1.a. Full Title: Meta-analysis of SNP interactions with age, hypertension, sex, and body mass index in relation to atrial fibrillation in CHARGE

b. Abbreviated Title (Length 26 characters): Gene-environment interactions for AF in CHARGE

2. Writing Group:
ARIC: Dan Arking, Alvaro Alonso, others welcome
Investigators from AGES, CHS, FHS and Rotterdam study

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

First author: Alvaro Alonso
Address: Division of Epidemiology & Community Health
Univ. of Minnesota
1200 S 2nd St, Suite 300
Minneapolis, MN 55454
Phone: 612 626 8597  Fax: 612 624 0315
E-mail: alonso@umn.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Dan E. Arking
Address: Johns Hopkins University School of Medicine
733 N. Broadway, BRB 453
Baltimore, MD 21205
Phone: 410-502-4867  Fax: 410-502-7544
E-mail: arking@jhmi.edu

3. Timeline:
ARIC data analysis: 1 month
Meta-analysis: 2 months
Manuscript preparation: 3 months

4. Rationale:
Atrial fibrillation (AF) is increasing in both incidence and prevalence, and is associated with substantial morbidity. Many risk factors for AF have been identified, including increasing age, male sex, hypertension, and obesity. Recently, associations between AF and common nucleotide variants on chromosome 4q25, and in \textit{KNCE4}, were identified in a genome-wide association study and candidate gene case-control study, respectively. In both studies, each single nucleotide polymorphism (SNPs) tested for association was assumed to have a constant effect on the log-odds of AF across values for each clinical covariate, and vice versa.

However, in the original genome-wide association study demonstrating an association between Chr 4q25 variants and AF, stratification of subjects by age suggested that the risk alleles for SNPs rs2200733 and rs10033464 were associated with greater risks of AF among younger subjects. Moreover, in a recent large-scale replication study and meta-analysis confirming the association between variation at chromosome 4q25 and AF, an exploratory analysis suggested that the effect of the risk allele for SNP rs2200733 was greater among individuals <60 years of age than for those >60 years of age in some but not all studied cohorts.

These findings suggest the possibility of interactions between genetic variants and clinical variables that may help to further refine existing associations. We therefore seek to test for interactions between SNPs and clinical covariates associated with AF on a genome-wide basis in several large cohorts of European ancestry.

5. Main Hypothesis/Study Questions:
We plan to conduct a meta-analysis of Age*SNP, Sex*SNP, Hypertension*SNP, and BMI*SNP interactions in relation to atrial fibrillation in adults of European ancestry in the CHARGE consortium.

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).

\textbf{Subjects:}
European/European-American subjects with GWAS data with or without atrial fibrillation.

\textbf{Exposures:}
1) Genotyped and imputed SNPs
2) Age
   \begin{itemize}
   \item Analysis of prevalent AF: at time of AF dx (case) / DNA draw (control)
   \end{itemize}
If age at time of AF diagnosis is unknown (CHS, Rotterdam study (RS)), age at time of DNA draw will be substituted
  o Incident: at baseline exam
3) Sex
4) Hypertension status
  o Prevalent: at time of AF dx (case) / DNA draw (control)
  o Incident: at baseline exam
5) Body mass index
  o Prevalent: at time of AF dx (case) / DNA draw (control)
  o Incident: at baseline exam

Outcomes:
1) Prevalent AF: subjects not defined as having incident AF
2) Incident AF: AF was not present at the baseline evaluation but was subsequently documented

ARIC will only provide incident AF cases since prevalence of AF at baseline was very low. Ascertainment of atrial fibrillation in ARIC has been detailed elsewhere. Briefly, we will use information from ECGs done in follow-up exams, hospital discharge codes (ICD-9 427.31 or 427.32) and death certificates (I48).

Primary statistical approach:
1. Both prevalent and incident analyses will be adjusted for age, body mass index, and hypertension status (at the time of AF diagnosis or DNA collection [prevalent] or at the baseline exam [incident]); sex; and if relevant site (ARIC, CHS) or cohort (FHS). Framingham will use robust variance methods to account for correlated family data.
2. The results of incident and prevalent AF associations will be analyzed separately. An analysis of pooled incident and prevalent AF will be conducted using logistic regression.
3. Population structure related to outcome will be assessed in each study and adjusted for as necessary using genome-wide principal components or similar methods.
4. Cohorts will be analyzed separately, and the results will be meta-analyzed using an inverse variance weighting method. Adjustment for the genomic control inflation factor within each study will be applied prior to meta-analysis.
5. Prevalent and incident analyses will be performed using logistic regression and Cox proportional hazards analyses, respectively.
6. Secondary analyses may adjust for important AF risk factors including diabetes, body mass index, heart failure, or myocardial infarction.

Interaction analyses:
1. Age*SNP analysis
   a. Prevalent analysis. Logistic regression will be used to model the log-odds of AF as a function of age at the time of AF diagnosis for cases (or DNA blood draw if age of AF onset not available) and age at the
time of DNA blood draw for controls, the additive effects of a SNP, and the interaction between these terms (age*SNP).

b. **Incident analysis.** Cox proportional hazards regression will be used to model the time to onset of AF/censoring as a function of age at the baseline exam, the additive effects of a SNP, and the interaction between these terms (age*SNP). Person time will be counted beginning at the time of the baseline examination (or DNA blood draw if not available).

2. **Sex*SNP**
   a. Models will include sex, the log-additive effects of each SNP, and a Sex*SNP interaction term.

3. **HTN*SNP**
   a. Models will include HTN, the log-additive effects of each SNP, and a HTN*SNP interaction term.

4. **BMI*SNP**
   a. Models will include BMI, the log-additive effects of each SNP, and a BMI*SNP interaction term.

**Statistical significance:**
A prespecified significance threshold of 5x10^-8 will be chosen to detect significant associations.

**Cohorts Included in Analysis:**
We will include the following cohorts: AGES, ARIC, CHS, FHS, RS, MONICA/KORA

**Anticipated limitations**
1. The sample is of European descent, and middle-aged to elderly; the generalizability to other races/ethnicities and younger individuals is unknown.
2. The study design is observational.
3. Significant SNPs may not be causal, rather in LD with causal SNPs.
4. Because the analyses are secondary, and involve examining for effect modification, our power to detect interactions is limited given the multiple testing penalty inherent in analyzing GWAS data.
5. In addition, the multiple testing of interactions will further inflate our false positives; we will need to replicate any findings in external cohorts.

7a. 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 ICTDER03 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 ICTDER03 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 #1396 (Arking) CHARGE GWAS for atrial fibrillation
    MS #1397 (Arking) CHARGE GWAS for lone atrial fibrillation
    These two manuscripts report results of genome-wide analysis of genetic
    markers for atrial fibrillation, but they do not evaluate gene-environment
    interactions. Dan Arking, first author in both proposals, is also a co-investigator in
    the current proposal. He will be in charge of performing analysis.

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
    ___   A. primarily the result of an ancillary study (list number*  ___________)
    ___   B. primarily based on ARIC data with ancillary data playing a minor
            role (usually control variables; list number(s)* 2008.09, 2006.03, 2007.02)

    *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

    adults: national implications for rhythm management and stroke prevention: the


