ARIC Manuscript Proposal # 1140

1.a. Full Title: Association of a Novel Prothrombin (FII) Variant (PT20209) in African Americans with Stroke and Myocardial Infarct

b. Abbreviated Title (Length 26 characters): PT20209 and Stroke

2. Writing Group:
Writing group members:
Lisa Vincent
Kelly Volcik
Eric Boerwinkle
Kachi Illoh
Thomas H. Mosley
Kenneth K. Wu
Aaron R. Folsom
James Grotta
Dianna Milewicz

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

First author: Lisa Vincent
Address: 6431 Fannin MSB 6.015
Houston, Texas 77030
Phone: 713-500-6724 Fax: 713-500-0693
E-mail: lvincent@uth.tmc.edu

Senior author: Dianna M. Milewicz, MD, Ph.D.
Address: 6431 Fannin MSB 4.202
Houston, Texas 77030
Phone: 713-500-6715 Fax: 713-500-0693
E-mail: dmilewicz@uth.tmc.edu

Corresponding author for ARIC: Eric Boerwinkle, Ph.D.
Address: 1200 Hermann Pressler, Suite E447
Houston, Texas 77030
Phone: 713-500-9816 Fax: 713-500-0900
E-mail: Eric.Boerwinkle@uth.tmc.edu
3. **Timeline:**
   a. Data Preparation: Feb06
   b. Data/Statistical Analysis: March06-April06
   c. Manuscript Preparation: April06-May06
   d. Manuscript Revision: June06
   e. Manuscript Submission: July06

4. **Rationale:**
   The regulation of coagulation, fibrinolytic, and other cellular processes involves a multitude of plasma proteins and platelets. Variants of the activated plasma protein prothrombin have been linked to severe hemorrhagic phenotypes. In contrast, two specific variations are associated with an increased risk of venous thrombosis in primarily Caucasians and African Americans. The G20210A mutation has been associated with an approximately threefold higher risk of venous thrombosis in Caucasians\(^1\). When screening for the G20210A mutation as part of another study, another prothrombin variant, C20209T, was found to be present in four unrelated African American individuals\(^3\). Three of these individuals had a history of venous thrombosis or stroke, while the fourth had severe liver disease, which could have potentially masked a predisposition to thrombotic disease. In addition, scanning of recent literature revealed a small pilot study that identified a low prevalence (1.6\%) of this variant in a UK black population with a history of VTE or DVT\(^4\). No study has investigated the role of this gene variant on incident stroke or CHD, and no study has examined its impact in a population based sample of African-Americans.

   We have identified a large African American kindred exhibiting significant clinical history of premature strokes and myocardial infarcts. The family consists of approximately 76 members, with at least 12 members exhibiting some history of stroke or thrombosis. The patient also had an emboli surgically removed from her left brachial artery during the same hospital visit. Consequently, the proband was tested for all known thrombotic coagulation disorders, and all tests returned as normal. Interestingly, the individual was found to have the 20209T prothrombin gene variant. For pilot data, we screened 198 African American stroke patients and 90 African American controls. The variant was observed in four of the African-American stroke cases and none of the controls. We hypothesize that this C20209T prothrombin gene variant, currently only seen in African-Americans, results in an increased risk to premature stroke, coronary heart disease and deep venous thrombosis.

4. **Main Hypothesis/Study Questions:**
   a. Estimate the frequency of the prothrombin variant 20209 (C20209T) in a population-based sample of African-Americans.
   b. Investigate the association of 20209 carrier status with the incidence of stroke, CHD, and venous thromboembolism in African Americans. Two models will be fit to the data: First, adjusting for age and sex only. And second, adjusting for established disease-specific risk factors.
   c. A subset of the ARIC cohort (including both cases and non-cases) have been measured for prothrombin levels. In that subset, repeat the analyses described
above, but including prothrombin levels as a potential covariate/confounder. Because of the weighted case-cohort design of this subset appropriately weighted survival analysis will be used.

6.a. Data (variables, time window, source, inclusions/exclusions):
ARIC’s stroke and CHD incident case status will be the primary dependent variables. Independent variables include, but are not limited to, the PT C20209T genotype status and traditional risk factors such as age, gender, race, smoking status, and hypertension. The prothrombin (FII) level/activity will be used for effect modification analysis.

6.b. Analysis Plan:
1. Allele frequencies will be estimated by gene counting. Goodness of fit to Hardy-Weinberg expectations will be carried out using a chi-square goodness of fit test.
2. Because the frequency of the 20209T allele is low, heterozygotes and homozygotes will be combined for all genotype-phenotype analyses (i.e. T carriers).
3. Cox proportional hazards modeling will be used to investigate the association between 20209T carriers and CHD. Model 1 will include age and sex as covariates. Model 2 will include age, sex, BMI, HDL and total-cholesterol, diabetes, hypertension, and smoking status. Because there is a specific a priori hypothesis, a one-side test will be used.
4. Cox proportional hazards modeling will be used to investigate the association between 20209T carriers and stroke. Model 1 will include age and sex as covariates. Model 2 will include age, sex, BMI, HDL and total-cholesterol, diabetes, hypertension, and smoking status. Because there is a specific a priori hypothesis, a one-side test will be used.
5. Based on the biology of F2, we will investigate the association between this variant and deep vein thrombosis. However, since the number of cases is expected to be small, we consider these exploratory or descriptive analyses.
6. Because factor two levels are only available on a subset of the cohort, and this subset is not a random sample, SUDAAN will be used to taking into account sample weights.

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

b. If Yes, is the author aware that the file ICTDER02 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 ICTDER02 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 ICTDER02 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)? All of the closest related proposals concern the PT 20210 variant and its effect on CVD and stroke. The methods used to test for PT20210 can not extend to identifying if there is a variation at PT20209, thus it is not necessary to form a collaboration because we are looking at a novel, unique variant.

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)*__777; 595; 793__) – Data from these studies may be used to compare the results of the PT20209 with the results of the PT20210.

*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 List


