ARIC Manuscript Proposal # 1457

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

1.a. Full Title: CHARGE GWAS for factors VII, VIII, and von Willebrand factor (vWF)

b. Abbreviated Title (Length 26 characters): Hemostatic factor GWAS

2. Writing Group: CHARGE WBC working group

Writing group members: Weihong Tang, Saonli Basu, Xiaoxiao Kong, Eric Boerwinkle, Aaron Folsom. 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. __WT___ **[please confirm with your initials electronically or in writing]**

F**irst author**: Weihong Tang

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

Phone: 612-626-9140 Fax: 612-624-0315
E-mail: tang0097@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: Submitting manuscript to journal: Winter 2008-Spring 2009

4. Rationale:

Hemostatic factors participate in atherogenesis and several (including factors VII, VIII, and von Willebrand factor) are associated positively with incidence of arterial and venous

thromboembolic diseases. Evidence suggests that most hemostatic factor levels are heritable. Some variants in identified genes determine plasma hemostatic factor levels, but additional genes likely contribute. There have been no genome wide association studies (GWAS) of the hemostatic factors targeted.

CHARGE (ARIC, CHS, Rotterdam, Framingham, and selected other cohorts) is doing a meta-analysis of GWAS findings related to factors VII, VIII, and von Willebrand factor. ARIC data analysis will take place in Minnesota. A meta-analysis will be conducted by Nicholas Smith at CHS.

5. Main Hypothesis/Study Questions:

Gene variants can be identified that associate with levels of factors VII, VIII, and von Willebrand factor.

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).

Participating groups with GWAS data and FVII, FVIII, or vWF phenotypes:

- Framingham Study
- Rotterdam Study
- ARIC (Note: ARIC will have complete data in early November)
- CHS (Note: CHS has complete data in whites and will have complete data in blacks by late October)
- B58C

Comparability of phenotypes across cohorts

	FVII	FVIII	wWF
ARIC	activity	activity	antigen
CHS	activity	activity	NA
FHS	antigen	NA	antigen
Rotterdam	activity	activity	antigen *(pending)*
B58C	NA	NA	antigen

Phenotypes:

- vWF, factor VII, factor VIII

Exclusions (applies to all cohorts):

- Use of anti-coagulation therapy at the time of phenotype measurement

Imputation:

- Imputation to HapMap 2.5 M for all 23 chromosomes (including X sex chromosomes)
- Note: X chromosome still needs to be imputed. Preference is to have imputed data for the X chromosome but will work with unimputed data until the imputed data are available. Hopefully these will be ready in time for the paper's submission. We are waiting for guidelines from the CHARGE Analysis Subcommittee regarding the best approach to impute X chromosome. Christopher J. O'Donnell will address the X-chromosome issue at the next analysis meeting.

Analysis Approach for Continuous Measures (FVII, FVIII, vWF)

1) Phenotypes:

- Unstandardized (recommended by CHARGE analysis committee)
- Untransformed (large numbers of observations will overcome problems encountered with small numbers of observations and a non-normal distribution)*
- Based on 1 measurement per cohort
- Note: there were discussions about the advantages of log transforming the FVII phenotype if needed to account for the different assays (antigen vs activity). Preliminary results appear sufficiently robust that log-transformation does not seem necessary.
- Trait distribution in ARIC data will be checked and extreme outliers will be excluded before GWAS.

2) Model:

- Linear regression
- Additive genetic model (for X chromosome women and men will be modeled differently)
- Robust variance estimates
- For X chromosome, we may consider stratifying results by sex

3) Covariates:

- Age and sex adjusted (+ cohort/center where appropriate)
- For possible secondary analyses we would use the following multivariate adjusted: Age (continuous), sex, (+ cohort/center where appropriate), smoker (current, former, never), BMI (continuous), diabetes (y,n), CVD (y,n), TG (continuous), HDL-C (continuous), total cholesterol (continuous), alcohol (continuous, with 0 for nondrinker), systolic blood pressure, htnrx (yes,no), HRT. Notes: Coding HRT determined by cohort. Inclusion of dummy variable for fasting status is optional (Framingham only fasting subjects; ARIC and CHS: part of subjects non-fasting, therefore include dummy variable, no exclusion of non-fasters, Rotterdam has only non-fasting subjects). Adjustment for population stratification will be conducted as deemed appropriate by individual cohorts: Framingham will adjust for principle components as needed. For the fibrinogen analyses, these adjustments did not add to

the interpretation of the results. Other cohorts will probably not pursue this for these phenotypes.

4) Meta-analysis: Fixed-effect meta-analysis with inverse-variance weighting based on 2.5 M observed and imputed SNPs. No screens are proposed for the continuous measures although we should consider a lower MAF screen for rare variants that are only represented in few cohorts. Conduct both beta and z-score estimates for FVII since there are 2 types of assays. Q-Q plot for individual cohorts and meta-analysis will be examined.

5) Genome-wide significance threshold: 5×10^{-8} (same as fibrinogen paper)

6) Sensitivity analyses: step-wise analysis in areas where markers cluster over multiple genes

7) Replication: Data from CHS blacks will be available for FVII and FVIII (vWF was not measured in CHS)

Analysis Approach for Binary Measures (These analyses might be included in a separate paper)

1) Phenotypes:

- Unstandardized (recommended by CHARGE analysis committee)
- Based on 1 measurement per cohort
- Create dichotomous variable for the continuous phenotypes. I think we are most interested in low FVIII and vWF levels
- Low will be defined as the lower 5% of the distribution
- Comparison group (normals) not yet decided; RS is looking into possible options which may include a 5% versus 95% comparison, a 5% versus interquartile range (25%-75%) comparison, or a 5% versus 50% and above comparison; there was also discussion about removing the top 5% when assembling the comparison group.

2) Model:

- Logistic regression
- Additive genetic model
- Robust variance estimates
- For X chromosome, we may consider stratifying results by sex

3) Covariates:

- Age and sex adjusted (+ cohort/center where appropriate)
- Secondary analyses: (see adjustment list above)

4) Meta-analysis: Since inverse-variance weighting can be very biased for logistic regression,[1] we will use fixed-effect meta-analysis with effective sample size weighting based on 2.5 M observed and imputed SNPs. No screens are proposed for the binary measures analyzed using logistic regression although we should consider a lower MAF

screen for rare variants that are only represented in few cohorts. For FVIII and vWF, the homogeneity of odds ratio across the cohorts will be checked. If the odds ratios are homogeneous across studies, we can then estimate the common odds ratio using Mantel-Haenszel method or by logistic regression. We will conduct both beta and z-score estimates for FVII since there are 2 types of assays.

5) Genome-wide significance threshold: 5×10^{-8} (same as fibrinogen paper)

6) Sensitivity analyses: step-wise analysis in areas where markers cluster over multiple genes

7) Replication: Data from CHS blacks will be available for FVII and FVIII (vWF was not measured in CHS)

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.

Reference:

1. Tang JL. Weighting bias in meta-analysis of binary outcomes. *J Clin Epidemiol.* 2000;53:1130-1136.