Genome-wide association meta-analysis of 30,000 samples identifies seven novel loci for quantitative ECG traits.

Running title: Meta-analysis identifies seven novel ECG loci

Corresponding author:
Jessica van Setten, PhD
Department of Cardiology, University Medical Center Utrecht
Room G03.550, PO Box 85500, 3508 GA Utrecht, The Netherlands
T: +31 88 755 7381 @: J.vanSetten@umcutrecht.nl

Jessica van Setten[1*], Niek Verweij[2,3], Hamdi Mbarek[4], Maartje N. Niemeijer[5], Stella Trompet[6], Dan E. Arking[7], Jennifer A. Brody[8], Ilaria Gandin[9], Niels Grarup[10], Leanne M. Hall[11,12], Daiane Hemerich[13,14], Leo-Pekka Lyytikäinen[15], Hao Mei[16], Martina Müller-Nurasyid[17,18,19], Bram P. Prins[20], Antonietta Robino[21], Albert V. Smith[22,23], Helen R. Warren[24,25], Folkert W. Asselbergs[1,26,27], Dorret I. Boomsma[4], Mark J. Caulfield[24,25], Mark Eijgelsheim[28,29], Ian Ford[30], Torben Hansen[10], Tamara B. Harris[31], Susan R. Heckbert[32], Jouke-Jan Hottenga[4], Anita Iorio[33], Jan A. Kors[34], Allan Linneberg[35,36], Peter MacFarlane[37], Thomas Meitinger[38,19,39], Christopher P. Nelson[11,12], Olli T. Raitakari[40], Claudia T. Silva Aldana[41,42,43], Gianfranco Sinagra[33], Moritz Sinner[18,19], Elsayed Z. Soliman[44], Monika Stoll[45,46,47], Andre Uitterlinden[48], Cornelia M. van Duijn[49], Melanie Waldenberger[50,19,51], Alvaro Alonso[52], Paolo Gasparini[53,54], Vilmundur Gudnason[22,23], Yalda Jamshidi[20], Stefan Kääb[18,19], Jørgen K Kanters[55], Terho Lehtimäki[15], Patricia B. Munroe[24,25], Annette Peters[19,51,56], Nilesh J. Samani[11,12], Nona Sotoodehnia[57], Sheila Ulivi[21], James G. Wilson[58], Eco J.C. de Geus[4], J. Wouter Jukema[59], Bruno Stricker[5], Pim van der Harst[60,61], Paul I.W. de Bakker[62,63], Aaron Isaacs[45,46,47*]

1 Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, University of Utrecht, the Netherlands
2 University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, The Netherlands
3 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, USA
4 Department of Biological Psychology, Amsterdam Public Health research institute, Vrije Universiteit Amsterdam, The Netherlands
5 Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
6 Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands
7 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD USA
8 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA
9 Research Unit, AREA Science Park, Trieste, Italy
10 The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
11 Department of Cardiovascular Sciences, University of Leicester

12 NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Groby Road, Leicester
13 Department of Cardiology, University Medical Center Utrecht, Utrecht, the Netherlands
14 CAPES Foundation, Ministry of Education of Brazil, Brasília DF 70040-020, Brazil
15 Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland
16 University of Mississippi Medical Center, Center of Biostatistics and Bioinformatics
17 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
18 Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany
19 DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
20 Human Genetics Research Centre, ICCS, St George's University of London, London, UK
21 Institute for Maternal and Child Health, IRCCS "Burlo Garofolo" , Trieste , Italy
22 Icelandic Heart Association, Kopavogur, Iceland
23 Faculty of Medicine, University of Iceland, Reykavik, Iceland
24 William Harvey Research Institute, Barts and The London School of Medicine & Dentistry, Queen Mary University of London
25 NIHR Barts Cardiovascular Research Centre, Barts and The London School of Medicine & Dentistry, Queen Mary University of London
26 Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, the Netherlands
27 Institute of Cardiovascular Science, Faculty of Population Health Sciences, and Farr Institute of Health Informatics Research and Institute of Health Informatics, University College London, London, United Kingdom
28 Department of Nephrology, University Medical Center Groningen, Groningen, The Netherlands
29 Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
30 Robertson Center for Biostatistics, University of Glasgow, United Kingdom
31 Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, MD, USA
32 Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, Washington, USA
33 Cardiovascular Department, "Ospedali Riuniti and University of Trieste", Trieste, Italy
34 Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands
35 Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark
36 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
37 Institute of Health and Wellbeing, University of Glasgow, United Kingdom
38 Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
39 Institute of Human Genetics, Technische Universität München, Munich, Germany
40 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, and Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland
41 Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands
42 Doctoral Program in Biomedical Sciences, Universidad del Rosario, Bogotá, Colombia
43 Institute of translational Medicine-IMT- Center For Research in Genetics and Genomics-CIGGUR. GENIUROS Research Group. School of Medicine and Health Sciences, Universidad del Rosario, Colombia.
44 Epidemiological Cardiology Research Center (EPICARE), Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC
45 CARIM School for Cardiovascular Diseases, Maastricht University, Maastricht, The Netherlands
46 Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands
47 Department of Biochemistry, Maastricht University, Maastricht, The Netherlands

48 Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands.
49 Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
50 Research unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
51 Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
52 Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA
53 DSM, University of Trieste, Italy
54 IRCCS-Burlo Garofolo Children Hospital, Via dell'Istria 65, Trieste, Italy
55 Laboratory of Experimental Cardiology, University of Copenhagen, Copenhagen, Denmark
56 German Center for Diabetes Research, Neuherberg, Germany
57 Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, Washington, USA
58 Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, USA
59 Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands
60 University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands
61 Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands
62 Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands
63 Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

* Corresponding authors

Please note: funding and acknowledgements are listed in the Supplementary Data, because of the large number of cohorts.

Conflicts of interest

Dr. de Bakker is currently an employee of and owns equity in Vertex Pharmaceuticals.

M.J. Caulfield is Chief Scientist for Genomics England a UK Government company.

Abstract

Genome-wide association studies (GWAS) of quantitative electrocardiographic (ECG) traits in large consortia have identified more than 130 loci associated with QT interval, QRS duration, PR interval, and heart rate (RR interval). In the current study, we meta-analyzed genome-wide association results from 30,000 mostly Dutch samples on four ECG traits: PR interval, QRS duration, QT interval, and RR interval. SNP genotype data was imputed using the Genome of the Netherlands reference panel encompassing 19 million SNPs, including millions of rare SNPs (minor allele frequency <5%). In addition to many known loci, we identified seven novel locus-trait associations: *KCND3*, *NR3C1*, and *PLN* for PR interval, *KCNE1*, *SGIP1*, and *NFKB1* for QT interval, and *ATP2A2* for QRS duration, of which six were successfully replicated. At these seven loci, we performed conditional analyses and annotated significant SNPs (in exons and regulatory regions), demonstrating involvement of cardiac-related pathways and regulation of nearby genes.

Key words: Electrocardiology (ECG), Genetic association studies, Imputation, Meta-analysis

Introduction

Quantitative electrocardiographic (ECG) traits have been well studied in large consortia, identifying over 130 significant loci. Some loci were associated with multiple traits. Nevertheless, these loci collectively explain only a small portion of the genetic variation of these traits.[1] Large GWAS meta-analyses on PR interval[2,3], RR interval/heart rate,[4,5] QRS duration,[6,7] and QT interval[8-10] were based on HapMap imputations.[11] Testing ~2.5 million SNPs, these studies provided good coverage of common variation in the genome. SNPs with lower allele frequencies (e.g. minor allele frequencies between 1% and 5%), however, are poorly covered.[12,13] While HapMap included only 270 samples (30 trios and 90 unrelated samples) from 3 continental populations,[11] the 1000 Genomes Project Phase 3 contains 2504 samples from 26 populations.[14] Larger reference panels cover a broader variety of haplotypes and, therefore, increase the quality of imputation in a GWAS sample. Moreover, the number of observed SNPs also increases, expanding the number available for imputation. This has led to novel findings in non-ECG related studies.[15]

In the current study, we meta-analyzed genome-wide data on four ECG traits in 30,000 predominantly Dutch samples. We tested over 19 million SNPs for association, which were imputed using the Genome of the Netherlands (GoNL) reference panel.[16] This dataset contains whole-genome sequencing data at 12x coverage collected in 250 families (trios and parents with two offspring). Nearly all polymorphic sites with a population frequency of more than 0.5% are captured. This makes it one of the largest single population sequencing efforts worldwide and the trio design ensures very accurate haplotype phasing. These features and the good match with the predominantly Dutch cohorts, make this dataset well suited as a reference panel for imputation. Using this approach, we had two aims: 1) the discovery of novel loci associated with ECG traits, and 2) the fine-mapping and

functional annotation of known regions associated with ECG traits. We increased our SNP density almost seven-fold compared to previous studies based on HapMap, enabling us to study key signals in much finer detail.

Methods

Individual cohort data

Eight cohorts were included in the discovery phase of this study, totaling approximately 30,000 samples (**Supplementary Tables 1 and 2, Supplementary Notes**). Most study participants were Dutch with the exception of most participants of PROSPER; this study included approximately 19% samples of Dutch origin, while the remaining samples were of other European descent. All cohorts performed stringent quality control to exclude low-quality samples and SNPs prior to imputation and also post-imputation. Imputation was performed using 998 phased haplotypes from the Genome of the Netherlands Project release 4 as the reference panel, encompassing 19,763,454 SNPs.[16] All genomic data in this manuscript is listed with respect to the hg19 (build37) reference genome.

We evaluated four phenotypes on the electrocardiogram: RR interval, PR interval, QRS duration, and QT interval. Seven out of eight cohorts contributed data to all four phenotypes; NTR only had data on RR interval available. Samples of non-European descent and samples with missing data were excluded, as well as individuals that fulfilled any of the exclusion criteria listed in **Supplementary Table 3**. SNPs were individually tested for association with each trait using linear models. For all four phenotypes we included age, sex, height, BMI, and study specific covariates (for instance to correct for study site, relatedness, or population stratification) as covariates. In addition, RR interval and hypertension (in those cohorts that had data available on this measure) were included as covariates for QT interval to reduce noise introduced by these factors. We chose these covariates to correspond with previously published GWAS on these four ECG traits.

Quality control and meta-analysis

Association results from all cohorts were collected at a single site and underwent quality control. SNPs with extreme values of beta (> 1000 or < -1000), standard error (SE) (> 1000), or imputation quality (< 0.1 or > 1.1) were removed and distributions of beta, SE, and *P*-values were manually checked. We made QQ-plots to test *P*-value distributions, which were stratified by minor allele frequency and by imputation quality. Aberrant subsets of SNPs (usually with very low frequency) were removed from downstream analyses.

Inverse-variance fixed-effect model meta-analyses were conducted for all four traits using MANTEL.[17] For each individual GWAS, genomic inflation factors (lambda) were calculated and, during meta-analysis, standard errors were adjusted accordingly to correct for population structure and technical errors. We did not correct for genomic inflation after meta-analysis. SNP associations were considered significant if $P \leq 5 \times 10^{-8}$.

Follow-up on known and novel loci

For each locus, we tested the number of independent signals using the LD structure from GoNL in GCTA-COJO, which was designed to allow conditional analyses based on summary level data.[18] Secondary hits had to fulfill two criteria: 1) genome-wide significant in the GWAS, and 2) $P < 1 \times 10^{-5}$ after conditioning to correct for multiple testing of 4,757 significant SNPs across all four traits. A novel locus for a trait was defined if the significant SNPs, or SNPs within a distance of 1 Mb upstream and downstream of the significant SNPs, had not been observed before in GWAS of the same trait. We performed a look-up of all novel loci in previous HapMap-based GWAS.

Replication of novel loci in CHARGE

We sought to replicate our findings in 13 independent cohorts taking part in the CHARGE consortium[19] (**Supplementary Tables 1 and 2, Supplementary notes**). Twelve studies (TwinsUK, CHS, ARIC, KORA F3, KORA S4, JHS, AGES, BRIGHT, YFS, INGI-FVG, and INGI-CARL) used 1000 Genomes Phase 1 as their imputation reference panel and a single study (Inter99) provided only genotyped data. All studies contained samples of European ancestry, except for JHS, which consisted only of African-American samples. The summary-level results for all novel SNPs determined in the discovery analysis were combined in inverse-variance fixed-effects meta-analyses. A two-sided *P*-value ≤ 0.05, in conjunction with a concordant effect direction, was considered significant.

***In silico* tests of possibly functional SNPs**

We looked up the functional annotations for all SNPs that reached genome-wide significance in any of the four traits. First, we checked whether SNPs were potentially damaging to protein function, testing all non-synonymous SNPs in SIFT[20] and PolyPhen-2.[21] Second, we used GREAT[22] to identify biological pathways in which regulatory SNPs are involved, testing the index SNPs for all locus-trait associations. Lastly, we tested all significant SNPs one by one for their possible effect on regulatory regions using RegulomeDB.[23]

Results

Meta-analysis detects novel loci

We conducted a GWAS meta-analysis comprising eight cohorts that together encompassed approximately 30,000 samples. Over 19 million SNPs, imputed using the GoNL reference panel, were assessed for association with four quantitative ECG traits: RR, PR, QRS, and QT. Considering all traits, we observed 52 locus-phenotype associations (17 for PR, 13 for QRS, 15 for QT, and 7 for RR; **Supplementary Figures 1 and 2, Supplementary Table 4**). A locus was defined as an associated region (containing one or more SNPs with $P \leq 5 \times 10^{-8}$) that is located at least 1Mb away from the next (i.e. if two associated SNPs are within 1Mb, they belong to the same locus). Of these 52 loci, 45 have been observed before in large GWAS meta-analyses[2-4,7-9] and seven are novel findings (Table 1). **Box 1** shows regional association plots and provides additional information on the seven novel loci. Imputation qualities of the index SNPs were 0.60 and 0.84 for the relatively rare *KCNE1* and *KCDN1* variants, respectively, and >0.96 for the remaining common index SNPs. The variance explained by each of these variants ranges between 0.09% and 0.23%.

Fine mapping of known loci

For each locus, we tested if more than one independent signal was present (**Supplementary Table 4**). Thirteen loci had suggestive evidence of having more than one independent signal; four locus-phenotype associations had five or more independent signals. The *SCN5A*/*SCN10A* locus was the most outstanding locus with eleven independent signals for PR, and six for QRS. *NOS1AP* for QT contained seven independent signals.

307 **Replication in CHARGE**

308 For six out of seven novel loci, we were able to conduct look-ups of the index SNP or
309 a proxy SNP in strong LD ($r^2 \geq 0.89$) in previous large-scale HapMap-based GWAS.
310 These GWAS contained over 70,000 samples each, and included many of the Dutch
311 cohorts from our current study. All six loci were associated with their respective traits
312 ($P \leq 0.004$). Next, we tested the seven novel loci for replication in 13 studies from the
313 CHARGE consortium. In contrast to the HapMap look-ups, this replication was
314 independent from the Dutch discovery sample. Results are shown in **Table 1**. Allele
315 frequencies were very similar to the discovery dataset, except for JHS, which
316 consists of individuals of African American descent. Effect directions for all seven
317 SNPs were concordant between our primary findings and replication, with effect sizes
318 between 0.2 and 1.5 times those of the betas in the discovery study. Six of seven loci
319 were replicated with $P < 0.05$, three of which pass Bonferroni correction, accounting
320 for seven tests.

321

322 **Functional SNPs in genes and regulatory regions**

323 All genome-wide significant SNPs were tested *in silico* for their potential effect on
324 gene expression and protein structure. Ten loci contained, in total, 15
325 nonsynonymous SNPs, which were tested using the prediction programs PolyPhen-2
326 and SIFT. According to PolyPhen-2, three SNPs were possibly damaging (rs1805128
327 in *KCNE1* for QT, rs12666989 in *UFSP1* for RR, and rs2070492 in *SLC22A14* for
328 PR). SIFT predicted only one SNP to be damaging to a protein (rs3746471 in
329 *KIAA1755* for RR).

330

331 We used GREAT to test all 100 index SNPs from the four ECG traits combined for
332 their biological function in *cis*-regulatory regions. Significant GO-terms (molecular
333 function, biological process, and cellular component), human phenotypes, and

disease ontologies are shown in **Supplementary Tables 5a-d**. In total, these index SNPs mapped to 103 genes.

Of 52 locus-phenotype associations, 34 contained significant SNPs that have a RegulomeDB score of 3 or better, meaning that they may affect protein binding (**Supplementary Table 6**). We observed 15 loci containing SNPs with scores of 1 (likely to affect binding and linked to the expression of a gene target), 15 loci containing SNPs with a maximum score of 2 (likely to affect binding), and four loci that have SNPs with a maximum score of 3 (less likely to affect binding). Eighteen loci contained only SNPs with scores from 4 to 6 (minimal binding evidence) and 7 (no data available).

Discussion

We imputed over 19 million SNPs using GoNL as the reference panel, and tested these SNPs for association with four traits in eight Dutch cohorts comprising roughly 30,000 samples. We observed 52 locus-phenotype associations, seven of which were novel (**Table 1**, **Box 1, Supplementary Table 4**).

Discovery of loci associated with quantitative ECG traits

We detected seven novel loci, three for PR interval, three for QT interval, and one for QRS duration (**Box 1**). No novel loci were found for RR interval, accounting for loci previously associated with either RR interval[4] or heart rate.[5] We replicated six out of seven novel loci utilizing 13 independent studies from the CHARGE consortium. Interestingly, the only variant that does not replicate is rs74640693 for PR interval, located close to *PLN* (phospholamban). Variants in this gene have been consistently associated with various QRS measures[6] but not with PR interval. The gene transcribes the phospholamban protein, which is important in calcium signaling in cardiac muscle cells.[24] Although a Dutch-specific pathogenic mutation, p.Arg14del, in the *PLN* gene has been described,[25] it is unlikely that this mutation drives the association signal in our study because the allele frequency of SNP rs74640693 is similar in our samples (4.9%) compared to other samples of European ancestry (4.6% in the 12 European CHARGE replication cohorts). Furthermore, the allele frequency of this SNP is ~5 times higher than that of the mutation and the SNP is located approximately 200kb upstream of the PLN gene, so, therefore, not in LD with these mutations. In addition, a recent large study of PR interval used the Illumina exome chip to identify a common variant (rs74640693, allele frequency 47%) in this region,[26] however, this variant is not in LD with the variant that we identified (r^2 = 0.04). To confirm that the lack of association was not caused by strand issues (because rs74640693 is an A/T variant), we tested the nearby proxy SNP

375 rs12210733 (which is an A/G variant, r^2 = 0.89) in the CHARGE replication cohorts,
376 and found it was also non-significant.
377
378 We looked up our top SNPs in previous, much larger, HapMap-based GWAS meta-
379 analyses to determine why our SNPs were not identified in those studies (**Table 1**).
380 Two loci contained rare SNPs with MAF < 5%. Low-frequency SNPs at *KCND3* were
381 not present in HapMap and could therefore not be tested. The functional SNP at
382 *KCNE1* was observed in a single cohort in a meta-analysis in 2009, but this result
383 could neither be replicated in other cohorts,[9] nor in later studies, because the
384 imputation quality was too low.
385
386 For common SNPs (MAF > 5%), it is much more difficult to define why they were not
387 previously observed at genome-wide significance. For many loci we may have better
388 tags of the causal variants because our coverage is almost seven-fold greater.
389 Indeed, the index SNPs at *PLN* (PR), *NFKB1* (QT), and *ATP2A2* (QRS) were not
390 tested in previous studies. Nevertheless, for all SNPs, proxies with r^2 > 0.9 were
391 available in the respective studies (**Table 1**). Common SNPs at *KCND3* (PR), *NR3C1*
392 (PR), and *SGIP1* (QT) were present in HapMap. Both proxies and directly imputed
393 SNPs were at least nominally significant in previous studies (*P*-values ranging from
394 10^{-3} to 10^{-6}) with typically high imputation quality.
395
396 In addition to the ”winner’s curse” effect, we expect that higher quality imputation due
397 to the considerably larger haplotype panel (compared to HapMap) and the ancestry
398 matching between GoNL and our Dutch cohorts will improve the power to detect a
399 true association signal, if present. Although combining multiple reference panels for
400 imputation is becoming the new standard[27], limitations to our study remain: (1) the
401 GoNL reference panel may not contain sufficient information on rare SNPs; (2) the
402 small sample size of individual cohorts may cause abnormal behavior of rare SNPs

as a group, requiring us to remove that subset of SNPs; or (3) the sample size or power of the overall study is still limited to detect rare variant associations.

Fine mapping of known loci

Although we did not sequence the loci containing the known and novel signals, we have a much denser interrogation of these regions compared to previous (HapMap-based) studies. In an attempt to fine map the significant loci, we annotated all significant SNPs with their predicted functional consequences.

First, we used SIFT and PolyPhen-2 to predict the effect of 15 nonsynonymous SNPs that were associated with one of the ECG traits at genome-wide significance. PolyPhen-2 classified three SNPs as possibly damaging and SIFT predicted only one SNP to be damaging. These were non-overlapping, raising questions as to the actual effect of these SNPs on their respective genes. Functional studies should be pursued to test the direct effect of these SNPs on protein structure.

Combining all index SNPs, we tested the function of those SNPs located in *cis*-regulatory regions using GREAT.[22] We identified 100 independent SNP-trait associations, which mapped to 103 genes. Interestingly, we find hundreds of significant nodes, of which the vast majority is related to cardiac functioning and heart disease (**Supplementary Tables 5a-e**). This shows that, indeed, many SNPs are located in *cis*-regulatory regions of genes that are critical in the functioning of the human heart, which implicates a regulatory function of these loci rather than a structural function changing the protein directly. One example is shown in **Supplementary Figure 3**; this figure contains all significant GO molecular function nodes. Most of these nodes are in the group of transporter activity, which includes all transmembrane channels that are known to be important for cardiac function.

Because the GREAT pathways show that many SNPs probably have their effect on the trait due to gene regulation, we extracted all significant SNPs from RegulomeDB to check which variants would likely affect binding in regulatory regions. A majority of loci contained at least one SNP that was expected to affect transcription factor binding (**Supplementary Table 6**). The score that is provided by RegulomeDB indicates that a SNP is likely (or less likely) located in a binding site. Interestingly, there are strong differences between loci in terms of the number of SNPs that may have a regulatory effect, and percentage of loci per trait that have a high score. For instance, seven out of 15 QT interval loci contains SNPs with a score of 1, while only a single PR interval locus contains a SNP with this score. The *SCN5A*/*SCN10A* locus is strongly associated with PR interval (best SNP $P = 4.9 \times 10^{-107}$) and contains over 450 significant SNPs. Nevertheless, only six SNPs have a score of 2 or 3, and none of the significant SNPs have a score of 1. However, many binding sites are tissue specific, and, therefore, testing SNPs with high scores systematically for their role in cardiac tissue could lead to more knowledge about their biological function.

Conclusions

Using the Genome of the Netherlands as imputation reference panel, we identified seven novel loci for quantitative ECG traits. Higher SNP density and higher imputation quality enabled us to annotate known loci, facilitating future studies to understand the biological effects of causal variants at many loci. Ultimately, combining imputation reference panels and increasing sample size for GWAS meta-analyses will continue to increase power for genetic discovery and novel target identification. With many sequencing efforts ongoing and large population-based cohorts being genotyped (such as UK Biobank, of which the first release data showed 46 novel loci for RR interval[28]), we can expect hundreds of novel loci for ECG phenotypes in the near future.

Conflicts of interest

Dr. de Bakker is currently an employee of and owns equity in Vertex Pharmaceuticals.

M.J. Caulfield is Chief Scientist for Genomics England a UK Government company.

References

1. Silva CT, Kors JA, Amin N *et al*. Heritabilities, proportions of heritabilities explained by GWAS findings, and implications of cross-phenotype effects on PR interval. *Hum Genet* 2015; **134:** 1211-1219.
2. Pfeufer A, van Noord C, Marciante KD *et al*. Genome-wide association study of PR interval. *Nat Genet* 2010; **42:** 153-159.
3. van Setten J, Brody JA, Jamshidi Y *et al*. PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat Commun* 2018; **9:** 2904.
4. Eijgelsheim M, Newton-Cheh C, Sotoodehnia N *et al*. Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet* 2010; **19:** 3885-3894.
5. den Hoed M, Eijgelsheim M, Esko T *et al*. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet* 2013; **45:** 621-631.
6. van der Harst P, van Setten J, Verweij N *et al*. 52 Genetic Loci Influencing Myocardial Mass. *J Am Coll Cardiol* 2016; **68:** 1435-1448.
7. Sotoodehnia N, Isaacs A, de Bakker PI *et al*. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet* 2010; **42:** 1068-1076.
8. Newton-Cheh C, Eijgelsheim M, Rice KM *et al*. Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet* 2009; **41:** 399-406.
9. Pfeufer A, Sanna S, Arking DE *et al*. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet* 2009; **41:** 407-414.
10. Arking DE, Pulit SL, Crotti L *et al*. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet* 2014; **46:** 826-836.
11. Frazer KA, Ballinger DG, Cox DR *et al*. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449:** 851-861.
12. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet* 2006; **38:** 659-662.
13. Pe'er I, de Bakker PI, Maller J, Yelensky R, Altshuler D, Daly MJ. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* 2006; **38:** 663-667.
14. 1000 Genomes Project Consortium, Auton A, Brooks LD *et al*. A global reference for human genetic variation. *Nature* 2015; **526:** 68-74.
15. de Vries PS, Sabater-Lleal M, Chasman DI *et al*. Comparison of HapMap and 1000 Genomes Reference Panels in a Large-Scale Genome-Wide Association Study. *PLoS One* 2017; **12:** e0167742.
16. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 201; **46:** 818-825.
17. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008; **17:** R122-128.
18. Yang J, Ferreira T, Morris AP *et al*. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44:** 369-375, S361-363.
19. Psaty BM, O'Donnell CJ, Gudnason V *et al*. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2009; **2:** 73-80.

20. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4:** 1073-1081.
21. Adzhubei IA, Schmidt S, Peshkin L *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7:** 248-249.
22. McLean CY, Bristor D, Hiller M *et al*. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; **28:** 495-501.
23. Boyle AP, Hong EL, Hariharan M *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22:** 1790-1797.
24. Luo W, Grupp IL, Harrer J *et al*. Targeted ablation of the phospholamban gene is associated with markedly enhanced myocardial contractility and loss of beta-agonist stimulation. *Circ Res* 1994; **75:** 401-409.
25. van der Zwaag PA, van Rijsingen IA, de Ruiter R *et al*. Recurrent and founder mutations in the Netherlands-Phospholamban p.Arg14del mutation causes arrhythmogenic cardiomyopathy. *Neth Heart J* 2013; **21:** 286-293.
26. Lin H, van Setten J, Smith AV *et al*. Common and Rare Coding Genetic Variation Underlying the Electrocardiographic PR Interval. *Circ Genom Precis Med* 2018; **11:** e002037.
27. McCarthy S, Das S, Kretzschmar W *et al*. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48:** 1279-1283.
28. Eppinga RN, Hagemeijer Y, Burgess S *et al*. Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nat Genet* 2016; **48:** 1557-1563.
29. Verweij N, Mateo Leach I, Isaacs A *et al*. Twenty-eight genetic loci associated with ST-T-wave amplitudes of the electrocardiogram. *Hum Mol Genet* 2016; **25:** 2093-2103.
30. Low SK, Takahashi A, Ebana Y *et al*. Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat Genet* 2017; **49:** 953-958.
31. Lundby A, Olesen SP. KCNE3 is an inhibitory subunit of the Kv4.3 potassium channel. *Biochem Biophys Res Commun* 2006; **346:** 958-967.
32. Delpon E, Cordeiro JM, Nunez L *et al*. Functional effects of KCNE3 mutation and its role in the development of Brugada syndrome. *Circ Arrhythm Electrophysiol* 2008; **1:** 209-218.
33. Lee YC, Durr A, Majczenko K *et al*. Mutations in KCND3 cause spinocerebellar ataxia type 22. *Ann Neurol* 2012; **72:** 859-869.
34. Kadmiel M, Cidlowski JA. Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol Sci* 2013; **34:** 518-530.
35. Rog-Zielinska EA, Thomson A, Kenyon CJ *et al*. Glucocorticoid receptor is required for foetal heart maturation. *Hum Mol Genet* 2013; **22:** 3269-3282.
36. Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. *J Allergy Clin Immunol* 2013; **132:** 1033-1044.
37. Lundmark R, Doherty GJ, Howes MT *et al*. The GTPase-activating protein GRAF1 regulates the CLIC/GEEC endocytic pathway. *Curr Biol* 2008; **18:** 1802-1808.
38. Doherty JT, Lenhart KC, Cameron MV, Mack CP, Conlon FL, Taylor JM. Skeletal muscle differentiation and fusion are regulated by the BAR-containing Rho-GTPase-activating protein (Rho-GAP), GRAF1. *J Biol Chem* 2011; **286:** 25903-25921.
39. Borkhardt A, Bojesen S, Haas OA *et al*. The human GRAF gene is fused to MLL in a unique t(5;11)(q31;q23) and both alleles are disrupted in three cases of myelodysplastic syndrome/acute myeloid leukemia with a deletion 5q. *Proc Natl Acad Sci U S A* 2000; **97:** 9168-9173.

40. Brittsan AG, Kranias EG. Phospholamban and cardiac contractile function. *J Mol Cell Cardiol* 2000; **32:** 2131-2139.
41. Nishimura M, Suzuki S, Satoh T, Naito S. Tissue-specific mRNA expression profiles of human solute carrier 35 transporters. *Drug Metab Pharmacokinet* 2009; **24:** 91-99.
42. Kranias EG, Hajjar RJ. Modulation of cardiac contractility by the phospholamban/SERCA2a regulatome. *Circ Res* 2012; **110:** 1646-1660.
43. Trevaskis J, Walder K, Foletta V *et al*. Src homology 3-domain growth factor receptor-bound 2-like (endophilin) interacting protein 1, a novel neuronal protein that regulates energy balance. *Endocrinology* 2005; **146:** 3757-3764.
44. Uezu A, Horiuchi A, Kanda K *et al*. SGIP1alpha is an endocytic protein that directly interacts with phospholipids and Eps15. *J Biol Chem* 2007; **282:** 26481-26489.
45. Cummings N, Shields KA, Curran JE *et al*. Genetic variation in SH3-domain GRB2-like (endophilin)-interacting protein 1 has a major impact on fat mass. *Int J Obes (Lond)* 2012; **36:** 201-206.
46. Vogel U, Jensen MK, Due KM *et al*. The NFKB1 ATTG ins/del polymorphism and risk of coronary heart disease in three independent populations. *Atherosclerosis* 2011; **219:** 200-204.
47. Zhou B, Rao L, Peng Y *et al*. Functional polymorphism of the NFKB1 gene promoter is related to the risk of dilated cardiomyopathy. *BMC Med Genet* 2009; **10:** 47.
48. Karban AS, Okazaki T, Panhuysen CI *et al*. Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis. *Hum Mol Genet* 2004; **13:** 35-45.
49. Tang T, Cui S, Deng X *et al*. Insertion/deletion polymorphism in the promoter region of NFKB1 gene increases susceptibility for superficial bladder cancer in Chinese. *DNA Cell Biol* 2010; **29:** 9-12.
50. Paulussen AD, Gilissen RA, Armstrong M *et al*. Genetic variations of KCNQ1, KCNH2, SCN5A, KCNE1, and KCNE2 in drug-induced long QT syndrome patients. *J Mol Med (Berl)* 2004; **82:** 182-188.

Figures

Figure legends:

Figure 1 (Box 1): Novel loci associated with PR, QRS, and QT

Seven novel loci were identified; three for PR, three for QT, and one for QRS. Information and regional association plots are shown for every locus. Each SNP is plotted with respect to its chromosomal location (hg19, x-axis) and its *P*-value (y-axis on the left). The tall blue spikes indicate the recombination rate (y-axis on the right) at that region of the chromosome.

KCND3, associated with PR interval (Fig 1a, 1b)

We observed two independent signals at the *KCND3* gene. The first signal consists of low-frequency SNPs (MAF < 3.8%, index SNP MAF = 2.4%) upstream of *KCND3* (**top**), while the second signal contains intronic SNPs with much higher allele frequencies (index SNP MAF = 19.6%, **bottom**). *KCND3* encodes voltage-gated potassium channel subunit $K_v4.3$. SNPs near *KCND3* have been associated with P-wave duration and ST-T wave amplitude,[29] and with Atrial Fibrillation in the Japanese population.[30] It is thought that *KCND3* overexpression may be involved in Brugada syndrome because of its direct interaction with *KCNE3*. This gene inhibits *KCND3*, and specific mutations in the latter gene lead to Brugada syndrome.[31,32] Moreover, it has been shown that mutations in *KCND3* cause spinocerebellar ataxia.[33]

ARHGAP26 and *NR3C1*, associated with PR interval (Fig 1c)

The association signal in this locus spans the *NR3C1* gene, with the two genome-wide significant SNPs located between *NR3C1* and *ARHGAP26*. Both SNPs are common, with MAFs of approximately 45%. *NR3C1* encodes the glucocorticoid receptor, which interacts with a wide variety of proteins, transcription factors, and other cellular compounds.[34] In mice, this gene is involved in cardiac development,[35]

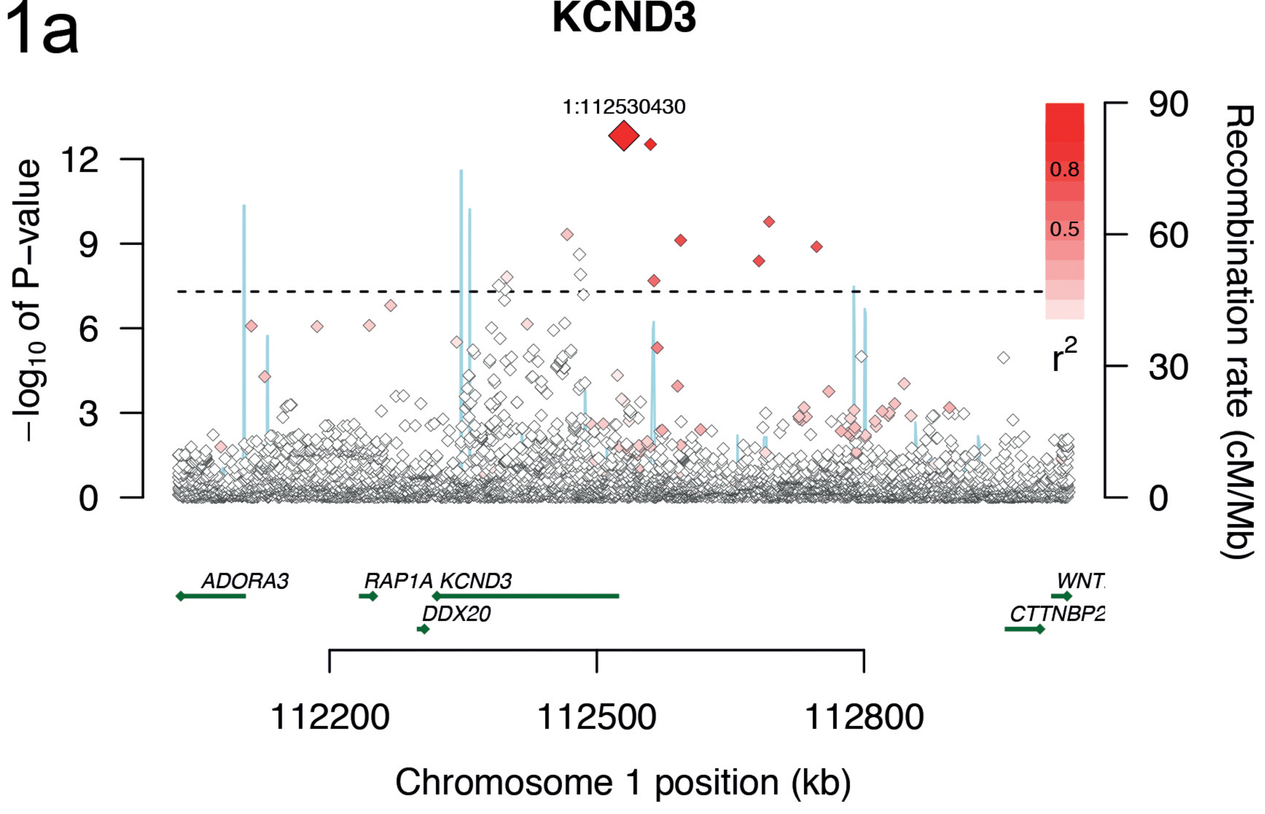
633 and overexpression causes ECG abnormalities,[36] which makes it likely that this is the
634 gene underlying the association signal. *ARHGAP26* encodes GRAF protein (GTPase
635 Regulator Associated with Focal Adhesion Kinase), which is required in specific exo-
636 and endocytosis pathways,[37] but also for muscle development.[38] Mutations in this
637 gene have been implicated in leukemia.[39]
638
639 *SLC35F1* and *PLN*, associated with PR interval (Fig 1d)
640 This locus has been associated previously with RR interval[4], QT interval[8,9], and QRS
641 duration.[7] The index SNP has a MAF of 5.4% and the association signals spans
642 *SLC35F1* and *PLN*. The latter gene encodes phospholamban, which is an important
643 regulator of cardiac contractility.[40] *SLC35F1* encodes a transporter protein that is
644 highly expressed in the human brain.[41]
645
646 *ATP2A2*, associated with QRS duration (Fig 1e)
647 Although only one (common, MAF = 32.2%) SNP reached genome-wide significance,
648 SNPs in strong LD with the index SNP span an area of almost 500kb, covering many
649 genes. This locus has been associated with QT interval previously.[10] Our most
650 significant SNP is located just downstream of *ATP2A2*, a strong candidate gene in
651 this region that encodes a SERCA Ca^{2+} ATPase, which is involved in calcium
652 transport in the human heart and under regulation of phospholamban.[42]
653
654 *SGIP1* and *TCTEX1D1*, associated with QT interval (Fig 1f)
655 This locus spans approximately 300kb in between two recombination hotspots.
656 Significant SNPs are in almost complete LD with each other, with minor allele
657 frequencies of approximately 15%. The locus spans two genes, *SGIP1* and
658 *TCTEX1D1*. *SGIP1* encodes a proline-rich endocytic protein that interacts with
659 endophilin and is involved in energy homeostasis.[43,44] This gene is mainly expressed

660 in the human brain[43] and has been associated with fat mass.[45] The *TCTEX1D1* gene
661 belongs to the dynein light chain Tctex-type family and has an unknown function.
662
663 *NFKB1*, associated with QT interval (Fig 1g)
664 The most significant SNPs in this locus are located upstream of the *NFKB1* gene,
665 encoding the NF-kappa-B p105 subunit. SNPs in this locus are common (MAF =
666 43.5%). An indel in the promotor of this gene has been associated with coronary
667 heart disease[46] and dilated cardiomyopathy[47]. This particular indel is in moderate LD
668 with the index SNP in this locus (r^2 in GoNL = 0.4). *NFKB1* is a transcription factor is
669 involved in many immune- and tumor-related processes, and has been associated
670 with ulcerative colitis[48] and bladder cancer.[49]
671
672 *KCNE1*, associated with QT interval (Fig 1h)
673 This locus contains a low frequency SNP (MAF = 1.7%) with a large effect on QT
674 interval. This SNP has been observed in GWAS before, but could not be replicated
675 (in this[8] and later studies[10]) because it was poorly imputed so only cohorts that
676 genotyped the SNP directly could be included.[8] *KCNE1* encodes a voltage-gated
677 potassium channel, and the index SNP encodes a pathogenic Asp to Asn amino acid
678 substitution at position 85 of *KCNE1*, causing long QT syndrome 5.[50]

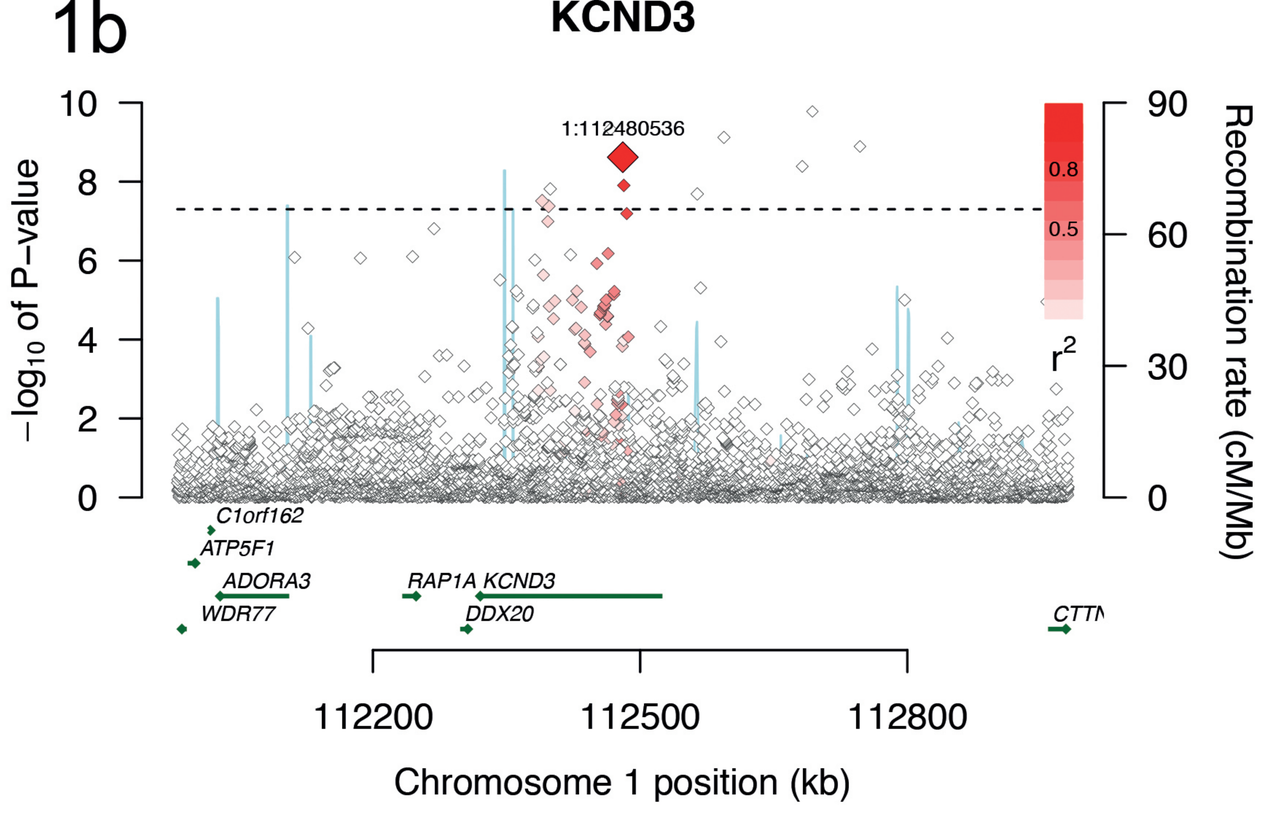
Table 1: Meta-analyses in 30,000 samples identify seven novel loci for PR interval, QRS duration, and QT interval. Using GoNL as reference panel in approximately 30,000 samples mostly of Dutch descent, we found seven loci not previously identified or (in the case of *KCNE1* for QT interval) not consistently replicated in previous genome-wide association studies. We conducted look-ups of these SNPs (or proxy SNPs in strong LD if the SNPs were not present in HapMap) in their respective HapMap-based meta-analyses and replicated six out of seven in a combined analysis of 13 CHARGE cohorts imputed with 1000 Genomes Phase 1. All effect estimates and allele frequencies are with respect to the coded allele.

SNP info					GoNL-imputed data							Previous HapMap-based meta-analysis				Replication in 13 CHARGE cohorts (1000 Genomes Phase 1 imputed)			
Locus	**Trait**	**Index SNP**	**Chr**	**Position (hg19)**	**Coded allele**	**Non-coded allele**	**Coded allele frequency**	**Beta**	**SE**	**P-value**	**Sample size**	**Proxy used**	**P-value**	**Sample size**	**Ref**	**Beta**	**SE**	**P-value**	**Sample size**
KCND3	PR	rs75013985	1	112530430	G	A	0.033	-4.090	0.554	1.5×10^{-13}	31695	No proxies available with $r^2 > 0.4$	N/A	92340	3	-5.967	0.985	1.4×10^{-9}	19302
NR3C1 / ARHGAP26	PR	rs17287745	5	142655015	G	A	0.425	1.011	0.185	4.2×10^{-8}	31695	No	1.9×10^{-6}	92340	3	0.585	0.193	0.002	24438
PLN / SLC35F1	PR	rs74640693	6	118684824	T	A	0.049	2.376	0.428	2.9×10^{-8}	31695	rs10457327 ($r^2 = 0.89$)	2.9×10^{-4}	92340	3	0.457	0.419	0.276	27106
SGIP1	QT	rs6588213	1	67107894	T	C	0.126	1.596	0.282	1.5×10^{-8}	26794	No	0.001	76061	10	0.757	0.199	1.4×10^{-4}	22663
NFKB1	QT	rs11097788	4	103407428	G	A	0.561	1.048	0.186	1.8×10^{-8}	26794	rs1598856 ($r^2 = 0.97$)	1.3×10^{-4}	76061	10	0.336	0.131	0.010	30504
KCNE1	QT	rs1805128	21	35821680	T	C	0.018	7.409	0.939	2.9×10^{-15}	26794	No	0.004	76061	10	4.874	0.671	3.7×10^{-13}	15896
ATP2A2 / ANAPC7	QRS	rs28637922	12	110819139	T	G	0.259	0.565	0.102	3.0×10^{-8}	25509	rs1502337 ($r^2 = 0.89$)	4.1×10^{-4}	73518	6	0.177	0.074	0.027	29427

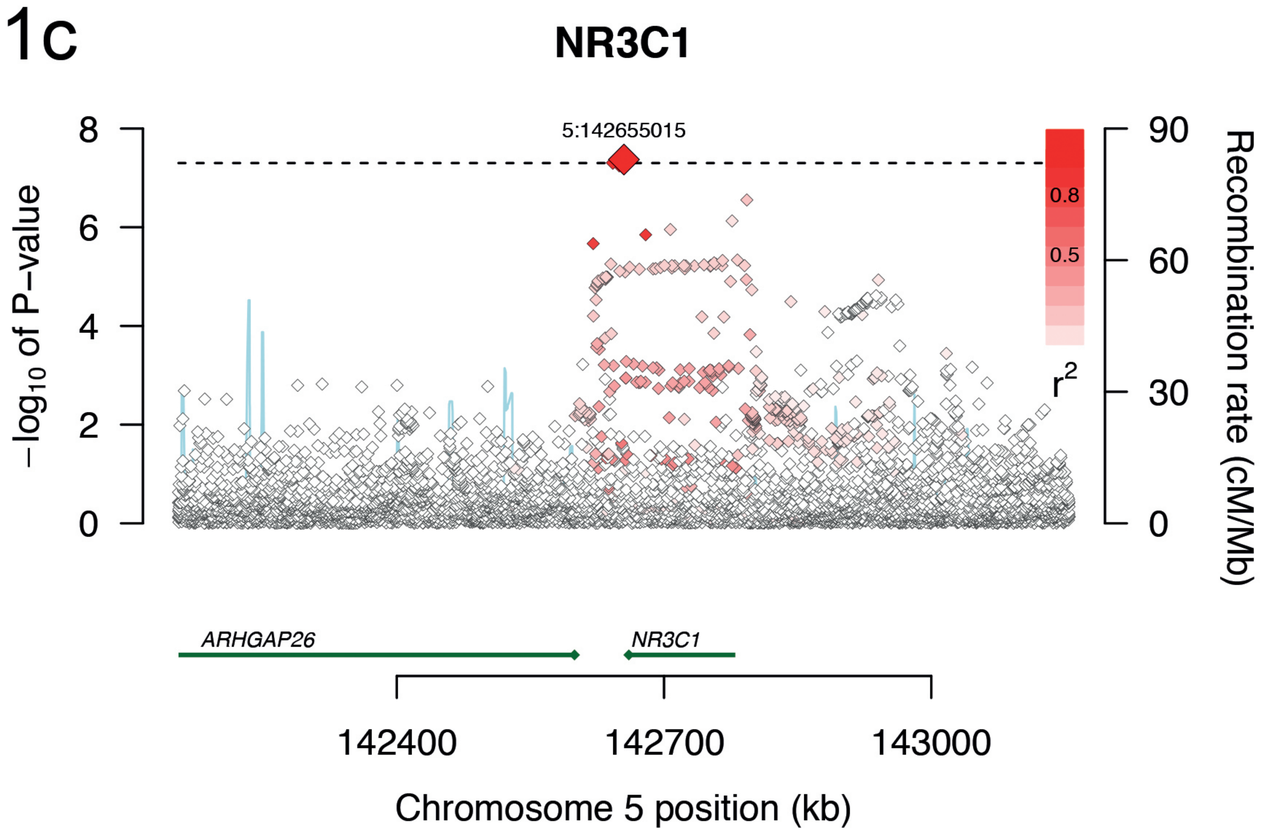
1a



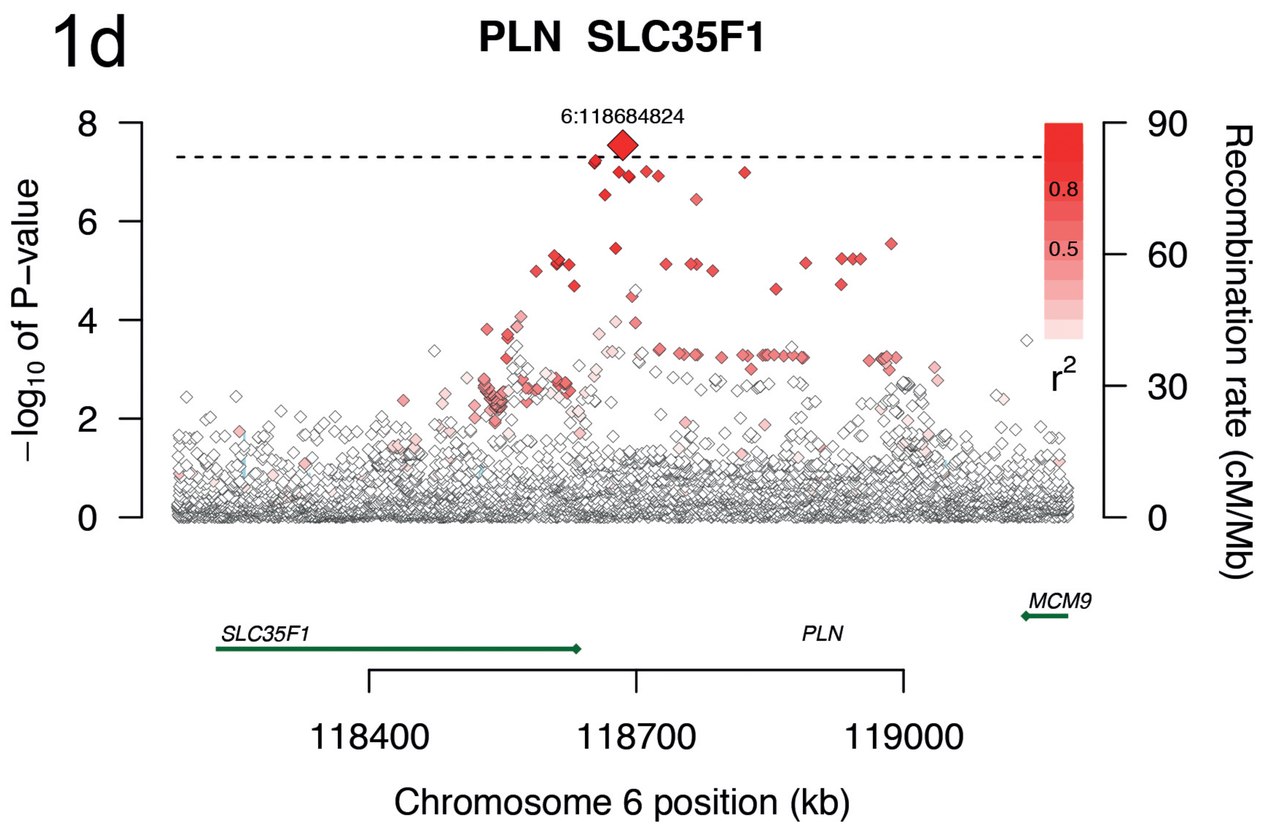
1b



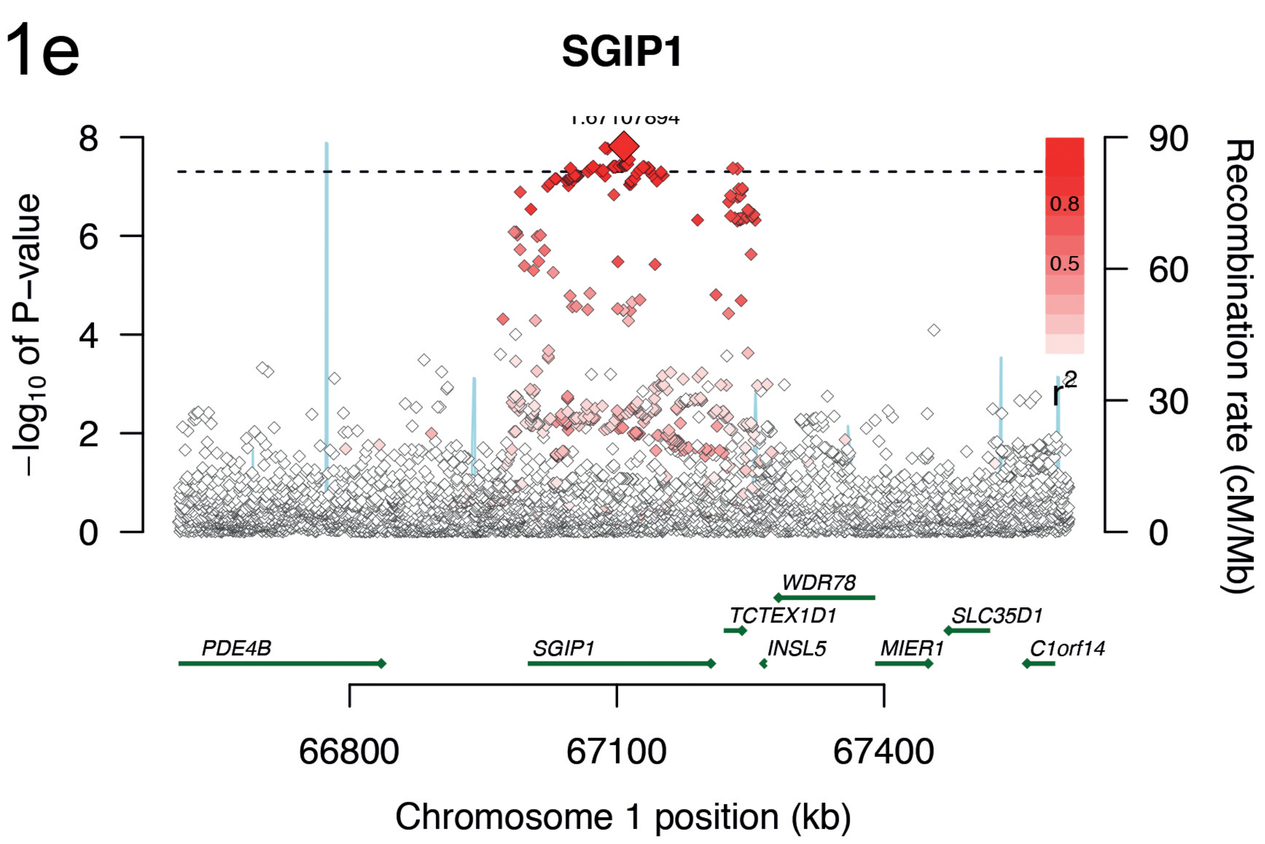
1c



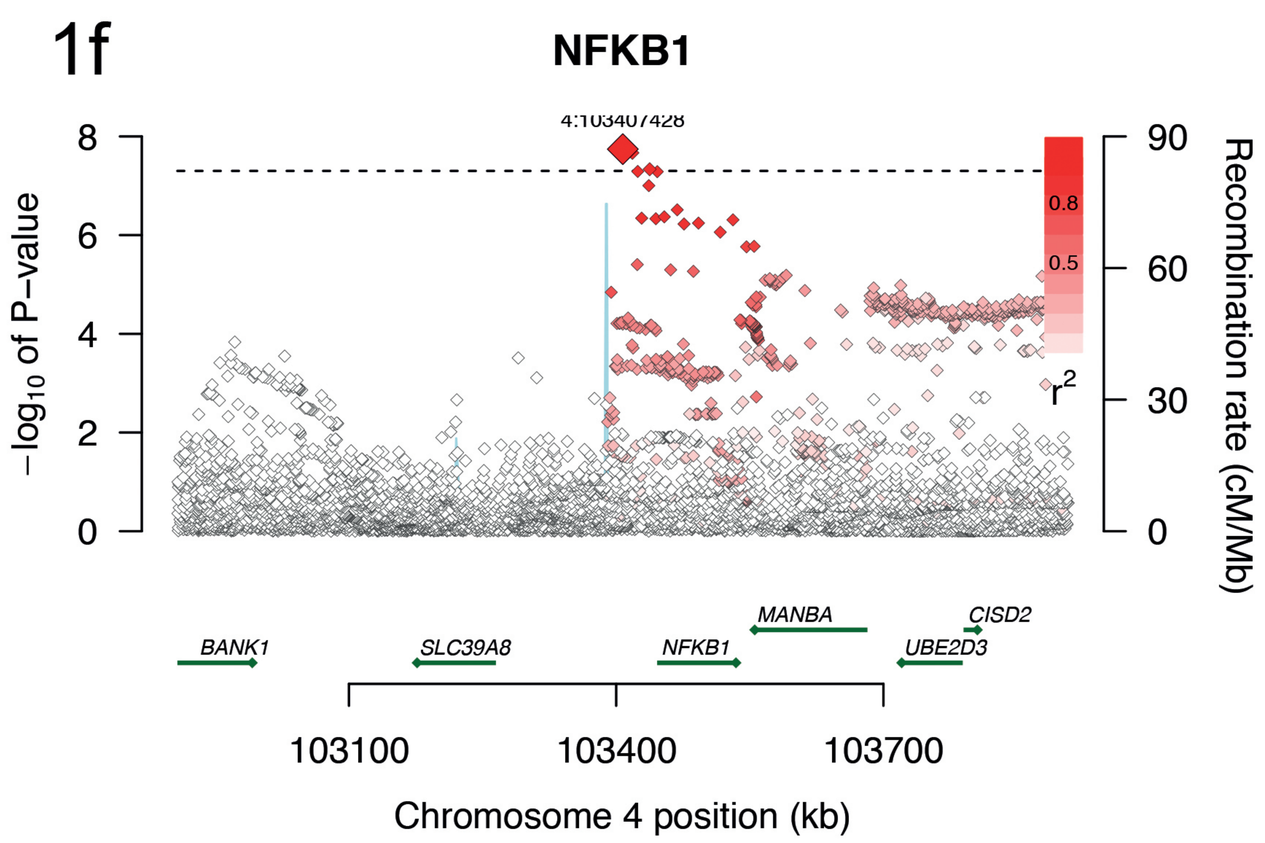
1d



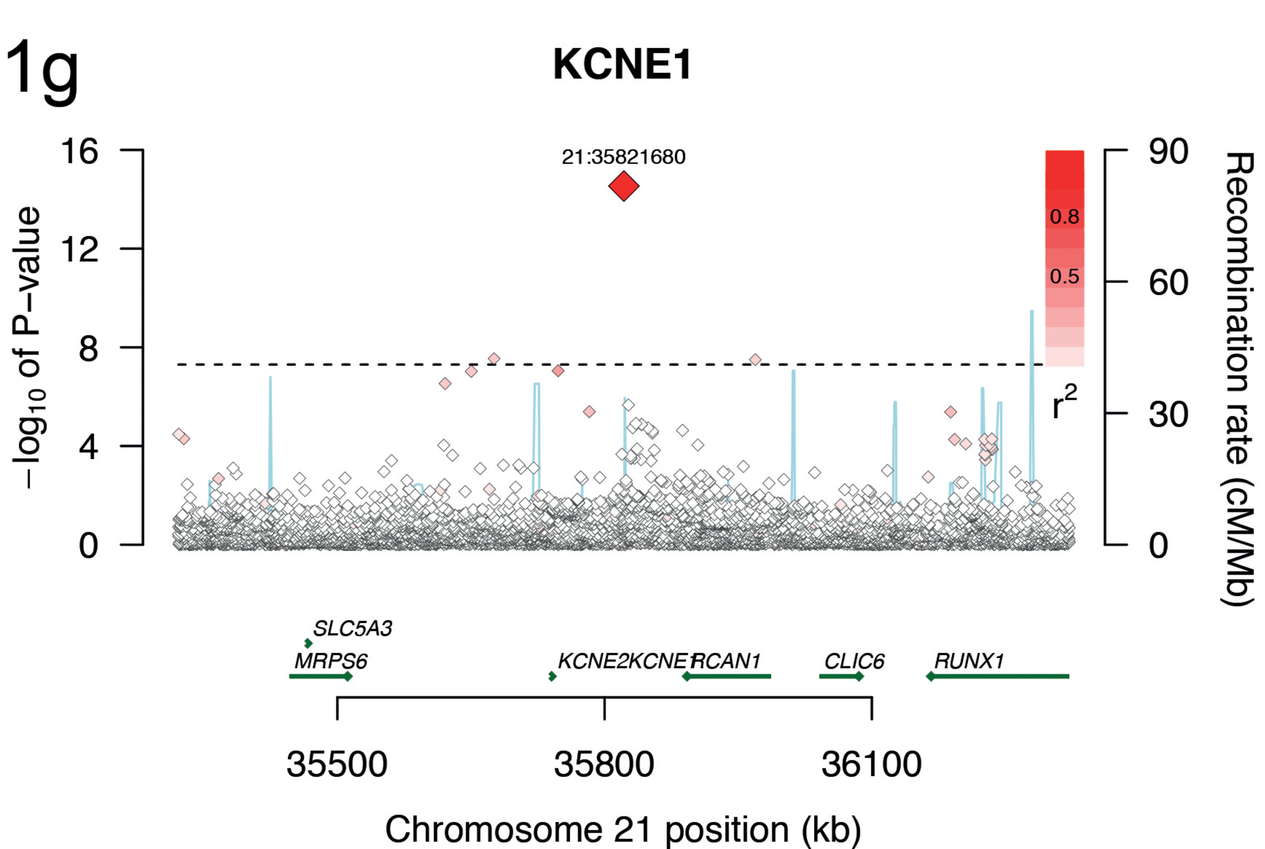
1e



1f



1g



1h