ARIC Manuscript Proposal # 1430

1a. Full Title: Genotype-by-smoking and the risk of atherosclerosis and its clinical sequelae: the ARIC Study

b. Abbreviated Title: G-S GWAS of atherosclerosis

2. Writing Group:

ARIC-CHD, stroke and IMT working group members from the CHARGE Consortium will be invited.

Thus far, ARIC writing group members include, Christy Avery, Kari North, Keri Monda, Gerardo Heiss, Richey Sharrett, Dan Arkin, Kelly Volcik, Anna Kottgen, David Couper, Christy Ballantyne, Tom Mosley, and Eric Boerwinkle (Senior Author)

Other authors from additional consortium cohorts will be included. Additional ARIC authors may be invited. 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. CA

First author: Christy Avery
Address: Department of Epidemiology
University of North Carolina at Chapel Hill
137 E. Franklin St, Suite 306
CB #8050
Chapel Hill, NC  27514
Phone: 919-966-8491    Fax: 919-966-9800
E-mail: christy_avery@unc.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):

Senior Author: Eric Boerwinkle
Human Genetics Center
The University of Texas Health Science Center at Houston
P. O. Box 20334
Houston, Texas 77225-0334
(713) 500-9816
Facsimile: (713) 500-0900
email: Eric.Boerwinkle@.uth.tmc.edu
3. Timeline:

- Statistical analyses: August – October, 2008
- Manuscript revision: January 2009
- Manuscript submission: February 2009

4. Rationale:

Several lines of evidence support an association between cigarette smoke exposure and atherosclerosis and its clinical endpoints. Cigarette smoke contains approximately 4,800 chemicals,\(^1\) of which more than 100 have been identified as carcinogenic and/or mutagenic.\(^2\) Cigarette smoke affects the initiation and progression of atherosclerosis through its effects on vasomotor dysfunction, inflammation, and lipid modification,\(^3\) factors that proceed any apparent structural and clinicopathologic disease manifestations.\(^4,5\) Nicotine, possibly the most-studied component of tobacco smoke, also has deleterious effects on the vasculature, and has been associated with increased platelet adhesiveness and aggregation,\(^6,7\) elevated fibrinogen,\(^8,9\) and decreased fibrinolysis.\(^10-12\) Although the association between cigarette smoking and coronary heart disease (CHD) is well established and consistent across age, sex, racial, and geographic strata,\(^13-20\) the mechanisms by which smoking initiates vascular disease are poorly understood.

It is well known that genetic variants may influence a population’s sensitivity to environmental stressors,\(^21-24\) for example through inherited disease resistance or mutations that moderate disease expression. A genetic effect may also be obscured unless the appropriate environmental context is considered.\(^25\) Although examined by few studies, several lines of evidence suggest that genetic variation may influence the association between cigarette smoke exposure and atherosclerosis and its clinical sequelae. Experimental animal research has demonstrated that the tobacco smoke mutagens benzo(a)pyrene and 1,3-butadiene can induce and stimulate a proliferative vascular smooth muscle cell phenotype.\(^26,27\) Studies investigating the role of DNA damage in atherogenesis also found higher levels of aromatic DNA adducts in vascular tissues (e.g. abdominal aorta and cardiac) of smokers\(^28,29\) and plasma cotinine levels were also predictive of bulky DNA adduct levels in humans.\(^30\) Thus, a mechanistic basis exists to examine how inherited genetic variation modifies the association between cigarette smoke exposure and atherosclerosis.

Genome-wide association studies (GWAS) interrogate whether variation across the human genome in the form of SNPs is associated with given phenotypes. GWAS are now widely recognized as powerful data-driven tools for identifying genetic variants related to common complex diseases such as atherosclerosis. Unlike GWAS of obesity traits, which have identified common variants (e.g. \textit{INSIG2}, \textit{FTO}) across numerous studies of adults and children,\(^31-34\) few concordant loci have been identified in GWAS of myocardial infarction (MI), stroke, or subclinical atherosclerosis measures. Exceptions include a GWAS in approximately 8,000 males and females from Iceland that identified \textit{ALOX5AP} and \textit{LTA4H}, both implicated in leukotriene B4 production,\(^35,36\) as well as a GWAS in the Wellcome Trust Case Control Consortium that identified nine loci strongly associated with coronary artery disease.\(^37\) Failure to detect an association or replicate it in an independent population could be due to many reasons including sample differences, inadequate statistical power, incomplete phenotype harmonization, or, importantly, failure to account for environmental interactions. In fact, genetic effects on
Atherosclerosis phenotypes are likely to be strongly influenced by the environment, suggesting the utility of approaches that incorporate environmental exposures.

We previously submitted a manuscript proposal (Kari North – MS #1045) to examine the association between approximately 100 candidate SNPs, cigarette smoking and three outcomes: coronary heart disease (CHD), stroke, and intimal–medial thickness. This proposal builds upon our earlier work by examining ~1,000,000 SNPs available on the ARIC sample through its collaboration with the Broad Institute. Phenotypes assessed included the following:

- Baseline IMT
- Incident CHD
- Incident stroke

5. Main Hypotheses/Study Questions:
   1) To evaluate evidence for genotype-by-smoking interaction and the following three outcomes:
      i. Baseline IMT
      ii. Incident CHD
      iii. Incident stroke

6. Design and Analysis:

Definitions and treatment of variables

Genotype: For these analyses we will likely employ a general model when sample size permits (cell size > 10). This model allows for flexibility of effect and has been shown to have good power to find an effect when the underlying mode of inheritance is unknown. In a general model two indicator variables are entered in the model with the major homozygote used as the referent category. This results in a regression coefficient for the heterozygote and the minor homozygote; an overall (or global) p-value is used to assess the statistical significance of the result.

Cigarette smoke: ARIC has several metrics for quantifying cigarette smoke exposure: intensity, duration, age at initiation, second hand smoke exposure, as well as current, past, and ever smoking status. We propose to initially measure cigarette smoke exposure using the ever-smoking metric. Although ever-smoking considers all participants who reported smoking > 400 cigarettes at study baseline as a homogeneous group, 90% of Caucasian and African American participants classified as ever-smokers reported ≥ 10 years of cigarette smoking. While there is sure to be some misclassification of exposure to cigarette smoke, the distribution of smoking duration and intensity indices in ARIC suggest that the majority of participants reporting ever-smoking actually experienced long-term exposure. Other smoking metrics (e.g. current smoking, duration and intensity of smoking) will also be evaluated.

Phenotype measures:
IMT- Baseline mean IMT will be defined as the weighted IMT average at the six carotid artery sites after adjustment for measurement drift and reader differences. Normality will be assessed prior to analysis.

To allow replication in other CHARGE populations, we will also consider additional IMT phenotypes, including: (1) Baseline measurements of the far-wall common carotid and internal carotid intima media thickness, defined as the ln(mean of max), (2) presence of plaque ("yes"/"no" or 25% stenosis); the working group will consider a second clinically-defined dichotomous variable of CCA IMT mean-max >=1.5 mm versus <1.5mm, after reviewing the numbers of subjects with this variable in the CHARGE cohorts.

Incident CHD- will be defined according to the CHARGE consensus definition

Incident Stroke- sudden onset focal neurological deficit of presumed vascular etiology lasting for at least 24 hours, or until death if the participant died less than 24 hours after onset of symptoms. Strokes were classified as ischemic, hemorrhagic or unknown type based on clinical and imaging criteria. For the analyses of total stroke, ischemic, hemorrhagic, and unknown strokes were included; subarachnoid hemorrhages were excluded.

Analysis strategy / statistical analysis

General linear models for IMT and hazard models for qualitative traits will include the main effects of SNP and ever-smoking as well as the interaction terms. Significant association of interaction will be assessed by means of a likelihood ratio test. We will test both minimally adjusted models (age, sex, field center) and more extensively adjusted models based on analysis of our directed acyclic graph. Systematic differences in ancestry will be addressed by stratification by race. We will account for population substructure within racial group using the principal components analysis method developed by Price and colleagues and implemented in the software EIGENSOFT. This method explicitly models ancestry and has higher power to detect true associations than other methods. Principal component factor scores will be incorporated into genetic models to account for population stratification in each of the samples.

Control for multiple comparisons

1 EFP; i.e., p< 1/# tests

Imputation

Imputation into HapMap 2.5 million SNPs using MACHv1.0.16.

Replication

Existing collaborations with the Framingham Heart Study (FHS), the Age Gene/Environment Susceptibility (AGES)-Reykjavik Study, the Rotterdam Study, and the Cardiovascular Health Study (CHS) will facilitate the replication of significant results. We
recognize that cigarette smoking history measures or prevalence may differ between these populations and we will seek to harmonize environmental characterization as best as possible. Brief descriptions of these studies follow.

FHS: The Framingham Heart Study is a longitudinal study of cardiovascular disease with the original participants (aged 30-62) recruited in 1948. A second-generation group composed of the original participants’ adult children and their spouses was enrolled in 1971. Finally, a third-generation was enrolled in 2002. Genotyping data is available via the Affymetrix 500K SNP chip on a sample size of approximately 9300 individuals. The FHS is a joint project of the NHLBI and Boston University.

AGES-Reykjavik: The Age Gene/Environment Susceptibility-Reykjavik Study was initiated in 2002 to examine risk factors in relation to disease and disability in old age. The sample is drawn from an established population-based cohort, the Reykjavik Study, of men and women born between 1907 and 1935 in Iceland. Genotyping data will be available in June via the Affymetrix chip.

Rotterdam: The Rotterdam Study is a prospective cohort study started in 1990 in Ommoord, a suburb of Rotterdam, among 10,994 men and women aged 55 and over. The main objective is to investigate the prevalence and incidence of and risk factors for chronic diseases in the elderly. Baseline measures were obtained between 1990 and 1993 and all participants were subsequently examined in follow-up examinations every 2-3 years. Genotyping data via the Illumina 370 chip are available for approximately 6000 individuals.

CHS: Study of risk factors for development and progression of CHD and stroke in people aged 65 years and older. The 5,888 study participants were recruited from four U.S. communities and have undergone extensive clinic examinations for evaluation of markers of subclinical cardiovascular disease. Genotype data is available on 2400 individuals free of clinical CVD at baseline at time of DNA collection; an additional 1600 individuals will be completed by mid-July 2008.

African Americans

While the above collaborations have been set up to facilitate replication in the white ARIC sample, it will not serve as a replication sample in the African American sample. The GWAS for the African American sample in ARIC is being conducted through the Candidate Gene Association Resource (CARe). We are of course very interested in this research question in the African American sample as well and will strive to set up collaborations with investigators associated with the other CARe cohorts. We will also take care to provide CARe with a copy of our manuscript proposal.

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 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?

___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”?

___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)?

Related manuscript proposals (#1045) were submitted by this group of investigators.

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 (AS #2006.03 & 2007.02)

___B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________)

2006.03 (Stampede and Geneva genotype funding in Caucasians) ;
2007.02 (CARe, genotyping in African Americans) --- IF INCLUDED

*ancillary studies are listed by number at http://www.csc.unc.edu/aric/forms/

The following ARIC Acknowledgements statement will appear in our paper-
The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

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.
References


the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316(5826):889-894.


