

**ARIC Manuscript Proposal #2482**

**PC Reviewed:** 12/9/14  
**SC Reviewed:** \_\_\_\_\_

**Status:** A  
**Status:** \_\_\_\_\_

**Priority:** 2  
**Priority:** \_\_\_\_\_

**1.a. Full Title:** Exonic Variants Associated with Blood Cell Traits

**b. Abbreviated Title (Length 26 characters):** Exome and blood cells

**2. Writing Group:**

Writing group members: ARIC writing group members include the participants in the CHARGE hematology working group and the Blood Cell Consortium and individuals involved in generation of the exome chip and exome sequence data (alphabetical): Eric Boerwinkle, Aaron Folsom, Santhi Ganesh, Alanna Morrison, Nathan Pankratz, Linda Polfus

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

<b>First author:</b> Nathan Pankratz, PhD	/ Linda Polfus, PhD
Address: University of Minnesota	University of Texas
Moos Tower, Room 1-156	1200 Pressler Street, Room E411
515 Delaware Street SE	Houston, TX 77030
Minneapolis, MN 55455	
Phone: (612)624-2456	Phone: (713)500-9012
Fax: N/A	Fax: 713-500-0900
E-mail: <a href="mailto:pankr018@umn.edu">pankr018@umn.edu</a>	Email: <a href="mailto:Linda.M.Whitaker@uth.tmc.edu">Linda.M.Whitaker@uth.tmc.edu</a>

**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: Santhi Ganesh  
Address: University of Michigan  
1150 West Medical Center Drive  
MSRBIII /7220A  
Ann Arbor, MI 48109  
Phone: (734)764-4500  
Fax: (734)936-8266  
E-mail: [sganesh@umich.edu](mailto:sganesh@umich.edu)

**3. Timeline:**

We anticipate completion of analyses and subsequent draft of manuscript(s) during 2015 calendar year, as described below under “Analysis Model.”

#### **4. Rationale:**

Circulating blood cell counts represent potentially important intermediate phenotypes for a variety of cardiovascular, pulmonary, hematologic, and immunologic diseases. Recent GWAS have begun to contribute to our understanding of the genetics of blood cell traits in European-and African ancestry populations. The CHARGE, HaemGen, and COGENT Consortia [1,2] have identified ~100 loci associated with blood cell traits including hemoglobin concentration, hematocrit, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC) red blood cell count (RBC), white blood count (WBC), platelet count, and mean platelet volume. Traits such as leukocyte count and hemoglobin are known to be associated with mortality and cardiovascular disease [3,4]. Traits such as hemoglobin and leukocyte count are known to differ by race [5]. Through admixture mapping, the DARC locus has been identified as a major determinant of the lower WBC in individuals of African descent compared with European-Americans [6].

Many deeply phenotyped cohorts have or are currently acquiring exome sequence data and genotype data for the 247,870 variants on “the exome chip.” Many of the exonic variants covered by these technologies affect protein structure, either through a change in an amino acid, truncation of the protein, or interruption of a splice site, thus providing a more direct interpretation of the underlying mechanism of a trait association than is often possible from a traditional GWAS. These same cohorts also have existing phenotype data for a range of blood cell traits with total sample sizes up to ~50,000 white samples and as many as 20,000 black samples for exome chip data and as many as 14,000 samples for exome sequencing data.

The overarching goal of this proposal is to determine and investigate novel, exonic genetic associations with blood cell traits through multi-ethnic consortia efforts related to blood cell traits. Our primary hypothesis is that systematic evaluation of variants in the exome, both common and rare, will uncover genes which play an important role in blood cell count and function.

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

- A. Rare and common variants in the exonic regions of the genome will be associated with blood cell counts and indices.
- B. Analyses will be performed using both single variant and gene-based approaches.

**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).**

Blood cell counts from Baseline: hematocrit, hemoglobin, WBC and differential, platelet count, red cell indices

Covariate data (baseline): Self-reported race, sex, age at measurement, renal disease, smoking status, cancer diagnosis, BMI and lipids (total cholesterol or lipid profile)

*Analysis Model:*

Exome Chip and Exome Sequence data will be meta-analyzed with other agreed upon large consortia, including 1) Exome chip meta-analysis in CHARGE led by Nathan Pankratz/Sanathi Ganesh (to be submitted for publication in January 2015), 2) Exome chip meta-analysis in the Blood Cell Consortium (Ganesh/Pankratz/Polfus from ARIC) estimated to be submitted for publication in late 2015, and 3) Exome Sequence Analysis in CHARGE/ESP led by Linda Polfus/Sanathi Ganesh estimated to be submitted in 2015, as the discovery analyses and follow-up experiments are completed.

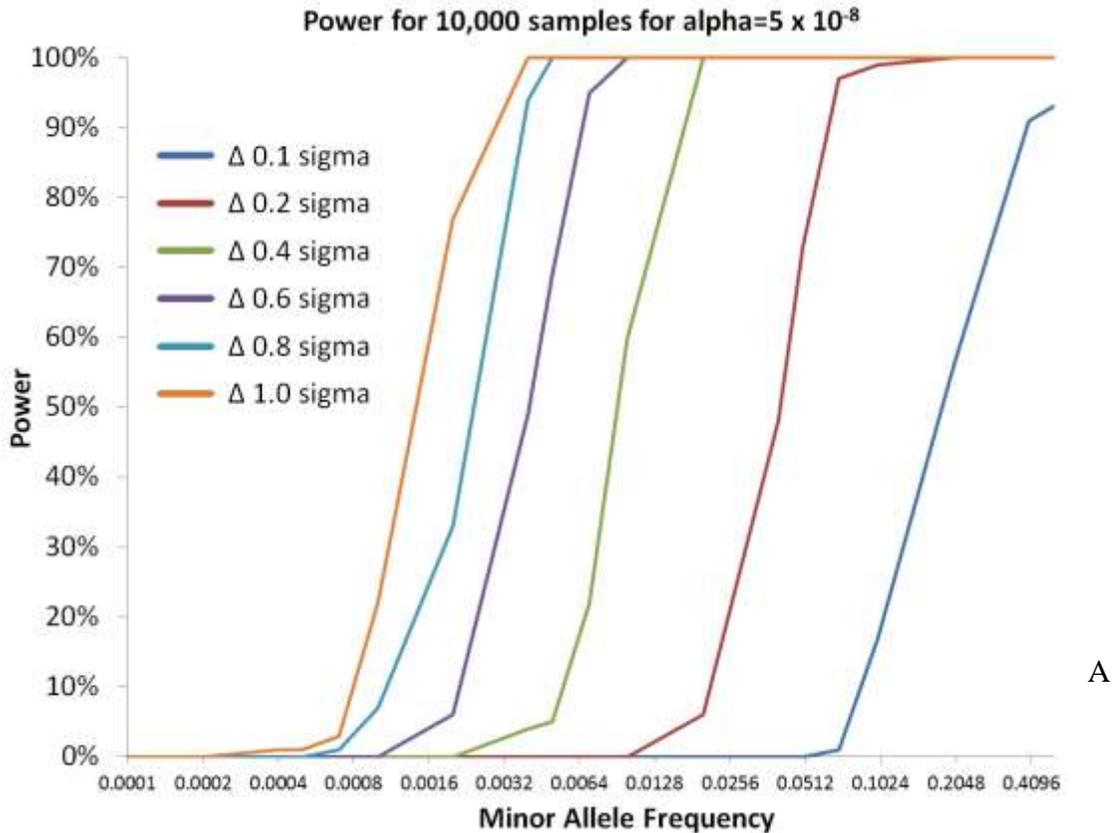
*Analysis Plan:*

Our primary analyses will focus on 4 main quantitative traits: hemoglobin, hematocrit, WBC, and platelet count. Additional traits will include red cell indices and WBC subtypes, which are available in smaller subsamples. Exclusion criteria are end-stage renal disease and congenital anemia, such as sickle cell and thalassemia as determined by the genetic data.

We will perform a number of preliminary QC steps, including plotting the distribution of betas and standard errors by study and plotting the allele frequencies of variants compared to the weighted average of the allele frequencies across all studies. QQ-plots will be examined at different points in the analysis in order to identify potential issues that would require additional refining of sample and variant filters.

Rare variant analyses can be carried out either using SeqMeta (for the CHARGE-led analyses) or RareMetalWorker/rvtests (for the Blood Cell Consortium). For the testing of single candidate regions for association of rare and common variants with of quantitative blood count traits, we will begin with using a single SNP analysis with a multiple linear regression approach. In addition to analyzing genetic variants individually, we will perform analyses collapsing rare variants over genes [7]. Rare (<1% allele frequency) and uncommon (1-5% allele frequency) variants will be collapsed gene-by-gene using a burden test. In its simplest form, the CMC (Combined Multivariate and Collapsing) method (Li et al. 2008), sums all the minor alleles in a specified window that are below a pre-specified threshold (typically <1% or <5%) for each individual. SNP-set (Sequence) Kernel Association Test (SKAT) is a variance components approach that takes into account heterogeneity in direction of effect for different variants in the same gene [8].

**Brief Discussion of Power**



sample size of 10,000 yields 80% power for a variety of minor allele frequency/effect size pairings using simple linear regression (see Figure 1). We currently anticipate inclusion of about 50,000 whites, 20,000 African Americans, and 10,000 Hispanic samples across participating cohorts.

7.a. Will the data be used for non-CVD analysis in this manuscript?  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  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?  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"?

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.csc.unc.edu/ARIC/search.php>

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 are no manuscript proposals for the evaluation of exome sequence data and blood cell traits. This proposal is similar to other manuscript proposals to evaluate exome sequence data and hematologic and other traits (e.g., EKG phenotypes, atrial fibrillation, pulmonary measures, blood pressure, hemostatic factors), and our group of investigators studying hematology traits are collaborating on these studies

- 11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  Yes  No

This proposal is related to the ancillary study proposal 2012.05 Exome and whole genome sequence analysis of blood cell traits

- 11.b. If yes, is the proposal

A. primarily the result of an ancillary study (list number\* 2012.05)

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.csc.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.csc.unc.edu/aric/index.php>, under

Publications, Policies & Forms. [http://publicaccess.nih.gov/submit\\_process\\_journals.htm](http://publicaccess.nih.gov/submit_process_journals.htm) shows you which journals automatically upload articles to Pubmed central.

### 13. References

- 1) Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, Kuhnel B, Aspelund T, Yang Q, Tanaka T, Jaffe A, Bis JC, Verwoert GC, Teumer A, Fox CS, Guralnik JM, Ehret GB, Rice K, Felix JF, Rendon A, Eiriksdottir G, Levy D, Patel KV, Boerwinkle E, Rotter JI, Hofman A, Sambrook JG, Hernandez DG, Zheng G, Bandinelli S, Singleton AB, Coresh J, Lumley T, Uitterlinden AG, Vangils JM, Launer LJ, Cupples LA, Oostra BA, Zwaginga JJ, Ouwehand WH, Thein SL, Meisinger C, Deloukas P, Nauck M, Spector TD, Gieger C, Gudnason V, van Duijn CM, Psaty BM, Ferrucci L, Chakravarti A, Greinacher A, O'Donnell CJ, Witteman JC, Furth S, Cushman M, Harris TB, Lin JP. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet.* 2009 Nov;41(11):1191-8.
- 2) Soranzo N, Spector TD, Mangino M, Kühnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M, Salo P, Voight BF, Burns P, Laskowski RA, Xue Y, Menzel S, Altshuler D, Bradley JR, Bumpstead S, Burnett MS, Devaney J, Döring A, Elosua R, Epstein SE, Erber W, Falchi M, Garner SF, Ghorri MJ, Goodall AH, Gwilliam R, Hakonarson HH, Hall AS, Hammond N, Hengstenberg C, Illig T, König IR, Knouff CW, McPherson R, Melander O, Mooser V, Nauck M, Nieminen MS, O'Donnell CJ, Peltonen L, Potter SC, Prokisch H, Rader DJ, Rice CM, Roberts R, Salomaa V, Sambrook J, Schreiber S, Schunkert H, Schwartz SM, Serbanovic-Canic J, Sinisalo J, Siscovick DS, Stark K, Surakka I, Stephens J, Thompson JR, Völker U, Völzke H, Watkins NA, Wells GA, Wichmann HE, Van Heel DA, Tyler-Smith C, Thein SL, Kathiresan S, Perola M, Reilly MP, Stewart AF, Erdmann J, Samani NJ, Meisinger C, Greinacher A, Deloukas P, Ouwehand WH, Gieger C. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet.* 2009 Nov;41(11):1182-90.
- 3) Zakai NA, McClure LA, Prineas R, Howard G, McClellan W, Holmes CE, Newsome BB, Warnock DG, Audhya P, Cushman M. Correlates of anemia in American blacks and whites: the REGARDS Renal Ancillary Study. *Am J Epidemiol.* 2009 Feb 1;169(3):355-64.
- 4) Zakai NA, Katz R, Hirsch C, Shlipak MG, Chaves PH, Newman AB, Cushman M. A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. *Arch Intern Med.* 2005 Oct 24;165(19):2214-20.
- 5) Zakai NA, Katz R, Jenny NS, Psaty BM, Reiner AP, Schwartz SM, Cushman M. Inflammation and hemostasis biomarkers and cardiovascular risk in the elderly: the Cardiovascular Health Study. *J Thromb Haemost.* 2007 Jun;5(6):1128-35.
- 6) Nalls MA, Wilson JG, Patterson NJ, Tandon A, Zmuda JM, Huntsman S, Garcia M, Hu D, Li R, Beamer BA, Patel KV, Akyzbekova EL, Files JC, Hardy CL, Buxbaum SG, Taylor HA, Reich D, Harris TB, Ziv E. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am J Hum Genet.* 2008 Jan;82(1):81-7.
- 7) Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet.* 2010 Jun 11;86(6):832-8.
- 8) Mukhopadhyay I, Feingold E, Weeks DE, Thalamuthu A. Association tests using kernel-based measures of multi-locus genotype similarity between individuals. *Genet Epidemiol.* 2010 Apr;34(3):213-21.