

SCIENTIFIC REPORTS



OPEN

Genome-wide association study and meta-analysis identify loci associated with ventricular and supraventricular ectopy

Melanie D. Napier¹ , Nora Franceschini¹, Rahul Gondalia¹, James D. Stewart^{1,2}, Raúl Méndez-Giráldez¹ , Colleen M. Sitlani³, Amanda A. Seyerle⁴, Heather M. Highland¹, Yun Li^{5,6,8}, Kirk C. Wilhelmsen^{5,7}, Song Yan^{5,6}, Qing Duan⁵, Jeffrey Roach⁹, Jie Yao^{10,11}, Xiuqing Guo^{10,11}, Kent D. Taylor^{10,11}, Susan R. Heckbert¹², Jerome I. Rotter^{10,13}, Kari E. North^{2,14}, Alexander P. Reiner^{15,16}, Zhu-Ming Zhang¹⁷, Lesley F. Tinker¹⁵, Duanping Liao¹⁸, Cathy C. Laurie¹⁹, Stephanie M. Gogarten¹⁹ , Henry J. Lin^{10,20}, Jennifer A. Brody³, Traci M. Bartz²¹, Bruce M. Psaty^{22,23}, Nona Sotoodehnia^{3,12}, Elsayed Z. Soliman²⁴, Christy L. Avery^{1,2} & Eric A. Whitsel^{1,25} 

The genetic basis of supraventricular and ventricular ectopy (SVE, VE) remains largely uncharacterized, despite established genetic mechanisms of arrhythmogenesis. To identify novel genetic variants associated with SVE/VE in ancestrally diverse human populations, we conducted a genome-wide association study of electrocardiographically identified SVE and VE in five cohorts including approximately 43,000 participants of African, European and Hispanic/Latino ancestry. In thirteen ancestry-stratified subgroups, we tested multivariable-adjusted associations of SVE and VE with single nucleotide polymorphism (SNP) dosage. We combined subgroup-specific association estimates in inverse variance-weighted, fixed-effects and Bayesian meta-analyses. We also combined fixed-effects

¹Department of Epidemiology, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, USA.

²Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA. ³Division of Cardiology, Department of Medicine, University of Washington, Seattle, WA, USA. ⁴Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA. ⁵Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. ⁶Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. ⁷Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC, USA. ⁸Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA. ⁹Research Computing Center, University of North Carolina, Chapel Hill, NC, USA. ¹⁰Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA. ¹¹Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California, USA. ¹²Cardiovascular Health Research Unit and the Department of Epidemiology, University of Washington, Seattle, WA, USA. ¹³Division of Genomic Outcomes, Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center, Torrance, California, USA. ¹⁴Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC, USA. ¹⁵Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ¹⁶Department of Epidemiology, University of Washington, Seattle, WA, USA. ¹⁷Epidemiological Cardiology Research Center, Department of Epidemiology and Prevention, Wake Forest University, Winston-Salem, NC, USA. ¹⁸Department of Public Health Sciences, Penn State University College of Medicine, Hershey, PA, USA. ¹⁹Department of Biostatistics, University of Washington, Seattle, WA, USA. ²⁰Division of Medical Genetics, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California, USA. ²¹Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, USA. ²²Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA. ²³Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. ²⁴Department of Epidemiology, School of Medicine, Wake Forest University, Winston-Salem, NC, USA. ²⁵Department of Medicine, University of North Carolina, Chapel Hill, NC, USA. Correspondence and requests for materials should be addressed to M.D.N. (email: mdnapier@email.unc.edu) or E.A.W. (email: eric_whitsel@med.unc.edu)

meta-analytic *t*-test statistics for SVE and VE in multi-trait SNP association analyses. No loci reached genome-wide significance in trans-ethnic meta-analyses. However, we found genome-wide significant SNPs intronic to an apoptosis-enhancing gene previously associated with QRS interval duration (*FAF1*; lead SNP rs7545860; effect allele frequency = 0.02; $P = 2.0 \times 10^{-8}$) in multi-trait analysis among European ancestry participants and near a locus encoding calcium-dependent glycoproteins (*DSC3*; lead SNP rs8086068; effect allele frequency = 0.17) in meta-analysis of SVE ($P = 4.0 \times 10^{-8}$) and multi-trait analysis ($P = 2.9 \times 10^{-9}$) among African ancestry participants. The novel findings suggest several mechanisms by which genetic variation may predispose to ectopy in humans and highlight the potential value of leveraging pleiotropy in future studies of ectopy-related phenotypes.

Supraventricular and ventricular ectopy (SVE, VE) are extra, abnormal depolarizations at non-sinus atrial, atrioventricular, or ventricular foci. The electrocardiographic hallmarks of SVE include absent or morphologically distinct P waves or PR intervals of different duration^{1,2}, while those of VE include widened, morphologically bizarre single or multiple QRS complexes not preceded by P waves^{1,2}.

As defined above, SVE and VE are common, but often occur as intermittent, asymptomatic, and/or clinically isolated events that increase in frequency with age^{3,4} depending on the method and duration of observation. On resting, supine, ten-second, standard twelve-lead electrocardiograms (ECGs), the prevalence of isolated SVE (<1%)⁵ and VE (~1%)⁶ is low, but higher in those with diseases of the heart³, lung⁷, brain⁸, kidney⁹, and/or exposure to medications used to treat them⁶. Moreover, SVE has been associated with ischemic heart disease mortality in persons free of such diseases¹⁰ and can trigger e.g. atrial fibrillation¹¹. VE also is associated with ventricular fibrillation and sudden cardiac death^{12,13}. Precipitants of SVE and VE are therefore of great clinical and public health interest.

Although behavioral and environmental precipitants (stress, tobacco, alcohol, caffeine, air pollution, exercise)^{6,14–16} have been studied, genetic predisposition to and heritability of SVE and VE in humans have not. Genome-wide association studies (GWAS) have nonetheless illuminated a shared genetic architecture at e.g. the *SCN5QA/10A* locus and distinct pathophysiological mechanisms underlying QT, PR, and QRS durations and atrial fibrillation¹⁷. Similarly, multi-trait analyses have identified novel, mechanistically important loci previously undetected by single-trait analyses¹⁸. Given established genetic mechanisms of arrhythmogenesis¹⁹ and familial aggregation of other supraventricular²⁰ and ventricular²¹ arrhythmias, we performed the first GWAS examining the hypothesis that genetic variation on a genome-wide scale influences electrocardiographic manifestation of SVE and VE in a large study of diverse ancestries.

Methods

Cohorts and participants. We studied SVE and VE among 42,976 and 44,131 participants who provided written, informed consent to use their genetic data. The participants were of European, African, or Hispanic/Latino ancestry and originated in five large, prospective cohorts: the Atherosclerosis Risk in Communities (ARIC) study²², the Women's Health Initiative Clinical Trial (WHI)²³, the Multi-Ethnic Study of Atherosclerosis (MESA)²⁴, the Cardiovascular Health Study (CHS)²⁵, and the Hispanic Community Health Study/Study of Latinos (HCHS/SOL)²⁶. We excluded first-degree relatives, and participants with low quality ECGs, pacemakers, or anti-arrhythmic medication use at each visit. Institutional Review Boards (IRB) at participating institutions approved all cohort-specific study protocols.

Briefly, ARIC is a prospective, longitudinal study of cardiovascular disease. Between 1987 and 1989, the study enrolled a probability sample of 15,792 men and women of African American and European descent aged 45–64 years²² from four US communities: Forsyth County, NC; Jackson, MS; suburban Minneapolis, MN; and Washington County, MD. Cohort members completed five visits: (1) 1987–1989; (2) 1990–1992; (3) 1993–1995; (4) 1996–1999; and (5) 2011–2013, at which ECGs were recorded.

WHI is randomized, controlled trial of hormone (estrogen +/- progestin) therapy, calcium/vitamin D supplementation, and dietary modification on the risk of breast and colorectal cancer, cardiovascular disease, and bone fracture^{23,27}. A total of 68,132 postmenopausal women aged 50–79 years old were enrolled from 40 clinical centers in the U.S. between 1993 and 1998. Eligible women were followed-up at years 1, 3, 6 and 9, during which ECGs were recorded. For the present study, we included European ancestry participants from three sub-cohorts: GARNET (Genomics and Randomized Trials Network), MOPMAP (Modification of Particulate-Matter-mediated Arrhythmogenesis in Populations), and WHIMS (Women's Health Initiative Memory Study). We included African and Hispanic/Latino ancestry participants from the SNP Health Association Resource (SHARe) project. Sub-cohorts are described in the Supplementary Methods.

MESA is a cohort study of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt disease²⁴. The 6,814 MESA participants were asymptomatic men and women aged 45–84 [38% European ancestry; 28% African ancestry; 22% Hispanic/Latino; and 12% Asian (mainly Chinese) ancestry]. During 2000–2002, participants were recruited from six US communities: (Forsyth County, NC; New York, NY; Baltimore, MD; St. Paul, MN; Chicago, IL, and Los Angeles County, CA). After an initial physical examination, there were four additional examination periods (17–20 months long). ECGs were recorded at exams 1 and 5 (2010–2011), but herein, we only use those from the former.

CHS is a cohort study of risk factors for coronary heart disease and stroke in adults ≥ 65 years conducted across four field centers: Sacramento County, CA; Washington County, MD; Forsyth County, NC; and Pittsburgh, PA²⁵. The original, predominantly European ancestry cohort of 5,201 persons was recruited in 1989–1990 from random samples of Medicare eligibility lists; subsequently, an additional predominantly-African-American

cohort of 687 persons were enrolled for a total sample of 5,888. ECGs were recorded at annual visits, but herein, we only used those from baseline.

HCHS/SOL is a study focused on describing the prevalence of risk and protective factors for chronic conditions, and to quantify all-cause mortality, fatal and non-fatal cardiovascular disease and pulmonary disease, and pulmonary disease exacerbation over time²⁶. From 2008–2011, 16,415 Hispanic/Latino individuals aged 18–74 were recruited from randomly selected households in four US communities: Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA using a stratified two-stage area probability sampling design. HCHS/SOL includes participants who self-identified as having Hispanic/Latino background, the largest groups being Central American (n = 1,730), Cuban (n = 2,348), Dominican (n = 1,460), Mexican (n = 6,471), Puerto-Rican (n = 2,728), and South American (n = 1,068). At the time of the present study, participants only had an ECG at the baseline visit.

Electrocardiography. Trained, certified technicians digitally recorded ECGs at visits 1–5 in ARIC²⁸; screening and annual visits 3, 6, and 9 in WHI²⁹; and, for this analysis, the baseline visit in MESA²⁴, CHS²⁵, and HCHS/SOL²⁶. Technicians used comparable procedures for preparing participants, placing electrodes, recording ECGs with MAC PC electrocardiographs (GE Marquette Electronics, Inc., Milwaukee, WI), and telephonically transmitting them to the Epidemiological Cardiology Research Center (Wake Forest School of Medicine, Winston Salem, NC) for inspection, identification of technical errors/inadequate quality, and analysis using the Marquette 12-SL program (2001 version, GE Marquette, Milwaukee, WI).

Identification of Supraventricular and Ventricular Ectopy. Since SVE and VE often occur intermittently and in isolation, presence of each phenotype on the ECG was determined independently at each visit. Supraventricular and ventricular ectopic beats were separately detected by computer algorithms based on the Minnesota Code (MC) and visually over-read by physicians (ARIC, WHI, MESA, HCHS/SOL). SVE was defined as ≥ 1 supraventricular ectopic beat (MC8.1.1, 8.1.3–8.1.5) and VE as ≥ 1 ventricular ectopic beat (MC8.1.2–8.1.3, 8.1.5) during the ten-second recording¹. Because few participants had ≥ 1 ectopic beat at a given visit, we analyzed SVE and VE as binary variables (0: absence, ≥ 1 : presence).

Genotyping, Quality Control, and Imputation. Each cohort or study performed genome-wide genotyping using Affymetrix or Illumina arrays and used similar quality control thresholds for excluding SNPs and samples (Supplementary Table S1). Genotypes were imputed using HapMap 2, HapMap 2 and 3, or 1000 Genomes Phase 1 (version 3, March 2012 release) reference panels. To enable cross-platform comparisons, Build 36 coordinates were converted to Build 37, and analyses were restricted to SNPs present in HapMap 2.

Statistical analysis. We stratified cohort participants by ancestry (and study) into thirteen subgroups of European (ARIC, CHS, MESA, WHI-GARNET, WHI-MOPMAP, WHI-WHIMS), African (ARIC, CHS, MESA, WHI-SHARe), and Hispanic/Latino (MESA, HCHS/SOL, WHI-SHARe) descent. For each of the thirteen ancestry-stratified subgroups, GWA analyses followed a standard protocol leveraging the availability of repeat ECGs, when available, to increase power. In cohorts with multiple ECGs per participant over time (ARIC, WHI), we estimated ectopy-SNP associations using generalized estimating equation (GEE) methods³⁰, a logit link, and an exchangeable working correlation structure to control for correlation of repeated measures (R geepack package). In studies with one ECG per participant (MESA, CHS), we estimated associations using logistic regression (SNPTEST, R geeglm package). Though multiple ECGs were available in MESA and CHS, only baseline visit data were used in accordance with analytic pipelines. In HCHS/SOL, we estimated associations among unrelated (at the 3rd degree level) participants (one per household) using a generalized linear model and a Firth test³¹ to account for small numbers of cases (R logistf package), assuming Census block group effects were negligible. We adjusted all models for age (year), sex (studies containing >1), season (quarter), study center (ARIC, CHS, MESA, HCHS/SOL) or geographic region (WHI), and ancestry principal components estimated using Eigenstrat³² (ARIC, CHS, MESA, WHI) or PC-AiR³³ (HCHS/SOL).

Within subgroups, we compared observed *P*-values for each SNP with expected values from a χ^2 distribution using quantile-quantile (Q-Q) plots and genomic inflation factors (λ). To eliminate statistical artifacts at low allele and ectopy frequencies, the comparisons excluded SNPs with an effective number of minor alleles present in exposed participants (defined as $2 \times$ number of exposed participants \times minor allele frequency \times imputation quality) <10 or a log odds ratio >10 . After filtering, thirteen and twelve subgroups contributed to the SVE and VE meta-analyses (MESA Hispanic/Latinos did not meet filtering thresholds due to infrequency of VE; n = 17).

Meta-analysis. We prioritized trans-ethnic analyses to maximize power and generalizability, given previous research suggesting that causal variants are typically relevant across populations³⁴, but also conducted ancestry-specific analyses given the potential for differences in linkage disequilibrium (LD) and allele frequency among populations. Analyses involved combining subgroup- and ancestry-specific summary results in 1) fixed-effects, inverse-variance-weighted meta-analyses with genomic control (METAL) and 2) trans-ethnic Bayesian meta-analysis (MANTRA)³⁵ to account for allelic heterogeneity among ancestry groups. MANTRA clustered similar populations according to allele frequencies, allowed for varying allele effects across populations, and produced Bayes' factors (BFs) for each ectopy-SNP association and its posterior probability of heterogeneity (P_{het}). We also performed multi-trait SNP association analyses that combined *t*-test statistics from fixed-effects meta-analyses of SVE and VE, using adaptive sum of powered score (aSPU) methods to investigate potential pleiotropy³⁶. While etiologies of SVE and VE may differ, combination was justified by extant knowledge of their shared precipitants^{14,15}, potential co-occurrence (MC 8.1.3 or 8.1.5)¹, and difficulty distinguishing them from each other³⁷. Multi-trait analyses provided *P*-values for genetic correlations among traits, but no effect estimates.

Supraventricular ectopy						Ventricular ectopy				
Ancestry	n	Female	Age, yr	Visits	SVE at any visit	n	Female	Age, yr	Visits	VE at any visit
		n (%)	mean (SD) [†]	mean (SD)	n (%)		n (%)	mean (SD) [†]	mean (SD)	n (%)
European	22,517					23,672				
ARIC	9,055	4,802 (53.0)	60.2 (8.3)	4.0 (1.0)	730 (8.1)	9,055	4,802 (53.0)	60.2 (8.2)	4.0 (1.0)	714 (7.9)
CHS	3,081	1,901 (61.7)	72.2 (5.3)	1.0 (0)	102 (3.3)	3,081	1,901 (61.7)	72.2 (5.3)	1.0 (0)	135 (4.4)
MESA	2,491	1,302 (52.3)	62.7 (10.2)	1.0 (0)	87 (3.5)	2,491	1,302 (52.3)	62.7 (10.2)	1.0 (0)	52 (2.1)
WHI-GARNET	2,091	2,091 (100)	66.5 (6.8)	2.6 (0.8)	160 (7.6)	1,944	1,944 (100)	69.2 (7.2)	2.6 (0.9)	41 (2.1)
WHI-MOPMAP	1,458 [‡]	1,458 (100)	64.1 (6.6)	2.8 (0.7)	103 (7.1)	2,957	2,957 (100)	68.5 (7.2)	3.0 (0.7)	1,475 (49.9)
WHI-WHIMS	4,341	4,341 (100)	70.8 (3.8)	2.5 (0.7)	436 (10.0)	4,144	4,144 (100)	73.6 (4.5)	2.5 (0.7)	51 (1.2)
African	9,274					9,274				
ARIC	2,381	1,495 (62.8)	58.7 (8.1)	3.6 (1.1)	224 (9.4)	2,381	1,495 (62.8)	58.7 (8.1)	3.6 (1.1)	179 (7.5)
CHS	767	487 (63.5)	72.8 (5.7)	1.0 (0)	55 (7.2)	767	487 (63.5)	72.8 (5.7)	1.0 (0)	58 (7.6)
MESA	1,600	859 (53.7)	62.2 (10.1)	1.0 (0)	72 (4.5)	1,600	859 (53.7)	62.2 (10.1)	1.0 (0)	33 (2.1)
WHI-SHARe	4,526	4,526 (100)	61.9 (6.8)	2.6 (0.8)	301 (6.6)	4,526	4,526 (100)	64.9 (7.2)	2.6 (0.8)	211 (4.7)
Hispanic/Latino	11,185					11,185				
HCHS/SOL	7,909	5,727 (27.4)	47.3 (13.3)	1.0 (0)	263 (2.1)	7,909	5,728 (27.4)	47.3 (13.2)	1.0 (0)	94 (0.7)
MESA	1,440	743 (51.6)	61.4 (10.2)	1.0 (0)	30 (2.1)	1,440 [§]	743 (51.6)	61.4 (10.2)	1.0 (0)	16 (1.1)
WHI-SHARe	1,836	1,836 (100)	60.6 (6.4)	2.6 (0.8)	61 (3.3)	1,836	1,836 (100)	63.7 (6.8)	2.6 (0.8)	58 (3.2)
Total	42,976	76%	66.3	2.2	2.1–10.0%	44,131	74%	67.7	2.2	0.7–49.9%

Table 1. Characteristics of the subgroups, by ancestry. ARIC = Atherosclerosis Risk in Communities study; CHS = Cardiovascular Health Study; GARNET = Genome-wide Association Research Network into Effects of Treatment; HCHS/SOL = Hispanic Community Health Study/Study of Latinos; MESA = Multi-Ethnic Study of Atherosclerosis; MOPMAP = Modification of PM-Mediated Arrhythmogenesis in Populations; n = number; SD = standard deviation; SHARe = SNP Health Association Resource; SVE = supraventricular ectopy; VE = ventricular ectopy; WHI = Women's Health Initiative; WHIMS = Women's Health Initiative Memory Study. [†]For studies with multiple visits, mean age is the age on date of ECG computed across all visits. [‡]Controls only in the trans-ethnic meta-analysis of SVE. [§]After filtering on expected heterozygosity, MESA Hispanics did not contribute to the trans-ethnic meta-analysis of VE.

By convention, we set genome-wide significance at $P < 5.0 \times 10^{-8}$ and suggestive significance at $P < 2.5 \times 10^{-6}$ for fixed-effects meta-analyses. For Bayesian meta-analyses, we used a $\log_{10}BF \geq 6.0$ as a genome-wide threshold for discovery (to approximate the performance of a $P < 5.0 \times 10^{-8}$)³⁸, a $P_{\text{het}} < 0.5$ as a liberal indicator of homogeneity among subgroups, and \geq two contributing racial/ethnic groups as a threshold for performing meta-analysis. Suggestive SNPs had $\log_{10}BF \geq 5.0$, $P_{\text{het}} < 0.5$, and \geq two contributing racial/ethnic groups. We report sub-threshold hits for trans-ethnic meta-analyses because they had the largest number of participants. We considered SNPs with ancestry-specific LD $r^2 < 0.2$ as independent. We summarized results from genomically controlled meta-analyses in Q-Q plots, Manhattan plots of the $-\log_{10}P$ value versus SNP position, and regional association plots. We functionally annotated lead and correlated SNPs ($r^2 \geq 0.8$) in relevant cardiac tissues using HaploReg v4.1³⁹ and visualized relevant tracks using the UCSC Genome Browser. We estimated heritability in European ancestry populations (ARIC, WHI-MOPMAP) using Genome-wide Complex Trait Analysis⁴⁰.

Data availability. Complete results are available on dbGAP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v5.p1. Primary data are available from the parent studies conditional on review and approval of requests by cohort-specific presentation and publication committees.

Results

Study characteristics. A total of 42,976 participants in thirteen subgroups contributed to the SVE analysis, of whom 22% were of African ancestry, 26% Hispanic/Latino, and 76% female (Table 1). On average, these participants were aged 66.3 years and contributed 2.2 visits (range:1–5), at which 2–10% of them had SVE at one or more visits. Estimated heritability (standard error(SE)) of SVE in ARIC was 3.2% (3.4%). A total of 44,131 participants in twelve subgroups contributed to the VE analysis, of whom 21% were of African ancestry, 25% Hispanic/Latino, and 74% female. On average, these participants were aged 67.7 years and also contributed 2.2 visits, during which 1–8% had VE at one or more visits, except in MOPMAP, which sampled VE cases and controls in equal proportions. Baseline prevalence of VE was $< 3\%$ in all subgroups, except in MOPMAP. Heritability of VE in ARIC and WHI-MOPMAP were 9.4% (3.4%) and 32% (14%). Lambdas from subgroup-specific Q-Q plots of SVE and VE ranged from 0.99 to 1.04 (Supplementary Figs S1 and S2).

Trans-ethnic meta-analyses. No SNP associations exceeded a genome-wide threshold for SVE or VE in trans-ethnic, fixed-effects meta-analyses; however, sub-threshold associations were identified for both

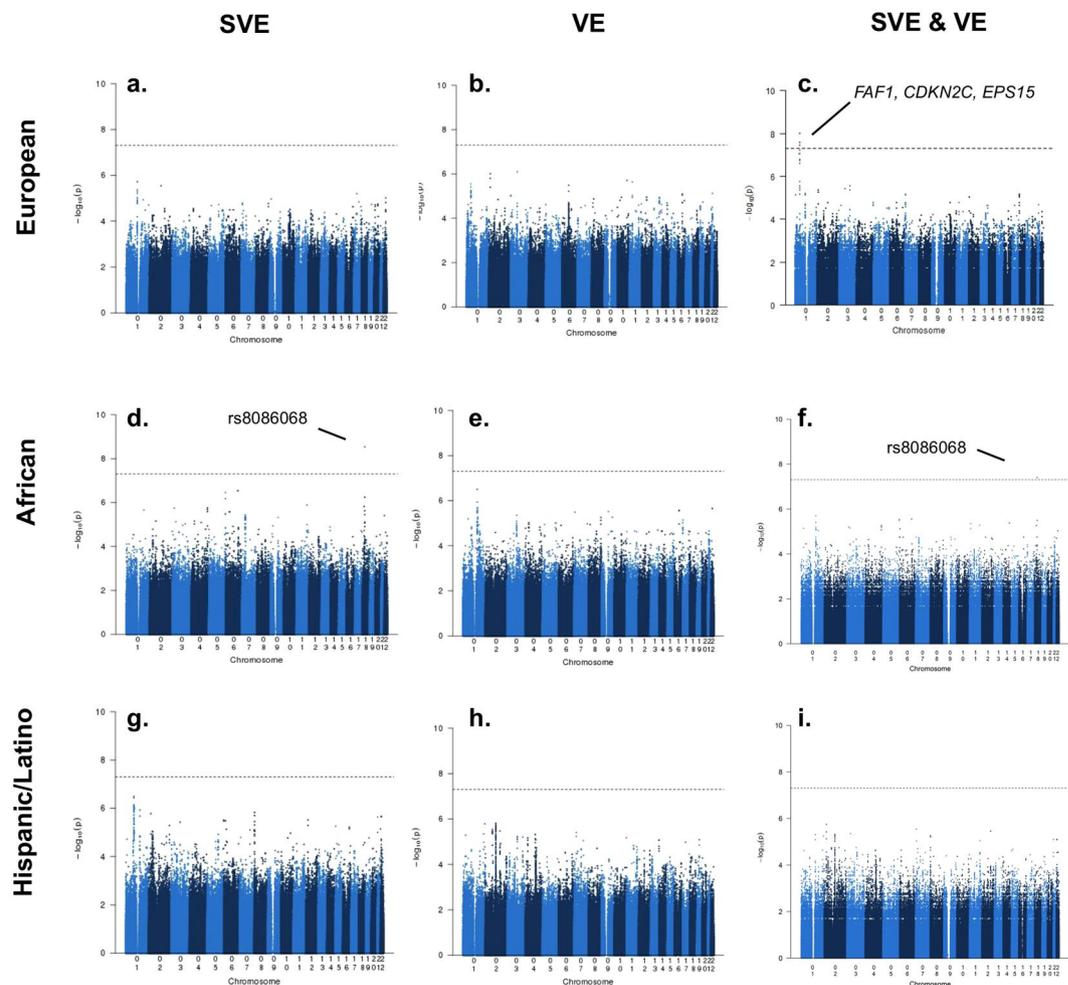


Figure 1. Manhattan plots of P -values from ancestry-specific fixed-effects meta-analyses of SVE and VE, and multi-trait analysis of SVE & VE. (Column 1, Panels a,d and g) Ancestry-specific fixed-effects meta-analyses of SVE; (Column 2, Panels b,e and h) Ancestry-specific fixed-effects meta-analyses of VE; (Column 3, Panels c,f and i) Multi-trait analysis of SVE & VE. Dotted horizontal line indicates the genome-wide significance threshold (5×10^{-8}).

Ancestry	SNP	Genomic Region (Build 37)	Gene	Alleles [†]	Effect Allele Frequency	P -value		
						SVE [§]	VE	SVE & VE
European	rs7545860 [‡]	C1:51341666	<i>FAF1,CDKN2C</i>	A/G	0.98	1.0×10^{-4}	2.8×10^{-6}	1.0×10^{-8}
African	rs8086068	C18:28363623	—	A/C	0.17	2.9×10^{-9}	8.3×10^{-1}	4.0×10^{-8}

Table 2. SNPs with genome-wide significant ($P < 5 \times 10^{-8}$) associations in ancestry-specific multi-trait analysis of SVE & VE. C = chromosome; SNP = single nucleotide polymorphism; SVE = supraventricular ectopy; VE = ventricular ectopy. [†]Effect/other. [‡]Associations with two SNPs in LD with rs7545860 (rs17106627[*EPS15*] and rs12022046[intergenic]) also were significant. [§]Direction in European ancestry = --???? (ARIC – WHI-WHIMS – WHI-MOPMAP controls – WHI-GARNET controls – CHS – MESA); Direction in African ancestry = ----(ARIC – WHI-SHARE – CHS – MESA). [¶]Direction in European ancestry = --???? (ARIC – WHI-MOPMAP – WHIMS – WHI-GARNET controls – MESA – CHS); Direction in African ancestry = + -?- (ARIC – WHI-SHARE – MESA – CHS).

phenotypes (Supplementary Table S2). Furthermore, Bayesian and multi-trait analyses (not shown) did not identify trans-ethnic loci (Supplementary Fig. S3).

Ancestry-specific meta-analyses: European. There were no genome-wide significant associations in fixed-effects meta-analyses of European ancestry studies of SVE or VE. However, multi-trait analysis identified a locus on chromosome 1 jointly associated with SVE and VE ($P = 2.0 \times 10^{-8}$; Panels A,B in Fig. 1). The lead SNP at this locus, rs7545860 (effect allele frequency [EAF] = 0.02), is an intron variant in *Fas Associated Factor 1 (FAF1)* and 92 kb 5' from *Cyclin-Dependent Kinase Inhibitor 2C (CDKN2C)* (Table 2; Panel C in Fig. 1; Supplementary

Fig. S4). Rs7545860, and correlated SNPs ($r^2 \geq 0.2$) including rs72692218 and rs66462949, reside in a genomic region including deoxyribonuclease (DNase I) hypersensitive sites, regulatory motifs, and putative enhancer/promoter histone signals in fetal heart, right atrium and ventricle, and/or aorta (Supplementary Fig. S5). This locus may also include the epidermal growth factor receptor pathway substrate 15 (*EPS15*) gene through SNPs in LD with rs7545860 (rs17106627 and rs12022046) (Supplementary Fig. S6). These SNPs are also in regions containing DNase I hypersensitivity sites, DNA methylation sites, putative enhancer/promoter histone marks, and regulatory motifs in cardiomyocytes and cardiac fibroblasts (Supplementary Fig. S7).

Ancestry-specific meta-analyses: African. Among African ancestry participants, fixed-effects meta-analysis of SVE identified a novel signal on chromosome 18 (lead SNP rs8086068; EAF = 0.17; $P = 2.87 \times 10^{-9}$), associated with a 75% increased odds of SVE per copy of the C allele (95% CI: 1.46–2.11) (Panel D in Fig. 1). This variant also was directionally consistent, if not significant, among European ancestry studies and one Hispanic/Latino ancestry subgroup (Supplementary Fig. S8). Multi-trait analyses identified this same lead SNP ($P = 4.0 \times 10^{-8}$), driven by its association with SVE (Table 2; Panel F in Fig. 1). While intergenic, rs8086068 is 206 kb 3' from the desmocolin 3 (*DSC3*) gene, one of a family of desmocolin genes clustered in the area, though the SNP is separated from the gene family by a recombination spike and may not interact with it (Supplementary Figs S8, S9). Functional annotation indicates that three SNPs in LD ($r^2 \geq 0.2$) with this lead SNP using the 1000G AFR referent population (rs2097047, rs17711533, rs17711559) occur within DNase I hypersensitivity sites in fetal heart tissue (Supplementary Fig. S10). No SNPs met the genome-wide threshold for significance among African ancestry studies in fixed-effects meta-analyses of VE (Panel E in Fig. 1).

Ancestry-specific meta-analyses: Hispanic/Latino. No SNPs met the genome-wide threshold among Hispanic/Latino ancestry studies in fixed-effects meta-analyses of SVE or VE, or in multi-trait meta-analyses (Panels G–I in Fig. 1). Complete results are available on dbGAP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v5.p1.

Discussion

This first GWAS of ectopy identified two biologically plausible loci among European and African ancestry individuals. It identified the *FAF1/CDKN2C/EPS15* locus (chromosome 1) in multi-trait meta-analyses of SVE and VE among European ancestry individuals. Earlier GWAS have associated this locus with QRS duration⁴¹. It also identified a second locus among African ancestry individuals, approximately 206 kb 3' from a desmocolin gene cluster that includes *DSC3* and *DSC2*, the latter previously associated with arrhythmogenic cardiomyopathy (ACM)⁴². Together, these findings provide insight into putative mechanisms underlying genetic susceptibility to ectopy.

Contrary to expectation, this GWAS of ectopy did not identify any loci meeting the genome-wide threshold for significance in trans-ethnic, fixed-effects or Bayesian meta-analyses of either phenotype. Restriction of analyses to HapMap 2 SNPs may be one reason why none were identified, given the admittedly limited genomic coverage of this reference panel, although restriction also enabled cross-platform comparisons. Heterogeneity of association among races/ethnicities due to differences in imputation quality or minor allele frequency may be another. Lastly, as large as our study is, an even larger study may be required to adequately power the identification of trans-ethnically important variants, as is further discussed, below.

The European ancestry locus identified by rs7545860 is intronic to *FAF1*, an apoptosis protein-encoding gene previously implicated in GWAS of QRS duration⁴³. Two SNPs in LD with rs7545860 (rs72692218; rs66462949) are intronic to the nearby gene *CDKN2C*, a cyclin-dependent kinase inhibitor dually implicated by that GWAS. The lead SNP is also in LD with rs17391905, a *FAF1* and *CDKN2C* SNP identified by Sotoodehnia *et al.* ($r^2 = 0.53$, multi-trait aSPU $P = 1.60 \times 10^{-7}$). Additional SNPs in LD with rs7545860 include intronic variants (rs17106627, rs12022046) of *EPS15*, a gene that encodes a calcium-binding protein involved in receptor-mediated endocytosis of epidermal growth factor, but has no previously established role in arrhythmogenesis. Functional annotation for these SNPs suggests potential involvement with histone modification and enhancer activity in fetal heart.

It is notable that the aforementioned European ancestry locus (*FAF1/CDKN2C/EPS15*) was only identified when using adaptive sum of powered score methods to investigate pleiotropy. This finding highlights the potential value of leveraging pleiotropic effects in future studies of ectopy-related phenotypes. Indeed, examining them may well improve understanding of biological mechanisms underlying correlated traits.

No GWAS has been published to date relating arrhythmia to genetic variation in desmocolin cluster genes, including *DSC3*. The desmocolin gene cluster is of interest because the desmocolins are calcium-dependent glycoproteins involved in cardiac intercellular connections and neighboring gene *DSC2* is associated with ACM, a congenital heart disorder characterized by right ventricular fibrofatty infiltration, myocardiocyte apoptosis, gap junction pathophysiology, supraventricular/ventricular arrhythmias, and sudden cardiac death⁴². Moreover, several SNPs in LD with lead SNP rs8086068 are located within DNase I hypersensitivity sites in fetal heart tissue, suggesting potential involvement in tissue-specific regulation. We also demonstrated that this variant was directionally consistent among European ancestry studies and one Hispanic/Latino ancestry subgroup, suggesting that differences in risk factors or allelic effects among races/ethnicities may explain the ancestral heterogeneity of effects, a possibility deserving further study. Several other loci that reached the threshold for suggestive significance in trans-ethnic meta-analyses also have biologically plausible relationships with ectopy (Supplementary Discussion).

In addition to the loci discussed above, this paper adds to the literature an estimate of heritability for SVE and VE in European ancestry populations. In lieu of available family-based data, we estimated heritability using two cohorts with the largest number of ectopy cases (ARIC; WHI-MOPMAP) and among European ancestry participants because of the difficulty obtaining minority reference populations. Our finding that the estimated heritability of VE, a binary phenotype, differed in ARIC (9.4%, SE = 3.4%) and WHI-MOPMAP (32%, SE = 14%) is partly attributable to the difference in VE prevalence (and design) between those populations^{44,45}. In our study,

WHI-MOPMAP sampled VE cases and controls in equal proportions (i.e. 50% of participants had ectopy), but among ARIC European ancestry participants, the prevalence of VE was only 7.9%. The SVE heritability estimate was likely influenced by the same factors. Although the generalizability of such estimates outside of ARIC and WHI-MOPMAP is unknown and the estimates are not directly comparable to those estimated from pedigree data^{44,45}, they remain notable findings from this study.

This work has several limitations that deserve consideration. Despite many participants in this meta-analysis, low prevalence of ectopy as measured by brief ECGs and limited genomic coverage of HapMap 2—especially in non-European ancestry populations—limited its overall power to identify trans-ethnic signals. By extension, our ability to detect ancestry-specific signals also was limited. Modest power is a well-known limitation of GWAS involving small populations, cross-sectional designs, infrequent outcomes, and brief ECG recordings. However, ectopy has not been examined in a multi-ethnic GWAS, so to examine it, we leveraged the following: (1) imputed genomic data from five cohorts including multiple ancillary studies and thirteen ancestry-stratified subgroups collectively representing >42,000 participants; (2) ECG data from up to five recordings per participant and eleven years of follow-up; (3) relatively powerful, longitudinal and meta-analytic methods that exploit ancestral heterogeneity³⁵; and (4) multi-trait SNP association methods that exploit phenotypic correlation³⁶. Leveraging multi-ethnic cohorts and these analytical methods powered the discovery and localization of ectopy-SNP associations, albeit based on ten-second ECG recordings.

We acknowledge that longer ECG recording durations are essential for detecting ectopy with sensitivity. Although the relative sensitivity of short ECG recordings for ectopy is low—even when repeated—paroxysmal arrhythmias frequent enough to be captured by insensitive, but highly specific, short recordings may have more prognostic significance than those so infrequent that they require long recordings to capture them¹⁵. Moreover, the bias of odds ratios reported here approaches zero because specificity of physician-verified ECGs for ectopy approaches 100%⁴⁶ while their sensitivity among participants with and without a given variant is identical⁴⁷. It is also possible that short ECG recordings capture frequent ectopy known to increase the risk of myocardial infarction, cardiac, and all-cause mortality *in addition* to infrequent ectopy associated with a relatively benign prognosis. The group to whom inferences can be made may therefore be heterogeneous. Finally, because independent replication was not feasible due to the current paucity of genotyped cohorts with physician-verified ectopy, we accordingly acknowledge that findings may be due to chance. These considerations underscore the need for further confirmation of our findings.

Conclusions

Given these limitations, we view the study findings as hypothesis-generating and have provided publicly accessible summary statistics from ancestry-specific fixed-effects meta-analyses on dbGAP to facilitate external replication. But under those hypotheses, we also provide evidence that variants in *FAF1/CDKN2C*, *EPS15*, *DSC2/3*, and *SCN5A* on chromosomes 1, 3, and 18 contribute to the genetic risk of supraventricular/ventricular ectopy and arrhythmogenesis in humans via plausible cellular, intercellular, and cationic mechanisms involving myocardial cyto apoptosis, desmosome-related gap junction abnormality, sodium channelopathy, and electrocardiographically manifest derangement of normal atrioventricular physiology.

References

- Prineas, R. J., Crow, R. S. & Zhang, Z.-M. *The minnesota code manual of electrocardiographic findings: Standards and procedures for measurement and classification* (Springer Verlag, New York, 2010).
- Rautaharju, P. M., Park, L. P., Chaitman, B. R., Rautaharju, F. & Zhang, Z.-M. The Novacode criteria for classification of ECG abnormalities and their clinically significant progression and regression. *J Electrocardiol* **31**, 157–187 (1998).
- Conen, D. *et al.* Premature atrial contractions in the general population: frequency and risk factors. *Circulation* **126**, 2302–2308 (2012).
- Simpson, R. J. *et al.* Prevalence of premature ventricular contractions in a population of African American and white men and women: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J* **143**, 535–540 (2002).
- Kusumoto, F. *ECG interpretation from pathophysiology to clinical application* (Springer, New York, 2009).
- Ng, G. A. Treating patients with ventricular ectopic beats. *Heart* **92**, 1707–1712 (2006).
- Kawano, Y., Tamura, A., Ono, K. & Kadota, J. Association between obstructive sleep apnea and premature supraventricular contractions. *J Cardiol* **63**, 69–72 (2014).
- Di Pasquale, G. *et al.* Holter detection of cardiac arrhythmias in intracranial subarachnoid hemorrhage. *Am J Cardiol* **59**, 596–600 (1987).
- Kimura, K., Tabei, K., Asano, Y. & Hosoda, S. Cardiac arrhythmias in hemodialysis patients. A study of incidence and contributory factors. *Nephron* **53**, 201–207 (1989).
- Qureshi, W., Shah, A. J., Salahuddin, T. & Soliman, E. Z. Long-Term Mortality Risk in Individuals With Atrial or Ventricular Premature Complexes (Results from the Third National Health and Nutrition Examination Survey). *Am J Cardiol* **114**, 59–64 (2014).
- Binici, Z., Intzilakis, T., Nielsen, O. W., Køber, L. & Sajadieh, A. Excessive supraventricular ectopic activity and increased risk of atrial fibrillation and stroke. *Circulation* **121**, 1904–1911 (2010).
- Ataklte, F., Erqou, S., Laukkanen, J. & Kaptoge, S. Meta-analysis of ventricular premature complexes and their relation to cardiac mortality in general populations. *Am J Cardiol* **112**, 1263–1270 (2013).
- Chen, A., Chen, X., Shen, Y. & Li, W. Modalities of ventricular pacing for cardiac resynchronization therapy in patients with heart failure: a meta-analysis and systematic review. *Arch Med Sci* **13**, 1006–1017 (2017).
- Sessler, C. N. & Cohen, M. D. Cardiac arrhythmias during theophylline toxicity. A prospective continuous electrocardiographic study. *CHEST Journal* **98**, 672–678 (1990).
- Liao, D. *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* **72**, 30–38 (2009).
- Porta, A. *et al.* Plasma B-type natriuretic peptide levels are poorly related to the occurrence of ischemia or ventricular arrhythmias during symptom-limited exercise in low-risk patients. *Arch Med Sci* **12**, 341–348 (2016).
- Sano, M. *et al.* Genome-wide association study of electrocardiographic parameters identifies a new association for PR interval and confirms previously reported associations. *Hum Mol Genet* **23**, 6668–6676 (2014).
- Kim, J., Zhang, Y. & Pan, W. & Alzheimer's Disease Neuroimaging Initiative Powerful and Adaptive Testing for Multi-trait and Multi-SNP Associations with GWAS and Sequencing Data. *Genetics* **203**, 715–731 (2016).
- George, A. L. Molecular and genetic basis of sudden cardiac death. *J Clin Invest* **123**, 75–83 (2013).
- Darbar, D. *et al.* Familial atrial fibrillation is a genetically heterogeneous disorder. *J Am Coll Cardiol* **41**, 2185–2192 (2003).

21. Arnett, D. K. *et al.* Genetic variation in NCAM1 contributes to left ventricular wall thickness in hypertensive families. *Circ Res* **108**, 279–283 (2011).
22. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687–702 (1989).
23. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* **19**, 61–109 (1998).
24. Bild, D. E. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *Am J Epidemiol* **156**, 871–881 (2002).
25. Fried, L. P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263–276 (1991).
26. Sorlie, P. D. *et al.* Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* **20**, 629–641 (2010).
27. Hays, J. *et al.* The Women's Health Initiative recruitment methods and results. *Ann Epidemiol* **13**, S18–S77 (2003).
28. ARIC Study Investigators <http://www2.csc.unc.edu/aric/cohort-manuals>.
29. WHI Study Group *Volume 2 (section 13): ECG procedures* (WHI Clinical Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA, 1994).
30. Zeger, S. L., Liang, K. Y. & Albert, P. S. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–1060 (1988).
31. Wang, X. Firth logistic regression for rare variant association tests. *Front Genet* **5**, 187 (2014).
32. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).
33. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol* **39**, 276–293 (2015).
34. Carlson, C. S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* **11**, e1001661 (2013).
35. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809–822 (2011).
36. Kim, J., Bai, Y. & Pan, W. An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics. *Genet Epidemiol* **39**, 651–663 (2015).
37. Drew, B. J., Scheinman, M. M. & Dracup, K. MCL1 and MCL6 compared to V1 and V6 in distinguishing aberrant supraventricular from ventricular ectopic beats. *Pacing Clin Electrophysiol* **14**, 1375–1383 (1991).
38. Stephens, M. & Balding, D. J. Bayesian statistical methods for genetic association studies. *Nature Rev Genet* **10**, 681–690 (2009).
39. Ward, L. D. & Kellis, M. HaploRegv4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877–D881 (2016).
40. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76–82 (2011).
41. van der Harst, P. *et al.* 52 Genetic Loci Influencing Myocardial Mass. *J Am Coll Cardiol* **68**, 1435–1448 (2016).
42. Marcus, F. I. *et al.* Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Circulation* **121**, 1533–1541 (2010).
43. Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet* **42**, 1068–1076 (2010).
44. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294–305 (2011).
45. Zaitlen, N. & Kraft, P. Heritability in the genome-wide association era. *Hum Genet* **131**, 1655–1664 (2012).
46. Meyer, M. L. *et al.* Repeatability of ectopic beats from 48-hr ambulatory electrocardiography: The Atherosclerosis Risk in Communities (ARIC) Study. *Ann Noninvasive Electrocardiol* **1–7** (2017).
47. Greenland, S. & Lash, T. L. In *Modern epidemiology* (eds Rothman, K. J., Greenland, S. & Lash, T. L.) 359 (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2008).

Acknowledgements

Study-specific acknowledgements and funding sources for participating studies are reported in the supplement.

Author Contributions

Conception and design: C.L.A., K.E.N., E.A.W. Data acquisition: C.L.A., S.R.H., D.L., Y.L., K.E.N., A.P.R., J.I.R., N.S., E.Z.S., K.D.T., L.F.T., E.A.W., K.C.W., Z.-M.Z. Analysis and interpretation: T.M.B., J.A.B., N.F., S.M.G., R.G., X.G., S.R.H., C.L., H.J.L., M.D.N., B.M.P., J.I.R., A.A.S., C.M.S., N.S., J.D.S., E.A.W., J.Y. Statistical methods and software development: Q.D., S.M.G., H.M.H., Y.L. R.M.-G., K.E.N., J.R., C.M.S., J.D.S., E.A.W., S.Y. Original draft preparation: N.F., M.D.N., E.A.W. Revision of critical content and final approval: All authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-23843-z>.

Competing Interests: Dr. Psaty serves on the Data and Safety Monitoring Board of a device clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. The remaining authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018