

## Fifteen Genetic Loci Associated with the Electrocardiographic P-wave

Ingrid E. Christophersen, MD, PhD,<sup>\*1-3</sup> Jared W. Magnani, MD, MSc,<sup>\*4,5</sup> Xiaoyan Yin, PhD,<sup>\*5,6</sup> John Barnard, PhD,<sup>7</sup> Lu-Chen Weng, PhD,<sup>1,2</sup> Dan E. Arking, PhD,<sup>8</sup> Maartje N. Niemeijer MD,<sup>9</sup> Steven A. Lubitz, MD, MPH,<sup>1,2</sup> Christy L. Avery, PhD,<sup>10</sup> Qing Duan, PhD,<sup>11</sup> Stephan B. Felix, MD,<sup>12,13</sup> Joshua C. Bis, PhD,<sup>14</sup> Kathleen F. Kerr, PhD,<sup>15</sup> Aaron Isaacs, PhD,<sup>16</sup> Martina Müller-Nurasyid, PhD,<sup>17-19</sup> Christian Müller, MSc,<sup>20,21</sup> Kari E. North, PhD,<sup>10,22</sup> Alex P. Reiner, MD, MSc,<sup>23,24</sup> Lesley F. Tinker, PhD RD,<sup>24</sup> Jan A. Kors, PhD,<sup>25</sup> Alexander Teumer, PhD,<sup>26</sup> Astrid Petersmann, MD,<sup>27</sup> Moritz F. Sinner, MD, MPH,<sup>18,19</sup> Petra Buzkova, PhD,<sup>15</sup> Jonathan D. Smith, PhD,<sup>7</sup> David R. Van Wagoner, PhD,<sup>7</sup> Uwe Völker, PhD,<sup>12,28</sup> Melanie Waldenberger, PhD,<sup>19,29,30</sup> Annette Peters, PhD,<sup>19,30,31</sup> Thomas Meitinger, MD,<sup>19,32,33</sup> Marian C. Limacher, MD,<sup>34</sup> Kirk C. Wilhelmsen, MD PhD,<sup>11,35</sup> Bruce M. Psaty, MD, PhD,<sup>14,36-38</sup> Albert Hofman, MD, PhD,<sup>9</sup> Andre Uitterlinden, PhD,<sup>9,39</sup> Bouwe P. Krijthe, MD, PhD,<sup>9</sup> Zhu-Ming Zhang, MD, MPH,<sup>40</sup> Renate B. Schnabel, MD, MSc,<sup>20,21</sup> Stefan Kääh, MD,<sup>18,19</sup> Cornelia van Duijn, PhD,<sup>16</sup> Jerome I. Rotter, MD,<sup>41</sup> Nona Sotoodehnia, MD, PhD,<sup>14,42</sup> Marcus Dörr, MD,<sup>12,13</sup> Yun Li, PhD,<sup>11,43</sup> Mina K. Chung, MD,<sup>7</sup> Elsayed Z. Soliman, MD, MSc, MS,<sup>40</sup> Alvaro Alonso, MD, PhD,<sup>44</sup> Eric A. Whitsel, MD, MPH,<sup>10,45</sup> Bruno H. Stricker, MMed, PhD,<sup>9,39</sup> Emelia J. Benjamin, MD, ScM,<sup>4,5,46</sup> Susan R. Heckbert, MD, PhD,<sup>36,37</sup> Patrick T. Ellinor, MD, PhD.<sup>1,2</sup>

\*These authors contributed equally to this manuscript.

### Affiliations

1. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA.
2. Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA.
3. Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Rud, Norway.
4. Department of Medicine, Boston University School of Medicine, Boston, MA.
5. NHLBI and Boston University's Framingham Heart Study, Framingham, MA.
6. Department of Biostatistics, Boston University School of Public Health, Boston, MA.
7. Departments of Cardiovascular Medicine, Cellular and Molecular Medicine, Molecular Cardiology, and Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH.
8. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.
9. Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.
10. Department of Epidemiology, University of North Carolina, Chapel Hill, NC.
11. Department of Genetics, University of North Carolina, Chapel Hill, NC.
12. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany.
13. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany.
14. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA.
15. Department of Biostatistics, University of Washington, Seattle, WA.

## Genetic Loci Associated with ECG P-wave Indices

16. Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands.
17. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
18. Department of Medicine I, University Hospital Munich, Ludwig-Maximilians University, Munich, Germany
19. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany.
20. Clinic for General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany.
21. German Center for Cardiovascular Research (DZHK e.V.), partner site Hamburg/Lübeck/Kiel, Germany.
22. Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC.
23. Department of Epidemiology, University of Washington, Seattle, WA.
24. Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA.
25. Dept of Medical Informatics, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.
26. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany.
27. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany.
28. Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany.
29. Research unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
30. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
31. German Center for Diabetes Research, Neuherberg, Germany.
32. Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
33. Institute of Human Genetics, Technische Universität München, Munich, Germany.
34. Department of Medicine, Division of Cardiovascular Medicine, University of Florida, Gainesville, FL.
35. Department of Neurology, University of North Carolina, Chapel Hill, NC.
36. Department of Epidemiology and Cardiovascular Health Research Unit, University of Washington, Seattle, WA.
37. Group Health Research Institute, Group Health Cooperative, Seattle, WA.
38. Department of Health Services, University of Washington, Seattle, WA.
39. Department of Internal Medicine, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.
40. Epidemiological Cardiology Research Center (EPICARE), Wake Forest School of Medicine, Winston Salem, NC.
41. Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics at Harbor-UCLA Medical Center, Torrance, CA.

## Genetic Loci Associated with ECG P-wave Indices

42. Cardiology Division, University of Washington, Seattle, USA.
43. Department of Biostatistics, Department of Computer Science, University of North Carolina, Chapel Hill, NC.
44. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA.
45. Department of Medicine, University of North Carolina, Chapel Hill, NC.
46. Department of Epidemiology, Boston University School of Public Health, Boston, MA.

**Short Title:** Genetic basis of the P-wave

**Word count:** 6890

**Corresponding authors:**

Jared W. Magnani, MD, MSc  
Boston University School of Medicine  
88 E. Newton Street  
Boston, MA, 02118  
Phone: 617 638 8714  
Fax: 617 638 8969  
jmagnani@bu.edu

Patrick T. Ellinor, MD, PhD  
Broad Institute of Harvard and MIT  
415 Ames Street  
Cambridge, MA 02124  
Phone: 617 724 8729  
ellinor@mgh.harvard.edu

**Journal subject codes:** Genetic, Association Studies; Genetics; Arrhythmias; Electrocardiology (ECG)

**Disclosures:** Dr. Ellinor is the PI on a grant from Bayer HealthCare to the Broad Institute focused on the genetics and therapeutics of atrial fibrillation. The remaining authors have no disclosures.

## Genetic Loci Associated with ECG P-wave Indices

### Abstract

**Background:** The P-wave on an electrocardiogram is a measure of atrial electrical function and its characteristics may serve as predictors for atrial arrhythmias. Increased mean P-wave duration and P-wave terminal force traditionally have been used as markers for left atrial enlargement and both have been associated with increased risk of atrial fibrillation. Here, we explore the genetic basis of P-wave morphology through meta-analysis of genome-wide association study (GWAS) results for P-wave duration and P-wave terminal force from twelve cohort studies.

**Methods and results:** We included 44,456 individuals, of which 6,778 (16%) were of African ancestry. Genotyping, imputation, and GWAS were performed at each study site. Summary level results were meta-analyzed centrally using inverse-variance weighting. In meta-analyses of P-wave duration, we identified six significant ( $P < 5 \times 10^{-8}$ ) novel loci, and replicated a prior association with *SCN10A*. We identified three loci at *SCN5A*, *TBX5*, and *CAV1/CAV2* that were jointly associated with the PR-interval, PR-segment, and P-wave duration. We identified six novel loci in meta-analysis of P-wave terminal force. Four of the identified genetic loci were significantly associated with gene expression in 329 left atrial samples. Finally, we observed that some of the loci associated with the P-wave were linked to overall atrial conduction, whereas others identified distinct phases of atrial conduction.

**Conclusion:** We have identified six novel genetic loci associated with P-wave duration and six novel loci associated with P-wave terminal force. Future studies of these loci may aid in identifying new targets for drugs that may modify atrial conduction or treat atrial arrhythmias.

**Keywords:** genome wide association study, electrocardiography, arrhythmia, genetic variation, atrial function

## Genetic Loci Associated with ECG P-wave Indices

### INTRODUCTION

The electrocardiographic P-wave is a measure of electrical activation of the atria and was first described by Starling and Bayliss over 120 years ago.<sup>1</sup> P-wave indices (PWI) describe atrial electrical function and are quantified as measures of duration and voltage from the surface electrocardiogram (ECG). The clinical utility of PWI have been demonstrated by their relation to atrial electrophysiology using invasive intracardiac studies,<sup>2, 3</sup> and by their assessment as risk factors for atrial fibrillation (AF).<sup>4-7</sup> P-wave terminal force – derived from the product of the duration of the negative phase of the P wave in lead V1 and its terminal depth (**Figure 1**) – has been used as an ECG marker for the presence of left atrial enlargement.<sup>8</sup> In addition to terminal force, P-wave duration has been used widely because of its accessibility from the electrocardiogram. Elucidation of the associations of genetic variants related to PWI may improve our understanding of the biology underlying atrial conduction and further inform the genetic determinants of risk of arrhythmias, including AF.

We sought to enhance our understanding of atrial electrical function by conducting large genome-wide association studies (GWAS) of PWI in individuals of European and African ancestry. Prior GWAS have distinguished genetic loci uniquely associated with P-wave duration from those associated with the PR interval and the PR-segment.<sup>9, 10</sup> The largest study evaluating common variants associated with P-wave maximal duration to date, included 16,468 individuals of European ancestry and identified three significant genetic loci.<sup>10</sup> There have been no reports on common genetic variants associated with P-wave terminal force.

In the present investigation, we examined the genome-wide associations of maximum P-wave duration and P-wave terminal force in twelve cohorts, including more than 44,000 individuals of European and African American ancestry. We hypothesized that the PWI examined have novel genetic associations, distinct from those previously identified as associated with the PR interval.

## Genetic Loci Associated with ECG P-wave Indices

### MATERIALS AND METHODS

#### Study cohorts

Participants of European or African ancestry from twelve studies contributed to the present analyses (**Table S1**): Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Erasmus Rucphen Family (ERF) Study, Framingham Heart Study (FHS), Cooperative Health Research in the Augsburg Region (KORA), Gutenberg Health Study I (GHS I), Multi-Ethnic Study of Atherosclerosis (MESA), Rotterdam Studies I, II, and III (RS I-III), Study of Health in Pomerania (SHIP), and 3 distinct studies affiliated with the Women's Health Initiative clinical trials (WHI CT): Genome-wide Association Research Network (GARNET), Modification of Particulate Matter-Mediated Arrhythmogenesis in Populations (MOPMAP), and SNP Health Association Resource Project (SHARe). The study was limited to participants with available DNA and consent for genetics research. We excluded individuals in whom we could not determine PWI and individuals with a history of AF, an implanted pacemaker or ICD, WPW syndrome, complete heart block, or who received medication altering atrioventricular nodal conduction (beta blockers, dihydropyridine calcium channel blockers, type I and III antiarrhythmic medication, and digoxin). Individuals with mitral valve disease were not excluded. All participants had given their written informed consent to participation and all studies were approved by their respective IRBs.

#### PWI measurement

Participants in all cohorts underwent a standardized, digital 12-lead ECG at rest in the supine position as a component of the cohort examination. All PWI were quantified using contemporary software algorithms from digitized tracings. Electrocardiograms from the ARIC, CHS, MESA, and data from the 3 WHI CT sub-studies were automatically processed at the Epidemiological Cardiology Research Center (EPICARE) of Wake Forest University School of Medicine (Winston Salem, NC) using General Electric (GE)

## Genetic Loci Associated with ECG P-wave Indices

12-SL software (GE, Milwaukee, WI) running under GE MUSE and Magellan Research Work Station. ECGs in FHS were read independently at FHS and analyzed using GE 12-SL software. PWI calculated using GE 12-SL software have previously been reported as having a repeatability of 100%.<sup>4</sup> In GHS I, ECGs were recorded using the Welch Allyn CardioPerfect electrocardiograph (Skaneateles Falls, NY) and PWI were calculated using the GE Healthcare software CASE, CardioSoft, version 6 (Palatine, IL). In ERF and RSI-III, electrocardiograms were recorded using the ACTA Gnosis IV ECG recorder (Esaote Biomedica, Florence, Italy) and in SHIP, electrocardiograms were recorded using the Personal 120LD (Esaote, Genova, Italy). In ERF, RSI-III, and SHIP, electrocardiograms were analyzed with the Modular ECG Analysis System.<sup>11</sup> KORA employed the AMEDTEC ECGpro system (AMEDTEC Medizintechnik Aue GmbH, Aue, Germany), and analyzed tracings using the Hannover ECG system (Corscience GmbH&Co. KG, Erlangen, Germany).

PWI were selected for their availability across cohorts. P-wave duration in each lead was calculated by summing the durations of the positive and negative phase of the P-wave in each lead. The maximum P-wave duration was selected as the highest value among the 12 leads. P-wave terminal force, specific to right precordial lead V1, was calculated as the product of the duration and voltage of the negative deflection of the P-wave in lead V1. Distributions of the PWI by cohort are presented in **Table S2**. P-wave terminal force was not available in KORA and GHS.

### Genotyping

Details on genotyping methods, quality control, imputation, and software are provided in detail in **Table S3**. All cohorts used genome-wide arrays for genotyping. Genotyping was performed independently in each cohort using the following arrays: Affymetrix 6.0 genome-wide array (ARIC, KORA, GHS, MESA, SHIP, WHI SHARE), Illumina 370 (CHS), Affymetrix 500K + 50K (FHS), Illumina Infinium (RSI-III), Illumina Quad (WHI GARNET), and Affymetrix Titan (MOPMAP). Samples with call rates <95% (ARIC, GHS, MESA,

## Genetic Loci Associated with ECG P-wave Indices

WHI MOPMAP, WHI SHARe), <97% (CHS, FHS), or <98% (KORA, Rotterdam, ERF, GARNET) at genotyped markers were excluded.

### eQTL analyses

A complete description is enclosed in the Supplemental Methods. In brief, eQTL analyses were performed from human left atrial tissue samples and from the Genotype-Tissue Expression (GTEx) database.

### Estimation of the variance explained by the significant genetic variants

Results from the meta-analyses were used to estimate the percentage of the variance that can be explained by the most significant SNP at each genetic locus, as described by Hibar *et al.*<sup>12</sup> We calculated estimated variance explained for SNPs identified in both separate and combined ancestry analyses and adjusted for age, sex, RR-interval and the maximum number of principal components adjusted for in any of the ancestry-matched GWAS.

### Evaluation of the genetic associations between P-wave duration and P-wave terminal force

We performed a summary-level analysis to test the association between the genetic variants associated with P-wave duration and the measured P-wave terminal force and vice versa, in the European ancestry group. In order to define independent signals, we performed LD clumping by selecting SNPs with  $p < 1 \times 10^{-5}$  and assigning all SNPs within 250 kb in each direction in moderate LD ( $r^2 > 0.5$ ) and  $p < 0.05$  to the same clump. LD clumping was iterated until all SNPs with  $p < 1 \times 10^{-5}$  had been clumped, using PLINK v.1.9.<sup>13</sup> To assess whether the selected LD threshold was sufficient to identify independent variants, we also performed LD clumping using lower thresholds ( $r^2 > 0.1$  and  $r^2 > 0.05$ ). LD information was obtained from the Framingham Heart Study. Genetic risk scores (GRS) were then created for both ECG traits,



## Genetic Loci Associated with ECG P-wave Indices

including only SNPs significantly associated with the trait and that were not in LD with any other significant SNPs. The scores were used to evaluate the association between the genetically determined P-wave duration and measured P-wave terminal force, and between the genetically determined P-wave terminal force and measured P-wave duration, using the R-package gtx.<sup>14</sup> The association results were then used to estimate the fraction of the total variance of each trait that could be explained by the GRS of the other trait.

### Statistical analysis

Each cohort conducted an independent association analysis relating genotype to P-wave duration and P-wave terminal force. Cohorts with participants of African ancestry performed separate analyses by race. Adjustment for principal components was performed in individual studies if appropriate. Cohorts used linear regression models that adjusted for participant age, sex, RR-interval, and cohort or site in a primary analysis and further adjusted for hypertension (defined as systolic blood pressure  $\geq 140$  mm Hg, diastolic blood pressure  $\geq 90$  mm Hg, or treatment with an antihypertensive) and body mass index (weight [kg]/height [m]<sup>2</sup>) in a secondary analysis.

Genetic variants were given a marker ID consisting of chromosome number and chromosomal position before meta-analysis, to avoid inconsistencies in dbSNP rsIDs across different builds. Cohort-specific summary statistics were pooled for inverse-variance weighted fixed-effect meta-analyses, which were performed separately for maximum P-wave duration and P-wave terminal force. Both ancestry-specific and combined ancestry analyses were performed for both traits. Variants that were not present in at least two studies were excluded. Associations with two-sided  $P < 5 \times 10^{-8}$  were considered genome-wide significant.

## Genetic Loci Associated with ECG P-wave Indices

### RESULTS

The study sample for maximum P-wave duration included 37,678 individuals of European ancestry and 6,778 of African ancestry. Since KORA and GHS did not contribute analyses of the P-wave terminal force, the total number of individuals of European ancestry for that meta-analysis was 33,955. Descriptive characteristics of the study participants are provided in **Table S2**. The three WHI CT cohorts included only women, while the remaining cohorts included both sexes. Mean age within-study ranged from 46±16 to 72±5 years. Mean P-wave duration was highly consistent across cohorts. In contrast, P-wave terminal force had a wider range with proportionately larger standard deviation. Details regarding study design, genotyping, GWAS, and imputation are listed in **Table S3**.

#### P-wave duration

In the primary analyses, we identified nine loci associated with P-wave duration, five of which have not been associated with P-wave duration previously. The most significant variant from each locus is listed in **Table S4** and **Table S5**. **Table S6** displays a comparison of the genome-wide significant loci across P-wave duration and P-wave terminal force analyses. **Figure 2** depicts the Manhattan plots for the P-wave duration analysis, stratified by ancestry. **Figure S1A** shows the Manhattan plot for pooled ancestries and **Figure S2** shows the regional plots for all genetic loci associated with P-wave duration in this study. **Table S7** summarizes the relationship between significant genetic loci in the present GWAS of P-wave duration and previous PWI GWAS.

#### *Genetic loci unique to P-wave duration*

We identified an association between P-wave duration and a variant on chromosome 5 (rs4276421) that resides ~100 kb upstream of *HCN1*. On chromosome 7, we identified a locus comprising *CAV1* and *CAV2*, with the most significant variant located intronic to *CAV1*. More than forty variants at this locus had

## Genetic Loci Associated with ECG P-wave Indices

significant or near significant eQTLs for *CAV1/CAV2* in European American left atrial samples (**Table S8**). The most significant variant in the combined ancestry analysis (rs3801995) was in moderate LD (CEU  $r^2=0.56$ ) with the most significant variants in previous GWAS of PR-interval;<sup>10, 15-17</sup> which are in perfect LD (rs3807989 and rs11773845, CEU  $r^2=1$ ) with each other. The most significant variant in Europeans (rs13242816) was not in LD (CEU  $r^2=0.11$ ) with previously identified variants at this locus. A third locus was identified on chromosome 2 and the top hit (rs11689011) was intronic to the gene *EPAS1*. On chromosome 1, we identified a locus (most significant SNP rs562408) surrounding the gene *SSBP3*. The lead SNP and five other variants were significant eQTLs for *SSBP3* in European American left atrial samples (**Table S8**) and the lead SNP in Europeans was the strongest cis eQTL for this gene; similar results were found in the GTEx database (**Table S9**). Another locus was located on chromosome 4 with rs2285703 intronic to the gene *CAMK2D*.

Last, in the meta-analysis of European and African-American ancestries combined, we identified an additional genetic locus associated with P-wave duration on chromosome 3. The top variant, rs1467026, was located upstream of *CAND2* and was in moderate LD (CEU  $r^2 = 0.67$ ) with a variant associated with AF by Sinner et al. in 2014.<sup>18</sup> Rs1467026 has been shown to be a significant eQTL for *CAND2* ( $p=7.5 \times 10^{-27}$ ) in skeletal muscle in the GTEx database (**Table S9**), though this was not found to be a significant eQTL in left atrial tissues.

### *Genetic loci previously associated with electrocardiographic traits.*

We identified an association between P-wave duration and the *SCN5A* region in both Europeans and African Americans. The most significant European variant (rs41312411) at the *SCN5A* locus was in strong LD with the most significant variant (rs11708996, CEU  $r^2=0.94$ ) from the first PR interval GWAS by Pfeufer et al.<sup>15</sup> and in moderate LD with a variant associated with both PR-segment<sup>10</sup> and PR-interval<sup>19</sup> (rs6599222, CEU  $r^2=0.55$ ). In African Americans, we replicated the association between SNP rs3922844

## Genetic Loci Associated with ECG P-wave Indices

and the PR-interval observed in two previous GWAS in African Americans;<sup>19, 20</sup> this variant also has been associated with the PR-segment in Europeans.<sup>10</sup>

In individuals of European ancestry, we identified an additional association signal in the adjacent *SCN10A* gene. The most significant variant at the *SCN10A* locus (rs6790396) was in strong LD with the following previously identified variants; 1) rs6800541 (CEU  $r^2=1$ ), which was previously associated with PR-interval;<sup>15</sup> 2) rs6795970 (CEU  $r^2=0.97$ ), which was associated with both PR-interval and P-wave duration in an Indian Asian population,<sup>21</sup> and PR-interval in Europeans;<sup>16</sup> 3) rs6801957 (CEU  $r^2=0.97$ ), which was associated with PR-interval,<sup>20, 22</sup> PR-segment,<sup>10</sup> and P-wave duration.<sup>10</sup> There also was strong (CEU  $r^2=0.87$ ) and moderate (YRI  $r^2=0.51$ ) LD between our lead *SCN10A* variant and a variant associated with PR-interval in African Americans (rs6798015).

In both Europeans and African Americans, we identified a locus with the most significant variant located intronic to the gene *TBX5*. The lead variants in all ancestry analyses were significant eQTLs for *TBX5* in both European American and African American left atrial samples (**Table S8**). The most significant SNP in the African American ancestry group (rs1895582) was also an eQTL for *TBX5* in left ventricular tissue in GTEx (**Table S9**). The same SNP, rs1895582, was in strong LD (YRI  $r^2=0.81$ ) with a variant previously associated with PR-interval in African Americans. The most significant variant (rs148020424) in Europeans was surrounded by a ~300 kb LD block, bordered by two recombination hotspots. There were no LD data available between rs148020424 and previously associated SNPs at this locus. However, the most significant variant in our combined ancestry analysis (rs7312625) was previously associated with PR-interval in African Americans. Rs7312625 was in strong LD (YRI  $r^2=0.81$ ) with the most significant variant in African Americans in our study and in moderate LD with rs1895585 (YRI  $r^2=0.78$ ), the most significant variant in the *TBX5* region from a previous GWAS on PR-interval in

## Genetic Loci Associated with ECG P-wave Indices

African Americans.<sup>20</sup> SNP rs7312625 was also in moderate LD with rs3825214 (CEU  $r^2=0.76$ ), an intronic variant in *TBX5* previously associated with heart rate in Europeans.<sup>16</sup>

We identified an association between P-wave duration and variant rs452036 on chromosome 14, in a region encompassing *MYH6* and *MYH7*, and the same variant was also significantly associated with P-wave terminal force in our study. Rs452036 was the most significant variant in a GWAS for heart rate by Eijgelsheim et al. in 2010<sup>23</sup> and was in strong LD with rs365990 (CEU  $r^2=0.96$ , YRI  $r^2=1$ ), which has been associated with heart rate in studies of both European and African American ancestry.<sup>16, 23-25</sup> Rs452036 was not in LD (CEU  $r^2=0.16$ ) with the heart rate-associated SNP rs223116, which is intronic to *MYH7* and thus most likely represents an independent locus for heart rate.<sup>23</sup>

### P-wave terminal force

In the primary analysis, we identified six loci associated with P-wave terminal force. The most significantly associated variants are listed in **Table S10** and **Table S11**; **Figure 3** depicts the Manhattan plots, stratified by ancestry. The Manhattan plot for the pooled analysis of Europeans and African Americans is shown in **Figure S1B** and regional plots for all genetic loci significantly associated with P-wave terminal force are shown in **Figure S3**. **Table S12** summarizes the shared associations of genetic loci between the present P-wave terminal force analysis and previous GWAS of PWI.

### *Genetic loci unique to the P-wave terminal force*

In the European study populations, the locus with the strongest association was identified on chromosome 1 (rs12090194), intronic to *KCND3*. Rs12090194 was in weak LD (CEU  $r^2=0.2$ ) with rs2798334, which has been associated with P-wave duration by Verweij and colleagues,<sup>10</sup> thus this locus appears to be specific to P-wave terminal force. A second locus was identified on chromosome 19, with the most significant variant (rs4435363) intronic to the gene *PPP5D1* and downstream of the gene

## Genetic Loci Associated with ECG P-wave Indices

*CALM3*. On chromosome 15q25, the most significant variant was an indel (rs201517563), intronic to *ALPK3*. This variant was not present in the 1000 Genomes database reference and so there were no available LD data. However, LD data were available for the second most significant variant at this locus (rs11073730), which was located between *ALPK3* and *ZNF592*. The 15q25 locus spanned several other genes, including *NMB*, for which there were multiple significant eQTLs in European American left atrial samples (**Table S13**), and rs11073730 was also a significant eQTL for several genes in various tissues in GTEx, including *ALPK3* (**Table S9**). In addition, two perfect proxies were associated with increased expression of *CSPG4P11* in tissue from the tibial nerve ( $p=1.4 \times 10^{-6}$  and  $2.2 \times 10^{-6}$ ) and rs11073730, along with 14 proxies in strong LD ( $r^2 = 0.8-1$ ), was associated with increased expression of *WDR73* in the muscular layer of the esophagus ( $p=3.3 \times 10^{-7}$ - $1.4 \times 10^{-9}$ ). We identified another locus in an intergenic region on chromosome 4, with the most significant variant (rs11099412) ~700 kb downstream of the closest gene *PCDH18*. Finally, a locus was identified on chromosome 6, ~150 kb downstream of the gene *C6orf195*. A strong proxy ( $r^2=0.9$ ) for the most significant variant was previously associated with orthostatic hypotension.<sup>26</sup>

In the African American population, one locus was identified on chromosome 11. The most significant variant (rs10832139) was located 44 kb upstream of the closest gene, *SPON1*.

### *Genetic loci associated with heart rate, P-wave duration and P-wave terminal force*

On chromosome 14, we identified a locus with the most significant variant located intronic to *MYH6*. Rs445754 was in moderate LD with the most significant variant from the P-wave duration analysis in the present study (rs452036, CEU  $r^2=0.7$ ) and the most significant variant from the first GWAS on heart rate (rs365990, CEU  $r^2=0.6$ ).<sup>16</sup> The variant rs365990 has been replicated in three subsequent GWAS of heart rate.<sup>16, 24, 25</sup>

## Genetic Loci Associated with ECG P-wave Indices

### Variance explained by the genetic loci associated with P-wave duration and P-wave terminal force

The most significant SNPs associated with P-wave duration explained 0.08-0.83% of the total variance (**Table S4 and Table S5**), while the variants associated with P-wave terminal force explained 0.09-0.47% of the total variance (**Table S10 and Table S11**).

### P-wave duration and P-wave terminal force are genetically associated

After LD-clumping ( $r^2 > 0.5$ ), 199 significant SNPs remained from the P-wave duration analysis and 85 significant SNPs remained from the P-wave terminal force analysis, which were included in the respective GRS. The P-wave terminal force GRS was associated with measured P-wave duration ( $\beta = 0.00095$ ;  $SE = 0.00014$ ;  $p = 2.5 \times 10^{-11}$ ) and the P-wave duration GRS was associated with measured P-wave terminal force ( $\beta = 8.33$ ;  $SE = 1.64$ ,  $p = 3.6 \times 10^{-7}$ ). Using lower  $r^2$  thresholds for the LD clumping produced similar results, with even lower p-values (**Supplemental Results**). One unit increase in the genetically determined P-wave terminal force was associated with a 0.00095 ms increase in P-wave duration, whereas one unit increase in the genetically determined P-wave duration was associated with an 8.33  $\mu\text{V} \times \text{ms}$  increase in P-wave terminal force. The estimated percentage of total variance of the measured P-wave terminal force explained by the P-wave duration GRS is 0.07% and conversely, the estimated fraction of the total variance of the measured P-wave duration explained by the P-wave terminal force GRS is 0.1%.

## DISCUSSION

In the current work, we sought to characterize the genetic basis of the P-wave and to integrate our findings with prior association studies of electrocardiographic traits quantifying atrial conduction. Our results reveal both unique and overlapping genetic loci for the different phases of atrial conduction (**Figure 4**). As expected, the phenotypic variance explained by the causal SNP from each genetic locus

## Genetic Loci Associated with ECG P-wave Indices

associated with P-wave duration and P-wave terminal force was small (0.08-0.83%). Similar estimates have been reported for common variants associated with other continuous traits, such as human height and BMI, and reflect the often highly polygenic nature of common traits which are thus explained by a large number of variants with small effect sizes. The larger percentage of variance explained by the variants identified in the African American ancestry groups is caused by the large effect sizes of these SNPs.

In the Supplemental Discussion, we have provided a detailed description of the loci that are unique to either P wave duration or P-wave terminal force and we have discussed the broad contribution of the sodium channels, caveolin, and TBX5 to atrial conduction.

### Interrelationship between P-wave duration and P-wave terminal force

P-wave duration and P-wave terminal force are both endophenotypes for AF; however, they represent different aspects of atrial conduction. P-wave duration reflects the depolarization of both right and left atrium, while P-wave terminal force, which is calculated from the last half portion of the P-wave, more specifically represents left atrial activation. Soliman et al. have shown that the mutual PWI correlations are weak to moderate in a study of more than 15,000 individuals.<sup>4</sup> In particular, P-wave duration and P-wave terminal force displayed the weakest correlation (Pearson's  $R$  0.08-0.29). In this study, we found that the genetically determined P-wave duration is highly associated with the measured P-wave terminal force and vice versa, but that the percentage of explained variance between the two PWI was small. Thus, we had expected to identify genetic loci uniquely associated with either right or left atrial activation, in addition to some overlap of genetic variants associated with both phenotypes in the present study. We did discover one genetic locus associated with both P-wave duration and P-wave terminal force, encompassing the gene *MYH6*. The 14q11 locus includes *MYH6*, the gene encoding myosin heavy chain 6 protein, which is important in actin and calmodulin binding in the sarcomere of



## Genetic Loci Associated with ECG P-wave Indices

striated muscles. The most significant variant at this locus, rs452036, and its proxy rs365990 (CEU  $r^2=0.96$ ) have been associated with heart rate in multiple GWAS<sup>16, 23, 24</sup> and most likely represent the same locus. *MYH6* has been implicated in congenital heart defects,<sup>27</sup> dilated and hypertrophic cardiomyopathy,<sup>28, 29</sup> AF,<sup>30</sup> atrial septal defects,<sup>31, 32</sup> and sick sinus syndrome.<sup>33</sup> All of these conditions may affect inter- or intra-atrial conduction and modify P-wave duration.

According to reports that P-wave terminal force is a predictor of left atrial enlargement,<sup>8</sup> one might have expected to see a predominance of structurally related genes associated with P-wave terminal force and a preponderance of genes involved in electrical properties of the atrium associated with P-wave duration. Indeed, some of the genes identified in the analyses of P-wave duration are more likely to produce “electrical” disturbances (*SCN5A*, *SCN10A*, *HCN1*, *CAV1/CAV2*, *EPAS1*, *CAND2*, and *CAMK2D*), while a selection of the genes identified in the analyses of P-wave terminal force are more likely to cause “structural” changes (*TBX5*, *ALPK3/NMB*, *MYH6*, and *SPON1*). However, recent studies have reported that P-wave terminal force was not associated with left atrial enlargement on echocardiography<sup>34</sup> or computed tomography volume measurement.<sup>35</sup> In a sample of 275 middle-aged Americans referred to cardiovascular magnetic resonance imaging between 2001-2004, electrocardiographic P-wave terminal force (negative P terminal force in V1 > 0.04 s x mm) had high specificity (88%), but low sensitivity (37%) and positive predictive value (47%) for left atrial enlargement, compared to volumetric assessment of left atrial size.<sup>36</sup>

### Strengths and limitations

Strengths of our study include the substantial power obtained by including more than 44,000 individuals from established studies with carefully phenotyped participants. Further, the PWI measurements were all measured digitally and calculated automatically. Last, since AF is thought to originate predominantly from the left atrium and pulmonary veins, evaluation of eQTLs in a large sample of left atrial samples

## Genetic Loci Associated with ECG P-wave Indices

provides a suitable model for functional interrogation of the genetic variants associated with AF in our study.

Our study was also subject to a number of limitations. First, both P-wave duration and P-wave terminal force measurements are less reproducible over time than PR interval, P-wave axis and P wave area.<sup>37</sup> Second, one could consider adjusting the P value for significant results for the two traits considered ( $5 \times 10^{-8} / 2 = 2.5 \times 10^{-8}$ ); however, the P-wave duration and P-wave terminal force are only weakly correlated so we used a conventional threshold for genome wide significance of  $5 \times 10^{-8}$ . Third, associations with p-values in the range of  $1 \times 10^{-8}$  to  $10^{-10}$  will require replication in independent cohorts. Fourth, although we included individuals of both European and African ancestry, our findings may not be generalizable to other ancestries. Fifth, the analyses in African Americans were underpowered compared with our analyses in European-ancestry individuals and require replication as other studies become available in the future. Last, it is important to remember that we have identified markers of genetic regions involved in the P-wave indices. The genes outlined above are the genes closest to the most significant variant in each genetic region, yet it remains unclear whether they truly are the effector genes. Other genes in the same region or even more distantly located genes might actually be the causal genes. Future studies including targeted mapping of these loci and functional analyses of the implicated genes may shed light on the underlying biology.

### Conclusion

We report 15 genetic loci associated with P-wave duration and P-wave terminal force, four of which harbor significant eQTLs in left atrial tissue. In the context of the existing knowledge of genetic factors for atrial conduction, a genetic architecture is emerging, in which specific genetic regions influence different phases of atrial conduction. Functional characterization of the novel loci may clarify the

## Genetic Loci Associated with ECG P-wave Indices

underlying pathophysiology and aid the discovery of targets for new drugs, which may modulate atrial conduction and thus potentially treat a range of arrhythmia disorders.

### Supplemental data description

The supplemental data include additional methods, results, discussion, funding information, eleven supplemental tables, three supplemental figures, and supplemental references.

### Web Resources

The Genotype-Tissue Expression (GTEx) browser, <http://www.gtexportal.org>

SNAP proxy search, <http://www.broadinstitute.org/mpg/snap/>

Locus zoom, <https://statgen.sph.umich.edu/locuszoom/>

### Conflict of Interest Disclosures

Dr. Ellinor is the PI on a grant from Bayer HealthCare to the Broad Institute focused on the genetics and therapeutics of atrial fibrillation. The remaining authors have no disclosures.

## Genetic Loci Associated with ECG P-wave Indices

### References

1. Bayliss W, Starling E. On the electrical variations of the heart in man. *J Physiol.* 1891;13:lviii-lix.
2. Sanders P, Morton JB, Davidson NC, Spence SJ, Vohra JK, Sparks PB, et al. Electrical remodeling of the atria in congestive heart failure: electrophysiological and electroanatomic mapping in humans. *Circulation.* 2003;108:1461-1468.
3. Sanders P, Morton JB, Kistler PM, Spence SJ, Davidson NC, Hussin A, et al. Electrophysiological and electroanatomic characterization of the atria in sinus node disease: evidence of diffuse atrial remodeling. *Circulation.* 2004;109:1514-1522.
4. Soliman EZ, Prineas RJ, Case LD, Zhang Z-m, Goff DC. 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.
5. 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.e51.
6. Robitaille GA, Phillips JH. An analysis of the P wave in patients with transient benign atrial fibrillation. *Dis Chest.* 1967;52:806-812.
7. Nielsen JB, Kuhl JT, Pietersen A, Graff C, Lind B, Struijk JJ, et al. P-wave duration and the risk of atrial fibrillation: Results from the Copenhagen ECG Study. *Heart Rhythm.* 2015;12:1887-1895.
8. Hazen MS, Marwick TH, Underwood DA. Diagnostic accuracy of the resting electrocardiogram in detection and estimation of left atrial enlargement: an echocardiographic correlation in 551 patients. *Am Heart J.* 1991;122:823-828.
9. Smith JG, Lowe JK, Kovvali S, Maller JB, Salit J, Daly MJ, et al. Genome-wide association study of electrocardiographic conduction measures in an isolated founder population: Kosrae. *Heart Rhythm.* 2009;6:634-641.
10. Verweij N, Mateo Leach I, van den Boogaard M, van Veldhuisen DJ, Christoffels VM, Hillege HL, et al. Genetic determinants of P wave duration and PR segment. *Circ Cardiovasc Genet.* 2014;7:475-481.
11. Leening MJG, Kavousi M, Heeringa J, van Rooij FJA, Verkroost-van Heemst J, Deckers JW, et al. Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study. *Eur J Epidemiol.* 2012;27:173-185.
12. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N, et al. Common genetic variants influence human subcortical brain structures. *Nature.* 2015;520:224-229.
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, 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.
14. Johnson T. *Efficient calculation for multi-SNP genetic risk scores.*: The Comprehensive R Archive Network 2013.
15. Pfeufer A, Noord Cv, Marcianti KD, van NC, Marcianti KD, Arking DE, et al. Genome-wide association study of PR interval. *Nat Genet.* 2010;42:153-159.
16. Holm H, Gudbjartsson DF, Arnar DO, Thorleifsson G, Thorgeirsson G, Stefansdottir H, et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nat Genet.* 2010;42:117-122.
17. Hong K-W, Lim JE, Kim JW, Tabara Y, Ueshima H, Miki T, et al. Identification of three novel genetic variations associated with electrocardiographic traits (QRS duration and PR interval) in East Asians. *Hum Mol Genet.* 2014;23:6659-6667.

## Genetic Loci Associated with ECG P-wave Indices

18. Sinner MF, Tucker NR, Lunetta KL, Ozaki K, Smith JG, Trompet S, et al. Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation*. 2014;130:1225-1235.
19. Smith JG, Magnani JW, Palmer C, Meng YA, Soliman EZ, Musani SK, et al. Genome-wide association studies of the PR interval in African Americans. *PLoS genetics*. 2011;7:e1001304.
20. Butler AAM, Yin X, Evans DDS, 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.
21. Chambers JC, Zhao J, Terracciano CM, Bezzina CR, Zhang W, Kaba R, et al. Genetic variation in SCN10A influences cardiac conduction. *Nat Genet*. 2010;42:149-152.
22. Sano M, Kamitsuji S, Kamatani N, Hong K-W, Han B-G, Kim Y, et al. Genome-wide association study of electrocardiographic parameters identifies a new association for PR interval and confirms previously reported associations. *Hum Mol Genet*. 2014;23:6668-6676.
23. Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, de Bakker PIW, Müller M, Morrison AC, et al. Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet*. 2010;19:3885-3894.
24. den Hoed M, Eijgelsheim M, Esko T, Brundel BJM, Peal DS, Evans DM, et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet*. 2013;45:621-631.
25. Deo R, Nalls MA, Avery CL, Smith JG, Evans DS, Keller MF, et al. Common genetic variation near the connexin-43 gene is associated with resting heart rate in African Americans: a genome-wide association study of 13,372 participants. *Heart Rhythm*. 2013;10:401-408.
26. Hong K-W, Kim SS, Kim Y. Genome-wide association study of orthostatic hypotension and supine-standing blood pressure changes in two Korean populations. *Genomics & Informatics*. 2013;11:129-134.
27. Granados-Riveron JT, Ghosh TK, Pope M, Bu'Lock F, Thornborough C, Eason J, et al. Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. *Hum Mol Genet*. 2010;19:4007-4016.
28. Carniel E, Taylor MRG, Sinagra G, Di Lenarda A, Ku L, Fain PR, et al. Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation*. 2005;112:54-59.
29. Hershberger RE, Norton N, Morales A, Li D, Siegfried JD, Gonzalez-Quintana J. Coding Sequence Rare Variants Identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 From 312 Patients With Familial or Idiopathic Dilated Cardiomyopathy. *Circ Cardiovasc Genet*. 2010;3:155-161.
30. Weeke P, Muhammad R, Delaney JT, Shaffer C, Mosley JD, Blair M, et al. Whole-exome sequencing in familial atrial fibrillation. *Eur Heart J*. 2014;35:2477-2483.
31. Ching Y-H, Ghosh TK, Cross SJ, Packham EA, Honeyman L, Loughna S, et al. Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet*. 2005;37:423-428.
32. Posch MG, Waldmuller S, Müller M, Scheffold T, Fournier D, Andrade-Navarro MA, et al. Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One*. 2011;6:e28872.
33. Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadóttir HT, Zanon C, et al. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat Genet*. 2011;43:316-320.
34. Petersson R, Berge HM, Gjerdalen GF, Carlson J, Holmqvist F, Steine K, et al. P-wave morphology is unaffected by atrial size: a study in healthy athletes. *Ann Noninvasive Electrocardiol*. 2014;19:366-373.

## Genetic Loci Associated with ECG P-wave Indices

35. Truong QA, Charipar EM, Ptaszek LM, Taylor C, Fontes JD, Kriegel M, et al. Usefulness of electrocardiographic parameters as compared with computed tomography measures of left atrial volume enlargement: from the ROMICAT trial. *J Electrocardiol.* 2011;44:257-264.
36. Tsao CW, Josephson ME, Hauser TH, O'Halloran TD, Agarwal A, Manning WJ, et al. Accuracy of electrocardiographic criteria for atrial enlargement: validation with cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* 2008;10:7.
37. Snyder ML, Soliman EZ, Whitsel EA, Gellert KS, Heiss G. Short-term repeatability of electrocardiographic P wave indices and PR interval. *J Electrocardiol.* 2014;47:257-263.

## Genetic Loci Associated with ECG P-wave Indices

### FIGURE LEGENDS

#### **Figure 1. Calculation of P-wave terminal force.**

Electrocardiogram with the P-wave highlighted in green and calculation of P-wave terminal force depicted in lead V1.

#### **Figure 2. Manhattan plot of meta-analysis results for genome-wide association of maximum P-wave duration for European and African American ancestry.**

**A.** European ancestry meta-analysis results. **B.** African American ancestry meta-analysis results. The dashed horizontal lines represent the genome-wide significance threshold. Genetic loci that reached genome-wide significance are highlighted in red.

#### **Figure 3. Manhattan plot of meta-analysis results for genome-wide association of P-wave terminal force for European and African American ancestry.**

**A.** European ancestry meta-analysis results. **B.** African American ancestry meta-analysis results. The dashed horizontal lines represent the genome-wide significance threshold. Genetic loci that reached genome-wide significance are highlighted in blue.

#### **Figure 4. Integrated view of the anatomical, electrical and genetic architecture of atrial conduction.**

Top: Heart, anterior four-chamber view, with the different phases of atrial conduction color-coded.

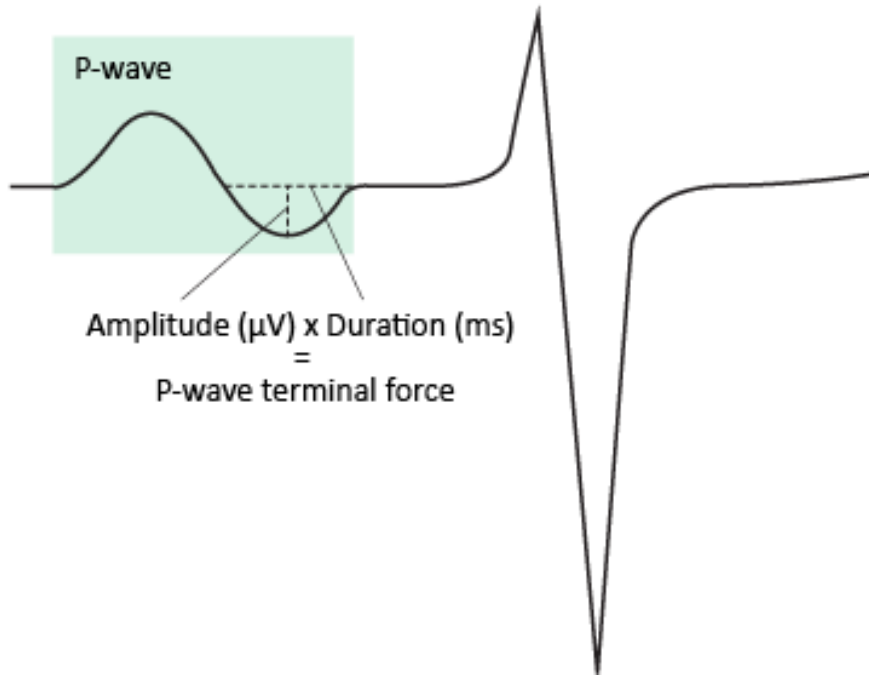
Purple, sinoatrial node activity (heart rate); green, P-wave duration; red, PR-segment; blue, PR-interval; grey, P-wave terminal force; yellow, overall atrial conduction.

Middle: Electrocardiogram corresponding to the different phases of atrial conduction.

Bottom: All identified risk genes for heart rate (R-R), PR-interval, and P-wave indices. For each trait, only genes uniquely associated with that trait are listed. Genes associated with multiple electrocardiographic traits are listed under multiple atrial conduction traits. Red font, novel loci in the present study; green font, loci identified in both the present study and previous literature; black font, loci only identified in previous studies; SA, Sinoatrial; AV, atrioventricular; RA, right atrium; LA, left atrium; RV, right ventricle; LV, left ventricle; PTF, P-wave terminal force.

**FIGURES**

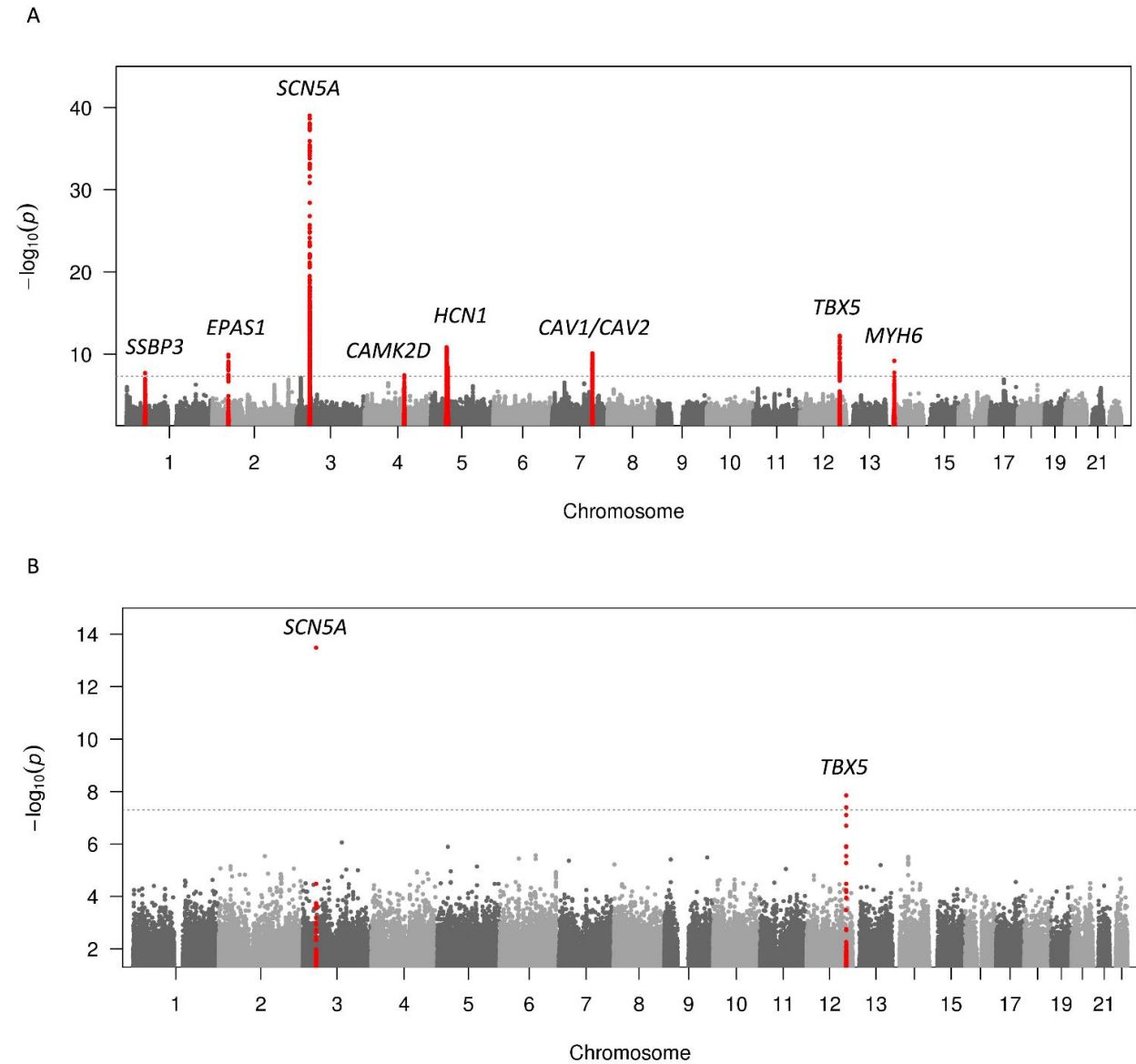
**Figure 1. Calculation of P-wave terminal force.**





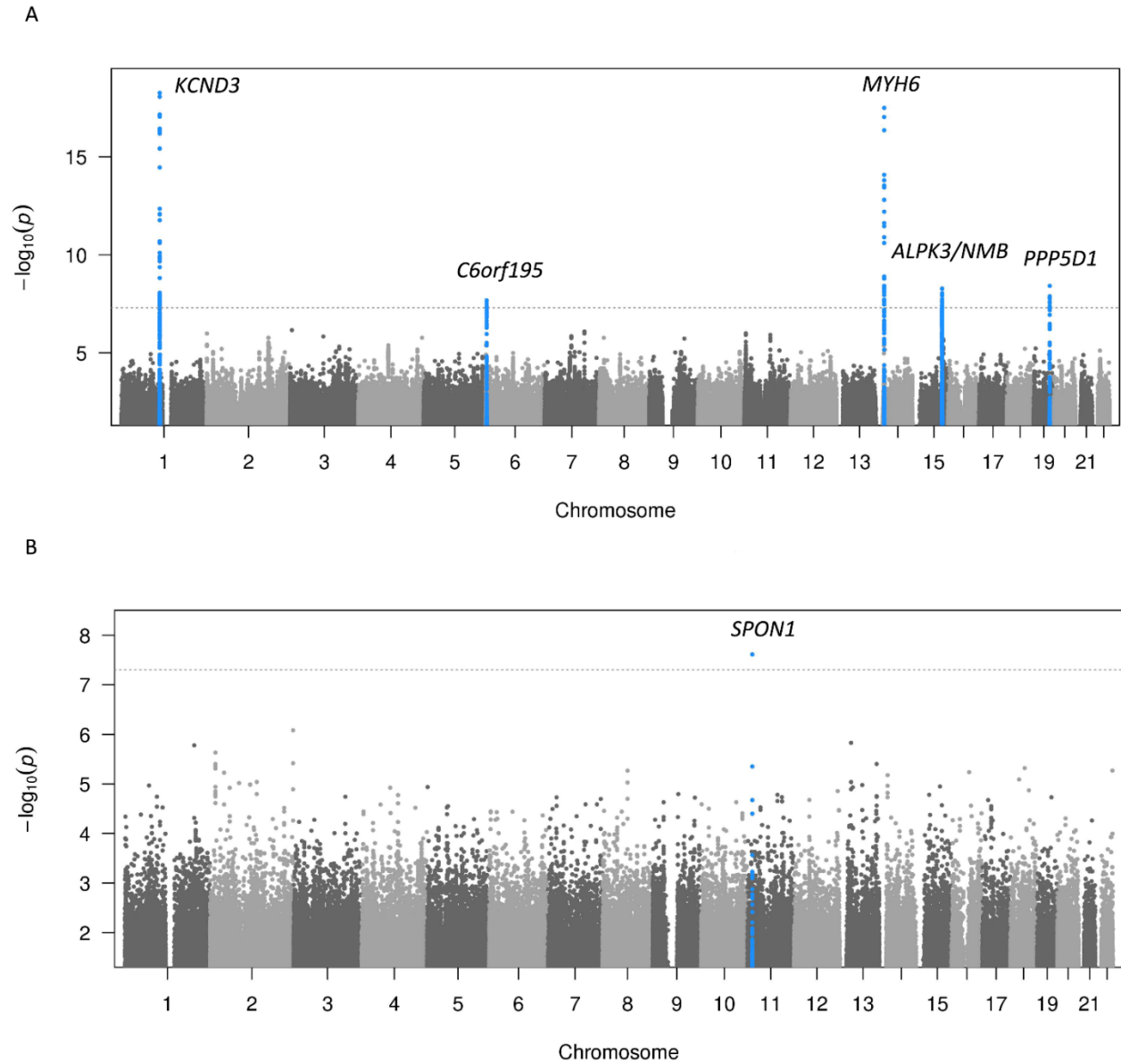
## Genetic Loci Associated with ECG P-wave Indices

Figure 2. Manhattan plot of meta-analysis results for genome-wide association of maximum P-wave duration.



### Genetic Loci Associated with ECG P-wave Indices

Figure 3. Manhattan plot of meta-analysis results for genome-wide association of P-wave terminal force.



## Genetic Loci Associated with ECG P-wave Indices

Figure 4. Integrated view of the anatomical, electrical and genetic architecture of atrial conduction.

