Circulation: Genomic and Precision Medicine

ORIGINAL ARTICLE

Multi-Ethnic Genome-Wide Association Study of Decomposed Cardioelectric Phenotypes Illustrates Strategies to Identify and Characterize Evidence of Shared Genetic Effects for Complex Traits

Antoine R. Baldassari, MSPH; Colleen M. Sitlani, PhD; Heather M. Highland, PhD; Dan E. Arking, PhD; Steve Buyske, PhD; Dawood Darbar, MD; Rahul Gondalia, PhD; Misa Graff, PhD; Xiuqing Guo, PhD; Susan R. Heckbert, MD, PhD; Lucia A. Hindorff, PhD; Chani J. Hodonsky, PhD; Yii-Der Ida Chen, PhD; Robert C. Kaplan, PhD; Ulrike Peters, PhD; Wendy Post, MD, MS; Alex P. Reiner, MD, MSC; Jerome I. Rotter, PhD; Ralph V. Shohet, MD; Amanda A. Seyerle, PhD; Nona Sotoodehnia, MD, PhD; Ran Tao, PhD; Kent D. Taylor, PhD; Genevieve L. Wojcik, PhD; Jie Yao, MD, MS; Eimear E. Kenny, PhD; Henry J. Lin, MD; Elsayed Z. Soliman, MD, MS, MSc; Eric A. Whitsel, MD; Kari E. North, PhD; Charles Kooperberg, PhD; Christy L. Avery, PhD

BACKGROUND: We examined how expanding electrocardiographic trait genome-wide association studies to include ancestrally diverse populations, prioritize more precise phenotypic measures, and evaluate evidence for shared genetic effects enabled the detection and characterization of loci.

METHODS: We decomposed 10 seconds, 12-lead electrocardiograms from 34 668 multi-ethnic participants (15% Black; 30% Hispanic/Latino) into 6 contiguous, physiologically distinct (P wave, PR segment, QRS interval, ST segment, T wave, and TP segment) and 2 composite, conventional (PR interval and QT interval) interval scale traits and conducted multivariable-adjusted, trait-specific univariate genome-wide association studies using 1000-G imputed single-nucleotide polymorphisms. Evidence of shared genetic effects was evaluated by aggregating meta-analyzed univariate results across the 6 continuous electrocardiographic traits using the combined phenotype adaptive sum of powered scores test.

RESULTS: We identified 6 novels (*CD36, PITX2, EMB, ZNF592, YPEL2*, and *BC043580*) and 87 known loci (adaptive sum of powered score test $P<5\times10^{-9}$). Lead single-nucleotide polymorphism rs3211938 at *CD36* was common in Blacks (minor allele frequency=10%), near monomorphic in European Americans, and had effects on the QT interval and TP segment that ranked among the largest reported to date for common variants. The other 5 novel loci were observed when evaluating the contiguous but not the composite electrocardiographic traits. Combined phenotype testing did not identify novel electrocardiographic loci unapparent using traditional univariate approaches, although this approach did assist with the characterization of known loci.

CONCLUSIONS: Despite including one-third as many participants as published electrocardiographic trait genome-wide association studies, our study identified 6 novel loci, emphasizing the importance of ancestral diversity and phenotype resolution in this era of ever-growing genome-wide association studies.

Key Words: cardiovascular diseases ■ electrophysiology ■ epidemiology ■ genome-wide association study ■ population

Correspondence to: Antoine R. Baldassari, MSPH, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 123 W Franklin St, Bldg C, Suite 4210, Chapel Hill, NC 27514. Email baldassa@email.unc.edu

The Data Supplement is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCGEN.119.002680.

For Sources of Funding and Disclosures, see page 279.

Circulation: Genomic and Precision Medicine is available at www.ahajournals.org/journal/circgen

Nonstandard Abbreviations and Acronyms

ARIC	Atherosclerosis Risk in Communities Study
aSPU	adaptive sum of powered score test
GWAS	genome-wide association study
HCHS/SOL	Hispanic Community Health Study/ Study of Latinos
MAF	minor allele frequency
MESA	Multi-Ethnic Study of Atherosclerosis
PAGE	Population Architecture Using Genomics and Epidemiology
SNP	single-nucleotide polymorphism

Genetic susceptibility underlies a majority of common diseases and traits, as demonstrated by genome-wide association studies (GWAS) that have identified thousands of genetic loci for cardiovascular, cardiometabolic, cancer, kidney, psychiatric, ocular, inflammatory, and neuromuscular traits.[1] Together, these GWAS have revealed common threads underlying the genetic architecture of complex diseases and traits, as well as research gaps. For example, evidence of shared genetic effects (ie, 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 in published GWAS, as the majority (>80%) of GWAS have been conducted in European ancestral populations.[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] Furthermore, the scale and collaborative nature of GWAS prioritize traits that are widely available, although these traits may not precisely capture phenotypic variation and underlying biology.[7,8] Together, these research gaps argue for expanding GWAS analyses to systematically examine evidence of shared genetic effects across a spectrum of biologically motivated traits in multi-ethnic populations.

Electrocardiograms measure a sequence of distinct electrophysiological processes in the myocardium that underlie cardiac conduction and repolarization. Electrocardiographic 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 of electrocardiographic traits.[12] Therefore, electrocardiographic 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 6 contiguous measures of the electrocardiogram waveform spanning an average heartbeat (Figure 1) using data from the multi-ethnic PAGE study (Population Architecture Using Genetic Epidemiology) and the MESA (Multi-Ethnic Study of Atherosclerosis). Our results illustrate the broad utility of multi-ethnic GWAS of carefully constructed individual and aggregate traits to illuminate the biology of complex diseases and traits.

METHODS

Methods for this article are detailed in Data Supplement.

Summary-level (PAGE) and individual-level (ARIC [Atherosclerosis Risk in Communities], HCHS/SOL [Hispanic Community Health Study/Study of Latinos], MESA, and PAGE) data are available at DbGaP (https://www.ncbi.nlm.nih.gov/gap/, accession numbers phs000090.v1.p1 [ARIC], phs000810.v1.p1 [HCHS/SOL], phs000293.v1 [MESA], phs000.56.v1.p1 [PAGE], and phs000200.v1 [Women's Health Initiative]).

The institutional review board of the University of North Carolina at Chapel Hill determined this study as exempt from review, further each participating study was approved by the institutional review board at the respective sites, and all participants provided written consent.

RESULTS

Sample Description

Of the 39 538 participants with GWAS and electrocardiogram data in ARIC, HCHS/SOL, MESA, and Women's Health Initiative, 34 668 (88%) met all inclusion criteria (Tables I and II in the Data Supplement). Seventy-five percent of eligible participants were female, the mean age was 55 years, and nearly half were either Hispanic/Latino (30%) or Black (15%; Table III in the Data Supplement). On average, participants were overweight (body mass index 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, PR segment and TP segment durations were the most strongly correlated among the 6 electrocardiographic traits (partial correlation ρ=−0.64), whereas T wave and P wave durations (ρ=−0.01) was largely uncorrelated (Table IV in the Data Supplement). P wave and QRS interval were the only 2 electrocardiographic traits with significant and positive genetic correlations (r_g=0.27; P=0.05; Table V in the Data Supplement).

Overview of Association Results

Approximately 22M single-nucleotide polymorphisms (SNPs) met our inclusion criteria (Table VI in the Data Supplement) and were evaluated in our combined phenotype multi-ethnic analysis of 6 contiguous

electrocardiographic traits, our primary analysis (Figures I and II in the Data Supplement). Lead SNPs at 82 of 149 loci (56%) previously reported by 26 interval scale electrocardiographic trait GWAS analyses (Table VII in the Data Supplement) were identified at genome-wide significance in our multi-ethnic population. The identification of known loci varied by trait (Table VIII in the Data Supplement), ranging from 21 of 45 (47%) loci for QRS interval to 9 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 electrocardiographic trait loci were identified.

An additional 6 loci identified by our primary analysis were >2 Mb away from all lead SNPs previously reported by interval scale electrocardiographic trait GWAS and are presented as novel (Table and Figure 2). 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.

Phenotype Decomposition

Of the 6 novel loci identified in our primary multi-ethnic combined phenotype analysis, accompanying univariate analyses indicated that lead SNPs primarily affected P wave (*PITX2* and *EMB*), TP segment (*CD36*), PR segment (*ZNF592*), T wave (*YPEL2*), and QRS interval (*BC043580*). None of the novel loci were associated with ST segment. Furthermore, the combined phenotype analysis did not identify novel loci beyond univariate analysis.

We then contrasted results for the 6 contiguous electrocardiographic traits with results from the 2 composite electrocardiographic traits, QT interval (QRS interval + ST segment + T wave) and PR interval (P wave + PR segment) (Figure III in the Data Supplement). *CD36* was the only novel locus identified for both a contiguous (TP segment) and a composite (QT interval) electrocardiographic trait (Table). We also examined evidence of consistency of SNP effects by grouping traits according to whether they affected atrial (PR interval, PR segment, and P wave) or ventricular (QT interval, QRS interval, T wave, and ST segment) conduction. For atrial traits, novel loci identified for the contiguous traits had varying directions of effects (Figure IIIA and Table IX in the Data Supplement), which when combined resulted in near-zero estimated effects for the composite trait. For example, every copy of the T allele for *PITX2* lead SNP rs13143308 increased P wave duration by 0.63 ms ($P_{univariate}=2\times10^{-11}$) but shortened the PR segment by 0.58 ms ($P_{univariate}=6\times10^{-4}$). However, when evaluated together as the composite trait PR interval, every copy of the rs13143308 T allele prolonged the PR interval by 0.03 ms ($P_{univariate}=0.84$). Similarly, among the 59 loci associated with ventricular conduction, 2 of the three novel loci (rs142166837 and rs13047360) had opposite effects on QRS interval and T wave duration, which did not reach genome-wide significance when summed for the composite trait QT interval (Figure IIIB and Table IX in the Data Supplement). There were no instances of either PR or QT interval identifying a novel locus not associated with any of the 6 contiguous traits at the genome-wide level.

Ancestral Diversity

Lead SNPs at 5 of the 6 novel loci were common (minor allele frequency [MAF]>5%) across ancestral populations, with modest evidence of heterogeneity of effect across race/ethnicity (Table X in the Data Supplement). One locus (*CD36*) showed evidence of population specificity, with lead SNP rs3211938 near monomorphic in European and Chinese populations (MAF<0.01%), infrequent in Hispanic/Latinos (MAF=1%), and common in Black (MAF=10%). Variant rs3211938 showed genome-wide significant associations with TP segment ($P_{univariate}=1\times10^{-13}$) and QT interval ($P_{univariate}=6\times10^{-10}$) and nominal associations with P wave ($P_{univariate}=2\times10^{-5}$), PR segment ($P_{univariate}=0.008$), and QRS interval ($P_{univariate}=0.005$). Although no GWAS of TP segment has been published, each copy of the rs3211938 G allele

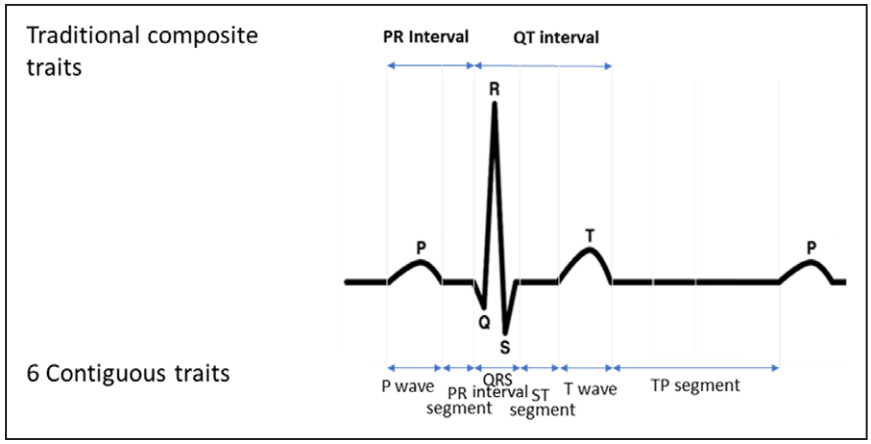

Figure 1. Illustration of the 6 contiguous (P wave, PR segment, QRS interval, ST segment, and TP segment) and 2 composite (QT interval and PR interval) ECG traits.

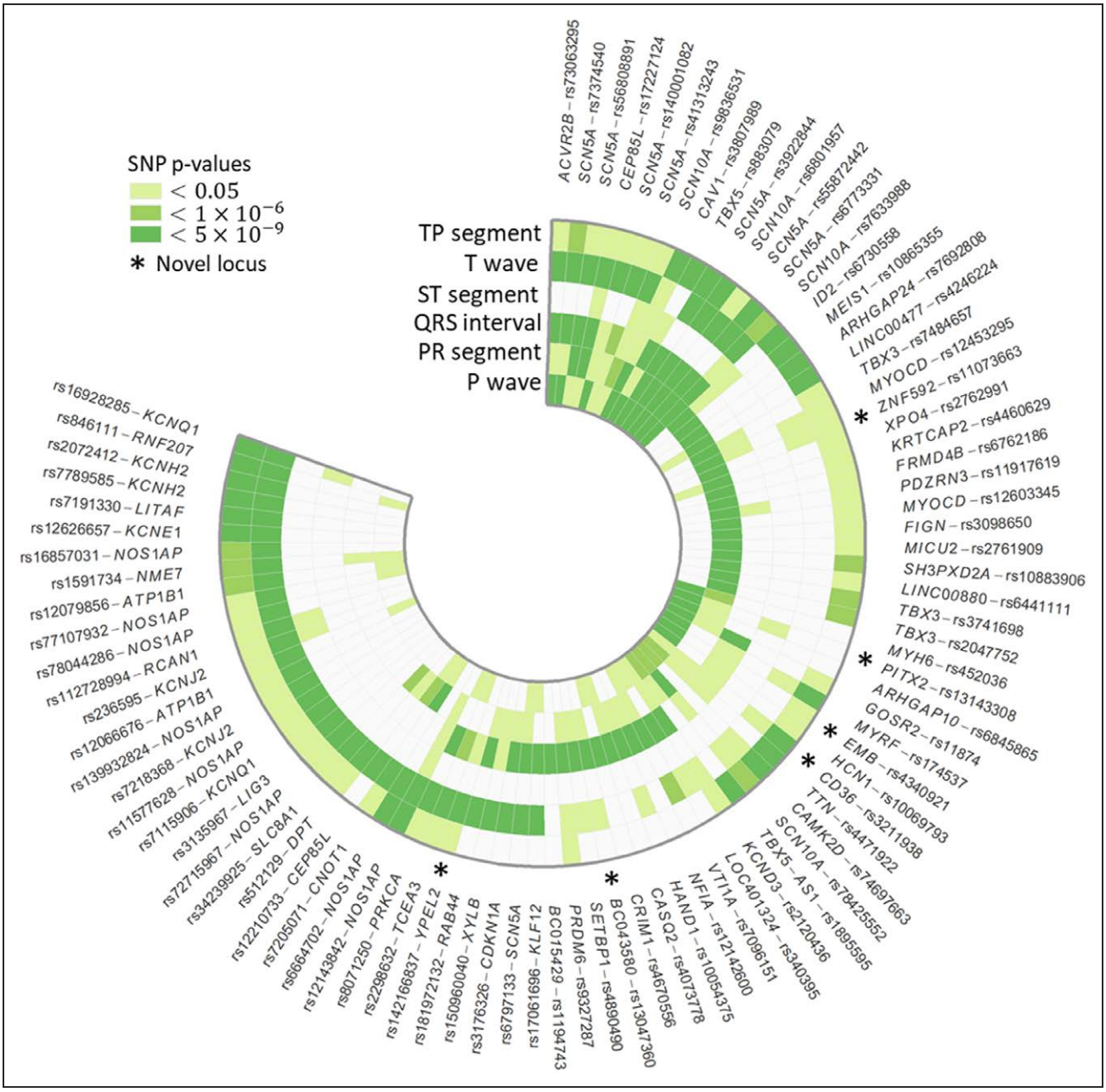

Figure 2. Lead single-nucleotide polymorphisms (SNPs) at 87 loci significantly associated (Paspu<5×10–9) with 6 contiguous ECG traits spanning an average heartbeat, in n=34 668 multi-ethnic participants in the PAGE study (Population Architecture Using Genomics and Epidemiology) and MESA (Multi-Ethnic 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.

increased QT interval by 3.70 ms. Reported effects for common (MAF>5%) QT lead SNPs range from 0.5 ms to 3.5 ms.[13] SNP rs3211938 was either genotyped or well-imputed across studies and ancestry groups (imputation quality >0.98, Table XI in the Data Supplement).

Combined Phenotype Analyses

We found widespread evidence of shared genetic effects across electrocardiographic traits, with adaptive sum of powered score test (aSPU) gamma scores that varied substantially across lead SNPs (Table IX in the Data Supplement). One-fourth of lead SNPs identified as genome-wide significant (P_{aspu}<5×10^{-9}) had univariate associations with at least 2 electrocardiographic traits ($P_{Univariate}$<5×10^{-9}). Lead SNPs at *ACVR2B, SCN5A, SCN10A, CEP85L, CAV1*, and *TBX5* were associated with 3 or more electrocardiographic traits at univariate genome-wide significance levels. As expected, traits that were more highly correlated also showed stronger

Table. Novel Genome Wide-Significant (P_{aSPU}<5×10−9) Loci Discovered in Genome-Wide Association Study of 6 Contiguous Electrocardiographic Traits That Decompose an Average Heartbeat in N=34 668 Participants From the Multi-Ethnic PAGE Study and the MESA

Lead SNP	Chr	Position	Coded Allele	Noncoded Allele	Locus	AA	EA	CHN	HIS	Contiguous ECG Traits						Composite ECG Traits	
										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%	2×10^{-11}*	5×10^{-4}	2×10^{-1}	3×10^{-2}	3×10^{-1}	1×10^{-1}	4×10^{-1}	8×10^{-1}
rs4340921	5	49687697	C	T	*EMB*	66%	46%	49%	44%	8×10^{-13}*	8×10^{-1}	1×10^{-3}	7×10^{-1}	2×10^{-1}	2×10^{-4}	7×10^{-1}	2×10^{-3}
rs3211938	7	80300449	G	T	*CD36*	10%	<0.01%	<0.01%	1%	2×10^{-5}	8×10^{-3}	5×10^{-3}	4×10^{-1}	1×10^{-5}	1×10^{-13}*	6×10^{-10}*	6×10^{-6}
rs11073663	15	85260268	A	G	*ZNF592*	27%	54%	19%	48%	4×10^{-1}	3×10^{-10}*	5×10^{-1}	4×10^{-1}	2×10^{-3}	7×10^{-3}	2×10^{-2}	6×10^{-7}
rs142166837	17	57471022	C	T	*YPEL2*	31%	52%	32%	49%	5×10^{-2}	6×10^{-1}	1×10^{-3}	1×10^{-1}	4×10^{-11}*	1×10^{-2}	4×10^{-7}	8×10^{-1}
rs13047360	21	28851580	G	A	*BC043580*	7%	17%	23%	16%	7×10^{-1}	3×10^{-2}	2×10^{-11}*	2×10^{-1}	5×10^{-2}	1×10^{0}	2×10^{-1}	3×10^{-2}

AA indicates African American; aSPU, adaptive sum of powered tests; CHN, Chinese-American; EA, European-American; HIS, Hispanic/Latinos; MESA, Multi-Ethnic Study of Atherosclerosis; and PAGE, Population Architectures Using Genomics and Epidemiology.
*Values exceed the genome-wide significance threshold ($P<5\times10^{-9}$).

evidence of shared genetic effects, with 10 of the 20 lead SNPs that were associated with PR segment also showing genome-wide associations with TP segment. However, evidence of shared genetic effects among uncorrelated traits was also observed. For example, 8 genome-wide significant SNPs at *SCN5A*, *SCN10A*, *CEP85L*, and *CNOT1* exhibited significant univariate associations with both the T wave and P wave, despite low correlation between the 2 traits (ρ=−0.01).

There also was evidence of allelic heterogeneity for multiple electrocardiographic traits. As an example, 5 signals in low linkage disequilibrium (r^2<0.1) were detected within a 500 kb region near the previously identified locus *TBX5*, each associated with a distinct combination of electrocardiographic traits. Lead variants at these 5 independent signals remained genome-wide significant after sequential conditional analyses (results not shown). Two of the 5 independent signals (rs3741698 and rs2047752) were largely specific to PR segment (Figure 3). The other 3 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 ≈3 orders of magnitude from the rs7484657 *P* value.

Bioinformatic Characterization

The rs3211938 variant is a well-known, nonsynonymous protein-coding variant causing CD36 deficiency.[14] For the 5 remaining novel SNPs, bioinformatics characterization found evidence of genetic regulation, including chromatin marks, and regulatory motifs (Table XII in the Data Supplement). Each of the novel lead SNPs were either rated as evolutionarily conserved based on the GRASP conservation score or were in high linkage disequilibrium with another SNP meeting that threshold (R^2 >0.9). In addition, rs1107366 was identified to regulate the expression of several long noncoding RNAs in electrocardiographic-relevant tissues (adipose, arterial, and atrial tissue).

DISCUSSION

In this study, 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 electrocardiographic trait loci. We identified 6 novel loci, 5 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

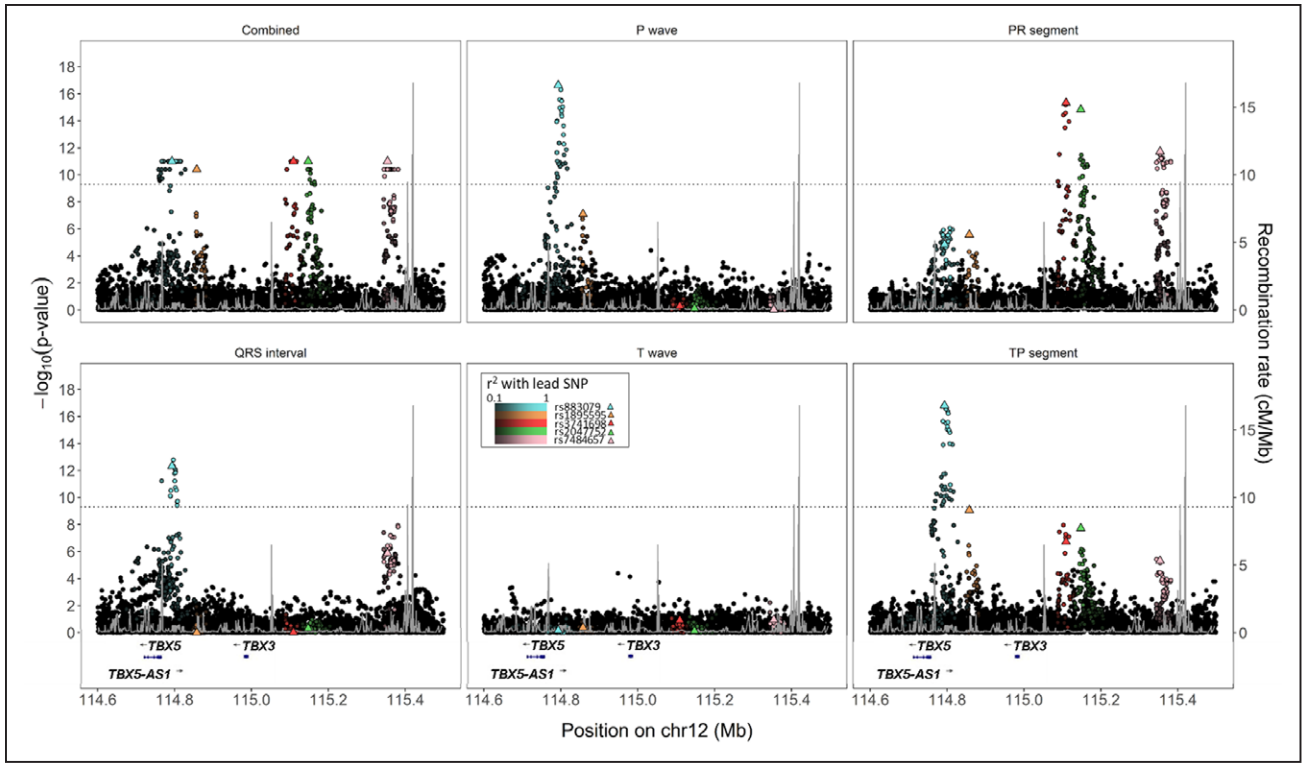

Figure 3. Regional single-nucleotide polymorphism (SNP) associations and linkage disequilibrium at 4 independent signals near TBX5 among 34 668 participants with electrocardiographic data in the PAGE study (Population Architectures Using Genomics and Epidemiology) and MESA (Multi-Ethnic Study of Atherosclerosis).

Lighter colors indicate greater linkage disequilibrium with lead SNPs, and black markers denote SNPs not in linkage disequilibrium (r^2<0.1) with any of the 4 lead SNPs. Combined phenotype *P* values are truncated at 1×10^{-11}.

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 electrocardiographic trait GWAS,[13,15] underscores the merits of prioritizing diversity and phenotype measurement.

Of the 3 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,[16] has resulted in the prioritization of commonly available traits (eg, body mass index) over traits that more precisely capture underlying biology (eg, direct measures of body fat[17]). In our case, composite electrocardiographic 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.[18] 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.[19]

The 6 traits we used in our electrocardiographic decomposition were motivated by their relations to physiology, and their coherence as an aggregate electrophysiological 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 electrocardiographic phenotypes. Further phenotypic specificity and additional biologic insight may be offered by GWAS of other electrocardiographic traits, including measures of waveform amplitudes, angles, or variability. For example, a recent, as yet unpublished of UK Biobank data used each of the sampled amplitudes recorded on the digital electrocardiographic, forming hundreds of distinct electrocardiographic measures for separate evaluation in GWAS (https://www.biorxiv.org/content/10.1101/648527v1). Another approach might focus on traits governed by a plausibly shared genetic

architecture, such as ion channel function or cardiac remodeling, potentially assisting efforts to map loci to specific biologic pathways. Further extending combined phenotype electrocardiographic trait GWAS to include other phenotypes and traits (eg, cardiometabolic traits or cardiovascular diseases) also is warranted, given evidence that these traits represent interrelated manifestations of common biologic mechanisms[12] and the success of prior combined phenotype studies to disentangle complex biology.[20] Overall, 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.

There has been mounting interest in combined phenotype statistical approaches; however, their merits for novel locus discovery and locus characterization remain largely untested in practice. Here, combined phenotype analysis of contiguous electrocardiographic traits did not identify novel loci that eluded traditional univariate analyses, despite the theoretical potential demonstrated for aSPU and related methods. Nonetheless, our evaluation of *TBX5*, a locus harboring multiple independent signals, suggested that combined phenotype approaches may be informative for fine-mapping. Supporting the use of combined phenotype methods to fine-mapping are methods that have been specifically developed for this challenge,[21] including fastPAINTOR. When compared with single trait fine-mapping, fastPAINTOR reduced the number of SNPs required for follow-up to capture 90% of the causal variants, from 23 SNPs per locus using a single trait to 12 SNPs when fine-mapping 2 traits simultaneously.

The lack of diversity in GWAS has long been described,[22] but the literature remains dominated by studies of 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.[23] 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,[24–31] the most intriguing finding is the potential linkage with sudden cardiac arrest, for which QT interval prolongation increases risk.[32] Sudden cardiac arrest accounts for ≈10% to 20% of total mortality in industrial countries,[33] and several decades of research have suggested a contributory role of impaired fatty acid uptake in cardiomyocytes.[14] Although genetic studies of *CD36* and sudden cardiac arrest were largely null,[34,35] 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 studies conducted in 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 electrocardiograms from which to derive the 6 contiguous electrocardiographic traits. However, this study is the largest multi-ethnic GWAS of electrocardiograms 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.[36,37] 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. Furthermore, while this study helps address the lack of diversity in electrocardiographic 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, in-depth fine-mapping was outside the scope of the proposed study, despite the value of multi-ethnic populations for fine-mapping.[38] We also were unable to leverage heterogeneity in allelic effects between ethnic groups to increase statistical power, as available methods are incompatible with the multi-phenotype methods presented herein.[39] Identification of allelic heterogeneity provides a clear impetus for future studies that leverage evidence of a shared genetic effects to further characterize the genetic architecture underlying electrocardiographic traits and complex traits in general.

This study illustrates 3 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.

ARTICLE INFORMATION

Received July 8, 2019; accepted May 26, 2020.

Affiliations

Gillings School of Global Public Health (A.R.B., H.M.H., R.G., M.G., C.J.H., A.A.S., E.A.W., K.E.N., C.L.A.) and Carolina Center for Genome Sciences (K.E.N.), University of North Carolina at Chapel Hill. Cardiovascular Health Research Unit, Department of Medicine (C.M.S.) and Cardiovascular Health Research Unit, Division of Cardiology, Department of Medicine (S.R.H., N.S.), University of Washington, Seattle. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD (D.E.A.). Department of Statistics and Biostatistics, Rutgers University, New Brunswick, NJ (S.B.). Department of Medicine, University of Illinois at Chicago (D.D.). Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance (X.G., Y.-D.I.C., J.I.R., K.D.T., J.Y., H.J.L.). Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA (X.G., Y.-D.I.C., J.I.R., K.D.T., J.Y., H.J.L.). Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD (L.A.H.). Albert Einstein College of Medicine, Bronx, NY (R.C.K.). Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA (U.P., A.P.R., C.K.). Departments of Medicine and Epidemiology, Johns Hopkins University, Baltimore, MD (W.P.). Center for Cardiovascular Research, John A. Burns School of Medicine, Honolulu, HI (R.V.S.). Department of Biostatistics, Vanderbilt University, Nashville, TN (R.T.). Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (G.L.W.). Center for Genomic Health (E.E.K.), Charles Bronfman Institute of Personalized Medicine (E.E.K.), Department of Genetics and Genomic Sciences (E.E.K.), and Department of Medicine (E.E.K.), Icahn School of Medicine at Mount Sinai, New York, NY. Epidemiological Cardiology Research Center (EPICARE), Wake Forest School of Medicine, Winston-Salem, NC (E.Z.S.).

Acknowledgments

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The PAGE (Population Architecture Using Genomics and Epidemiology) 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 data sets 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). We thank the researchers and research participants who made this dataset available to the community. We thank the Women's Health Initiative (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. We thank the staff and participants of the ARIC (Atherosclerosis Risk in Communities) 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, CA) and the Broad Institute of Harvard and MIT (Boston, MA) 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).

Sources of Funding

The PAGE (Population Architecture Using Genomics and Epidemiology) program is funded by the National Human Genome Research Institute with co-funding from the National Institute on Minority Health and Health Disparities. Assistance with data management, data integration, data dissemination, genotype imputation, ancestry deconvolution, population genetics, analysis pipelines, and general study coordination was provided by the PAGE Coordinating Center (NIHU01HG007419). Genotyping services were provided by the Center for Inherited Disease Research, which is fully funded through a federal contract from the National Institutes of Health (NIH) to The Johns Hopkins University, contract number HHSN268201200008I. Genotype data quality control and quality assurance services were provided by the Genetic Analysis Center in the Biostatistics Department of the University of Washington, through support provided by the Center for Inherited Disease Research contract. PAGE data and materials included in this report were funded through following studies and organizations: (1) HCHS/SOL (Hispanic Community Health Study/Study of Latinos): Primary funding support to Dr North and colleagues is provided by U01HG007416. Additional support was provided via R01DK101855 and 15GRNT25880008. The HCHS/SOL was performed as a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contribute to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution—Office of Dietary Supplements. (2) PAGE Global Reference Panel: The creation of the Stanford Global Reference Panel was funded by the National Human Genome Research Institute PAGE program (NIH U01HG007419). (3) Women's Health Initiative (WHI): Funding support for the Exonic variants and their relation to complex traits in minorities of the WHI study is provided through the National Human Genome Research Institute PAGE program (NIH U01HG007376). The WHI program is funded by the NHLBI, NIH, US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. (4) ARIC study (Atherosclerosis Risk in Communities): The ARIC study has been funded in whole or in part with Federal funds from the NHLBI, NIH, Department of Health and Human Services (contract numbers HHSN2682017000001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, and HHSN268201700005I), R01HL087641, R01HL086694; National Human Genome Research Institute contract U01HG004402; and NIH contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the NIH and NIH Roadmap for Medical Research. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are conducted and supported by the NHLBI in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN2682015000003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, CA) and the Broad Institute of Harvard and MIT (Boston, MA) using the Affymetrix Genome-Wide Human SNP Array 6.0. Drs Avery, Baldassari, and Sitlani were supported by R01HL142825; Drs Avery, Baldassari, Hodonsky, and Highland were supported by NIH grant U01HG007416; Dr Kenny is supported by NIH grants U01HG009080, U01HG009610, U01HG109391, and R01HL104608; Dr Hodonsky and Highland were supported by T32HL129982; Dr Baldassari was supported by NIH grant R01ES020836; Dr Sotoodehnia was supported by R01HL116747 and R01HL111089. Dr Darbar was supported by R01HL138737 and T32HL139439.

Disclosures

None.

REFERENCES

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet.* 2017;101:5–22. doi: 10.1016/j.ajhg.2017.06.005
2. LeBlanc M, Zuber V, Andreassen BK, Witoelar A, Zeng L, Bettella F, Wang Y, McEvoy LK, Thompson WK, Schork AJ, et al; CARDIoGRAM Consortium. Identifying novel gene variants in coronary artery disease and shared genes with several cardiovascular risk factors. *Circ Res.* 2016;118:83–94. doi: 10.1161/CIRCRESAHA.115.306629
3. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet.* 2018;50:390–400. doi: 10.1038/s41588-018-0047-6
4. Price AL, Spencer CCA, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. *Proc Biol Sci.* 2015;282:20151684. doi: 10.1098/rspb.2015.1684
5. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016;538:161–164. doi: 10.1038/538161a
6. Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, Haiman CA, Heiss G, Kooperberg C, Marchand LL, et al; PAGE Study. The next PAGE in understanding complex traits: design for the analysis of

population architecture using genetics and epidemiology (PAGE) study. *Am J Epidemiol.* 2011;174:849–859. doi: 10.1093/aje/kwr160
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.* 2016;283:20160569. doi: 10.1098/rspb.2016.0569
8. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet.* 2018;19:110–124. doi: 10.1038/nrg.2017.101
9. Silva CT, Kors JA, Amin N, Dehghan A, Witteman JC, Willemsen R, Oostra BA, van Duijn CM, Isaacs A. Heritabilities, proportions of heritabilities explained by GWAS findings, and implications of cross-phenotype effects on PR interval. *Hum Genet.* 2015;134:1211–1219. doi: 10.1007/s00439-015-1595-9
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;70:61134–61135.
11. MacRae CA, Vasan RS. Next-generation genome-wide association studies: time to focus on phenotype? *Circ Cardiovasc Genet.* 2011;4:334–336. doi: 10.1161/CIRCGENETICS.111.960765
12. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2017;45(D1):D896–D901. doi: 10.1093/nar/gkw1133
13. Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, Sotoodehnia N, Rossin EJ, Morley M, Wang X, et al; CARe Consortium; COGENT Consortium; DCCT/EDIC; eMERGE Consortium; HRGEN Consortium. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet.* 2014;46:826–836. doi: 10.1038/ng.3014
14. Love-Gregory L, Sherva R, Schappe T, Qi JS, McCrea J, Klein S, Connelly MA, Abumrad NA. Common CD36 SNPs reduce protein expression and may contribute to a protective atherogenic profile. *Hum Mol Genet.* 2011;20:193–201. doi: 10.1093/hmg/ddq449
15. van Setten J, Brody JA, Jamshidi Y, Swenson BR, Butler AM, Campbell H, Del Greco FM, Evans DS, Gibson Q, Gudbjartsson DF, et al. PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat Commun.* 2018;9:2904. doi: 10.1038/s41467-018-04766-9
16. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM; GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet.* 2018;27:3641–3649. doi: 10.1093/hmg/ddy271
17. Blundell JE, Dulloo AG, Salvador J, Frühbeck G; EASO SAB Working Group on BMI. Beyond BMI–phenotyping the obesities. *Obes Facts.* 2014;7:322–328. doi: 10.1159/000368783
18. Syeda F, Kirchhof P, Fabritz L. PITX2-dependent gene regulation in atrial fibrillation and rhythm control. *J Physiol.* 2017;595:4019–4026. doi: 10.1113/JP273123
19. 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. *Genet Med.* 2017;19:743–750. doi: 10.1038/gim.2016.183
20. VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ, Han Y, Spitz MR, Shete S, Wu X, Gaborieau V, Wang Y, McLaughlin J, et al. Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. *Am J Epidemiol.* 2012;175:1013–1020. doi: 10.1093/aje/kwr467
21. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, Pasaniuc B. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 2014;10:e1004722. doi: 10.1371/journal.pgen.1004722
22. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature.* 2011;475:163–165. doi: 10.1038/475163a
23. 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.* US Census Bureau, Washington, D.C.; 2015.
24. Shibao CA, Celedonio JE, Ramirez CE, Love-Gregory L, Arnold AC, Choi L, Okamoto LE, Gamboa A, Biaggioni I, Abumrad NN, et al. A common CD36 variant influences endothelial function and response to treatment with phosphodiesterase 5 inhibition. *J Clin Endocrinol Metab.* 2016;101:2751–2758. doi: 10.1210/jc.2016-1294
25. Avery CL, He Q, North KE, Ambite JL, Boerwinkle E, Fornage M, Hindorff LA, Kooperberg C, Meigs JB, Pankow JS, et al. A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains. *PLoS Genet.* 2011;7:e1002322. doi: 10.1371/journal.pgen.1002322
26. Gautam S, Agrawal CG, Banerjee M. CD36 gene variants in early prediction of type 2 diabetes mellitus. *Genet Test Mol Biomarkers.* 2015;19:144–149. doi: 10.1089/gtmb.2014.0265
27. Love-Gregory L, Abumrad NA. CD36 genetics and the metabolic complications of obesity. *Curr Opin Clin Nutr Metab Care.* 2011;14:527–534. doi: 10.1097/MCO.0b013e32834bbac9
28. Elbers CC, Guo Y, Tragante V, van Iperen EP, Lanktree MB, Castillo BA, Chen F, Yanek LR, Wojczynski MK, Li YR, et al. Gene-centric meta-analysis of lipid traits in African, East Asian and Hispanic populations. *PLoS One.* 2012;7:e50198. doi: 10.1371/journal.pone.0050198
29. Musunuru K, Romaine SP, Lettre G, Wilson JG, Volcik KA, Tsai MY, Taylor HA Jr, Schreiner PJ, Rotter JI, Rich SS, et al. Multi-ethnic analysis of lipid-associated loci: the NHLBI CARe project. *PLoS One.* 2012;7:e36473. doi: 10.1371/journal.pone.0036473
30. Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, Keating BJ, Durda P, Fox ER, Palmer CD, 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;133:985–995. doi: 10.1007/s00439-014-1439-z
31. Love-Gregory L, Sherva R, Sun L, Wasson J, Schappe T, Doria A, Rao DC, Hunt SC, Klein S, Neuman RJ, et al. Variants in the CD36 gene associate with the metabolic syndrome and high-density lipoprotein cholesterol. *Hum Mol Genet.* 2008;17:1695–1704. doi: 10.1093/hmg/ddn060
32. O'Neal WT, Singleton MJ, Roberts JD, Tereshchenko LG, Sotoodehnia N, Chen LY, Marcus GM, Soliman EZ. Association between QT-interval components and sudden cardiac death: the ARIC study (Atherosclerosis Risk in Communities). *Circ Arrhythm Electrophysiol.* 2017;10:e005485. doi: 10.1161/CIRCEP.117.005485
33. Chugh SS. Sudden cardiac death in 2017: spotlight on prediction and prevention. *Int J Cardiol.* 2017;237:2–5. doi: 10.1016/j.ijcard.2017.03.086
34. Lemaitre RN, Johnson CO, Hesselson S, Sotoodehnia N, Sotoodhenia N, McKnight B, Sitlani CM, Rea TD, King IB, Kwok PY, et al. Common variation in fatty acid metabolic genes and risk of incident sudden cardiac arrest. *Heart Rhythm.* 2014;11:471–477. doi: 10.1016/j.hrthm.2014.01.008
35. Johnson CO, Lemaitre RN, Fahrenbruch CE, Hesselson S, Sotoodehnia N, McKnight B, Rice KM, Kwok PY, Siscovick DS, Rea TD. Common variation in fatty acid genes and resuscitation from sudden cardiac arrest. *Circ Cardiovasc Genet.* 2012;5:422–429. doi: 10.1161/CIRCGENETICS.111.961912
36. van de Bunt M, Cortes A, Brown MA, Morris AP, McCarthy MI; IGAS Consortium. Evaluating the performance of fine-mapping strategies at common variant GWAS Loci. *PLoS Genet.* 2015;11:e1005535. doi: 10.1371/journal.pgen.1005535
37. Mensah-Ablorh A, Lindstrom S, Haiman CA, Henderson BE, Marchand LL, Lee S, Stram DO, Eliassen AH, Price A, Kraft P. Meta-analysis of rare variant association tests in multiethnic populations. *Genet Epidemiol.* 2016;40:57–65. doi: 10.1002/gepi.21939
38. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature.* 2019;570:514–518.
39. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol.* 2011;35:809–822. doi: 10.1002/gepi.20630