ARIC Manuscript Proposal # 1446

PC Reviewed: 11/11/08	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: CHARGE GWAS for white blood cell count

b. Abbreviated Title (Length 26 characters): White blood cell GWAS

2. Writing Group: CHARGE WBC working group

Writing group members: Aaron Folsom, David Couper, Anna Kottgen, Joe Coresh. Other authors from additional CHARGE cohorts. The plan is to maintain symmetry across cohorts.

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

First author: Aaron Folsom

Address: Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, 1300 South 2nd Street, Suite 300, Minneapolis, MN 55454

Phone: 612-626-8862 Fax: 612-624-0315
E-mail: folso001@umn.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: Aaron Folsom

Address: Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, 1300 South 2nd Street, Suite 300, Minneapolis, MN 55454

Phone: 612-626-8862 Fax: 612-624-0315
E-mail: folso001@umn.edu

3. Timeline: Summer 2008

4. Rationale:

White blood cells (WBCs) participate in atherogenesis and WBC count is associated positively with incidence of arterial thromboembolic diseases. Evidence suggests that WBC level is heritable. Some variants in identified genes determine plasma WBC levels,

but additional genes likely contribute. There have been no genome wide association studies (GWAS) of WBC level.

CHARGE (ARIC, CHS, Rotterdam, Framingham, AGES, and selected other cohorts) is doing a meta-analysis of GWAS findings related to WBC. The analysis is focusing on WBC level and subtypes (leukocytes, monocytes, lymphocytes, basophils and eosinophils). A meta-analysis will be conducted by Michael Nalls at NIA.

5. Main Hypothesis/Study Questions:

Gene variants can be identified that associate with levels of WBC count and WBC subtypes.

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).

Design: meta-analysis of GWAS studies.

Participating groups:

Framingham Study
Rotterdam Study
ARIC – analysis performed by David Couper
CHS
AGES

Necessary data:

- ~2.2 million HapMap Imputed SNPS
- Complete white blood cell data for a single cross-sectional timepoint in the format of absolute cell counts (which can be extrapolated by multiplying total white cell count by differential percentages if not already available).
    - total white blood cell count
    - neutrophil count
    - eosinophil count
    - basophil count
    - lymphocyte count
    - monocyte count
- Covariates
    - Age at hematology assay
    - Smoking status at hematology assay (0 = no smoking/1 = current smoker)
    - Gender

- Exclusion factors
  - Missing any white blood cell measures or covariate data
  - Greater than 2 standard deviations from population mean for any single white blood cell measure
    - This will remove likely pathological (and sub-clinically ill) outliers that can have a serious impact on WBC analyses
    - This exclusion will also help to normalize data distribution
- Data transformations to log of phenotype do not seem to be an issue, and are at the discretion of the individual study, as distributions may vary between study/instruments. Meta-analysis should methodologically compensate for slight inter-study differences. The other option would be to all transform or all not transform, then we could perform either a z-score based meta or an inverse-variance weighted meta of the betas.

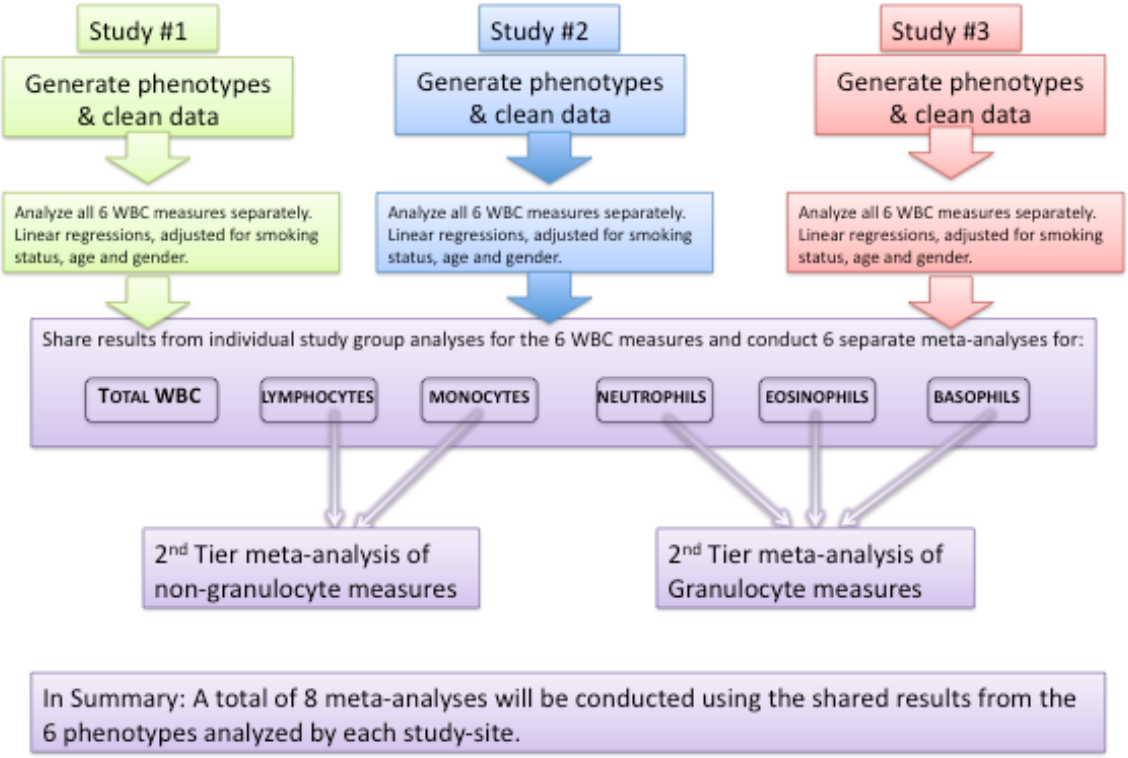
Primary Data Analysis:

- Data analysis for the white blood cell measures will include …
  - Each group should conduct their own analyses of the separate WBC measures
  - All available genotypes analyzed preferably using PLINK or a R loop if possible
  - Model format should be linear regression incorporating primary predictor of minor allele dosage (additive model) per SNP using all listed covariates
  - Model example
    - Phenotype ~ SNPmaf + age + gender + smoking
  - Results of each study-group's six phenotypic analyses should then be shared so meta-analyses can begin

Meta-Analyses:

- Once each study-group involved has shared their results, meta-analyses will be undertaken at NIA using the results from the different studies
- METAL will be used for meta-analyses
- Each individual phenotypic measure will undergo meta-analyses to compare across studies
- 2nd tier of meta-analyses will compare results from meta analyses for non-granulocyte measures (lymphocytes and monocytes) to determine SNPs specifically associated with a generalized non-granulocyte effect
  - This 2nd tier of analysis will be repeated for results from the granulocyte measures (neutrophils, eosinophils and basophils)

The flow chart on the next page summarizes the analysis plan.



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 the participants with RES_DNA = 'not for profit' restriction must be excluded if the data are used by a for profit group?
__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)?

None.

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

___ A. primarily the result of an ancillary study (list number* _____)
__X__ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 2006.03, 2007.02**)**

*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.