18 Colin N.A. Palmer18*

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

First author: Weihua Meng
Address: Division of Population Health Sciences, School of Medicine, University of Dundee, Dundee, UK, DD2 4BF
Phone: 44 1382383419 Fax: 44 1382383802
E-mail: w.meng@dundee.ac.uk

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 2 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 States. 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. To facilitate identification of the genetic factors associated with severe DR, we performed this GWAS using a homogenous Scottish diabetic population in the first stage and multiple Caucasian and African American DR cohorts in the replication stage.

5. **Main Hypothesis/Study Questions:** A well-powered genome-wide association study of severe 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).**

**Participants In the discovery cohort**

The datasets from the GoDARTS project will be analyzed as the discovery cohort in this study. The GoDARTS project mainly recruits type 2 diabetic patients and non-diabetic controls throughout Tayside, Scotland to identify genetic susceptibility to diabetes including its complications and response to treatment. Participants will undertake a simple baseline clinical examination and complete a lifestyle questionnaire in addition to providing biological samples such as blood and urine. The participants provide informed consent at the time of recruitment which allows the use of their data and samples (including extracted DNA) for research purposes as well as link the data anonymously to their medical records. These records include the Scottish Care Information-Diabetes Collaboration (SCI-DC) and Scottish Diabetic Retinopathy Screening Collaborative – electronic health records used by health care professionals throughout Scotland for the care of patients with diabetes. Further information, including data access procedures, can be found at [http://diabetesgenetics.dundee.ac.uk/](http://diabetesgenetics.dundee.ac.uk/). The GoDARTS study has been approved by Tayside Committee on Medical Research Ethics and informed consent was obtained from all patients (REC reference 053/04).

So far, the project has recruited 9,439 diabetic patients and 6,927 of them have been genotyped. For this study, we extracted the DR screening records of all GoDARTS individuals from June,
1996 until June 2011 as well as information on age, gender, BMI, HbA1c and duration of diabetes.

**DR grading in Scotland**
Retinal screening has been undertaken in Tayside since 1990 and the DR screening protocol has previously been described. According to the Scottish Diabetic Retinopathy Grading Scheme, DR status can be graded into five levels: level R0: No DR; level R1: mild background retinopathy; level R2: moderate background retinopathy; level R3: severe background retinopathy; level R4: PDR. NPDR consists of mild, moderate and severe background retinopathy. In addition, the status of the macula was recorded as with or without diabetic maculopathy. However, the status of the macula will not be taken into account in this study when defining DR.

**Definition of severe diabetic retinopathy cases and controls in GoDARTS**
A severe DR case is defined in this study as a type 2 diabetic individual with at least one eye that has previously been coded as severe background retinopathy (level R3) or PDR (level R4); or with a history of laser photocoagulation treatment in the e-health records. A control is defined as a type 2 diabetic individual with DR longitudinal screening records for both eyes, which were only graded as normal (level R0) or mild background retinopathy (level R1). In addition, controls had no record of laser photocoagulation treatment. To maintain homogeneous case and control populations, we will remove type 2 diabetic individuals whose severest DR screening records were moderate background retinopathy (level R2) from both cases and controls. In simple words, this study compares severe DR cases (level R3 and R4) with controls (level R0 and R1).

**DR definitions in the Multiple Caucasian and African American DR cohorts**

**DR definitions for SDR:**

DR cases: type 2 diabetic patients with panretinal laser therapy, or at least moderate background retinopathy, or mild background retinopathy with duration of diabetes at retinopathy assessment <10 years.

Controls: patients with no recorded background retinopathy, maculopathy or panretinal laser therapy. These individuals should have at least 4 years duration of diabetes.

**DR definitions for FinnDiane:**

DR cases: type 1 diabetic patients with panretinal laser therapy, or at least moderate background retinopathy, or mild background retinopathy with duration of diabetes at retinopathy assessment <10 years.

Controls: patients with no recorded background retinopathy, maculopathy or panretinal laser therapy. These individuals should have at least 4 years duration of diabetes.

**DR definitions for GokinD and EDIC:**

DR definitions: type 1 diabetes with self-reported prior laser photocoagulation treatment history for GokinD and type 1 diabetes with ETDRS grading score > 61 for EDIC.
Controls: all the remaining type 1 diabetic subjects in the cohorts.

DR definitions for AUST, BMES, CHS, AAPDR, JHS, ARIC, MESA-AA:
DR definitions: type 2 diabetes with ETDRS grading score > 14
Controls: type 2 diabetes with ETDRS grading score < 14

**Genotyping and quality control**
The GoDARTS diabetic individuals have been genotyped by either Affymetrix SNP6.0 chips (3,673 patients) funded by the Wellcome Trust Case Control Consortium 2 (WTCCC2) project or by Illumina OmniExpress chips (3,254 patients) funded by the Surrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT) project. Standard protocols were used for genotyping quality controls for the WTCCC2 studies and the SUMMIT studies. ARIC participants underwent genome-wide genotyping on the Affymetrix 6.0 platform. ARIC African Americans participants were genotyped as part of the CARE project. ARIC will provide summary statistic data for replication of the SNPs with P value less than 1 X 10⁻⁵ in the discovery cohort.

**Statistical analysis**
The imputation of non-genotyped SNPs in the Affymetrix SNP6.0 chips and Illumina OmniExpress chips will be done by SHAPEIT and IMPUTE2 using reference files from the 1000 genome phase I datasets (53,54). The recommended $r^2 > 0.3$ will be used to filter out badly imputed SNPs. Routine quality control steps will be frequently applied using PLINK (removing SNPs with over 5% genotyping missing, or with minor allele frequency less than 1%, or those that failed Hardy-Weinberg tests $P < 0.000001$, and removing individuals with more than 5% genotype data missing). SNPs on the X and Y chromosomes and mitochondrial SNPs will be excluded from analyses. Population stratification analysis will be based on multidimensional scaling integrated in PLINK to detect any difference in ancestry within the cohort, with a lambda value indicating the level of stratification. Removal of related samples will be based on pi-hat > 0.125 in PLINK. The $P$ values for SNP associations will be generated based on logistic regression test integrated in PLINK with age, gender, HbA1c and duration of diabetes. A $P$ value of less than $5 \times 10^{-8}$ will be considered to be an association, warranting further exploration. The LD scores (R-squared) among significant SNPs will be later calculated by PLINK. The positive SNPs generated from the first stage will then be meta-analyzed using multiple replication cohorts (including ARIC) where $P$ values of these SNPs were generated by logistic regression adjusting with relevant covariates. Multiple meta-analyses will performed by GWAMA combining Caucasian cohorts and African American cohorts. SNP functional annotations will be applied by SNPnexus and the Manhattan plot will be generated by HaploView. LocusZoom will be used for regional visualization. SNPEVG will be used to generate the corresponding Q-Q plot, a tool to evaluate differences between cases and controls caused by potential confounders (different genotyping lab, different DNA extraction methods, etc).

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.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)?
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.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.

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