Genome-wide association analysis reveals insights into the genetic architecture of right ventricular structure and function

Nay Aung[1,2,3], Jose D Vargas[4], Chaojie Yang[5], Kenneth Fung[1,2,3], Mihir M Sanghvi[1,2,3], Stefan K Piechnik[6], Stefan Neubauer[6], Ani Manichaikul[5], Jerome I Rotter[7], Kent D Taylor[7], Joao A C Lima[8], David A Bluemke[9], Steven M Kawut[10], Steffen E Petersen[1,2,3#*], Patricia B Munroe[1,2#*]

[1] William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[2] National Institute for Health Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK

[3] Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, UK

[4] Medstar Heart and Vascular Institute, Medstar Georgetown University Hospital, Washington DC, USA

[5]Center for Public Health Genomics, University of Virginia, Charlottesville, USA

[6] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

[7] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA USA

[8]Division of Cardiology, Johns Hopkins University, Baltimore, MD, USA

[9]Department of Radiology, University of Wisconsin, Madison

[10]Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[#] Joint senior author

[*] Corresponding authors

Address for correspondence

Professors Patricia B Munroe and Steffen Petersen, William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK.

Email: p.b.munroe@qmul.ac.uk; s.e.petersen@qmul.ac.uk

Telephone: +44 (0) 20 7882 7188

Word count: 5,901

Abstract

Right ventricular (RV) structure and function play a key role in mediating the morbidity and mortality from coronary artery disease (CAD), dilated cardiomyopathy (DCM), pulmonary hypertension and heart failure. No previous study has evaluated the genetic basis of RV measurements. We perform genome-wide association analyses of four clinically relevant RV phenotypes (RV end-diastolic volume, RV end-systolic volume, RV stroke volume, RV ejection fraction) from cardiovascular magnetic resonance images, using a state-of-the-art deep learning algorithm in 29,506 UK Biobank participants. We identify 25 unique loci associated with at least one RV phenotype at $P < 2.27 \times 10^{-8}$. In a combined meta-analysis (N = 41,830), 17 out of 25 loci are validated. Of these, 10 loci are not known to be associated with left ventricular phenotypes. Several candidate genes overlap with Mendelian cardiomyopathy genes and are involved in cardiac muscle contraction and cellular adhesion. The RV polygenic risk scores are associated with DCM, CAD and hypothyroidism. The findings represent a significant advance in our understanding of the genetic underpinning of RV measurements.

Introduction

The vital role of cardiac right ventricular (RV) structure and function in congestive heart failure, arrhythmia, pulmonary hypertension and sudden death is increasingly recognised[1–5]. The RV ejection fraction is an independent predictor of morbidity and mortality in the settings of acute myocardial infarction, ischaemic cardiomyopathy and all-cause heart failure[6–11]. Reduction in RV longitudinal function is associated with an increased risk of mortality or cardiac transplantation in patients with myocarditis[12]. In non-ischaemic dilated cardiomyopathy (DCM), RV end-diastolic and end-systolic volumes and RV ejection fraction are predictive of heart failure hospitalisation and death[13,14]. RV hypertrophy has been associated with heart failure or death in a multi-ethnic population free of clinical cardiovascular disease at baseline[15] and RV longitudinal function has been found to predict cardiovascular death in the general population[16], even after adjusting for the corresponding LV parameters.

Morphologically, the right ventricle possesses a complex anatomy which appears triangular when viewed laterally and semi-lunar when viewed in cross-section[17]. Functionally, the main purpose of right ventricle is to propel systemic venous blood into the low-resistance pulmonary circulation. Despite the physiological and clinical significance of RV structure and function, there is a dearth of data on the genetic basis of RV imaging measurements. Previous studies investigating the genetic architecture of ventricular imaging traits focused solely on the left ventricle[18–22]. The major obstacle to accurate phenotyping in large studies is the complexity of RV geometry which defies conventional assessment with two-dimensional echocardiography. As a result, there has been no evidence to-date from a large-scale genome-

wide analysis of RV imaging phenotype. Cardiovascular magnetic resonance (CMR) imaging is considered as the gold standard for comprehensive evaluation of the right heart due to its superior image quality and reproducibility[23] and the lack of geometric assumptions compared to conventional echocardiography. The absence of ionising radiation in CMR compared to multidetector computed tomography also makes it ideal for large-scale population studies. The UK Biobank, one of the largest population imaging studies, has acquired both high-quality standardised CMR examinations and dense genotype data, offering a tremendous opportunity to investigate the as-yet-unknown genetic determinants of RV parameters.

In this study, we aim to investigate the genetic basis of four clinically relevant and prognostically important RV imaging phenotypes (RV end-diastolic volume [RVEDV], RV end-systolic volume [RVESV], RV stroke volume [RVSV] and RV ejection fraction [RVEF]). We report single nucleotide polymorphism (SNP)-based heritability estimates of 19% to 34% and a total of 46 locus-trait associations (25 unique loci) associated with RV structure and function. The RV GWAS loci are enriched in the components of cell adhesion and several putative candidate genes associated with RV phenotypes are linked to inherited cardiomyopathy and intra-cellular calcium handling. The phenome-wide scanning with RV polygenic risk scores shows associations with DCM, ischaemic heart disease and hypothyroidism. Overall, this study substantially enhances our knowledge of the genetic underpinning of RV structure and function and underscores their shared genetic architecture with intrinsic heart muscle disease and arrhythmia development.

Results

Derivation of high-quality RV phenotypes enabled by deep learning

A total of 32,581 CMR studies were available at the time of analysis. The first 5,065 studies were manually segmented by eight human analysts to create a reference dataset for four RV structural and functional phenotypes (RVEDV, RVESV, RVSV and RVEF). This expert-annotated dataset was used to train a deep fully convolutional neural network with a U-net like architecture[24,25], which subsequently performed automatic segmentation of the remaining 27,516 CMR studies (Methods). Both manual and automatic techniques produced highly accurate and reproducible RV segmentation and derived measurements as indicated by their Dice scores and intra-class correlation coefficients (ICC) (manual Dice = 0.87, manual ICC: 0.77-0.92 and automatic Dice = 0.90, automatic ICC: 0.90-0.96)[26,24]. Following the exclusion of poor image quality and sample quality-control procedures, 29,506 European participants free from a diagnosis of myocardial infarction or heart failure remained (Supplementary Fig. 1). The average age of the cohort at the time of imaging visit was 63 years and 47% were men. The mean values of RVEDV, RVESV, RVSV and RVEF were 157 ml, 68 ml, 89 ml and 57 %, respectively (Supplementary Table 1 and Supplementary Fig. 2). An overview of study design is presented in Fig. 1.

Heritability and genotypic correlation

We estimated the proportion of RV phenotypic variation attributable to the genotypes (also known as SNP-based heritability) by the variance component analysis (Methods). The RV structural and functional measurements were moderately heritable with SNP-based heritability of 31% for RVEDV, 34% for RVESV, 19% for RVSV and 21% for RVEF. The magnitude of genotypic correlation (r_g) between the RV phenotypes was moderate to high for all traits except for the correlation between RVEDV and RVEF (r_g = -0.13). The strongest

positive genotypic correlation was observed between RVEDV and RVSV (r_g = 0.83) and the strongest negative genotypic correlation was found between RVESV and RVEF (r_g = -0.67). The overall pattern of RV genotypic correlations closely mirrored the corresponding phenotypic correlations (Supplementary Fig. 3).

Genomic loci associated with RV phenotypes

The conventional single-trait genome-wide association analyses were conducted in ~ 9.9 million variants with minor allele frequency (MAF) ³ 0.01 and INFO score > 0.3. Additionally, to leverage on the increased statistical power afforded by highly correlated RV phenotypes, we performed paired multi-trait analyses using the single-trait GWAS summary statistics. Specifically, RVEDV was paired with RVSV and RVESV was paired with RVEF. The single-trait and multi-trait analyses yielded a total of a total of 46 susceptibility loci – 12 loci for RVEDV, 14 loci for RVESV, 5 loci for RVSV and 15 loci for RVEF – at P < 2.27 x 10^{-8} (conventional GWAS P value 5 x 10^{-8} divided by 2.2, effective number of tests for correlated RV phenotypes) (Table 1 and Fig. 2). Out of 46 loci, 34 loci were discovered by single trait analyses and the remaining 12 loci were obtained from joint multi-trait analyses. The LocusZoom region plots for all RV loci are depicted in Supplementary Fig. 4. Twelve loci (*TTN*, *ATXN2*, *PTPN11*, *ACTN4*, *RBL2*, *LUC7L2*, *AK097794*, *BAG3*, *GOSR2*, *SLC6A6*, *OBSCN*, *FHOD3*) were shared across more than one RV phenotype at our pre-defined GWAS P value (2.27 x 10^{-8}) resulting in 25 unique loci (Supplementary Fig. 5). Furthermore, all loci except for *SVIL* and *CCDC85C* for RVEF were associated with at least one other RV phenotype at a suggestive significant P < 1 x 10^{-5} (Supplementary Table 2), reflecting the strength of underlying phenotypic and genetic correlations. There was no evidence of confounding from population stratification or cryptic relatedness in our GWASs as

demonstrated by low genomic inflation factor (λ = 1.071 – 1.102), small LD score regression intercept (1.00 – 1.01) and the quantile-quantile plots (Supplementary Fig. 6). The lead variants explained a small proportion of trait variance (R^2 for RVEDV: 0.41%, RVESV: 0.96%, RVSV: 0.30%, RVEF: 1.48%). The clumping procedure at linkage disequilibrium (LD) $r^2 < 0.1$ in each genomic locus produced independent variants for several loci as indicated in the footnote of Table 1.

Meta-analysis of RV loci

We sought validation of the RV loci in the new UK Biobank sample which became available at the end of our discovery analysis (Nmax = 11,073) and Multi-Ethnic Study of Atherosclerosis (MESA) (N European ancestry = 1,251). We performed a meta-analysis combining the association summary statistics from the discovery analysis, the new UK Biobank sample and MESA, in a total of up to 41,830 individuals. In this analysis, 17 out of 25 unique loci achieved validation by attaining genome-wide significance at $P_{meta\text{-}analysis} < 5x10^{-8}$ (Table 1 and Supplementary Table 3). Four out of five loci highly specific to RV phenotypic variations (*OBSCN*, *PALLD*, *AK311445*, *SVIL*) were validated by the meta-analysis. The lead variant of *CCDC85C* locus (rs79884713) which failed to achieve validation ($P_{meta\text{-}analysis}$ = 1.8x10-4) tags few LD proxies (Supplementary Fig. 4.4) and the genes in this locus were not convincingly supported by our subsequent bioinformatic analyses.

Shared genetic architecture with known LV loci

No prior study had investigated the genetic architecture of RV imaging phenotypes; hence, all genetic loci reported in this study are new observations. However, due to interdependent nature of LV and RV chambers and the importance of RV remodelling in the context of left heart disease, we first sought to quantify the strength of their genetic correlations by LD score regression using our RV summary statistics and the summary data from a recently published LV GWAS[22]. The LV and RV imaging phenotypes were highly correlated except for the relationships between functional traits (RV and LV stroke volume and ejection fraction) where negligible genetic correlation was observed (Supplementary Fig. 7). We then looked up our RV lead variants in the published GWAS data of CMR-derived LV phenotypes to identify common genetic loci. The lead variants at 7 RV loci (*TTN*, *SLC6A6*, *PTPN11*, *BAG3*, *ATXN2*, *SLC35F1 and CLCNKA)* were also associated with CMR LV phenotypes at P $< 5 \times 10^{-8}$ (Supplementary Table 4). Additionally, 4 RV loci (*AK097794*, *GOSR2*, *TPM2* and *FHOD3*) were associated with at least one LV imaging phenotype at a suggestive $P < 1 \times 10^{-5}$. The remaining 6 loci (*AK311445*, *HSPA4*, *OBSCN*, *PALLD*, *PLEC*, *SVIL*) did not have strong evidence of association with LV phenotypes.

Pleiotropic associations with other complex traits

We interrogated the Phenoscanner database[27] to explore pleiotropic associations between our lead variants (and their close proxies at LD $r^2 \geqslant 0.8$) and other complex traits. Variants in 8 loci (*TTN, BAG3*, *ATXN2*, *PTPN11*, *GOSR2*, *SLC35F1*, *PLEC* and *HSPA4*) were associated with several traits including cardiovascular risk factors and disease phenotypes (Fig. 3). To highlight a few relevant associations, our GWAS variant in the *BAG3* locus (rs2234962) has been implicated in DCM and the variants in *ATXN2* and *PTPN11* loci have multiple pleiotropic relationships with CAD and risk factors such as hypertension, diabetes mellitus

and lipids (Supplementary Table 5). We observed variants at 2 loci to be associated with CV traits only (*TTN* and *GOSR2*). Four RV loci (*OBSCN*, *PALLD*, *AK311445*, *SVIL*) showed very limited evidence of association with other traits including LV phenotypes, thus, appeared more specific to the RV phenotypes (Supplementary Fig. 8).

Functional annotation of variants

The RV GWAS loci harboured a total of 2756 candidate variants in 99% credible sets (2.1% exonic variants, 47% intronic variants, 31% intergenic variants, and the remainder are non-coding RNA, untranslated, upstream and downstream variants). Four exonic variants were predicted to be damaging by two or more in-silico prediction tools. These variants are rs16866380 in the *TTN* gene, rs10497529 in the *CCDC141* gene, rs2234962 in the *BAG3* gene and rs34674752 in the *SHARPIN* gene. Among the non-coding variants, 25 variants were considered functionally important according to RegulomeDB[28] or Combined Annotation Dependent Depletion (CADD) score[29]. The eCaviar[30] colocalisation analysis, which uses the gene expression data in cardiovascular tissues, detected at least one causal variant in 16 out of 17 validated RV loci. The putative causal variant differed from the lead variant in the majority (72%) of cases (Table 1). The functional characteristics of GWAS variants are outlined comprehensively in Supplementary Table 6. We next investigated the enrichment of our GWAS loci in a wide spectrum of known biological annotations using the DEPICT[31] (Data-driven Expression Prioritized Integration for Complex Traits) framework. In this analysis, the GWAS summary statistics for each RV phenotype at $P < 1x10^{-5}$ was inputted into the software. Enrichment in the components of cell adhesion (in particular cell-substrate junctions that anchors cells to the extracellular matrix) was observed at false discovery rate < 0.05 for RVEDV and RVSV (Supplementary Table 7).

Candidate gene identification

We sought to identify candidate genes influencing RV phenotypic variation using an integrative approach supported by multiple lines of evidence. Based on the downstream analysis of the discovery GWAS summary statistics, eight genes were prioritised by the presence of damaging non-synonymous variants, 44 and 37 genes were identified by transcriptome-wide analysis of predicted gene expression and splicing, respectively, in S-MultiXcan[32] (Supplementary Tables 8-9), 32 genes were associated with alteration of myocardial phenotypes in mouse knockout models, 28 were discovered by the long-range chromatin interaction analyses (Supplementary Table 10) and 129 were identified by position (± 100kb of the lead variant). The gene-based analysis in MAGMA[33] (Multi-marker Analysis of GenoMic Annotation) which also incorporates rare variants by burden scoring identified 59 additional genes (Supplementary Table 11) including *CAV3* and *MYH6* (both associated with RVSV), two well-known causative genes for familial hypertrophic cardiomyopathy (HCM). In total, 221 candidate genes were mapped as detailed in Supplementary Table 12. Some notable candidate genes include *FHOD3*, *MYL4* and *TMEM43* known to be associated with inherited cardiovascular disease.

The functional profiling of prioritised genes in the 17 validated RV loci using g:Profiler revealed enrichment in the biological pathways associated with cardiac morphogenesis, structural constituents of heart muscle, cardiac muscle contraction and adrenergic signalling in cardiomyocytes (Fig. 4 and Supplementary Table 13).

Phenome-wide association study (PheWAS)

We investigated the relationships between weighted polygenic risk scores (PRSs) constructed from the lead and secondary variants of the validated loci for each RV phenotype and 1041 disease phenotypes derived predominantly from the hospital episodes data. The PheWASs showed significant associations with ischaemic heart disease, myocardial infarction, primary/intrinsic cardiomyopathy and heart failure (Supplementary Table 14). A strong signal of association was noted between RVEF PRS and non-ischaemic DCM (odds ratio 0.70, 95% confidence interval: 0.64 – 0.77, adjusted $P = 5.64 \times 10^{-12}$, per 1 standard deviation increase in PRS) (Fig. 5). Interestingly, beyond the circulatory system, a negative association was identified between the genetic risk scores of RVEDV, RVESV and RVSV and hypothyroidism, which may reflect the perturbation in preload and cardiac output often seen in this condition[34].

Discussion

This is the first study to examine the genetic determinants of clinically relevant RV structural and functional phenotypes, robustly derived from high-quality CMR examinations. Both single-trait genome-wide association and multi-trait joint analyses were conducted using ~9.9 million common genetic variants obtained from ~29,000 individuals free from pre-existing heart failure or myocardial infraction. This approach yielded a total of 25 unique loci (46 locus-trait associations) for RV structure and function. A follow-up European ancestry meta-analysis (Nmax = 41,830) validated 17 out 25 loci, 10 were not previously known to be associated with LV phenotypes. The discovered GWAS loci are enriched in the processes of cell-matrix adhesion. Extensive multi-layered bioinformatic annotations identified candidate genes involved in inherited cardiomyopathy and muscle contraction. The phenome-wide

association scanning with RV polygenic scores demonstrated the relationships between genetically determined RV structure and function and ischaemic heart disease, non-ischaemic dilated cardiomyopathy and heart failure, further underscoring the prognostic importance of RV imaging phenotypes.

Prior to this study, little was known about the heritability and genetic basis of RV measurements primarily due to difficulties in imaging the complex geometry. Delineation of RV cavity to extract quantitative phenotypes such as RV volume is equally challenging and requires experienced analysts to perform a time-consuming annotation process called segmentation. In this study, acquisition of highly standardised CMR images coupled with the state-of-the-art automatic segmentation technique revolutionised by deep learning permitted characterisation of the genetic architecture of clinically relevant RV measurements and provided novel biological insights. Firstly, our investigations found that a significant proportion of RV phenotypic variability is explained by the underlying genetics as indicated by heritability estimates ranging from 19% to 34%. The RV phenotypic and genotypic correlations are strong and almost identical in magnitude. There is a very high level of genetic overlap between four RV phenotypes, reflecting their intimate physiological relationship.

A large proportion of the RV loci were associated with previously reported loci for LV imaging phenotypes, an expected finding due to the interdependent nature LV and RV chambers and their strong genetic correlations as indicated by the LD score regression analysis. This observation reinforces the rationale for investigating the genetic basis of RV phenotypes as a complementary gateway to understanding the drivers of LV remodelling which is more strongly linked with adverse outcomes. Furthermore, some RV GWAS loci

showed evidence of pleiotropic associations polygenic traits such as hypertension (e.g., *GOSR2* locus for RVEDV, RVESV and RVEF), highlighting the shared genetic pathway influencing cardiovascular physiology and morphology. *GOSR2* a candidate gene at this locus encodes a vesicle docking protein, and recent functional studies validate its importance in cardiovascular development. Knockout studies in fish indicate numerous abnormalities including abnormal heart morphology and reverse looped heart[35]. Lahm and colleagues have reported a variant in high LD (r^2>0.8) with our lead variant associated with congenital heart disease, and differences in gene expression between human fetal and adult tissues, indicating a possible role for *GOSR2* in cardiac development[36].

At loci demonstrating some RV specificity in our analysis sample there are candidate genes with an important role in myocyte integrity and cell anchoring. *SVIL* (supervillin) binds the actin cytoskeleton and the plasma membrane and has been shown to regulate cell adhesion, cytokinesis and cell motility[37]. Loss of function *SVIL* mutations have been associated with the development of a type of skeletal myopathy with cardiac manifestations[37]. The *OBSCN* (Obscurin) gene in the *OBSCN* locus for RVESV and RVEF encodes a giant myofibrillar protein which mediates cellular adhesion and support sarcolemmal integrity[38]. Mutations in this gene had been shown to result in perturbation of calcium cycling and spontaneous arrhythmia[39]. *PALLD* (palladin) encodes a cytoskeletal protein that functions as a scaffolding molecule and is important for actin polymerization and assembly and has a role in cell morphology, motility and adhesion[40]. On the whole, our results suggest the involvement of protein complexes regulating the cell-matrix junction in RV remodelling in individuals without overt cardiac disease. This finding is somewhat analogous to the mutations in desmosomes (intercellular adhesives) associated with arrhythmogenic right ventricular cardiomyopathy (ARVC)[41], an extreme Mendelian form of RV disease.

There is a preponderance of Mendelian genes implicated in the pathogenesis of inherited cardiomyopathy at the RV loci. Beyond well-recognised cardiomyopathy-associated genes such as *TTN* and *BAG3*, we identified several other candidate genes associated with RV phenotypes (*FHOD3*, *MYH6*, *MYL4*, *TMEM43*) linked to heart muscle disease and arrhythmia. The *FHOD3* (Formin Homology 2 Domain Containing 3) gene in the *FHOD3* locus for RVESV and RVEF is involved in actin filament polymerisation during myofibrillogenesis[42] and is required for post-natal development and maintenance of the adult mouse heart[43]. A missense mutation (rs2303510) with G to A allele polymorphism was found to be protective of sporadic DCM in an exome-wide association study[44]. In our GWAS, the same variant was associated with smaller RVESV suggesting a protective effect. *MYH6* identified by the MAGMA analysis for RVSV and *MYL4* indicated by the sQTL and MAGMA analyses for RVESV are genes that encode proteins in the subunits of cardiac muscle myosin. Mutations in *MYH6* cause familial HCM and DCM[45]. *MYL4* encodes atrial-selective essential myosin light chain and has been implicated in familial atrial fibrillation[46]. Lastly, the *TMEM43* gene in the *SLC6A6* locus for RVESV and RVEF is a known causative gene for a fully penetrant, high risk subtype of arrhythmogenic right ventricular cardiomyopathy (ARVC)[47]. These results support the role of RV imaging measurements as intermediate endophenotypes which share a genetic relationship with inherited cardiac conditions. Interestingly, other known ARVC genes such as *PKP2*, *DSG2*, *JUP*, *DSP* and *DSC2* were not prioritised by our analyses. The culprit pathogenic variants in these genes are often ultra-rare in a general population. As such, our GWAS of common variants (MAF $\geq$ 0.01) is not designed to identify such variants. Future exome-wide association studies a large sample may shine light on how these genes modulate RV phenotypic variation.

The association between RV structural and functional adaptation and worse heart failure outcomes in both ischaemic and non-ischaemic cardiomyopathy has been reported previously[6,7,13]. Although there is a paucity of data pertaining to RV parameters among the general population, large prospective studies of individuals free from clinical cardiovascular disease have reported associations between RVEF and RV mass and increased incidence of heart failure as well as between RV function and increased risk of cardiovascular death[15,16]. Our phenome-wide association analyses using RV PRSs shed light on the linkage between genetically determined RV phenotypic variation and the incidence of all-cause heart failure and non-ischaemic DCM. Specifically, higher genetically determined RVEF was associated with lower rates of hospitalisation for congestive heart failure and DCM, which further reinforced the importance of right ventricle in maintaining efficient circulatory physiology. This observation is overall concordant with the study by Bai et al.[48] which investigated the associations between cardiac imaging measurements and a wide range of outcomes including cardiac diseases.

We acknowledge some limitations in our study. Our study is of European descent and validation of our findings in more diverse cohorts is needed to assess their applicability to non-European populations. Furthermore, although we have demonstrated statistical support for the loci which are highly specific for RV phenotypic variation, future experimental studies using gene-editing techniques such as CRISPR in cellular and animal models are required to fully validate the functional significance of highlighted candidate genes and mechanisms modulating RV structure and function.

Conclusion

In this first genome-wide association study to-date of CMR-derived RV phenotypes, we report 17 genetic loci validated in a combined meta-analysis; highlight the role of cellular adhesion in determining RV phenotypic variation and indicate candidate genes linked to inherited cardiomyopathy. Altogether, the findings represent a significant advance in our understanding of the genetic architecture of RV phenotypes in a general population and provides a foundation to characterise the genetic drivers of RV remodelling in the future.

Methods

Study cohort

The discovery analysis was performed in the UK Biobank study. The UK Biobank is a large population-based prospective cohort study which has collected a wealth of information on health and lifestyle data, physical measurements, biological samples, and cardiac imaging phenotypes derived from CMR. This ambitious project aims to provide resources to disentangle the genetic and environmental determinants of complex diseases affecting middle and old age. The study protocol has been described in detail previously[49].

Derivation of RV phenotypes

Four RV parameters (RVEDV, RVESV, RVSV and RVEF) were measured from the short-axis SSFP cine images of UK Biobank CMR studies acquired using 1.5 Tesla scanners. The full CMR study protocol has been published previously[50]. In brief, the short-axis cine images were acquired using balanced steady-state free precession (bSSFP) sequence with typical parameters of TR/TE = 2.6/1.1 ms, flip angle 80°, Grappa factor 2, slice thickness 8mm, slice

gap 2mm, voxel size 1.8 mm × 1.8 mm × 8 mm. The actual temporal resolution of 32 ms was interpolated to 50 phases per cardiac cycle (~20 ms).

A combination of manual segmentation and automatic annotation with a deep learning algorithm was used to extract the measurements as previously described[24]. In brief, the short axis cine images of the first 5,065 UK Biobank CMR studies were manually segmented by eight analysts in two core laboratories using a pre-defined standard operating procedure[26]. After removing poor quality images, the remaining 4,875 manually annotated studies were used to train and validate a deep convolutional neural network adapted from the U-net architecture[51] with the VGG-16[52] backbone. For training, the images were cropped to the size of 192×192 and intensity normalised to the range of [0,1]. On-the-fly data augmentation was performed by applying random translation, rotation, scaling and intensity variation to each mini-batch (20 image slices) of images. We used the Adam optimiser[53] with a learning rate of 0.001 and iteration number of 50,000. The network was trained using TensorFlow[54] library in Python on a Nvidia Tesla K80 GPU. The Dice metric of automated contours vs manual contours for RV cavity was calculated in 600 CMR studies (hold-out test sample) and the Dice metric of manual contours for two human analysts was calculated in 50 CMR studies. The inter-observer variability of image-derived RV phenotypes was estimated by intra-class correlation coefficient. Exemplary CMR images annotated by manual and automatic methods are presented in Supplementary Fig. 9.

Quality control

Manually annotated CMR studies (N = 5,065) were quality checked visually by the analyst at the time of manual segmentation. Out of these 5,065 studies, 153 studies were excluded due

435 to poor image quality. The following are the reasons for exclusion: (i) incomplete coverage of
436 the ventricles (N = 78 [51%]), (ii) blurred image due to ECG mis-triggering commonly
437 caused by irregular R-R intervals (arrhythmia) (N = 26 [17%]), (iii) miscellaneous causes
438 including motion artefacts due to poor breath-hold and missing short-axis images (N = 43
439 [28%]).
440
441 As for the automatically analysed CMR studies, the segmentation and image quality was
442 reviewed in a bespoke image visualisation application written in R "Shiny" library[55]. We
443 used a "statistical and prior knowledge" approach to select studies that required visual quality
444 check. In this process, CMR studies with (i) outlying values (defined as values more than
445 three interquartile range [IQR] above the first quartile or below the third quartile) before and
446 after indexing for body surface area and height$^{2.7}$, (ii) non-physiological measurements
447 (RVEDV < 75ml and RVESV < 25ml) and, (iii) measurements within the abnormal zones as
448 defined in the published reference ranges paper[26] were visually reviewed using the custom
449 "Shiny" visualisation tool for segmentation errors, image artefacts and incomplete coverage
450 of ventricles. As a sanity check, 5,000 automatically segmented images were reviewed
451 visually and only five additional cases (above and beyond what we have identified in the
452 review process of pre-selected cases) were identified as poor quality due to incomplete
453 coverage rather than segmentation error. Therefore, the statistical and prior knowledge
454 approach of only reviewing the segmented images with outlying and non-physiological
455 values and measurements outside the normal bounds of reference ranges was accepted as a
456 good compromise which is adequate to ensure the quality of output data.
457
458 Sample selection
459

460 Out of 32,581 participants with available CMR studies, 29,506 European individuals free
461 from pre-existing heart failure, myocardial infarction were included in the analysis following
462 the quality control procedures outlined in Supplementary Fig. 1.
463
464 Single-trait GWAS
465
466 The SNP-based heritability was estimated by the directly genotyped variants using the
467 variance component method implemented in BOLT-REML software[56]. Single-trait GWASs
468 were performed under an additive linear mixed-effects model using ~9.9 million well-
469 imputed variants with MAF ³ 1% and INFO > 0.3 in BOLT-LMM software. The analysis
470 models were adjusted for age, sex, height, weight, SBP corrected for anti-hypertensive
471 medication use (by adding 15mmHg)[57], phenotype-derivation method (automatic/manual),
472 array type (UK Biobank vs UK BiLEVE array), and imaging centre. Due to the non-normal
473 distribution, we performed the rank-based inverse normal transformation of the residuals of
474 RV phenotypes.
475
476 Multi-trait analysis
477
478 Paired multi-trait analyses of GWAS summary statistics were performed using MTAG[58].
479 MTAG boosts statistical power by capitalising on the correlated effect estimates across
480 different traits. We paired the GWAS summary data of RVEDV with RVSV and RVESV
481 with RVEF as input because of their strong genotypic correlations (r_g = 0.83 and -0.67,
482 respectively). The MAF filter of 0.01 and the INFO filter of 0.3 were applied.
483

The cut-off for genome-wide significance was set at $P < 2.27 \times 10^{-8}$ (conventional GWAS P value threshold of 5×10^{-8} divided by 2.2, effective number of tests for multiple correlated phenotypes). LD score regression (LDSC) was performed using the ldsc python package v1.0.1 using the default settings to obtain the LDSC intercept[59] which represents the extent of confounding form population stratification and genetic correlations[60]. A genomic risk locus was defined as 500kb upstream and downstream of the most significant variant. In each genomic locus, we performed clumping using plink '--clump' function with r^2 threshold of 0.1. Therefore, each locus may contain multiple lead SNPs and their proxies at r^2 cut-off of 0.1. The proportion of variance explained by the genome-wide significant variants for each RV phenotype was calculated by the difference in the adjusted R^2 between the linear regression model containing all covariates plus all lead variants for the trait and the model containing only the analysis covariates.

Cross-referencing RV loci in other traits

We looked up our RV lead variants in the summary data available from a recently published genome-wide association studies[22] of CMR derived LV phenotypes with comparable sample size. Additionally, we cross-referenced our RV lead variants and their close proxies (LD $r^2 \geq 0.8$) in the 99% credible sets with the genome-wide association results of other traits in the Phenoscanner[61] database v2 (http://www.phenoscanner.medschl.cam.ac.uk/).

European ancestry meta-analysis

We obtained the lookup results of the RV lead variants in the UK Biobank European sample which became available at the end of our discovery analysis (N = 11,073) and Multi-Ethnic

Study of Atherosclerosis (MESA) (N = 1,251). In the new UK Biobank cohort, we removed individuals with relatives (third-degree or closer; Kinship coefficient ≥ 0.044) in our discovery sample to get unbiased estimates of association.

MESA is a longitudinal study of subclinical cardiovascular disease and risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. Between 2000 and 2002, MESA recruited 6,814 men and women 45 to 84 years of age from Forsyth County, North Carolina; New York City; Baltimore; St. Paul, Minnesota; Chicago; and Los Angeles. Exclusion criteria were clinical cardiovascular disease, weight exceeding 136 kg (300 lb.), pregnancy, and impediment to long-term participation. RV imaging and measurements in MESA have been previously described[62].

MESA participants who consented to genetic analyses were genotyped in 2009 using the Affymetrix Human SNP array 6.0. Following genotype quality control for these data including filter on SNP level call rate < 95%, individual level call rate < 95%, heterozygosity > 53%, 897,981 SNPs remained. The University of Michigan imputation server was used for pre-phasing and imputation using the 1,000 Genomes Phase 3 integrated variant set. Among the MESA participants with RV phenotypes available, we stratified by race/ethnic group and constructed a subset of unrelated individuals by retaining at most one individual from each family. We further excluded those individuals with top principal components (PCs) of ancestry > 3.5 SD from the mean within any race/ethnic group. The genetic association analysis in the European subset of MESA (N = 1,251) was conducted on the rank-based inverse normal transformed residuals model adjusted for age, sex, study site, top 3 principal components of ancestry, height, weight and medication-adjusted systolic blood pressure in ProbABELv0.5.0[63]. Finally, a fixed-effect meta-analysis of the association

summary statistics from UK Biobank discovery sample, UK Biobank additional sample and MESA European cohort was performed in GWAMA v2.2.2[64].

Bioinformatic analyses

Variant-level annotation

The lead variants and their proxies at $r^2 < 0.01$ were first filtered to 99% credible sets using the Bayesian approach previously described by Wakefield[65]. All variants in the 99% credible sets were interrogated with ANNOVAR[66] database to describe their locations and predicted function using several risk prediction tools including SIFT and PolyPhen-2. The non-synonymous variants were classified as 'damaging' if two or more methods predicted detrimental effects and 'probably damaging' if indicated by a single prediction tool. The non-coding variants were annotated with CADD (v1.6)[29] and RegulomeDB (v2.0)[28]. Variants with scaled CADD score > 20 or RegulomeDB score ≤ 1f were considered functionally important. Finemapping of causal variants within credible sets was performed by the colocalisation analysis of GWAS and cis-eQTL signals from cardiovascular tissues (aorta, coronary artery, tibial artery, left atrial appendage and left ventricle) in GTEx (v7)[67] using eCaviar[30]. Variants with colocalisation posterior probability (CLPP) value higher than 0.01 were identified as candidate causal variants. We performed the pathway enrichment analysis of GWAS signals using DEPICT[31] (Data-driven Expression-Prioritised Integration for Complex Traits) using the GWAS summary statistics of each RV trait as the input files with an association P value threshold of 1×10^{-5} as recommended by the authors of DEPICT.

Gene-level annotation

Transcriptome-wide association study

We investigated the influence of our GWAS variants on gene expression and splicing using the multi-tissue transcriptome datasets from Genotype-Tissue Expression project (GTEx v8)[68]. In this analysis, we estimated the genetically regulated gene expression and splicing using the S-MultiXcan tool[32] which integrates the evidence from the GWAS summary data and expression and splicing quantitative trait loci studies across multiple tissues to prioritise candidate genes. We corrected the P values for multiple testing by adjusting for the number of genes x the effective number of tests for correlated RV phenotypes (2.2).

Gene-based association study

Genome-wide gene-based association analysis was performed using Multi-marker Analysis of GenoMic Annotation (MAGMA v1.07b)[33]. In MAGMA, the variants in raw genotype data were first assigned to the genes based on the overlap of their genomic location with a gene window of 35kb upstream and 10kb downstream to include regulatory elements. Rare variants with MAF < 0.01 were inputted into the model as burden scores. The association between variants in gene units and RV phenotypes was tested by principal component regression adjusted for the same covariates as the primary GWAS analysis. The MAGMA values were corrected for multiple testing by adjusting for the number of genes x the effective number of tests for correlated RV phenotypes (2.2).

Hi-C analysis

584 Long-range chromatin interaction analysis was performed using the Hi-C data in FUMA[69]
585 web-based platform v1.3.3c. In the chromatin interaction analysis, we only considered target
586 genes with evidence of significant enhancer-promoter interactions at FDR < $1x10^{-6}$ in the left
587 and right ventricular and aortic tissues. The analysis was filtered to our regulatory GWAS
588 variants with RegulomeDB score £ 5 (where low scores indicate greater evidence of
589 functional significance) locating within these enhancer and promotor regions obtained from
590 the Roadmap Epigenomic project[70].
591
592 Finally, all candidate genes were ranked based on evidence from:
593 i. presence of damaging coding variant in the loci;
594 ii. genes prioritised by S-MultiXcan analyses in expression and splicing quantitative trait
595 loci (eQTL and sQTL) datasets;
596 iii. availability of knockout model from International Mouse Phenotyping Consortium
597 (http://www.mousephenotype.org/) and the Mouse Genome Informatics database
598 (http://www.informatics.jax.org/);
599 iv. targets genes from Hi-C data;
600 v. genes locating within the 100kb window of the lead variant
601 vi. genes prioritised by MAGMA
602
603 *Enrichment and pathway analyses*
604
605 Functional enrichment and pathway characterisation of candidate genes in the validated RV
606 loci were done in the g:profiler tool which leverages on the diverse sources of biological
607 evidence including Gene Ontology (GO), Human Phenotype Ontology (HPO), Reactome,
608 KEGG and Wikipathway[71]. We used the candidate genes supported by at least 2 lines of

evidence (as described above) as g:Profiler input. In g:profiler, multiple-testing correction was performed by the bespoke ontology-focused g:SCS (Set Counts and Sizes) method[72] at 5% threshold.

Phenome-wide association scan

We conducted PheWASs using the polygenic risk scores of each RV phenotype derived from the lead and secondary variants of genome-wide significant loci validated in the meta-analysis. Logistic regression adjusted for age, sex and the first five genetic principal components was performed to test for the associations between each polygenic risk score and a total of 1041 phenotypes with prevalence of ≥ 200 cases in the remaining UK Biobank sample excluding the RV GWAS discovery cohort and their relatives (third-degree or closer with Kinship coefficient ≥ 0.044) (n = 354,307). The PheWAS phenotypes were derived from ICD-10, ICD-9 and OPCS4 codes from hospital episode data, death reports, and self-reported medical history which were last updated on 5th March 2020. Out of 1041 phenotypes, 1029 phenotypes were defined according to the phecode system as previously described[73] and 12 cardiovascular phenotypes (hypertrophic cardiomyopathy, heart failure, myocardial infarction, CAD, non-ischaemic DCM, other inherited cardiomyopathy, stroke, atrial fibrillation/flutter, bradyarrhythmia, ventricular tachycardia, insertion of implantable cardioverter-defibrillator and congenital heart disease) were manually curated using the ICD10, ICD9 and OPCS4 codes detailed in Supplementary Table 15. Multiple testing correction by adjusting for 1041 phenotypes and 2.2 (effective number of tests for correlated RV phenotypes) yielded a significant p value threshold of 2.2×10^{-5} (0.05/2290).

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application 2964. The authors wish to thank all UK Biobank participants and staff. Dr Aung recognizes the National Institute for Health Research (NIHR) Integrated Academic Training programme which supports his Academic Clinical Lectureship post. Professors Piechnik, Neubauer and Petersen acknowledge the British Heart Foundation for funding the manual analysis to create a cardiovascular magnetic resonance imaging reference standard for the UK Biobank imaging resource in 5000 CMR scans (PG/14/89/31194). Prof. Petersen also acknowledges support from the "SmartHeart" Engineering and Physical Sciences Research Council programme grant (EP/P001009/1). Professors Piechnik and Neubauer are supported by the Oxford National Institute for Health Research Biomedical Research Centre and the Oxford British Heart Foundation Centre of Research Excellence. This work was part of the portfolio of translational research of the National Institute for Health Research Biomedical Research Centre at Barts and The London School of Medicine and Dentistry; Dr Aung, Professors Munroe and Petersen acknowledge support from this center. This project was enabled through access to the Medical Research Council eMedLab Medical Bioinformatics infrastructure, supported by the Medical Research Council (grant No. MR/L016311/1). The UK Biobank was established by the Wellcome Trust medical charity, the Medical Research Council, the Department of Health, the Scottish Government, and the Northwest Regional Development Agency. It has also received funding from the Welsh Assembly Government and the British Heart Foundation. MESA and the MESA SHARe projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006,

N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. Also supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. RV phenotyping in MESA was funded by National Institutes of Health K24 HL103844 and R01 HL086719.

Contributions

N.A. conceived the study, designed the methodology, performed manual and automated segmentation of CMR studies, carried out genetic and bioinformatic analyses, and drafted and finalised the manuscript. J.D.V. conceived the study and revised the manuscript. C.Y. and A.M. performed statistical analysis of MESA data. J.I.R., K.D.T., J.AC.L., D.A.B. substantively revised the manuscript. K.F. performed manual segmentation of CMR studies. M.M.S. performed manual segmentation of CMR studies and contributed to the writing of specific sections. S.K.P. and S.N. supervised CMR image segmentation and acquired funding. S.E.P. and P.B.M. conceived the study, designed the methodology, provided supervision, acquired funding and edited the manuscript. All authors read the paper and contributed to its final form.

References

1. Walker, L. A. & Buttrick, P. M. The Right Ventricle: Biologic Insights and Response to Disease: Updated. *Curr. Cardiol. Rev.* **9**, 73–81 (2013).
2. Chatterjee Neal A. *et al.* Right Ventricular Structure and Function Are Associated With Incident Atrial Fibrillation. *Circ. Arrhythm. Electrophysiol.* **10**, e004738 (2017).
3. Naksuk Niyada *et al.* Right Ventricular Dysfunction and Long-Term Risk of Sudden Cardiac Death in Patients With and Without Severe Left Ventricular Dysfunction. *Circ. Arrhythm. Electrophysiol.* **11**, e006091 (2018).
4. Noordegraaf, A. V. & Galiè, N. The role of the right ventricle in pulmonary arterial hypertension. *Eur. Respir. Rev.* **20**, 243–253 (2011).
5. Voelkel Norbert F. *et al.* Right Ventricular Function and Failure. *Circulation* **114**, 1883–1891 (2006).
6. Shah, P. K. *et al.* Variable spectrum and prognostic implications of left and right ventricular ejection fractions in patients with and without clinical heart failure after acute myocardial infarction. *Am. J. Cardiol.* **58**, 387–393 (1986).
7. Polak, J. F., Holman, B. L., Wynne, J. & Colucci, W. S. Right ventricular ejection fraction: An indicator of increased mortality in patients with congestive heart failure associated with coronary artery disease. *J. Am. Coll. Cardiol.* **2**, 217–224 (1983).
8. Groote, P. de *et al.* Right ventricular ejection fraction is an independent predictor of survival in patients with moderate heart failure. *J. Am. Coll. Cardiol.* **32**, 948–954 (1998).
9. Di Salvo, T. G., Mathier, M., Semigran, M. J. & Dec, G. W. Preserved right ventricular ejection fraction predicts exercise capacity and survival in advanced heart failure. *J. Am. Coll. Cardiol.* **25**, 1143–1153 (1995).

10. Gavazzi, A. *et al.* Value of right ventricular ejection fraction in predicting short-term prognosis of patients with severe chronic heart failure. *J. Heart Lung Transplant. Off. Publ. Int. Soc. Heart Transplant.* **16**, 774–785 (1997).
11. Ghio, S. *et al.* Independent and additive prognostic value of right ventricular systolic function and pulmonary artery pressure in patients with chronic heart failure. *J. Am. Coll. Cardiol.* **37**, 183–188 (2001).
12. Mendes, L. A. *et al.* Right ventricular dysfunction: An independent predictor of adverse outcome in patients with myocarditis. *Am. Heart J.* **128**, 301–307 (1994).
13. Juillière, Y. *et al.* Additional predictive value of both left and right ventricular ejection fractions on long-term survival in idiopathic dilated cardiomyopathy. *Eur. Heart J.* **18**, 276–280 (1997).
14. Gulati Ankur *et al.* The Prevalence and Prognostic Significance of Right Ventricular Systolic Dysfunction in Nonischemic Dilated Cardiomyopathy. *Circulation* **128**, 1623–1633 (2013).
15. Kawut, S. M. *et al.* Right ventricular structure is associated with the risk of heart failure and cardiovascular death: the Multi-Ethnic Study of Atherosclerosis (MESA)--right ventricle study. *Circulation* **126**, 1681–1688 (2012).
16. Modin Daniel, Møgelvang Rasmus, Andersen Ditte Madsen, & Biering-Sørensen Tor. Right Ventricular Function Evaluated by Tricuspid Annular Plane Systolic Excursion Predicts Cardiovascular Death in the General Population. *J. Am. Heart Assoc.* **8**, e012197 (2019).
17. Apostolakis, S. & Konstantinides, S. The Right Ventricle in Health and Disease: Insights into Physiology, Pathophysiology and Diagnostic Management. *Cardiology* **121**, 263–273 (2012).

733 18. Vasan, R. S. *et al.* Genetic Variants Associated With Cardiac Structure and Function.
734 *Jama* **302**, 168–178 (2009).
735 19. Wild, P. S. *et al.* Large-scale genome-wide analysis identifies genetic variants associated
736 with cardiac structure and function. *J. Clin. Invest.* **127**, 1798–1812 (2017).
737 20. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links
738 cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
739 21. Aung Nay *et al.* Genome-Wide Analysis of Left Ventricular Image-Derived Phenotypes
740 Identifies Fourteen Loci Associated With Cardiac Morphogenesis and Heart Failure
741 Development. *Circulation* **140**, 1318–1330 (2019).
742 22. Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000
743 individuals yields genetic insights into dilated cardiomyopathy. *Nat. Commun.* **11**, 2254
744 (2020).
745 23. Grothues, F. *et al.* Interstudy reproducibility of right ventricular volumes, function, and
746 mass with cardiovascular magnetic resonance. *Am. Heart J.* **147**, 218–223 (2004).
747 24. Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully
748 convolutional networks. *J. Cardiovasc. Magn. Reson.* **20**, 65 (2018).
749 25. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical
750 Image Segmentation. *ArXiv150504597 Cs* (2015).
751 26. Petersen, S. E. *et al.* Reference ranges for cardiac structure and function using
752 cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank
753 population cohort. *J. Cardiovasc. Magn. Reson.* **19**, 18 (2017).
754 27. Staley, J. R. *et al.* PhenoScanner: a database of human genotype–phenotype associations.
755 *Bioinformatics* **32**, 3207–3209 (2016).
756 28. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using
757 RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).

29. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
30. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
31. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
32. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genet.* **15**, e1007889 (2019).
33. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
34. Udovcic, M., Pena, R. H., Patham, B., Tabatabai, L. & Kansara, A. Hypothyroidism and the Heart. *Methodist DeBakey Cardiovasc. J.* **13**, 55–59 (2017).
35. Meyer, H. V. *et al.* Genetic and functional insights into the fractal structure of the heart. *Nature* **584**, 589–594 (2020).
36. Lahm, H. *et al.* Congenital heart disease risk loci identified by genome-wide association study in European patients. *J. Clin. Invest.* **131**, (2021).
37. Hedberg-Oldfors, C. *et al.* Loss of supervillin causes myopathy with myofibrillar disorganization and autophagic vacuoles. *Brain* **143**, 2406–2420 (2020).
38. Kontrogianni-Konstantopoulos, A., Ackermann, M. A., Bowman, A. L., Yap, S. V. & Bloch, R. J. Muscle giants: molecular scaffolds in sarcomerogenesis. *Physiol. Rev.* **89**, 1217–1267 (2009).
39. Hu, L.-Y. R. *et al.* Deregulated Ca2+ cycling underlies the development of arrhythmia and heart disease due to mutant obscurin. *Sci. Adv.* **3**, (2017).

782 40. Parast, M. M. & Otey, C. A. Characterization of Palladin, a Novel Protein Localized to
783 Stress Fibers and Cell Adhesions. *J. Cell Biol.* **150**, 643–656 (2000).
784 41. Garrod, D. & Chidgey, M. Desmosome structure, composition and function. *Biochim.*
785 *Biophys. Acta BBA - Biomembr.* **1778**, 572–587 (2008).
786 42. Marston, S. Obscurin variants and inherited cardiomyopathies. *Biophys. Rev.* **9**, 239–243
787 (2017).
788 43. Ushijima, T. *et al.* The actin-organizing formin protein Fhod3 is required for postnatal
789 development and functional maintenance of the adult heart in mice. *J. Biol. Chem.* **293**,
790 148–162 (2018).
791 44. Esslinger, U. *et al.* Exome-wide association study reveals novel susceptibility genes to
792 sporadic dilated cardiomyopathy. *PLOS ONE* **12**, e0172995 (2017).
793 45. Carniel, E. *et al.* Alpha-myosin heavy chain: a sarcomeric gene associated with dilated
794 and hypertrophic phenotypes of cardiomyopathy. *Circulation* **112**, 54–59 (2005).
795 46. Peng, W. *et al.* Dysfunction of Myosin Light-Chain 4 (MYL4) Leads to Heritable Atrial
796 Cardiomyopathy With Electrical, Contractile, and Structural Components: Evidence
797 From Genetically-Engineered Rats. *J. Am. Heart Assoc. Cardiovasc. Cerebrovasc. Dis.*
798 **6**, (2017).
799 47. Merner, N. D. *et al.* Arrhythmogenic Right Ventricular Cardiomyopathy Type 5 Is a
800 Fully Penetrant, Lethal Arrhythmic Disorder Caused by a Missense Mutation in the
801 TMEM43 Gene. *Am. J. Hum. Genet.* **82**, 809–821 (2008).
802 48. Bai, W. *et al.* A population-based phenome-wide association study of cardiac and aortic
803 structure and function. *Nat. Med.* 1–9 (2020) doi:10.1038/s41591-020-1009-y.
804 49. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a
805 wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

50. Petersen, S. E. *et al.* UK Biobank's cardiovascular magnetic resonance protocol. *J. Cardiovasc. Magn. Reson.* **18**, 8 (2015).
51. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv150504597 Cs* (2015).
52. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs* (2015).
53. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).
54. Abadi, M. *et al.* TensorFlow: A system for large-scale machine learning. in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* 265–283 (2016).
55. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. *shiny: Web Application Framework for R*. (2018).
56. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
57. Tobin, M. D., Sheehan, N. A., Scurrah, K. J. & Burton, P. R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935 (2005).
58. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
59. Bulik-Sullivan, B. K. *et al.* LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. *Nat. Genet.* **47**, 291–295 (2015).
60. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

61. Staley, J. R. *et al.* PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
62. Chahal, H. *et al.* Relation of cardiovascular risk factors to right ventricular structure and function as determined by magnetic resonance imaging (results from the multi-ethnic study of atherosclerosis). *Am. J. Cardiol.* **106**, 110–116 (2010).
63. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
64. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
65. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
66. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
67. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
68. Aguet, F. *et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues*. http://biorxiv.org/lookup/doi/10.1101/787903 (2019) doi:10.1101/787903.
69. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
70. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
71. Reimand, J. *et al.* g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).

72. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).

73. Wu, P. *et al.* Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med. Inform.* **7**, e14325 (2019).

860 Table 1. Genomic loci identified for CMR-derived RV phenotypes

CMR phenotypes	Locus name	Lead variant	Co-localised variant[†]	CHR	Position (hg19)	EA	NEA	EAF	BETA Discovery	SE Discovery	P Discovery	BETA Meta-analysis	SE Meta-analysis	P Meta-analysis
RVEDV	***OBSCN*[††]**	**rs12126782[*]**	**rs78529941**	**1**	**228613648**	**T**	**G**	**0.64**	**-0.051**	**0.008**	**1.37E-09**	**-0.047**	**0.007**	**2.19E-11**
RVEDV	***TTN*[a]**	**rs2042995**	**rs967507, rs16866380**	**2**	**179558366**	**T**	**C**	**0.78**	**0.070**	**0.010**	**1.50E-13**	**0.053**	**0.008**	**7.27E-11**
RVEDV	*SLC6A6*	rs754020	rs754020	3	14306081	T	C	0.45	-0.050	0.010	1.80E-09	-0.034	0.008	9.67E-06
RVEDV	***AK097794***	**rs2276773**	**rs6792449**	**3**	**158287455**	**A**	**G**	**0.50**	**-0.049**	**0.008**	**8.61E-10**	**-0.041**	**0.007**	**5.28E-10**
RVEDV	*BC038750*	rs752650	rs2714972	4	120632387	T	C	0.65	-0.050	0.010	1.40E-08	-0.032	0.008	4.62E-05
RVEDV	*LUC7L2*	rs144567740	–	7	139099813	T	G	0.80	-0.060	0.010	1.20E-08	-0.038	0.008	5.22E-06
RVEDV	*BAG3*	rs11199043	rs4751742	10	121367249	A	G	0.52	-0.047	0.008	5.11E-09	-0.033	0.007	5.97E-07
RVEDV	***ATXN2*[b]**	**rs35350651**	**rs7137828**	**12**	**111907431**	**A**	**AC**	**0.49**	**-0.060**	**0.010**	**9.80E-15**	**-0.055**	**0.008**	**8.42E-13**
RVEDV	***PTPN11***	**rs11066320**	**rs11066320**	**12**	**112906415**	**A**	**G**	**0.42**	**-0.050**	**0.010**	**2.90E-11**	**-0.045**	**0.008**	**7.89E-09**
RVEDV	*RBL2*	rs375730363	–	16	53440590	CT T	C	0.58	-0.050	0.010	8.60E-09	-0.040	0.008	3.07E-07
RVEDV	***GOSR2***	**rs76774446**	**rs78033733**	**17**	**45046368**	**C**	**A**	**0.86**	**-0.070**	**0.010**	**1.50E-08**	**-0.072**	**0.009**	**1.77E-16**
RVEDV	*ACTN4*	rs11083473	–	19	39179934	A	G	0.45	-0.046	0.008	6.50E-09	-0.033	0.007	8.21E-07
RVESV	***OBSCN*[††]**	**rs12126782[*]**	**rs78529941**	**1**	**228613648**	**T**	**G**	**0.64**	**-0.060**	**0.010**	**4.40E-13**	**-0.056**	**0.008**	**2.18E-12**
RVESV	***TTN*[c]**	**rs2042995**	**rs2366920**	**2**	**179558366**	**T**	**C**	**0.78**	**0.080**	**0.010**	**4.00E-18**	**0.062**	**0.008**	**8.07E-14**
RVESV	***SLC6A6*[d]**	**rs9856926**	**rs2128163, rs13061705**	**3**	**14280450**	**C**	**A**	**0.57**	**-0.060**	**0.010**	**1.40E-15**	**-0.049**	**0.008**	**2.18E-10**
RVESV	***AK097794***	**rs2276773**	**rs6792449**	**3**	**158287455**	**A**	**G**	**0.50**	**-0.060**	**0.010**	**1.60E-13**	**-0.053**	**0.008**	**8.38E-12**
RVESV	*CAMK2D*	rs55754224	–	4	114428714	C	T	0.74	0.060	0.010	7.20E-10	0.044	0.008	6.58E-08
RVESV	***PALLD*[††]**	**rs11357121**	**rs12509709**	**4**	**169847115**	**TA**	**T**	**0.20**	**0.060**	**0.010**	**4.60E-09**	**0.049**	**0.008**	**7.87E-09**

RVESV	***BAG3*[e]**	**rs72840788**	**rs4751742**	**10**	**121415685**	**G**	**A**	**0.78**	**0.080**	**0.010**	**2.70E-17**	**0.069**	**0.008**	**1.10E-16**
RVESV	***ATXN2*[f]**	**rs35350651**	**rs4766578**	**12**	**111907431**	**A**	**AC**	**0.49**	**-0.060**	**0.010**	**7.90E-15**	**-0.056**	**0.008**	**5.16E-13**
RVESV	***PTPN11***	**rs11066320**	**rs11066320**	**12**	**112906415**	**A**	**G**	**0.42**	**-0.050**	**0.010**	**5.20E-10**	**-0.045**	**0.008**	**1.36E-08**
RVESV	*RBL2*	rs61400540	rs61400540	16	53442606	G	A	0.68	-0.050	0.010	1.10E-08	-0.037	0.008	3.74E-06
RVESV	***GOSR2***	**rs17608766**	**rs17608766**	**17**	**45013271**	**T**	**C**	**0.85**	**-0.080**	**0.010**	**6.10E-12**	**-0.085**	**0.009**	**2.76E-22**
RVESV	*FHOD3*	rs503274	rs561021	18	34253745	C	T	0.30	-0.050	0.010	2.80E-09	-0.042	0.008	1.64E-07
RVESV	*ACTN4*	rs554462699	rs28488032	19	39164393	AT T	A	0.52	-0.050	0.010	1.20E-09	-0.034	0.008	1.06E-05
RVESV	*RSPH6A*	rs10402263	rs10412574	19	46313758	G	C	0.65	-0.050	0.010	3.00E-09	-0.039	0.008	7.01E-07
RVSV	*TTN*	rs2303838	rs967507	2	179444939	C	T	0.83	0.061	0.010	3.13E-09	0.039	0.009	8.62E-06
RVSV	***SLC35F1***	**rs3951016**	**rs6912208**	**6**	**118559658**	**T**	**A**	**0.53**	**0.050**	**0.010**	**3.10E-09**	**0.049**	**0.008**	**2.92E-10**
RVSV	*LUC7L2*	rs144567740	–	7	139099813	T	G	0.80	-0.057	0.010	1.13E-08	-0.041	0.009	1.49E-06
RVSV	***ATXN2***	**rs653178**	**rs7137828**	**12**	**112007756**	**C**	**T**	**0.48**	**-0.052**	**0.008**	**1.62E-11**	**-0.050**	**0.006**	**2.53E-15**
RVSV	***PTPN11***	**rs11066320**	**rs11066320**	**12**	**112906415**	**A**	**G**	**0.42**	**-0.047**	**0.008**	**2.39E-09**	**-0.044**	**0.006**	**1.15E-11**
RVEF	***CLCNKA***	**rs9442216**	**rs10927878**	**1**	**16353400**	**T**	**C**	**0.33**	**0.048**	**0.008**	**3.77E-09**	**0.046**	**0.007**	**9.09E-11**
RVEF	***OBSCN*[††]**	**rs78529941**	**rs55756479**	**1**	**228551488**	**G**	**A**	**0.62**	**0.050**	**0.010**	**3.50E-10**	**0.050**	**0.008**	**1.45E-10**
RVEF	***TTN*[g]**	**rs2042995**	**–**	**2**	**179558366**	**T**	**C**	**0.78**	**-0.066**	**0.009**	**1.62E-12**	**-0.059**	**0.008**	**5.59E-14**
RVEF	***SLC6A6*[h]**	**rs55834511**	**rs9856926, rs11715111**	**3**	**14273414**	**G**	**C**	**0.79**	**0.070**	**0.010**	**5.30E-13**	**0.070**	**0.009**	**2.05E-16**
RVEF	***AK097794***	**rs2276773**	**rs6792449**	**3**	**158287455**	**A**	**G**	**0.50**	**0.050**	**0.010**	**5.20E-09**	**0.054**	**0.008**	**2.89E-12**
RVEF	***HSPA4***	**rs72801474**	**rs72801474**	**5**	**132444128**	**G**	**A**	**0.90**	**-0.075**	**0.013**	**2.05E-08**	**-0.069**	**0.011**	**1.17E-09**
RVEF	*SLC23A1*	rs6876106	rs10063949	5	138710030	G	A	0.71	-0.050	0.010	2.10E-09	-0.038	0.008	2.24E-06
RVEF	***PLEC***	**rs11786896**	**rs11786896**	**8**	**145018354**	**C**	**T**	**0.95**	**-0.110**	**0.020**	**4.20E-09**	**-0.113**	**0.016**	**6.21E-12**
RVEF	***TPM2***	**rs2789750**	**–**	**9**	**35683473**	**C**	**G**	**0.68**	**0.060**	**0.010**	**8.40E-11**	**0.048**	**0.008**	**5.71E-09**
RVEF	***AK311445*[††]**	**rs12006440**	**rs12006440**	**9**	**107703337**	**C**	**T**	**0.97**	**0.140**	**0.020**	**2.40E-09**	**0.117**	**0.017**	**1.84E-11**

RVEF	***SVIL*†††**	**rs11007712**	**rs10826702**	**10**	**30009340**	**A**	**G**	**0.75**	**-0.060**	**0.010**	**8.50E-09**	**-0.046**	**0.008**	**4.93E-08**
RVEF	***BAG3***	**rs72840788**	**–**	**10**	**121415685**	**G**	**A**	**0.78**	**-0.090**	**0.010**	**8.00E-19**	**-0.081**	**0.008**	**6.33E-22**
RVEF	*CCDC85C*††	rs79884713	–	14	99987456	G	A	0.98	-0.160	0.030	3.70E-09	-0.092	0.025	1.81E-04
RVEF	***GOSR2***	**rs17608766**	**–**	**17**	**45013271**	**T**	**C**	**0.85**	**0.067**	**0.011**	**1.15E-09**	**0.071**	**0.009**	**4.52E-14**
RVEF	***FHOD3***	**rs501740**	**rs2644266**	**18**	**34244174**	**G**	**T**	**0.26**	**0.060**	**0.010**	**5.90E-11**	**0.054**	**0.008**	**4.97E-11**

*The locus name indicates the nearest annotated gene. Single-trait analysis was performed in a conventional genome-wide association framework using BOLT-LMM. Multi-trait analysis was performed by a joint-analysis of the summary statistics of correlated phenotypes using MTAG (*multi-trait analysis of GWAS). The loci additionally supported by the combined meta-analysis at $P_{meta\text{-}analysis} < 5x10^{-8}$ are highlighted in bold.
CMR, cardiovascular magnetic resonance; RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVEF, right ventricular ejection fraction; CHR, chromosome; EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency in UKBB cohort; SE, standard error
The effect size (BETA) represents the change in rank-transformed phenotype per effect allele.
[a]rs143720616 is an independent variant in this locus. [b]rs2339905 is an independent variant in this locus. [c]rs4894062 and rs115150884 are independent variants in this locus. [d]rs13061705 is an independent variant in the locus. [e]rs10718041 is an independent variant in the locus. [f]rs2339905 is an independent variant in this locus. [g]rs10930846 is an independent variant in this locus. [h]rs758925815 is an independent variant in this locus.
*[*]The nearest gene is HIST3H3 but the lead variant (*rs12126782) is in high LD (r^2 0.75) with rs78529941 (*OBSCN* locus) associated with RVEF.
[†]Co-localisation of GWAS variants and cis-eQTL variants was performed in eCaviar; [††]Loci showing some RV specificity.

Figure Legends

Figure 1. Flowchart of analysis strategy for RV GWASs

RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVSV, right ventricular stroke volume; RVEF, right ventricular ejection fraction; UKBB, UK Biobank; CMR, cardiovascular magnetic resonance; IVNT, rank-based inverse normal transformation; MAF minor allele frequency; INFO, imputation quality score; LD, linkage disequilibrium; GCTA, genome-wide complex trait analysis; LV, left ventricle; PheWAS, phenome-wide association study; PRS, polygenic risk score; SIFT, Sorting Intolerant From Tolerant score; PolyPhen-2, Polymorphism Phenotyping score 2; CADD, Combined Annotation Dependent Depletion score; eCaviar, eQTL and GWAS CAusal Variants Identification in Associated Regions tool; MTAG, multi-trait analysis of genome-wide association; MAGMA, Multi-marker Analysis of GenoMic Annotation; KO, knockout; Hi-C, long-range chromatic interaction

Figure 2. Manhattan plots of genomic loci associated with CMR-derived RV phenotypes

The red line indicates the genome-wide significant threshold at $P < 2.27x10^{-8}$.
RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVSV, right ventricular stroke volume; RVEF, right ventricular ejection fraction

Figure 3. Pleiotropic associations between the RV GWAS variants and other traits

The left semicircle represents the RV GWAS loci and the right semicircle represents previously reported GWAS traits in the Phenoscanner database.

GWAS, genome-wide association studies; CV, cardiovascular; CVD, cardiovascular disease; DCM, dilated cardiomyopathy

Figure 4. Enrichment of genes associated with RV phenotypes in g:Profiler

Gene sets and pathways results were corrected for multiple testing by the g:Profiler "sets counts and sizes" method at 5% threshold.

GO: BP, Gene Ontology Biological Process; GO:CC, Gene Ontology Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway

Figure 5. Phenome-wide association analysis of RV polygenic risk scores

A total of 1041 phenotypes were evaluated for associations with RV polygenic risk scores adjusted for age, sex and the first 5 genetic principal components by logistic regression. The red line represents the significant threshold after accounting for multiple testing (-log10 of 2.2×10^{-5}). The upright triangles indicate positive correlations and the inverted triangles indicate negative correlations.

RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVSV, right ventricular stroke volume; RVEF, right ventricular ejection fraction; NOS, not otherwise specified

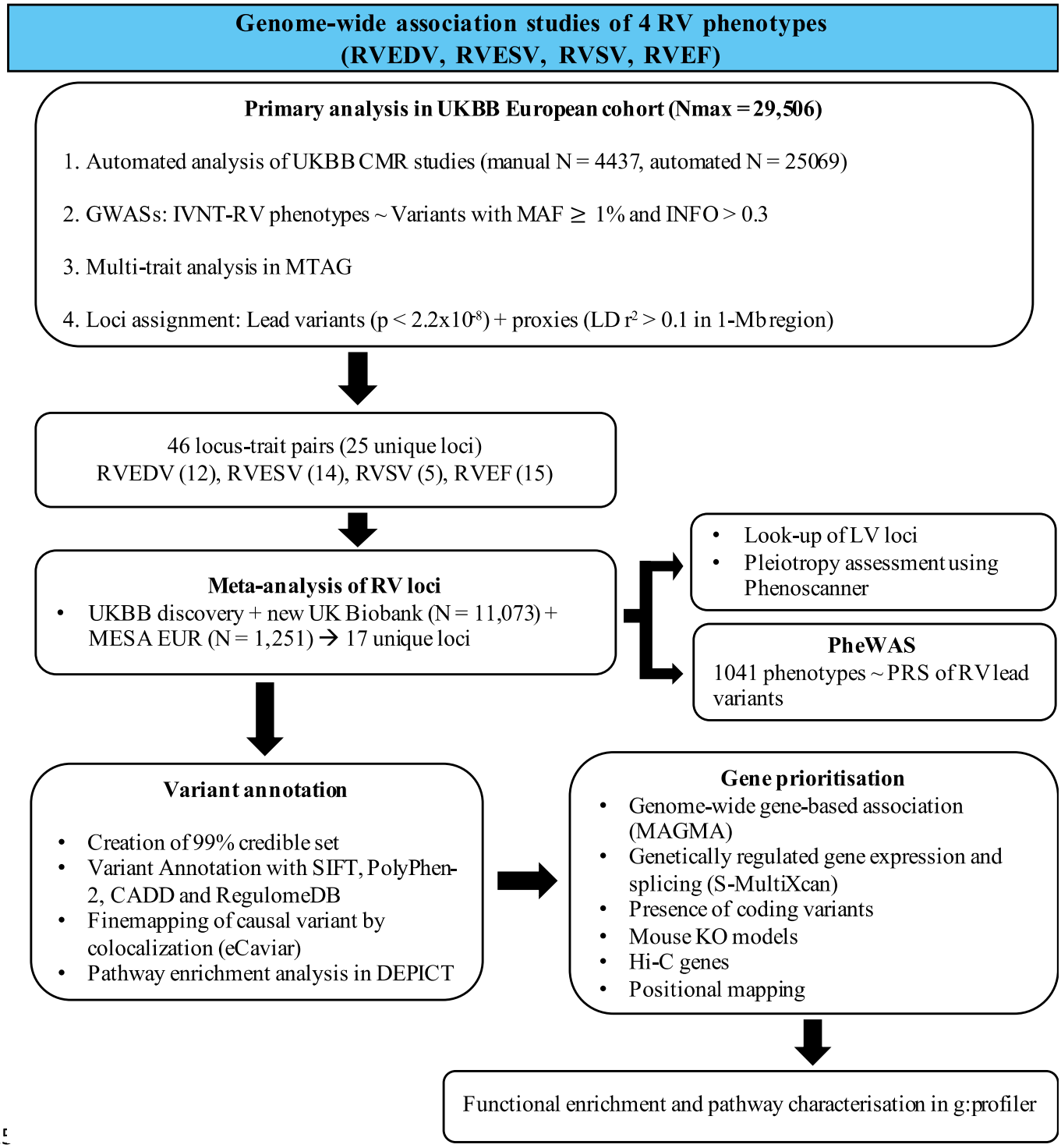

Figure 1. Flowchart of analysis strategy for RV GWASs

RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVSV, right ventricular stroke volume; RVEF, right ventricular ejection fraction; UKBB, UK Biobank; CMR, cardiovascular magnetic resonance; IVNT, rank-based inverse normal transformation; MAF minor allele frequency; INFO, imputation quality score; LD, linkage disequilibrium; GCTA, genome-wide complex trait analysis; LV, left ventricle; MESA, Multi-Ethnic Study of Atherosclerosis; EUR; European; PheWAS, phenome-wide association study; PRS, polygenic risk score; SIFT, Sorting Intolerant From Tolerant score;

PolyPhen-2, Polymorphism Phenotyping score 2; CADD, Combined Annotation Dependent Depletion score; eCaviar, eQTL and GWAS CAusal Variants Identification in Associated Regions tool; MTAG, multi-trait analysis of genome-wide association; MAGMA, Multi-marker Analysis of GenoMic Annotation; KO, knockout; Hi-C, long-range chromatic interaction

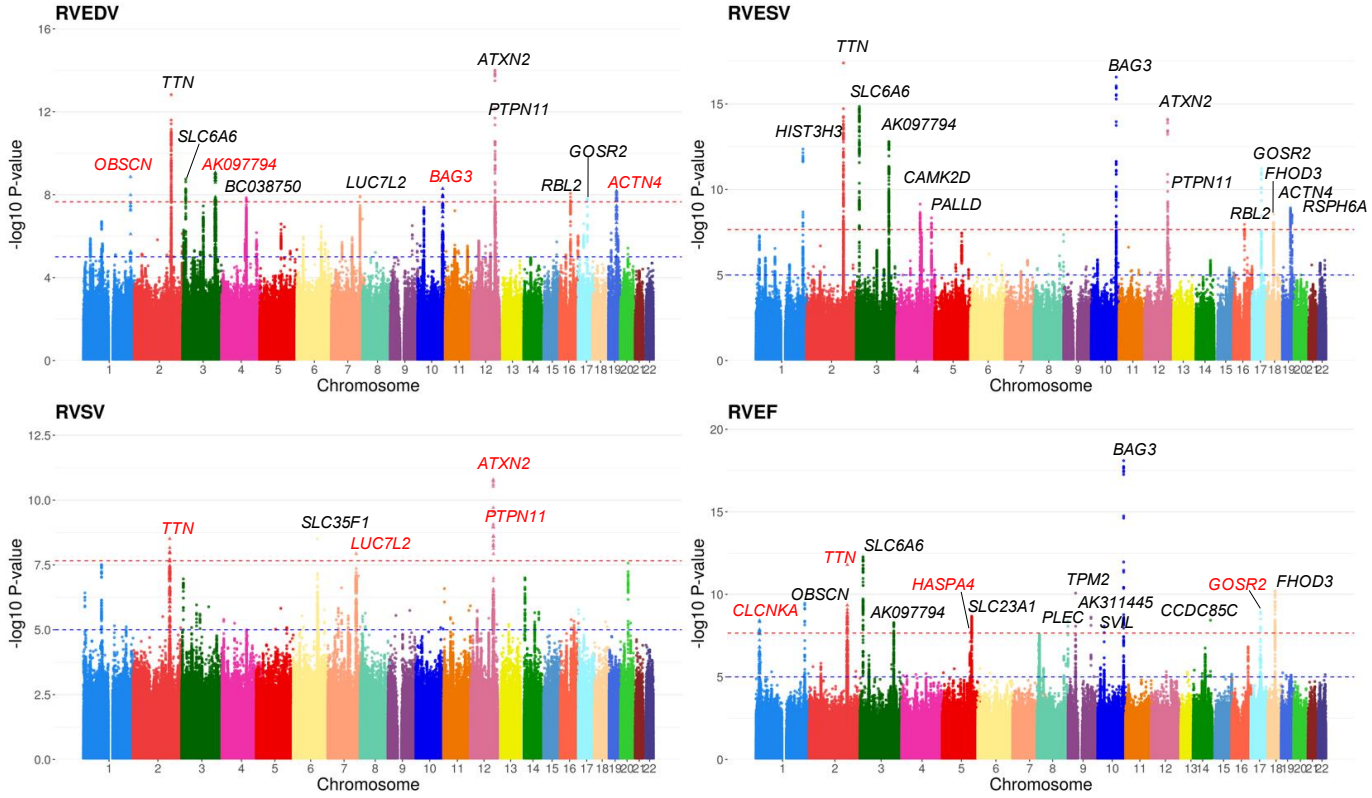

Figure 2. Manhattan plots of genomic loci associated with CMR-derived RV phenotypes

The red line indicates the genome-wide significant threshold at $P < 2.27x10^{-8}$. The loci discovered by the multi-trait analysis are shown in red.
RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVSV, right ventricular stroke volume; RVEF, right ventricular ejection fraction

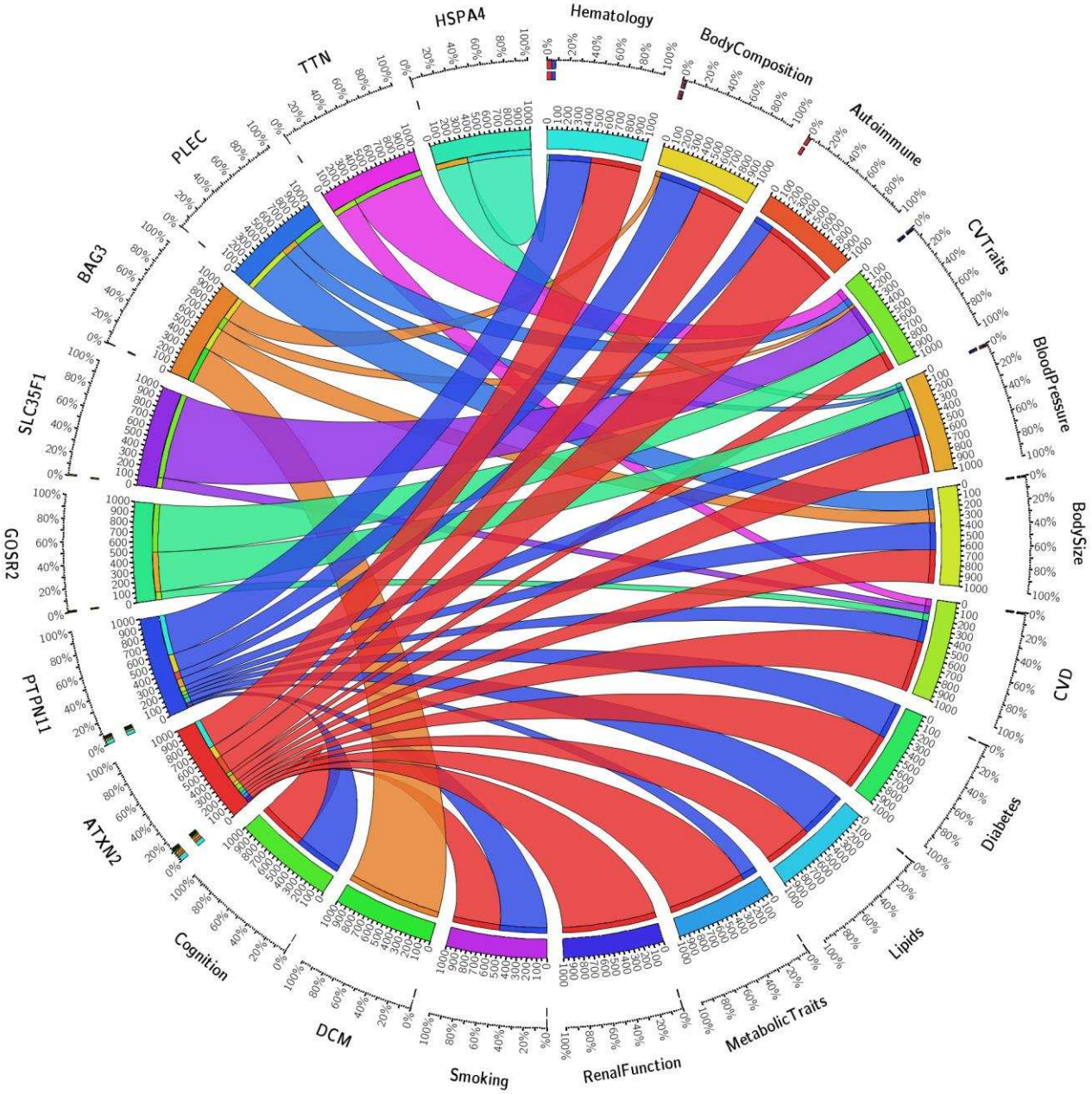

Figure 3. Pleiotropic associations between the RV GWAS variants and other traits

The left semicircle represents the RV GWAS loci and the right semicircle represents previously reported GWAS traits in the Phenoscanner database.

GWAS, genome-wide association studies; CV, cardiovascular; CVD, cardiovascular disease; DCM, dilated cardiomyopathy

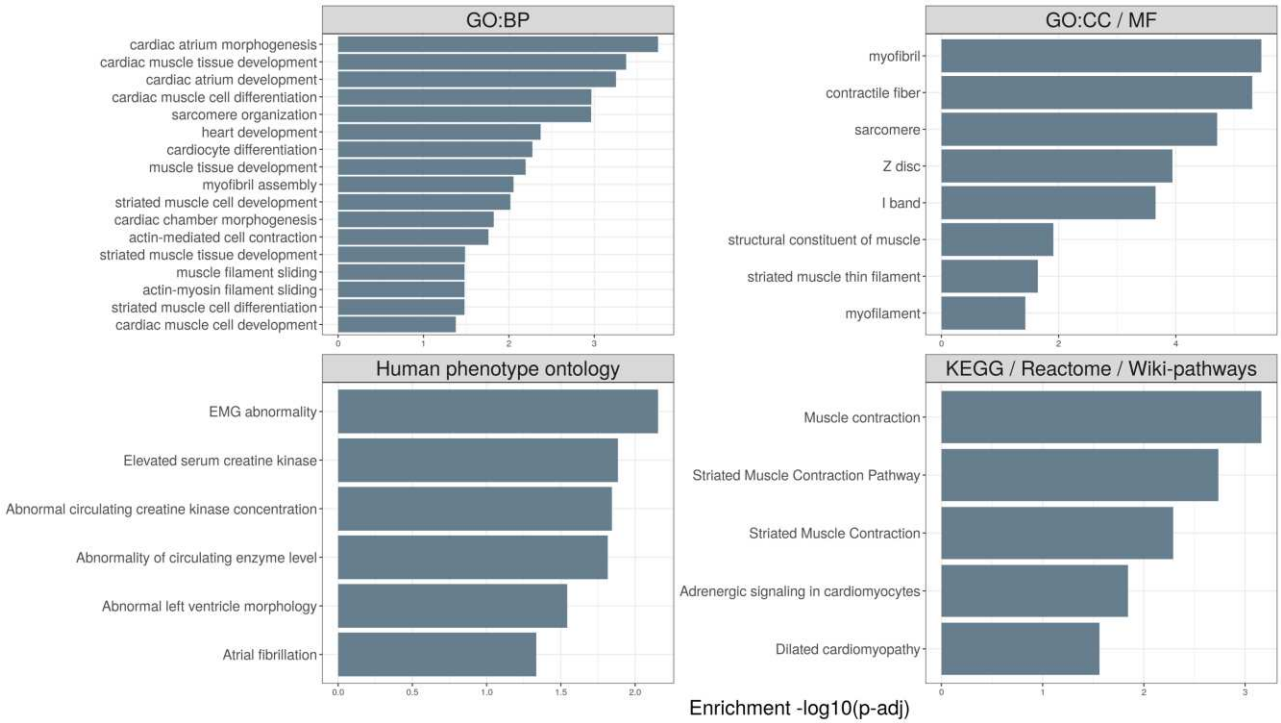

Figure 4. Enrichment of genes associated with all RV phenotypes in g:Profiler

Gene sets and pathways results were corrected for multiple testing by the g:Profiler "sets counts and sizes" method at 5% threshold. The results are available in a tabular format in Supplementary Table 13.

GO: BP, Gene Ontology Biological Process; GO:CC, Gene Ontology Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway

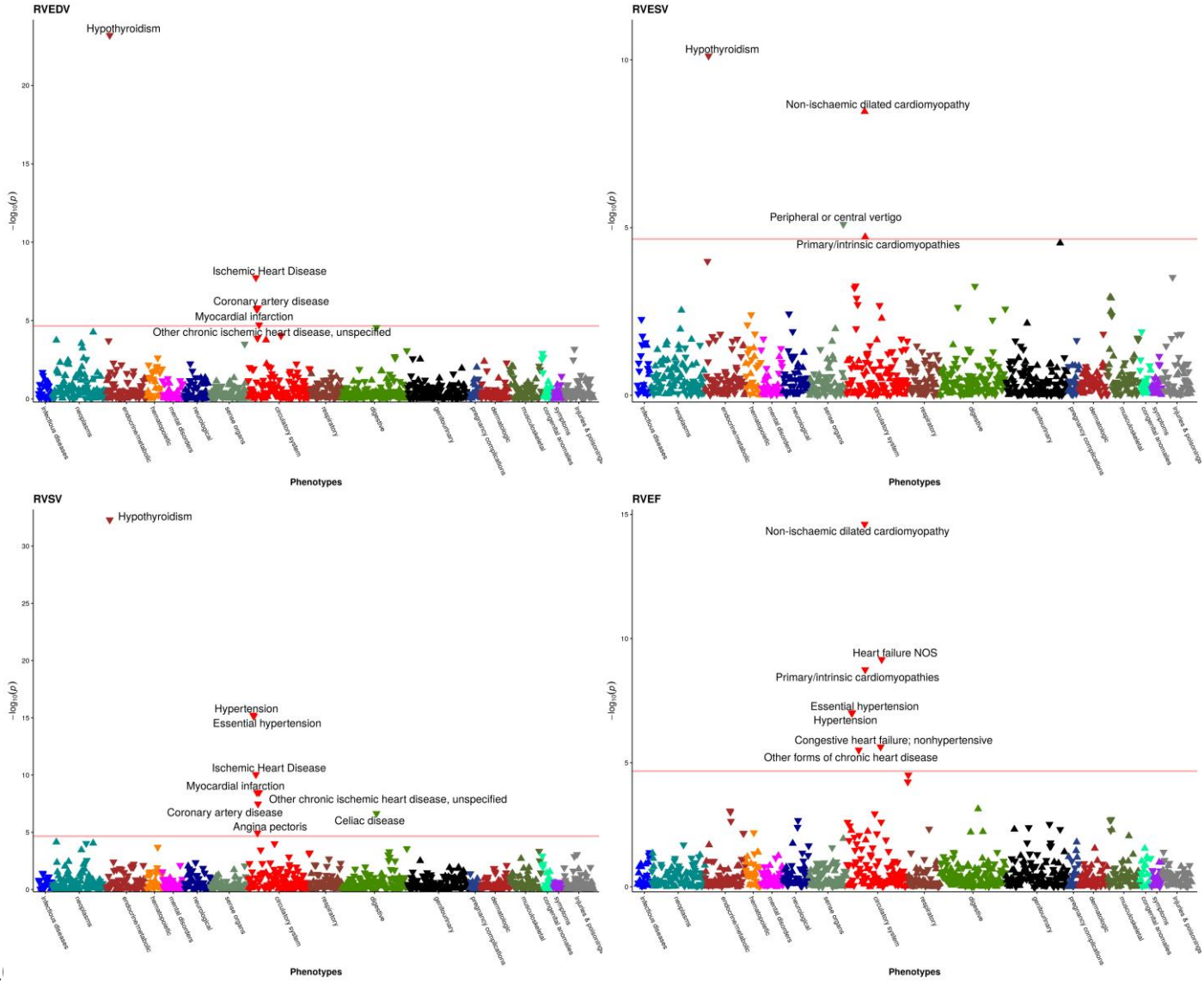

Figure 5. Phenome-wide association analysis of RV polygenic risk scores

A total of 1041 phenotypes were evaluated for associations with RV polygenic risk scores adjusted for age, sex and the first 5 genetic principal components by logistic regression. The red line represents the significant threshold after accounting for multiple testing (-log10 of 2.2 x 10^{-5}). The upright triangles indicate positive correlations and the inverted triangles indicate negative correlations.

RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVSV, right ventricular stroke volume; RVEF, right ventricular ejection fraction; NOS, not otherwise specified