1.a. Full Title: Genetic risk score for early-onset CHD in the Atherosclerosis Risk in Communities Study.

b. Abbreviated Title (Length 25 characters):
Genetics of early-onset CHD

2. Writing Group:
Dov Shiffman, James Devlin, Charley Rowland, Lance Bare, Eric Boerwinkle and other authors from Celera, ARIC, and from a validation study group.

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 for ARIC internal review.

4. Rationale:

Although genetic variants contribute to risk of coronary heart disease (CHD), the algorithms currently used to predict risk of CHD do not consider individual genetic variants since, in common complex disease like CHD, the contribution from any single gene variant to the disease is small. Morrison et al. (MP#1095) and Bare et al. (MP#1142) have shown that aggregating the risk associated with a number of genetic variants into a composite genetic risk score (GRS) can be used to assess CHD risk in ARIC. Early-onset CHD is known to have a larger genetic component than late onset CHD\(^1\) and has been shown to be poorly diagnosed by current risk assessment algorithms\(^2\). Therefore, we propose to calculate a genetic risk score for early-onset CHD (<65 years of age for females, <60 for males), and validate this risk score in an additional, external study. This genetic risk score manuscript differs from those previously published because: it considers only early-onset CHD, it introduces the newly discovered variant in 9p21, it considers alternative methods for assembly of risk score, and finally, the risk score will be validated in an external population.

5. Main Hypothesis/Study Questions:

Main hypothesis: A combination of SNPs can predict early-onset CHD in ARIC and be validated in an external population.

Study questions:

a) Are SNPs that were shown to be associated with CHD in multiple studies in addition to ARIC, also associated with early-onset CHD in ARIC?

b) What is the risk for early-onset CHD based on an early-onset risk score in ARIC?

c) Can we validate in an external population an early-onset risk score that was developed in ARIC?

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 December 31, 2004.

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.
SNPs tested in ARIC will be the 5 SNPs described by Bare et al. For these 5 SNPs, the risk allele was associated with risk of CHD in at least two antecedent studies as well as in ARIC. One more SNP (rs10757274) described by McPherson et al. also meets these criteria and will be included in the analysis.

**Inclusions/exclusions:** In ARIC exclusions prior to analysis 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 early-onset CHD endpoint. Early onset CHD is defined as first event CHD occurring prior to age 65 in females and prior to age 60 in males. Since there are only about 100 incident CHD events among white males under age of 55, we chose 60 as the age cutoff for early onset CHD in males. There are about 220 early onset CHD events among white females and about 300 early onset CHD events among white males using these cutpoints. The secondary outcome will be time from enrollment to the first occurrence of early onset MI (females <65 years of age, males <60 years of age).

**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).

**Data analysis**

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

**Survival analyses:** Analysis will be carried out in white and African American population separately. Event-free participants will be followed until the earliest of the date participants reach age 60 if male or 65 if female, December 31, 2004, the date of last contact, or death. Incidence rates of early-onset CHD 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 early-onset CHD 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 (systolic and diastolic blood pressure, hypertension medication use, total cholesterol, HDL-cholesterol, LDL-cholesterol, diabetes status, smoking status, gender and family history), and relevant potential intermediate variables depending on the putative function of the gene in which the SNP is located.

**Early-onset CHD risk score**

SNPs that are found to be associated with early-onset CHD (P<0.1 in an additive model) will be included in the early-onset risk panel. SNPs in the panel will be coded as additive, unless the evidence from ARIC and all the antecedent studies unanimously indicate a different model.

Early onset risk score will be calculated for each individual based on the genotype for the SNPs in the risk score panel. For a SNP in an additive model, risk homozygotes will be coded as -1, heterozygotes will be coded as 0, and nonrisk homozygotes will be coded as 1. For a SNP in a recessive model, the coding will be -1, -1, and 1 for nonrisk homozygotes, heterozygotes and risk homozygotes, respectively. For SNP in a dominant model the coding will be -1, 1, 1. This coding was chosen so that the risk homozygotes would be coded 2 levels higher than the non risk homozygotes in all three inheritance models considered (additive, recessive and dominant). The association with early-onset CHD will be determined using a Cox model with the early-onset risk score as a continuous variable. Cut points that compare individuals with high risk to those with low risk will also be considered.

**Validation in an external population**

We will test the association of the early-onset risk score with early-onset CHD in another study using the SNPs and coding that was developed in ARIC. We are in discussions with several groups, the criteria we are considering in choosing a validation study are: definition of early onset CHD, power to replicate findings in ARIC, prospective vs. case-control study design, and ethnicity.

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
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

