Discovery and fine mapping of candidate effector genes for heart rate

Julia Ramírez[1,2,3*], Stefan van Duijvenboden[3,4,5*], William J. Young[3,7], Yutang Chen[3,6], Tania Usman[8], Michele Orini[5], Pier Lambiase[5,7], Andrew Tinker[3,10], Christopher Bell[3], Andrew P Morris[11,12], Patricia B. Munroe[3,10]

[1]Aragon Institute of Engineering Research, University of Zaragoza, Spain.

[2]Centro de Investigación Biomédica en Red – Bioingeniería, Biomateriales y Nanomedicina, Spain.

[3]William Harvey Research Institute, Barts and the London Faculty of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, U.K.

[4]Nuffield Department of Population Health, University of Oxford, Oxford, U.K.

[5]Institute of Cardiovascular Science, University College London, London, U.K.

[6]Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland

[7]Barts Heart Centre, St Bartholomew's Hospital, London, EC1A 7BE, U.K.

[8]Kings College London, London, U.K.

[9]Khalifa University, Abu Dhabi, United Arab Emirates.

[10]National Institute of Health and Care Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, EC1M 6BQ, U.K.

[11]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, The University of Manchester, Manchester, U.K.

[12]National Institute of Health and Care Research, Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, U.K.

*These authors contributed equally to this work

Corresponding authors:

Patricia B Munroe, p.b.munroe@qmul.ac.uk

Charterhouse square, EC1M 6BQ

Queen Mary University of London, London, United Kingdom

Julia Ramírez, Julia.Ramirez@unizar.es

c/Mariano Esquillor s/n, 50018

University of Zaragoza, Zaragoza, Spain

Stefan van Duijvenboden, stefan.vanduijvenboden@ndph.ox.ac.uk

Old Road Campus, Headington, OX3 7LF, United Kingdom

University of Oxford, Oxford, United Kingdom

Current wordcount (max 4,000): 4,367.

Abstract

An elevated resting heart rate (RHR) is associated with an increased risk of cardiovascular mortality, but the mechanisms linking RHR to risk are still debated. Genome-wide association studies (GWAS) have identified over 350 loci for RHR explaining about 5% trait variance, but the effector genes and biological pathways at these loci are mostly unknown. We performed a meta-analysis of GWAS for RHR in 403,518 individuals from UK Biobank, including 388,237, 6,714 and 8,567 individuals of European, African, and South Asian ancestry, respectively, and performed annotation-informed fine-mapping in the European GWAS dataset. We discovered 318 RHR loci, including 23 not previously reported, which were validated in a larger meta-analysis. Annotation-informed fine mapping in Europeans indicated 442 signals at 307 loci. For 20% (90 signals), a single variant accounted for >75% posterior probability of being causal, 22 of which were annotated as missense. In disease-relevant tissues, 39 signals colocalised with cis-eQTLs, 3 with cis-pQTLs, and 75 had promoter interactions with HiC variants. In total, 262 candidate genes were highlighted, and enrichment analyses indicated cellular component, cardiovascular and autonomic nervous system amongst the top pathways. Twenty-three genes had additional support from human or mouse disorder and RNA or protein differential expression. Druggability analyses revealed potential drug repurposing opportunities, highlighting *ACHE*, *CALCRL*, *MYT1* and *TDP1* as potential targets. Our study provides support for 23 novel loci for RHR, and we highlight 262 effector genes which, with support from functional evidence, biological pathways and druggability analyses, unravel new mechanisms underlying RHR for therapeutic strategies.

Keywords (3 to 6): fine mapping, effector gene, resting heart rate, UK Biobank, genome-wide association study

Introduction

An elevated resting heart rate (RHR) has been associated with an increased risk of cardiovascular mortality and morbidity independent of traditional risk factors[1-3], but the exact mechanisms linking RHR to risk are still not clear. For example, observational studies have reported a U-shaped association between RHR and atrial fibrillation (AF), but mendelian randomization studies have found an inverse causal association between both[4,5]. Several potential mechanisms by which RHR may influence cardiovascular risk have been proposed, including the effect of RHR on coronary blood flow, cardiac contractility, energy expenditure, but the exact mechanisms are not fully understood.

Reduction of RHR using pharmacological inhibition of the pacemaker current has been shown to reduce the number of clinical events[6], suggesting that RHR is a modifiable risk factor. Genetics contributes to up to 20% of the interindividual variance in RHR[5,7] and, at present, genome-wide association studies (GWAS) and exome-array wide association studies have identified >350 loci for RHR explaining >5% of this variance[5,7-11]. The actual underlying effector genes remain unknown for most of these loci, which limits the understanding of the genetic and biological mechanisms of RHR. Consequently, identification of the responsible effector genes and biological pathways may reveal the mechanisms underlying RHR and its relation to cardiovascular risk, as well as novel opportunities to improve treatment and prevention of cardiovascular disease.

In the present work, we conducted a meta-analysis of RHR GWAS in European, African and South Asian ancestry in UK Biobank and an annotation-informed fine-mapping analysis to identify causal variants and candidate effector genes. Candidate effector genes prioritisation was based on evidence from functional annotation, colocalisation analyses with expression and protein quantitative loci (eQTLs and pQTLs) and promoter interactions in relevant RHR tissues. We investigated the biological pathways of the prioritised effector genes and searched for additional evidence of support from mouse and human phenotypes and differential expression. Finally, we assessed the potential of the prioritised effector genes for

drug target identification or repurposing opportunities. An overview of the study is shown in Figure 1.

Materials and Methods

UK Biobank

UK Biobank (UKB) is a prospective study of 502,366 volunteers, comprising relatively even numbers of men and women aged 40–69 years old at recruitment (2006-2008), with extensive baseline and follow-up clinical, biochemical, genetic and outcome measures. The study has approval from the North West Multi-Centre Research Ethics Committee, and all participants provided informed consent[12]. The work was undertaken as part of UKB application 8256.

Phenotypic data

Two replicate measurements of resting pulse rate, which is equivalent to RHR, were recorded (data field 102) at baseline assessment. After calculating the mean of these two measurements, we removed participants with extreme values (< 40 or > 120 bpm) or with a diagnosis of cardiovascular disease (with International Code of Disease – tenth revision matching those in Supplementary Table 1). We also excluded participants using beta-blockers (Supplementary Table 2) given their potential confounding influence on heart rate. After exclusions, we applied *kmeans* clustering to the first and second principal components data[13], resulting in a total of 388,237 individuals with European ancestry, 6,714 with African ancestry and 8,567 with South Asian ancestry.

Genetic QC

Following QC procedures already carried out centrally by UKB (Supplemental Methods), we exclude discordant SNVs and samples with QC failures, gender discordance, low imputation quality (INFO score ≤ 0.3) and high heterozygosity/missingness (Hardy-weinberg equilibrium test with $P < 10^{-6}$ and missingness > 0.015).

Ancestry-specific single variant RHR association analyses

Previously, we had performed a GWAS in the 388,237 individuals of European ancestry[14]. In this work, we additionally performed two GWAS separately for individuals with African and South Asian ancestry. We used SNPTest[15] in the African and South Asian GWAS (after removing related individuals (N = 381 and N = 531, respectively). Each GWAS included the following covariates: age, age squared, sex, body mass index, the first 10 principal components, and a binary indicator variable for UK Biobank versus UK BiLEVE to adjust for the different genotyping arrays. We only considered single-nucleotide variants (SNVs) with minor allele frequency (MAF) ≥ 1%. A *P* value threshold of $< 5 \times 10^{-8}$ was used as the genome-wide significance level.

Meta-Analysis of RHR

We undertook a one-stage, single-discovery design using an inverse variance-weighted, fixed effects model using METAL (version released 2011-03-25[16]) to meta-analyse each of the three ancestry GWAS summary statistics. Quality control was performed using the EasyQC R package (version 23.8[17]). Genomic control was applied during meta-analysis to ancestries in which the inflation factor (λ) was > 1.0. All variants with $P < 5 \times 10^{-8}$ were compiled and mapped to individual loci based on genomic distance of >500 kb to each side of another variant. If multiple variants fit the selection criteria for a single region, only the variant with the smallest *P* value was considered for follow-up.

A locus was declared potentially novel if there were no previous studies[5,7,10] reporting genome-wide significant variants in the 1Mb region. To validate the novelty of loci, we meta-analysed the summary statistics of the lead SNVs from our meta-analysis RHR with those from the International Consortium for Resting Heart Rate (IC_RHR) summary statistics[5], using METAL. Novelty was declared if the loci were genome-wide significant in the meta-analysis (across UK Biobank and IC_RHR), and the direction of effect was concordant.

Identification of distinct associations signals for fine-mapping

We conducted fine-mapping analysis of the GWAS summary statistics from individuals with European ancestry. We initially defined loci as mapping 500kb up- and down-stream of each lead SNV (at genome-wide significance, $P < 5 \times 10^{-8}$). However, where loci overlapped, they were combined as a single locus. We then performed approximate conditional analyses using GCTA-COJO[18] to detect distinct association signals at each locus, using European ancestry haplotypes from the 1000 Genomes Project (Phase 3, October 2014 release)[19] as a reference for linkage disequilibrium (LD). Within each locus, variants attaining genome-wide significance in the joint GCTA-COJO model were selected as index SNVs for distinct association signals.

Enrichment of RHR associations for genomic annotations

We used fGWAS[20] to identify genomic annotations from a total of 253 functional and regulatory annotations[21,22], enriched for RHR association signals (Supplemental Methods). We then used an iterative approach to identify a joint model of enriched annotations using a forward-selection approach. At each iteration, we added the annotation to the joint fGWAS model that maximised the improvement in the penalised likelihood. We continued until no additional annotations improved the fit of the joint model ($P < 0.00020$, Bonferroni correction for 253 annotations).

Fine-mapping distinct association signals for RHR

For each jth variant at the ith distinct signal, we first estimated its prior probability of causality using an annotation-informed prior model:

$$\gamma_j = \exp\left[\sum_k \hat{\beta}_k \, z_{jk}\right],$$

where the summation is over the enriched annotations, $\hat{\beta}_k$ is the estimated log-fold enrichment of the kth annotation from the final joint fGWAS model, and z_{jk} is an indicator

variable taking the value 1 if the jth variant maps to the kth annotation, and 0 otherwise. We then approximated the Bayes' factor, Λ_{ij}, using the European ancestry summary statistics, as previously described[23,24] (Supplemental Methods). Finally, we estimated the jth variant posterior probability of causality as $\pi_{ij} \propto \gamma_j \Lambda_{ij}$.

Finally, we derived a 99% credible set[25] for the ith distinct association signal by: (i) ranking all SNVs according to π_{ij}; and (ii) including ranked variants until their cumulative posterior probability attains or exceeds 99%[23,24]. The credible set would, then, include the minimum number of variants that jointly explained >99% of the posterior probability of driving the RHR association under the annotation-informed prior. We defined high-confidence causal variants as single variants from the credible sets accounting for more than 75% of the posterior probability.

Functional annotation of variants

We used variant-effect predictor (VEP) analysis[26] to annotate the high-confidence causal variants from the credible sets, and selected those annotated as missense variants.

Colocalisation with gene expression data

We integrated genetic fine-mapping data with expression quantitative trait loci (*cis*-eQTL) in adrenal gland, artery, heart, nerve and brain tissues from the GTEx Consortium[27]. The tissue selection was informed by tissue enrichment analysis from prior GWAS (artery, heart and adrenal gland[10]) and biological mechanisms known to regulate RHR (nerve, and brain). We first did a lookup of significant lead eQTL variants in the 99% credible sets. For each signal where we detected overlap, we formally assessed whether the annotation informed Bayes' factor for the credible set variants of the corresponding signal colocalised with the eQTL results, as previously described[24]. We used publicly available eQTL results from GTEx version 8[27] (Supplemental Methods).

Long-range chromatin interaction (Hi–C) analyses

We identified potential target genes of regulatory SNVs using long-range chromatin interaction (Hi–C) data from adrenal gland, aorta, left and right ventricles, hippocampus and cortex, similar tissues as selected for eQTL analysis. Hi–C data was corrected for genomic biases and distance using the Hi–C Pro and Fit-Hi-C pipelines according to Schmitt et al.[28]. We followed a similar procedure as in the eQTL colocalization analysis. We first identified signals where potential regulatory SNVs (RegulomeDB score ≤2) where in the 99% credible set. Then, we reported the interactors with the SNVs of highest regulatory potential to annotate the signals.

Colocalisation with protein expression data

We additionally integrated genetic fine-mapping data with protein quantitative trait loci (*cis*-pQTL) in plasma[29]. We performed the same Bayesian statistical procedure as for eQTL colocalisation to assess whether those signals for which a 99% credible set variant was the lead pQTL variant colocalised with pQTL results.

Prioritisation of candidate effector RHR genes

A full list of candidate effector genes for RHR was collated from the results of our fine-mapping pipeline and computational approaches, as done recently[24]. A gene was indicated for a signal if there was support from a coding and high-confidence variant in the gene at the locus, or if the gene was indicated from eQTL, pQTL colocalization or Hi-C analyses.

Effector gene pathway analysis

We used the Gene2Function analysis tool in FUMA (v1.4.0) to perform geneset enrichment on the prioritised list of candidate genes, and to identify significantly associated Gene Ontology (GO) terms and pathways[30]. Redundant GO terms were removed using the

Reduce and Visualize Gene Ontology (REVIGO) web application[31]. REVIGO uses a hierarchical clustering method to remove highly similar terms, incorporating enrichment P-values in the selection process. Default settings (dispensability cut off <0.7) were used in this analysis.

Additional evidence for effector genes from mouse and human phenotypes and differential expression

We collated additional information for each prioritised candidate gene using data from GeneCards[32] (https://genealacart.genecards.org). This included the following: 1) a mouse model from Mouse Genome Informatics which has a cardiovascular phenotype. 2) A cardiovascular or vascular phenotype described for the candidate gene in the Human Phenotype Ontology database. 3) Differential RNA expression of the candidate gene in the GTEx database in cardiovascular or vascular tissues, only genes with fold changes >4 in a tissue were selected. 4) Differential protein expression of the candidate gene based on 69 integrated normal proteomics datasets in HIPED (the Human Integrated Protein Expression Database). Genes with a fold change value of >6 and protein abundance value of >0.1 PPM in an anatomical were selected.

Druggability of prioritised effector genes

To identify candidate druggable targets, a look-up was done of the prioritised list of candidate genes in a previously published database of the druggable genome developed by Finan et al[33]. To identify potential opportunities for drug repurposing, a look-up of each candidate gene was performed for Tier 1 to identify any existing drug targets (Supplemental Methods).

Results

RHR GWAS results

We identified 318 genome-wide significant loci in the UK Biobank meta-analysis (Supplementary Table 3), and there was no evidence of heterogeneity of test statistics (Supplementary Figure 1).

The European-only GWAS discovered 307 genome-wide significant loci (Supplementary Table 4), 295 of which were genome-wide significant in the full meta-analysis. The African-only GWAS identified one genome-wide significant locus (*GJA1-HSF2)*, which had been previously reported in Europeans (Supplementary Table 5). No genome-wide significant loci were found in the South Asian-only GWAS.

Novel loci for RHR

Thirty-six potentially novel loci were identified in the meta-analysis GWAS (9 of these were not genome-wide significant in the European GWAS). From these, 23 were genome-wide significant in the meta-analysis with IC_RHR summary statistics (Table 1, Supplementary Table 6, Supplementary Figure 2), and declared as novel.

Fine-mapping and genomic annotation reveals high-confidence causal variants for RHR

Using European GWAS[14] results, through approximate conditional analyses, we partitioned RHR associations at the 307 loci into a total of 442 distinct association signals that were genome-wide significant (Supplementary Table 7).

We next mapped SNVs to functional and regulatory annotations. We observed significant joint enrichment for RHR associations mapping to protein coding exons and 5' UTRs, enhancers in the heart, and promoters in the right ventricle (Main Figure 2, Supplementary Table 8).

Using these enriched annotations, for each of the 442 distinct signals, we derived 99% credible sets of variants. The median 99% credible set size for RHR was 26 variants (Supplementary Table 7). For 90 (20.4%) RHR signals, a single SNV accounted for >75% of the posterior probability of driving the RHR association under the annotation-informed prior, which we defined as "high-confidence" for causality (Main Figure 3, Supplementary Table 9).

Missense variants implicate causal candidate genes

From the 90 high-confidence variants, 22 were missense variants (Main Table 2), of which six were annotated as damaging and deleterious by PolyPhen and SIFT, respectively. Four variants annotated as probably damaging and deleterious were in *APOE (*rs7412, p.Arg176Cys*), LAMB1* (rs80095409, p.Arg795Gly) and *CCDC141* (rs17362588, p.Arg935Trp and rs10497529, p.Ala141Val), Main Table 2). The *APOE* variant, rs7412, determines the *APOE2* isoform, which has been shown in both human and animal studies to be protective against Alzheimer Disease[34] and increase longevity[35]. Previous studies suggest that laminins have important roles in human heart development and function[36,37]. In particular for *LAMB1,* zebrafish embryos had mild morphogenetic defects and progressive cardiomegaly, as well as a limited heart size during cardiac development[38]. *CCDC141* is a less characterised gene involved in axon guidance and cell adhesion and plays a critical role in radial migration and centrosomal function.

Three missense variants were annotated as possibly damaging and deleterious by PolyPhen and SIFT, respectively (*GAB1* (rs28925904, p.Pro311Leu), *ARHGEF40* (rs12889267, p.Lys293Glu) and *FHOD3* (rs61735998, p.Val647Phe), respectively). The variants in *ARHGEF40* and *FHOD3* had a 100% posterior probability of driving the RHR signal. *GAB1* is an adapter protein that plays a role in intracellular signaling cascades triggered by activated receptor-type kinases[39]. Cardiac *GAB1* deletion has been reported to lead to dilated cardiomyopathy associated with mitochondrial damage and cardiomyocyte apoptosis[40]. *ARHGEF40* has previously been associated with all-cause mortality[10], but is less functionally characterised and encodes a protein similar to guanosine nucleotide exchange factors for Rho GTPases. Finally, *FHOD3* is essential for myofibrillogenesis at an early stage of heart development[41].

Effector genes identified using gene expression in disease relevant tissues

Convincing support for colocalization with gene expression in at least one tissue was identified for 39 distinct high confidence signals at 37 loci (Supplementary Table 10). There was a total of 36 genes with tissue-specific colocalizations, of which 29 (81%) were in heart or arterial tissue, and 7 (19%) in brain. There were no specific colocalisations in adrenal gland tissues (Supplementary Figure 3). An interesting candidate gene with heart-specific colocalisation is *PLEC*. *PLEC* mouse models show right bundle branch block and abnormal heart morphology, and a missense variant in *PLEC* has been reported to increase risk of AF[42] in humans.

There were a few genes with brain-specific colocalisations, including *NKX2-5*, *LEMD2* and *UCK1*. *NKX2-5* is an eQTL in brain in our data, but its specific function is not well understood. There are mouse and zebrafish models with cardiovascular phenotypes[43], and previous GWAS have identified variants significantly associated with CAD[44] and AF[44-46], as well as for ECG traits, including the QT interval[47] and the T-peak-to-T-end interval[48]. There are *LEMD2* mouse and human models that include an arrhythmogenic cardiomyopathy phenotype[49]. Finally, *UCK1* phosphorylates uridine and cytidine to uridine monophosphate and cytidine monophosphate (GeneCards), but there is no data indicating association with RHR or cardiovascular phenotypes.

Identification of effector genes using promotor-centred long-range chromatin interactions

Promoter interactions and candidate genes were identified for 75 unique high confidence signals at 66 loci (Supplementary Table 11). A total of 107 genes were indicated in a single tissue, of which 14 (13%) where in left or right ventricle, 83 (78%) in brain, and 10 (9%) in adrenal gland. From the genes specifically indicated in heart tissue, *SCN10A* has been thoroughly characterised as a RHR modifier[10,13,50] and *CASZ1* is involved in cardiac morphogenesis and development[51], and there are abnormal mouse phenotype including congenital and structural cardiomyopathies[52-54]. There are also some genes indicated that do not have experimental support for cardiovascular traits, these include *CABLES1*, *BET1* and *SLC22A17*. *CABLES1* encodes a protein involved in regulation of the cell cycle through

interactions with several cyclin-dependent kinases. *BET1* encodes a golgi-associated membrane protein that participates in vesicular transport from the endoplasmic reticulum to the Golgi complex. *SLC22A17* is a cell surface receptor for *Lipocalin 2*, an antibacterial protein that acts by sequestering iron during bacterial infection and has recently been reported to be involved in various pathophysiological conditions in various organs and tissues, including the heart and brain[55].

Two genes indicated in brain tissue with a low (2a) regulome score include *CEP68*, and *CISD3 (*Supplementary Table 11**)**. *CEP68* has a mouse cardiovascular phenotype (increased heart weight), and variants at this locus have previously been associated with AF[56,57]. *CISD3* may play a role in regulating electron transport and oxidative phosphorylation[58], and diseases associated with this gene include Wolfram Syndrome (OMIM number 222300).

Effector genes identified using protein expression

We identified significant pQTLs for 3 genes, *GCKR*, *ENO3* and *MXRA7* (Supplementary Table 12), which also had support from missense annotation (*GCKR*) and eQTL analyses (*ENO3* and *MXRA7*). *GCKR* regulates glucokinase by forming an inactive complex with this enzyme[59,60]. Postprandial triglyceridemia is an emerging risk factor for cardiovascular disease and *GCKR* gene polymorphism affects postprandial lipemic response in a dietary intervention study[61]. *ENO3* has demonstrated increased differential expression in the left ventricle in rats[62]. Finally, the role *MXRA7* (matrix-remodeling-associated protein 7) in modulating RHR is less known.

Candidate gene prioritisation

From the complementary fine mapping and computational approaches (high-confidence missense, colocalised eQTLs and pQTLs and Hi-C interactions), we prioritised a total of 262 candidate genes for RHR that had at least one line of evidence (Main Table 3**,** Supplementary Table 13).

Biological pathways

To gain insights into the biological role of the 262 candidate genes for RHR, we performed gene-set enrichment analyses. We found significant enrichment for 41 unique GO biological processes (Supplementary Table 14). The most significant biological processes included cellular component morphogenesis ($P = 2.1 \times 10^{-4}$), neuron differentiation ($P = 4.6 \times 10^{-4}$), and neuron development ($P = 8.9 \times 10^{-4}$).

Additional functional evidence

We observed 74 of the 262 prioritised candidate genes (28.2%) had support from mouse model data and 45 (17.2%) from human cardiovascular phenotypes (25 genes had support from both mouse and human cardiovascular phenotypes). We also found 11 candidate genes (4.2%) had support from RNA or protein differential expression. In total, 23 candidate genes (9.2%) had additional functional evidence from mouse or human phenotypes and RNA or protein differential expression (Main Table 3).

Drug target identification and repositioning opportunities

We found 21 of the 262 candidate effector genes were existing targets of small molecules or biotherapeutics and clinical drug candidates (Tier 1, Supplementary Table 15). Of these, *CACNA1D*, *MYH6* and *SCN10A* are the top gene targets for a cardiovascular disease (sinoatrial node dysfunction, hypertrophic cardiomyopathy and AF, respectively). The remaining 18 genes are existing targets of drugs for diseases not involving the cardiovascular system, suggesting potential drug repurposing.

Discussion

In the present work, we used the enhanced statistical power of our multi-ancestry meta-analysis in UK Biobank to discover 23 previously unreported loci. In parallel, we employed a robust statistical framework for identifying genomic annotations that are most

relevant to the biology of RHR and used this information to advance from initial broadly associated genomic regions to the prioritisation of 262 candidate effector genes for RHR.

Whilst preparing this manuscript, Van de Vegte et al. published the largest genome-wide meta-analysis comprising 100 studies in up to 835,465 individuals[5]. Their analysis revealed 493 independent genetic variants in 352 loci[5]. Despite their larger sample size, our analysis discovered 36 loci, of which 23 validated that were not reported in their work. Potential explanations in discordancy of results between studies could be that we meta-analysed ancestry specific GWASs in our discovery analysis whereas they conducted single GWASs with all ancestries combined, including UK Biobank. Their analysis was also not limited to three ancestries, like ours.

A review of candidate genes for the 23 novel loci highlighted two interesting genes implicated in the cardiovascular system, *ETV1* and *KCNE3*. *ETV1* is a critical transcription factor in determining fast conduction physiology in the heart[63,64], as well as atrial remodelling associated with atrial arrhythmias[65,66]. Functional studies have reported that mutations in *KCNE3* can underlie the development of Brugada syndrome, an inherited syndrome associated with a high incidence of sudden cardiac arrest[67]. We also found relevant genes implicated in the nervous system*, FAM155A, ZNF804A* and *HLA-DRB5. FAM155A* belongs to a large protein complex termed the *NALCN* channelosome, which is a sodium leak channel expressed mainly in the central nervous system that is responsible for the resting Na+ permeability that controls neuronal excitability[68]. *ZNF804A* affects brain structure and function, modulating neurodevelopment and translation[69]. Finally, the *HLA-DRB5* gene was also highlighted in a previous study of our team investigating the genetic architecture of two phenotypes reflecting the sympathovagal balance, HR response to exercise and recovery[13].

We observed relatively strong support for two loci from individuals of European and African ancestry, *METTL15,* a novel protein necessary for efficient translation in human mitochondria[70], and *KCNE3*, described earlier. Support for the *SIM1* locus came from European and South Asian ancestry samples (Main Table 1). *SIM1* is involved in the development and function of the paraventricular nucleus of the hypothalamus.

Our genetic fine-mapping pipeline prioritised 262 candidate effector genes for RHR for future focused functional validation, and for 23 of these genes we found additional functional evidence from existing mouse or human phenotypes and RNA or protein differential expression data. We highlight two genes from our refined list of 23 genes not previously prioritised as candidate genes for RHR; *CACNA1D* and *RBM20*. *CACNA1D* is present in the membrane of most excitable cells and mediates calcium influx in response to depolarization. Associated diseases include sinoatrial node dysfunction and deafness. *RBM20* acts as a regulator of mRNA splicing of a subset of genes encoding key structural proteins involved in cardiac development, such as *TTN*, *CACNA1C*, *CAMK2D* or *PDLIM5/ENH*[71-80]). Mutations in this gene have been associated with familial dilated cardiomyopathy[81].

Van de Vegte et al did not perform fine-mapping, but used a scoring system for gene prioritisation based on four criteria: proximity, coding, eQTL and DEPICT, highlighting 39 genes with high levels of support, *PHACTR2*, *ENO3* and *SENP2* being their top effector genes[5]. A direct comparison with our results is complicated, because our methodology ranks variants based on posterior probability, instead of P-values, and also the definition of signal might differ. We found 12 of their 39 genes were amongst the 262 candidate genes prioritised by our fine-mapping pipeline, and 5 of these were in our list of 23 genes with additional evidence. Moreover, at six genomic regions, the candidate gene prioritised by our pipeline was different. Our method also offered more granularity in suggesting candidate genes at loci, prioritising additional candidate genes at five genomic regions (Supplemental Table 16).

Our enrichment analyses indicated cellular component and animal organ morphogenesis, neuron differentiation and development and circulatory system differentiation and development as top biological pathways of the 262 candidate genes. Twenty-nine of the 41 significant biological pathways identified using the 262 prioritised candidate genes implicate at least one of the 23 candidate genes with additional functional evidence, strengthening further our fine-mapping pipeline to prioritise RHR effector genes for subsequent functional assessment. Previous studies have reported enrichment of

associations in pathways involved in cardiac tissue development, muscle cell differentiation and pro-arrhythmic pathways[5,9,11]. The prominence of nervous system pathways observed in our findings, in addition to cardiac, supports the validity of our effector genes as modulators of RHR.

Druggability analyses highlighted the role of *CACNA1D*, *MYH6* and *SCN10A* as top gene targets for cardiovascular disease, all genes which are well known to play a role in the cardiovascular system. We also found four top gene targets for a neurological disease with repurposing potential, *ACHE*, *CALCRL*, *MYT1* (autism spectrum disorder) and *TDP1*. *ACHE* is a target of drugs for Alzheimers disease including Donepezil and Galantamine, which cause bradycardia as a side effect. *CALCRL* is a target of drugs for migraine disorder. It is a receptor for adrenomedullin, together with *RAMP2*[82,83]. One of the reported mouse phenotypes is differences in heart rate of heterozygous *CALCRL* female and male mice[84]. *MYT1* is a drug target for autism spectrum disorder[85] and is less characterised, it binds to the promoter regions of proteolipid proteins of the central nervous system and plays a role in the developing nervous system[86,87]. Finally, *TDP1* is a drug target for spinocerebellar ataxia type 1 with axonal neuropathy[88] and mouse models have identified abnormal retina vessel phenotypes.

We performed annotation-informed fine mapping analyses on the European GWAS, despite finding more loci in the multi-ancestry meta-analysis. This decision was made to avoid the calibration issues of meta-analysis fine-mapping over individual GWAS[89]. Despite this, we are aware that the lack of population diversity in our fine-mapping approach is a limitation. An additional weakness is that whilst benefitting from dense genotyping and imputation of common variants, this is not exhaustive in capturing all the potential phenotypically associated genetic variation within each locus. This will miss the possible impact of rare variants, as well as any poorly tagged larger variants (copy number variants, short tandem repeats, inversions, etc.).

In conclusion, we have identified 23 novel genes for RHR, and prioritised 262 candidate genes using annotation-informed fine mapping. Functional evidence, biological pathways and druggability analyses provide support to these genes and unravel mechanisms

underlying RHR. Our findings will further investigations to improve the functional understanding of the biology underlying RHR and enable novel preventive and therapeutic opportunities.

Data Availability Statement

References

1. Fox, K. *et al.* Resting Heart Rate in Cardiovascular Disease. *Journal of the American College of Cardiology* **50**, 823-830 (2007).
2. Cooney, M.T. *et al.* Elevated resting heart rate is an independent risk factor for cardiovascular disease in healthy men and women. *American Heart Journal* **159**, 612-619.e3 (2010).
3. Zhang, D., Shen, X. & Qi, X. Resting heart rate and all-cause and cardiovascular mortality in the general population: a meta-analysis. *Canadian Medical Association Journal* **188**, E53-E63 (2016).
4. Klevjer, M. *et al.* Insight into the relationship between resting heart rate and atrial fibrillation: a Mendelian randomization study. *EP Europace* **25**(2023).
5. van de Vegte, Y.J. *et al.* Genetic insights into resting heart rate and its role in cardiovascular disease. *Nature Communications* **14**, 4646 (2023).
6. Cargnoni, A., Ceconi, C., Stavroula, G. & Ferrari, R. Heart rate reduction by pharmacological If current inhibition. *Adv Cardiol* **43**, 31-44 (2006).
7. Guo, Y. *et al.* Genome-Wide Assessment for Resting Heart Rate and Shared Genetics With Cardiometabolic Traits and Type 2 Diabetes. *Journal of the American College of Cardiology* **74**, 2162-2174 (2019).
8. Eijgelsheim, M. *et al.* Genome-wide association analysis identifies multiple loci related to resting heart rate. *Human Molecular Genetics* **19**, 3885-3894 (2010).
9. den Hoed, M. *et al.* Identification of heart rate–associated loci and their effects on cardiac conduction and rhythm disorders. *Nature Genetics* **45**, 621-631 (2013).
10. Eppinga, R.N. *et al.* Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nature Genetics* **48**, 1557-1563 (2016).
11. van den Berg, M.E. *et al.* Discovery of novel heart rate-associated loci using the Exome Chip. *Human Molecular Genetics* **26**, 2346-2363 (2017).

12. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779 (2015).
13. Ramírez, J. *et al.* Thirty loci identified for heart rate response to exercise and recovery implicate autonomic nervous system. *Nature Communications* **9**, 1947 (2018).
14. Mensah-Kane, J. *et al.* No Clinically Relevant Effect of Heart Rate Increase and Heart Rate Recovery During Exercise on Cardiovascular Disease: A Mendelian Randomization Analysis. *Frontiers in Genetics* **12**(2021).
15. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* **11**, 499-511 (2010).
16. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
17. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols* **9**, 1192-1212 (2014).
18. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369-375 (2012).
19. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
20. Pickrell, Joseph K. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics* **94**, 559-573 (2014).
21. Harrow, J. *et al.* GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Research* **22**, 1760-1774 (2012).
22. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
23. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics* **50**, 1505-1513 (2018).
24. van Duijvenboden, S. *et al.* Integration of genetic fine-mapping and multi-omics data reveals candidate effector genes for hypertension. *bioRxiv*, 2023.01.26.525702 (2023).
25. Maller, J.B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics* **44**, 1294-1301 (2012).
26. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).

27. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
28. Schmitt, Anthony D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports* **17**, 2042-2059 (2016).
29. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nature Genetics* **53**, 1712-1721 (2021).
30. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
31. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).
32. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics* **54**, 1.30.1-1.30.33 (2016).
33. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine* **9**, eaag1166 (2017).
34. Li, Z., Shue, F., Zhao, N., Shinohara, M. & Bu, G. APOE2: protective mechanism and therapeutic implications for Alzheimer's disease. *Molecular Neurodegeneration* **15**, 63 (2020).
35. Shinohara, M. *et al.* APOE2 is associated with longevity independent of Alzheimer's disease. *eLife* **9**, e62199 (2020).
36. Haag, T.A., Haag, N.P., Lekven, A.C. & Hartenstein, V. The Role of Cell Adhesion Molecules inDrosophilaHeart Morphogenesis:Faint Sausage, Shotgun/DE-Cadherin, andLaminin AAre Required for Discrete Stages in Heart Development. *Developmental Biology* **208**, 56-69 (1999).
37. Yarnitzky, T. & Volk, T. Laminin Is Required for Heart, Somatic Muscles, and Gut Development in the Drosophila Embryo. *Developmental Biology* **169**, 609-618 (1995).
38. Derrick, C.J. *et al.* Lamb1a regulates atrial growth by limiting second heart field addition during zebrafish heart development. *Development* **148**(2021).
39. Yousaf, R. *et al.* Modifier variant of METTL13 suppresses human GAB1–associated profound deafness. *The Journal of Clinical Investigation* **128**, 1509-1522 (2018).
40. Zhao, J. *et al.* Cardiac Gab1 deletion leads to dilated cardiomyopathy associated with mitochondrial damage and cardiomyocyte apoptosis. *Cell Death & Differentiation* **23**, 695-706 (2016).
41. Kan-O, M. *et al.* Mammalian formin Fhod3 plays an essential role in cardiogenesis by organizing myofibrillogenesis. *Biology Open* **1**, 889-896 (2012).
42. Thorolfsdottir, R.B. *et al.* A Missense Variant in PLEC Increases Risk of Atrial Fibrillation. *Journal of the American College of Cardiology* **70**, 2157-2168 (2017).

43. Harrington, J.K., Sorabella, R., Tercek, A., Isler, J.R. & Targoff, K.L. Nkx2.5 is essential to establish normal heart rate variability in the zebrafish embryo. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **313**, R265-R271 (2017).
44. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics* **53**, 1415-1424 (2021).
45. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nature Genetics* **50**, 1225-1233 (2018).
46. Nielsen, J.B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nature Genetics* **50**, 1234-1239 (2018).
47. van Duijvenboden, S. *et al.* Genomic and pleiotropic analyses of resting QT interval identifies novel loci and overlap with atrial electrical disorders. *Human Molecular Genetics* **30**, 2513-2523 (2021).
48. Ramírez, J. *et al.* Common Genetic Variants Modulate the Electrocardiographic Tpeak-to-Tend Interval. *The American Journal of Human Genetics* **106**, 764-778 (2020).
49. Gerull, B. & Brodehl, A. Genetic Animal Models for Arrhythmogenic Cardiomyopathy. *Frontiers in Physiology* **11**(2020).
50. Delaney, J.T. *et al.* Common SCN10A variants modulate PR interval and heart rate response during atrial fibrillation. *EP Europace* **16**, 485-490 (2013).
51. Liu, Z. *et al.* Essential Role of the Zinc Finger Transcription Factor Casz1 for Mammalian Cardiac Morphogenesis and Development*. *Journal of Biological Chemistry* **289**, 29801-29816 (2014).
52. Kennedy, L. *et al.* Formation of a TBX20-CASZ1 protein complex is protective against dilated cardiomyopathy and critical for cardiac homeostasis. *PLOS Genetics* **13**, e1007011 (2017).
53. Huang, R.-T. *et al.* CASZ1 loss-of-function mutation associated with congenital heart disease. *Gene* **595**, 62-68 (2016).
54. Spielmann, N. *et al.* Extensive identification of genes involved in congenital and structural heart disorders and cardiomyopathy. *Nature Cardiovascular Research* **1**, 157-173 (2022).
55. Lim, D., Jeong, J.-h. & Song, J. Lipocalin 2 regulates iron homeostasis, neuroinflammation, and insulin resistance in the brains of patients with dementia: Evidence from the current literature. *CNS Neuroscience & Therapeutics* **27**, 883-894 (2021).
56. Christophersen, I.E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature Genetics* **49**, 946-952 (2017).

57. Sinner, M.F. *et al.* Integrating Genetic, Transcriptional, and Functional Analyses to Identify 5 Novel Genes for Atrial Fibrillation. *Circulation* **130**, 1225-1235 (2014).
58. Wiley, S.E., Murphy, A.N., Ross, S.A., van der Geer, P. & Dixon, J.E. MitoNEET is an iron-containing outer mitochondrial membrane protein that regulates oxidative capacity. *Proceedings of the National Academy of Sciences* **104**, 5318-5323 (2007).
59. Pautsch, A. *et al.* Crystal Structure of Glucokinase Regulatory Protein. *Biochemistry* **52**, 3523-3531 (2013).
60. Choi, J.M., Seo, M.-H., Kyeong, H.-H., Kim, E. & Kim, H.-S. Molecular basis for the role of glucokinase regulatory protein as the allosteric switch for glucokinase. *Proceedings of the National Academy of Sciences* **110**, 10171-10176 (2013).
61. Shen, H. *et al.* Glucokinase regulatory protein gene polymorphism affects postprandial lipemic response in a dietary intervention study. *Human Genetics* **126**, 567-574 (2009).
62. Giusti, B. *et al.* Gene expression profile of rat left ventricles reveals persisting changes following chronic mild exercise protocol: implications for cardioprotection. *BMC Genomics* **10**, 342 (2009).
63. Shekhar, A. *et al.* Transcription factor ETV1 is essential for rapid conduction in the heart. *The Journal of Clinical Investigation* **126**, 4444-4459 (2016).
64. Shekhar, A. *et al.* ETV1 activates a rapid conduction transcriptional program in rodent and human cardiomyocytes. *Scientific Reports* **8**, 9944 (2018).
65. Rommel, C. *et al.* The Transcription Factor ETV1 Induces Atrial Remodeling and Arrhythmia. *Circulation Research* **123**, 550-563 (2018).
66. Yamaguchi, N. *et al.* Cardiac Pressure Overload Decreases ETV1 Expression in the Left Atrium, Contributing to Atrial Electrical and Structural Remodeling. *Circulation* **143**, 805-820 (2021).
67. Delpón, E. *et al.* Functional Effects of KCNE3 Mutation and Its Role in the Development of Brugada Syndrome. *Circulation: Arrhythmia and Electrophysiology* **1**, 209-218 (2008).
68. Xie, J. *et al.* Structure of the human sodium leak channel NALCN in complex with FAM155A. *Nature Communications* **11**, 5831 (2020).
69. Zhou, Y. *et al.* Interactome analysis reveals ZNF804A, a schizophrenia risk gene, as a novel component of protein translational machinery critical for embryonic neurodevelopment. *Molecular Psychiatry* **23**, 952-962 (2018).
70. Haute, L.V. *et al.* METTL15 introduces N4-methylcytidine into human mitochondrial 12S rRNA and is required for mitoribosome biogenesis. *Nucleic Acids Research* **47**, 10267-10281 (2019).
71. Guo, W. *et al.* RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nature Medicine* **18**, 766-773 (2012).

72. Maatz, H. *et al.* RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *The Journal of Clinical Investigation* **124**, 3419-3430 (2014).
73. Khan, M.A.F. *et al.* RBM20 Regulates Circular RNA Production From the Titin Gene. *Circulation Research* **119**, 996-1003 (2016).
74. Beqqali, A. *et al.* A mutation in the glutamate-rich region of RNA-binding motif protein 20 causes dilated cardiomyopathy through missplicing of titin and impaired Frank–Starling mechanism. *Cardiovascular Research* **112**, 452-463 (2016).
75. Wyles, S.P. *et al.* Modeling structural and functional deficiencies of RBM20 familial dilated cardiomyopathy using human induced pluripotent stem cells. *Human Molecular Genetics* **25**, 254-265 (2015).
76. Murayama, R. *et al.* Phosphorylation of the RSRSP stretch is critical for splicing regulation by RNA-Binding Motif Protein 20 (RBM20) through nuclear localization. *Scientific Reports* **8**, 8970 (2018).
77. Bertero, A. *et al.* Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. *Nature Communications* **10**, 1538 (2019).
78. Gaertner, A. *et al.* Cardiomyopathy-associated mutations in the RS domain affect nuclear localization of RBM20. *Human Mutation* **41**, 1931-1943 (2020).
79. Vieira-Vieira, C.H., Dauksaite, V., Sporbert, A., Gotthardt, M. & Selbach, M. Proteome-wide quantitative RNA-interactome capture identifies phosphorylation sites with regulatory potential in RBM20. *Molecular Cell* **82**, 2069-2083.e8 (2022).
80. Fenix, A.M. *et al.* Gain-of-function cardiomyopathic mutations in RBM20 rewire splicing regulation and re-distribute ribonucleoprotein granules within processing bodies. *Nature Communications* **12**, 6324 (2021).
81. Hoogenhof, M.M.G.v.d. *et al.* RBM20 Mutations Induce an Arrhythmogenic Dilated Cardiomyopathy Related to Disturbed Calcium Handling. *Circulation* **138**, 1330-1342 (2018).
82. Kusano, S. *et al.* Structural basis for extracellular interactions between calcitonin receptor-like receptor and receptor activity-modifying protein 2 for adrenomedullin-specific binding. *Protein Science* **21**, 199-210 (2012).
83. Mackie, D.I. *et al.* hCALCRL mutation causes autosomal recessive nonimmune hydrops fetalis with lymphatic dysplasia. *Journal of Experimental Medicine* **215**, 2339-2353 (2018).
84. Pawlak, J.B., Wetzel-Strong, S.E., Dunn, M.K. & Caron, K.M. Cardiovascular effects of exogenous adrenomedullin and CGRP in Ramp and Calcrl deficient mice. *Peptides* **88**, 1-7 (2017).

85. Weigel, B. *et al.* MYT1L haploinsufficiency in human neurons and mice causes autism-associated phenotypes that can be reversed by genetic and pharmacologic intervention. *Molecular Psychiatry* (2023).
86. Lee, J. *et al.* A Myt1 family transcription factor defines neuronal fate by repressing non-neuronal genes. *eLife* **8**, e46703 (2019).
87. Melhuish, T.A. *et al.* Myt1 and Myt1l transcription factors limit proliferation in GBM cells by repressing YAP1 expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1861**, 983-995 (2018).
88. Hirano, R. *et al.* Spinocerebellar ataxia with axonal neuropathy: consequence of a Tdp1 recessive neomorphic mutation? *The EMBO Journal* **26**, 4732-4743 (2007).
89. Kanai, M. *et al.* Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genomics* **2**, 100210 (2022).

Acknowledgments

JR acknowledges funding from the European Union-NextGenerationEU, fellowship RYC2021-031413-I from MCIN/AEI/10.13039/501100011033, and from the European Union ‘‘NextGenerationEU/PRTR’’ and from grant PID2021-128972OA-I00, funded by MCIN/AEI/ 10.13039/501100011033. PBM, AT and WJY acknowledge the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330); a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George’s University Hospitals NHS Foundation Trust and St George’s University of London, and Medical Research Council grant MR/N025083/1.

Author Contribution Statement

Ethical Approval

Competing Interests

Figures

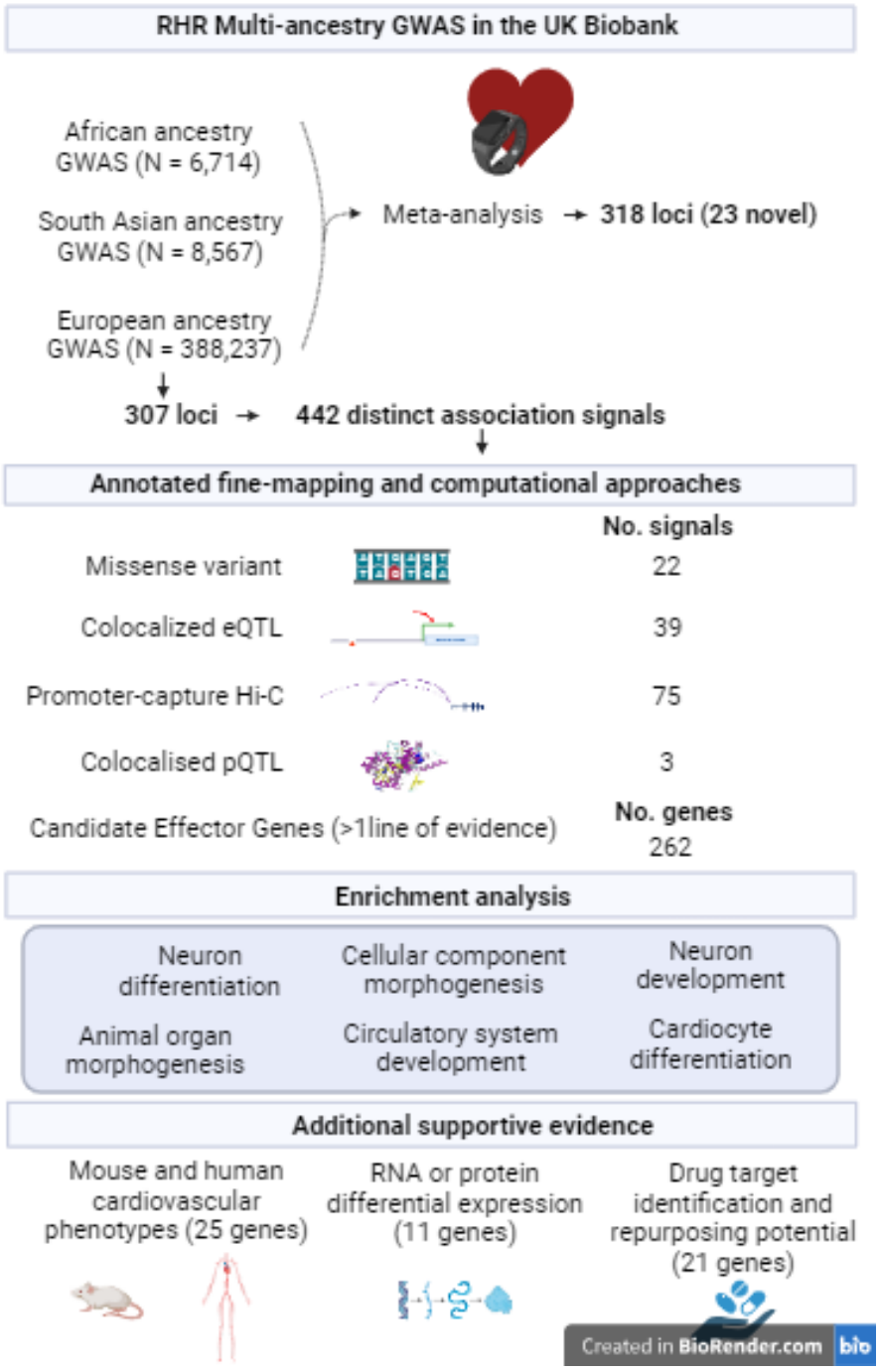

Figure 1. Study Overview and summary of main findings.

eQTL, expression quantitative locus; GWAS, genome-wide association study; Hi-C, long-range chromatin interaction; pQTL, protein quantitative locus; RHR, resting heart rate; RNA, ribosomal nucleic acid.

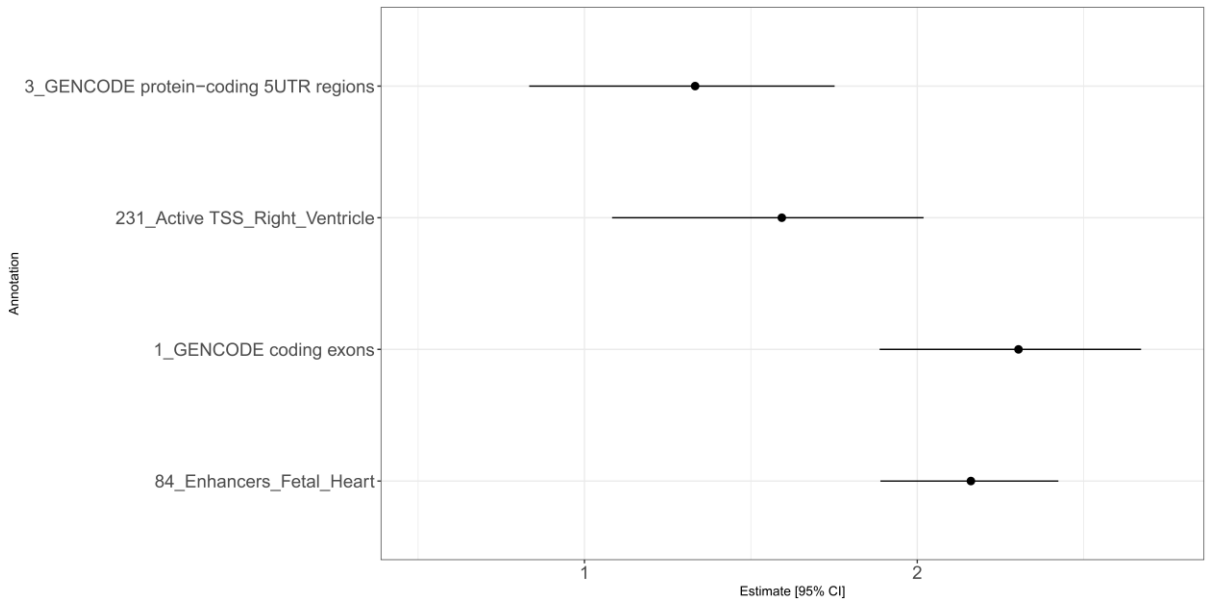

Figure 2. Results from genomic enrichment annotation for RHR. Estimate and 95% confidence interval of the enrichment at the most significant annotations for RHR, calculated using functional GWAS.

UTR, untranslated region; TSS, transcription start site; RHR, resting heart rate.

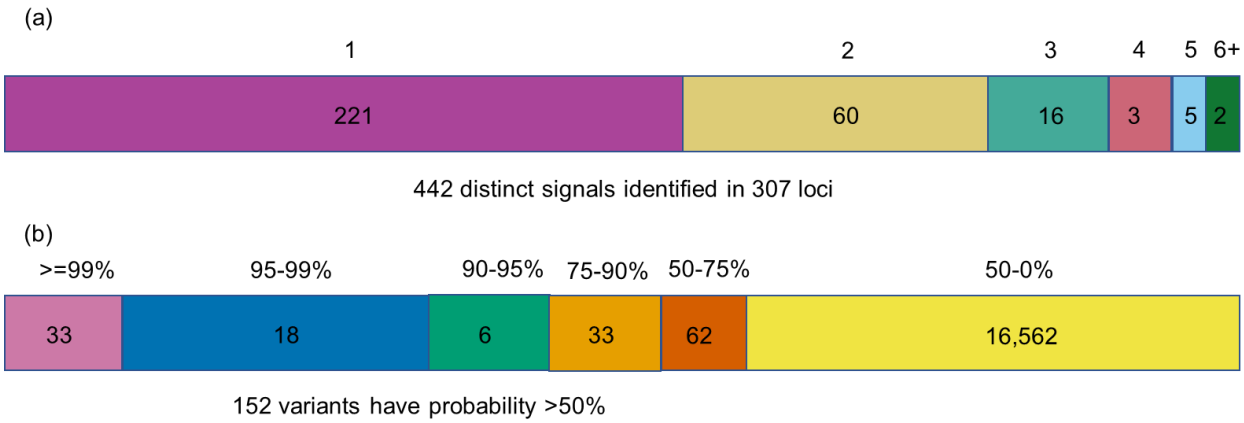

Figure 3. Distinct RHR association signals.

(a) Summary of distinct association signals for RHR. A single signal at 221 genomic regions and at least two at 52. (b) Distribution of the posterior probability of causality of the variants in credible sets. RHR, resting heart rate.

Tables

Table 1. Novel Genes.

Table 2. High-confidence missense variants.

Table 3. Summary of candidate effector genes for RHR.