Genome-wide association analysis identifies multiple loci related to resting heart rate


1Department of Epidemiology, 2Department of Internal Medicine, 3Department of Medical Informatics and 4Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands, 5Center for Human Genetic Research, Cardiovascular Research Center and 6Cardiology Division, Massachusetts General Hospital, Boston, MA, USA, 7Program in Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA, 8National Heart, Lung and Blood Institute’s Framingham Heart Study, Framingham, MA, USA, 9Division of Cardiology, Department of Medicine, School of Medicine, 10Cardiovascular Health Research Unit, Metropolitan Park East Tower, 11Department of Medicine, 12Department of Biostatistics, 13Department of Epidemiology and 14Department of Health Services, University of Washington, Seattle, WA, USA, 15Division of Genetics, Department of Medicine, Brigham’s and Women’s Hospital, Boston, MA, USA, 16Institute of Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany, 17Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology and 18Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany, 19Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA, 20Icelandic Heart Association Research Institute, Kopavogur, Iceland, 21Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy, 22Department of Internal Medicine B, 23Interfaculty Institute for Genetics and Functional Genomics and

‡These authors contributed equally to this work.

∗To whom correspondence should be addressed at: Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA, Rotterdam, The Netherlands. Tel: +31 107043489; Fax: +31 107044657; Email: b.stricker@erasmusmc.nl (B.H.Ch.S.); Framingham Heart Study, 73 Mt Wayte Avenue, Suite No. 2, Framingham, MA 01702-5827, USA. Tel: +1 5089353435; Fax: +1 5086261262; Email: odonnellc@nihbi.nih.gov (C.J.O.).

†These authors contributed equally to this work.
Higher resting heart rate is associated with increased cardiovascular disease and mortality risk. Though heritable factors play a substantial role in population variation, little is known about specific genetic determinants. This knowledge can impact clinical care by identifying novel factors that influence pathologic heart rate states, modulate heart rate through cardiac structure and function or by improving our understanding of the physiology of heart rate regulation. To identify common genetic variants associated with heart rate, we performed a meta-analysis of 15 genome-wide association studies (GWAS), including 38,991 subjects of European ancestry, estimating the association between age-, sex- and body mass-adjusted RR interval and 2.5 million markers. Results with \( P < 5 \times 10^{-8} \) were considered genome-wide significant. We constructed regression models with multiple markers to assess whether results at less stringent thresholds were likely to be truly associated with RR interval. We identified six novel associations with resting heart rate at six loci: 6q22 near \( \text{GJA1} \); 14q12 near \( \text{MYH7} \); 12p12 near \( \text{SOX5} \), \( \text{c12orf67} \), \( \text{LRMP} \), and \( \text{UfSp1} \); 7q22 near \( \text{SLC12A9} \) and \( \text{UfSp1} \); and 11q12 near \( \text{FADS1} \). Associations at 6q22 400 kb away from \( \text{GJA1} \), at 14q12 \( \text{MYH6} \) and at 1q32 near \( \text{CD34} \) identified in previously published GWAS were confirmed. In aggregate, these variants explain \( 0.7\% \) of RR interval variance. A multivariate regression model including 20 variants with \( P < 10^{-8} \) essentially doubled the explained variance to \( 1.6\% \), suggesting that some loci falling short of genome-wide significance are likely truly associated. Future research is warranted to elucidate underlying mechanisms that may impact clinical care.

**INTRODUCTION**

Higher resting heart rate is associated with increased risk of cardiovascular disease (1), cardiovascular mortality (2,3), including sudden death (4), and all-cause mortality independent of traditional risk factors (5–9). However, it is not known whether heart rate directly impacts mortality or merely reflects unrecognized subclinical disease (1,7,8). Recently it was shown that physical exercise reduces the increased cardiovascular mortality risk associated with...
higher heart rate (3). This suggests that heart rate is a clinically relevant and potentially modifiable risk factor.

Heart rate is a complex trait, determined by multiple environmental, genetic and other endogenous factors. Heredity plays a substantial role in the inter-individual variation of resting heart rate, accounting for 26–32% of heart rate variation in prior studies (10–13). Twin studies report even higher heritability estimates up to 55–63% (14,15). Candidate gene approaches have identified multiple loci associated with heart rate (12,13,16–20), but the results have been inconsistent and difficult to replicate. Genome-wide genotyping arrays of single-nucleotide polymorphisms (SNPs) assay common variation in the human genome and can identify genetic variants with modest influences on a complex trait such as heart rate, as shown by two recent genome-wide association studies (GWAS) that identified common variation at or near MYH6, GJA1 and CD34 associated with heart rate (21,22). These chromosomal loci identified in an unbiased genome-wide study may represent novel risk factors for cardiovascular disease and mortality. This knowledge may also have an impact on clinical care (i) by identifying novel factors that cause pathologic heart rate states (such as sick sinus syndrome or other arrhythmias), (ii) by identifying factors that influence cardiac structure or function (e.g. stroke volume) and thereby modulate heart rate (since cardiac output = heart rate × stroke volume) or (iii) by improving our understanding of the physiologic basis of heart rate regulation. Altogether this will generate insights into the underlying mechanisms of heart rate as a well-established, but poorly understood, risk indicator for cardiovascular disease and mortality.

To identify additional genetic determinants of heart rate, we performed a meta-analysis of GWAS of resting heart rate, measured as the RR interval on the electrocardiogram (ECG) in 38,991 individuals of European ancestry derived from 15 studies in the RRGEN consortium.

RESULTS

There were 38,991 individuals available for genotype–phenotype association analysis after exclusions. Subject characteristics are shown in Table 1. While cohort-specific quantile–quantile plots of P-value distributions approximated expectations under the null, the meta-analysis of all results showed a clear excess of low P-values (Supplementary Material, Fig. S1). Study-specific genomic inflation lambda values ranged from 0.98 to 1.05, suggesting that population stratification or other technical artifacts were minimal (Supplementary Material, Table S1).

Meta-analysis of results from all studies resulted in 156 SNPs reaching the pre-specified genome-wide statistical significance level (P < 5 × 10⁻⁸; Supplementary Material, Table S2) before applying post-meta-analysis genomic control. In total, nine independent genome-wide significant signals were observed across seven chromosomal loci (Table 2, Fig. 1), with all but one locus harboring multiple SNPs reaching the genome-wide significance threshold (Supplementary Material, Table S2; Fig. 2A–E).

Genome-wide significant associations were observed at 6q22 nearest to GJA1; 14q12 near MYH6 and MYH7; 12p12 near SOX5, c12orf67, BCAT1, LRMP and CASC1; 6q22 near SLC35F1, 6orf204 and PLN (>3 Mb away from 6q22 GJA1); 7q22 near SLC12A9; 11q12 near FADS1; and 1q32 near CD34. The genomic inflation lambda value of the meta-analysis was 1.05. After applying post-meta-analysis genomic control, the signal at 1q34 near CD34 and a second independent signal at 6q22 GJA1 lost genome-wide significance.

The strongest association was observed for an intergenic SNP 370 kb upstream of GJA1 at 6q22 rs9398652 [minor allele frequency (MAF) = 0.10] with a 12.6 ms shorter RR interval per minor A allele, which is equivalent to a 0.95 bpm higher heart rate based on the baseline mean heart rate of 66.8 bpm, across all studies (P = 8.0 × 10⁻¹⁶, Pmeta-gc = 3.8 × 10⁻¹⁵; Table 2, Fig. 2A). Cho et al. (21) observed genome-wide significant association between rs12110693 and pulse rate in an Asian population-based GWAS. This SNP is in perfect linkage disequilibrium (r² = 1) with rs939862 in both Caucasian and Asian HapMap reference populations and thus reflects the same signal 370 kb from GJA1. A second SNP only 8 kb away from GJA1 also reached genome-wide significance pre-genomic control, rs11154022 (MAF = 0.33) with a 5.8 ms longer RR interval (0.43 bpm lower heart rate) per A allele (P = 3.5 × 10⁻⁸, Pmeta-gc = 7.2 × 10⁻⁸; Fig. 2B, Table 2). This SNP had low correlation with rs9398652 (r² = 0.006 in HapMap CEU), suggesting a novel-independent association signal. Since HapMap is limited to 90 subjects, we assessed the linkage disequilibrium in our data. All observed r² values ranged between 0.0001 and 0.004, which is lower than seen in the HapMap CEU reference population. Conditional analysis confirmed that these two signals are independent with Pconditional = 2.4 × 10⁻¹¹ and Pconditional = 3.3 × 10⁻³, respectively, for rs9398652 and rs1154022 in the subset (n ≤ 33,846) used for this analysis.

The second locus with two signals in low correlation reaching genome-wide significance is located on chromosome 14q12. The strongest association at this locus was observed for rs452036 located in intron 19 of MYH6 [MAF = 0.36, with a 7.8 ms shorter RR interval (0.58 bpm higher heart rate) per C allele, P = 8.1 × 10⁻¹⁵, Pmeta-gc = 3.8 × 10⁻¹⁴, Fig. 2C]. This replicates the finding by Holm et al. (22) who previously described an association between rs452036 and heart rate. The non-synonymous coding variant rs365990, the gene product, is a possible functional variant since it showed strong correlation with rs452036 (Table 2). A second SNP located near MYH7 rs223116 (MAF = 0.24), associated with a 7.4 ms shorter RR interval per A allele (0.55 bpm higher heart rate, P = 1.1 × 10⁻⁸, Pmeta-gc = 2.5 × 10⁻⁸, Fig. 2D) was in low correlation with rs452036 (r² = 0.08 in HapMap CEU) and reflects a novel association. We observed r² values similar to HapMap CEU with values ranging from 0.03 to 0.08. Conditional analysis confirmed the presence of two independent signals with Pconditional = 7.9 × 10⁻¹⁴ and Pconditional = 3.7 × 10⁻³ for rs452036 and rs223116, respectively.

For the other five loci, only a single signal of association met our genome-wide significance threshold, meaning that all other genome-wide significant SNPs have an r² > 0.1 in HapMap CEU to the most significant SNP at that locus. For these loci,
the MAF of the index SNPs ranged from 15 to 50%, effect sizes ranged between 5.4 and 8.6 ms and \( P_{\text{meta-ge}} \) ranged between \( 2.1 \times 10^{-9} \) and \( 1.6 \times 10^{-10} \) (Table 2, Fig. 2C–H). The \( CD34 \) locus lost genome-wide significance upon meta-analytic genomic control but replicates the association reported by Cho et al. (21). Of these five loci, only the index SNP at \( 7q22 \) \( SLC12A9 \) shows strong correlation with a non-synonymous coding SNP. This coding SNP (rs12666989, \( r^2 = 0.88 \) to rs314370 in HapMap CEU) results in a Leucine coding SNP (rs12110693, \( r^2 = 0.88 \) to rs4314370 in HapMap CEU) results in a Leucine > Valine substitution at amino acid position 41 in the \( UfSp1 \) gene product.

Additionally, an association result that just missed our genome-wide significance threshold (\( P = 5.2 \times 10^{-8}, P_{\text{meta-ge}} = 1.1 \times 10^{-7} \)) was observed for a non-synonymous coding SNP in \( CCDC141 \) at 2q31. This SNP, rs17362588 (MAF = 0.12, with a 8.3 ms shorter RR interval per A allele), results in an amino acid substitution at position 360 (Tryptophan > Arginine) of the encoded protein.

**Evidence for additional causal loci not reaching genome-wide significance**

We used polygenic modeling methods to quantify the genetic variance explained and to indicate if loci falling short of genome-wide significance are likely to harbor additional variants that influence heart rate. The explained variance of resting heart rate in RS-II (first extended Rotterdam Study cohort; \( n = 1589 \)) was \( \sim 0.7\% \) when the score was calculated based on genome-wide significant signals. Inclusion of 20 independent variants with \( P < 1 \times 10^{-5} \) resulted in the maximal proportion of explained variance of 1.6% (Supplementary Material, Fig. S2). These 20 variants included signals from the 7 genome-wide significant loci and signals from an additional 12 loci, including several loci with genes of potential cardiac relevance (see Supplementary Material, Table S3, for full list).

**DISCUSSION**

The application of genome-wide association methods in a large sample of subjects of European ancestry identified common variants at multiple loci associated with inter-individual variation in resting heart rate. Genetic determinants of heart rate could alter the function of the sinus node (the dominant pacemaker in the normal heart) either directly through altered pacemaking activity (23) or indirectly through sympathetic or parasympathetic inputs to the heart. We were therefore encouraged to find that genes at some of the loci identified here are cardiac ion channels or their regulatory proteins. Besides a direct effect on sinus node function, effects on cardiac structure—either developmental or through remodeling—and function could underlie the observed associations.

The most strongly related locus on chromosome region \( 6q22 \) included two independent association signals 8 and 370 kb away from \( GJA1 \). The latter finding is in line with a recent GWAS that described an association between rs12110693 and pulse rate in an Asian sample (21). rs9398652 described in the present study is in perfect linkage disequilibrium (\( r^2 = 1 \)) with rs12110693 in both the Caucasian and Asian HapMap reference populations and thus reflects the same signal 370 kb from \( GJA1 \). \( GJA1 \) encodes Cx43, a connexin family protein and a major component of the cardiac gap junction, crucial in electrical coupling of myocytes (24). Mutations in \( GJA1 \) cause a Mendelian inherited hypoplastic left heart syndrome (25). To our knowledge, a role for Cx43 in pacemaker function in the adult sinus node has not been reported.

The 14q12 \( MYH6 \) locus was also associated with resting heart rate in the RRGEN study. Its gene product, the myosin heavy chain-6 protein, is a component of the hexameric myosin protein. Mutations in \( MYH6 \) have been related to Mendelian forms of hypertrophic cardiomyopathy (26), atrial septal defect (27) and dilated cardiomyopathy (28). The index SNP was strongly correlated with an amino acid-altering common variant in \( MYH6 \). The association of the coding SNP with heart rate raises the possibility that it is the causal variant and \( MYH6 \) the causal gene. This finding replicates the result of Holm et al. (22) who performed a GWAS on heart rate in an Icelandic population-based sample. We do report a
novel-independent signal located near MYH7. Future research is warranted to define the allelic architecture of this locus in relation to heart rate. Of note, cardiac-specific microRNAs encoded within intronic regions of MYH6 (miR-208a) and MYH7 (miR-208b) have regulatory effects on cardiac conduction (29,30).

The locus on chromosome 12p12 includes several genes, but without a clear candidate for association with heart rate in the associated interval. The closest genes are BCAT1, which encodes a cytosolic form of the branched-chain amino acid transaminase enzyme that catalyzes transamination of branched-chain amino acids to their respective alpha-keto acids essential for cell growth and protein synthesis and SOX5, a member of the Sox family (31) and of incompletely understood function. Sox genes play a major role in cell fate modulation through transcriptional activity but without a clear cardiac role for SOX5 (32). The variant at 12p12 we describe here, rs17287293, is a perfect proxy for rs11047543 which was associated with the PR interval (reflecting atrial depolarization duration and atrioventricular nodal conduction time) in a GWAS (33), independent of heart rate, suggesting pleiotropic electrophysiological effects.

The second locus on 6q22 is located near SLC35F1 and PLN, which encodes phospholamban. This locus is located 3-5 Mb away from the loci near GJA1 on 6q22. We have previously described the common variation at this locus to be associated with heart rate corrected QT interval (34–36). The index SNP reported here was in moderate linkage disequilibrium with two SNPs associated with heart rate-adjusted QT interval duration [rs11756438 (35), $r^2 = 0.43$ in HapMap CEU, RR interval $P = 6.5 \times 10^{-6}$ and rs11970286 (34), $r^2 = 0.58$ in HapMap CEU, RR interval $P = 2.7 \times 10^{-8}$], resulting in overlapping signals in independent phenotypes.

In addition to higher heart rate and longer QT interval, this locus has also been associated with decreased end-diastolic left ventricular diameter in the EchoGen Study (37). In humans, the sinoatrial node shows comparable expression of phospholamban compared with atrial myocytes (23). However, basal cAMP-mediated, protein kinase A (PKA)-dependent phosphorylation of phospholamban is elevated in sinoatrial nodal pacemaker cells compared with other cardiac cell types (38). Basal PKA-dependent phosphorylation is obligatory for spontaneous basal pacemaking activity, and graded changes in the phosphorylation of phospholamban cause graded changes of the pacemaker cell basal heart rate (38,39). The location of the top SNP at this locus in an intron of SLC35F1, which is mainly expressed in the brain (40), raises the possibility that causal variation influencing this gene in fact underlies the association at the locus. However, the observation that heart rate, QT interval and left ventricular structure are all associated with genetic variation at this locus makes phospholamban the best candidate in light of its role in excitation–contraction coupling and intracellular calcium signaling, critical to action potential development in both the sinoatrial node and ventricles.

A locus without a clear candidate to explain its strong association with resting heart rate is located on chromosome 7q22 with the index SNP in SLC12A9 encoding a cation-CI–co-transporter-interacting protein. However, this SNP was strongly correlated with genome-significant SNPs in TRIP6 encoding a thyroid receptor-interacting protein, AChE, which encodes acetylcholinesterase and a non-synonymous-coding SNP in Usp1. This gene encodes an ubiquitin-fold modifier protease (41,42).

The exact same signal that we report from the FADS1 locus on 11q12 was previously associated with cholesterol levels.
The direct product of the reaction catalyzed by FADS1 is arachidonyl-CoA, which has been shown to release Ca\textsuperscript{2+} from the sarcoplasmic reticulum (48).

The 1q32 CD34/C1orf132 locus has no clear potential mechanism through which it is related to heart rate and it loses genome-wide significance after applying post-meta-analysis genomic control. However, it is likely to be truly associated with pulse rate since it was previously identified in a non-Caucasian sample (21), and our study thus represents a replication of this finding in a different ethnic group.

The 2q31 CCDC141 locus just missed our genome-wide significance threshold, but the most significant SNP is a non-synonymous coding variant within C\textsuperscript{DC}CDC141, a coiled-coil domain containing protein of unknown function. Interestingly, the nearby TTN gene encodes Titin, which is expressed in cardiac and skeletal myocytes (49) and plays a key role in muscle assembly, force transmission and maintenance of resting tension (50).

In additional analysis, we have shown that based on the results of this GWAS, the explained variance could be increased to 1.6% with the inclusion of additional signals with \( P < 10^{-5}\). The goal of this analysis was to describe whether additional variants associated with heart rate are likely to be present. Although true positives are among these loci, we cannot discriminate them from the false-positives. We did observe that the loci contributing to the maximal explained variance include loci of specific interest. Three loci that stand out are 3q26 near GNB4, 12p13 near C\textsuperscript{AC}NA1C and 14q11 near PRKD1. G\textsuperscript{NB}4 encodes G\textsubscript{B}4 known to influence G-protein-activated inwardly rectifying K+ channels (GIRK) that play an important role in heartbeat regulation (51–54) and are activated on binding of acetylcholine to the muscarinic M2-receptor present in the sinoatrial node (23,54). C\textsuperscript{AC}NA1C encodes the alpha-1 subunit of the voltage-dependent calcium channel and is related to Timothy’s syndrome including, among other traits, prolonged QT interval, high arrhythmia risk and with a single observation of \textit{in utero} bradycardia (55). Lastly, the PRKD1 gene product is relevant for calcium-/calmodulin-dependent kinase affecting cardiac remodeling and contraction (56). Validation studies in independent samples have to indicate whether these loci are truly associated with heart rate.

**Strengths and limitations**

The large sample derived from several population-based cohort studies allowed us to identify common variants with modest effects. In addition, these cohort studies have extensive data on covariates and disease status that allowed us to harmonize exclusion criteria and phenotype modeling prior to analyzing the data.

Additional signals may have been missed due to random sampling variation, restriction to autosomal SNPs, poor coverage of certain genomic regions or rare alleles by the genotyping platforms used or a lack of power to detect even smaller effects. Lastly, heart rate is a very dynamic trait. Strong environmental influences such as chronic physical activity or training as well as variability in the time at rest or posture at the moment of measurements and other factors, such as anemia or anxiety, which we have not accounted for, would add noise to the phenotype, which would be expected to bias our study toward the null but not toward the false inference of association.

**Conclusion**

RRGEN identified nine signals at seven loci at which common genetic variation is associated with resting heart rate. Six of these signals at six loci are novel, while the other three signals replicate previous findings from GWAS. Several of these loci include genes encoding ion channel regulator proteins with known involvement in heart rate regulation, or proteins with cardiovascular relevant functions and known associated Mendelian disorders. These variants may impact clinical care by identifying novel factors that influence pathologic heart rate states, are relevant for cardiac structure and thereby modulate heart rate, or by improving our understanding of the physiologic basis of heart rate regulation.
Figure 2. (A–I) Regional association plots covering 1 Mb surrounding the index SNP. Statistical significance of SNPs is shown on the $-\log_{10}(P)$ scale as a function of chromosomal position. The primary SNP is annotated by rs number. Correlation of each plotted SNP to the primary SNP is indicated on a color scale from white (minimal correlation) to red (maximal correlation). Non-synonymous coding SNPs resulting in amino acid changes in the encoded protein are plotted as blue circles. Estimated recombination rates from HapMap (blue lines) and RefSeq gene annotations (green arrows) are shown.
MATERIALS AND METHODS

Study participants

The RRGEN sample consisted of subjects from the five participating studies in the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (57)—comprised of the Age, Gene, Environment Susceptibility (AGES) Study, the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS) and the Rotterdam Study (RS)—as well as the Cooperative Health Research in the Region Augsburg (KORA) study, the SardiNIA study, the Study of Health in Pomerania (SHIP), TwinsUK, Netherlands Study of Depression and Anxiety (NESDA) and three population isolate studies in the European Special Populations Network (EUROSPAN), the Erasmus Rucphen Family (ERF), the South Tyrolean Micro-Isolate (MICROS) and the Orkney Complex Disease Study (ORCADES). For RRGEN, both the baseline Rotterdam Study (RS-I) and first extended cohort (RS-II) were used (58,59). KORA subjects in RRGEN were drawn from the F3 and S4 cohorts.

Individuals were excluded if they were non-Caucasian, had atrial fibrillation, a second- or third-degree atrioventricular block, a pacemaker, a diagnosis of prevalent myocardial infarction or prevalent heart failure, used beta-adrenergic blocking agents, non-dihydropyridine calcium antagonists or digoxin, or had a heart rate <50 (RR interval >1200 ms) or >100 bpm (RR interval <600 ms).

To adhere to STrenghtening the REporting of Genetic Associations studies (STREGA) statement guidelines (60), we include an online supplement with additional information (Supplementary Material). All studies have approval from their institutional review committee, and the subjects of all cohorts provided written informed consent. A more detailed description of cohorts is given in Supplementary Material, Appendix.

RR interval measurement methods

All cohorts recorded 12-lead ECGs from which the RR interval (which equals the inverse heart rate) was measured. For cohort-specific details on RR interval measurement methods, please refer to the Supplementary Material, Appendix.

Genotyping and imputation

Affymetrix and Illumina arrays were used for genotyping. Using genotype information generated on these platforms, all cohorts imputed genotypes for a common set of ∼2.5 million autosomal SNPs based on linkage disequilibrium patterns observed in HapMap CEU reference samples (Utah residents of Northern and Western European descent). The genetic trait analyzed was the imputed allele dosage, a fractional value between 0 and 2, reflecting the estimated number of allele copies of an SNP for each subject. A more detailed description is given in the Supplementary Material, Appendix and Table S1.

Statistical methods

The resting RR interval was adjusted for age, sex and body mass index. For each SNP, we tested the genotype for association with the covariate-adjusted RR interval under an additive genetic model using linear regression models (Supplementary Material, Appendix). We then conducted a fixed-effects, inverse variance weighted meta-analysis using beta estimates and standard errors from each of the cohorts with applying genomic control on a per study basis and additionally post-meta-analysis (Pmeta-q). Genomic control refers to the correction made to the test statistics to account for any inflation of the test statistic distribution, which can result from unaccounted population substructure or other technical biases (61,62). We mapped all SNPs to dbSNP build 129, resulting in a unique set of 2 650 552 autosomal SNPs, after confirming the consistency of the coded allele across all studies. Scripts used for this meta-analysis are available online (http://www.broadinstitute.org/~debakker/meta.html).

Genome-wide significance was defined as $P < 5 \times 10^{-8}$ based on the estimated multiple testing burden for all common variants in populations of European ancestry (63). To identify independent signals reaching the genome-wide significance threshold within a locus, genome-wide association meta-analysis results were aggregated into bins by index SNP at a linkage disequilibrium $r^2$ threshold of 0.1, such that all results within a given bin were correlated to the index SNP at $r^2 \geq 0.1$ but to any index SNP in other bins at $r^2 < 0.1$. Four index SNPs at two loci were subsequently analyzed in conditional regression models ($n \leq 33 846$) to assess statistical independence.

Finally, we adopted the polygenic regression modeling approach as recently described by Purcell et al. (64) and implemented in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink) to estimate the genetic variance explained by associated loci at progressively less stringent $P$-value thresholds within the RS-II sample. The outcome of this analysis is a $P$-value threshold at which the explained variance is maximized. The list of loci included in the score yielding the maximum explained variance will include non-genome-wide loci that in aggregate contribute to the model’s performance and indicate that additional true-positive signals are likely to be present within that list. For this analysis, we removed the RS-I and RS-II data from the discovery meta-analysis to remove the risk of correlation between the discovery and validation sample and overestimation of the explained variance. Subsequently, we obtained a list of independent signals to be included in the model based on their statistical significance from the meta-analysis without RS-I or RS-II (PLINK-clump option, $r^2 \geq 0.05$, 1 Mb window) (65). We summarized variation across associated loci (using significance thresholds from $P < 5 \times 10^{-8}$ to $P < 0.05$) into quantitative scores per individual. The score was then used as a predictor in a linear regression analysis in the RS-II sample ($n = 1589$) and the resulting $r^2$ is reported as the measure for explained variance for each $P$-value threshold.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank all participants and the study staff of the Age, Gene/Environment Susceptibility Reykjavik Study (AGES); the
Atherosclerosis Risk in Communities (ARIC) study; the Cardiovascular Health Study (CHS); the Erasmus Rucphen Family (ERF) study; the Framingham Heart Study (FHS); the Cooperative Gesundheitsforschung in der Region Augsburg (KORA); the Micros Study; the Netherlands Study of Depression and Anxiety (NESDA); the Orkney Complex Disease Study (Orpacades); the Rotterdam Study (RS); the SardiNIA study; the Study of Health in Pomerania (SHIP); the TwinsUK study as well as the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium, for their invaluable contributions.

Conflict of Interest statement. None declared.

FUNDING
This work has been made possible by the generous financial support of a variety of funding sources to participating studies and individual scientists. A detailed list of acknowledgments and funding sources, as well as an author contribution list, is included in Supplementary Material.

REFERENCES


