Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development

Running title: Genome-wide association study of LV phenotypes

Nay Aung, MBBS, MRCP[1,2,3#]; Jose D Vargas, MD, DPhil[4#]; Chaojie Yang, MSc[5]; Claudia P Cabrera, PhD[6]; Helen R Warren, PhD[1,2]; Kenneth Fung, MBBS, MRCP[1,2,3]; Evan Tzanis, PhD[6]; Michael R Barnes, PhD[6]; Jerome I Rotter, MD[7]; Kent D Taylor, MD[7]; Ani W Manichaikul, PhD[5]; Joao AC Lima, MD[8]; David A Bluemke, MD[9]; Stefan K Piechnik, D.Sc[10]; Stefan Neubauer, MD, FRCP, FMedSci[10]; Patricia B Munroe, PhD[1,2#*]; Steffen E Petersen, MD, DPhil, MPH, FRCP[1,2,3#*]

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. Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

7. The Institute for Translational Genomics and Population Sciences, Division of Genomics Outcomes, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA

8. Division of Cardiology, Johns Hopkins University, Baltimore, Maryland, USA

9. Department of Radiology, University of Wisconsin, Madison, Wisconsin, USA

10. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

[#] These authors contributed equally to this work.

[*] Corresponding authors

Address for correspondence

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

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

Telephone: +44 (0) 20 7882 7188

Total word count: 4,993

Abstract

Background

The genetic basis of left ventricular (LV) image-derived phenotypes, which play a vital role in the diagnosis, management and risk stratification of cardiovascular diseases, is unclear at present.

Methods

The LV parameters were measured from the cardiovascular magnetic resonance (CMR) studies of the UK Biobank. Genotyping was done using Affymetrix arrays, augmented by imputation. We performed genome-wide association studies (GWASs) of six LV traits – LV end-diastolic volume (LVEDV), LV end-systolic volume (LVESV), LV stroke volume (LVSV), LV ejection fraction (LVEF), LV mass (LVM) and LV mass to end-diastolic volume ratio (LVMVR). The replication analysis was performed in Multi-Ethnic Study of Atherosclerosis (MESA). We identified the candidate genes at GWAS loci based on the evidence from extensive bioinformatic analyses. Polygenic risk scores (PRSs) were constructed from the GWAS summary statistics to predict the heart failure events.

Results

The study comprised 16,923 European UK Biobank participants (mean age: 62.5 years; 45.8% men) without prevalent myocardial infarction or heart failure. We discovered fourteen genome-wide significant loci – three loci each for LVEDV, LVESV and LVMVR; four loci

for LVEF and one locus for LVM – at a stringent $p < 1 \times 10^{-8}$. Three loci were replicated at Bonferroni significance and seven loci at nominal significance ($P < 0.05$ with concordant direction of effect) in the MESA study (N = 4,383). Follow-up bioinformatic analyses identified 28 candidate genes which were enriched in the cardiac developmental pathways and regulation of the LV contractile mechanism. Eight genes (*TTN, BAG3, GRK5, HSPB7, MTSS1, ALPK3, NMB* and *MMP11*) supported by at least two independent lines of in-silico evidence were implicated in the cardiac morphogenesis and heart failure development. The PRSs of LV phenotypes were predictive of heart failure in a hold-out UK Biobank sample of 3,106 cases and 224,134 controls (odds ratio 1.41, 95% CI: 1.26 – 1.58, for the top quintile vs the bottom quintile of the LVESV risk score).

Conclusions

We report fourteen genetic loci and indicate several candidate genes which not only enhance our understanding of the genetic architecture of prognostically important LV phenotypes but also shed light on potential novel therapeutic targets for LV remodelling.

Key words

Genome Wide Association Study; Left ventricle; Left ventricle geometry; Heart failure

Clinical Perspective

What is New?

- Prognostically important left ventricular imaging phenotypes are highly heritable (~22% to 39%).
- A total of fourteen genetic susceptibility loci (eight of which are unique) enriched in the cardiac developmental pathways and regulation of contractile mechanism, are discovered in the largest genome-wide association study of CMR-derived left ventricular phenotypes.
- The polygenic risk scores of left ventricular phenotypes are predictive of heart failure events independently of clinical risk factors.

What Are the Clinical Implications?

- The findings from this study not only enhance our understanding of the genetic basis of prognostically important LV phenotypes in the general population but also underscore the intricate genetic relationship between these endophenotypes and the pathogenesis of the heart failure syndrome.
- The prioritized genes in the genome-wide significant loci should be followed up in the functional studies to aid the development of potential novel therapies for heart failure.
- The polygenic risk scores of left ventricular phenotypes may have a role in personalized risk stratification pending further validation of clinical robustness in future studies.

Non-standard Abbreviations and Acronyms

BMI, body mass index

CMR, cardiovascular magnetic resonance

DBP, diastolic blood pressure

DCM, dilated cardiomyopathy

ECG, electrocardiogram

ECHO, echocardiography

eQTL, expression quantitative trait locus

FDR, false discovery rate

GTEx, genotype-tissue expression dataset

GWASs, genome-wide association studies

INFO, imputation quality score

INT, rank-based inverse normal transformation

LD, linkage disequilibrium

LV, left ventricle/left ventricular

LVEDV, left ventricular end-diastolic volume

LVEF, left ventricular ejection fraction

LVESV, left ventricular end-systolic volume

LVH, left ventricular hypertrophy

LVM, left ventricular mass

LVMVR, left ventricular mass to end-diastolic volume ratio

LVSV, left ventricular stroke volume

MAF, minor allele frequency

MESA, multi-ethnic study of atherosclerosis

PRS, polygenic risk score

SBP, systolic blood pressure

Introduction

Heart failure is a clinically heterogeneous condition associated with a substantial mortality, morbidity and economic burden to the society[1]. Globally, both incidence and prevalence of heart failure are increasing due to improved survival from other contributory cardiovascular diseases in an ageing population. The diagnosis and treatment of heart failure is in part based on left ventricular (LV) functional and structural parameters derived from cardiac imaging. Although the impact of modifiable risk factors on LV structure and function is well established, our current understanding of the genetic component of these imaging phenotypes is limited.

Previous large-scale genetic association studies of LV imaging phenotypes[2,3] were hampered by the lack of a standardised measurement protocol in the phenotyping process and reliance on two-dimensional echocardiography (ECHO) with inherent dependency on geometric assumptions and adequate acoustic window. These shortcomings can be overcome by using the individual-level data from a single large study such as the UK Biobank, which also provides accurate and reproducible imaging phenotypes from cardiovascular magnetic resonance (CMR) imaging, considered to be a reference standard for the assessment of cardiac morphology[4].

Although multiple indices of LV structure and function can be measured from CMR images, five parameters – LV mass (LVM), LV end-diastolic volume (LVEDV), LV end-systolic volume (LVESV), LV stroke volume (LVSV) and LV ejection fraction (LVEF) – are frequently used in clinical practice and carry prognostic information[5,6]. In addition, LV mass to end-diastolic volume ratio (LVMVR) represents geometric remodelling of the left ventricle

and an elevated LVMVR reflects concentric remodelling or hypertrophy associated with adverse outcomes[7]. Systematic genome-wide scanning for loci associated with the LV image-derived measurements is a vital first step, which will advance our understanding of their genetic basis in a general population and may inform on novel diagnostic and targeted therapeutic opportunities. In this study, we conducted genome-wide association studies (GWASs) to identify the genetic loci for six clinically relevant CMR-derived LV imaging phenotypes.

Methods

Data access

The data including GWAS summary statistics, analytic methods, and study materials will be returned to the UK Biobank. The UK Biobank will make these data available to all bona fide researchers for all types of health-related research that is in the public interest, without preferential or exclusive access for any person. Please see the UK Biobank's website for the detailed access procedure (http://www.ukbiobank.ac.uk/register-apply/).

UK Biobank

The UK Biobank is a large population-based prospective cohort study of half a million people aged between 40-69 years at the time of initial recruitment between 2006 and 2010. It has collected a wealth of information on health and lifestyle data, physical measurements, biological samples, genotype and cardiac phenotypes derived from CMR. The study protocol has been described in detail previously[8]. The study complies with the Declaration of Helsinki

and was approved by our institutional review body. All participants provided informed written consent.

Genetic data

Genotypes directly called by two closely-related, purpose-built arrays known as UK Biobank Axiom™ Array (825,927 markers) and UK BiLEVE Axiom™ Array (807,411 markers) were imputed to ~92 million variants using two reference panels (Supplemental Methods, *UK Biobank genetic data*).

CMR phenotypes

The detailed CMR protocol and analysis methods have been described previously[9] and further details are available in Supplemental Methods, *UK Biobank CMR phenotypes* and Supplemental Figure 1. Since our primary aim was to investigate the genetic basis of LV image-derived phenotypes, we excluded individuals with prevalent myocardial infarction, heart failure or LVEF < 50% to minimise the confounding influence of these pre-existing conditions. Additional sample quality control measures were outlined in Supplemental Methods, *Sample quality control* and Supplemental Figure 2).

Genetic analyses

Primary analyses

The detailed analysis strategy is outlined in Supplemental Methods, *UK Biobank genetic association analysis* and Supplemental Figure 3. We estimated the heritability explained by the genotyped variants (h_g^2 SNP) and bivariate genetic correlation (r_g) using a variance component method implemented in BOLT-REML[10]. We next performed the discovery GWAS of each LV trait using a linear mixed-model method by BOLT-LMM[11], under an additive genetic model with ~ 7 million imputed genetic variants with minor allele frequency (MAF) ≥ 5% and imputation quality score (INFO) > 0.3. Both heritability analysis and GWAS models were adjusted for age, sex, height, weight, systolic blood pressure (SBP) corrected for anti-hypertensive medication use (by adding 15mmHg), phenotype-derivation method (automatic/manual), array type (UK Biobank vs UK BiLEVE array), and imaging centre. (See Supplemental Methods, *Definitions of covariates*). The untransformed LV phenotypes which showed evidence of positive skewness normalised well after rank-based inverse normal transformation (INT) (Supplemental Figure 4). We used these INT phenotypes in all primary analysis models for heritability, genotypic correlation and GWAS. In addition to consideration for the multiple testing of genotype involved in a single GWAS, we had to consider multiple-hypothesis testing with our six distinct, albeit correlated, LV traits. The effective number of phenotype-association tests estimated by Galwey method was 3.3. However, we set a more stringent threshold for genome-wide significance at $p < 5 \times 10^{-8}/5 = 1 \times 10^{-8}$ given the absence of a large replication cohort of comparable size. Genome-wide significant loci were defined by the most significant variant (known as the sentinel or lead variant) and their proxies (correlated variants) in linkage disequilibrium (LD) $r^2 > 0.1$ in 1Mb region.

We also performed conditional analysis to determine and the presence of secondary independent signals within the GWAS loci and nested linear regression analyses to calculate

the percentage variance explained by the lead variants (Supplemental Methods, *Conditional analysis* and *Percentage Variance*).

Secondary analyses

As a sensitivity analysis, we adjusted the primary analysis models with additional cardiovascular risk factors, namely diastolic blood pressure (DBP) corrected for anti-hypertensive medication use (by adding 10mmHg), body mass index (BMI) as a measure of obesity (replacing weight in the primary model to avoid collinearity), smoking status, regular alcohol use, dyslipidaemia and diabetes (See Supplemental Methods, *Definitions of covariates*). We also repeated the association analyses with untransformed LV traits while controlling for the same covariates as the primary analysis to obtain more comprehensible effect sizes (β) which are in the same unit of measurements as the LV traits (Supplemental Methods, *Secondary analyses*). Since the ratio of LV mass to $LVEDV^{0.67}$ ($concentricity^{0.67}$) was previously shown to be more correlated with both LV wall thickness and SBP than the standard definition of concentricity (LVMVR)[12], we repeated the association analysis with this phenotype. Furthermore, potential mediating effects of SBP and sex were explored by incorporating the lead variant * covariate interaction terms in the primary models, using the mixed-model method implemented in the MMAP software (https://mmap.github.io/). Bonferroni correction was applied on the interaction p values to adjust for the number of variants tested.

Replication analyses in the MESA cohort

We performed the association analyses for all our sentinel variants in both European and non-European ancestries of the Multi-Ethnic Study of Atherosclerosis (MESA)[13]. After sample quality control (See sample selection flowchart in Supplemental Figure 5), a total of 4,383 individuals from MESA study (European = 1,742, African American = 1,083, Chinese = 586, Hispanic = 972) were included in the look-up analysis. Further details on the design and analysis of MESA cohort are available in Supplemental Methods.

Pleiotropy analyses

We searched PubMed to collate all genome-wide significant variants ($p < 5x10^{-8}$) reported in published literature on closely related phenotypes (ECHO-derived LV measurements, left ventricular hypertrophy [LVH] identified by electrocardiogram [ECG]) and performed a lookup of these variants in our GWAS results. We cross-referenced our sentinel variants and their close proxies (LD $r^2 \geq 0.8$) with Phenoscanner[14] database v2 (http://www.phenoscanner.medschl.cam.ac.uk/) and presented the variants which showed strong associations with other traits at $p < 5x10^{-8}$. We also interrogated the GeneAtlas[15] database (http://geneatlas.roslin.ed.ac.uk/) to assess the associations of our variants with other traits in the UK Biobank.

Functional annotation

We employed an integrative bioinformatics approach to compile the functional information at both variant and gene level. Significant genomic loci were annotated using multiple lines of evidence including presence of coding variant, gene expression data, chromatin interaction

analyses, knockout models and literature review (Supplemental methods, *Bioinformatic annotation*).

Polygenic scoring

We used the LDPred tool[16] to construct the polygenic risk score (PRS) of each LV trait based on the effect sizes derived from the LV GWASs and predicted the risk of heart failure event in the remainder of the UK Biobank cohort. LDPred considers the tuning parameter known as the fraction of causal variant ($\boldsymbol{\rho}$). In order to choose the best unbiased $\boldsymbol{\rho}$, we first split the UK Biobank dataset into the training (2,033 cases; 149,461 controls) and test (3,106 cases; 224,134 controls) sets. The final PRS constructed from the best-fit $\boldsymbol{\rho}$ value was used to predict heart failure in the hold-out test dataset using a logistic regression model controlled for age, sex, BMI, SBP and DBP adjusted for anti-hypertensive medication use, smoking status, regular alcohol use, dyslipidaemia, diabetes, and 15 genetic principal components (PCs). We also explored the prospective association between LV-PRS quintiles and incident heart failure in the test dataset using multivariate Cox proportional hazards models adjusted for the same covariates as the logistic regression models. Further information on polygenic risk score is available in Supplemental Methods, *Polygenic risk prediction of heart failure events*.

Results

The overall study design is illustrated in Figure 1 and the summary characteristics of the UK Biobank CMR cohort are presented in Supplemental Table 1. The mean±SD age of the cohort 62.5±7.5 years and 45.8% were men. The primary analysis comprised a total of 16,923

European individuals with a maximum sample size of: LVEDV (n = 16,920), LVESV (n = 16,920), LVSV (n = 16,917), LVEF (n = 16,923), LVM (n = 16,920) and LVMVR (n = 16,884). Approximately 25% of the CMR studies were manually analysed and the remainder were segmented by a deep learning algorithm. The reproducibility of both manual and automatic measurements was very high (intra-class correlation coefficient ranged from 0.88 to 0.98).

Heritability and genotypic correlation

The highest genome-wide heritability (h^2_g SNP) estimates were observed for the structural traits such as LVEDV and LVESV (both at 39%), followed by LVM at 34% and LVMVR at 33% and the functional traits such as LVSV and LVEF had lower heritability (25% and 22%, respectively). The genotypic correlations between LV traits ranged from very high (r_g = 0.92 between LVEDV and LVSV) to very low (r_g = -0.01 between LVSV and LVEF) (Figure 2).

Genomic loci associated with LV phenotypes

We discovered a total of fourteen genomic loci defined by a 1MB region – three loci each for LVEDV, LVESV and LVMVR; four loci for LVEF and one locus for LVM – at a stringent $p < 1 \times 10^{-8}$ as summarised in Table 1 and Figure 3. There was no evidence of population stratification or cryptic relatedness (genomic inflation factor, $\lambda = 1.047 – 1.097$, quantile-quantile plots in Supplemental Figure 6). We assigned a single "sentinel" variant with the lowest GWAS p value for each locus and plotted the LocusZoom plots for all sentinel variants (Supplemental Figure 7). Variants at several loci were associated with more than one LV trait. The *TTN* locus was associated with four LV traits (LVEDV, LVESV, LVEF, LVM),

and the *BAG3* and *MTSS1* loci were shared across more than one trait (Figure 4). The LV remodelling trait, LVMVR, had three distinct loci (*CDKN1A*, *DERL3* and *ZNF592*) which were not shared with other LV traits. No significant locus was found for LVSV. In total, eight unique loci were identified. There was no evidence of secondary independent variants achieving the pre-specified criteria at any loci in conditional analyses (Supplemental Table 2). The percentages of trait variance explained by the sentinel variants were small as expected given the limited number of significant loci for each trait (LVEDV: 0.21%, LVESV: 0.32%, LVEF:0.38%, LVM: 0.10%, LVMVR: 0.46%).

Secondary analyses

The secondary analysis models additionally adjusted for other cardiovascular risk factors produced results generally congruous with the primary analyses, except for the *SH2B3* locus for LVEDV where the GWAS p value of sentinel variant (rs7310615) became significantly attenuated (GWAS p secondary = 1.5×10^{-5}, -log10 p primary/-log10 p secondary = 1.8; see Supplemental Table 3). The β estimates and the p values of our primary results and the analyses with untransformed LV phenotypes were highly correlated ($\rho = 0.98$ to 1.0 for β and $\rho = 0.94$ to 0.99 for p) with no significant attenuation of GWAS p values (-log10 p primary/-log10 p secondary < 1.5) (Supplemental Table 4). The absolute effect sizes in the models with untransformed phenotypes ranged between 1 to 2ml for LVEDV and LVESV and ~0.5% for LVEF per allele (Table 1). The lead variants for LVMVR loci remained genome-wide significant ($p < 1 \times 10^{-8}$) in the sensitivity analyses with concentricity$^{0.67}$ index. There was no significant interaction between SBP and the lead variants (Supplemental Table 5). In contrast, male sex significantly mediated the effects of lead variants in the TTN locus in the direction of larger LVEDV (interaction **β** = 2.63 ml per allele, $p = 8.9 \times 10^{-6}$) and LVESV

(interaction **β** = 1.43 ml per allele, p = 9.3 x 10^{-6}) and higher LVM (interaction **β** = 1.42 g per allele, p = 2.7 x 10^{-5}) (Supplemental Table 6).

Follow up of loci in an independent multi-ethnic cohort

Out of fourteen locus-trait associations discovered in the UK Biobank, we validated three loci at Bonferroni significance (P < 0.0036 [0.05/14]) and seven loci at nominal significance (P < 0.05 with concordant direction of effect) in at least one of the MESA ethnic sub-cohorts (summary characteristics in Supplemental Table 7). The rs200712209 variant associated with LVESV and rs34866937 variant associated with LVEF – both in *MTSS1* locus – were replicated in the European subset of MESA study after Bonferroni correction. Three other variants tagging *TTN*, *BAG3* and *ZNF592* loci were nominally associated with LVESV, LVEF and LVMVR traits in the MESA European cohort (Supplemental Table 8).

,

Genetic relationships between CMR LV phenotypes with other related traits

For ECHO traits, two previously reported variants in the *SH2B3* and *MTSS1* loci were genome-wide significant and four other variants were nominally significant (p < 0.05 with concordant directionality) for the corresponding CMR traits in our GWAS (Supplemental Table 9). The rs10774625 variant reported by Wild et al.[2] for ECHO-derived LV end-diastolic diameter (LVEDD) was in high LD (r^2 = 0.95) with the sentinel variant at *SH2B3* in our LVEDV GWAS. Previously unvalidated rs34866937 variant at the *MTSS1* locus, reported by Kenai et al.[3] for ECHO-derived LV end-systolic diameter (LVESD) and ECHO-derived LVEF in a Japanese population was genome-wide significant for CMR-derived LVESV and LVEF in our European GWAS. For ECG-LVH traits, thirteen previously

reported variants were nominally associated with our CMR-derived LVM (Supplemental Table 10).

We also explored the association between our sentinel variants and their proxies with other traits from published GWASs using Phenoscanner[14]. The variants in the *SH2B3* locus for LVEDV was associated with multiple risk factors which could mediate the cardiac remodelling process such as blood pressure/hypertension, cholesterol/LDL level, diabetes and smoking status. The variants in the *CLCNKA* and *BAG3* loci for LVEF and LVESV were associated with dilated cardiomyopathy (DCM) (Supplemental Table 11). We also interrogated the Gene Atlas PheWAS database which reported the association results of hundreds of traits in the UK Biobank with our sentinel variants. The sentinel variant in the *SH2B3* locus (rs7310615) was again found to be highly associated with the presence of hypertension ($p = 5.9 \times 10^{-46}$) and ischaemic heart disease ($p = 4.8 \times 10^{-14}$) in the UK Biobank (Supplemental Table 12).

Functional annotation of variants

At variant-level annotation, we identified a total 238 candidate variants for all LV traits (7.6% exonic variants, 45% intronic variants, 30% intergenic variants; see Supplemental Figure 8). Out of 18 exonic variants, 12 were nonsynonymous and were located in the *BAG3*, *ALPK3*, *TTN*, *SH2B3*, *NMB* and *WDR73* genes. The missense variant, rs2234962, in the *BAG3* gene for LVESV and LVEF was predicted to be damaging by at least two prediction tools.

Among 220 non-coding variants, 52 were located in promotor histone marks, 148 in enhancer histone marks, 82 in DNase I sites and 13 altered the binding sites of regulatory proteins. We also found 6 non-coding variants which were highly conserved in vertebrates according to SiPhy[17]. RegulomeDB[18] ranked three non-coding variants in the *ZNF592* locus as class 1f (strong support for functional importance). Of these variants, two (rs2175567 and rs17598603) were intronic for the *NMB* gene and the other (rs7237) was located in the 3' UTR region of the *WDR73* gene. The expression quantitative trait locus (eQTL) analysis in Genotype-Tissue Expression (GTEx) dataset[19] revealed that rs35006907, a close proxy of the sentinel variants in the *MTSS1* locus for LVESV and LVEF, was associated at false discovery rate (FDR) < 0.05 with the *MTSS1* gene expression in left ventricle and left atrial appendage tissues. Additionally, rs2070458, a proxy variant in the *DERL3* locus for LVMVR was associated with the *MMP11* gene expression in left ventricular tissue (Supplemental Table 13).

We explored the long-range interaction influence of the sentinel variants and their proxies in the Hi-C data of aorta, left ventricle and right ventricle and found 11 potential target genes (Supplemental Table 14). A summary of all variant-level annotations is presented in Supplemental Table 15. The gene prioritisation, gene-set enrichment and pathway analyses for the LV traits did not yield any significant results. However, in the tissue-specific enrichment analysis using the data from the Roadmap Epigenomics project, there was a significant enrichment of our genome-wide significant LV variants within the regions of DNase I hotspots in foetal heart tissue (Supplemental Figure 9).

At gene-level annotation, we identified a total of 28 candidate genes at eight unique loci: five genes were prioritised by presence of nonsynonymous variant (sentinel or proxies with LD r^2

≥ 0.8), six were prioritised by eQTL data in cardiovascular tissues, 11 were prioritised based on long-range chromatin interactional analyses and six were prioritised by availability of knockout models (Supplemental Table 16). A summary of prioritised genes for all LV traits is depicted in Figure 5. Evaluation of these genes as input in GeneNetwork[20] pathway analysis revealed enrichment for terms related to "heart development", "regulation of the force of heart contraction" and "abnormality of the cardiovascular system" (Supplemental Table 17).

Polygenic risk score analyses

We explored the predictive ability of polygenic risk scores (PRSs) derived from the genetic variants associated with LV traits to predict heart failure events in an independent test sample of the UK Biobank cohort (3,106 cases, 224,134 controls). All LV-PRSs (except LVM-PRS) were significantly associated with heart failure. The PRS quintiles of LVEDV and LVESV were associated with higher odds of heart failure (OR: 1.25 – 1.41 for the top 20% vs bottom 20% group) while the opposing pattern was observed for LVEF- and LVMVR-PRSs across all levels of adjustments for potential confounders including age, sex, body size and cardiovascular risk factors (Table 2). The sensitivity analyses including only incident cases produced similar patterns of association between LV-PRS quintiles and incidence of heart failure (Supplemental Figure 10).

Discussion

This is the largest individual-level genome-wide association study to date investigating the genetic architecture of prognostically important LV phenotypes derived from CMR. The key

strength of this study is the combination of standardised, highly precise and reproducible phenotyping process with a high-quality dense genotype dataset in a cohort of ~ 17,000 individuals free from myocardial infarction and heart failure. This strategy yielded a total of fourteen locus-trait associations. Three loci (*TTN*, *BAG3* and *MTSS1*) were shared between more than one LV trait resulting in the discovery of eight unique loci, of which six were novel for LV imaging traits. The follow-up analysis of our sentinel variants in an independent multi-ethnic cohort (MESA) showed a strong support for two loci (*MTSS1* and *BAG3*) at the level of Bonferroni-significance while three other loci (*TTN*, *SH2B3* and *ZNF592*) achieved a nominal support. Our enrichment analyses revealed that the regulatory variants associated with the LV loci are highly enriched in the foetal heart tissue and the candidate genes for these loci are involved in the cardiac developmental pathways and regulation of the LV contractile mechanism.

Although the LV image-derived phenotypes were known to be heritable, their reported heritability varied widely, ranging from 15%[21] to 84%[22] for LVM, which may reflect the methods used including monozygotic and dizygotic twins, siblings, nuclear and complex families. In this study, we provide a robust estimate of the proportion of LV phenotypic variance explained by the genotype (h^2_g SNP heritability) in a large population of unrelated Caucasian individuals. A significant proportion of LV phenotypic variability was explained by the genotype (ranging from 22% to 39%) and the structural traits such as LVEDV or LVM had a noticeably higher heritability than more functional traits such as LVEF. The relatively lower heritability estimates in these functional traits could be due to the inflated inter-personal variations secondary to differences in the loading conditions and chronotropic and inotropic states[23]. The heritability estimates of CMR-derived LV traits were overall higher than previously reported values in other complex cardiovascular traits such as resting heart

rate (h^2_g = 21%)[24] and comparable ECHO traits (for example, LVM h^2_g = 14.8% for ECHO vs. 39% for CMR)[3]. However, the heritability of the ECHO trait was calculated using the summary-level data with LD score regression approach, which tends to produce lower estimates than the variance component method (which requires the individual-level data) as in our study. The pattern and magnitude of genotypic correlations between LV traits mostly mirrored the corresponding phenotypic correlations except for the relationship between LVSV and LVEF, where the genotypic correlation was absent despite a moderate phenotypic correlation (r_p = 0.39 vs r_g = -0.01). The finding may suggest dissimilar genetic architecture between these two functional traits which should be validated in an independent cohort. The majority of genomic loci observed in our study were specific for the LV traits except for the *SH2B3* locus (associated with LVEDV) which appeared to be highly pleiotropic and associated with multiple cardiovascular risk factors. This finding may explain the substantial weakening of the sentinel variant's association p value in our secondary analysis additionally adjusted for a wider range of cardiovascular risk factors.

Among several candidate genes, we highlighted the functional roles of a few potential causal genes based on the bioinformatic analyses and literature review. The *TTN* (titin) gene emerged as a strong candidate causal gene for four LV traits (LVEDV, LVESV, LVEF and LVM). The pivotal function of *TTN* in the maintenance of sarcomere assembly, stretch sensing and signalling, passive stiffness adjustment and active force generation[25] corresponds well with our finding of *TTN* being an important gene for both structural and functional LV imaging traits. Several titin-truncating variants (TTNtv) have been found in 10% to 20% of DCM cases and may have a role in the pathogenesis of DCM in these individuals[26]. These TTNtv have been reported to occur at much lower frequencies in the general population (1.1% of alleles in the 1000 Genomes project)[27] and were found to be associated with larger

LV volumes[28]. In our study including variants with minor allele frequency ≥ 5%, we found several missense variants (but no protein truncation variants) in the *TTN* locus (Supplemental Table 13) which were all predicted to be benign by the pathogenicity prediction tools.

The *BAG3* (BCL2 associated athanogene 3) gene appears to be a likely candidate gene which is involved in modulation of the two volumetric traits (LVEDV and LVESV) and the derived LV functional index (LVEF). *BAG3* encodes an anti-apoptotic protein expressed in the heart and skeletal muscle and serves as a co-chaperone of heat-shock protein (HSP) family[29]. It is essential for the homeostasis of filamin[30] and influences myocyte contraction through interaction with the b1-adrenegic receptor and the L-type calcium channel[31]. Mutations in *BAG3* gene have been implicated in DCM pathogenesis with the myocardial tissues displaying evidence of myofibril disarray and relocation of *BAG3* protein in the sarcomeric Z-disc[32]. In our cohort of individuals with preserved LVEF, free from known heart failure or DCM, we found a missense mutation (rs2234962, LD r^2 0.99 with the sentinel variant) in the *BAG3* gene which was reported to be damaging by two in-silico prediction tools. The same missense variant has been indicated as the sentinel variant in a DCM GWAS[33], underlining the shared genetic basis of LV volumetric phenotypes and DCM. Our long-range chromatin interaction (Hi-C) analysis flagged up another potential candidate gene at the *BAG* locus called *GRK5* (G Protein-Coupled Receptor Kinase 5) gene. *GRK5* regulates cardiac development through the mTOR pathway in Zebrafish[34,35]. Its expression level in the myocardium is elevated in heart failure[36] and a non-synonymous polymorphism of *GRK5* appears to be protective of heart failure through inhibition of β-Adrenergic receptor signalling[37]. In addition, *GRK5* is a known drug target for β-blockers and anti-hypertensive agents (Supplemental Table 14). At the *CLCNKA* locus for LVEF, we indicate *HSPB7* (Heat Shock Protein Family B Member 7) as a likely candidate given its recognised role in

cardiogenesis by modulating actin filament assembly[38,39] and known association with heart failure development[40]. Decreased cardiomyocyte proliferation and abnormal sarcomere morphology were observed in *HSPB7* knockout mice.

The *MTSS1* locus for CMR-derived LVESV and LVEF has previously been reported for comparable ECHO-derived LV traits[2,3]. The sentinel variant at the *MTSS1* locus is in the intergenic region. However, its close proxy (rs35006907, $r^2 > 0.98$) is located in the DNase I hypersensitive site in the cardiac tissue and is associated with expression of the *MTSS1* (Metastasis Suppressor Protein 1) gene in left ventricular and left atrial appendage tissues. This gene is involved in cytoskeletal signalling pathway[41] and encodes a scaffold protein that regulates actin dynamics[42].

The GWAS of remodelling phenotype, LVMVR, produced three genetic loci which were not shared with other LV traits. The *ALPK3* (Alpha Kinase 3) gene in the *ZNF592* locus is expressed in the developing heart and involves in cardiomyocyte differentiation[43]. The *ALPK3* knockout mouse model exhibited evidence of cardiac-specific phenotypic changes such as LV dilatation and hypertrophy. Another potential candidate gene at the same locus is *NMB* (Neuromedin B) gene which is associated with regulation of eating behaviour and obesity[44]. This gene has previously been highlighted as a candidate gene in a GWAS investigating the ECG indices of LV hypertrophy[45], in which the sentinel variant was moderately correlated (LD r^2 0.39) with our LVMVR sentinel variant. Lastly, the proxy variant (rs2070458) of the *DERL3* locus for LVMVR was in cis-eQTL with the *MMP11* (Matrix Metallopeptidase 11) gene expression in left ventricle. *MMP11* belongs to the family of proteolytic enzymes that regulate extracellular matrix and play a role in the development of myocardial fibrosis and ventricular remodelling[46,47].

Altogether, several key candidate genes in our GWAS loci (such as *TTN*, *BAG3*, *MTSS1*) which were shared across multiple structural and functional LV traits, encode essential proteins involved in the construction and maintenance of sarcomeric infrastructure. In contrast, the variation of LVMVR, the LV remodelling trait, was determined by the genetic loci containing candidate genes implicated in cardiomyocyte differentiation and extracellular matrix homeostasis.

Despite a limited number of loci found in this study, the polygenic risk scores (PRSs) constructed from LVEDV, LVESV, LVEF and LVMVR were predictive of heart failure events in the remainder of the UK Biobank cohort. This finding reinforces the fundamental role of the LV imaging endo-phenotypes in the pathogenesis of heart failure. The directionality of association between PRSs and heart failure event was generally concordant with prior expectations (alleles associated with larger LVEDV and LVESV and lower LVEF were predictive of increased risk of heart failure) except for the PRS derived from LVMVR where higher scores were correlated with lower odds of heart failure. LVMVR is a geometric phenotype where higher values are indicative of concentric remodelling or concentric hypertrophy of left ventricle. Despite the conventional wisdom of higher LVMVR being associated with an increased risk of adverse cardiovascular outcomes, a prospective longitudinal study in general population did not find a positive correlation between LVMVR and heart failure events[7]. Furthermore, previous studies have reported that LV concentricity indicated by an increased LVMVR did not commonly lead to impaired LVEF especially in the absence of interval myocardial infarction[48,49]. The significant negative association between LVMVR-PRS and heart failure in this study may reflect the possible dominance of dilatative (rather than hypertrophic) phenotype in the heart failure cases in our cohort.

Interestingly and contrary to the epidemiological evidence of association between LVM and incident heart failure[7,50], the LVM-PRS was not predictive of heart failure in our study. Although LVM appeared to be highly heritable in our genome-wide heritability analysis (h^2_g SNP = 34%), only a single locus was discovered at $p < 1x10^{-8}$ (percentage variance explained = 0.1%). Thus, the limited statistical power may have curtailed the predictive ability of LVM-PRS.

We acknowledge some limitations in our study. Despite being the largest GWAS of CMR image-derived LV phenotypes, the relatively small discovery sample size translated to the discovery of a few loci explaining < 0.5% of trait variance per trait. Additionally, the current sample size restricted our analysis to common variants with MAF ≥ 5%. However, the expected expansion of the CMR sample size to 100,000 in the UK Biobank, together with the highly optimised automatic image segmentation pipeline will lead to future studies with statistical power to detect more genetic loci at a lower MAF threshold. Furthermore, the upcoming exome sequencing data from the UK Biobank may allow us to investigate the role of rarer, protein truncating variants in a general population. Of note, three out of eight unique loci discovered in the UK Biobank were not replicated in MESA. Therefore, the evidence for these susceptibility loci should be considered preliminary. Second, although we have performed a limited look-up of our loci in MESA for additional support, our findings should be formally validated in a larger cohort.

In summary, the findings from this study not only enhance our understanding of the genetic basis of prognostically important LV phenotypes in the general population but also underscore the intricate genetic relationship between these endophenotypes and the

pathogenesis of heart failure syndrome, which may lead to potential novel therapeutic targets and personalised risk stratification strategy in the future.

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application 2964. The authors wish to thank all UK Biobank participants and staff.

Sources of funding

Dr Aung is supported by a Wellcome Trust Research Training Fellowship (203553/Z/16/Z). Drs Petersen, Neubauer, and Piechnik 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). This work was part of the portfolio of translational research of the National Institute for Health Research (NIHR) Biomedical Research Centre at Barts and The London School of Medicine and Dentistry; Drs Cabrera, Barnes, Warren, Munroe and Petersen acknowledge support from this centre. Dr Petersen also acknowledges support from the "SmartHeart" Engineering and Physical Sciences Research Council program grant (EP/P001009/1). Drs 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 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). Dr Fung is supported by The Medical College of Saint Bartholomew's Hospital Trust, an independent registered charity that promotes and advances medical and dental education and research at Barts and The London School of Medicine and Dentistry. The UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, 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 project 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 HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. 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.

Disclosures

S.E.P. provides consultancy to Circle Cardiovascular Imaging Inc, Calgary, Canada. Other authors have no conflicts of interest to declare.

Affiliations

William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK (N.A., H.R.W., K.F., P.B.M., S.E.P.)
National Institute for Health Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK (N.A., H.R.W., K.F., P.B.M., S.E.P.)

Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, UK (N.A., K.F., S.E.P.)

Medstar Heart and Vascular Institute, Medstar Georgetown University Hospital, Washington DC, USA (J.D.V.)

Center for Public Health Genomics, University of Virginia, Charlottesville, USA (C.Y., A.W.M.)

Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK (C.P.C., E.T., M.R.B.)

The Institute for Translational Genomics and Population Sciences, Division of Genomics Outcomes, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA (J.I.R., K.D.T.)

Division of Cardiology, Johns Hopkins University, Baltimore, Maryland, USA (J.AC.L.)

Department of Radiology, University of Wisconsin, Madison, Wisconsin, USA (D.A.B.)

Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK (S.K.P., S.N.)

References

1. Snipelisky D, Chaudhry S-P, Stewart GC. The Many Faces of Heart Failure. *Card Electrophysiol Clin*. 2019;11:11–20.

2. Wild PS, Felix JF, Schillert A, Teumer A, Chen M-H, Leening MJG, Völker U, Großmann V, Brody JA, Irvin MR, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J Clin Invest*. 2017;127:1798–1812.

3. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet*. 2018;50:390–400.

4. Grothues F, Smith GC, Moon JCC, Bellenger NG, Collins P, Klein HU, Pennell DJ. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. *Am J Cardiol*. 2002;90:29–34.

5. Levy D, Garrison RJ, Savage DD, Kannel WB, Castelli WP. Prognostic Implications of Echocardiographically Determined Left Ventricular Mass in the Framingham Heart Study. *N Engl J Med*. 1990;322:1561–1566.

6. Yeboah J, Bluemke DA, Hundley WG, Rodriguez CJ, Lima JAC, Herrington DM. Left ventricular dilation and incident congestive heart failure in asymptomatic adults without cardiovascular disease: multi-ethnic study of atherosclerosis (MESA). *J Card Fail*. 2014;20:905–911.

7. Bluemke DA, Kronmal RA, Lima JAC, Liu K, Olson J, Burke GL, Folsom AR. The Relationship of Left Ventricular Mass and Geometry to Incident Cardiovascular Events: The MESA Study. *J Am Coll Cardiol*. 2008;52:2148–2155.

8. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, 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 Med*. 2015;12:e1001779.

9. Petersen SE, Aung N, Sanghvi MM, Zemrak F, Fung K, Paiva JM, Francis JM, Khanji MY, Lukaschuk E, Lee AM, et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J Cardiovasc Magn Reson*. 2017;19:18.

10. Loh P-R, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ, Schizophrenia Working Group of the Psychiatric Genomics Consortium, de Candia TR, Lee SH, Wray NR, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet*. 2015;47:1385–1392.

11. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284–290.

12. Khouri Michel G., Peshock Ronald M., Ayers Colby R., de Lemos James A., Drazner Mark H. A 4-Tiered Classification of Left Ventricular Hypertrophy Based on Left Ventricular Geometry. *Circ Cardiovasc Imaging*. 2010;3:164–171.

13. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, JacobsJr. DR, Kronmal R, Liu K, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *Am J Epidemiol*. 2002;156:871–881.

14. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, et al. PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics*. 2016;32:3207–3209.

15. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50:1593.

16. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015;97:576–592.

17. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*. 2009;25:i54–i62.

18. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790–1797.

19. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585.

20. Fehrmann RSN, Karjalainen JM, Krajewska M, Westra H-J, Maloney D, Simeonov A, Pers TH, Hirschhorn JN, Jansen RC, Schultes EA, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet*. 2015;47:115–125.

21. Chien K-L, Hsu H-C, Su T-C, Chen M-F, Lee Y-T. Heritability and major gene effects on left ventricular mass in the Chinese population: a family study. *BMC Cardiovasc Disord*. 2006;6:37.

22. Busjahn CA, Schulz-Menger J, Abdel-Aty H, Rudolph A, Jordan J, Luft FC, Busjahn A. Heritability of left ventricular and papillary muscle heart size: a twin study with cardiac magnetic resonance imaging. *Eur Heart J*. 2009;30:1643–1647.

23. Boissier F, Razazi K, Seemann A, Bedet A, Thille AW, de Prost N, Lim P, Brun-Buisson C, Mekontso Dessap A. Left ventricular systolic dysfunction during septic shock: the role of loading conditions. *Intensive Care Med*. 2017;43:633–642.

24. Eppinga RN, Hagemeijer Y, Burgess S, Hinds DA, Stefansson K, Gudbjartsson DF, van Veldhuisen DJ, Munroe PB, Verweij N, van der Harst P. Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nat Genet*. 2016;48:1557–1563.

25. Granzier Henk L., Labeit Siegfried. The Giant Protein Titin. *Circ Res*. 2004;94:284–295.

26. Tayal U, Newsome S, Buchan R, Whiffin N, Halliday B, Lota A, Roberts A, Baksi AJ, Voges I, Midwinter W, et al. Phenotype and Clinical Outcomes of Titin Cardiomyopathy. *J Am Coll Cardiol*. 2017;70:2264–2274.

27. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, Buchan RJ, Walsh R, John S, Wilkinson S, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med*. 2015;7:270ra6.

28. Schafer S, de Marvao A, Adami E, Fiedler LR, Ng B, Khin E, Rackham OJL, van Heesch S, Pua CJ, Kui M, et al. Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet*. 2017;49:46–53.

29. Knezevic T, Myers VD, Gordon J, Tilley DG, Sharp TE, Wang J, Khalili K, Cheung JY, Feldman AM. BAG3: a new player in the heart failure paradigm. *Heart Fail Rev*. 2015;20:423–434.

30. Ulbricht A, Höhfeld J. Tension-induced autophagy. *Autophagy*. 2013;9:920–922.

31. Feldman AM, Gordon J, Wang J, Song J, Zhang X-Q, Myers VD, Tilley DG, Gao E, Hoffman NE, Tomar D, et al. BAG3 regulates contractility and Ca2+ homeostasis in adult mouse ventricular myocytes. *J Mol Cell Cardiol*. 2016;92:10–20.

32. Domínguez F, Cuenca S, Bilińska Z, Toro R, Villard E, Barriales-Villa R, Ochoa JP, Asselbergs F, Sammani A, Franaszczyk M, et al. Dilated Cardiomyopathy Due to BLC2-Associated Athanogene 3 (BAG3) Mutations. *J Am Coll Cardiol*. 2018;72:2471–2481.

33. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, et al. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J*. 2011;32:1065–1076.

34. Burkhalter MD, Fralish GB, Premont RT, Caron MG, Philipp M. Grk5l Controls Heart Development by Limiting mTOR Signaling During Symmetry Breaking. *Cell Rep*. 2013;4:625–632.

35. Philipp M, Berger IM, Just S, Caron MG. Overlapping and Opposing Functions of G Protein-coupled Receptor Kinase 2 (GRK2) and GRK5 during Heart Development. *J Biol Chem*. 2014;289:26119–26130.

36. Traynham CJ, Hullmann J, Koch WJ. Canonical and Non-Canonical Actions of GRK5 in the Heart. *J Mol Cell Cardiol*. 2016;92:196–202.

37. Liggett SB, Cresci S, Kelly RJ, Syed FM, Matkovich SJ, Hahn HS, Diwan A, Martini JS, Sparks L, Parekh RR, et al. A GRK5 Polymorphism that Inhibits β-Adrenergic Receptor Signaling is Protective in Heart Failure. *Nat Med*. 2008;14:510–517.

38. Rosenfeld GE, Mercer EJ, Mason CE, Evans T. Small heat shock proteins Hspb7 and Hspb12 regulate early steps of cardiac morphogenesis. *Dev Biol*. 2013;381:389–400.

39. Wu T, Mu Y, Bogomolovas J, Fang X, Veevers J, Nowak RB, Pappas CT, Gregorio CC, Evans SM, Fowler VM, et al. HSPB7 is indispensable for heart development by modulating actin filament assembly. *Proc Natl Acad Sci U S A*. 2017;114:11956–11961.

40. Garnier S, Hengstenberg C, Lamblin N, Dubourg O, De Groote P, Fauchier L, Trochu J-N, Arbustini E, Esslinger U, Barton PJ, et al. Involvement of BAG3 and HSPB7 loci in various etiologies of systolic heart failure: Results of a European collaboration assembling more than 2000 patients. *Int J Cardiol*. 2015;189:105–107.

41. Mattila PK, Salminen M, Yamashiro T, Lappalainen P. Mouse MIM, a Tissue-specific Regulator of Cytoskeletal Dynamics, Interacts with ATP-Actin Monomers through Its C-terminal WH2 Domain. *J Biol Chem*. 2003;278:8452–8459.

42. Xie F, Ye L, Ta M, Zhang L, Jiang WG. MTSS1: a multifunctional protein and its role in cancer invasion and metastasis. *Front Biosci Sch Ed*. 2011;3:621–631.

43. Hosoda T, Monzen K, Hiroi Y, Oka T, Takimoto E, Yazaki Y, Nagai R, Komuro I. A Novel Myocyte-specific Gene MidoriPromotes the Differentiation of P19CL6 Cells into Cardiomyocytes. *J Biol Chem*. 2001;276:35978–35989.

44. Spálová J, Zamrazilová H, Vcelák J, Vanková M, Lukásová P, Hill M, Hlavatá K, Srámková P, Fried M, Aldhoon B, et al. Neuromedin beta: P73T polymorphism in overweight and obese subjects. *Physiol Res*. 2008;57 Suppl 1:S39-48.

45. Shah S, Nelson CP, Gaunt TR, Harst P van der, Barnes T, Braund PS, Lawlor DA, Casas J-P, Padmanabhan S, Drenos F, et al. Four Genetic Loci Influencing Electrocardiographic Indices of Left Ventricular HypertrophyClinical Perspective. *Circ Cardiovasc Genet*. 2011;4:626–635.

46. Li YY, McTiernan CF, Feldman AM. Interplay of matrix metalloproteinases, tissue inhibitors of metalloproteinases and their regulators in cardiac matrix remodeling. *Cardiovasc Res*. 2000;46:214–224.

47. Creemers Esther E.J.M., Cleutjens Jack P.M., Smits Jos F.M., Daemen Mat J.A.P. Matrix Metalloproteinase Inhibition After Myocardial Infarction. *Circ Res*. 2001;89:201–210.

48. Drazner MH, Rame JE, Marino EK, Gottdiener JS, Kitzman DW, Gardin JM, Manolio TA, Dries DL, Siscovick DS. Increased left ventricular mass is a risk factor for the development of a depressed left ventricular ejection fraction within five years: The Cardiovascular Health Study. *J Am Coll Cardiol*. 2004;43:2207–2215.

49. Milani RV, Drazner MH, Lavie CJ, Morin DP, Ventura HO. Progression from Concentric Left Ventricular Hypertrophy and Normal Ejection Fraction to Left Ventricular Dysfunction. *Am J Cardiol*. 2011;108:992–996.

50. de Simone G, Gottdiener JS, Chinali M, Maurer MS. Left ventricular mass predicts heart failure not related to previous myocardial infarction: the Cardiovascular Health Study. *Eur Heart J*. 2008;29:741–747.

Figure titles and legends

Figure 1. Flowchart of analysis strategy

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVSV, left ventricular stroke volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction; SBP, systolic blood pressure; REML, restricted maximal likelihood; SNV, single nucleotide variant; MAF, minor allele frequency; INFO, imputation quality score; LD, linkage disequilibrium; GCTA, genome-wide complex trait analysis; MESA, multi-ethnic study of atherosclerosis

Figure 2. SNP-heritability and genotypic and phenotypic correlations between LV traits

The upper triangle of correlogram represents the degree of genotypic correlation and the lower triangle represents the degree of phenotypic correlation. Heritability estimated from genotyped SNPs is presented on the right-hand side of the figure. * $p < 0.0001$; $^{\#}$ $p > 0.05$; All other correlation estimates had $p < 1x10^{-16}$

SNP, single nucleotide polymorphism; LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVSV, left ventricular stroke volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction; SD, standard deviation

Figure 3. Genomic loci associated with CMR-LV phenotypes

Circular Manhattan plot depicting the GWAS results of all LV traits. The red line indicates the genome-wide significant threshold at $p < 1x10^{-8}$. No genome-wide significant locus was found for LVSV. The significant genomic loci are denoted by the red dots.

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVSV, left ventricular stroke volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction

Figure 4. Venn diagram of LV loci

The locus name indicates the nearest annotated gene.

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction

Figure 5. Summary of genes associated with LV traits

Genes are ranked on the basis of the supporting evidence summarised in Supplemental Table 11 based on the presence of nonsynonymous variant, gene expression data, chromatin interaction analyses, literature review and knockout models.

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVEF, left ventricular ejection fraction; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ration

* coding variant gene, † knockout phenotype, ‡ Previously reported cardiovascular biology or strong functional rationale, § expression quantitative trait locus (eQTL) gene, ‖ Hi-C long-range interaction gene

The illustration used elements with permission from Servier Medical Art.

Table 1. Genomic loci identified for CMR-derived LV phenotypes

CMR trait	Locus Name	Lead variant	CHR	Position (hg19)	EA	NEA	EAF	BETA INT	SE INT	P INT	BETA raw	SE raw
LVEDV	***TTN***	**rs2042995†**	**2**	**179558366**	**T**	**C**	**0.779**	**0.087**	**0.013**	**2.3E-11**	**1.956**	**0.295**
LVEDV	***BAG3***	**rs7071853**	**10**	**121311606**	**T**	**C**	**0.767**	**-0.075**	**0.013**	**3.9E-09**	**-1.672**	**0.289**
LVEDV	*SH2B3*	rs7310615†	12	111865049	C	G	0.481	-0.066	0.011	1.4E-09	-1.409	0.247
LVESV	***TTN***	**rs2042995†**	**2**	**179558366**	**T**	**C**	**0.779**	**0.118**	**0.013**	**8.4E-20**	**1.431**	**0.161**
LVESV	*MTSS1*	rs200712209*	8	125858538	GA	G	0.687	0.079	0.012	1.7E-11	0.953	0.145
LVESV	***BAG3***	**rs72840788†**	**10**	**121415685**	**G**	**A**	**0.778**	**0.109**	**0.013**	**5.6E-17**	**1.352**	**0.162**
LVEF	***CLCNKA***	**rs945425**	**1**	**16348412**	**T**	**C**	**0.326**	**0.075**	**0.011**	**8.6E-11**	**0.368**	**0.057**
LVEF	***TTN***	**rs2042995†**	**2**	**179558366**	**T**	**C**	**0.779**	**-0.091**	**0.013**	**2.5E-12**	**-0.446**	**0.064**
LVEF	*MTSS1*	rs34866937*	8	125859850	G	A	0.686	-0.076	0.012	6.8E-11	-0.379	0.058
LVEF	***BAG3***	**rs72840788***	**10**	**121415685**	**G**	**A**	**0.778**	**-0.104**	**0.013**	**3.4E-15**	**-0.507**	**0.065**
LVM	***TTN***	**rs2255167†**	**2**	**179558282**	**T**	**A**	**0.812**	**0.103**	**0.014**	**8.3E-14**	**1.246**	**0.168**
LVMVR	***CDKN1A***	**rs146170154**	**6**	**36646768**	**C**	**CTA**	**0.802**	**-0.092**	**0.014**	**2.6E-11**	**-0.007**	**0.001**
LVMVR	***ZNF592***	**rs149369954†**	**15**	**85348961**	**TTTTG**	**T**	**0.745**	**0.085**	**0.013**	**1.9E-11**	**0.006**	**0.001**
LVMVR	***DERL3***	**rs6003909**	**22**	**24181652**	**A**	**G**	**0.172**	**0.111**	**0.014**	**9.7E-15**	**0.008**	**0.001**

The novel loci are highlighted in bold face. The locus name indicates the nearest annotated gene. The p values were calculated from the BOLT-LMM χ^2 tests statistics obtained from the linear mixed-effects models adjusted for age, sex, height, weight, systolic blood pressure corrected for anti-hypertensive medication use, phenotype-derivation method (automatic/manual), array type (UK Biobank vs UK BiLEVE array), and imaging centre. The sample size for each phenotype are: LVEDV (n = 16,920), LVESV (n = 16,920), LVSV (n = 16,917), LVEF (n = 16,923), LVM (n = 16,920) and LVMVR (n = 16,884).

CMR, cardiovascular magnetic resonance; LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVEF, left ventricular ejection fraction; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; CHR, chromosome; EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency in UKBB cohort; SE, standard error; INT, rank-based inverse normal transformation

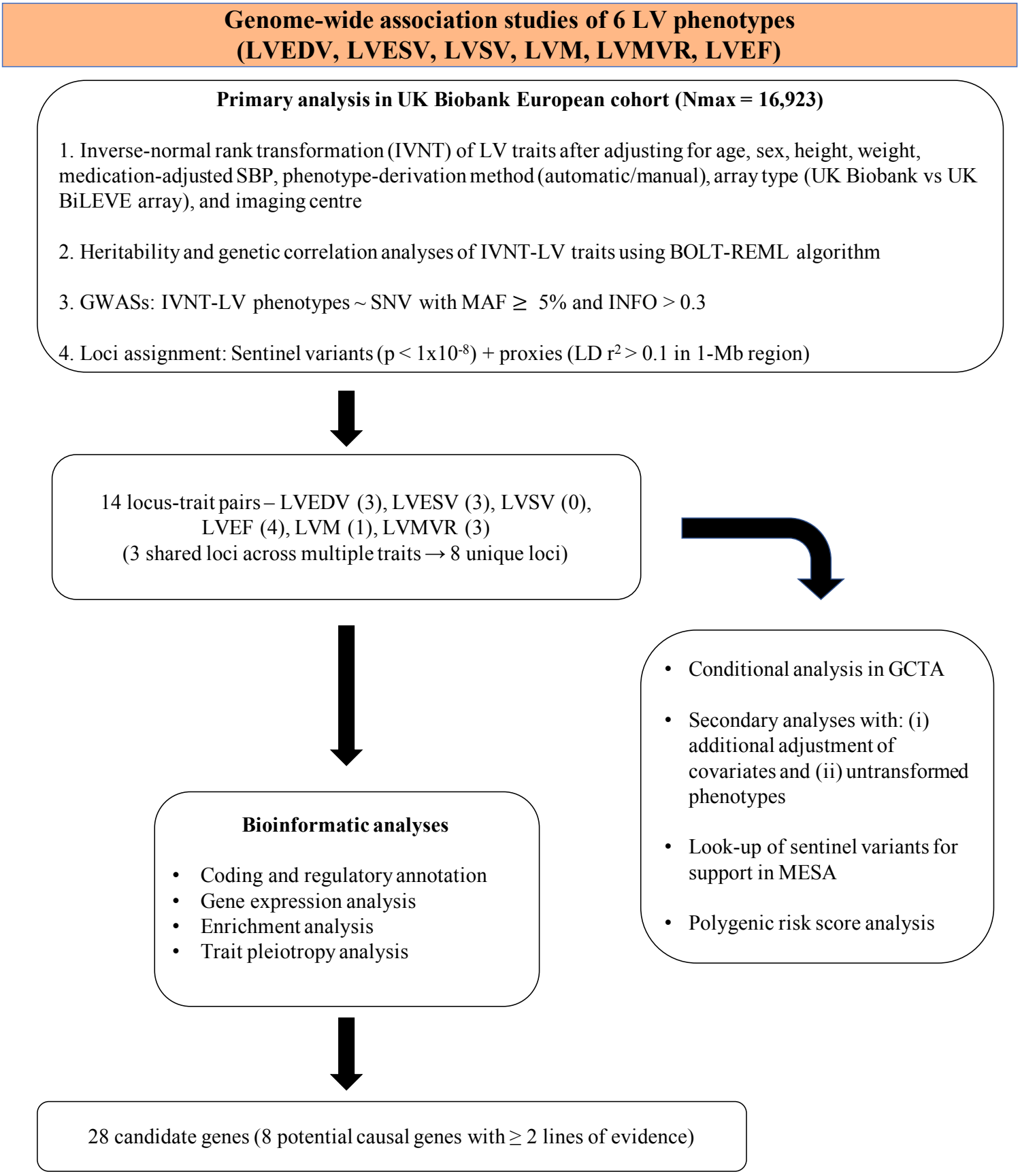
* Support in MESA cohort at $p < 0.0036$ [Bonferroni-correction=0.05/14]; † Nominal Support in MESA cohort at $p < 0.05$ with concordant direction of effect

Table 2. Polygenic risk prediction of heart failure using PRSs constructed from variants associated with LV traits

Phenotype (mean ± SD)	PRS quintiles	OR (95% CI)	P
LVEDV (126 ± 23 ml)	Q1 (Reference)	1	–
LVEDV (136 ± 24 ml)	Q2	1.07 (0.95-1.20)	2.74E-01
LVEDV (144 ± 26 ml)	Q3	1.16 (1.04-1.31)	1.05E-02
LVEDV (153 ± 27 ml)	Q4	1.19 (1.06-1.34)	3.53E-03
LVEDV (177 ± 35 ml)	Q5	1.25 (1.11-1.40)	1.49E-04
LVESV (47 ± 11 ml)	Q1 (Reference)	1	–
LVESV (53 ± 12 ml)	Q2	1.13 (1.004-1.28)	4.32E-02
LVESV (57 ± 13 ml)	Q3	1.16 (1.03-1.31)	1.50E-02
LVESV (63 ± 14 ml)	Q4	1.21 (1.08-1.36)	1.60E-03
LVESV (75 ± 17 ml)	Q5	1.41 (1.26-1.58)	4.76E-09
LVEF (56 ± 4 %)	Q1 (Reference)	1	–
LVEF (58 ± 4 %)	Q2	0.88 (0.78-0.98)	1.95E-02
LVEF (60 ± 4 %)	Q3	0.84 (0.75-0.94)	1.82E-03
LVEF (62 ± 4 %)	Q4	0.85 (0.76-0.95)	3.17E-03
LVEF (65 ± 4 %)	Q5	0.77 (0.69-0.86)	6.64E-06
LVM (78 ± 17 g)	Q1 (Reference)	1	–
LVM (80 ± 19 g)	Q2	0.96 (0.86-1.08)	5.19E-01
LVM (83 ± 20 g)	Q3	0.96 (0.86-1.08)	5.23E-01
LVM (87 ± 21 g)	Q4	0.90 (0.80-1.01)	7.37E-02
LVM (99 ± 26 g)	Q5	1.06 (0.94-1.18)	3.48E-01
LVMVR (0.51 ± 0.06 g/ml)	Q1 (Reference)	1	–
LVMVR (0.55 ± 0.06 g/ml)	Q2	1.05 (0.94-1.17)	3.93E-01
LVMVR (0.58 ± 0.06 g/ml)	Q3	0.93 (0.83-1.04)	2.14E-01
LVMVR (0.61 ± 0.07 g/ml)	Q4	0.90 (0.80-1.01)	6.23E-02
LVMVR (0.67 ± 0.09 g/ml)	Q5	0.80 (0.71-0.90)	2.66E-04

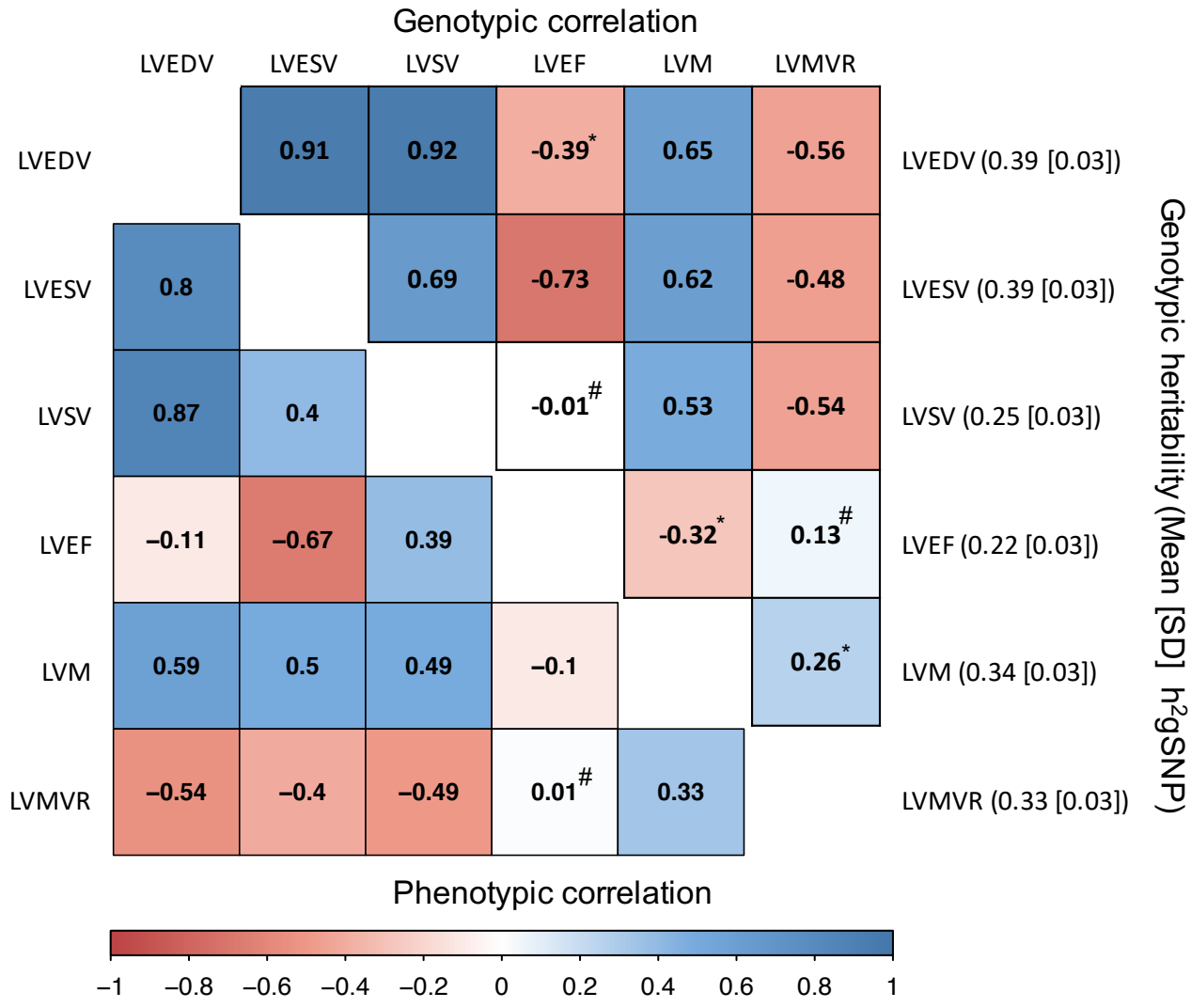
The association between the polygenic risk scores calculated from the summary statistics of LV GWASs and heart failure development was estimated by logistic regression in the hold-out test sample of UK Biobank cohort (Total N = 227,240 [3,106 cases, 224,134 controls]). The PRS models were adjusted for age, sex, BMI, SBP and DBP corrected for anti-hypertensive medication use, smoking status, regular alcohol use, dyslipidaemia, diabetes and 15 PCs.
The mean± SD of each LV phenotype for each PRS quintile is presented in the parenthesis.
LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVEF, left ventricular ejection fraction; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; PRS, polygenic risk score; SD, standard deviation; OR, odds ratio; CI, confidence interval; PCs, genetic principal components

Figure 1. Flowchart of analysis strategy



LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVSV, left ventricular stroke volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction; SBP, systolic blood pressure; REML, restricted maximal likelihood; SNV, single nucleotide variant; MAF, minor allelic frequency; INFO, imputation quality score; LD, linkage disequilibrium; GCTA, genome-wide complex trait analysis; MESA, multi-ethnic study of atherosclerosis

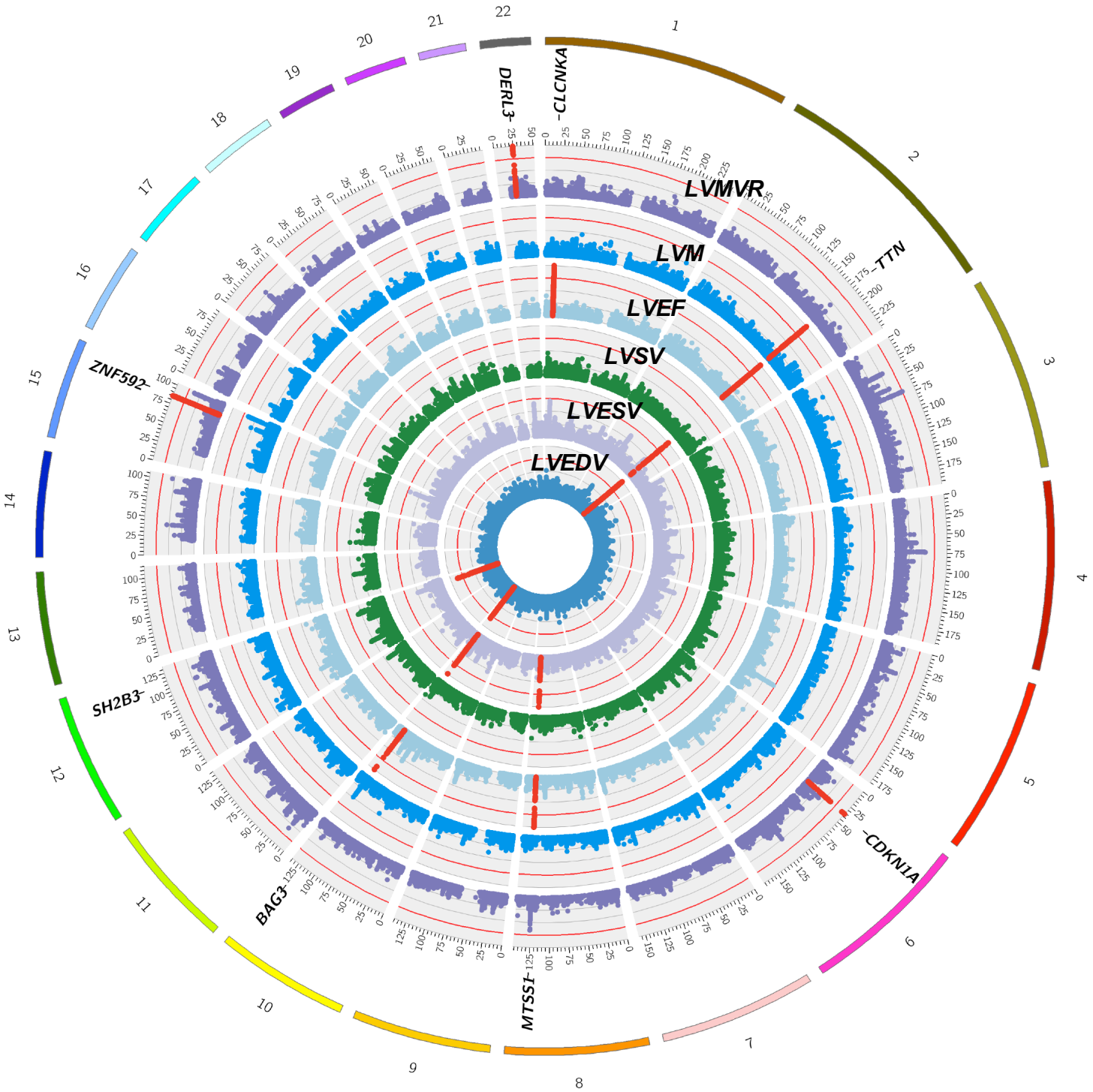
Figure 2. SNP-heritability and genotypic and phenotypic correlations between LV traits



The upper triangle of correlogram represents the degree of genotypic correlation and the lower triangle represents the degree of phenotypic correlation. Heritability estimated from genotyped SNPs is presented on the right-hand side of the figure. * $p < 0.0001$; # $p > 0.05$; All other correlation estimates had $p < 1x10^{-16}$

SNP, single nucleotide polymorphism; LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVSV, left ventricular stroke volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction; SD, standard deviation

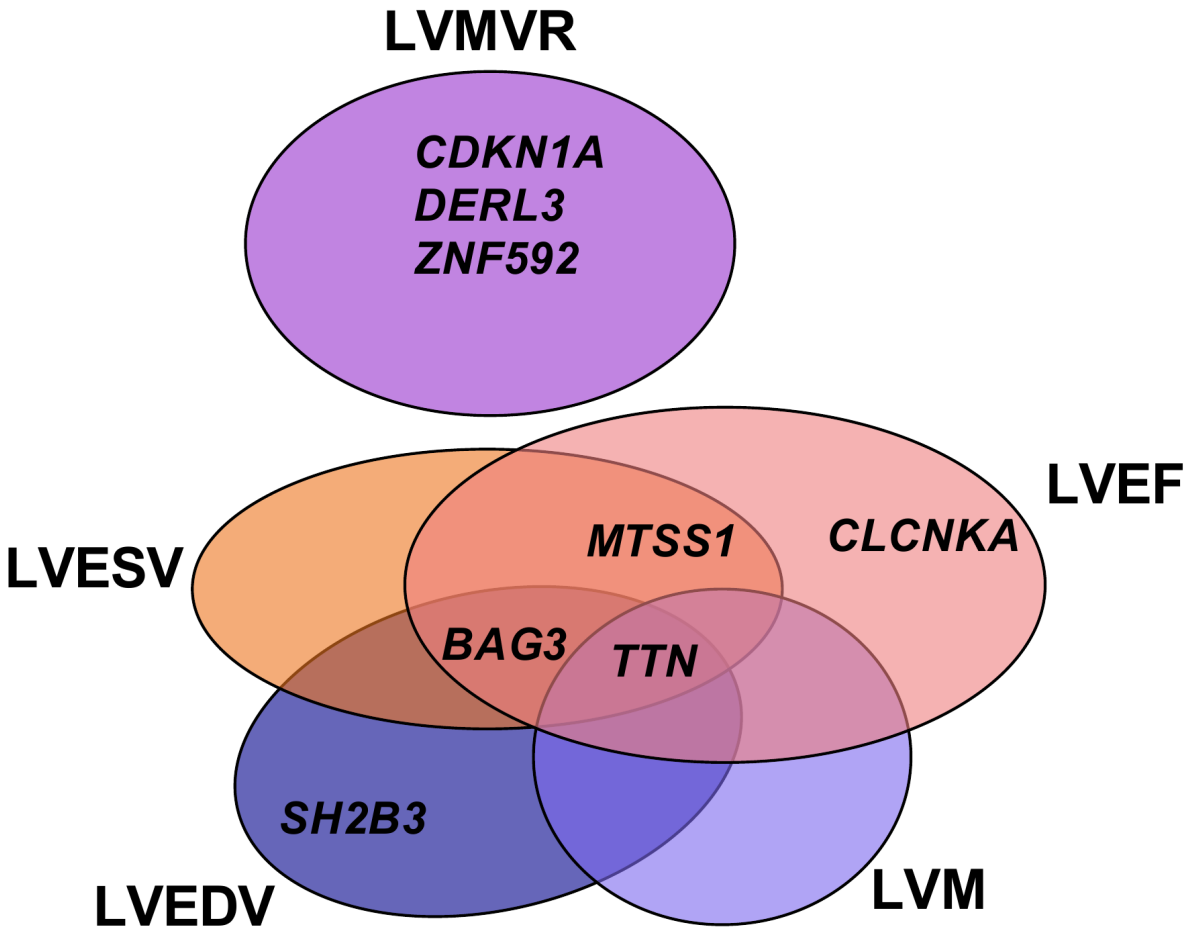
Figure 3. Genomic loci associated with CMR-LV phenotypes



Circular Manhattan plot depicting the GWAS results of all LV traits. The red line indicates the genome-wide significant threshold at $p < 1x10^{-8}$. No genome-wide significant locus was found for LVSV. The significant genomic loci are denoted by the red dots.

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVSV, left ventricular stroke volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction

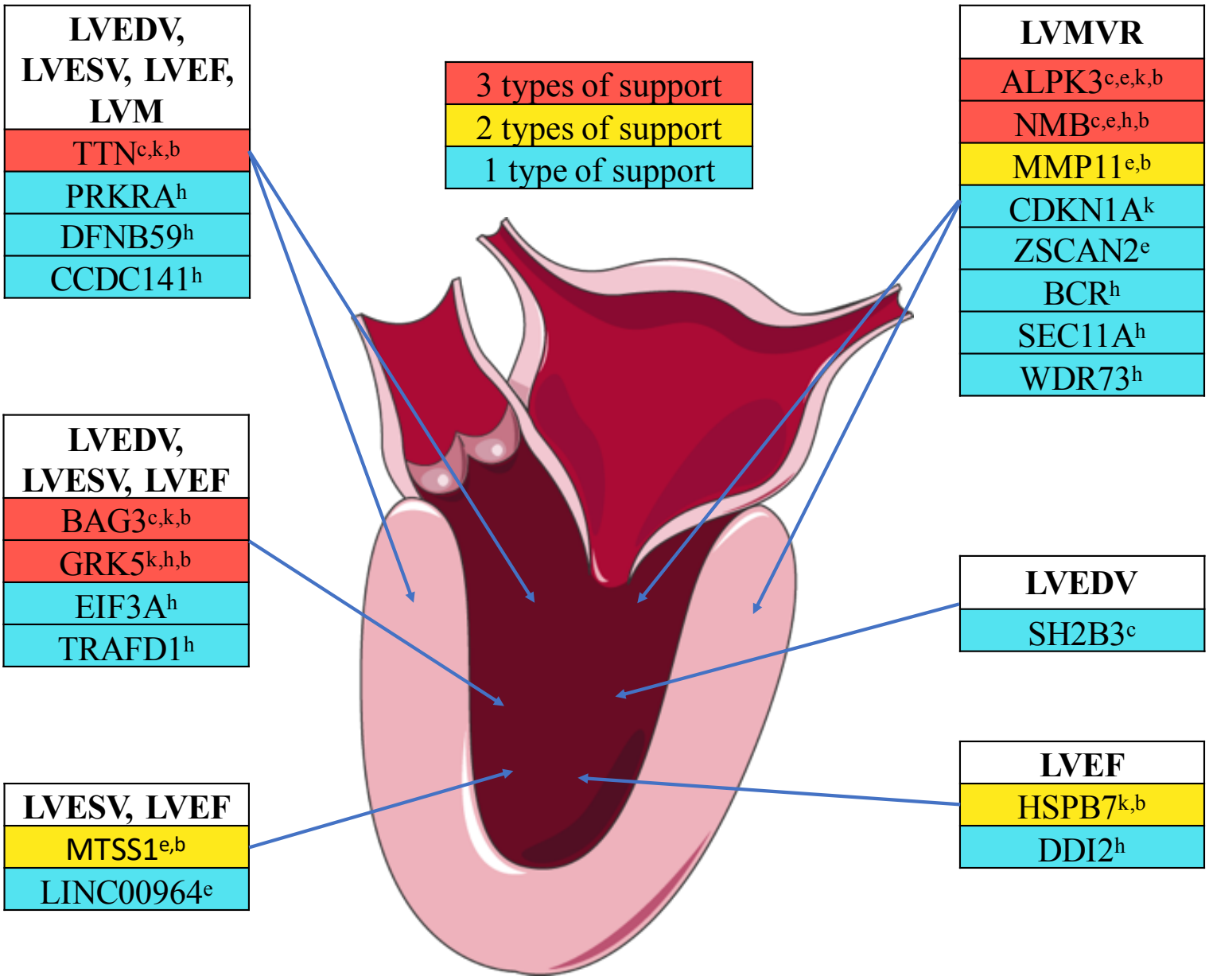
Figure 4. Venn diagram of LV loci



The locus name indicates the nearest annotated gene.

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ratio; LVEF, left ventricular ejection fraction

Figure 5. Summary of genes associated with LV traits



Genes are ranked on the basis of the supporting evidence summarised in Supplemental Table 11 based on the presence of nonsynonymous variant, gene expression data, chromatin interaction analyses, literature review and knockout models.
LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LVEF, left ventricular ejection fraction; LVM, left ventricular mass; LVMVR, left ventricular mass to end-diastolic volume ration
[b] Previously reported cardiovascular biology or strong functional rationale, [c] coding variant gene, [e] expression quantitative trait locus (eQTL) gene, [h] Hi-C long-range interaction gene, [k] knockout phenotype
The illustration used elements with permission from Servier Medical Art.