Original Article

Investigating the Genetic Architecture of the PR Interval Using Clinical Phenotypes

Jonathan D. Mosley, MD, PhD; M. Benjamin Shoemaker, MD, MSCI; Quinn S. Wells, MD, PhamD, MSc; Dawood Darbar, MD, PhD; Christian M. Shaffer, BS; Todd L. Edwards, PhD; Lisa Bastarache, MS; Catherine A. McCarty, PhD, MPH; Will Thompson, PhD; Christopher G. Chute, MD, DrPH; Gail P. Jarvik, MD, PhD; David R. Crosslin, PhD; Eric B. Larson, MD, MPH; Iftikhar J. Kullo, MD; Jennifer A. Pacheco, BA; Peggy L. Peissig, PhD, MBA; Murray H. Brilliant, PhD; James G. Linneman, BA; John S. Witte, PhD; Josh C. Denny, MD, MS; Dan M. Roden, MD

Background—One potential use for the PR interval is as a biomarker of disease risk. We hypothesized that quantifying the shared genetic architectures of the PR interval and a set of clinical phenotypes would identify genetic mechanisms contributing to PR variability and identify diseases associated with a genetic predictor of PR variability.

Methods and Results—We used ECG measurements from the ARIC study (Atherosclerosis Risk in Communities; n=6731 subjects) and 63 genetically modulated diseases from the eMERGE network (Electronic Medical Records and Genomics; n=12 978). We measured pairwise genetic correlations (rG) between PR phenotypes (PR interval, PR segment, P-wave duration) and each of the 63 phenotypes. The PR segment was genetically correlated with atrial fibrillation (rG=−0.88; *P*=0.0009). An analysis of metabolic phenotypes in ARIC also showed that the P wave was genetically correlated with waist circumference (rG=0.47; *P*=0.02). A genetically predicted PR interval phenotype based on 645 714 single-nucleotide polymorphisms was associated with atrial fibrillation (odds ratio=0.89 per SD change; 95% confidence interval, 0.83–0.95; *P*=0.0006). The differing pattern of associations among the PR phenotypes is consistent with analyses that show that the genetic correlation between the P wave and PR segment was not significantly different from 0 (rG=−0.03 [0.16]).

Conclusions—The genetic architecture of the PR interval comprises modulators of atrial fibrillation risk and obesity.

(*Circ Cardiovasc Genet*. 2017;10:e001482. DOI: 10.1161/CIRCGENETICS.116.001482.)

Key Words: atrial fibrillation ■ biomarker ■ cardiac electrophysiology ■ molecular epidemiology ■ PR interval ■ risk factors

The PR interval is an electrophysiological parameter derived from a cardiac ECG and measures the duration of conduction through the atrium and atrioventricular node. The PR interval comprises 2 components: the P wave, which primarily measures atrial conduction, and the PR segment, which primarily reflects atrioventricular nodal conduction. One potential use for the PR interval is as a biomarker for future disease risk. For instance, a prolonged PR interval is associated with an increased risk for atrial fibrillation (AF).[1,2] If such associations are driven by heritable variation affecting both phenotypes, then a risk classifier based on genetic factors modulating the PR interval could be used to identify individuals at high risk for AF. Because heritable genetic risk is determined at birth, genetic classifiers can be evaluated at early time points, thereby enhancing early prevention and risk stratification strategies.

See Clinical Perspective

To date, a relatively small number of single-nucleotide polymorphisms (SNPs) associated with the PR interval has been identified by genome-wide association studies (GWAS),[3–5] and these SNPs only account for a small portion of the underlying genetic variability. Hence, building and evaluating a robust genetic classifier for the PR interval based on known SNPs is

Received April 20, 2016; accepted March 3, 2017.

From the Department of Medicine (J.D.M., M.B.S., Q.S.W., C.M.S., J.C.D., D.M.R.), Vanderbilt Epidemiology Center (T.L.E.), Department of Biomedical Informatics (L.B., J.C.D., D.M.R.), Department of Pharmacology (D.M.R.), Vanderbilt University, Nashville, TN; Division of Cardiology, University of Illinois at Chicago (D.D.); Essentia Institute of Rural Health, Duluth, MN (C.A.M.); Center for Biomedical Research Informatics, NorthShore University Health System, Evanston, IL (W.T.); School of Medicine (C.G.C.), School of Public Health (C.G.C.), and School of Nursing (C.G.C.), Johns Hopkins University, Baltimore, MD; Division of Medical Genetics, Department of Medicine (G.P.J.), Department of Genome Sciences (G.P.J.), Department of Biomedical Informatics (D.R.C.), Department of Medical Education (D.R.C.), University of Washington; Group Health Research Institute, Seattle, WA (E.B.L.); Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN (I.J.K.); Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL (J.A.P.); Biomedical Informatics Research Center (P.L.P.), Center for Human Genetics (M.H.B., J.G.L.), Marshfield Clinic Research Foundation, WI; and Department of Epidemiology and Biostatistics, University of California, San Francisco (J.S.W.).

The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.116.001482/-/DC1.

Correspondence to Jonathan D. Mosley, MD, PhD, Department of Medicine, Vanderbilt University School of Medicine, 1285 Medical Research Bldg IV, Nashville, TN 37232. E-mail jonathan.d.mosley@vanderbilt.edu

not feasible. Newer genetic approaches, such as those based on generalized linear mixed models (GLMM) that measure the contribution of large numbers of SNPs to a phenotype, can circumvent this limitation.[6–8] Furthermore, these methods can also identify genetically related phenotypes across data sets by measuring genetic correlations based on additive genetics between pairs of phenotypes.[8–10] A phenotype that is genetically correlated (ie, has a nonzero genetic correlation) with the PR interval likely shares common physiological mechanisms and, potentially, can be predicted by a PR interval–based genetic classifier.

We used mixed modeling approaches to probe the additive genetic architecture of the PR interval based on the extent to which its architecture was shared by a set of clinically recognized diseases. This approach identifies both clinical diagnoses genetically related to the PR interval and genetically mediated disease mechanisms underlying PR variability. Specifically, we used a discovery-oriented approach whereby we measured genetic correlations between PR interval phenotypes and a collection of clinical phenotypes. To ensure that associations are attributable to shared genetic risk factors and not environmental factors, we tested associations across populations: PR interval phenotypes were from the prospectively studied ARIC cohort (Atherosclerosis Risk in Communities),[11] and clinical phenotypes were from the eMERGE network (Electronic Medical Records and Genomics), a consortium of medical centers with observational electronic health record (EHR)–linked DNA biobank data sets.[12] We show distinct patterns of genetic disease associations among the PR phenotypes and that PR interval variability is driven by genetic factors associated with electrophysiological and metabolic phenotypes.

Materials and Methods

An overview of the analyses is presented in Figure 1.

Study Populations

Analysis Data Sets

ARIC: The ARIC population was derived from 13 113 genotyped adult subjects and comprised 6732 unrelated European ancestry (EA) subjects with normal ECGs.[11] Genetic and phenotypic data were downloaded from dbGaP (phs000280.v3.p1). eMERGE: The eMERGE population comprised 12 978 unrelated EA adult subjects collected by the eMERGE phase I network (Vanderbilt University, Marshfield Clinic, Northwestern University, Mayo Clinic, and Group Health Research Institute), a consortium of medical centers using EHRs as a tool for genomic research.[13] Genetic data for the eMERGE network is available through dbGaP (phs000360.v2.p1).

Replication Data Sets

BioVU AF registry: The Vanderbilt Lone AF registry data set comprised 1690 EA patients between 18 and 65 years of age enrolled through Vanderbilt's inpatient and outpatient services, as previously described, and had 1022 AF cases and 668 control subjects.[14] Of the cases, 220 have lone AF, 444 have paroxysmal AF, and 259 have persistent AF. BioVU Vanderbilt Electronic Systems for Pharmacogenomic Assessment (VESPA) data set: The BioVU VESPA study population comprised 1206 AF adult cases and 2405 controls from Vanderbilt University's collection of genotyped patients.[15,16]

All data sets were predominantly composed of self-reported whites, so only EA subjects were evaluated, defined using STRUCTURE[17] in conjunction with ancestry informative markers, with EA defined as >80% (eMERGE subjects) or >90% (all other data sets) probability of being in the HapMap CEU cluster.

Genetic Data

ARIC: Genotype data were acquired on the Affymetrix 6.0 SNP array. Quality control steps for the ARIC data set followed the guidelines accompanying the dbGaP release and included removing SNPs with preidentified chromosomal anomalies, with >5 discordant calls in replicate samples, and used a subset of unrelated subjects identified by the ARIC study. eMERGE: SNP genotype data were acquired on the Illumina Human660W-Quadv1_A. BioVU AF data set: Subjects were genotyped on the Illumina 610-quad Beadchip.[14] BioVU VESPA: Subjects were genotyped on the Illumina HumanOmni1-Quad and HumanOmni5-Quad platforms. Quality control steps for the eMERGE and BioVU data sets used established protocols[18] including filtering for a sample missingness rate <2.0%, a SNP missingness rate <2.0%, and a SNP deviation from Hardy–Weinberg <0.001.

All data sets were imputed to the October 2014 release of the 1000 Genomes cosmopolitan reference haplotypes. SNPs were prephased using SHAPEIT[19] and imputed using IMPUTE2.[20] The genetic correlation analyses used an intersection of the unimputed ARIC and imputed eMERGE data set and contained 503 404 SNPs with an minor allele frequency (MAF) >1.0%. The Bayesian sparse linear mixed modeling (BSLMM) analyses used an linkage-disequilibrium reduced (r^2=0.9) set of SNPs with an MAF >1.0% present on all platforms (n=645 714 SNPs).

Phenotype Data

The clinical phenotypes for the eMERGE and BioVU data sets were based on PheWAS Phecodes, which are collections of related *International Classification of Disease, Ninth revision, Clinical Modification* diagnosis codes.[21–24] There are >1600 defined Phecodes, described at http://PheWAScatalog.org. For each Phecode, cases are subjects with ≥2 instances of the phenotype appearing in their medical record on 2 separate dates.[23] Controls with no instances of the phenotype were randomly selected. There were 315 phenotypes with >500 cases in the eMERGE data set. AF cases and controls were based on PheWAS code 427.21 (atrial fibrillation), which has been previously used in other genetic studies.[22,23,25]

ARIC phenotypes came from the GENEVA substudy (pht000114.v2.p1) and from ECG measurements taken at visit 1 (pht004071.v1.p1). Subjects with a baseline ECG diagnosis of AF, atrioventricular block other than first degree, Wolff–Parkinson–White syndrome, a nonsinus rhythm, or a pacemaker were excluded. Subjects on atrioventricular nodal blocking drugs were also excluded. The PR interval was extracted from the ECG. The P-wave duration was based on lead aVR, and the PR segment duration was calculated as the difference between the PR segment and P-wave duration. Subjects with a PR interval ≤80 or ≥320 ms were excluded, as were subjects with a P-wave duration ≤50 or ≥140 ms. Phenotype definitions for metabolic phenotypes were based on previously described thresholds for the ARIC data set[26]: elevated waist circumference (≥102 cm [men] or ≥88 cm [women]); insulin resistance (fasting glucose ≥100 mg/dL or history of diabetes mellitus); hypertension (systolic blood pressure ≥130 mm Hg, diastolic blood pressure ≥85 mm Hg, or use of antihypertensive medications); elevated triglycerides (≥150 mg/dL or use of medications for elevated lipids); low high-density lipoprotein cholesterol (<40 mg/dL [men] or <50 mg/dL [women] or use of medications for elevated lipids); and metabolic syndrome (≥3 abnormal metabolic components). Subjects who were not fasting for >8 hours at their first visit were excluded from the analysis of metabolic phenotypes (n=150).

Analyses

Linear mixed models and GLMM estimate the additive genetic variance and liability, respectively, attributable to a collection of common SNPs among unrelated individuals by modeling the genetic similarity between pairs of individuals as random effects.[6,7,27] The linear mixed model is expressed as:

$$y = X\beta + g_G + \varepsilon \text{ and } \mathrm{var}(g_G) = A_G\sigma_G^2$$

where y is phenotype vector, X is a vector of fixed effects (covariates and principal components [PCs]), and ε is a vector of errors. The term

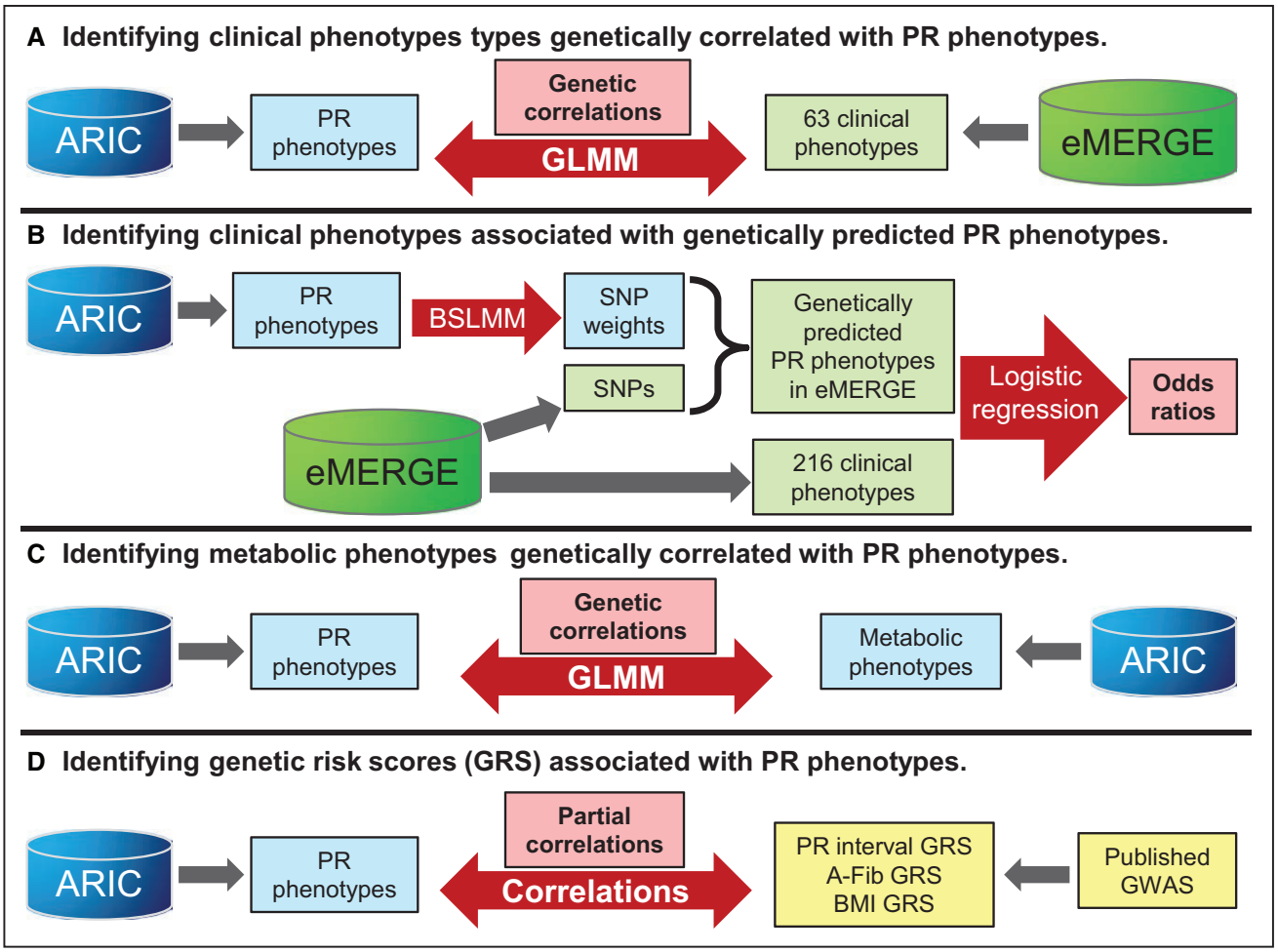


Figure 1. Overview of approach. **A**, Generalized linear mixed models (GLMMs) were used to measure the pairwise genetic correlations (rG) between each ARIC (Atherosclerosis Risk in Communities) PR phenotype (PR interval, PR segment, and P-wave duration) and each genetic eMERGE (Electronic Medical Records and Genomics) phenotype. **B**, Bayesian sparse linear mixed modeling (BSLMM) analyses were used to compute single-nucleotide polymorphism (SNP) weights for each PR phenotype. These weights were used to compute genetically predicted PR phenotype values in eMERGE subjects. Logistic regression was used to test the association between the predicted PR phenotypes and eMERGE clinical phenotypes. **C**, GLMMs were used to compute rG between ARIC PR phenotypes and metabolic phenotypes. **D**, Partial correlation coefficients were calculated between ARIC PR phenotypes and genetic risk scores based on SNPs identified by previous genome-wide association studies (GWAS). A-Fib indicates atrial fibrillation; and BMI, body mass index.

g_G is a vector of random polygenic effects, and A_G is often referred to as the genetic relationship matrix, with each element in the matrix defined by the equation $(1/N)\ \Sigma^N i=1\left(x_{ij}-2p_i\right)\left(x_{ik}-2p_i\right)/2p_i(1-p_i)$, where N in the number of SNPs analyzed, x is the genotype at that SNP (coded 0, 1 or 2) for individuals j and k, and p is the allele frequency. The variance components are estimated by a restricted maximum likelihood algorithm. These analyses used LMMs and GLMMs as implemented in the Genome-wide Complex Trait Analysis program.[6,7,9,27,28] To ensure only unrelated subjects are analyzed, subjects with a genetic relatedness score >0.05 were excluded. Genetic liability estimates, adjusting for birth decade, sex, and 20 PCs, were computed for each eMERGE PheWAS phenotype[29] with >500 cases, and phenotypes with a genetic liability estimate P<0.05 (n=63) were used for the exploratory genetic correlation analyses (Table I in the Data Supplement).

A bivariate extension of the GLMM was used to undertake the exploratory genetic correlation analyses (Table I in the Data Supplement). Here, y is now composed of pairs of phenotype vectors. For each pair of traits (t1 and t2), the bivariate GLMM estimates the genetic variance (σ^2_G) for the phenotypes and the genetic covariance between the phenotypes cov_g(G_t1,G_t2).[9,10] The genetic covariance is a measure of how much pairs of traits change together based on the additive genetic effects from common SNPs. This model is most commonly applied to data from 2 different nonoverlapping samples, where the trait (y) values are simply set to missing when not observed (eg, for subjects in the study of trait t1, their t2 values are set to missing).[9,10] The genetic correlation between pairs of traits is then defined as $r_G = \text{cov}_g\left(G\_t1, G\_t2\right)\ /\ \text{sqrt}\left[\left(\sigma^2_{Gt1}\right)\left(\sigma^2_{Gt2}\right)\right]$. This genetic correlation is a measure of the extent to which the additive genetic effects estimated from common SNPs are shared between a pair of traits. r_G is computationally analogous to a Pearson correlation coefficient and has a value of −1 to +1. Genetic correlations were computed between the ARIC PR phenotypes (PR interval, the P wave, and PR segment) and each eMERGE PheWAS phenotype (n=63), adjusting for age, sex, and 20 PCs. P values for genetic correlations were determined using a likelihood ratio test comparing the bivariate GLMM to a model where the genetic correlation was fixed at 0. Although SEs are given for rG point estimates, the 95% confidence intervals (CIs) surrounding these estimates under the assumption of asymptotic normality may fall outside the range of plausible values for rG. False discovery rate (FDR)–adjusted P values (Q values) were determined using a Benjamini–Hochberg adjustment. Although not all PheWAS phenotype pairs are independent, the test statistics meet Benjamini–Hochberg criteria by the positive regression–dependent criterion.[30]

BSLMM was used to compute genetically predicted levels of PR phenotypes in the eMERGE and BioVU data sets. BSLMM

uses a hybrid of GLMM and sparse regression models.[31] In general, this method estimates the proportion of variance explained by a set of SNPs and the distribution of effect sizes for the SNPs and then jointly models the contribution of all SNPs to the phenotypic variance. The posterior SNP weights generated by this approach can be used in conjunction with SNP genotypes to compute a genetically predicted value for a phenotype. Each PR phenotype in the ARIC data set was first adjusted for age, sex, and 3 PCs using linear regression. BSLMM was then used to generate SNP effect sizes (α and β) for the PR phenotype residuals. These effect sizes were then used to compute the genetically predicted value for a PR phenotype for an individual in the eMERGE and BioVU data sets using the equation:

$$\text{Predicted phenotype} = \sum(\alpha_i + \beta_i\gamma_i) * \left(\text{number of reference alleles for SNP}_i\right)$$

where α is the small SNP effect, βγ is the large SNP effect.

Multivariable logistic regression adjusting for 3 PCs, age, and sex was used to test the association between the predicted phenotype levels and the EMR PheWAS and AF phenotypes. The predicted phenotypes were set to have an SD of 1, so odds ratios (ORs) reflect risk per SD increase in the predicted phenotype. An FDR-adjusted Q value <0.1 was considered significant.

Genetic risk scores (GRS) based on either OR or β coefficients for previously reported SNPs reaching genome-wide significance ($P<5\times10^{-8}$) were computed for the PR interval (n=9 SNPs), body mass index (BMI; n=98 SNPs), and AF (n=10 SNPs).[5,14,32] The SNPs used to compute the GRS in each data set are shown in Tables II and III in the Data Supplement. A GRS based on previously the published GWAS association statistics was computed for each individual using the formulas:

$$\text{PR interval and BMI}: \text{score} = \sum \beta_i * \left(\#\text{ of ref. alleles for SNP}_i\right)$$

$$\text{AF}: \text{score} = \sum \log\left(\text{OR}_i\right) * \left(\#\text{ of ref. alleles for SNP}_i\right)$$

Only 8 of 9 SNPs for the PR GRS passed QC protocols and were used in the calculations. To ascertain whether the GRS are differentially associated with the PR phenotypes, partial correlation coefficients between each PR phenotype and each GRS were computed using PROC CORR (SAS) and adjusted for age and sex.

All quality control analyses and SNP association analyses were performed using PLINK v1.07.[34] Genetic liability and correlation estimates were computed using the Genome-wide Complex Trait Analysis v1.24.[27] BSLMM is part of the GEMMA v0.94.1 program package.[31] All other analyses were performed using SAS v9.3 (SAS Institute, Cary, NC).

Ethics Statement

The eMERGE study has been approved by the institutional review board at each site.[12,15] Vanderbilt's BioVU resource operates as non-human subjects research according to the provisions of 45 Code of Federal Regulations, part 46, with oversight by Vanderbilt's Institutional Review Board, as previously described.[15] Institutional review board approval for the current study was obtained through Vanderbilt's Institutional Review Board.

Results

The ARIC population comprised 6731 unrelated EA subjects with a normal ECG. Their median age was 54 years, and 45% of subjects were men (Table IV in the Data Supplement). Almost a quarter of subjects had ≥3 metabolic syndrome phenotypes. The eMERGE data set comprised 12 978 subjects, of which 48% were men, with an average of 44 clinical diagnoses per subject (Table I in the Data Supplement).

Clinical Phenotypes Genetically Correlated With PR Phenotypes

The estimated heritability explained by the SNPs for the PR interval in the ARIC data set was 0.23 (SE 0.05; Table 1). We measured the genetic correlation (rG) between the PR interval and 63 eMERGE phenotypes (listed in Table I in the Data Supplement). The strongest genetic correlations were with AF/atrial flutter (rG=–0.59; P=0.02) and AF (rG=–0.57; P=0.02) but were not significant (FDR Q>0.1) after adjusting for multiple testing (Figure 2A; Table 2; characteristics of the AF cases and controls are shown in Table V in the Data Supplement).

We next examined the P wave and the PR segment durations (Table 1), which comprise the PR interval. The point estimate of the genetic correlation between the PR segment and the PR interval (rG=0.89 [0.04]) was larger than that for the P wave and the PR interval (rG=0.49 [0.16]). The genetic correlation between the P wave and the PR segment was not significantly different from 0 (rG=–0.03 [0.16]). The PR segment showed a similar pattern of genetic correlations with the eMERGE phenotypes as the PR interval, with the exception that the genetic correlation with AF was significant after multiple testing correction (rG=–0.88; 95% CI, –1.6 to –0.19; P=0.0009; FDR Q=0.047; Table 2; Figure 2B). For both the PR interval and PR segment, the AF correlation was negative indicating that genetic factors associated with a longer interval are associated with a decreased risk of AF. There were no significant genetic correlations with the P wave (Figure 2C). The most strongly genetically correlated phenotype was type 2 diabetes mellitus (rG=0.49; P=0.008; FDR Q=0.26).

We examined the impact of adjusting for PR phenotypes on the genetic correlation between the PR interval duration and AF. Adjusting for the P-wave duration minimally impacted the genetic correlation between the PR interval and AF (rG=–0.84; P=0.001; Table 2). In contrast, adjusting for the PR segment further attenuated the P-wave–AF correlation

Table 1. Heritability/Liability Estimates for the ARIC Cohort

Characteristic	Heritability/Liability (SE)*
EKG parameters (mean (SD)	
PR interval duration, ms)	0.23 (0.05)
P-wave duration, ms	0.19 (0.05)
PR segment duration, ms	0.18 (0.05)
Metabolic trait, n (%)†	
Waist circumference	0.14 (0.04)
Insulin resistance	0.05 (0.04)
Hypertension	0.18 (0.04)
Triglycerides	0.15 (0.04)
HDL cholesterol	0.14 (0.04)
Metabolic syndrome	0.21 (0.04)

ARIC indicates Atherosclerosis Risk in Communities; and HDL, high-density lipoprotein.

*Heritability or liability estimates or metabolic traits and ECG phenotypes are based on genetic linear mixed models adjusting for age, sex, and 20 principal components.

†See Materials and Methods for metabolic trait definitions.

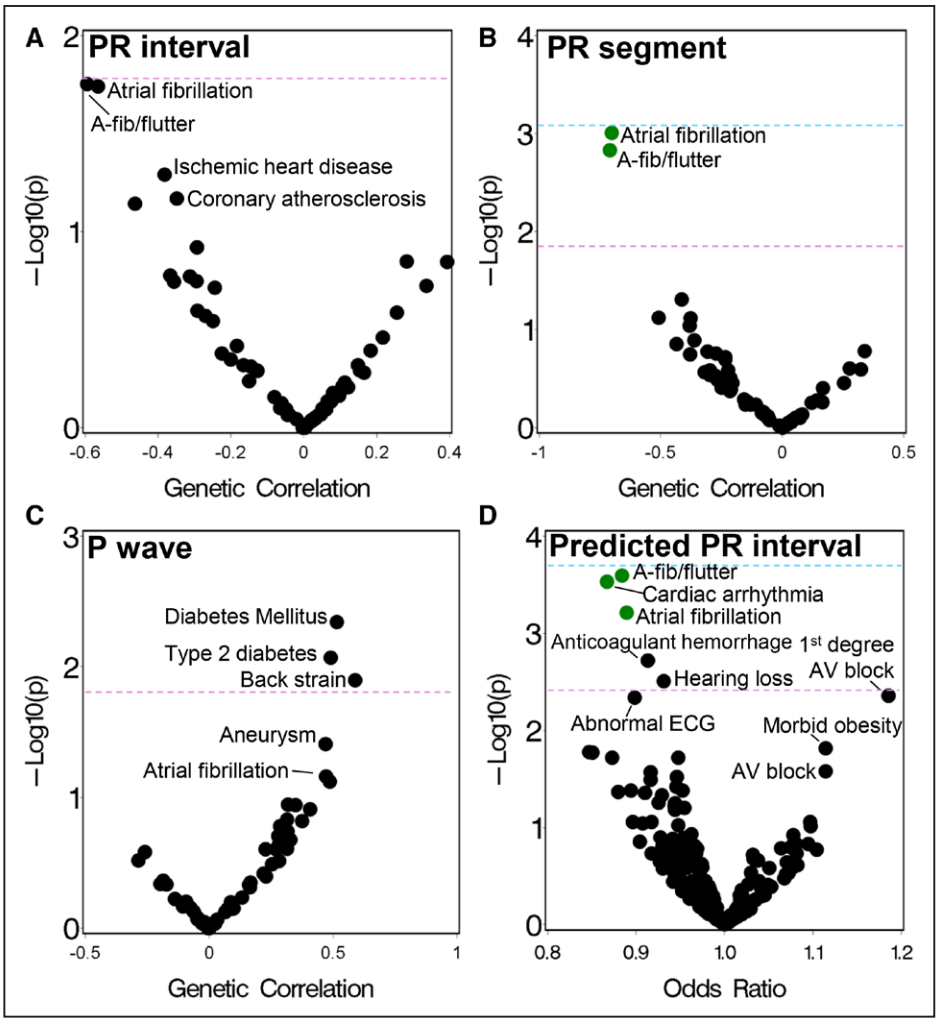


Figure 2. Phenotypes genetically associated with PR phenotypes. **A–C**, Genetic correlations between the (**A**) PR interval, (**B**) PR segment, and (**C**) P-wave duration, measured in ARIC (Atherosclerosis Risk in Communities) subjects and 63 phenotypes measured in eMERGE (Electronic Medical Records and Genomics) subjects, adjusted for age, sex and principal components (PCs). *P* values are based on a likelihood ratio test comparing a full model to a model where the genetic correlation was fixed at 0. **D**, Plot of *P* value vs odds ratio (OR) from logistic regression analyses of the association between a genetically predicted PR interval and 261 eMERGE phenotypes, adjusted for age, sex, and PCs. ORs represent the risk per SD change in the predicted PR interval. For all graphs, each point represents a phenotype, and green-colored points have a false discovery rate (FDR) *Q*<0.1. The purple dotted line denotes the expected false-positive rate, and the blue line corresponds to a Bonferroni correction. A-Fib indicates atrial fibrillation; and AV, atrioventricular.

(rG=0.33; *P*=0.22; Table 2). Thus, the genetic signal in the PR interval that is associated with AF is most strongly captured by the PR segment.

Associations With a Genetically Predicted PR Interval

The mixed models analyses indicate that a highly polygenic SNP-based genetic classifier could capture up to ≈23% of the variability of the PR interval. We used BSLMM[31] to construct a highly polygenic SNP classifier for the PR interval using the ARIC data set. This classifier was used to impute a genetically predicted PR interval for each subject in the eMERGE population. We then tested for an association between the predicted PR interval and 261 clinical phenotypes (with >250 cases and a genetic liability *P*<0.2).[35] Significant associations were seen with arrhythmia phenotypes including AF (OR=0.89; 95% CI, 0.83–0.95; *P*=0.0006; FDR *Q*=0.04; Table 3; Figure 2D). Thus, a genetically predicted prolonged PR interval is associated with decreased AF risk. The magnitude of this association was modestly attenuated when adjusting for GRS based on significant GWAS SNP associations for the PR interval and AF or when adjusting for these SNPs (n=17) as covariates, although the *P* value was no longer significant in the latter model (Table 3). Although no associations with an opposite direction of effect were significant, the strongest associations were with first-degree atrioventricular block, a diagnosis of a prolonged PR interval, and morbid obesity (Figure 2D). Analyses using a genetically predicted PR segment or P-wave duration did not identify any significant associations, although the top associations for the PR segment were the same as those seen for the genetically predicted PR interval (Figure I in the Data Supplement).

Validating the AF Association

To confirm the genetic correlation between the PR interval and AF, we tested the association between the genetically predicted PR interval and AF in 2 independent data sets. A second

Table 2. Genetic Correlations (rG)* Between PR Components and Atrial Fibrillation

	PR Interval		P Wave		PR Segment	
Adjustment	rG (SE)	*P* Value	rG (SE)	*P* Value	rG (SE)	*P* Value
None	–0.57 (0.27)	0.02	0.48 (0.29)	0.07	–0.88 (0.35)	0.0009
P wave adjusted†	–0.84 (0.34)	0.001	…	…	–0.84 (0.34)	0.001
PR segment adjusted†	0.33 (0.28)	0.22	0.32 (0.28)	0.22	…	…

AF indicates atrial fibrillation; ARIC, Atherosclerosis Risk in Communities; eMERGE, Electronic Medical Records and Genomics; and PC, principal component.

*Genetic correlations between the PR component measured in ARIC subjects and AF measured in eMERGE. Genetic correlations were adjusted for age, sex, and 20 PCs.

†Additional covariates for ARIC subjects added to the model.

EHR-derived data set (1206 AF cases and 2405 controls) that used the same AF phenotype definition as the discovery set had a significant association (OR=0.90; 95% CI, 0.85–0.98; *P*=0.01; Table 3). A comparable result was seen using subjects (1022 cases and 668 controls) from Vanderbilt's AF registry (OR=0.90; 95% CI, 0.81–0.99; *P*=0.03; Table 3). There was a similar magnitude and direction of effect when the results were stratified by AF subtypes (lone, paroxysmal, and persistent AF; Table 3).

PR Components and Metabolic Syndrome Phenotypes

Other than AF, the strongest genetic correlations for the PR phenotypes were with metabolic phenotypes (diabetes mellitus and obesity). Epidemiological studies have also shown that P-wave duration is positively associated with metabolic syndrome phenotypes.[26] We measured the genetic correlations between each PR interval component and metabolic phenotypes in the ARIC subjects. The PR interval and PR segment were not genetically correlated with any metabolic phenotype (Table 4). The P wave was positively genetically correlated with waist circumference (rG=0.47; *P*=0.03).

Table 3. Association Between a Genetically Predicted PR Interval and AF

Data Set	Subjects	Cases/ Controls	OR (95% CI)*	*P* Value
eMERGE	All AF cases	1547/3128	0.89 (0.83–0.95)	0.0006
	All AF cases, GRS adjusted†	1547/3128	0.90 (0.83–0.98)	0.02
	All AF cases, SNP adjusted‡	1547/3128	0.90 (0.81–1.01)	0.06
BioVU EHR set	All AF	1206/2405	0.90 (0.85–0.98)	0.01
BioVU AF registry	All AF	1022/668	0.90 (0.81–0.99)	0.03
	Lone AF	220/668	0.87 (0.74–1.03)	0.1
	Paroxysmal	444/668	0.90 (0.80–1.02)	0.09
	Persistent	259/668	0.93 (0.80–1.08)	0.34

AF indicates atrial fibrillation; BioVU, ; CI, confidence interval; EHR, electronic health record; eMERGE, Electronic Medical Records and Genomics; GRS, genetic risk scores; GWAS, genome-wide association studies; OR, odds ratio; PC, principal component; and SNP, single-nucleotide polymorphism.

*The OR is per SD increase in the genetically predicted PR interval. All association models are adjusted for age, sex, and 3 PCs.

†Adjusted for GRS for AF and the PR interval.

‡Adjusted for SNPs (n=17) previously associated with AF or the PR interval by GWAS.

Associations Between PR Components and Genetic Risk Scores

Finally, we examined whether there was a differential association between GRS based on known genetic modulators of the PR interval, AF, weight (measured by BMI), and the PR phenotypes. The PR GRS was significantly linearly correlated with each PR phenotype and had the largest linear correlations with PR interval and PR segment (Table 5). The AF GRS was weakly correlated with the P-wave duration (partial *r*=0.024; *P*=0.049), whereas the BMI GRS was correlated with both the PR interval (partial *r*=0.035; *P*=0.004) and P wave (partial *r*=0.048; *P*<0.001; Table 5).

Discussion

We used a discovery-oriented approach to identify clinical phenotypes modulated by genetic factors that also modulate the PR interval. We found that AF risk was genetically correlated with the PR interval, and this association was also observed using a highly polygenic risk score derived from the PR interval. We also observed genetic correlations with metabolic phenotypes including measures of adiposity. Thus, the genetic architecture underlying PR interval variability is driven, in part, by SNP variation that predisposes to AF risk and SNP variation that modulates body mass. Our analyses also found that the constitutive components of the PR interval (the PR segment and the P wave) were associated with different phenotypes, and further characterizing their individual genetic architectures may enable the development of better genetic risk prediction tools.

Although the PR interval is a genetically modulated measure of cardiac conduction, relatively few SNPs associated with this phenotype have been identified.[3–5] This paucity is not unexpected, as the genetic variability underlying many complex phenotypes is driven by numerous SNPs with small effect sizes that are difficult to detect by GWAS. We used modeling approaches that analyze the contributions of large number of SNPs to broadly characterize the genetic architecture of the PR interval. We found that common SNP variation accounted for at least 23% of phenotypic variability in the PR interval, indicating that much of the additive heritability of PR interval is currently hidden. When we examined the individual constituents of the PR interval, we found that the genetic correlation

Table 4. Genetic Correlations (rG)* Between PR Components and Metabolic Syndrome Phenotypes in ARIC

	PR Interval		P Wave		PR Segment	
Metabolic Phenotype	rG (SE)	*P* Value	rG (SE)	*P* Value	rG (SE)	*P* Value
Waist circumference	0.16 (0.19)	0.42	0.47 (0.21)†	0.03†	–0.06 (0.22)	0.80
Insulin resistance	–0.21 (0.30)	0.47	–0.14 (0.32)	0.66	–0.21 (0.33)	0.52
Hypertension	–0.10 (0.17)	0.56	0.23 (0.19)	0.22	–0.23 (0.19)	0.22
Triglycerides	–0.17 (0.18)	0.34	–0.14 (0.20)	0.46	–0.11 (0.20)	0.57
HDL cholesterol	–0.15 (0.19)	0.44	–0.15 (0.21)	0.62	–0.11 (0.22)	0.31
Metabolic syndrome	–0.05 (0.16)	0.76	–0.06 (0.17)	0.72	–0.03 (0.17)	0.66

ARIC indicates Atherosclerosis Risk in Communities; HDL, high-density lipoprotein; and PC, principal component.
*Analyses are adjusted for age, sex, and 20 PCs.
†Indicates a non-zero genetic correlation at $P<0.05$.

between PR segment and P-wave durations was not significantly different for zero, suggesting that they have differing genetic architectures. This observation is consistent with GWAS studies that have found that these intervals are associated with different SNPs.[36] To further characterize the genetic architectures of the PR phenotypes, we examined their genetic correlations with a large number of clinical phenotypes.

The individual PR phenotypes were not uniformly genetically correlated with the same clinical phenotypes. The most significant association was between the PR interval and PR segment and AF. The genetic correlation was negative, indicating that a genetically prolonged PR interval is associated with decreased risk of AF. This finding was not anticipated, as epidemiological studies have frequently observed that a prolonged PR interval is associated with an increased risk of AF.[1,2] This epidemiological association is attributed, in part, to prolongation in the PR interval because of acquired structural changes to the atrium that manifest as slowed atrial conduction and lead to increased atrial arrhythmogenicity.[37] Indeed, the PR interval duration increases with age, cardiac diseases,[38] and metabolic phenotypes such as obesity and hypertension.[39–41] These increases are most pronounced for the P wave.[26] These epidemiological associations are consistent with the trends in the genetic correlations that we observed when analyzing P-wave duration. The P wave was most strongly genetically correlated with metabolic phenotypes including waist circumference and type 2 diabetes mellitus, and a genetically predicted P-wave duration was most strongly associated with a diagnosis of obesity. Although not significant, the genetic correlation between the P wave and AF was positive, suggesting that a prolonged P-wave duration is associated with an increased risk of AF. In turn, these results indicate that there are genetic factors, such as those that modify BMI, which prolong the PR interval by affecting the P wave and which increase the risk of AF.

Table 5. Partial Correlation Coefficients (*r*)* Between PR Interval, AF, and BMI GRS and PR Phenotypes

	PR GRS		AF GRS		BMI GRS	
Phenotype	Partial *r*	*P* Value	Partial *r*	*P* Value	Partial *r*	*P* Value
PR interval	0.17	<0.0001	0.014	0.25	0.04	0.004
P-wave duration	0.07	<0.0001	0.024	0.049	0.05	<0.001
PR segment	0.14	<0.0001	0.002	0.87	0.01	0.36

AF indicates atrial fibrillation; BMI, body mass index; and GRS, genetic risk scores.
*Correlations are adjusted for age and sex.

The epidemiological association between PR interval and AF is U-shaped, as a short PR interval is also associated with increased AF risk.[42–45] Hence, our observation that a genetically shorter PR interval and PR segment is associated with an increased AF risk suggest that the inverse association is genetically mediated and that a short PR interval represents an accumulation of PR-shortening genetic variants, some of which also predispose to AF risk. Our results also suggest that the genetic mechanisms modulating the PR interval duration modulate AF risk in different directions. Thus, the genetic risk relationships between AF and each PR phenotype should be evaluated individually to better define this association. Another approach to examining the U-shaped relationship between the PR interval and AF is to use nonlinear statistical models. However, we think that ascribing the nonlinear association to the individual effects of the PR phenotypes is biologically more plausible than nonlinear additive genetic effects underlying the PR interval. Our findings also indicate that there is opportunity for more discoveries. For instance, a GRS comprised of known AF SNPs predominantly reflected the genetic risk associated with the P wave, but not the PR segment. Thus, identifying and evaluating additional SNP variants associated with the PR segment may reveal additional genetic mechanisms contributing to AF risk.

A significant genetic correlation between a pair of phenotypes suggests that they are modulated by a common set of genetic factors. Hence, a genetic predictor derived from one phenotype should associate with the other phenotype, provided that that predictor is able to capture a sufficient portion of the underlying genetic architecture of the first phenotype. We used BSLMM, which models phenotypes based on large numbers of SNP, to compute genetically predicted PR intervals in 3 data sets. This genetically predicted PR interval was associated with AF risk in each data set, and the direction was consistent with that observed with the genetic correlations analyses. As larger sample sizes become available and new

polygenic modeling techniques are developed, it may be possible to develop a PR interval–derived genetic classifier that can robustly predict AF risk and can offer sufficient lead time to maximize the benefit of intervention strategies.

There are several limitations to this study. We used phenotypes derived from EHR data sets, which often lack rigid phenotype definitions and can have incomplete ascertainment. Incomplete ascertainment and phenotype misclassification can attenuate associations. In support of the validity of our EHR AF phenotype, we note that it has been used for several genetic studies and has been shown to replicate known SNP associations.[22,23,25] It is possible that the genetic correlations we observed are spurious and are caused by SNPs simultaneously tagging disparate causative genetic variants that impact the phenotypes through distinct mechanisms.[46] However, all of our genetic correlations are supported by epidemiological observations, so this is unlikely for the phenotypes we identified. Our AF cases also had more comorbidities compared with our controls, which could inflate genetic correlation estimates for risk factors related to the metabolic syndrome. We also did not have sufficient individuals of other ancestries to evaluate and validate our findings in these other racial groups.

In conclusion, we used mixed models to characterize the genetic architecture of the PR interval. We found that SNP variants that predispose to AF and elevated body mass modulate the PR interval and that these variants differentially influence the P-wave and PR segment durations. Future GWAS should examine the constitutive PR phenotypes separately to more fully define the genetic modulators of the PR interval. Furthermore, focusing on genetic variation underlying the PR segment may identify novel AF genetic risk factors and mechanisms, which may lead to better AF risk prediction models.[47] Finally, a portion of the genetic predisposition toward AF is driven by genetic factors for metabolic risk factors including obesity, highlighting the continued need for aggressive risk modification and treatment for these predisposing conditions.

Acknowledgments

We thank the staff and participants of the ARIC study for their important contributions.

Sources of Funding

This work was supported by a career development award from the Vanderbilt Faculty Research Scholars Fund (Dr Mosley), American Heart Association (15MCPRP25620006 and 16FTF30130005;Dr Mosley), PGRN (P50 GM115305), R01 LM010685, K23 HL127704 (Dr Shoemaker), and R01 HL092217 (Dr Darbar). BioVU is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. The eMERGE Network is funded by National Human Genome Research Institute (NHGRI) and National Institute of General Medical Sciences (NIGMS) through the following grants: U01-HG-004610 (Group Health Cooperative/University of Washington); U01-HG-004608 (Marshfield Clinic Research Foundation and VUMC); U01-HG-04599 (Mayo Clinic); U01-HG-004609 (Northwestern University); U01-HG-006378, U01-HG-04603, and U01-HG-004438 (CIDR); and U01-HG-004424 (the Broad Institute) serving as Genotyping Centers. ARIC is supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Funding for GENEVA was provided by NHGRI grant U01HG004402 (E. Boerwinkle).

Disclosures

None.

References

1. Cheng S, Keyes MJ, Larson MG, McCabe EL, Newton-Cheh C, Levy D, et al. Long-term outcomes in individuals with prolonged PR interval or first-degree atrioventricular block. *JAMA*. 2009;301:2571–2577. doi: 10.1001/jama.2009.888.
2. Cheng M, Lu X, Huang J, Zhang S, Gu D. Electrocardiographic PR prolongation and atrial fibrillation risk: a meta-analysis of prospective cohort studies. *J Cardiovasc Electrophysiol.* 2014. doi:10.1111/jce.12539.
3. Smith JG, Magnani JW, Palmer C, Meng YA, Soliman EZ, Musani SK, et al; Candidate-gene Association Resource (CARe) Consortium. Genome-wide association studies of the PR interval in African Americans. *PLoS Genet*. 2011;7:e1001304. doi: 10.1371/journal.pgen.1001304.
4. Butler AM, Yin X, Evans DS, Nalls MA, Smith EN, Tanaka T, et al. Novel loci associated with PR interval in a genome-wide association study of 10 African American cohorts. *Circ Cardiovasc Genet*. 2012;5:639–646. doi: 10.1161/CIRCGENETICS.112.963991.
5. Pfeufer A, van Noord C, Marciante KD, Arking DE, Larson MG, Smith AV, et al. Genome-wide association study of PR interval. *Nat Genet*. 2010;42:153–159. doi: 10.1038/ng.517.
6. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet*. 2011;43:519–525. doi: 10.1038/ng.823.
7. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–569. doi: 10.1038/ng.608.
8. Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet*. 2012;8:e1002637. doi: 10.1371/journal.pgen.1002637.
9. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012;28:2540–2542. doi: 10.1093/bioinformatics/bts474.
10. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet.* 2013;45:984–994.
11. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129:687–702.
12. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al; eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med*. 2013;15:761–771. doi: 10.1038/gim.2013.72.
13. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al; eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4:13. doi: 10.1186/1755-8794-4-13.
14. Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet*. 2012;44:670–675. doi: 10.1038/ng.2261.
15. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008;84:362–369. doi: 10.1038/clpt.2008.89.
16. Bowton E, Field JR, Wang S, Schildcrout JS, Van Driest SL, Delaney JT, et al. Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med*. 2014;6:234cm3. doi: 10.1126/scitranslmed.3008604.
17. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–959.
18. Zuvich RL, Armstrong LL, Bielinski SJ, Bradford Y, Carlson CS, Crawford DC, et al. Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. *Genet Epidemiol*. 2011;35:887–898. doi: 10.1002/gepi.20639.
19. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10:5–6. doi: 10.1038/nmeth.2307.
20. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44:955–959. doi: 10.1038/ng.2354.
21. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide

scan to discover gene-disease associations. *Bioinformatics*. 2010;26:1205–1210. doi: 10.1093/bioinformatics/btq126.
22. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010;86:560–572. doi: 10.1016/j.ajhg.2010.03.003.
23. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31:1102–1110. doi: 10.1038/nbt.2749.
24. Israel RA. The International Classification of Disease. Two hundred years of development. *Public Health Rep*. 1978;93:150–152.
25. Weeke P, Denny JC, Basterache L, Shaffer C, Bowton E, Ingram C, et al. Examining rare and low-frequency genetic variants previously associated with lone or familial forms of atrial fibrillation in an electronic medical record system: a cautionary note. *Circ Cardiovasc Genet*. 2015;8:58–63. doi: 10.1161/CIRCGENETICS.114.000718.
26. Magnani JW, Lopez FL, Soliman EZ, Maclehose RF, Crow RS, Alonso A. P wave indices, obesity, and the metabolic syndrome: the atherosclerosis risk in communities study. *Obesity (Silver Spring)*. 2012;20:666–672. doi: 10.1038/oby.2011.53.
27. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82. doi: 10.1016/j.ajhg.2010.11.011.
28. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol*. 2013;1019:215–236. doi: 10.1007/978-1-62703-447-0_9.
29. Mosley JD, Witte JS, Larkin EK, Bastarache L, Shaffer CM, Karnes JH, et al. Identifying genetically driven clinical phenotypes using linear mixed models. *Nat Commun*. 2016;7:11433. doi: 10.1038/ncomms11433.
30. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29:1165–1188.
31. Zhou X, Carbonetto P, Stephens M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet*. 2013;9:e1003264. doi: 10.1371/journal.pgen.1003264.
32. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197–206. doi: 10.1038/nature14177.
33. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795.
35. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48:245–252. doi: 10.1038/ng.3506.
36. Verweij N, Mateo Leach I, van den Boogaard M, van Veldhuisen DJ, Christoffels VM, Hillege HL, et al; LifeLines Cohort Study. Genetic determinants of P wave duration and PR segment. *Circ Cardiovasc Genet*. 2014;7:475–481. doi: 10.1161/CIRCGENETICS.113.000373.
37. Soliman EZ, Cammarata M, Li Y. Explaining the inconsistent associations of PR interval with mortality: the role of P-duration contribution to the length of PR interval. *Heart Rhythm*. 2014;11:93–98. doi: 10.1016/j.hrthm.2013.10.003.
38. Alonso A, Soliman EZ, Chen LY, Bluemke DA, Heckbert SR. Association of blood pressure and aortic distensibility with P wave indices and PR interval: the multi-ethnic study of atherosclerosis (MESA). *J Electrocardiol*. 2013;46:359.e1–359.e6. doi: 10.1016/j.jelectrocard.2013.01.009.
39. Magnani JW, Johnson VM, Sullivan LM, Gorodeski EZ, Schnabel RB, Lubitz SA, et al. P wave duration and risk of longitudinal atrial fibrillation in persons ≥ 60 years old (from the Framingham Heart Study). *Am J Cardiol*. 2011;107:917–921.e1. doi: 10.1016/j.amjcard.2010.10.075.
40. Magnani JW, Gorodeski EZ, Johnson VM, Sullivan LM, Hamburg NM, Benjamin EJ, et al. P wave duration is associated with cardiovascular and all-cause mortality outcomes: the National Health and Nutrition Examination Survey. *Heart Rhythm*. 2011;8:93–100. doi: 10.1016/j.hrthm.2010.09.020.
41. Magnani JW, Zhu L, Lopez F, Pencina MJ, Agarwal SK, Soliman EZ, et al. P-wave indices and atrial fibrillation: cross-cohort assessments from the Framingham Heart Study (FHS) and Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J*. 2015;169:53–61.e1. doi: 10.1016/j.ahj.2014.10.009.
42. Magnani JW, Wang N, Nelson KP, Connelly S, Deo R, Rodondi N, et al; Health, Aging, and Body Composition Study. Electrocardiographic PR interval and adverse outcomes in older adults: the Health, Aging, and Body Composition study. *Circ Arrhythm Electrophysiol*. 2013;6:84–90. doi: 10.1161/CIRCEP.112.975342.
43. Soliman EZ, Prineas RJ, Case LD, Zhang ZM, Goff DC Jr. Ethnic distribution of ECG predictors of atrial fibrillation and its impact on understanding the ethnic distribution of ischemic stroke in the Atherosclerosis Risk in Communities (ARIC) study. *Stroke*. 2009;40:1204–1211. doi: 10.1161/STROKEAHA.108.534735.
44. Nielsen JB, Pietersen A, Graff C, Lind B, Struijk JJ, Olesen MS, et al. Risk of atrial fibrillation as a function of the electrocardiographic PR interval: results from the Copenhagen ECG Study. *Heart Rhythm*. 2013;10:1249–1256. doi: 10.1016/j.hrthm.2013.04.012.
45. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102. doi: 10.1161/JAHA.112.000102.
46. Gianola D, de los Campos G, Toro MA, Naya H, Schön CC, Sorensen D. Do molecular markers inform about pleiotropy? *Genetics*. 2015;201:23–29. doi: 10.1534/genetics.115.179978.
47. Everett BM, Cook NR, Conen D, Chasman DI, Ridker PM, Albert CM. Novel genetic markers improve measures of atrial fibrillation risk prediction. *Eur Heart J*. 2013;34:2243–2251. doi: 10.1093/eurheartj/eht033.

CLINICAL PERSPECTIVE

Biomarkers that predict disease risk enable risk stratification and disease prevention. Because many biomarkers and diseases are modulated by underlying genetic risk, it is possible to associate them based on this shared genetic risk. Importantly, these genetic associations can be assessed across different data sets, as long as all subjects have genotypic data, and the approach can be used to study relationships between potential biomarkers and disease. Here, we measured genetic correlations, a measure of genetic association, between a potential biomarker, the PR interval (and its individual components, the P wave and the PR segment), and 63 electronic health record disease phenotypes. The ECG phenotypes were analyzed in the ARIC cohort (Atherosclerosis Risk in Communities), and the electronic health record phenotypes in the eMERGE network (Electronic Medical Records and Genomics). We found that a genetically predicted PR interval was associated with atrial fibrillation risk, consistent with previous epidemiological studies but with an opposite direction of association. The individual components had different genetic architectures, were not correlated with each other, and atrial fibrillation risk was predominantly associated with genetically determined PR segment. This study establishes that the shared genetic architectures of clinical phenotypes like atrial fibrillation and putative biomarkers like the PR and its components can identify epidemiological associations, validate the biomarkers, and point to disease mechanisms.

Investigating the Genetic Architecture of the PR Interval Using Clinical Phenotypes

Jonathan D. Mosley, M. Benjamin Shoemaker, Quinn S. Wells, Dawood Darbar, Christian M. Shaffer, Todd L. Edwards, Lisa Bastarache, Catherine A. McCarty, Will Thompson, Christopher G. Chute, Gail P. Jarvik, David R. Crosslin, Eric B. Larson, Iftikhar J. Kullo, Jennifer A. Pacheco, Peggy L. Peissig, Murray H. Brilliant, James G. Linneman, John S. Witte, Josh C. Denny and Dan M. Roden

Circ Cardiovasc Genet. 2017;10:
doi: 10.1161/CIRCGENETICS.116.001482
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268

SUPPLEMENTAL MATERIAL

Supplementary Figures

Supplementary Figure 1. Associations between genetically predicted PR segment and P wave durations and EHR phenotypes.

Supplementary Tables

Supplementary Table 1: Characteristics of the eMERGE population.

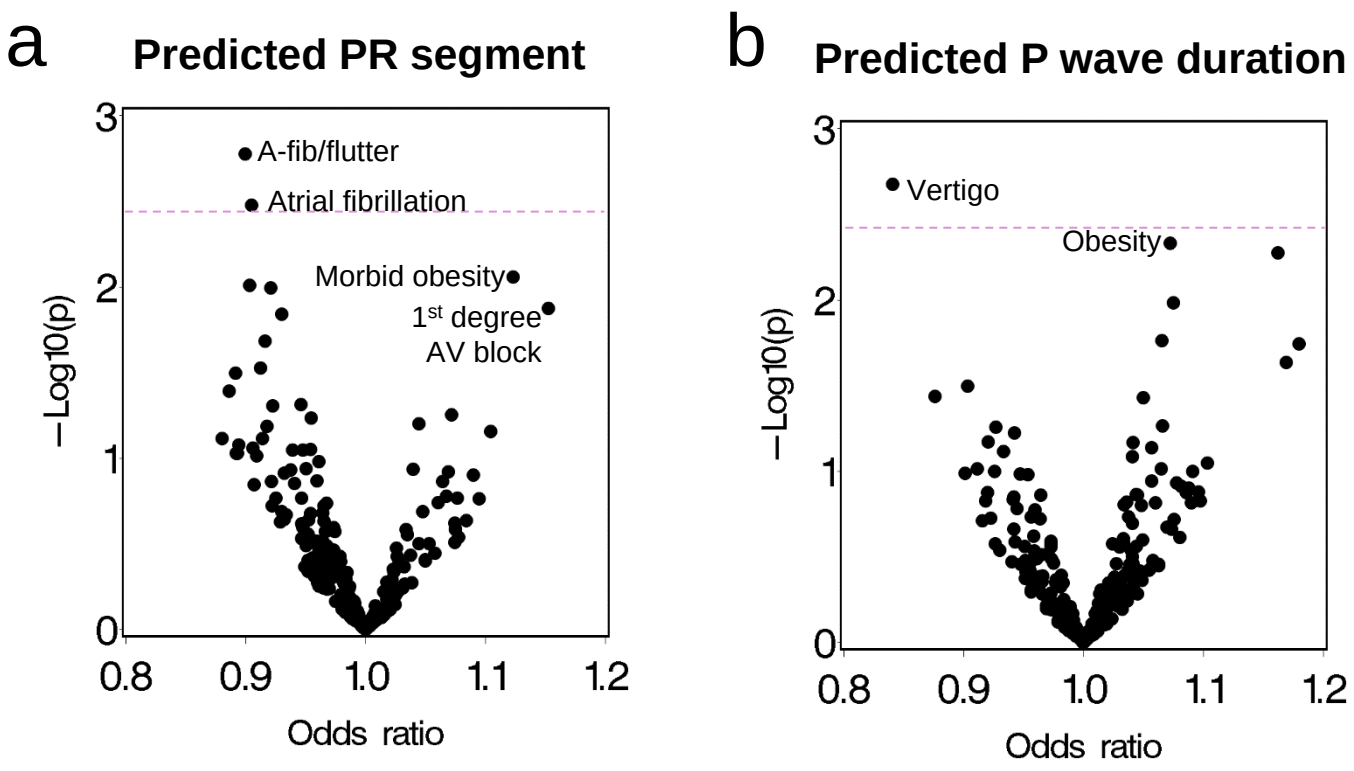
Supplementary Table 2: List of SNPS identified by prior GWAS used to compute genetic risk scores for the PR interval and atrial fibrillation.

Supplementary Table 3: List of SNPS identified by prior GWAS used to compute genetic risk scores for body mass index (BMI).

Supplementary Table 4: Characteristics of the ARIC population.

Supplementary Table 5: Demographics for the eMERGE atrial fibrillation cases and controls.

Supplementary Figure 1



Supplementary Figure 1. Plot of p-value versus odds-ratio (OR) from logistic regression analyses of the association between a genetically predicted a) PR segment or b) P wave duration and 261 eMERGE phenotypes, adjusted for sex, gender and principal components. ORs represent the risk per standard deviation change in the predicted value. For all graphs, each point represents a phenotype and green colored points have an FDR q<0.1. The purple line denotes the expected false positive rate (=1/261).

Supplementary Table 1: Characteristics of the eMERGE population.

Characteristic		
Gender [n (%)]		
Males	6,253 (48.2)	
Females	6,725 (51.8)	
Birth Decade		
median (IQR)	1940 (1920-1950)	
Distinct PheWAS codes per subject		
median (IQR)	44 (21-76)	
Phenotypes used for genetic correlation analyses	**Cases**	**Controls**
Hypertension	6,354	3,425
Essential hypertension	6,321	3,425
Disorders of lipid metabolism	6,042	3,896
Hyperlipidemia	5,966	3,896
Osteoarthrosis	3,989	5,348
Pain in limb	3,851	5,357
Peripheral enthesopathies and allied syndromes	3,473	4,805
Osteoarthrosis NOS	3,391	5,128
Hypercholesterolemia	3,180	3,611
Ischemic Heart Disease	3,160	5,606
Acute upper respiratory infections of multiple or unspecified sites	2,984	5,486
Coronary atherosclerosis	2,717	5,282
Overweight, obesity and other hyperalimentation	2,610	6,730
Diabetes mellitus	2,488	6,532
GERD	2,442	5,621
Type 2 diabetes	2,408	6,504
Obesity	2,301	6,265
Benign neoplasm of colon	2,244	6,592
Diverticulosis and diverticulitis	2,179	4,194
Hemorrhoids	2,049	4,612
Dizziness and giddiness (Light-headedness and vertigo)	2,036	6,247
Diverticulosis	1,966	3,975
Anxiety, phobic and dissociative disorders	1,870	5,101
Glaucoma	1,866	3,866
Other headache syndromes	1,711	6,695
Other peripheral nerve disorders	1,698	6,587
Atherosclerosis	1,695	3,770
Impacted cerumen	1,611	4,998

Atrial fibrillation and flutter	1,600	3,166
Atrial fibrillation	1,547	3,128
Peripheral vascular disease	1,539	4,415
Sprains and strains of back and neck	1,533	2,484
Diseases of sebaceous glands	1,533	5,898
Tobacco use disorder	1,527	7,094
Nausea and vomiting	1,442	6,713
Peripheral vascular disease, unspecified	1,361	3,812
Atherosclerosis of the extremities	1,318	3,128
Other arthropathies	1,227	6,186
Urinary incontinence	1,098	3,990
Otitis media and Eustachian tube disorders	1,074	4,941
Angina pectoris	1,029	3,755
Other chronic ischemic heart disease, unspecified	953	3,382
Sebaceous cyst	924	4,526
Atherosclerosis of native arteries of the extremities with intermittent claudication	890	1,878
Dementias	889	2,179
Varicose veins	824	3,203
Other upper respiratory disease	813	3,322
Otitis media	801	3,790
Adjustment reaction	777	3,079
Other persistent mental disorders due to conditions classified elsewhere	760	2,709
Varicose veins of lower extremity	748	3,034
Disorders of function of stomach	729	3,831
Symptoms involving cardiovascular system	678	3,201
Fracture of unspecified bones	662	3,831
Dyspepsia and other specified disorders of function of stomach	655	3,484
Corns and callosities	649	3,284
Other aneurysm	617	2,780
Morbid obesity	602	3,600
Cellulitis and abscess of leg, except foot	598	3,668
Alzheimer's disease	590	1,860
Hyposmolality and/or hyponatremia	583	3,375
Type 2 diabetes with neurological manifestations	525	3,479
Lymphadenitis	501	3,679

Supplementary Table 2: List of SNPS available in each data set for use in computing genetic risk scores for either the PR interval or atrial fibrillation. Shown are the published weights used from computing GRSs. In addition, association statistics for either the PR interval or AF, respectively, are shown.

Phenotype	SNP	Beta (s.e.) for the PR interval from prior GWAS*	OR (95% CI) for AF in the eMERGE data set
PR interval	rs11897119	1.4 (0.2)	0.98 (0.89-1.08)
	rs11708996	3.0 (0.3)	0.92 (0.81-1.06)
	rs6800541	3.8 (0.2)	0.89 (0.81-0.98)
	rs7692808	-2.0 (0.2)	1.05 (0.95-1.17)
	rs251253	-1.5 (0.2)	0.92 (0.83-1.02)
	rs3807989	2.3 (0.2)	0.92 (0.84-1.01)
	rs4944092	-1.2 (0.2)	1.06 (0.96-1.18)
	rs1896312	2.0 (0.2)	0.98 (0.88-1.09)

			Association in ARIC data set		
Phenotype	**SNP**	**OR (95% CI) for AF previously reported by GWAS****	**PR interval Beta (s.e.)**	**PR segment Beta (s.e.)**	**P wave duration Beta (s.e.)**
Atrial fibrillation	rs6666258	1.18 (1.13–1.23)	-0.69 (0.45)	-0.43 (0.44)	-0.28 (0.23)
	rs3903239	1.14 (1.10–1.18)	0.63 (0.42)	0.64 (0.40)	0.03 (0.21)
	rs6817105	1.64 (1.55–1.73)	1.37 (0.66)	0.91 (0.64)	0.54 (0.33)
	rs2040862	1.15 (1.09–1.21)	0.00 (0.55)	-0.14 (0.53)	0.13 (0.27)
	rs3807989	0.88 (0.84–0.91)	2.18 (0.43)	1.51 (0.42)	0.61 (0.21)
	rs10821415	1.13 (1.08–1.18)	-0.29 (0.42)	-0.44 (0.41)	0.18 (0.21)
	rs10824026	0.85 (0.81–0.90)	-0.39 (0.58)	-1.10 (0.56)	0.67 (0.29)
	rs1152591	1.13 (1.09–1.18)	0.71 (0.43)	0.55 (0.41)	0.19 (0.21)
	rs7164883	1.16 (1.10–1.22)	-1.30 (0.57)	-1.32 (0.55)	0.06 (0.28)
	rs2106261	1.24 (1.17–1.30)	-0.37 (0.55)	-0.62 (0.54)	0.28 (0.28)

* Pfeufer A, van Noord C, Marciante KD, Arking DE, Larson MG, Smith AV et al. Genome-wide association study of PR interval. Nat Genet 2010; 42: 153–159.

**Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nat Genet 2012; 44: 670–675.

Supplementary Table 3: List of previously* published SNPs and weights used to compute a BMI genetic risk score.

*Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.

SNP	Beta
rs1558902	0.082
rs6567160	0.056
rs13021737	0.060
rs10938397	0.040
rs543874	0.048
rs2207139	0.045
rs11030104	0.041
rs3101336	0.033
rs7138803	0.032
rs10182181	0.031
rs3888190	0.031
rs1516725	0.045
rs12446632	0.040
rs2287019	0.036
rs16951275	0.031
rs3817334	0.026
rs2112347	0.026
rs12566985	0.024
rs3810291	0.028
rs7141420	0.024
rs13078960	0.030
rs10968576	0.025
rs17024393	0.066
rs657452	0.023
rs12429545	0.033
rs12286929	0.022
rs13107325	0.048
rs11165643	0.022
rs7903146	0.023
rs10132280	0.023
rs17405819	0.022
rs6091540	0.019
rs1016287	0.023
rs4256980	0.021
rs17094222	0.025
rs12401738	0.021
rs7599312	0.022
rs2365389	0.020
rs205262	0.022
rs2820292	0.020
rs12885454	0.021
rs9641123	0.019

rs12016871	0.030
rs16851483	0.048
rs1167827	0.020
rs758747	0.023
rs1928295	0.019
rs9925964	0.019
rs11126666	0.021
rs2650492	0.021
rs6804842	0.019
rs12940622	0.018
rs7164727	0.018
rs11847697	0.049
rs4740619	0.018
rs492400	0.016
rs13191362	0.028
rs3736485	0.018
rs17001654	0.031
rs11191560	0.031
rs2080454	0.017
rs7715256	0.016
rs2176040	0.014
rs1528435	0.018
rs2075650	0.026
rs1000940	0.019
rs2033529	0.019
rs11583200	0.018
rs7239883	0.016
rs2836754	0.016
rs9400239	0.019
rs10733682	0.017
rs11688816	0.017
rs11057405	0.031
rs9914578	0.020
rs977747	0.017
rs2121279	0.025
rs29941	0.018
rs11727676	0.036
rs3849570	0.019
rs9374842	0.019
rs6477694	0.017
rs4787491	0.016
rs1441264	0.018
rs7899106	0.040
rs2176598	0.020
rs2245368	0.032
rs17203016	0.021
rs17724992	0.019
rs7243357	0.022
rs16907751	0.035

rs1808579	0.017
rs13201877	0.023
rs2033732	0.019
rs9540493	0.017
rs1460676	0.020
rs6465468	0.017

Supplementary Table 4: Characteristics of the ARIC population.
See Methods for metabolic trait definitions.

Characteristic	Value
Gender [n (%)]	
Males	3,040 (45.2)
Females	3,691 (54.8)
Age at first visit (years)	
median (IQR)	54 (49-59)
Metabolic traits [n (%)]	
Waist circumference	3,242 (48.1)
Insulin Resistance	444 (6.6)
Hypertension	1,341 (19.9)
Triglycerides	1,983 (29.5)
HDL cholesterol	2,698 (44.1)
Metabolic syndrome	1,435 (21.3)
EKG parameters [mean (s.d.)]	
PR interval duration (ms)	159 (23)
P wave duration (ms)	105 (12)
PR segment duration (ms)	55 (28)

Supplementary Table 5: Demographics for the eMERGE atrial fibrillation cases and controls. Comorbidities are based on ICD-9 pheWAS codes, as described in the methods.

Characteristic	AF cases (n=1,547)	AF controls (n=3,128)
Gender [n (%)]		
Males	867 (56.0)	1,642 (52.5)
Females	680 (44.0)	1,486 (47.5)
Birth Decade		
median (IQR)	1925 (1915-1935)	1935 (1925-1945)
Comorbidities[1] [n (%)]		
Essential hypertension	1249 (80.7)	1467 (46.9)
Coronary atherosclerosis	851 (55.0)	380 (12.1)
Congestive heart failure	809 (52.3)	111 (3.5)
Hemorrhage due to anticoagulants	789 (51.0)	151 (4.8)
Cerebrovascular disease	602 (38.9)	360 (11.5)
Heart valve disorders	536 (34.6)	152 (4.9)
Renal failure	499 (32.3)	195 (6.2)
Type 2 diabetes	499 (32.3)	526 (16.8)
Chronic airway obstruction	442 (28.6)	256 (8.2)
Myocardial infarction	394 (25.5)	84 (2.7)
Cardiomegaly	365 (23.6)	54 (1.7)
Tobacco use disorder	246 (15.9)	361 (11.5)