ARIC Manuscript Proposal # 1152

PC Reviewed: 04_/_18_/06 Status: __A__ Priority: _2__
SC Reviewed: _04/19/06___ Status: __A__ Priority: __2__

1.a. Full Title: **Genomic Predictors of Sudden Cardiac Death**
b. Abbreviated Title (Length 26 characters): **SCD Genomics**
2. Writing Group:
Writing group members:

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

1. **First author**: **Wendy Post, M.D., M.S.**

Blalock 910
Johns Hopkins Hospital
600 N. Wolfe Street
Baltimore, MD 21287
Phone: 410 955-1780 Fax: 443 287-0121
E-mail: wpost@jhmi.edu

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

Co- PI on ancillary study:
Aravinda Chakravarti, Ph.D.
McKusick - Nathans Institute of Genetic Medicine
Professor of Medicine, Pediatrics, Molecular Biology & Genetics
Johns Hopkins University School of Medicine
Broadway Research Building, Suite 579
733 N. Broadway
Baltimore, MD 21205
tel: (410) 502-7525, fax: (410) 502-7544
aravinda@jhmi.edu

CHS sponsor:
Ronald J Prineas MD PhD
Professor
Department of Public Health Sciences
Wake Forest University Health Sciences
2000 W. First Street, Suite 505
Winston-Salem NC 27104
(336)716-7441 Fax: (336)716-0834
rprineas@wfubmc.edu

Collaborators: **(in alphabetical order)**

a) Dan Arking, Ph.D.
Post-doctoral Fellow
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine
733 N. Broadway, Room 580
Baltimore, MD 21205
(410) 502-7531, Fax (410) 502-7544
arking@jhmi.edu
b) Eric Boerwinkle, Ph.D.
Health Science Center at Houston
UT School of Public Health
Human Genetics Center
1200 Herman Pressler Road, Suite E453
Houston, TX 77030-
(713) 500-9816 Fax: (713) 500-0900
eric.boerwinkle@uth.tmc.edu
c) Gregory L. Burke, M.D., M.Sc.
Professor and Chair
Department of Public Health Sciences
Wake Forest University School of Medicine
Medical Center Blvd.
Winston-Salem, NC 27157
Phone: (336) 716-2930, Fax: (336) 716-5425
gburke@wfubmc.edu
d) Josef Coresh, M.D.
Professor of Epidemiology and Medicine
Johns Hopkins University
2024 E. Monument Street
Baltimore, MD 21287
410 955-0495
coresh@jhu.edu
e) Linda Kao, Ph.D.
Assistant Professor of Epidemiology
615 N. Wolfe St.
Room W6513
Baltimore, MD 21205
410-614-0945, fax 410-955-0863
wkao@jhsph.edu
f) Eduardo Marban, M.D., Ph.D.
Director
Johns Hopkins Reynolds Clinical Cardiovascular Center
Division of Cardiology, Dept of Medicine
Blalock 910
600 N. Wolfe St.
Baltimore, MD 21287

410 955-3097, fax 410 955-7953
marban@jhmi.edu
g) Bruce M. Psaty, MD, PhD
Professor, Medicine & Epidemiology
University of Washington
Cardiovascular Health Research Unit
1730 Minor Avenue, Suite #1360
Seattle, WA 98101-1448
206/287-2777, fax: 206/287-2662
Email: psaty@u.washington.edu
h) Tom Rea MD MPH
Assistant Professor of Medicine
Cardiovascular Health Research Unit
Metropolitan Park, East Tower Suite 1360
1730 Minor Ave
Seattle WA 98101-1448
206.287.2777 fax 206.287.2662
i) David Siscovick, MD, MPH
Professor of Medicine
University of Washington
Cardiovascular Health Research Unit
1730 Minor Avenue, Suite 1360
UW Box 358085
Seattle, WA 98101
206-287-2777fax: 206-287-2662
dsisk@u.washington.edu
i) Nona Sotoodehnia, MD, MPH
Division of Cardiology, University of Washington,
Cardiovascular Health Research Unit
1730 Minor Avenue, Suite 1360
UW Box 358085
Seattle, WA 98101
206-287-2777 fax: 206-287-2662
nsotoo@u.washington.edu
k) Peter Spooner, Ph.D.
Executive Associate Director
Johns Hopkins Reynolds Clinical Cardiovascular Center
Division of Cardiology, Dept of Medicine
Blalock 910
600 N. Wolfe St.
Baltimore, MD 21287
410 614-5745
pspoone1@jhmi.edu
l) Gordon Tomaselli, M.D.
Professor of Medicine
Vice-Chair for Research, Dept. of Medicine

Ross Research Bldg 844
720 North Rutland Ave
Baltimore, MD 21205-2196
(410) 955-2774, fax (410) 502-2096
gtomasel@jhmi.edu

3. Timeline: Analysis to begin soon after approval is obtained.

4. Rationale:

Background and rationale

This manuscript proposal involves a collaborative effort funded through the Johns Hopkins University Donald W. Reynolds Clinical Cardiovascular Research Center. Please see the ancillary study proposal for details of the background and rationale. We propose to evaluate the associations between CAPON SNP's (and 4 other genes/loci) and SCD (and QT interval) in the SCD collaborative network (ARIC/CHS). These genes/loci were identified from our genome wide association study.

Genome wide analysis: We employed a novel technology to perform genome wide association (GWA) studies using ~115,000 SNPs distributed across the genome to map genetic determinants of SCD susceptibility. This study was performed in collaboration with Dr. Stefan Kaab of the Ludwig-Maximilians University in Munich, Germany in the KORA ("Cooperative Research in the Region of Augsburg") population. We chose to study the QT interval as our first phenotype of interest since it is a continuous trait, for which we could sample from the extremes of a general population. In addition, known monogenic disorders of SCD are associated with a prolonged QT interval (Long QT syndrome), the QT interval is known to be moderately heritable in a general population, electronic measurement of the QT interval is very reliable, and QT interval is associated with risk for CVD events in the general population. [1,2]

To minimize false positive findings, we took a 3 stage approach. In stage 1, we chose individuals from each of the extremes of the QT interval (7.5^{th} and 92.5^{th} percentile) between the ages of 25 and 75 years from among 2171 women in the KORA S4 study. Women were selected to avoid heterogeneity due to sex difference. Linear regression analysis was performed to "adjust" the QT interval for age and heart rate (RR interval) prior to choosing the extremes in the distribution. One hundred individuals from each extreme were genotyped for ~100,000 SNPs using Affymetrix Centurion arrays. While no single SNP reached genome-wide significance, several loci exhibited multiple SNPs with large differences in allele frequency between extremes of QT-interval ($p < 0.0001$). In stage 2, the 10 most significant SNPs (all $p < 0.0001$) from the genome wide screen were genotyped in the next 200 females from each extreme. Based on our current understanding of the biology of cardiac repolarization, we also selected *a priori* 45 candidate genes, implicated in short QT syndrome (SQTS) or LQTS, cardiac cellular electrophysiology, or homologous to the selected genes, each having at least one SNP represented on the array within 10 kb of its 5' or 3' UTR. We used a less stringent significance threshold for candidate gene SNPs to be followed up in stage II ($p<0.01$) since their prior probability of involvement is higher than that for anonymous markers. In stage III, anonymous SNPs significant at $p<0.005$ and candidate gene SNPs with $p<0.01$ in stage II, were genotyped in the remaining 3,366 subjects of both genders.

Importantly, significance tests were performed separately on the men and women specific to stage III (i.e., excluding the 600 females analyzed in stages I and II), allowing stage III to serve as a validation study for stages I and II.
One SNP, in intron 1 of the CAPON gene, a regulator of nNOS which plays a key role in modulating cardiac contractility, was significantly associated with QTc RAS ($p < 10^{-7}$, excluding the women in the extremes of QT) with ~5 ms difference in QTc RAS between the two homozygous genotype groups. The results with CAPON were validated in two independent samples of 2,646 subjects from Germany (KORA F3) and 1,805 subjects from the US Framingham Heart Study. Results of this study demonstrate the feasibility and the potential of genome-wide SNP association studies for identification of novel genes for complex traits.
We propose to evaluate the associations between CAPON SNP's and SCD (and QT interval) in the SCD collaborative network (ARIC/CHS). Since CAPON is a large gene, we are unable cover it comprehensively. We have elected to cover the region of the gene that showed the strongest associations with the QT interval. Using 21 SNP's, selected using the HAPMAP phase 2 data, we can cover the associated region and 1LD block 5' (containing OLFM2B) and 2 LD blocks 3', including all of intron 1, exon 2, and intron 2. The total region covered is ~170 kb. In addition we will evaluate the associations between selected SNP's in four other genes that were associated with QT interval in the early stages of our genome wide association study. The genes that we are interested in studying in the SCD collaborative network at this time include:

1. CAPON carboxy-terminal PDZ ligand of nNOS which regulates the distribution of nNOS (21 SNPs)
2. CACNA2D1 calcium channel alpha 2/delta subunit 1(5 SNPs)
3. Unknown gene on chromosome 5 (QTc_5.3) (one SNP)
4. KCNK1 TWIK-1 potassium channel (one SNP)
5. FGFR2 Fibroblast Growth Factor Receptor 2 (4 SNPs)

By identifying common variants that influence QT interval and therefore ventricular repolarization, it is hypothesized that these same genetic variants are likely to directly influence risk for SCD, or at least identify genes likely to play a role in SCD. Identification of genomic determinants of SCD might be used in the future to target our most aggressive therapies towards those at greatest risk (i.e. implantable cardiac defibrillators, ICD) or to develop pharmacologic interventions targeting specific proteins. Survival from cardiac arrest is generally less than 10%, thus early identification of increased risk and effective intervention is essential. By studying large prospective cohorts like ARIC and CHS, we will be able to determine the attributable risk for SCD susceptibility genes as well as examine the role of gene-environment interactions of SCD

5. Main Hypothesis/Study Questions:
Specific aims: 1) Evaluate the association between specific genes/loci and sudden cardiac death in the SCD network (ARIC and CHS).
2) Evaluate the association between specific genes/loci and QT interval at baseline in the SCD network (ARIC and CHS).

6. Data (variables, time window, source, inclusions/exclusions):
Sample size justification: All subjects from CHS (and ARIC) who consented to genetic analysis will be included in this study. It is important to include a large sample size in

order to minimize the likelihood of type 1 or type 2 errors. The positive findings from genetic association studies are less likely to be reproduced in other populations when they originate from studies with a small sample size. Since SCD is a complex phenotype, including the entire cohort will allow us the opportunity to compare the genotype frequency in the cases compared with a variety of control groups (see below). Including the entire cohort will allow greater power to examine rare variants or haplotypes and especially to analyze potential interactions and confounders. The large cohort size is also necessary for us to determine absolute risks within the population and will allow a powerful analysis of intermediate traits such as QT interval.

Methods: Population: All subjects who gave informed consent to be involved in genetic analyses will be included in this study. The primary analysis will compare the genotypes in the subjects with SCD to those who do not experience SCD during the follow-up period. However, often SCD occurs as a result of an acute coronary syndrome. We will compare the results of the primary analysis with that seen in secondary analyses comparing the genotypes of those with SCD to those with nonfatal acute coronary syndromes (MI and unstable angina) who do not experience SCD, and those with fatal CHD who do not experience SCD. Since many people with SCD have underlying atherosclerosis we will also compare the genotypes in those with SCD to those with high carotid intima media thickness who have not yet had an event.

Definition of SCD: All cases of fatal MI and fatal CHD (both inpatient and out of hospital) have been reviewed to determine if they meet criteria for SCD. Sudden cardiac death is defined as a sudden unexpected death that appears to be related to an arrhythmic etiology. Cases of SCD are being identified as definite, or possible, which includes cases complicated by other co-morbidities, such as ESRD, CHF or liver failure. For these cases, the patient must have been clinically stable prior to a sudden cardiac arrest . For all events, the individual must have been seen alive within 24 hours and had symptoms for less than one hour.

Genetic analyses: Currently, there is no commercially available cost effective and rapid method to genotype numerous SNPs in hundreds to thousands of samples. BioTrove has developed a novel platform for large-scale genotyping, relying upon a high-density, through-hole, nanotiter plate in conjunction with traditional TaqMan assays from ABI. This technology has the added benefit of reduced reagent usage and allows for up to 64 SNPs per sample (32 SNPs in duplicate) to be assayed in a single step.

Statistical analyses: The analyses will proceed as described below with the standard assumptions regarding genome distributions and other assumptions as outlined:

1. Hardy-Weinberg equilibrium among genotypes will be checked by calculating expected frequencies of genotypes and using the chi-square goodness-of-fit test.
2. All analyses will first be stratified by ethnicity to test for interaction. If no interaction is detected, pooled analyses will be performed.
3. For analyses of SCD, event times will be computed as the time to SCD. For participants without SCD, censoring times will be calculated according to the last date of follow-up or the date of death (non-SCD).
4. Genotype will be coded as 0 (zero copies of the candidate allele), 1 (one copy of the candidate allele), or 2(two copies of the candidate allele). An additive genetic model will be assumed unless indicated otherwise by results of the analysis or unless the allele

frequency of a given candidate variant is low, in which case, a dominant model combining the risk of heterozygotes and homozygotes will be used. Kaplan-Meier estimates of mortality will be computed, and log-rank tests will be used to compare curves among the genotypes. Cox proportional hazards models will be constructed to control for the effects of correlated SNPs and of other potential confounders. For analysis of QT intervals, mean QT intervals will be estimated and compared for the three genotypes at each locus using ANOVA. Multiple linear regression models will be constructed to account for effects of potential confounders.

5. We will use two popular and well-tested approaches for haplotype estimation. The first approach will be a Bayesian-based method by Stephens et al. (PHASE). [3] It will be used to estimate and assign individual specific probabilities of haplotypes for each individual. This method estimates the uncertainty associated with each phase call, and these probabilities will be used as weights in subsequent analyses. Prospective analysis of haplotypes will be performed in a similar manner as analysis of a single SNP except the independent variable will be haplotype instead of genotype. Second, we will also use the generalized linear model (GLM) method of Schaid et al (Haplo.glm), which employs an E-M algorithm to estimate phase while also estimating haplotype risk effect parameters in a GLM setting, allowing for adjustment of covariates and incorporation of interaction terms. [4] In this program, haplotype phase parameters and regression parameters are iteratively updated. While using PHASE-estimated haplotypes in survival analysis does not account for the uncertainties of the haplotype estimations, the GLM method has been suggested to be less accurate and does not allow for survival analysis. Because neither method is perfect for analysis of haplotypes, both will be used to ensure consistency of results across methods. For genes with putatively causal alleles, the haplotype containing those SNPs will be used as the reference group when testing for statistical significance; otherwise, an overall omnibus likelihood ratio test for all haplotypes having a null effect will be conducted.

6. We will further examine the interactions among candidate genes, SCD, and potential effect modifiers, including, coronary heart disease at baseline, heart rate, age, sex, BMI, and common carotid and maximum internal carotid wall thickness. We will examine the presence of detectable gene-gene and gene-environment interactions first by stratification and then by standard regression techniques. Interactions hypothesized in the literature will be tested, followed by interactions between genes in the same biologic pathways.

References:

1) Dekker JM, Crow RS, Hannan PJ, Schouten EG, Folsom AR. Heart rate-corrected QT interval prolongation predicts risk of coronary heart disease in black and white middle-aged men and women: the ARIC study. J Am Coll Cardiol 2004;43(4):565-71.

2) Okin PM, Devereux RB, Howard BV, Fabsitz RR, Lee ET, Welty TK. Assessment of QT interval and QT dispersion for prediction of all-cause and cardiovascular mortality in American Indians: The Strong Heart Study. Circulation 2000;101(1):61-6.

3) Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 2003;73:1162-9.

4) Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet.* 2002;70:425-434.

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? __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)? The ancillary study was already approved. There were nominations from the steering committee for additional investigators (Eric Boerwinkle).

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 (list number* __2005.07___)

___ B. primiarly 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.