1.a. Full Title: Sequencing to follow up hematologic trait associations

b. Abbreviated Title (Length 26 characters): Sequencing of HCT GWAS Loci

2. Writing Group:
   Writing group members:
   
   Santhi K. Ganesh
   Anna Kottgen
   Susan Furth
   Georg Ehret
   Aravinda Chakravarti
   Linda Kao
   Josef Coresh
   Eric Boerwinkle

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. SKG

First author: Santhi K. Ganesh
Address: 1500 West Medical Center Drive
         MSRB III / 7220A
         University of Michigan Health System
         Ann Arbor, MI 48109
         Phone: 734-764-4500    Fax: 734-936-8266
         E-mail: sganesh@umich.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: Linda Kao
Address: 615 N. Wolfe St., Room 6513
         Baltimore, MD 21205
         Phone: 410-614-0945    Fax: 410-955-0863
         E-mail: wkao@jhsph.edu

3. Timeline: 1 year
4. **Rationale:** In an analysis of six participating cohorts, including ARIC, the CHARGE Hematology Working Group identified 23 loci associated with six red blood cell traits (hemoglobin concentration (Hb), hematocrit (Hct), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC) and red blood cell count (RBC)). (Ganesh et al, Nature Genetics 2009). A number of these loci associated with population-level variation in these red blood cell traits have well-described rare gene mutations that cause hematologic diseases (Supplementary Table 4 of the Nature Genetics paper). The next step in understanding how these genetic loci influence these erythrocyte traits is to more precisely define variation within these genomic loci we have identified. In the CHARGE-S project, sequencing based follow-up will be conducted for GWAS signals identified in our hematologic studies.

5. **Main Hypothesis/Study Questions:** The primary hypothesis of this study proposal is that loci identified through GWAS and meta-analysis of erythrocyte traits, chiefly hematocrit, harbor causal variants that influence hematocrit levels, and these causal variants may be identified through sequencing approaches. Accordingly, our primary study question is whether there are rare or putatively functional variants within the genomic loci identified through GWAS that may be more precisely defined by sequencing.

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 design of CHARGE-S is a case cohort study in which 200 cases (100 from the ARIC Study, 50 from the Cardiovascular Health Study (CHS) and 50 from the Framingham Heart Study (FHS)) are selected from the lower end of the Hct distribution only. The cohort sampling includes 1000 ARIC participants, 500 CHS participants, and 500 FHS participants so we are using the same proportions in the Hct case group (100 ARIC participants, 50 CHS participants, 50 FHS participants). In the primary CHARGE GWAS, we analyzed individuals with Hct values within 3 SD of the population mean in each cohort, to remove individuals with likely co-morbidities or secondary causes of anemia so we will select for these analyses individuals with hematocrit levels within 3SD of the ARIC population mean. Individuals with history of malignancy are excluded from the analysis.

Raw sequence data will be analyzed with the assistance of the sequencing center, and ultimately a flat file of sequencing-identified variants will be provided to ARIC investigators. These data will then be analyzed against the phenotypic data as described, for association testing for variants identified through sequencing. We will seek independent replication of signals identified in this study.

We will then conduct case control association analysis as the primary analysis, adjusting for age, gender and enrollment site. In a secondary analysis, we will examine hct as a continuous trait. Within each cohort with subjects sequenced in this study, we will calculate the residuals from
linear regression of hematocrit as a continuous trait, with adjustment for age, gender and enrollment site.

Secondary analyses of other hematologic traits will be conducted as well in CHARGE-S, also for follow up of previously defined associations.

7.a. Will the data be used for non-CVD analysis in this manuscript?  ____ Yes  _X___ No

The primary analysis is for hematocrit, a hematologic trait. Epidemiologic data shows that red blood cell traits are associated with blood pressure and hypertension, cardiovascular disease prevalence and outcomes, as well as all-cause mortality. In addition to the analysis plan outlined above, we are interested in participating in studies of pleiotropy, to determine and investigate genetic variants associated with vascular diseases and red blood cell traits. We chose Hct as the primary trait of interest, in part, because hematocrit has been extensively studied in large population-based cohorts and is associated with cardiovascular disease and all-cause mortality, therefore having the greatest relevance to the CHARGE mission to understand cardiovascular diseases.

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 ICTDER03 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)? ARIC MS 1458

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.cscce.unc.edu/aric/forms/](http://www.cscce.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.