ARIC Manuscript Proposal # 3129

PC Reviewed: 03/20/2018	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Analyses of Exome Chip and Exome Sequencing Data with FXI, D-dimer, and VTE

b. Abbreviated Title (Length 26 characters): exome with FXI, D-dimer, and VTE

2. Writing Group:

Writing group members: Weihong Tang, Nathan Pankratz, Mary Stimson, Susan Heckbert, Mary Cushman, James Pankow, Saonli Basu, and 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]**

F**irst author**: **Weihong Tang**
Address: Division of Epidemiology and Community Health
University of Minnesota

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 R. Folsom
Address: Division of Epidemiology and Community Health
University of Minnesota

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

3. Timeline:

Data analysis: as soon as possible upon the approval of the manuscript proposal
First draft of the manuscript: 3-4 months from manuscript approval date.

4. Rationale:

FXI is an important component of the coagulation pathway. D-dimer is a fibrin degradation product and marker of thrombin generation. Plasmas levels of FXI and D-dimer were associated with future risk of VTE in LITE.[1, 2] Large GWAS studies have

reported common variants (frequency ≥ 1%) associated with FXI and D-dimer.[3, 4] These variants together explained a modest amount of variance of the traits. The unexplained heritability could be attributable to low frequency, and rare variants.

Similarly, large scale candidate-gene and GWAS studies have associated common variants at more than a dozen of loci with VTE risk in unrelated patients or those ascertained from the general population. [5-14] However, most of the variants were associated with modestly increased risk for VTE (risk estimate < 2.0).[15] It is possible that there are additional genetic influences that are yet discovered, including rare variation in known and new genes. A newly published study, which applied whole exome sequencing approach to 64 VTE patients recruited from Hematology clinics, showed that rare, coding variants contributed to VTE risk, beyond what was identified by clinical thrombophilia testing.[16]

5. Main Hypothesis/Study Questions:

The analytical plan outlined here encompasses exome chip variants genotyped for ARIC Whites and Blacks and exome sequencing 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 inter-individual variation of FXI and D-dimer and risk of VTE;

(2) Identify low frequency or rare coding variants in new genes and known candidate genes that influence these phenotypes.

As part the LITE study, we will conduct the following analyses to address these questions.

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.

We will use two approaches to analyze the proposed phenotypes:

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 the continuous phenotypes in linear regression models and with VTE in binomial

function models using seqMeta, which can meta-analyze association results across cohorts. Covariate adjustment is described below.

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 (ie, genetic burden test). The SKAT test will be performed for the continuous phenotypes and T1 (or T5) test will be performed for VTE. We will select functional variants on the array or sequencing data, defined as missense, stop-gain, stop-loss, or splice site changes. When only longitudinal cohort studies are involved (eg, ARIC and CHS for VTE), a Cox regression will also be performed to analyze time-to-event data for VTE.

3) Meta-analysis
We will conduct collaboration with the CHARGE and INVENT[11] (for VTE) consortiums to meta-analyze association results across cohorts. Meta-analysis of single-variant and gene-based tests will be conducted using methodologies developed through the CHARGE consortium (i.e., seqMeta). A Bonferroni-correction will be applied based on the number of variants or genes tested.

Phenotypes: FXI, D-dimer, and VTE. Distribution of the continuous 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, sex, field center and principal components of ancestry if appropriate.

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 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?
_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

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

_X_ A. primarily the result of an ancillary study (list number* 2001.15 LITE)

___ 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/

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.cscc.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.

13. Per Data Use Agreement Addendum for the Use of Linked ARIC CMS Data, approved manuscripts using linked ARIC CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ___ Yes __X__ No.

References:

1. Folsom AR, Tang W, Roetker NS, Heckbert SR, Cushman M and Pankow JS. Prospective study of circulating factor XI and incident venous thromboembolism: The Longitudinal Investigation of Thromboembolism Etiology (LITE). *Am J Hematol*. 2015;90:1047-51.
2. Cushman M, Folsom AR, Wang L, Aleksic N, Rosamond WD, Tracy RP and Heckbert SR. Fibrin fragment D-dimer and the risk of future venous thrombosis. *Blood*. 2003;101:1243-8.
3. Sennblad B, Basu S, Mazur J, Suchon P, Martinez-Perez A, van Hylckama Vlieg A, Truong V, Li Y, Gadin JR, Tang W, Grossman V, de Haan HG, Handin N, Silveira A, Souto JC, Franco-Cereceda A, Morange PE, Gagnon F, Soria JM, Eriksson P, Hamsten A, Maegdefessel L, Rosendaal FR, Wild P, Folsom AR, Tregouet DA and Sabater-Lleal M. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet*. 2017;26:637-649.
4. Smith NL, Huffman JE, Strachan DP, Huang J, Dehghan A, Trompet S, Lopez LM, Shin SY, Baumert J, Vitart V, Bis JC, Wild SH, Rumley A, Yang Q, Uitterlinden AG, Stott DJ, Davies G, Carter AM, Thorand B, Polasek O, McKnight B, Campbell H, Rudnicka AR, Chen MH, Buckley BM, Harris SE, Peters A, Pulanic D, Lumley T, de Craen AJ, Liewald DC, Gieger C, Campbell S, Ford I, Gow AJ, Luciano M, Porteous DJ, Guo X, Sattar N, Tenesa A, Cushman M, Slagboom PE, Visscher PM, Spector TD, Illig T, Rudan I, Bovill EG, Wright AF, McArdle WL, Tofler G, Hofman A, Westendorp RG, Starr JM, Grant PJ, Karakas M, Hastie ND, Psaty BM, Wilson JF, Lowe GD, O'Donnell CJ, Witteman JC, Jukema JW, Deary IJ, Soranzo N, Koenig W and Hayward C. Genetic predictors of fibrin D-dimer levels in healthy adults. *Circulation*. 2011;123:1864-72.
5. Bezemer ID, Bare LA, Doggen CJ, Arellano AR, Tong C, Rowland CM, Catanese J, Young BA, Reitsma PH, Devlin JJ and Rosendaal FR. Gene variants associated with deep vein thrombosis. *Jama*. 2008;299:1306-14.
6. Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR and Psaty BM. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *Jama*. 2007;297:489-98.
7. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC, Zelenika D, Cambien F, Tiret L, Bertrand M, Dupuy AM, Letenneur L, Lathrop M, Emmerich J, Amouyel P, Tregouet DA and Morange PE. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One*. 2011;6:e25581.
8. Tregouet DA, Heath S, Saut N, Biron-Andreani C, Schved JF, Pernod G, Galan P, Drouet L, Zelenika D, Juhan-Vague I, Alessi MC, Tiret L, Lathrop M, Emmerich J and Morange PE. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood*. 2009;113:5298-303.
9. Heit JA, Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM and M DEA. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J Thromb Haemost*. 2012;10:1521-31.

10. Tang W, Teichert M, Chasman DI, Heit JA, Morange PE, Li G, Pankratz N, Leebeek FW, Pare G, de Andrade M, Tzourio C, Psaty BM, Basu S, Ruiter R, Rose L, Armasu SM, Lumley T, Heckbert SR, Uitterlinden AG, Lathrop M, Rice KM, Cushman M, Hofman A, Lambert JC, Glazer NL, Pankow JS, Witteman JC, Amouyel P, Bis JC, Bovill EG, Kong X, Tracy RP, Boerwinkle E, Rotter JI, Tregouet DA, Loth DW, Stricker BH, Ridker PM, Folsom AR and Smith NL. A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Genet Epidemiol*. 2013;37:512-21.
11. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, de Andrade M, de Visser MC, Wiggins KL, Suchon P, Saut N, Smadja DM, Le Gal G, van Hylckama Vlieg A, Di Narzo A, Hao K, Nelson CP, Rocanin-Arjo A, Folkersen L, Monajemi R, Rose LM, Brody JA, Slagboom E, Aissi D, Gagnon F, Deleuze JF, Deloukas P, Tzourio C, Dartigues JF, Berr C, Taylor KD, Civelek M, Eriksson P, Psaty BM, Houwing-Duitermaat J, Goodall AH, Cambien F, Kraft P, Amouyel P, Samani NJ, Basu S, Ridker PM, Rosendaal FR, Kabrhel C, Folsom AR, Heit J, Reitsma PH, Tregouet DA, Smith NL and Morange PE. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet*. 2015;96:532-42.
12. Hinds DA, Buil A, Ziemek D, Martinez-Perez A, Malik R, Folkersen L, Germain M, Malarstig A, Brown A, Soria JM, Dichgans M, Bing N, Franco-Cereceda A, Souto JC, Dermitzakis ET, Hamsten A, Worrall BB, Tung JY and Sabater-Lleal M. Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. *Hum Mol Genet*. 2016.
13. Hernandez W, Gamazon ER, Smithberger E, O'Brien TJ, Harralson AF, Tuck M, Barbour A, Kittles RA, Cavallari LH and Perera MA. Novel genetic predictors of venous thromboembolism risk in African Americans. *Blood*. 2016;127:1923-9.
14. Heit JA, Armasu SM, McCauley BM, Kullo IJ, Sicotte H, Pathak J, Chute CG, Gottesman O, Bottinger EP, Denny JC, Roden DM, Li R, Ritchie MD and de Andrade M. Identification of unique venous thromboembolism-susceptibility variants in African-Americans. *Thromb Haemost*. 2017;117:758-768.
15. Tregouet DA and Morange PE. What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies. *Br J Haematol*. 2017.
16. Lee EJ, Dykas DJ, Leavitt AD, Camire RM, Ebberink E, Garcia de Frutos P, Gnanasambandan K, Gu SX, Huntington JA, Lentz SR, Mertens K, Parish CR, Rezaie AR, Sayeski PP, Cromwell C, Bar N, Halene S, Neparidze N, Parker TL, Burns AJ, Dumont A, Yao X, Chaar CIO, Connors JM, Bale AE and Lee AI. Whole-exome sequencing in evaluation of patients with venous thromboembolism. *Blood Adv*. 2017;1:1224-1237.