TITLE

Multi-ethnic genome-wide association study of decomposed cardioelectric phenotypes illustrates strategies to identify and characterize evidence of shared genetic effects for complex traits

AUTHORS AND AFFILIATIONS

Antoine R. Baldassari[1*], Colleen M. Sitlani[2], Heather M. Highland[1], Dan E. Arking[3], Steve Buyske[4], Dawood Darbar[5], Rahul Gondalia[1], Misa Graff[1], Xiuqing Guo[6,7], Susan R. Heckbert[8], Lucia A. Hindorff[9], Chani J. Hodonsky[1], Yii-Der Ida Chen[6,7], Robert C. Kaplan[10], Ulrike Peters[11], Wendy Post[12], Alex P. Reiner[11], Jerome I. Rotter[6,7], Ralph V. Shohet[13], Amanda A. Seyerle[14], Nona Sotoodehnia[8], Ran Tao[15], Kent D. Taylor[6,7], Genevieve L Wojcik[16], Jie Yao[6,7], Eimear E. Kenny[17,18,19,20], Henry J. Lin[6,7], Elsayed Z. Soliman[21], Eric A. Whitsel[1], Kari E. North[1,22], Charles Kooperberg[11], Christy L. Avery[1]

[1] Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill NC
[2] Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA
[3] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[4] Department of Statistics & Biostatistics, Rutgers University, New Brunswick, NJ
[5] Department of Medicine, University of Illinois at Chicago, Chicago, IL
[6] Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA
[7] Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA
[8] Cardiovascular Health Research Unit, Division of Cardiology, Department of Medicine, University of Washington, Seattle, WA
[9] Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
[10] Albert Einstein College of Medicine, Bronx, NY
[11] Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA
[12] Departments of Medicine and Epidemiology Johns Hopkins University, Baltimore, MD
[13] Center for Cardiovascular Research, John A. Burns School of Medicine, Honolulu, HI
[14] UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC
[15] Department of Biostatistics, Vanderbilt University, Nashville, TN
[16] Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA
[17] Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York City, NY
[18] Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY
[19] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York City, NY
[20] Department of Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY
[21] Epidemiological Cardiology Research Center (EPICARE), Wake Forest School of Medicine, Winston-Salem, NC
[22] Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill NC

*Corresponding author; Antoine Baldassari, 123 W. Franklin Street, Building C, Suite 4210, Chapel Hill, NC, 27514, USA, baldassa@email.unc.edu

ABSTRACT

Background: Published genome-wide association studies (GWAS) are mainly European-centric, examine a narrow view of phenotypic variation, and infrequently interrogate genetic effects shared across traits. We therefore examined the extent to which a multi-ethnic, combined trait GWAS of phenotypes that map to well-defined biology can enable detection and characterization of complex trait loci.

Methods: With 1000 Genomes Phase 3 imputed data in 34,668 participants (15% African American; 3% Chinese American; 51% European American; 30% Hispanic/Latino), we performed covariate-adjusted univariate GWAS of six contiguous electrocardiogram (ECG) traits that decomposed an average heartbeat and two commonly reported composite ECG traits that summed contiguous traits. Combined phenotype testing was performed using the adaptive sum of powered scores test (aSPU).

Results: We identified six novel and 87 known ECG trait loci (aSPU p-value < 5E-9). Lead SNP rs3211938 at novel locus CD36 was common in African Americans (minor allele frequency=10%) and near-monomorphic in European Americans, with effect sizes for the composite trait, QT interval, among the largest reported. Only one novel locus was detected for the composite traits, due to opposite directions of effects across contiguous traits that summed to near-zero. Combined phenotype testing did not detect novel loci unapparent by univariate testing. However, this approach aided locus characterization, particularly when loci harbored multiple independent signals that differed by trait.

Conclusions: Despite including one-third as few participants as the largest published GWAS of ECG traits, our study identifies multiple novel ECG genetic loci, emphasizing the importance of ancestral diversity and phenotype measurement in this era of ever-growing GWAS.

65 **AUTHOR SUMMARY**

66 We leveraged a multiethnic cohort with precise measures of cardioelectric function to identify novel
67 genetic loci affecting this complex, multifaceted phenotype. The success of our approach stresses the
68 importance of phenotypic precision and participant diversity for future locus discovery and
69 characterization efforts, and cautions against compromises made in genome-wide association studies to
70 pursue ever-growing sample sizes.

71

INTRODUCTION

Genetic susceptibility underlies a majority of common diseases and traits, shown by genome-wide association studies (GWAS) that have identified thousands of genetic loci for traits such as cardiovascular, cardiometabolic, cancer, kidney, psychiatric, ocular, inflammatory, and neuromuscular conditions(1). The many GWAS reports have revealed both common threads underlying the genetic architecture of complex diseases and traits, as well as research gaps. For example, evidence of shared genetic effects (i.e., pleiotropy) is widespread, even for traits with few known etiologic links(2,3). Yet few studies have systematically examined evidence of shared genetic effects, thereby missing opportunities to identify and characterize master regulators as strong candidates for intervention(2,4). There is also limited racial/ethnic diversity, inasmuch as the majority (>80%) of GWAS participants have been of European ancestry[3]. Limited diversity leads to a biased view of human variation that hinders translation of genetic associations into clinical and public health applications for all populations(5,6). Further, the scale and collaborative nature of current GWAS analyses prioritize traits that are widely available across studies, which may not precisely capture phenotypic variation and its underlying biology(7,8). Together, these research gaps argue for expanding GWAS analyses to systematically assess for shared genetic effects across a spectrum of biologically important traits in multi-ethnic populations.

Electrocardiograms (ECG) measure a sequence of distinct electrophysiologic processes in the myocardium that underlie cardiac conduction and repolarization. ECG traits have high heritability(9), are relevant to cardiovascular health,(10) and allow opportunities for dense phenotyping(11). Moreover, there are few racially/ethnically diverse GWAS reports on ECG traits(12). Therefore, ECG traits are well suited for assessing the degree to which increased racial/ethnic diversity, evaluation of genetic effects shared across phenotypes, and improved phenotype resolution can enhance locus identification and characterization. We therefore examined individual and shared genetic effects underlying six contiguous measures of the ECG waveform spanning an average heartbeat. Analyses used data from the multi-ethnic

Population Architecture using Genetic Epidemiology (PAGE) study and the Multi-Ethnic Study of Atherosclerosis (MESA). Our results illustrate the broad utility of multi-ethnic GWAS analyses of carefully constructed individual and aggregate traits to illuminate the biology of complex diseases and traits.

RESULTS

Sample description

Of the 39,538 participants with GWAS and ECG data in ARIC, HCHS/SOL, MESA, and WHI, 34,668 (88%) met all inclusion criteria (Tables S1, S2). Seventy-five percent of eligible participants were female, the mean age was 55 years, and nearly half were either Hispanic/Latino (30%) or African American (15%) (Table S4). On average, participants were overweight (BMI mean = 29 kg/m^2) and had high serum low-density lipoprotein cholesterol (mean = 135 mg/dL). There was a high prevalence of hypertension (49%). Holding all adjustment variables constant, P wave and TP segment durations were the most strongly correlated among the six ECG traits (partial correlation ρ=-0.65), whereas T wave and P wave durations (ρ=0.01) as well as QRS and PR durations (ρ=-0.02) were largely uncorrelated (Table S5).

Overview of association results

Approximately 22M SNPs met our inclusion criteria and were used to assess for associations with the six contiguous and two composite ECG traits (Figures S1, S2). Lead SNPs at 82 of 149 loci (56%) previously reported by 26 interval scale ECG trait GWAS analyses (Table S6) were identified at genome-wide significance in our multi-ethnic population. The successful identification of previously recognized loci varied by trait (Table S7), ranging from 21 of 45 (47%) loci for QRS interval, to nine of 14 (64%) loci for P wave. When using a lower significance threshold of p_{aSPU}<0.0003 (0.05/149), 123 of the 149 (83%) previously recognized interval scale ECG trait loci were identified.

An additional six loci were >500 kb away from all lead SNPs previously reported by interval scale ECG trait GWAS and are presented as novel (Table 1, **Figure 1**). As described below, our results highlight the utility of phenotype decomposition, ancestral diversity, and combined-phenotype testing for the identification and characterization of complex trait loci.

Table 1. Novel genome wide-significant ($p_{aSPU} < 5x10^{-9}$) loci discovered in genome-wide association study of six contiguous electrocardiographic traits that decompose an average heartbeat in N=34,668 participants from the multi-ethnic Population Architectures using Genomics and Epidemiology study and the Multi-Ethnic Study of Atherosclerosis

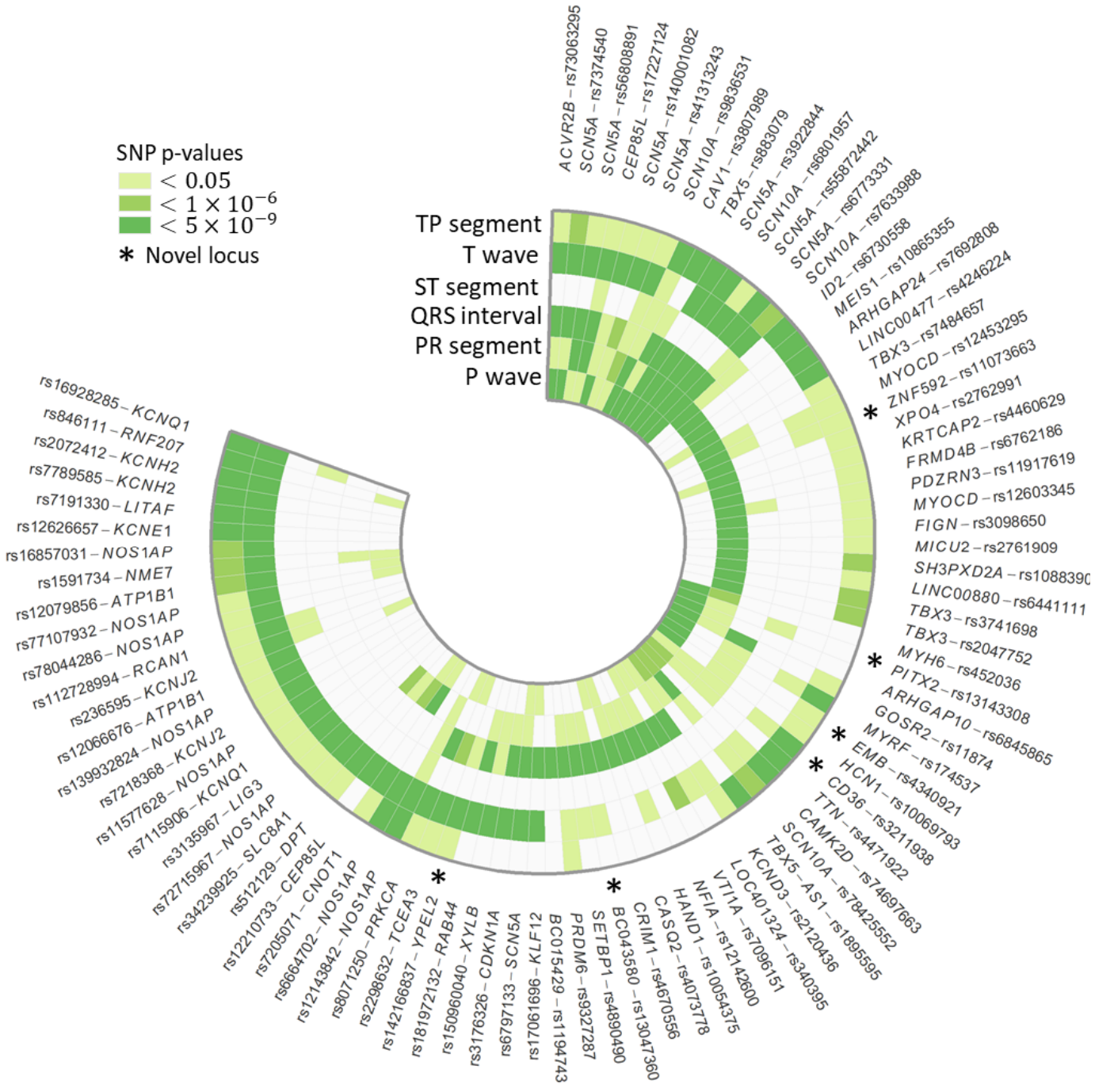
										Contiguous ECG traits						Composite ECG traits	
Lead SNP	Cnr	Position	Coded allele	Non-coded allele	Locus	AA	EA	CHN	HIS	P wave	PR segment	QRS interval	ST segment	T wave	TP segment	QT interval	PR interval
rs13143308	4	111714419	T	G	*PITX2*	30%	21%	74%	39%	**$2x10^{-11}$**	$5x10^{-4}$	$2x10^{-1}$	$3x10^{-2}$	$3x10^{-1}$	$1x10^{-1}$	$4x10^{-1}$	$8x10^{-1}$
rs4340921	5	49687697	C	T	*EMB*	66%	46%	49%	44%	**$8x10^{-13}$**	$8x10^{-1}$	$1x10^{-3}$	$7x10^{-1}$	$2x10^{-1}$	$2x10^{-4}$	$7x10^{-1}$	$2x10^{-3}$
rs3211938	7	80300449	G	T	*CD36*	10%	<0.01%	<0.01%	1%	$2x10^{-5}$	$8x10^{-3}$	$5x10^{-3}$	$4x10^{-1}$	$1x10^{-5}$	**$1x10^{-13}$**	**$6x10^{-10}$**	$6x10^{-6}$
rs11073663	15	85260268	A	G	*ZNF592*	27%	54%	19%	48%	$4x10^{-1}$	**$3x10^{-10}$**	$5x10^{-1}$	$4x10^{-1}$	$2x10^{-3}$	$7x10^{-3}$	$2x10^{-2}$	$6x10^{-7}$
rs142166837	17	57471022	C	T	*YPEL2*	31%	52%	32%	49%	$5x10^{-2}$	$6x10^{-1}$	$1x10^{-3}$	$1x10^{-1}$	**$4x10^{-11}$**	$1x10^{-2}$	$4x10^{-7}$	$8x10^{-1}$
rs13047360	21	28851580	G	A	*BC043580*	7%	17%	23%	16%	$7x10^{-1}$	$3x10^{-2}$	**$2x10^{-11}$**	$2x10^{-1}$	$5x10^{-2}$	$1x10^{0}$	$2x10^{-1}$	$3x10^{-2}$

aSPU: adaptive sum of powered tests

AA: African American, CHN: Chinese-American, EA: European-American, HIS: Hispanic/Latinos

Bolded values exceed the genome-wide significance threshold ($p<5\times10^{-9}$)

Figure 1. Lead SNPs at 87 loci significantly associated (p_{aspu}<5x10^{-9}) with six contiguous ECG traits spanning an average heartbeat, in n=34,668 multi-ethnic participants in the Population Architecture Using Genomics and Epidemiology study and Multi-Ethnic Study of Atherosclerosis. Outer stars denote novel loci and darker shades of green indicate lower p-values. To aid interpretation, lead SNPs were organized into broadly similar groups using hierarchical cluster analysis.



138 Phenotype decomposition

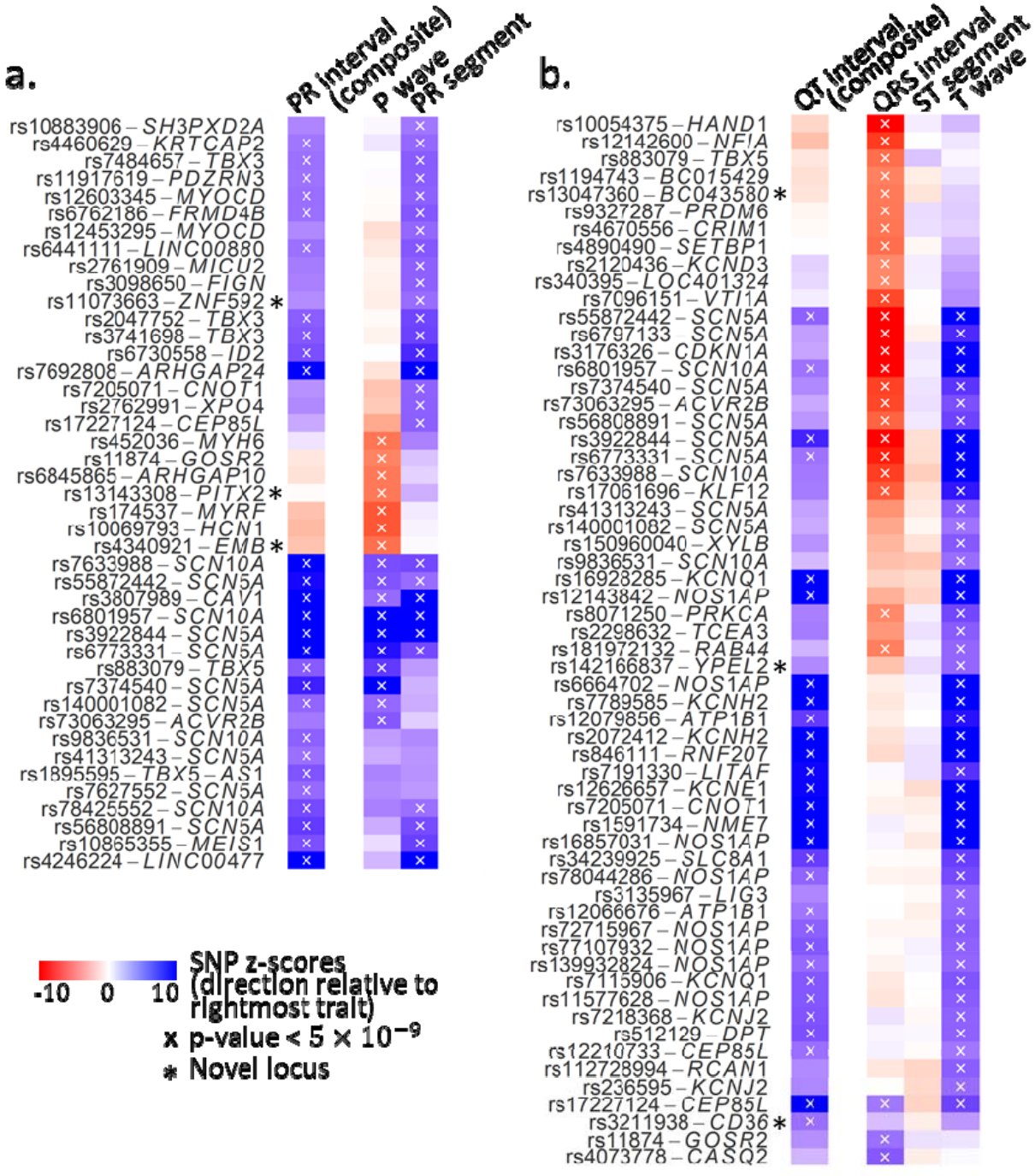
139 Of the six novel loci identified for the contiguous traits, two loci were identified for P wave and one locus
140 each was identified for PR segment, QRS interval, T wave, and TP segment. None of the novel loci were
141 associated with ST segment or affected multiple contiguous ECG traits at genome-wide significance
142 levels.

143

144 We then contrasted results for the six contiguous ECG traits with results from the two composite ECG
145 traits, QT interval (QRS interval + ST segment + T wave) and PR interval (P wave + PR segment) (Figure
146 **2**). *CD36* was the only novel locus identified for both a contiguous (TP segment) and a composite (QT
147 interval) ECG trait (Table 1). We also examined evidence of consistency of SNP effects by grouping
148 traits according to whether they affected atrial (PR interval, PR segment, and P wave) or ventricular (QT
149 interval, QRS interval, T wave, and ST segment) conduction. For atrial traits, novel loci identified for the
150 contiguous traits had varying directions of effects (Figure 2a, Table S9), which when combined resulted
151 in near-zero estimated effects for the composite trait. For example, every copy of the T allele for *PITX2*
152 lead SNP rs13143308 increased P wave duration by 0.63 milliseconds [ms] ($p_{univariate}=2x10^{-11}$), but
153 shortened the PR segment by 0.58 ms ($p_{univariate}=6x10^{-4}$). However, when evaluated together as the
154 composite trait PR interval, every copy of the rs13143308 T allele prolonged the PR interval by 0.03 ms
155 ($p_{univariate}=0.84$). Similarly, among the 59 loci associated with ventricular conduction, two of the three
156 novel loci (rs142166837 and rs13047360) had opposite effects on QRS interval and T wave duration,
157 which did not reach genome-wide significance when summed for the composite trait QT interval (Figure
158 2, Table S9). There were no instances of either PR or QT interval identifying a novel locus not associated
159 with any of the six contiguous traits at the genome-wide level.

160

Figure 2. Lead SNPs at all loci significantly ($P_{aSPU}<5x10^{-9}$) associated with (a) atrial and (b) ventricular conduction measures, in n=34,668 multi-ethnic Population Architecture Using Genomics in Epidemiology (PAGE) Study and Multi-Ethnic Study of Atherosclerosis participants. Outer stars denote novel loci; blue and red shading signifies opposite directions of effect, relative to PR interval and QT interval. To aid interpretation, lead SNPs were organized into broadly similar groups using hierarchical cluster analysis.



168 Ancestral diversity

169 Lead SNPs at five of the six novel loci were common (MAF >5%) across ancestral populations, with little
170 evidence of heterogeneity of effect across race/ethnicity (Table S8). One locus (*CD36*) showed evidence
171 of population specificity, with lead SNP rs3211938 near-monomorphic in European and Chinese
172 populations (MAF<0.01%), infrequent in Hispanic/Latinos (MAF=1%), and common in African
173 Americans (MAF=10%). Variant rs3211938 showed genome-wide significant associations with TP
174 segment ($p_{univariate}=1\times10^{-13}$) and QT interval ($p_{univariate}=6\times10^{-10}$) and nominal associations with P wave
175 ($p_{univariate}=2\times10^{-5}$), PR segment ($p_{univariate}=0.008$), and QRS interval ($p_{univariate}=0.005$). Although no GWAS
176 of TP segment has been published, each copy of the rs3211938 G allele increased QT interval by 3.70 ms.
177 Reported effects for common (MAF>5%) QT lead SNPs range from 0.5 ms to 3.5 ms(13). SNP
178 rs3211938 was either genotyped or well-imputed across studies and ancestry groups (imputation quality >
179 0.89, Table S10).

180

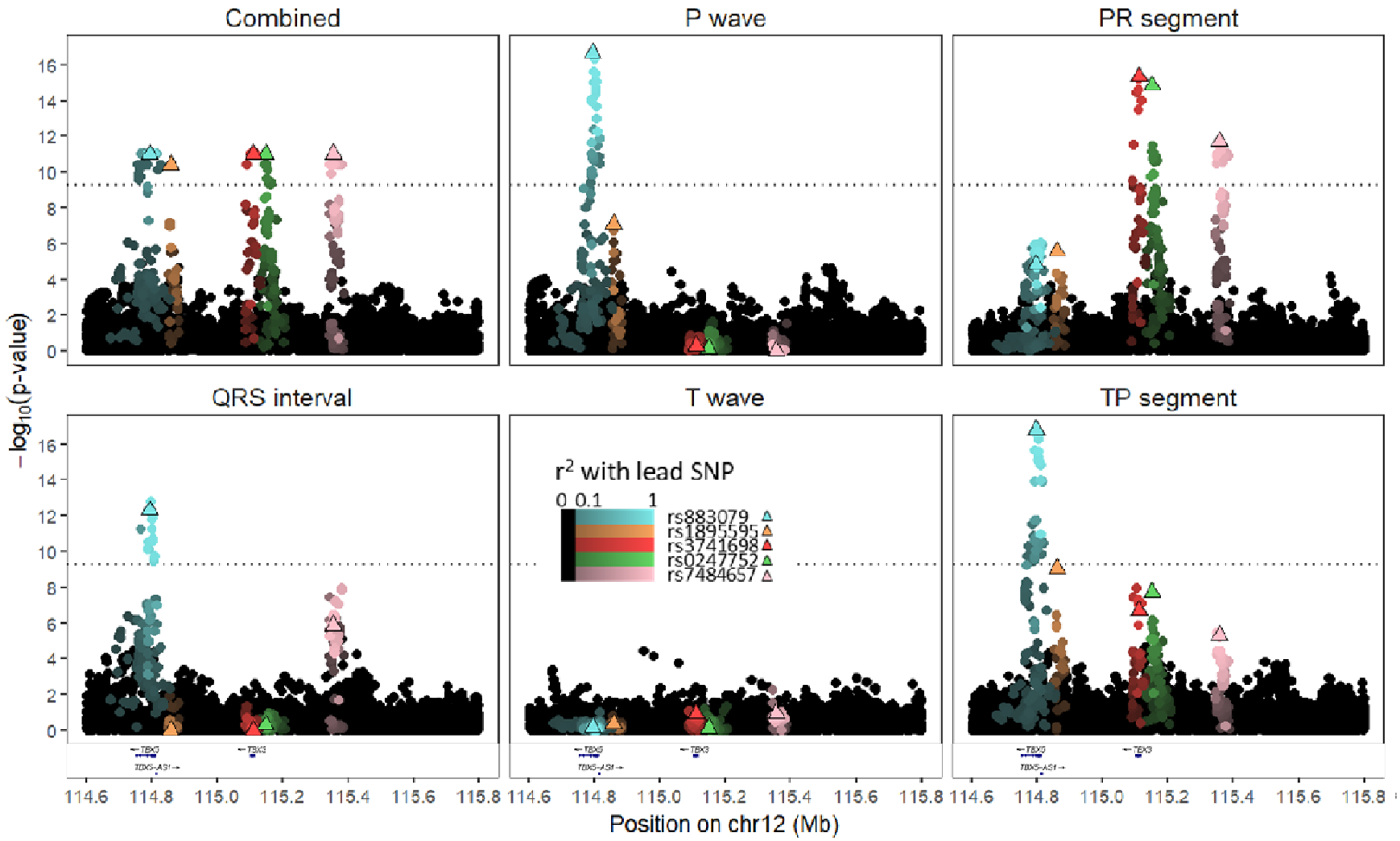
181 Combined phenotype analyses

182 We found widespread evidence of shared genetic effects across ECG traits. One fourth of lead SNPs
183 identified as genome-wide significant (P_{aspu}<h5x10^{-9}) had univariate associations with at least two ECG
184 traits ($p_{Univariate}$<5x10^{-9}). Lead SNPs at *ACVR2B, SCN5A, SCN10A, CEP85L, CAV1*, and *TBX5* were
185 associated with three or more ECG traits at univariate genome-wide significance levels. As expected,
186 traits that were more highly correlated also showed stronger evidence of shared genetic effects, with 10 of
187 the 20 lead SNPs that were associated with PR segment also showing genome-wide associations with TP
188 segment. However, evidence of shared genetic effects among uncorrelated traits was also observed. For
189 example, eight genome-wide significant SNPs at *SCN5A, SCN10A, CEP85L*, and *CNOT1* exhibited
190 significant univariate associations with both the T wave and P wave, despite low correlation between the
191 two traits ($\rho = 0.02$).

192

There also was evidence of allelic heterogeneity for multiple ECG traits. As an example, five signals in low LD ($r^2 < 0.1$) were detected within a 500 kb region near the previously identified locus *TBX5*, each associated with a distinct combination of ECG traits. Two of the five independent signals (rs3741698 and rs2047752) were largely specific to PR segment (**Figure 3**). The other three signals involved PR segment and QRS interval (lead SNP rs4784657), P wave, QRS interval, and TP segment (lead SNP rs883079), and the combined phenotype only (lead SNP rs1895595). Lead SNPs also showed some evidence of variation across traits, including the locus identified by lead SNP rs7484657, for which p-values for the QRS interval lead SNP differed by approximately three orders of magnitude from the rs7484657 p-value.

Figure 3. Regional SNP associations and linkage disequilibrium at four independent signals near *TBX5* among 34,668 participants with electrocardiographic data in the Population Architectures using Genomics and Epidemiology (PAGE) study and Multi-Ethnic Study of Atherosclerosis. Lighter colors indicate greater linkage disequilibrium with lead SNPs, and black markers denote SNPs not in LD ($r^2 < 0.1$) with any of the four lead SNPs. Combined phenotype p-values are truncated at $1x10^{-11}$.



DISCUSSION

Using cardiac conduction and repolarization traits as an example, we examined the extent to which combined multi-ethnic GWAS analyses of carefully selected phenotypes that map to well-defined biology improved detection and characterization of loci. We identified six novel loci, five of which were detected only when examining the more precisely defined phenotypes, and a sixth locus that was specific to African ancestral populations. We also showed how leveraging evidence of a shared genetic architecture aided the characterization of known loci, particularly when loci harbored multiple independent signals that differed by trait. In this mega-GWAS era involving predominantly European ancestral populations, this study, conducted in a population one-third the size of the largest published ECG trait GWAS (13,14), underscores the merits of prioritizing diversity and phenotype measurement.

Of the three GWAS challenges we examined, our deliberate selection of phenotype measures mapping to well-defined biology largely drove locus discovery, challenging current trends in GWAS that emphasize increased sample size. The growing scale of GWAS, which today can surpass one million participants(15), has resulted in the prioritization of commonly available traits (e.g., body mass index) over traits that more precisely capture underlying biology (e.g., direct measures of body fat(16)). In our case, composite ECG traits PR interval and QT interval have been most commonly interrogated by GWAS. However, these traits represent aggregates of physiologically distinct mechanisms, which may obscure loci with effects localized to, or inconsistent across, individual contiguous traits. This phenomenon was illustrated by the *PITX2* locus, a locus associated with atrial fibrillation.(17) Because *PITX2* lead SNP rs13143308 had opposing associations with the contiguous P wave and PR segment, a standard approach using the composite PR interval yielded a near-zero effect, obscuring the potential importance of the locus on atrial function regardless of sample size. These results emphasize the need to balance ongoing investments in large-scale genome measurement with use of precision phenotyping, for instance through efforts like the ongoing Precision Medicine Initiative's *All of Us* Research Program(18).

The six traits we used in our ECG decomposition were motivated by their relations to physiology, and their coherence as an aggregate electrophysiologic phenotype. While an important complement to traditional, coarser measures like the PR and QT intervals, our phenotype decomposition approach that identified novel loci and improved characterization of known loci captured but a fraction of the full variation in ECG phenotypes; further phenotypic specificity and additional biologic insight may be offered by GWAS of other ECG traits, including measures of waveform amplitudes, angles, or variability. A complementary approach might select traits that are governed by a shared genetic architecture, such as ion channel function or cardiac remodeling when applied to ECG trait GWAS, potentially assisting efforts to map loci to specific biologic pathways. Further extending combined phenotype ECG trait GWAS to include other phenotypes and traits (e.g. cardiometabolic traits or cardiovascular diseases) also is warranted, given evidence that these traits represent interrelated manifestations of common biologic mechanisms (12)a s well as success of prior combined phenotype studies to disentangle complex biology(19). Overall, the question of how to select intermediate traits and integrate such traits with other phenotypic data, including clinical and prognostic information, remains an open question, with best practices that likely will vary across complex traits.

Despite mounting interest in combined phenotype statistical approaches, their merits for novel locus discovery and locus characterization remain largely untested. Here, combined phenotype analysis of contiguous ECG traits did not identify novel loci that eluded traditional univariate analyses, despite the purported ability of combined phenotype methods to increase statistical power. However, preliminary evaluation of *TBX5*, a locus harboring multiple independent signals, suggested that combined phenotype approaches may be particularly informative for fine-mapping. Supporting the extension of combined phenotype methods to fine-mapping are methods that have been specifically developed for this challenge(20), including fastPAINTOR. When compared with single trait fine-mapping, fastPAINTOR

reduced the number of SNPs required for follow-up in order to capture 90% of the causal variants, from 23 SNPs per locus using a single trait to 12 SNPs when fine-mapping two traits simultaneously. Other potentially fruitful avenues may extend combined-phenotype studies of ECG traits to include cardiovascular disease.

The lack of diversity in GWAS has long been described(21), but the literature remains dominated by studies on European ancestral populations. As a result, genomics research is confined to a narrow sliver of human genetic diversity, even as the US population becomes more diverse(22). Our deliberate selection of an ancestrally diverse population enabled the identification of a novel *CD36* locus, which was common only in populations of African descent. Lead SNP, rs3211938, had a large effect on QT interval, among the largest effects reported to-date,(13) although winner's curse may be a concern(13). Variant rs3211938, a ClinVar-indexed missense mutation known to cause CD36 deficiency, encodes a scavenger receptor central to formation and cellular uptake of long-chain fatty acids. Although *CD36* and rs3211938 have been associated with a spectrum of cardiometabolic phenotypes,(23–30) the most intriguing finding is the potential linkage with sudden cardiac arrest (SCA), for which QT interval prolongation increases risk(31). SCA accounts for approximately 10-20% of total mortality in industrial countries(32), and several decades of research have suggested a contributory role of impaired fatty-acid uptake in cardiomyocytes. Although genetic studies of *CD36* and SCA were largely null(33,34), the use of predominantly European ancestral populations constrained evaluation of rs3211938, which is near monomorphic in all populations except those of African descent. Overall, these results highlight the potential for racially/ethnically diverse studies to provide novel biological insights, beyond the reach of predominantly European ancestral populations.

Limitations of our study point to several promising directions for future work. First, we lacked a replication cohort, reflecting the rarity of multi-ethnic studies with high-resolution ECGs from which to derive the six contiguous ECG traits. However, this study is the largest multi-ethnic GWAS of ECGs to-

date, with excellent statistical power, and we identified loci that are biologically plausible. Second, we limited our evaluation to common variants, although previous studies have demonstrated the utility of interrogating rare variants, particularly in the context of multi-ethnic studies(35,36). Our focus on common variants reflects the current limitations of combined phenotype methods for interrogating rare variants in complex samples or with summary data. Widespread interest in this approach suggests that this gap may be closed soon. Further, while this study helps address the lack of diversity in ECG trait GWAS, the small number of Chinese American participants limited our ability identify loci that were common only in populations of East Asian descent. Future efforts that further expand population racial/ethnic diversity represents an important next both for cardiac conduction studies and GWAS more broadly. Finally, we did not perform in-depth fine-mapping, although approaches that leverage multiple phenotypes have been described(20). Identification of allelic heterogeneity provides a clear impetus for future studies that leverage evidence of a shared genetic effects to disentangle the genetic architecture underlying ECG traits and, more broadly, complex traits.

This study illustrates three strategies to improve the efficiency of locus discovery. Of these, our findings emphasize the importance of carefully constructed phenotypes and of ancestral diversity for novel locus identification. In contrast, combined phenotype methods did not enable identification of novel loci unapparent using traditional approaches, although combined phenotype studies did inform characterization of known loci. As researchers contemplate the next generation of genomics studies, increased phenotype resolution and ancestral diversity will be crucial to understanding the ever-expanding "phenome," while ensuring equitable access to precision medicine.

MATERIAL AND METHODS

Study population

Data sources

The multi-ethnic PAGE study(6) is a consortium funded by the National Human Genome Research Institute (NHGRI) to examine the genetic underpinnings of common complex diseases and traits in ancestrally-diverse populations (online supplement). PAGE data used in this study included African American, European American, and Hispanic/Latino participants enrolled in the Women's Health Initiative (WHI)(37) Clinical Trial, and in the population-based Atherosclerosis Risk in Communities Study (ARIC)(38), and Hispanic Community Health Study/Study of Latinos (HCHS/SOL)(39). Data for African American, Chinese American, European American, and Hispanic/Latino populations also were contributed by the population-based Multi-Ethnic Study of Atherosclerosis (MESA)(40) (online supplement).

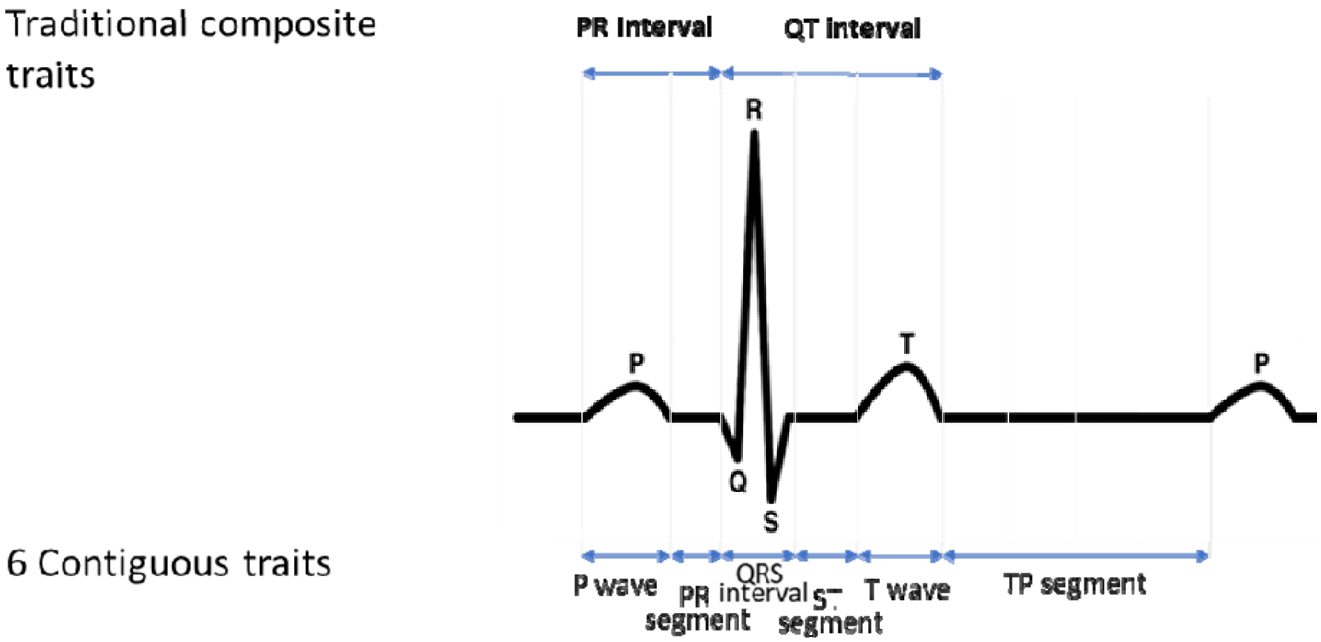
Electrocardiography

Ten-second, resting, standard 12-lead ECGs were collected following standard protocols (online supplement). To enable direct comparison with the published literature, we used exclusion criteria from published GWAS of QT interval, QRS duration, PR interval, and P wave . Poor-quality ECGs and ECGs exhibiting severe abnormalities were excluded (Table S1, S2). Also excluded were participants with prevalent heart failure or coronary heart disease, and participants taking any class I or III anti-arrhythmic medication.

We temporally decomposed the ECG waveform into six contiguous ECG traits, each reflecting distinct physiological processes: P wave, PR segment, QRS interval, ST segment, T wave, and TP segment (Figure 4). Three of the six contiguous traits were measured directly from maximum readings across all

12 ECG leads (P wave, QRS interval, T wave), and were used to construct the remaining three traits, as detailed in supplementary material. We also evaluated two composite ECG traits that are widely available and have been examined in numerous GWAS analyses: the QT interval(13) and the PR interval(14). The PR interval combines the contiguous P wave and PR segment, reflecting conduction through the atria and the AV-node. The QT interval spans the contiguous QRS interval, ST segment, and the T wave, representing ventricular activity.

Figure 4. Illustration of the six contiguous (P wave, PR segment, QRS interval, ST segment, and TP segment) and two composite (QT interval and PR interval) ECG traits.



Genotyping and Imputation

Imputation was performed across studies, using the 1000 Genomes Project phase 3 reference panel (Table S3). In addition to study-specific quality control protocols, we excluded SNPs meeting any of the following criteria: minor allele frequency (MAF) <1%; imputation quality score < 0.3; or effective sample size < 30, calculated as $2 \times MAF(1 - MAF) \times N \times Q$, where Q is imputation quality and N is study size.

Statistical Methods

Overview

Statistical analyses were carried out in two stages. First, genome-wide univariate associations for each of six contiguous ECG traits were calculated in twelve substudies defined by genotyping platform, race/ethnicity, and study (Table S3). The pooled PAGE sample of Hispanic/Latino and African American participants from the HCHS/SOL and WHI genotyped together(41) on the Multi-Ethnic Genotyping Array (MEGA) array was retained as a single substudy. The remaining 11 substudies were defined by ancestry (African American, Chinese American, and European American) and study (ARIC, MESA, and WHI). Substudy-specific univariate results for the six contiguous ECG traits were then combined via trait-specific inverse-variance meta-analysis(42). Next, meta-analyzed contiguous ECG trait-specific univariate associations were evaluated in aggregate using the adaptive sum of powered scores method(43) (aSPU, described below), yielding a combined-trait p-value for each SNP. All univariate and combined-phenotype analyses were performed in the multi-ethnic population as demonstrated previously in the PAGE study(44), and, in sensitivity analyses, by self-identified race/ethnicity.

358 *Univariate associations*

359 Univariate associations for the six contiguous ECG traits and the two composite ECG traits were
360 estimated assuming an additive genetic model of inheritance and adjusting for linear effects of age at
361 study exam, sex, study site or region, inverse heart rate, and ancestral principal components(45). The
362 pooled PAGE sample was analyzed using generalized estimating equations allowing correlated errors
363 among households sharing first or second-degree relatives, and independent error distributions
364 by self-reported ancestry group(46). Linear regression was used for the other 11 analytic groups.
365 Models were implemented using SUGEN (46) (ARIC, pooled PAGE sample, WHI) and SNPTest
366 (MESA). For each continuous and composite ECG trait, results were combined across substudies using
367 inverse-variance weighted, genomic inflation-corrected meta-analysis(42). SNP effect heterogeneity
368 across substudies and, in sensitivity analyses, among self-reported race/ethnicity was measured with the
369 Cochran's Q test. For each ECG trait, SNP meta-analysis p-values were assessed by calculating genomic
370 inflation factors (λ) and plotting the expected distribution against observed results.

371

372 *Combined phenotype analyses*

373 To evaluate evidence of shared genetic effects across all six contiguous ECG traits, meta-analyzed
374 univariate results were combined with aSPU(43). The aSPU procedure was selected based on its
375 accommodation of effect direction across traits, low type I error rate in simulation studies when compared
376 with other available procedures (data not shown), and its scalability to 1000 genomes imputed data. We
377 implemented aSPU using Julia 1.0 to optimize efficiency, and made the code freely available as a Julia
378 package (https://github.com/kaskarn/JaSPU).

379

aSPU uses meta-analyzed univariate summary z-scores, calculated for each SNP across all *k* traits. Briefly, the procedure estimates Σ, the $k \times k$ correlation of null z-scores across univariate results and draws 10^{11} Monte-Carlo samples from the multivariate $N_k(0, \Sigma)$ distribution. For each SNP *j*, the results for all *k* traits $z_{j1}, z_{j2}, \ldots, z_{jk}$ are used to form the sequence of powered scores: $SPU(\gamma) = z_1^\gamma + z_2^\gamma + \ldots + z_k^\gamma$, where $\gamma = 0, 1, \ldots, 8, \infty$. This sequence is compared to values produced by all 10^{11} draws from the simulated null distribution, and a Monte-Carlo p-value is derived for the best-powered SPU(γ) to calculate an overall SNP- p-value (p_{aSPU}), ranging between $1/(1+10^{11})$ and 1. The test is adaptive in the sense that the γ yielding the maximal SPU(γ) can vary across SNPs, so that the test maintains power against a number of different alternative hypotheses.

Reporting

Multi-ethnic combined phenotype results were presented as the primary findings, employing Bonferroni correction assuming 10M SNP tests (i.e. $p_{aSPU}<5\times10^{-9}$). Ancestry-specific and ECG-trait specific analyses were performed to aid interpretation of results and are considered sensitivity analyses.

Previously-reported SNPs were identified through review of the NHGRI-EBI GWAS Catalog(12) as of January 22, 2019 (Table S6), for the following interval scale ECG phenotypes: PR interval, PR segment, QRS interval, P wave, QT interval, and T wave, supplemented by a literature review to identify publications not indexed by the NHGRI-EBI GWAS Catalog(14,44,47). There were no published GWAS analyses of ST segment or TP segment durations.

We defined a locus to encompass all SNPs within 500 kb of, and in linkage-disequilibrium (LD) with, the lead SNP (multi-ethnic $r^2 > 0.2$ in the unrelated PAGE study participants). SNPs > 500 kb away from all previously reported ECG trait loci were considered novel. Locus-specific lead SNPs were identified as the SNP with the lowest p_{aSPU}. Tied lowest p_{aSPU} were resolved using the lowest locus-specific univariate p-value across contiguous ECG traits. Complete summary level results from all analyses are made available through dbGaP (phs000356).

Ethics Statement

All data were analyzed anonymously, and this study was exempted from review under 45 CFR 46.101(b) by the University of North Carolina institutional review board (IRB #18-1764).

411 **SUPPORTING INFORMATION**

412 Supporting material and methods, tables (S1-9), and figures (S1-2) are included in two additional files

413 **Figure S1. Quantile-Quantile plots for trans-ancestry meta-analyzed associations of SNPs with each**
414 **of six ECG traits (decomposed ECG phenotype)** in n=34,668 participants from the Population
415 Architectures using Genomics and Epidemiology (PAGE) study and the Multi-Ethnic Study of
416 Atherosclerosis (MESA). Black markers and lambda values represent to p-values for all SNPs that passed
417 quality control. Blue markers represent the subset of SNPs >500kb from any previously-reported ECG
418 lead SNP.

419 **Figure S2 Manhattan plots for univariate, trans-ethnic ECG trait GWAS** in 34,668 participants from
420 the Population Architectures using Genomics and the and Multi-Ethnic Study of Atherosclerosis
421 (MESA).. Genome-wide ($p_{univariate} < 5\times10^{-9}$) significant loci within 500kb of previously-reported ECG
422 GWAS results are shown in yellow; genome-wide ($p_{univariate} < 5\times10^{-9}$) significant loci >500 kb from
423 previously reported ECG loci are shown in blue.

424

425 **Table S1.** Exclusion criteria for genome-wide study of all evaluated electrocardiographic traits.

426 **Table S2**. Participant exclusions (cumulative), by study and ancestry group

427 **Table S3**. Genotyping, imputation, and quality control by participating study

428 **Table S4. Participant characteristics by study of origin, and ancestry group**, among N=34,668
429 participants from the Population Architecture using Genomics and Epidemiology (PAGE) study and the
430 Multiethnic Study of Atherosclerosis (MESA).
431
432 **Table S5**. **Partial correlations between six contiguous traits decomposing the ECG**, in n=10,618
433 eligible Hispanic Community Health Study/Study of Latinos participants. Partial correlations were
434 adjusted for RR interval, gender, study site, and ancestry principal components.
435
436 **Table S6. Published genome-wide association studies of electrocardiographic traits**, indexed on the NHGRI
437 GWAS catalog (Sept. 30, 2018).
438
439 **Table S7**. **Previously-reported associations of SNPs with temporal electrocardiographic traits**, and
440 corresponding results in N=34,668 participants in the Population Architecture using Genomics and
441 Epidemiology (PAGE) study and the Multiethnic Study of Atherosclerosis (MESA).

442 **Table S8 Trait-specific direction of effects, and meta-analysis heterogeneity p-values for univariate**
443 **associations (punivariate < 5×10-9) of loci discovered in combined-phenotype analyses of six**
444 **electrocardiographic traits (paSPU < 5x10-9),** in 34,668 participations from the Population
445 Architecture usingGenomics and Epidemiology study (PAGE) and Multi-Ethnic Study of Atherosclerosis
446 (MESA).
447

448 **Table S9 Trait-specific trans-ethnic meta-analyzed effect estimates and standard errors for**
449 **univariate associations (punivariate < 5×10-9) of loci discovered in combined-phenotype analyses of**
450 **six electrocardiographic traits (paSPU < 5x10-9)**, in 34,668 participations from the Population
451 Architecture using Genomics and Epidemiology study (PAGE) and the Multi-Ethnic Study of
452 Atherosclerosis (MESA).
453
454 **Table S10** Imputation quality scores of all novel variants reported in this study.

ACKNOWLEDGMENTS

The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions. The complete list of PAGE members can be found at http://www.pagestudy.org.

The Stanford Global Reference Panel was created by Stanford - contributed samples and comprises multiple datasets from multiple researchers across the world designed to provide a resource for any researchers interested in diverse population data on the Multi-Ethnic Global Array (MEGA). The authors thank the researchers and research participants who made this dataset available to the community.

The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.

The authors thank the staff and participants of the ARIC study for their important contributions. More detail about the ARIC study may be found at: https://sites.cscc.unc.edu/aric/.

Genotyping for MESA was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. We also thank the other investigators, the staff and the participants of MESA for their valuable contributions. A full list of participating MESA investigators and institutions can be found online (http://www.mesa-nhlbi.org).

REFERENCES

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics. 2017 Jul;101(1):5–22.

2. LeBlanc M, Zuber V, Andreassen BK, Witoelar A, Zeng L, Bettella F, et al. Identifying Novel Gene Variants in Coronary Artery Disease and Shared Genes With Several Cardiovascular Risk Factors. Circ Res. 2016 Jan 8;118(1):83–94.

3. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet. 2018 Mar;50(3):390–400.

4. Price AL, Spencer CCA, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. Proc Biol Sci. 2015 Dec 22;282(1821):20151684.

5. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016 13;538(7624):161–4.

6. Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, et al. The Next PAGE in Understanding Complex Traits: Design for the Analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. Am J Epidemiol. 2011 Oct 1;174(7):849–59.

7. Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proc Biol Sci [Internet]. 2016 Jul 27 [cited 2018 Oct 1];283(1835). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4971198/

8. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nature Reviews Genetics. 2018 Feb;19(2):110–24.

9. Silva CT, Kors JA, Amin N, Dehghan A, Witteman JCM, Willemsen R, et al. Heritabilities, proportions of heritabilities explained by GWAS findings, and implications of cross-phenotype effects on PR interval. Hum Genet. 2015 Nov 1;134(11–12):1211–9.

10. Food and Drug Administration, HHS. International Conference on Harmonisation; guidance on E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs; availability. Notice. Fed Regist. 2005 Oct 20;70(202):61134–5.

11. MacRae CA, Vasan RS. Next Generation GWAS: Time to Focus on Phenotype? Circ Cardiovasc Genet. 2011 Aug 1;4(4):334–6.

12. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017 Jan 4;45(D1):D896–901.

13. Arking DE, Pulit SL, Crotti L, Harst P van der, Munroe PB, Koopmann TT, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. Nature Genetics. 2014 Aug;46(8):826–36.

14. van Setten J, Brody JA, Jamshidi Y, Swenson BR, Butler AM, Campbell H, et al. PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. Nat Commun [Internet]. 2018 Jul 25 [cited 2018 Dec 13];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6060178/

15. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. 2018 Oct 15;27(20):3641–9.

16. Blundell JE, Dulloo AG, Salvador J, Frühbeck G. Beyond BMI - Phenotyping the Obesities. Obes Facts. 2014 Oct;7(5):322–8.

17. Syeda F, Kirchhof P, Fabritz L. PITX2-dependent gene regulation in atrial fibrillation and rhythm control. J Physiol (Lond). 2017 15;595(12):4019–26.

18. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genetics in Medicine. 2017 Jul;19(7):743–50.

19. VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ, Han Y, Spitz MR, Shete S, et al. Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. Am J Epidemiol. 2012 May 15;175(10):1013–20.

20. Kichaev G, Yang W-Y, Lindstrom S, Hormozdiari F, Eskin E, Price AL, et al. Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. PLOS Genetics. 2014 Oct 30;10(10):e1004722.

21. Bustamante CD, De La Vega FM, Burchard EG. Genomics for the world. Nature. 2011 Jul 14;475(7355):163–5.

22. Colby SL, Ortman JM. Projections of the Size and Composition of the U.S. Population: 2014 to 2060. Population Estimates and Projections. Current Population Reports. P25-1143 [Internet]. US Census Bureau; 2015 [cited 2018 Dec 8]. Available from: https://eric.ed.gov/?id=ED578934

23. Shibao CA, Celedonio JE, Ramirez CE, Love-Gregory L, Arnold AC, Choi L, et al. A Common CD36 Variant Influences Endothelial Function and Response to Treatment with Phosphodiesterase 5 Inhibition. J Clin Endocrinol Metab. 2016;101(7):2751–8.

24. Avery CL, He Q, North KE, Ambite JL, Boerwinkle E, Fornage M, et al. A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains. PLoS Genet. 2011 Oct;7(10):e1002322.

25. Gautam S, Agrawal CG, Banerjee M. CD36 gene variants in early prediction of type 2 diabetes mellitus. Genet Test Mol Biomarkers. 2015 Mar;19(3):144–9.

26. Love-Gregory L, Abumrad NA. CD36 genetics and the metabolic complications of obesity. Curr Opin Clin Nutr Metab Care. 2011 Nov;14(6):527–34.

27. Elbers CC, Guo Y, Tragante V, van Iperen EPA, Lanktree MB, Castillo BA, et al. Gene-centric meta-analysis of lipid traits in African, East Asian and Hispanic populations. PLoS ONE. 2012;7(12):e50198.

28. Musunuru K, Romaine SPR, Lettre G, Wilson JG, Volcik KA, Tsai MY, et al. Multi-ethnic analysis of lipid-associated loci: the NHLBI CARe project. PLoS ONE. 2012;7(5):e36473.

29. Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, et al. Large multiethnic Candidate Gene Study for C-reactive protein levels: identification of a novel association at CD36 in African Americans. Hum Genet. 2014 Aug;133(8):985–95.

30. Love-Gregory L, Sherva R, Sun L, Wasson J, Schappe T, Doria A, et al. Variants in the CD36 gene associate with the metabolic syndrome and high-density lipoprotein cholesterol. Hum Mol Genet. 2008 Jun 1;17(11):1695–704.

31. O'Neal WT, Singleton MJ, Roberts JD, Tereshchenko LG, Sotoodehnia N, Chen LY, et al. Association Between QT-Interval Components and Sudden Cardiac Death: The ARIC Study (Atherosclerosis Risk in Communities). Circulation: Arrhythmia and Electrophysiology [Internet]. 2017 Oct [cited 2019 Feb 28];10(10). Available from: https://www.ahajournals.org/doi/10.1161/CIRCEP.117.005485

32. Chugh SS. Sudden cardiac death in 2017: Spotlight on prediction and prevention. Int J Cardiol. 2017 Jun 15;237:2–5.

33. Lemaitre RN, Johnson CO, Hesselson S, Sotoodehnia N, Sotoodhenia N, McKnight B, et al. Common variation in fatty acid metabolic genes and risk of incident sudden cardiac arrest. Heart Rhythm. 2014 Mar;11(3):471–7.

34. Johnson CO, Lemaitre RN, Fahrenbruch CE, Hesselson S, Sotoodehnia N, McKnight B, et al. Common variation in fatty acid genes and resuscitation from sudden cardiac arrest. Circ Cardiovasc Genet. 2012 Aug 1;5(4):422–9.

35. Bunt M van de, Cortes A, Consortium I, Brown MA, Morris AP, McCarthy MI. Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. PLOS Genetics. 2015 Sep 25;11(9):e1005535.

36. Mensah-Ablorh A, Lindstrom S, Haiman CA, Henderson BE, Marchand LL, Lee S, et al. Meta-Analysis of Rare Variant Association Tests in Multiethnic Populations. Genetic Epidemiology. 2016 Jan 1;40(1):57–65.

37. Design of the Women's Health Initiative Clinical Trial and Observational Study. Controlled Clinical Trials. 1998 Feb 1;19(1):61–109.

38. THE ATHEROSCLEROSIS RISK IN COMMUNIT (ARIC) STUI)Y: DESIGN AND OBJECTWES. Am J Epidemiol. 1989 Apr 1;129(4):687–702.

39. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, et al. Design and Implementation of the Hispanic Community Health Study/Study of Latinos. Annals of Epidemiology. 2010 Aug 1;20(8):629–41.

40. Bild DE, Bluemke DA, Burke GL, Detrano R, Roux D, V A, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. Am J Epidemiol. 2002 Nov 1;156(9):871–81.

41. Wojcik G, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic Diversity Turns a New PAGE in Our Understanding of Complex Traits. 2017 Sep 15;

42. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1;26(17):2190–1.

43. Kim J, Bai Y, Pan W. An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics. Genet Epidemiol. 2015 Dec 1;39(8):651–63.

590 44. Wojcik G, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. The PAGE Study: How
591 Genetic Diversity Improves Our Understanding of the Architecture of Complex Traits. Nature. (In
592 press);188094.

593 45. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components
594 analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006
595 Aug;38(8):904–9.

596 46. Lin D-Y, Tao R, Kalsbeek WD, Zeng D, Gonzalez F, Fernández-Rhodes L, et al. Genetic Association
597 Analysis under Complex Survey Sampling: The Hispanic Community Health Study/Study of Latinos.
598 The American Journal of Human Genetics. 2014 Dec 4;95(6):675–88.

599 47. Prins BP, Mead TJ, Brody JA, Sveinbjornsson G, Ntalla I, Bihlmeyer NA, et al. Exome-chip meta-
600 analysis identifies novel loci associated with cardiac conduction, including ADAMTS6. Genome Biol
601 [Internet]. 2018 Jul 17 [cited 2018 Dec 13];19. Available from:
602 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6048820/

603