ARIC Manuscript Proposal # 1045

PC Reviewed: _11/04/04	**Status: _A___**	**Priority: _2___**
SC Reviewed: _11/04/04_	**Status: _A___**	**Priority: __2__**

1. a.Full Title: Gene-by-smoking interaction, subclinical atherosclerosis, incident CHD and stroke in ARIC (AS # 2002.06).

b. Abbreviated Title (Length 26 characters): Gene-by-smoking interaction

2. Writing Group (list individual with lead responsibility first):

Lead: Kari E. North
Assistant Professor
Department of Epidemiology
University of North Carolina Chapel Hill
Bank Of America Center
137 E. Franklin St., Suite 306
CB #8050
Chapel Hill, NC 27514
(919) 966-2148 (Voice)
(919) 966-9800 (Fax)
(919) 671-1029 (Cell)
kari_north@unc.edu

Writing group members: Molly S. Bray, Heejung Bang, Andrew Olshan, Gerardo Heiss, Danyu Lin (other interested ARIC investigators are welcome)

3. Timeline:

- Anticipate data will be available: Immediately
- Analysis to be completed: March, 2005
- First draft anticipated: December, 2005

4. Rationale:

While cigarette smoking is a well-established and potent risk factor for atherosclerotic vascular disease, individual susceptibility to smoking varies considerably, suggesting that modifiers such as genomic variation may influence an individual's responsivity to physiological and environmental toxins. Several key enzymes involved in the activation and detoxification of mutagenic tobacco smoke compounds, oxidative stress, and DNA damage are expressed in the tissues of the heart and vasculature and represent diverse mechanistic pathways for tobacco-induced pathology. Many of these enzymes have common polymorphisms (≥10% prevalence in the population) with known functional effects. Although restricted to a few enzymes and

hampered by shortcomings in design, a small number of studies have suggested that enzymatic activation and detoxification of tobacco smoke modifies the risk of certain cardiovascular outcomes associated with cigarette smoking.

The main goal of the proposed study is to evaluate common genetic polymorphisms that, in combination with exposure to tobacco smoke, may modify the risk of atherosclerosis and its clinical sequelae. An average of six polymorphisms, selected on the basis of their prevalence and functional significance, expression in relevant tissues, evaluation in previous studies and biologic plausibility, within 10 genes involved in activation, detoxification, oxidative stress, and DNA repair pathways will be evaluated . Four endpoints quantifying subclinical atherosclerosis and validated clinical atherosclerotic events will be studied in case-cohort/case-control mode: incident coronary heart disease, increased carotid artery intima-media thickness, peripheral arterial disease, and incident stroke. The proposed investigation is well-designed to address not only issues of how DNA sequence polymorphisms may promote or inhibit the damaging effects of smoking on the development of atherosclerosis but also larger issues of how genetic variation influences response to environmental insults and toxins that may ultimately result in chronic disease. The findings are expected to be of clinical and public health significance.

5. Main Hypothesis/Study Questions:

Genotyping has been completed for the first phase of this study. The genes and SNPs targeted to date are detailed in the table below.

Table 1: The 63 Single nucleotide polymorphisms (SNPs) considered in 10 candidate genes

Gene	Chromosome, Size	*Number of SNPs*	*Validated SNPs (dbSNP ID)*
Cytochrome P450, subfamily I, polypeptide 1, CYP1A1	*15q24.1, 6.1 kb*	5	rs1048943, rs4986884, rs4646421, rs1799814, rs4986879
Cytochrome P450, subfamily I, polypeptide 1, CYP1B1	*2p22.2, 8.5 kb*	*7*	*rs10916, rs1056836, rs10012, rs1800440, rs1056827, rs2855658, rs4987137*
Cytochrome P450, subfamily IIE, polypeptide 1, CYP2E1	10q26.3, 11.8 kb	8	rs2070673, rs2070674, rs2515641, rs6413419, rs2864985, rs2070677, rs2031920, rs915908
Myeloperoxidase, MPO	*17q23.2, 11.1 kb*	*6*	*rs2071409, rs2333227, rs2243828, rs2856857, rs7208693, rs2759*

Gene	Chromosome, Size	*Number of SNPs*	*Validated SNPs (dbSNP ID)*
Cytochrome P450, subfamily 2A, polypeptide 6, CYP2A6	*19q13.2*	*2*	*rs1801272, rs4986892*
X-ray repair complementing group 1, XRCC1	*19q13.2, 32.3 kb*	9	rs3213245, rs25487, rs1475933, rs25486, rs1799778, rs25489, rs915927, rs1799782, rs3213282
8-oxoguanine DNA glycosylase, OGG1	*3p26.2, 17.6 kb*	5	rs1052133, rs1805373, rs2072668, rs3219000, rs3219008
Apurinic endonuclease, APEX1	*14q11.2, 2.6 kb*	5	rs1048945, rs1760944, rs3136817, rs3136820
Excision repair cross-complementing rodent repair deficiency, XPD	*19q13.32, 20.7 kb*	7	rs50871, rs1052555, rs1052559, rs1618536, rs3916892, rs238406, rs1799793
X-ray repair complementing defective repair in Chinese hamster cells 3, XRCC3	*14q32.33, 17.85 kb*	9	rs861539, rs3212057, rs1799795, rs3212024, rs3212057, rs861531, rs1799794, rs1799796, rs3212038

Specific Aim 1. To evaluate the relationship between polymorphisms of selected Phase I (activation) enzymes, Phase II (detoxification) enzymes, Oxidate stress, and DNA repair enzymes and the risk incident coronary heart disease, increased carotid artery intima-media thickness, peripheral arterial disease, and incident stroke. The main effect for each polymorphism and the joint effect (interaction) of each polymorphism with tobacco exposure will be estimated.

Specific Aim 2. To evaluate the relationship between multiple genes (identified in aim 1) and the risk of 4 CVD endpoints. The effect of the polymorphisms in aim 2will be determined by estimating the joint effect (interaction) of selected polymorphisms and their interaction with tobacco exposure.

Analytical Strategy to Address Specific Aims

1. Assessment of Population Stratification

Tests of Hardy-Weinberg equilibrium (HWE) will be performed for each polymorphism in the cohort using the control samples.

2. Haplotype analysis

An important aspect of this project will be the ability to test multiple genotypes in predicting CVD and associated traits. One possible approach is to consider haplotypes. It is important to account for phase uncertainty in the analysis of haplotype-disease associations. We will use the best available methods and develop new methods if necessary. The general methodology of Zeng and Lin (2004) enables one to estimate haplotype-by-smoking interactions in cohort and case-control studies. We are currently developing similar methods for case-cohort studies.

Reference: Zeng D, Lin DY (2004). Estimating haplotype-disease association with potentially pooled genotype data. Genetic Epidemiology, in press.

Specific Aim 1. To evaluate the relationship between polymorphisms of selected Phase I (activation) enzymes, Phase II (detoxification) enzymes, Oxidate stress, and DNA repair enzymes and the risk of the 4 CVD endpoints. The main effect for each polymorphism and the joint effect (interaction) of each polymorphism with tobacco exposure will be estimated.

We will use the Cox proportional hazards (PH) model to analyze the effect of genetic variation within the context of smoking status on time to incident CHD and incident stroke. The Cox PH model allows for multifactorial designs and the inclusion of continuous (and categorical) covariates and their joint effects with censored survival time as the outcome variable. Logistic regression methods will be used to estimate the odds ratio in the cross-sectional analysis of the IMT and PAD endpoints. Given the case-cohort design, the weighted analytic technique described by Prentice will be applied. Interaction between genotype and smoking on the risk of 4 CVD endpoints will be examined on both the multiplicative and additive scales. A Wald χ^2 will be used to test for multiplicative interaction and an additive interaction will be assessed by testing the interaction contrast ratio (ICR). The presence of interaction will be evaluated both with and without covariate adjustment.

Specific Aim 2. To evaluate the relationship between multiple genes (identified in aim 1) and the risk of 4 CVD endpoints. The effect of the polymorphisms in aim 2will be determined by estimating the joint effect (interaction) of selected polymorphisms and their interaction with tobacco exposure.

Informed by the association analyses outlined above and published data on the biological effects associated with each pathway of interest, specific gene-gene interaction analysis will be explored. All interaction testing will be hypothesis-driven based on biologically plausible interactions between pathways, and not be conducted in an exploratory manner. As above, the Cox PH model will be applied to the incident disease phenotypes and logistic regression analysis will be applied to the other 2 CVD endpoints. If power allows, we will evaluate the relationship between polymorphisms of selected tobacco activation and detoxification enzymes and the risk of CVD among subgroups defined by age, gender, and ethnicity/race.

Because multiple metabolic polymorphisms will be evaluated in these analyses, the potential for Type I error resulting from multiple comparisons exists. We plan to address this concern by implementing haplotype analysis as described above and hierarchical regression (HR), also known as random effect models. HR has been shown to greatly improve the accuracy of unstable estimates, especially when multiple exposures (in this case, polymorphisms) are under study and data are limited. In general, we will use restricted maximum likelihood (REML) methods for parameter estimation. REML methods provide less biased variance component estimates than ML methods with finite samples. To test various relevant hypotheses, a Wald-type test as well as a likelihood ratio test will be employed. Different covariance structures will also be examined.

All of the analyses summarized above will be conducted uncer the lead author's direction using the analysis programs provided by the ARIC Coordinating Center, or in the case of haplotype analysis, provided by Danyu Lin. All results will be reviewed by the writing group to guide additional analyses and determine the number of manuscripts to be written.

6. Data (variables, time window, source, inclusions/exclusions):

a. Genotype data collected as part of the ARIC ancillary study AS #2002.06. The specific genes considered and SNPs genotypes are listed in Table 1.

b. Case status on four endpoints quantifying subclinical atherosclerosis and clinical atherosclerotic events from 1987-1998, including: (1) incident coronary heart disease (CHD)(non-fatal and fatal MI, acute coronary syndromes, and revascularization procedures), (2) incident cerebrovascular disease (stroke and transient ischemic attacks), (3) prevalent peripheral arterial disease (PAD)(ankle-arm blood pressure index assessed by DINAMAP, and (4) prevalent carotid atherosclerosis (increased carotid intima-media thickness assessed by B-mode ultrasound).

c. Control status for these four endpoints: (1) random cohort control (for incident stroke and CHD), (2) intima media thickness prevalent controls, and (3) prevalent peripheral arterial disease controls.

d. Information on demographic covariates (e.g., age, gender, race, center, etc.)

e. Information on other cardiovascular disease risk factors, to use as covariates (e.g., smoking, LDL cholesterol levels, BMI, etc.)

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? __X__ 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)?

None

11. 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.