1.a. Full Title: Association between gene variants and incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 25 characters):
Genetic risk for incident CHD

2. Writing Group:
Dov Shiffman, Alanna Morrison, Charles Rowland, Judy Louie, Lance Bare, David Ross, Joseph Catanese, Andre Arellano, James Pankow, Josef Coresh, Mary Malloy, John Kane, Stephen Ellis, James Devlin, Eric Boerwinkle and other ARIC authors, if desired.

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

Dov Shiffman
Address: Celera
1401 Harbor Bay Parkway
Alameda, CA 94502
Phone: (510) 749-4354 Fax: (510) 749-6200
E-mail: dov.shiffman@celera.com

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

Eric Boerwinkle
Address: Human Genetics Center
University of Texas at Houston Health Science Center
1200 Herman Pressler; Suite 453E
Houston, TX 77030
Phone: (713) 500-9800 Fax: (713) 500-0900
E-mail: Eric.Boerwinkle@uth.tmc.edu

3. Timeline: All analyses will be carried out at the University of Texas Health Science Center at Houston under the supervision of Dr. Eric Boerwinkle. Genotyping is part of ancillary study 2004.11 and will start immediately; analyses and manuscript preparation is projected to take place over the next year. The data will be sent to the ARIC
coordinating center at the time that a draft manuscript is circulated to ARIC internal review.

4. Rationale:

Coronary heart disease (CHD), the development and progression of coronary atherosclerosis and the resultant morbidity, is the most common cause of death in industrialized countries and is expected to be the most common cause of death worldwide by 2020\(^1\). CHD is a complex disease with an underlying genetic susceptibility\(^2\). Identifying gene variants associated with CHD can improve our understanding of disease pathophysiology and improve risk assessment. Many genetic polymorphisms have been reported to be associated with CHD\(^3\) and these reports frequently prompt investigation of the association between the genetic polymorphism and CHD in other populations. However, many subsequent studies fail to replicate the initially observed association. Because only those polymorphisms that are reproducibly associated with disease will contribute to our understanding of CHD, initial observations of association between polymorphisms and disease should be examined in population-based prospective studies that have sufficient power to detect the hypothesized associations.

We have been investigating the association between 20,009 putative functional single nucleotide polymorphisms (SNPs) and MI in two case control studies. We have identified 77 SNPs whose risk alleles are nominally significantly associated with MI in both of these studies and the effect is in the same direction in both studies. Thirty-four of these SNPs have previously been genotyped and investigated for an association with CHD in ARIC [Manuscript proposal #1095 & 1142]. We propose to investigate the remaining 43 SNPs in the ARIC Study for association with incident CHD and incident MI.

5. Main Hypothesis/Study Questions:

Main hypothesis: Genetic variants that have been associated with MI in two antecedent case-control studies are associated with incident CHD and/or incident MI in the ARIC Study.

Study questions:

1. Are the gene variants tested in this study associated with incident CHD and/or incident MI?
2. How will such associations change after adjustment for traditional CHD risk factors?
3. Will the association of such genetic variants with incident CHD and/or incident MI be present in both African American and Caucasian individuals?

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).

Study design: Prospective follow-up of all ARIC participants meeting the inclusion criteria from baseline (visit 1, 1987-1989) through January 1, 2003.

The SNPs will be tested in ARIC as part of collaboration between scientists at Celera and Dr. Boerwinkle as described in ARIC Ancillary Study 2004.11. These SNPs previously demonstrated significant association with MI in two antecedent case-control studies in white participants. These two case-control studies which compare cases with a history of myocardial infarction to controls with no history of myocardial infarction will be described in the manuscript.

Inclusions/exclusions: Exclusions prior to analysis in ARIC involve the removal of individuals who at baseline had a positive or unknown history of stroke or stroke symptoms, positive history or missing data for prevalent CHD, Blacks not from Jackson, MS or Forsyth County, NC, race other than Black or White, and individuals with restricted DNA use. 14,215 participants remain after these exclusions. The number of individuals excluded due to missing or unknown genotype will depend on the SNP under investigation.

Outcome: The primary outcome measure will be time from enrollment to the first occurrence of a component of the CHD endpoint. Incident CHD cases up to 2004 (or the latest available data) will be identified from the inc_by04 dataset. In secondary analyses, MI endpoint will be defined definite or probable MI (n = 842), silent MI between examinations (ascertained by electrocardiogram) (n = 110), definite fatal MI (n = 52).

Other variables of interest: Traditional risk factors used to adjust estimates of genetic risk include the following baseline information: systolic and diastolic blood pressure, hypertension medication use, total cholesterol, HDL-cholesterol, LDL-cholesterol, diabetes status, smoking status, gender and family history (age of mother’s MI, age of father’s MI). Other covariates may be selected, depending on the specific SNP of interest and the hypothesized function of the gene containing this SNP.

Data analysis

Data checks: Hardy-Weinberg equilibrium will be checked by race for each SNP by using the chi-square goodness-of-fit test as well as using the Fisher exact test.

Survival analyses: Event-free participants will be followed until the earliest of December 31, 2004, the date of last contact, or death. Incidence rates of CHD or MI will be calculated using person-time methods. Kaplan-Meier estimates of event free survival will be computed, and log-rank tests will be used to compare survival curves among the genotypes.
In regression analyses, an additive genetic model will be assumed (unless pre-specified otherwise based on data from antecedent studies). Risk allele will be determined based on antecedent studies. Genotype will be coded as 0 (zero copies of the risk increasing allele), 1 (one copy of the risk increasing allele), or 2 (two copies of the risk increasing allele).

Each SNP will be tested for association with incident CHD and incident MI in Cox proportional hazard analyses separately for each race. Those with a p-value of < 0.1 (using a pre-specified risk allele) in these analyses will be considered for further analyses. Cox proportional hazard regression will then be used to estimate the effect size (hazard ratio of incident CHD and of incident MI). Subsequent multivariate models will include basic variables (age, sex), traditional risk factors at baseline (hypertension, diabetes mellitus, LDL-cholesterol, HDL-cholesterol), and relevant potential intermediate variables depending on the putative function of the gene in which the SNP is located.

**Determination of statistical significance:**
We recognize the limitations of screening a large panel of genetic variants as risk factors for CHD. Rather than using a Bonferroni correction and controlling the overall type-I error rate (family-wise error rate) at a level of 0.05, we propose to critically evaluate significant findings at the \( \alpha \)-level of 0.05 after multivariate adjustment for each SNP by applying the false positive report probability (FPRP) as proposed by Wacholder et al.\(^5\) The FPRP is the probability of no true association between a genetic variant and disease given a statistically significant finding. It is based on 1) the prior probability of a true association of the tested genetic variant with the disease, 2) the observed p-value, and 3) the statistical power to detect the effect size of the alternative hypothesis at the given \( \alpha \) level, based on sample size, allele frequency, and the specified effect size for the presumed association under the alternative hypothesis. Since prior probability estimates could be somewhat subjective, a range of prior probabilities will be considered for each variant. We assume that 1 out of 3,000 SNPs with a potential to affect gene product function confers measurable risk of CHD, we further assume that each of the two antecedent studies resulted in a 10 fold enrichment of true positive SNPs, thus for each SNP in the study we initially assume a prior probability range of 0.003 to 0.3. This prior probability would be further modified by considering the specific potential of each SNP to affect protein function (e.g. non-synonymous SNPs that cause a non-conservative amino acid substitution would have higher prior probability than SNPs that cause a conserved amino acid substitution) and by the nature of the antecedent studies data (e.g. SNPs that are associated in both antecedent studies using an unstratified analysis would be considered to have higher prior probability).

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 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
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)?

    MS 1095: Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) Study using a genetic risk score  
    MP 1142: Genetic risk of Coronary Heart Disease in the Atherosclerosis Risk in Communities (ARIC) study: Application of a Genetic Risk Score

Both of these manuscripts are part of this same ancillary study. Therefore, the investigators can assure lack of overlap or duplication.

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 (2004.11)     
    ______ 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/

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


