ARIC Manuscript Proposal # 1458

PC Reviewed: 12/9/08	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Genome-wide Association Study of Single Nucleotide Polymorphisms with RBC phenotypes and Anemia

b. Abbreviated Title (Length 26 characters): GWAS of RBC phenotypes and Anemia

2. Writing Group:
Writing group members: Susan Furth, Anna Kottgen, Jeffrey Fadrowski, Meredith Atkinson, Eric Boerwinkle, Linda Kao, Josef Coresh, Man Li (core ARIC group). An equal number of authors from each of the other cohorts involved (currently FHS, CHS, Rotterdam) will be included, as well as individuals contributing to the analysis and writing as needed for the specific project. Other cohorts may be added depending on the availability of RBC data.

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

First author: Susan Furth
Address: Welch Center for Prevention, Epidemiology, and Clinical Research
2024 E Monument St, Suite 2-600
Baltimore, MD 21287

Phone: (410) 502-7964 Fax: (410) 955 0476
E-mail: sfurth@jhmi.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: Josef Coresh
Address: as above

Phone: (410) 955 0495 Fax: (410) 955 0476
E-mail: coresh@jhu.edu

3. Timeline: Data analysis to start immediately upon receipt of approval, first draft of manuscript May 2009.

4. Rationale:

Anemia is a common clinical problem. A recent analysis of the National Health and Nutrition Examination Survey (NHANES 1999-2004) data[1] found that overall, anemia was present in approximately 6% of the US population, with African Americans at greater risk than whites. Up to 10% of non-hospitalized men and women older than 65 years of age in the U.S. population are anemic. The risk of anemia is almost twice as high in individuals with chronic kidney disease[2]. Among the approximately 10 million individuals with CKD, anemia affects more than half[3], and is associated with increased risk of cardiovascular disease, increased mortality, and decreased quality of life[2].

Red blood cell count (rbcc), red blood cell size (mean corpuscular volume or MCV), and mean corpuscular hemoglobin (MCH) are major criteria for evaluating anemia. Environmental factors influence RBC count and other RBC indices, however previous studies have suggested that each of these measures have a substantial genetic component.[4-6] Assessment of potential genetic variants for RBC traits is a first step toward understanding the genetic variability leading to susceptibility to anemia. Anemia is a complex disease, with multiple environmental and inherited factors contributing to its etiology. The evidence for some genetic basis for hemoglobin concentration includes twin heritability studies, differences in hemoglobin norms for blacks vs whites which persist after matching for age and sex, and recent studies suggesting hemoglobin heritability within similar Western European populations.[7] Our Framingham study colleagues have performed GWAS studies of RBC traits including microsatellite linkage scans and 100K GWAS demonstrating the heritability of these traits. As part of the CHARGE consortium, the FHS has performed further analyses and has asked ARIC to perform similar RBC GWAS analyses and meta-analysis to evaluate variability in RBC Count and RBC indices. We therefore propose to assess possible genetic variation contributing to hematocrit (HCT), hemoglobin (HB), red blood cell count (RBCC), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and mean corpuscular hemoglobin concentration (MCHC), adjusted for age and gender as part of the CHARGE consortium.

5. Main Hypothesis/Study Questions:

In association with the CHARGE consortium, we propose to study the association of SNPs from the Affy 6.0 array in ARIC Study participants and the following rbc traits:

1. hematocrit (HCT)
2. hemoglobin (HGB)
3. red blood cell count (RBCC)
4. mean corpuscular volume (MCV)
5. mean corpuscular hemoglobin (MCH)
6. mean corpuscular hemoglobin concentration (MCHC)

In subsequent analyses, we propose to study anemia stratified by CKD.

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 and inclusion/exclusion: subjects and sample size

Individuals who did not consent to genetic research, those of self-reported race other than "black" or "white", and those without baseline rbc measurements will be excluded from analyses. The final baseline sample size will therefore consist of all successfully genotyped individuals among 11,357 white and 4,091 black ARIC participants. Race-specific cross-sectional analyses (study visit 1, visit 2 and visit 3) will be conducted. For the CHARGE consortium analysis of RBC phenotypes, we will

(1) exclude those not within WHO normal ranges as outlined below:

Hct: 45% to 52% for men and 37% to 48% for women

Hb: 13 to 18 grams per deciliter for men and 12 to 16 for women (international units 8.1 to 11.2 millimoles/liter for men, 7.4 to 9.9 for women)

RBCC: 4.2 to 5.9 million cells/cmm (international units 4.2 to 5.9 x 1012 cells per liter)

MCV: 80 to 100 femtoliters (a fraction of one millionth of a liter).

MCH: 27 to 32 picogram

MCHC: 32% to 36%

(2) Include everyone

(3) Exclude individuals using mean + 3SD

Covariates for all analyses include age, gender and study center

Sample sizes of the other CHARGE consortium members in this working group are:
FHS: ~9000 related and unrelated white individuals aged 30-99 years, Affy 500K chip + 50K supplemental chip
Rotterdam: ~6000 white individuals aged 55+ years at baseline, population-based, Illumina 500K chip
CHS: currently ~4000 white individuals 65+ years at baseline, population-based, no coronary heart disease, angina, or heart failure at time of DNA collection, Illumina 370duo chip.

Publication Strategy

Publication strategy will follow a working-group specific document. Currently, the document includes agreement to replicate and meta-analyze findings across cohorts with the above named traits for white samples only. Manuscripts will most likely include one or more for the most significant association signals replicating across cohorts and one or more for the meta-analysis of the entire association scan across cohorts. Data on black ARIC participants may be included if it increases the scientific impact of the paper and does not overlap other existing papers following CARE rules. Specifically, loci discovered in whites will also be examined in blacks, but loci discovered in blacks that do not replicate among whites will be the focus of other papers pursued under the CARE rules.

The publications committee will be notified by an addendum to this proposal if/when additional papers are produced from the science covered in this proposal.

Exposure Measurements and Definitions
This manuscript proposal is concentrated on the analyses of the Affimetrix 6.0 SNP data (~900,000 SNPs and imputed SNP's).

Quality Control of Genotyping Data
Clean genotype data will be used. Exclusions were data-driven and include excessive missing data, lack of chromosomal coordinates, sample mismatches, excessive autosomal heterozygosity, and genetic outliers.

Outcome Measurements and Definitions
In primary analyses, we will evaluate the following continuous traits: HCT, HGB, MCV. The measurements from study visit 1 (which include HCT and HGB) will be used in the primary analyses in order to maximize the sample size. At study visit 2, HGB, HCT and MCV are available for analysis. At study visit 3, all RBC phenotypes are available, RBCC, HGB, HCT, MCV, MCH and MCHC, however the sample size will be smaller, approximately 3,000 as these measures were limited to the Jackson site.

For subsequent analyses, we will evaluate the dichotomous trait anemia. Anemia will be defined as HGB <13.5 g/dl for men and <12g/dl for women, approximating the lowest sex-specific deciles. In order to maximize power, the case definition of anemia will include the first occurrence of low HGB at either study visit 1, 2, or 3.

Statistical Analysis
Analyses will be conducted stratified by race. Genotypes will be modeled using an additive genetic model. Analyses will further be divided into primary and secondary analyses.

Prior to the primary association analyses on the traits as listed above, age and sex-specific multivariable-adjusted standardized residuals will be generated for the analysis of continuous traits using linear regression analyses. Covariates will include age, center and gender. Dichotomous outcomes will be evaluated using logistic regression with adjustment for the same covariates. Analyses will use dosage estimates of genotypes and the probable program running under R.

Secondary analyses include those stratified on CKD status. To define CKD status, estimated GFR in ml/min/1.73m^2 will be calculated from calibrated serum creatinine measured at study visit 1, or 2, using the CKD-Epi equation[8]. CKD will be defined as eGFR <60 ml/min/1.73m^2. In order to maximize power, the case definition of CKD will include the first occurrence of CKD at either study visit 1, 2 or 3. Further secondary analyses will likely be necessary depending on the nature of the findings.

Meta-analysis on all SNPs will be conducted among all studies on all 2.5 million imputed SNPs. The genome-wide significance threshold will be 5*10^-8. Finally, we will use bioinformatics tools to obtain information about informative SNPs, both with respect to linkage disequilibrium as well as to function. SNPs will be evaluated for their location, SNP type, across-species conservation, and association with gene expression. For

nonsynonymous coding SNPs, we will evaluate the predicted consequence of the amino acid change. Collaborations with additional scientists to functionally study promising SNPs may be initiated as scientifically justified and agreed upon within the working group.

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 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 ____ 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

8.c. If yes, is the author aware that some DNA data is not allowed to be used by 'for profit' groups. Is this data being used by a 'for profit' organization? _X_No If yes, is the author aware that the participants with RES_DNA = 'not for profit' restriction must be excluded? ____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)?
There is a proposal to look at WBC variability using GWAS submitted by Aaron Folsom but there is no overlap with this proposal. Dr. Coresh has been in touch with Dr. Folsom.

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 2006.03 (Stampede, genotyping in Caucasiona); 2007.02 (CARe, genotyping in African Americans)
___ 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

1. McFarlane SI, Chen SC, Whaley-Connell AT, Sowers JR, Vassalotti JA, Salifu MO, Li S, Wang C, Bakris G, McCullough PA, Collins AJ, Norris KC, Kidney Early Evaluation Program Investigators. Prevalence and associations of anemia of CKD: Kidney Early Evaluation Program (KEEP) and National Health and Nutrition Examination Survey (NHANES) 1999-2004. *Am J Kidney Dis.* 2008; 51: S46-556.

2.Astor BC, Muntner P, Levin A, Eustace JA, Coresh J. Association of Kidney Function with Anemia: The Third National Health and Nutrition and Evaluation Survey. Arch Intern Med 2002; 162:1401-1408.

3. Deghan A, Kottgen A, Yang Q,Hwang Sj, Kao WL, Rivadeneira F, Boerwinkle E, Levy D, Hofman A, Astor BC, Benjamin EJ, van Dujin CM, Wittman JC, Coresh J, Fox CS. Association of three genetic loci with uric acid concentration and risk of gout: a genome wide association study. Lancet 2008, Sept 30, epub ahead of print

4.Yang Q, Kathiresan S, Lin JP, Tofler GH, O'Donnell CJ. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. BMC Med Genet. 2007 Sep 19;8 Suppl 1:S12.

5. Lin JP, O'Donnell CJ, Jin L, Fox C, Yang Q, Cupples LA.Evidence for linkage of red blood cell size and count: genome-wide scans in the Framingham Heart Study. Am J Hematol. 2007 Jul;82(7):605-10.

6. Lin JP, O'Donnell CJ, Levy D, Cupples LA. Evidence for a gene influencing haematocrit on chromosome 6q23-24: genomewide scan in the Framingham Heart Study J Med Genet. 2005 Jan;42(1):75-9.

7. Sala C, Ciulla M, Lanzara C, NutileT, Bione S, Massacane R, d'Adamo P, Gasparini P, Toniololo D, Camashella C. Variation in Hemoglobin levels in normal Italian populations from genetic isolates. Haematologica 2008; 93:1372-5

8. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, Kusek JW, Eggers P, Van Lente F, Green T, Coresh J. A New Equation to Estimate GFR from Serum Creatinine: Improved Accuracy and Updated Estimates of Prevalence of Chronic Kidney Disease in the United States. J Am Soc Nephrol 2008; 19(F-FC160):36A