ARIC Manuscript Proposal #2078

PC Reviewed: 2/12/11	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Genome-wide Association Study of Particulate Matter and Ventricular Ectopy

b. Abbreviated Title (Length 26 characters): GWAS of PM & Ventricular Ectopy

2. Writing Group:

Writing group members: Candidates currently include co-investigators who have been involved in the planning, execution of, or assembly of data for ARIC AS#2009.08 and its WHI clinical trial sister study AS#264: Whitsel EA, Avery CL, Li Y, Yan S, Liao D, Lin D, North KE, Smith RL, Tinker L, Vernon M, Wilhelmsen K, Wu M, Kabisa S, and Zhang Z

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

First author: Melanie D. Napier
Address: University of North Carolina – Chapel Hill
Department of Epidemiology
Cardiovascular Disease Program
Bank of America Center, Suite 306-E
137 East Franklin Street
Chapel Hill, NC 27514
Phone: (919) 966-3168
Fax: (919) 966-9800
E-mail: mdnapier@email.unc.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).
Name: Eric A. Whitsel
Address: University of North Carolina – Chapel Hill
Departments of Epidemiology and Medicine
Cardiovascular Disease Program
Bank of America Center, Suite 305-C
137 East Franklin Street
Chapel Hill, NC 27514
Phone: (919) 966-3168
Fax: (919) 966-9800
E-mail: eric_whitsel@unc.edu

3. Timeline:
First pass analyses will begin as soon as requisite electrocardiographic, environmental, genetic and covariate data have been assembled. Anticipated completion is six months once data are available.

4. Rationale:

Ventricular premature beats (VPBs)—aka premature ventricular contractions—are early-occurring depolarizations of the right or left ventricle. This form of ventricular ectopy is manifest on resting, standard, twelve-lead electrocardiograms as widened, morphologically bizarre QRS complexes not preceded by P waves. Although ventricular tachycardia, the succession of three or more such complexes at a rate of at least 100 beats/min, and ventricular fibrillation, the total absence of properly formed QRS complexes and P waves, are ventricular arrhythmias associated with ominous prognoses, VPBs are much more common.[1] Despite being common, VPBs are also associated with risk of incident cardiac events, such as acute myocardial infarction and sudden cardiac death,[2] and have not been studied by published GWAS of related phenotypes.[3] A genetic basis for ectopy in mice has been suggested,[4, 5] yet no study has characterized the genetic basis of ventricular ectopy in human populations.

Even less is known about how genes modify associations between ventricular ectopy and environmental factors like particulate matter (PM). For example, we previously found that exposure to ambient PM air pollution is directly associated with the odds of ventricular ectopy among 57,422 Women's Health Initiative clinical trial (WHI CT) participants who were examined between 1999 and 2002, but who were not using anti-arrhythmic medication at the time.[6] However, the role of genetic factors in susceptibility to PM-mediated ectopy has not been examined to date. This is partly because estimating interactive effects between genotype and environment requires strong phenotypic and environmental data, and a much larger sample than that available in the typical candidate gene studies.[7]

The proposed manuscript directly addresses the paucity of information on the genetic basis of ventricular ectopy and the complex interplay of SNPs and PM on ventricular ectopy. The well-characterized ARIC and WHI CT cohorts are ideal for GWA analyses designed to leverage major, existing, and funded resources. Genomic, environmental and electrocardiographic data from these two cohorts can support a well-powered, rigorous examination of main SNP effects on ventricular ectopy and the degree to which SNPs modify PM-ventricular ectopy associations, as illustrated by the power calculations that follow in the supplemental material. Filling this gap in knowledge will (1) advance understanding of genetic susceptibility to and the pathophysiological mechanisms underlying PM-mediated arrhythmogenesis in ethnically and geographically diverse populations and (2) inform epidemiologists, environmental health scientists, and federal regulators responsible for evaluating air quality standards in terms of their ability to protect cardiovascular health.

5. Main Hypothesis/Study Questions:
We therefore propose a GWAS to (1) examine associations between genetic variation and ventricular ectopy, (2) explore gene-by-PM effects as they relate to ventricular ectopy

and (3) evaluate the consistency of main and interactive effects among and across race and gender groups.

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).

Overview. The proposed study will be conducted in the ARIC and WHI CT cohorts, based on the foundation provided by ARIC Ancillary Study #2009.08, "*Modification of PM-Mediated Arrhythmogenesis in Populations*" (R01- ES017794; Whitsel, PI) and two WHI Ancillary Studies: #140 "*The Environmental Epidemiology of Arrhythmogenesis in WHI*" (R01-ES012238; Whitsel, PI) and #264 "*Genetic Modification of PM-Mediated Arrhythmogenesis*" (R01-ES017794; Whitsel, PI).

Study population & Inclusion/Exclusion criteria. The focus will be on approximately 26,000 uniformly well-characterized and consenting participants living in the contiguous 48 U.S. states (U.S. Environmental Protection Agency Regions 1-10) within 500 miles of their exam site who had one or more high quality ECGs between 1987 and 1999 and consented to use of DNA for genetic research. The population will include seven distinct subpopulations: (1) black women, (2) Hispanic women, and (3) white women in the WHI CT; and (4) black women, (5) white women, (6) black men, and (7) white men in ARIC. These seven subpopulations will be used to independently identify gene main effects and gene-by-PM interactions for ventricular ectopy between approximately 2.5 million SNPs using imputed data from the Affymetrix 6.0 platform and daily mean ambient PM concentrations spatially interpolated at geocoded participant addresses. Participants with poor quality ECG, electronic pacemaker, or anti-arrhythmic medication use will be excluded.

Environmental exposure: Particulate matter. The proposed work will focus on location-specific daily mean concentrations of $PM_{2.5}$ and PM_{10} at geocoded addresses of ARIC participants between 1987 and 1999. We estimated the concentrations using a practical approach to spatial interpolation developed and validated by our group.[7]

Genetic exposure. SNPs were imputed using MACH (version 1.0.16) and the publicly available phased haplotypes from HapMap (release 22, build 36) served as a reference panel.

Phenotype: Ventricular Ectopy. Ventricular ectopy is defined as one or more ventricular ectopic beats during the ten-second ECG. Beats were detected by computer algorithms of the Minnesota Code (8.1.2, 8.1.3, or 8.1.5)[8] and visually over-read by physicians. In this population, few participants have more than one in their ten-second ECGs. Therefore, the phenotype will be analyzed as a binary variable: presence vs. absence of ectopy.

Model. To identify main and interactive effects, we propose a stratified, longitudinal analysis of ventricular ectopy. The initial strategy is to longitudinally model the average of repeated outcomes and thereby facilitate estimation of effects by increasing power using methods we have identified, tested and applied to large-scale genomic data in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Pharmacogenetics Working Group over the last two years under ARIC AS#2009.10. In this context, the methods will involve estimating genetic main effects and gene-by-PM interactions using conventional generalized estimation equations (GEE). Although other structures can be accommodated, an independence correlation structure will be used in this context to ensure consistency of the GEE estimates in the presence of time-varying covariates, and protect against potential bias related to the putative effects of past ectopy on future PM exposure.[9, 10] Pan and Wall's small-sample GEE extension[11] of Satterthwaite's method of approximating the degrees of freedom[12] associated with the *t* reference distribution will be implemented in R using the bossWithdf package. We assume separate analyses for the seven subpopulations that are then combined by meta-analysis.

More powerful tests of gene-by-PM interaction across ancestral populations will be based on an extension of kernel machine regression methods[13-15] that aggregate SNP-level score test statistics within genes. Such methods are particularly useful in genetically diverse populations where different SNPs may be in linkage disequilibrium with causal SNP(s). They also can accommodate complex SNP interactions, permit covariate adjustment, and do not penalize SNPs with opposing associations within a gene.

The analysis plan described above relies on identical electrocardiographic, genetic and environmental data in the ARIC study and WHI CT, i.e. the same ECG measures estimated by the same ECG reading center using the same methods; approximately 2.5 million SNPs imputed to the same HapMap reference panel; and daily mean PM concentrations estimated at all geocoded participant addresses using the same kriging methods. Although issues related to phenotype and genotype harmonization have been minimized by design, opportunities for replication of genes implicated in gene-by-PM interaction analyses outside the ARIC study and the WHI CT may be limited to those populations with comparable measures, e.g. the Multiethnic Study of Atherosclerosis.

Adjustments. Analyses will be stratified by study, race, and within the ARIC study, gender. All analyses will be appropriately adjusted for both ancestral admixture (using race-specific principal components in statistical models) and multiple comparisons. To control for potential influences of season, day of week, time of day, health, and weather in models examining genetic modification of the PM-ectopy association, we will also add temporal, sociodemographic, clinical and weather co-variables to models examining the SNP-PM interactions.

Power. We have adequate power to detect genetic main and interactive effects using the above methods. For SNP main effects, considerably lower ORs are detectable with the same power as for SNP-$PM_{2.5}$ and SNP-PM_{10} interactions. See supplemental material for power calculations.

Meta-analysis. We will examine heterogeneity among subpopulations as a function of

study, race, and gender. Although we expect modest power to detect heterogeneity, meta-analytic methods will be used to combine gene-level results across subpopulations, when appropriate. We also highlight our exploration of potentially more powerful gene-based tests of association, as facilitated by the KMR methods referenced above. These methods will allow us to harness the benefits of gene-based tests (e.g. fewer statistical tests and accommodation of a multiplicity of variants) while avoiding the potential for allelic heterogeneity and population-specific patterns of linkage disequilibrium possible when meta-analytically combining summary results across race/ethnicity.

Genome-wide Significance Level. 1 ÷ number of tests

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 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
(This file ICTDER 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 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)?

#1940 – Whitsel, *Modification of PM-Associated Heart Rate Variability in AS #2009.08.* The lead author of that proposal (Whitsel) is a co-author on this proposal. The focus of #1940 is on gene-PM interactive effects related to HRV, while this proposal focuses on genetic main effects and gene-PM interactive effects ventricular ectopy.

#1780 – Whitsel, *Genome-Wide Association Analyses of Respiratory Sinus Arrhythmia / Heart Rate Variability.* The lead author of that proposal (Whitsel) is a co-author on this proposal. The focus of #1780 is SNP main effects related to HRV, while this proposal focuses on genetic main effects and gene-PM interactive effects ventricular ectopy.

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* _#2009.08_)

___ 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/

12a. 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.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.

References

1. Rautaharju PM, Kooperberg C, Larson JC, and LaCroix A. Electrocardiographic abnormalities that predict coronary heart disease events and mortality in postmenopausal women: the Women's Health Initiative. Circulation. 2006, Jan 31;113(4):473-80.
2. Engel G, Cho S, Ghayoumi A, Yamazaki T, Chun S, Fearon WF, and Froelicher VF. Prognostic significance of PVCs and resting heart rate. Ann Noninvasive Electrocardiol. 2007, Apr;12(2):121-9.
3. Bezzina CR, Pazoki R, Bardai A, Marsman RF, de Jong JS, Blom MT, et al. Genome-wide association study identifies a susceptibility locus at 21q21 for ventricular fibrillation in acute myocardial infarction. Nat Genet. 2010, Aug;42(8):688-91.
4. Gutstein DE, Danik SB, Lewitton S, France D, Liu F, Chen FL, et al. Focal gap junction uncoupling and spontaneous ventricular ectopy. Am J Physiol Heart Circ Physiol. 2005, Sep;289(3):H1091-8.
5. Fabritz L, Kirchhof P, Franz MR, Nuyens D, Rossenbacker T, Ottenhof A, et al. Effect of pacing and mexiletine on dispersion of repolarisation and arrhythmias in DeltaKPQ SCN5A (long QT3) mice. Cardiovasc Res. 2003, Mar 15;57(4):1085-93.

6. Liao D, Whitsel EA, Duan Y, Lin HM, Quibrera PM, Smith R, et al. Ambient particulate air pollution and ectopy--the environmental epidemiology of arrhythmogenesis in Women's Health Initiative Study, 1999-2004. J Toxicol Environ Health A. 2009;72(1):30-8.
7. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005, Apr;6(4):287-98.
8. Prineas R, Crow R, and Blackburn H. The Minnesota code manual of electrocardiographic findings . Littleton, MA: John Wright-PSG; 1982.
9. Pepe MS, and Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Communications in Statistics-Simulation and Computation. 1994;23(4):939-951.
10. Pan W, Louis TA, and Connett JE. A note on marginal linear regression with correlated response data. The American Statistician. 2000, Aug;54(3):191-195.
11. Pan W, and Wall MM. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. Stat Med. 2002, May 30;21(10):1429-41.
12. Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics. 1946, Dec;2(6):110-4.
13. Lee S, Wu MC, and Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics. 2012;13(4):762-775.
14. Wu MC, Lee S, Cai T, Li Y, Boehnke M, and Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011, Jul 15;89(1):82-93.
15. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, and Lin X. Powerful SNP-set analysis for case-control genome-wide association studies. The American Journal of Human Genetics. 2010;86(6):929-942.
16. Liang K-Y, and Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika. 1986, Apr;73(1):13.
17. Eberle MA, Ng PC, Kuhn K, Zhou L, Peiffer DA, Galver L, et al. Power to detect risk alleles using genome-wide tag SNP panels. PLoS Genet. 2007, Oct;3(10):1827-37.
18. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, and Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009, Jun 9;106(23):9362-7.
19. National Institutes of Health. National Human Genome Research Institute. Office of Population Genomics. A Catalog of Genome-Wide Association Studies [Internet]. Available from: http://www.genome.gov/26525384.

Supplemental Material

Statistical Power

Overview. For power calculations, we assumed a binary response D_{ij} (0 or 1) and a logistic regression model, $\log(P(D_{ij}=1)/P(D_{ij}=0)) = \beta_0 + \beta_g G_i + \beta_e E_{ij} + \beta_{ge} G_i E_{ij}$. In this case $\exp(\beta_0)/(1+ \exp(\beta_0))$ is the overall prevalence, while β_g, β_e, β_{ge} are the natural logarithms

of the odds ratios associated with the gene effect, the environment effect and the gene-environment interaction, respectively. Power is (at least asymptotically) independent of β_g and β_e. The design is assumed to be an unmatched case-control study, in which a fixed number of cases (participants i for whom at least one j has $D_{ij}=1$) and controls (all $D_{ij}=0$) are generated and analyzed by logistic regression. We assume that repeated measures within an individual have common correlation 0.2, and power is calculated by using the asymptotic distribution of the estimator given by Liang and Zeger.[16]

Based on preliminary data, we will have approximately 1860 cases with ventricular ectopy; various values of the true odds ratio (OR) associated with a 10 $\mu g/m^3$ rise in PM_{10} or $PM_{2.5}$; standard deviation of PM_{10} or $PM_{2.5}$ equal to 10 or 7 $\mu g/m^3$ respectively; minor allele frequency (MAF) of 0.05, 0.15, 0.25; and a type I error probability of $0.05/10^{-6}$ (set to a very low value to avoid problems of multiple testing).

We assume separate analyses for the seven subpopulations that are then combined by meta-analysis. For power calculations, we make the simplifying assumption that there is no inter-subgroup variance in the true parameters of interest. We make separate calculations corresponding to the main effect, the interaction of SNP and PM_{10}, and the interaction for SNP and $PM_{2.5}$, the latter calculation based only on data available beginning in 1999. We also ignore exam site because it is not expected to markedly affect power, but we intend to include in the actual analyses.

Power Calculations. Results are shown in **Figure 1**. In each case we have shown three power curves (solid lines, colored black for MAF=0.05, red for MAF=0.15, blue for MAF=0.25) for our proposed analysis combining the seven subpopulations.
If P is a lower bound on the power for detecting a single risk allele, then the probability that at least one of K risk alleles produces a statistically significant result, assuming independence of alleles, is at least $1-(1-P)^K$ (see e.g. Reference 17). Using this transformation, the power curves for the SNP-$PM_{2.5}$ interactions are redrawn in **Figure 2** at three values representing a plausible range of K: 1, 5, and 10. At K=10 and K=5, the resulting power curves are shifted progressively to the left of those at K=1, illustrating that comparable power is achieved at substantially lower odds ratios (see **Table 1).** The detectable ORs are comparable to those of GWAS-identified SNPs catalogued by the NHGRI: median OR = 1.33 (interquartile range, 1.20-1.61), median MAF = 36% (21%-53%).[18, 19] For SNP main effects, considerably lower ORs are detectable with the same power.

Figure 1. Power curves for tests of PM-SNP interactions & SNP main effects as they relate to the study of ventricular ectopy & RMSSD

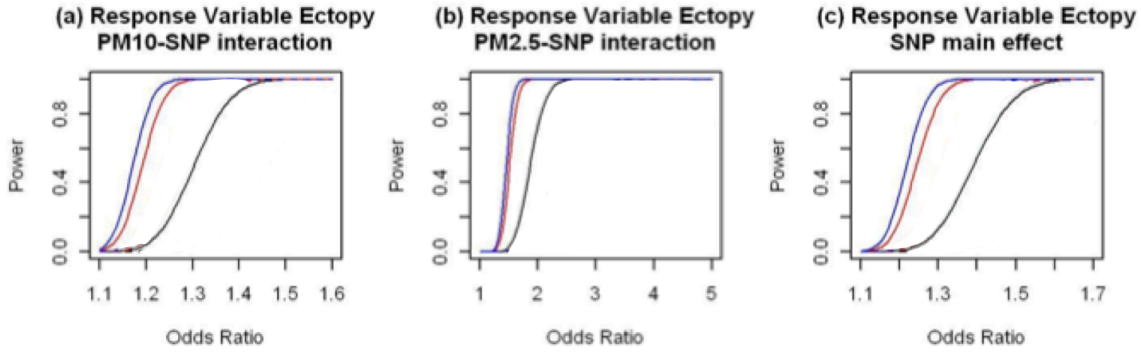

Figure 2. Power curves for tests of $PM_{2.5}$-SNP interactions at K = 1, 5 & 10

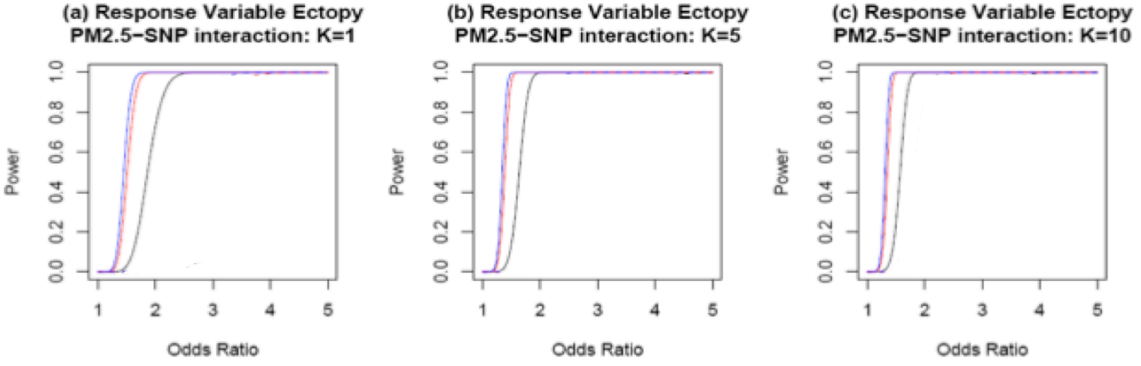

Table 1. Minimum odds ratios detected with power ≥ 0.8

Effect of Interest	K	MAF=0.05	MAF=0.15	MAF=0.25
SNP-PM_{10}	1	1.4	1.2	1.2
Interaction	5	1.3	1.2	1.2
	10	1.2	1.2	1.1
SNP-$PM_{2.5}$	1	2.1	1.6	1.5
Interaction	5	1.7	1.4	1.4
	10	1.7	1.4	1.4

MAF = minor allele frequency. K = number of independent risk alleles. PM10 and PM2.5 = concentrations particulate matter < 10 μm and 2.5 μm in diameter. SNP = single nucleotide polymorphism.