

ARIC Manuscript Proposal #2540

PC Reviewed: 4/14/15
SC Reviewed: _____

Status: A
Status: _____

Priority: 2
Priority: _____

1.a. Full Title: Analysis of exome array and exome sequence data with Activated Partial Thromboplastin Time (aPTT) and Protein C – the ARIC Study

b. Abbreviated Title (Length 26 characters): exome data, aPTT and protein C

2. Writing Group: Weihong Tang, James Pankow, Nathan Pankratz, Weihua Guan, Saonli Basu, Lu-Chen Weng, Mary Cushman, Eric Boerwinkle, Aaron Folsom, others are welcome...

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]

First author: Weihong Tang, MD, PhD

Address: Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
1300 South Second Street, WBOB 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, MD, MPH

Address: the same as for Dr. Tang

Phone: (612) 626-8862

Fax: (612) 624-0315

E-mail: folso001@umn.edu

3. Timeline: Data analysis to start immediately (February, 2015). First draft of manuscript expected in six months depending on the timeline for obtaining replication data for aPTT in the other cohorts of CHARGE Consortium.

4. Rationale:

Venous thromboembolism (VTE) is a common disease with a high mortality rate. It is the third most common life-threatening cardiovascular disease after coronary heart disease and stroke.¹ Both environmental and genetic risk factors are important in the etiology of VTE.^{2, 3,4,5,6}

aPTT is a commonly used coagulation test to screen for deficiencies in the coagulation cascade. Protein C is one of the most important anticoagulant regulators of the coagulation pathway. Reduced levels of aPTT and protein C are important risk factors for VTE.⁷⁻¹⁰ In the ARIC Study, the risk of VTE was 2-3 times higher for participants with aPTT below the median value⁸ and 3.3 times higher for 1.1% of participants with plasma level of protein C values <2.0 mg/L at baseline compared with participants with higher values.⁹

Twin and family studies suggest that aPTT and protein C are heritable, with heritability of 0.36-0.50 for protein C^{11,12} and 0.43-0.83 for aPTT.^{11,12} Genetic linkage analysis identified a region on chromosome 16 that was strongly linked to protein C level in Spanish families.¹³ Since aPTT and protein C are important risk factors, it is possible that genetic factors influencing the levels of aPTT and protein C also influence the risk of VTE. Therefore, identification of genetic factors for aPTT and protein C may shed light on genetic etiology of VTE.

In ARIC, aPTT and protein C were measured at baseline in 11,422 whites and 4089 African Americans. We have conducted genome-wide association studies for aPTT¹⁴ in CHARGE and protein C in ARIC,^{15,16} as well as a candidate gene analysis of aPTT with the CArE IBC SNP array in ARIC.¹⁷ For protein C, common variants in *GCKR*, *PROC*, *PROCR*, and *EDEM2* loci have been associated with protein C levels and explained 14%-15% of the variance in protein C in EAs and AAs;^{15,16} for aPTT, common variants in *F5*, *HRG*, *KNG1*, *F11*, *F12*, *ABO*, and *VWF* have been reported and explained 28%-31% of the variance in aPTT in EAs and AAs.^{14,17}

Since common polymorphisms identified to-date explain only some of the heritability of both aPTT and protein C, it is possible there are additional genetic influences that are yet discovered, including rare variation in known and new genes. In addition, some of the identified variants are located in introns or outside of genes and do not seem to influence the structure or function of corresponding genes, which might reflect linkage disequilibrium with underlying functional variants that have not yet been investigated, but many of which are now available in ARIC through exome chip and exome sequence data.

5. Main Hypothesis/Study Questions:

The analytical plan outlined here encompasses exome chip variants genotyped for ARIC Whites and Blacks and exome sequence data generated for ARIC Whites and

Blacks through the NHLBI Exome Sequencing Project (ESP) and the CHARGE consortium. Analysis of the exome data will provide an opportunity to: (1) identify novel genes/loci that contribute to the interindividual variation of protein C and aPTT; (2) identify low frequency or rare variation in new genes and known candidate genes (e.g., those previously identified by GWAS or CARE IBC analyses) that influence these phenotypes, and investigate the additional contribution of these variants to the variance of these phenotypes.

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).

We will utilize the analytical approach and pipeline that have been established in ARIC as part of CHARGE (<http://depts.washington.edu/chargeco/wiki/SKATmeta>, renamed to seqMeta last year). Analyses will be conducted separately in ARIC Whites and Blacks. The skatCohort portion of the analytic pipeline produces Rdata objects for the ARIC study that may then be meta-analyzed with other cohorts, or across projects (e.g., meta-analysis of CHARGE and ESP exomes).

We will use two approaches to analyze protein C and aPTT:

1) Single variant association analyses

Single markers of a suitable frequency depending on the total sample size (e.g., >1% minor allele frequency or >10 minor allele count) will be analyzed for their association with protein C or aPTT in linear regression models. The genetic association analysis for protein C will be performed in PLINK. The analysis for aPTT will be performed in seqMeta which can meta-analyze association results across cohorts. Covariate adjustment is described below.

For protein C, we will combine the exome data with the CARE IBC array as the CARE IBC array data focused on common variants of candidate genes, including functional variants, and thus serves as a good supplement to the exome data. Unlike aPTT, analyses of protein C with CARE IBC array data in ARIC have not yet been implemented or published.

2) Gene-based analyses

For variants of low frequency (e.g., <1% minor allele frequency or another suitable cutpoint), we will evaluate the rare variants in aggregate within a gene in R using the seqMeta package. The SKAT, T1 and T5 tests will be performed. Rare variants may be further filtered to only include those of possible functional consequence (e.g., nonsynonymous, splicing, stopgain, or stoploss).

Meta-analysis

For aPTT, meta-analysis will be conducted across the participating studies for single variants using methodologies developed through the CHARGE consortium (i.e., seqMeta). So far, we have identified cohorts including LBC, MICROS, and CHRIS that have the aPTT and exome array data and are willing to collaborate. Meta-analysis of gene-based tests will also be conducted using methodologies developed through the CHARGE consortium (i.e., seqMeta). For protein C, to the best of our knowledge, ARIC is the only cohort with exome array or sequence data that has measured this phenotype. We will conduct collaboration and meta-analysis if new cohorts are identified for protein C and the exome data.

Phenotypes: protein C and aPTT. Distribution of the phenotypes will be evaluated for normality and the following techniques will be used to correct non-normality: log-transformation, exclusion of outliers, or winsorization of outliers.

Covariates: age at baseline, sex, and field center or principal components if appropriate.

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 to evaluate exome sequence and exome array data with protein C and aPTT.

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

11.b. If yes, is the proposal

- A. primarily the result of an ancillary study (list number* 1998.03)**
 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 shows you which journals automatically upload articles to Pubmed central.

References

1. Goldhaber SZ. Epidemiology of pulmonary embolism and deep vein thrombosis. In: Bloom AL, Forbes, C.D., Thomas, D.P., Tuddenham, E.G.D., ed. *Haemostasis and Thrombosis (3rd Edition)*. New York: Churchill Livingstone; 1994:1327-1333.
2. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. Apr 3 1999;353(9159):1167-1173.
3. Rosendaal FR. Risk factors for venous thrombotic disease. *Thromb Haemost*. Aug 1999;82(2):610-619.
4. Anderson FA, Jr., Spencer FA. Risk factors for venous thromboembolism. *Circulation*. Jun 17 2003;107(23 Suppl 1):I9-16.
5. Lowe GD. Can haematological tests predict cardiovascular risk? The 2005 Kettle Lecture. *Br J Haematol*. May 2006;133(3):232-250.
6. Cushman M. Epidemiology and risk factors for venous thrombosis. *Semin Hematol*. Apr 2007;44(2):62-69.
7. Tripodi A, Chantarangkul V, Martinelli I, et al. A shortened activated partial thromboplastin time is associated with the risk of venous thromboembolism. *Blood*. Dec 1 2004;104(12):3631-3634.
8. Zakai NA, Ohira T, White R, et al. Activated partial thromboplastin time and risk of future venous thromboembolism. *Am J Med*. Mar 2008;121(3):231-238.
9. Folsom AR, Aleksic N, Wang L, et al. Protein C, antithrombin, and venous thromboembolism incidence: a prospective population-based study. *Arterioscler Thromb Vasc Biol*. Jun 1 2002;22(6):1018-1022.

10. Koster T, Rosendaal FR, Briet E, et al. Protein C deficiency in a controlled series of unselected outpatients: an infrequent but clear risk factor for venous thrombosis (Leiden Thrombophilia Study). *Blood*. May 15 1995;85(10):2756-2761.
11. Souto JC, Almasy L, Borrell M, et al. Genetic determinants of hemostasis phenotypes in Spanish families. *Circulation*. Apr 4 2000;101(13):1546-1551.
12. Warren DM, Soria JM, Souto JC, et al. Heritability of hemostasis phenotypes and their correlation with type 2 diabetes status in Mexican Americans. *Hum Biol*. Feb 2005;77(1):1-15.
13. Buil A, Soria JM, Souto JC, et al. Protein C levels are regulated by a quantitative trait locus on chromosome 16: results from the Genetic Analysis of Idiopathic Thrombophilia (GAIT) Project. *Arterioscler Thromb Vasc Biol*. Jul 2004;24(7):1321-1325.
14. Tang W, Schwienbacher C, Lopez LM, et al. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J Hum Genet*. Jul 13 2012;91(1):152-162.
15. Munir MS, Weng LC, Tang W, et al. Genetic markers associated with plasma protein C level in African Americans: the atherosclerosis risk in communities (ARIC) study. *Genet Epidemiol*. Dec 2014;38(8):709-713.
16. Tang W, Basu S, Kong X, et al. Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood*. Dec 2 2010;116(23):5032-5036.
17. Weng LC, Cushman M, Pankow JS, et al. A genetic association study of activated partial thromboplastin time in European Americans and African Americans: the ARIC Study. *Hum Mol Genet*. Dec 30 2014;pii: ddu732. [Epub ahead of print].