Published in final edited form as:
Nat Cardiovasc Res. 2022 April ; 1(4): 361–371. doi:10.1038/s44161-022-00048-2.

Genetic and environmental determinants of diastolic heart function

Marjola Thanaj[1,+], **Johanna Mielke**[3,+], **Kathryn A. McGurk**[1,4], **Wenjia Bai**[2,7], **Nicolò Savioli**[1,2], **Antonio de Marvao**[1], **Hannah V. Meyer**[5], **Lingyao Zeng**[3], **Florian Sohler**[3], **R. Thomas Lumbers**[8], **Martin R. Wilkins**[4], **James S. Ware**[1,4], **Christian Bender**[3], **Daniel Rueckert**[2,6], **Aidan MacNamara**[3], **Daniel F. Freitag**[3,*], **Declan P. O'Regan**[1,*,C]

[1]MRC London Institute of Medical Sciences, Imperial College London, London, UK

[2]Department of Computing, Imperial College London, London, UK

[3]Bayer AG, Research & Development, Pharmaceuticals, Wuppertal, Germany

[4]National Heart and Lung Institute, Imperial College London, London, UK

[5]Cold Spring Harbor Laboratory, Simons Center for Quantitative Biology, USA

[6]Institute for Artificial Intelligence and Informatics, Klinikum rechts der Isar, Technical University of Munich, Germany

[7]Department of Brain Sciences, Imperial College London

[8]Institute of Health Informatics, University College London, UK

Abstract

Diastole is the sequence of physiological events that occur in the heart during ventricular filling and principally depends on myocardial relaxation and chamber stiffness. Abnormal diastolic function is related to many cardiovascular disease processes and is predictive of health outcomes, but its genetic architecture is largely unknown. Here, we use machine learning cardiac motion analysis to measure diastolic functional traits in 39,559 participants of the UK Biobank and perform a genome-wide association study. We identified 9 significant, independent loci near genes that are associated with maintaining sarcomeric function under biomechanical stress and genes implicated in the development of cardiomyopathy. Age, sex and diabetes were independent predictors of diastolic function and we found a causal relationship between genetically-determined ventricular stiffness and incident heart failure. Our results provide insights into the genetic

[C]Corresponding author: declan.oregan@imperial.ac.uk.
[+]These authors contributed equally to this work
[*]These authors jointly supervised this work

Author contributions

M.T. and J.M. performed the formal analyses and co-wrote the paper; W.B, N.S., A.de M. and D.R. performed the image analysis; K.A.McG., J.S.W., H.V.M., L.Z., F.S., R.T.L. and C.B. performed or interpreted the genetic analyses; A.MacN. performed the GWAS; M.R.W. interpreted the pharmacology findings; D.F.F. and D.P.O'R conceived the study, managed the project and revised the manuscript. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

and environmental factors influencing diastolic function that are relevant for identifying causal relationships and potential tractable targets.

Diastole is not a passive phase of the cardiac cycle, but is a complex sequence of inter-related physiological processes dependent on myocardial relaxation, stiffness and recoil, that are modulated by loading conditions, heart rate, and contractile function. Diastolic function therefore plays a central role in determining left ventricular filling and stroke volume with dysfunction shown to be a predictor of major adverse cardiovascular events and all cause mortality[1]. Decline in diastolic function is also a hallmark of cardiac ageing which occurs through multiple pro-fibrotic and energetic pathways[2, 3]. While several candidate genes have been implicated in various systolic function phenotypes through genome wide association studies (GWAS)[4, 5], the genetic architecture of diastolic function and causal associations with disease are largely unknown. Efforts to better define the molecular mechanisms of diastolic dysfunction could enable the development of innovative therapies for many cardiovascular disease states.

Pre-clinical models of diastolic dysfunction are associated with alterations in left ventricular stiffness on atomic force microscopy that occur at the level of the cardiomyocyte sarcomere as well as due to extracellular matrix protein expansion[6]. Such tissue level changes can be assessed at macroscopic scale in human populations through analysis of diastolic mechanics. Here we use data from participants in UK Biobank with cardiac magnetic resonance imaging (CMR)[7], and apply deep learning computer vision techniques for precision motion analysis to derive image-based phenotypes of diastolic function[8, 9]. In a GWAS of diastolic traits we identify associated loci that map to genes involved in actin assembly, cardiac myocyte survival, and heart failure phenotypes. We also describe the relationship between diastolic function and cardiovascular risk factors, and identify potential causal relationships with disease through Mendelian randomisation.

Results

Study Overview

We analysed CMR data from 39,559 participants in UK Biobank using machine learning segmentation and motion tracking to measure three validated parameters of diastolic function - radial and longitudinal peak early diastolic strain rate ($PDSR_{rr}$ and $PDSR_{ll}$) (Fig. 1), and maximum body surface area-indexed left atrial volume (LAV_{max_i})[10]. A flow chart of the analysis steps is depicted in Extended Data Fig 1. Baseline characteristics of the population are shown in Extended Data Table 1. For the GWAS the population was partitioned into discovery and validation sets by the release of data tranches by UK Biobank. To assess the association between these diastolic function traits and other clinical measurements, we further considered a broad selection of 30 imaging and 110 non-imaging phenotypes that included biophysical data and circulating biomarkers (Supplementary Data 1). Independent GWASs were undertaken for each image-derived phenotype and heritability estimated. We used a phenome-wide association study (PheWAS) to identify multiple phenotypes associated with a polygenic instrumental variable score (PIVS) for diastolic function. Potentially causal associations were examined using 2-sample Mendelian

Europe PMC Funders Author Manuscripts

randomisation (MR). The results are reported in accordance with GWAS reporting guidelines and a checklist is provided as Supplementary Information.

Imaging and non-imaging phenotype associations

Strain rates declined with age and were lower in men ($P < 10^{-16}$ for both associations) (Fig. 2), but no univariable association was observed between age and LAV_{max_i} (Extended Data Fig 2). Multiple linear regression analysis was used to develop a model for predicting each diastolic trait from demographic, haemodynamic and cardiovascular risk factors (Fig. 3a, Extended Data Fig 3a). In this multivariable analysis strain rate and left atrial volumes were negatively associated with age, male sex and pulse rate in the full model ($P < 10^{-16}$ for all associations). Significant associations were also observed for body surface area (BSA) and systolic blood pressure (SBP). Diabetes also added significantly to the associations with the diastolic function traits in the model ($PDSR_{ll}$: $P = 2.36 \times 10^{-8}$; $PDSR_{rr}$: $P = 9.98 \times 10^{-6}$; LAV_{max_i}: $P = 1.04 \times 10^{-3}$).

We investigated the association between image-derived measures of atrial, ventricular and aortic function with a broader range of non-imaging phenotypes using regularized regression analysis (Extended Data Fig 4, Fig. 3b, Extended Data Fig 3b) (see Supplementary Material for further information).

C-reactive protein (CRP), a circulating biomarker of inflammation, showed a positive relationship with serum triglycerides, but we found no circulating biomarkers independently associated with diastolic function. We found that reduced peak diastolic strain rates were associated with reduced LAV_{max_i}. Left atrial function was related to indicators of right ventricular function emphasising their functional interdependence[11].

Genetic architecture of diastolic function traits

Genome-wide common and rare variant association analyses of diastolic function traits—The SNP-based heritability (i.e. the proportion of variance per trait explained by all considered SNPs) was 12% for $PDSR_{ll}$, 13% for $PDSR_{rr}$, and 21% for LAV_{max_i}. The observed genetic correlation between the diastolic function traits was 0.22 (SE 0.07) between $PDSR_{ll}$ and LAV_{max_i}, 0.12 (SE 0.08) between $PDSR_{rr}$ and LAV_{max_i}, and 0.85 (SE 0.04) between $PDSR_{ll}$ and $PDSR_{rr}$.

In total, we identified 9 independent loci from our GWAS analyses, 5 loci for $PDSR_{rr}$, 4 for $PDSR_{ll}$, and 2 for LAV_{max_i} (2 loci are shared between $PDSR_{rr}$ and $PDSR_{ll}$). Within the discovery set, we identified 5 independent loci (LAV_{max_i}, 1; $PDSR_{rr}$, 3; $PDSR_{ll}$, 1) reaching genome-wide significance ($P = 5 \cdot 10^{-8}$; Supplementary Fig 3), which were also significant in the validation dataset also ($P < 0.05/5$). Considering the full dataset, the number of significant independent loci increased to 9 with 2 additional loci associating with $PDSR_{rr}$, 1 additional with LAV_{max_i}, and 1 additional with $PDSR_{ll}$ (Fig. 4).

Variant annotation—Summary information for the 9 loci identified using the full GWAS dataset and two predicted loss of function variants are presented in Table 1 (further information can be found in Supplementary Material, Supplementary Fig 5 and Supplementary Table 1). The closest gene to each locus is depicted, with further variant to

gene mapping presented as the "likely gene" given by evidence of a functional effect on a gene (details in Supplementary Material), additional heart-related phenotype associations, or a previously reported mechanism linking the gene to diastolic function. Taking lead variants identified from GWAS and the loss of function analysis were able to highlight several structural genes associated with diastolic function that also have a known role in myocardial contractility (e.g. *TTN*, *PLN*, *GJA1*), and in the functional maintenance and stress-response of the cytoskeleton (e.g. *FHOD3*, *BAG3*)[12]. Moreover, we were also able to identify a novel link between the *NPR3* locus and left atrial volume. The signal co-localizes with a previously discovered association with blood pressure traits (systolic, diastolic and mean arterial blood pressure). The C-allele of the lead SNP (rs1173727) at this locus increases *NPR3* expression, and is associated with increased blood pressure and LAV_{max_i}, and an increase in risk of heart failure (Supplementary Material). The *NPR3* gene encodes the C-type natriuretic peptide receptor, which has a high drug tractability score (https://platform.opentargets.org/target/ENSG00000113389), making it a potential therapeutic target.

The relationship between common variants in *NPR3* and genes encoding other proteins in the natriuretic peptide pathway with traits linked to the lead SNP (rs1173727) are shown in Supplementary Fig 6, and an abridged version is provided in Extended Data Fig 5.

Potential causes and consequences of diastolic function

Creation of polygenic instrumental variable scores (PIVS and PheWAS)—

PIVSs for each diastolic function trait consisted of 20 SNPs for $PDSR_{rr}$, 15 SNPs for $PDSR_{ll}$, and 8 for LAV_{max_i}. The PIVS explained 1.5 % of the variability of $PDSR_{rr}$, 1.1 % of $PDSR_{ll}$ and 0.2 % of LAV_{max_i}. There was good agreement between the distribution of the PIVS in the UK Biobank participants with and without CMR indicating no systematic bias in genetic architecture (Supplementary Fig 9). The Pearson correlation coefficient for the PIVS for $PDSR_{ll}$ and $PDSR_{rr}$ was 0.35 whereas the correlation coefficient between LAV_{max_i} and $PDSR_{ll}$ or $PDSR_{rr}$, respectively, was much lower (<0.01). PheWAS was undertaken and we considered traits that have been previously associated with cardiac phenotypes in the literature, but in addition included an unbiased selection of phenotypes for exploration. In total, we considered 71 quantitative phenotypes and 63 (binary) disease endpoints (Supplementary Data 1). Out of these, 31 phenotypes were significantly associated ($P_{adj} < 0.05$) with at least one of the diastolic function PIVSs after leave-one-out cross validation (Fig. 5). Some of the identified PheWAS associations are consistent with the phenotype correlation analysis (e.g. pulse rate and blood pressure). We also confirmed associations between diastolic function and previously reported biomarkers of heart failure (e.g. SHBG[13] and IGF-1[14]). Furthermore, we identified an association of $PDSR_{rr}$ to heart failure, cardiomyopathy and dilated cardiomyopathy, implicating diastolic function in cardiovascular endpoints.

Mendelian randomization—

Diastolic dysfunction is a substrate for the subsequent development of heart failure and, in observational studies, diabetes and hypertension are associated risk factors[15]. Here we used Mendelian Randomization (MR) to identify potential causal relationships between diastolic function as an exposure and two key clinical outcomes

(mixed aetiology heart failure and atrial fibrillation). We also assessed causal effects of biochemical, metabolic and haemodynamic exposures on diastolic function. These were chosen on the basis of clinical plausibility and the findings of the phenotype correlation analysis.

We tested a number of MR techniques, each addressing different assumptions, and excluded potentially confounding instruments. A strong bidirectional causal relationship was observed between pulse rate and $PDSR_{rr}$, $PDSR_{ll}$ and LAV_{max_i} (Extended Data Fig 6, Supplementary Figs 12 to 14, Supplementary Tables 2 to 4), consistent with findings from pre-clinical models[16]. Diastolic blood pressure was causally associated with $PDRS_{rr}$, and had a bidirectional association with $PDSR_{ll}$. Systolic blood pressure was causally associated $PDSR_{ll}$, but not $PDSR_{rr}$. In addition, higher total peripheral resistance was strongly associated with higher $PDSR_{ll}$, $PDSR_{rr}$ and LAV_{max_i}, adding to the evidence implicating ventriculo-vascular coupling in the development of the diastolic dysfunction[17].

We also identified a potential causal relationship between lower $PDSR_{rr}$ (stiffer ventricle) and increased risk of heart failure Supplementary Fig 11, which was further corroborated using GWAS summary results[18] from the HERMES consortium (Supplementary Table 5), a GWAS meta-analysis from 47309 heart failure cases and 930014 controls. The magnitude of the effect observed in the MR analysis is consistent with the observational epidemiological estimate, derived from correlating $PDSR_{rr}$ with incident heart failure (Extended Data Fig 6). We found no causal relationship between longitudinal $PDSR_{ll}$ and heart failure, and neither was one observed in our epidemiological analysis (Extended Data Fig 6).

Diastolic dysfunction is frequently present in diabetic patients[19], however the effects are mostly mediated by an increased risk of coronary artery disease[18]. We found parameter estimates that support a causal relationship between diabetes as an exposure and diastolic function as an outcome, as well as a potential link with instruments for lipid profiles.

Lastly, we found a causal association between LAV_{max_i} and an outcome of atrial fibrillation[20], but there was no evidence that ventricular stiffness also has a causal association.

Discussion

Diastole is a complex series of molecular, biophysical and electro-mechanical processes that initiate contractile deactivation and promote efficient ventricular filling. Impairment of these coordinated mechanisms may lead to diastolic dysfunction which is associated with the presence of multiple cardiovascular risk factors leading to reduced quality of life and higher mortality[21, 22]. Here, we used deep learning cardiac motion analysis to perform the first reported GWAS of diastolic function traits with the aim of determining tractable causative mechanisms. We found that diastolic function was a heritable trait with associations in loci related to myofilament mechanics, protein synthesis during mechanical stress and regulation of cardiac contractility. Furthermore, we find a role for a gene implicated in endothelium-derived signalling in diastolic function that is a potential therapeutic target[23]. Lastly, through

Mendelian randomisation we observe a causal relationship between genetically-determined diastolic function and heart failure outcomes.

A decline in diastolic function is a feature of the ageing heart and we found that age was a strong independent predictor of diastolic function, with a greater decrease present in males. Outcome studies have suggested that this is a prognosticallybenign feature of healthy ageing that is not related to adverse effects of cardiac senescence[2,24,25]. Changes in titin protein phosphorylation, myocardial redox state and impairment of nitric oxide signalling have been proposed as potential mechanisms[26], and clinical studies indicate that age-related myocardial fibrosis, cardiomyocyte hypertrophy, and reduced microvascular density, may be a consequence rather than an initiating cause of diastolic dysfunction[27]. Non-invasive imaging biomarkers of fibrosis have also shown promise in identifying biologically relevant pathways for myocardial fibrosis in adult hearts[28].

We found that diabetes was causally associated with impaired diastolic function after excluding potentially confounding instruments. In epidemiological analyses this relationship was independent of age, BSA, and systolic blood pressure. Increased myocardial stiffness is recognised as one of the earliest, and potentially reversible, manifestations of myocardial dysfunction in diabetes[29]. Several underlying mechanisms related to insulin resistance have been proposed that include altered cardiac energetics and accumulation of advanced glycation end products that promote ventricular stiffness[30]. We also observed a unidirectional causal relationship between genetically-determined diastolic function and an outcome of heart failure, as well as associations with cardiovascular endpoints and circulating biomarkers of heart failure through PheWAS. Longitudinal cohort studies have suggested that persistence or progression of diastolic dysfunction is a risk factor for subsequent heart failure[15], and our findings suggest that ventricular stiffness is a substrate for the evolution of mixed aetiology heart failure. We also found a unidirectional causal association between left atrial volume and atrial fibrillation, suggesting that it is atrial remodelling that drives this arrhythmic outcome[31]. Lipid profiles are associated with adverse changes in cardiac structure and systolic function, and our findings extend that causal association to diastolic traits[32].

Our study provides insights into the biological basis of diastolic function with potential implications for therapy development. We identified common variants within genes implicated in cardiomyopathies (e.g. *BAG3*, *FHOD3*, *PLN*), suggesting sarcomere homeostasis during mechanical stress may affect diastolic function in both health and disease[33]. Phospholamban (*PLN*) is a key regulator of cardiac diastolic function, which modulates sarcoplasmic reticulum calcium-ATPase activity[34]. Common variants in this gene are also associated with trabeculation which has been implicated in promoting ventricular filling[9]. Speckle-tracking echocardiography of *Pln* knockout mice reveals alterations in longitudinal strain but not radial strain[35], which is concordant with our observed associations with diastolic function and may relate to associated changes in ventricular geometry[36]. Although there is a genetic correlation between strain rate vectors the majority of SNPs used as polygenic instruments were independent of each other for these traits. We also identified a potential therapeutic target through the association of variants at the locus of *NPR3* influencing diastolic function and risk of heart failure. Previous studies have

highlighted its role in blood pressure control[37], and in mediating the cardioprotective effects of cardiomyocyte and fibroblast-released CNP[23].

This analysis has some limitations. UK Biobank is a large-cross sectional study that is subject to selection bias and latent population stratification, however risk factor associations appear to be broadly generalisable[38]. The population is predominantly European and further work is required to explore diastolic traits and outcomes in people of diverse ancestries. Echocardiography has been the cornerstone of assessing diastolic function by characterising features of ventricular relaxation, stiffness and recoil[39]. However, feature-tracking CMR has excellent agreement with speckle-tracking echocardiography[40] and invasive measures of diastolic function[41]. While analysis of myocardial deformation is performed throughout the cardiac cycle the measures of early diastolic strain rate may not capture variation in active relaxation prior to ventricular filling. While the relationship between quantitative and dichotomous outcomes may be non-linear such a relationship has not been observed between other genetically-driven diastolic traits and outcomes[42].

In conclusion, we found that diastolic function is a heritable trait that is causally upstream of incident heart failure. Associated common variants are related to genes that maintain functional homeostasis under biomechanical stress. We also identify a gene encoding an atrial natriuretic peptide receptor as a potential therapeutic target for modulating aspects of diastolic function.

Methods

All analyses in this study can be found here https://github.com/ImperialCollegeLondon/diastolic_genetics/[43] and were conducted with R version > 3.6.0.

Participants

For UK Biobank, approximately 500,000 community-dwelling participants aged 40–69 years were recruited across the United Kingdom between 2006 and 201044. All subjects provided written informed consent for participation in the study, which was also approved by the National Research Ethics Service (11/NW/0382). Our study was conducted under terms of access approval number 28807 and 40616. A range of available data were included in this study comprising genotyping arrays and whole exome sequencing, cardiac imaging, health-related diagnoses, and biological samples.

There are 488,252 genotyped participants of which 200,640 have whole exome sequencing. We partitioned 39,559 participants with both CMR imaging and genotyping array data into two tranches by date of release from UK Biobank providing a discovery dataset of 26,893 participants and a validation dataset of 12,666 participants.

Imaging protocol

A standardised CMR protocol was followed to assess cardiac structure and function using two-dimensional retrospectively-gated cine imaging on a 1.5T magnet (Siemens Healthineers, Erlangen, Germany). A contiguous stack of images in the left ventricular short-axis plane from base to apex was acquired, with long axis cine imaging in the two and

four chamber views. Each cine sequence had 50 cardiac phases with an acquired temporal resolution of 31 ms7. Transverse cine imaging was also performed in the ascending and descending thoracic aorta. All imaging phenotypes used for the analysis underwent quality control assessment[8]. Participants also underwent a resting 12 lead electrocardiogram which was automatically analysed using proprietary software (CardioSoft, GE Healthcare).

Cardiac image analysis

Segmentation of the short-axis and long-axis cine images in UK Biobank was made using fully convolutional networks, a type of deep learning neural network, which predict a pixel-wise image segmentation by applying a number of convolutional filters onto each input image for feature extraction and classification[9]. The accuracy of image segmentation on the UK Biobank dataset is equivalent to expert human readers[45]. End-diastolic volume, end-systolic volume, stroke volume, and ejection fraction were determined for both ventricles. Left ventricular myocardial mass was calculated from the myocardial volume assuming a density of 1.05 g.ml $^{-1}$. Left atrial volume was calculated from the segmented images using the biplane area–length formula $V = \frac{8}{3\pi} \cdot \frac{A_{2Ch} \cdot A_{4Ch}}{L}$, where A_{2Ch} and A_{4Ch} indicate the atrial area on the two and four-chamber cines respectively, and L indicates the longitudinal diameter averaged across two views. Measurements were indexed to body surface area (BSA) according to the Du Bois formula: 0.20247 * ($Weight^{0.425}$) * ($Height^{0.725}$), with weight in kg and height in m. The heart was divided into 16 standardised anatomical segments, excluding the true apex, according to American Heart Association nomenclature[46].

The aorta was segmented on the cine images using a spatio-termporal neural network[47]. The maximum and minimum cross-sectional areas were derived from the segmentation and distensibility calculated using estimates of central blood pressure obtained using peripheral pulse-wave analysis (Vicorder, Wuerzburg, Germany)[8].

Motion tracking was performed on the cine images using non-rigid image registration between successive frames (in GitHub repository ukbb_cardiac)[48, 49]. To reduce the accumulation of registration errors motion tracking was performed in both forward and backward directions from the end-diastolic frame and an average displacement field calculated[8]. This motion field was then used to warp the segmentation contours from end-diastole onto successive adjacent frames. Circumferential (E_{cc}) and radial (E_{rr}) strains were calculated on the short axis cines by the change in length of respective line segments (Fig. 1A) as $E_{dir} = \frac{\Delta L_{dir}}{L_{dir}}$, where dir represents the direction, L_{dir} the length of a line segment along this direction and ΔL_{dir} its change over time. Motion tracking was also performed on the long-axis four-chamber cines to derive longitudinal (E_{ll}) strain. Peak strain for each segment and global peak strain were then calculated (Fig. 1B). Strain was measured from slices acquired at basal, mid-ventricular, and apical levels. For comparison between each component absolute strain values are reported. Strain rate was estimated as the first derivative of strain and peak early diastolic strain rate in radial ($PDSR_{rr}$) and longitudinal ($PDSR_{ll}$) directions was detected using an algorithm to identify local maxima (in GitHub repository peak_detection) (Fig. 1C).

Non-imaging phenotypes

In total we consider 110 non-imaging cardiovascular-related phenotypes in UK Biobank participants for the phenotype regression analysis and the genetic analysis. These phenotypes contain information acquired by touch screen questionnaire, interview, biophysical measurement, hospital episode statistics, primary care data and biochemical analysis of venous blood. Details of how each phenotype was acquired are available on the UK Biobank Showcase (http://biobank.ctsu.ox.ac.uk/crystal/). It should be noted that the biochemical markers used here were acquired at the initial assessment visit that preceded imaging assessment. Also, note that not all phenotypes were used in both the phenotype and the genetic analysis (e.g., due to lack of available data at the imaging visit). We refer to the Supplementary Material both for details on the definition of the considered phenotypes and for information on the inclusion of specific phenotypes for each analysis.

Statistical significance testing and multiplicity control

We consider in all phenotype analysis a P-value < 0.05 as significant. Where not stated otherwise, we control the false discovery rate with the Benjamini-Hochberg adjustment. Significance thresholds and decision criteria for GWAS significant loci and causality assessment (Mendelian Randomization) are described in the respective sections and/or in the Supplementary Material.

Phenotype Association Analysis

Continuous variables are expressed as mean ± standard deviation (SD). Differences in continuous variables between groups were performed using Student's t-test. Univariable and multiple linear regression analysis was used to explore the phenotype relationship between each diastolic parameter and cardiovascular risk factors. To identify relationships between diastolic function and a broader range of imaging and non-imaging phenotypes, including circulating biomarkers, we used the least absolute shrinkage and selection operator (LASSO) with stability selection, to optimise the model coefficients. We then ran regression diagnostics on the model with the selected variables, to exclude a possible collinearity inappropriately influencing our model (see Supplementary Material for details on the phenotype analysis and LASSO analysis procedure).

Genotyping and sample QC

Genotyping of UK Biobank participants has been described elsewhere in detail[50]. Briefly, UK Biobank genotyping for 488,252 subjects was performed on the UK BiLEVE or UK Biobank Axiom arrays. Imputation was based on the HaplotypeReference Consortium panel and the UK10K+1000 Genomes panel. In this study, UK Biobank Imputation V3 (in GRCh[37] coordinates) were used. Whole exome sequencing (WES) was performed on data released in 2020 collected from 200,640 UK Biobank participants[51]. The sequencing methods and variant calling procedures have been described in detail[52]. In the present study, genotypes in their released PLINK-format files are utilized, and samples were restricted to the European population. Quality control of the genetic data was performed as recommended by UK Biobank (see Supplementary Material for details on the procedure and number of excluded samples).

GWAS analysis

For the genetic analysis, there were 34,242 participants of European ancestry (see Supplementary Material for criteria) providing a discovery dataset of 23,321 participants and a validation set of 10,924 participants. GWAS analyses for the three diastolic function traits and additional quantitative traits of interest (as described for the causality assessment) were performed with BOLT-LMM (version 2.3.2) which accounts for ancestral heterogeneity, unknown population structure, and sample relatedness[53, 54]. GWAS analyses were adjusted for imaging traits for the first ten genetic principal components, sex, age at time of MRI, the genotyping array and the MRI assessment center and for non-imaging quantitative traits for the first ten principal components, sex, age at measurement of the trait and the genotyping array. GWAS analyses for clinical endpoints of interest (binary endpoints) were conducted with PLINK2 and adjusted for the first ten principal components, sex, age at baseline and the genotyping array. Post-GWAS filtering removed any SNPs with a Hardy-Weinberg equilibrium p-value < 0.05 and $MAF < 0.005$.

Assessment of shared genetic architecture

For the assessment of shared genetic architecture between diastolic function traits, LD score regression (LDSC (LD SCore) v1.0.1, PMID 25642630) was used to obtain a genetic correlation score between each pair of traits.

Variant annotations

Lead variants for each locus were assigned causal genes, where possible, using a combination of variant annotations and additional functional genomic data sources (colocalisation). Each lead variant was systematically tested for any evidence of functional consequence using VEP. In addition, QTL evidence was extensively searched using Open Targets Genetics[55]. Where eQTL data was available for the locus, the full summary statistics were downloaded to assess colocalisation (see Suppplementary Material).

Variant Effect Predictor (VEP)[56] and Loss-of-Function Transcript Effect Estimator (LOFTEE)[57] plugin were applied on all genomic variants of WES data. In the present study, we considered the genomic variants predicted by LOFTEE with high-confidence label "HC", non-dubious (no "LoF flag" such as variants that located in poorly conserved exons, or splice variants that affect NAGNAG sites or non-canonical splice regions), and minor allele frequency < 0.05, as a Loss-of-function (LoF) mutation.

LoF association analysis

A Loss-of-Function (LoF) carrier indicator was created for each WES sample and each of the human protein-coding genes based on the collapsed information of LoF annotations. A subject was considered as an LoF carrier of the gene if there was at least one LoF mutation (based on methods in the variant annotation section), and a non-carrier if there was none. We then conducted the association test between LoF carrier indicator and the three diastolic function imaging phenotypes. Linear regression was performed with the adjustment of sex, age at time of MRI, and the top ten genetic principal components. The association results were further filtered as those with at least two carriers and the endpoint available. The association was considered significant after multiple testing correction at $\alpha = 0.05$ (FDR,

calculated for three diastolic function traits). We identified 18,660 participants with both whole exome sequencing data and CMR imaging data.

Polygenic instrumental variable scores (PIVS)

Candidate variants for PIVS for the three diastolic function traits ($LAV_{max,I}$, $PDSR_{II}$, $PDSR_{III}$) were obtained based on the respective GWAS (full imaging cohort) results by performing clumping (PLINK 1.9) using an LD threshold of $R^2 = 0.1$ (in a window of 1000kb) and considering all SNPs with $P < 10^{-6}$. Unlike more traditional polygenic risk scores we do not use thousands of variants as instruments but aim to identify a set of instrumental variables that are minimally correlated. This comes with the price of a relatively small set of instruments that explains less variability of a trait, but can be used as proper instruments for the Mendelian randomization analysis. Candidate variants were included in multivariate linear modelling evaluated on the European subset of the full imaging cohort with the first ten genetic principal components, age at MRI, sex, genotyping array and the MRI center as additional covariates and the respective diastolic function trait as dependent variables. The diastolic function traits were scaled to standard deviation (sd) 1 prior to the model estimation - therefore, a unit change in the PIVS score represents a change of 1 sd unit in the respective diastolic function trait. PIVS estimates per individual were then calculated by multiplying the observed genotype with the estimated beta from the multivariate linear model for each SNP and summing these values up. Missing genotypes were imputed using a mean imputation. The variance explained for the PIVS is measured by R^2, estimated in a linear regression with the PIVS as only variable and the respective diastolic function trait as endpoint.

Next, we conducted a PheWAS using the obtained PIVS (see above and Supplementary Material for a full definition of included phenotypes in the PheWAS). Evaluation of the PIVS were performed in the European non-imaging cohort, i.e. an independent set of subjects compared to the PIVS construction set. Only results are shown that are significant after multiple testing correction at $\alpha = 0.05$ (FDR, calculated per diastolic function trait) and, as a sensitivity analysis, for which all leave-one-SNP out cross validations analysis lead to a significant result at $\alpha = 0.05$ after multiple testing correction (FDR) for the number of considered phenotypes. The latter condition is supposed to exclude spurious results which are only driven by one single variant. Leave-one-SNP out cross-validation is performed by excluding one SNP from the list of candidate variants, then re-estimating the PIVS and performing the PheWAS as described above. For the leave-one-SNP out cross-validation, FDR adjustment is performed per combination of diastolic trait and phenotype, considering the number of included SNPs.

Mendelian randomization

For exploring the causes and consequences of diastolic function parameters, we used a bidirectional Mendelian randomization (MR) approach, i.e. two MR analysis are performed: first, an MR analysis using the first chosen trait as exposure is conduced and secondly a MR analysis using the selected second trait is run. By considering both results, evidence can be gathered for a one-directional causal relationship, a bi-directional causal relationship or no causal relationship at all. We performed this analysis taking into account one diastolic

and one non-diastolic function trait and for that, we selected non-diastolic function traits of interest by taking into account the results from the observational correlation analysis and clinical expertise. This approach lead to the consideration of six dichotomous risk factors associated with diastolic dysfunction, Arteriosclerosis, Atrial Fibrillation, Heart failure, Hypertension and Diabetes - considering Type I and Type II separately. Further, we considered four physiological variables as potential causes or consequences of changes in diastolic function, as well as five quantitative lipid traits as surrogate for arteriosclerotic risks as potential confounder source for changes in diastolic function. In total we analysed 15 non-diastolic phenotypes and the 3 diastolic phenotypes in our MR.

We established a workflow for the MR analysis which briefly described in this section. Full details are provided in the Supplementary Material. Genetic instrumental variables were selected from the UK Biobank GWAS results generated -as described above- via clumping with PLINK1.9 as described for the PIVS approach. The candidate SNP set prior to clumping was restricted to the intersection between the SNP sets of the pair of GWAS results (hypothesised causal trait GWAS and hypothesised consequence trait GWAS). A full list of the instrumental variables is contained in supplementary table file SupplementaryTable_InstrumentalVariantsMR.xlsx.

We aimed to remove potential confounding instruments by two filtering steps. First, we ran phenotype association analysis to identify and remove instruments that associate significantly with any of the traits Arteriosclerosis, Triglycerides, Apoliprotein B and LDL-Cholesterol. Second, we ran Steiger Filtering to remove instruments with potentially wrongly inferred causal directions.

All MR analysis are based on the point estimates and standard deviations obtained from the respective GWAS. We follow a similar approach to van Oort et al.[58] by using inverse-variance weighted method (IVW) as the main analysis and applying several other MR methods for ensuring robustness of the obtained results as sensitivity analyses. We used weighted median-based methods, MR-PRESSO and MR-Egger. Consistent effect estimates across the different methods improves our confidence in a truly causal effect. We consider an association as "potential causal" if the main analysis indicates a causal relationship ($P < 0.01$), at least two of the sensitivity analyses indicate at least a suggestive causal relationship ($P < 0.05$) and none of the sensitivity analyses indicate associations with inconsistent effect directionality, i.e. none of the methods showed a suggestive association with conflicting directionality ($P < 0.05$). No explicit multiplicity adjustment is performed for MR experiments. For "potential causal" associations, we next conducted a supplementary sensitivity analysis using published GWAS results as described in the Supplementary Material - if published GWAS data was available.

All analysis, which involved diastolic and non-diastolic function traits, were conducted in a two-sample approach, i.e., the diastolic function trait GWAS was calculated in the full imaging cohort and the non-diastolic function trait GWAS was calculated in the non-imaging cohort.

For comparison of the effect estimates from the MR-analysis to the observed correlation of diastolic function measurement and disease status, we restricted the analysis population to subjects which were disease-free at the CMR visit. We then fitted a logistic regression model by coding subjects who experienced a first event of the selected disease during follow-up time as 1 and event-free subjects during follow-up as 0. As covariates, we included age at CMR visit, gender, diabetes status, diastolic blood pressure and BMI. Note that this analysis was only performed for relationships judged as potential causal and that involves a disease endpoint (and not a quantitative measurement like pulse rate).

NPR3 Pathway Analysis

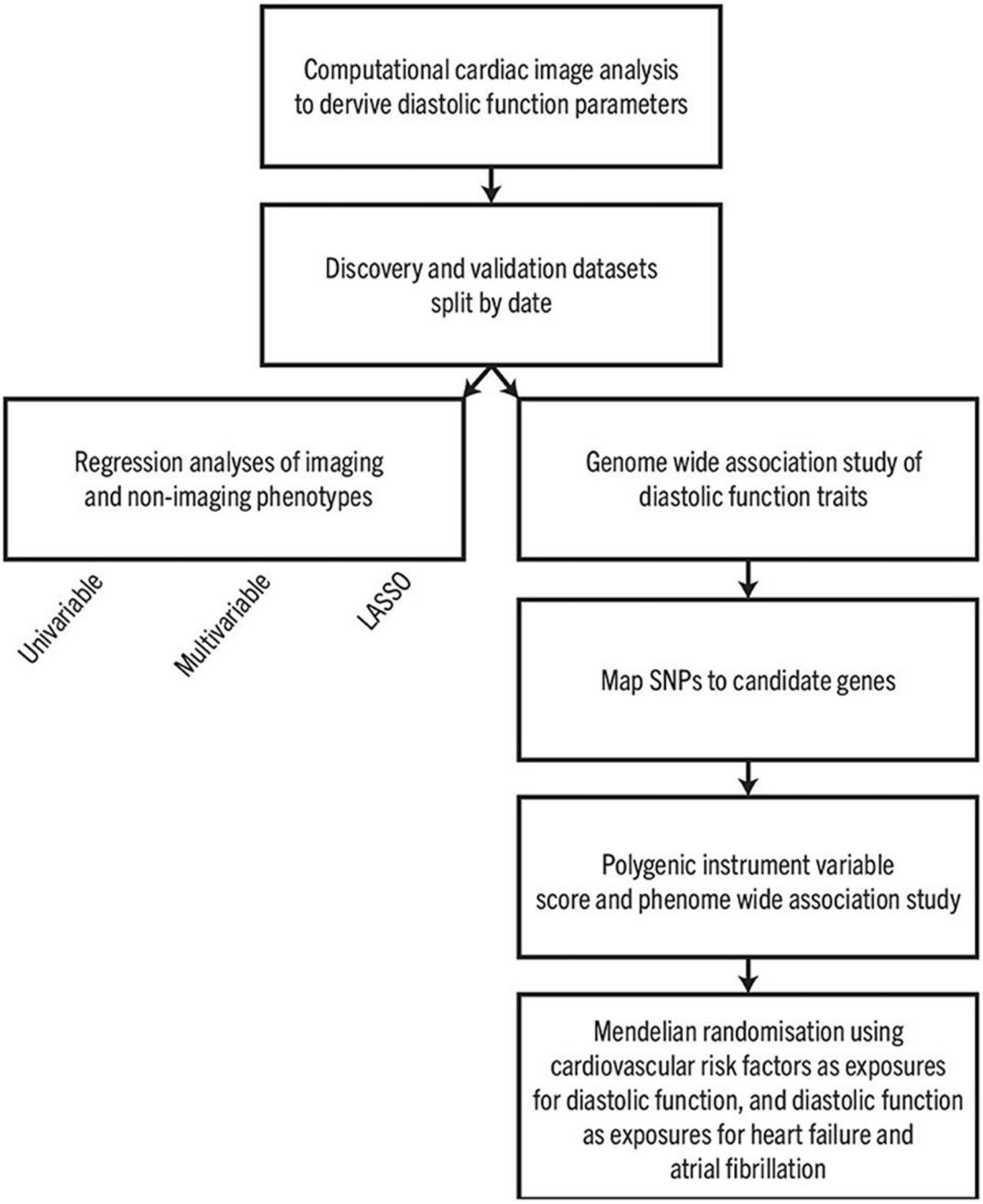
In order to increase our understanding of the association of NPR3 with LAV_{max_i} and to further characterize the role of natriuretic peptides, we looked for additional genetic associations within genes of the natriuretic peptide pathway (so in addition to *NPR3* - *NPR1*, *NPR2*, *NPPA*, *NPPB*, and *NPPC*). We conducted GWAS using BOLT-LMM for all imaging traits listed in Extended Data Table 1 as described above, as well as any non-imaging traits associated with rs1173727 (the lead variant for *NPR3*) across the 4 loci (*NPPA* and *NPPB* share the same locus). The GWAS summary statistics were filtered to a 1MB window around each gene (for *NPPA/B*, the gene used for centering was *NPPA*). Across these summary statistics, we performed clumping with a p-value threshold of 10^{-5} and $R^2 < 0.1$.

For the identified tag SNPs and associated variants in LD from the clumping analysis, we then tested which of these variants we could confidently link to the natriuretic gene in the locus. If any variant was classified as missense, we selected that variant directly. For eQTL variants, we used colocalisation analysis to link these SNPS to the natriuretic genes in each locus. Relevant eQTL and pQTL data was used (eQTL summary statistics were taken from eQTL Catalog[59] and pQTL data from Sun et al.[60]) and SNPs with only a clear association with the gene of interest and traits of interest were kept (i.e. $p < 1^{-4}$ for association with gene or protein expression, $P < 10^{-5}$ for association with the trait, and $H_{12} > 0.5$ was used as a threshold for the co-localization analysis).

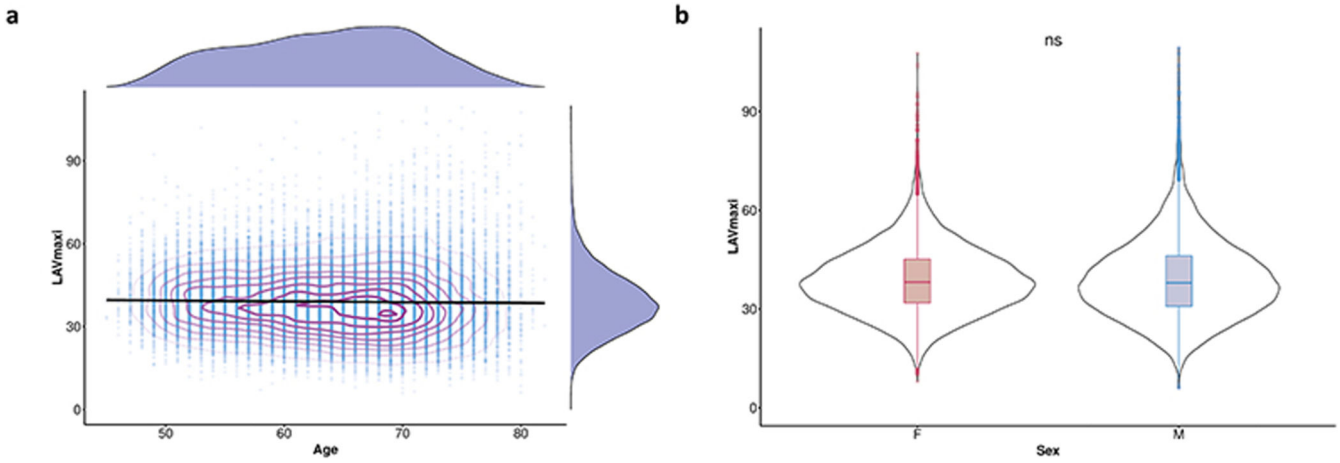
Hierarchical clustering was then performed on the $-\log(P) \times \beta$ values with the βs aligned to have a negative sign on the diastolic blood pressure. Extended Data Fig 5 shows all SNPs and traits with a genome-wide significant association. The SNPs and traits with suggestive associations ($P < 10^{-5}$) are shown in the Supplementary Material (Supplementary Fig 6).

Europe PMC Funders Author Manuscripts

Extended Data



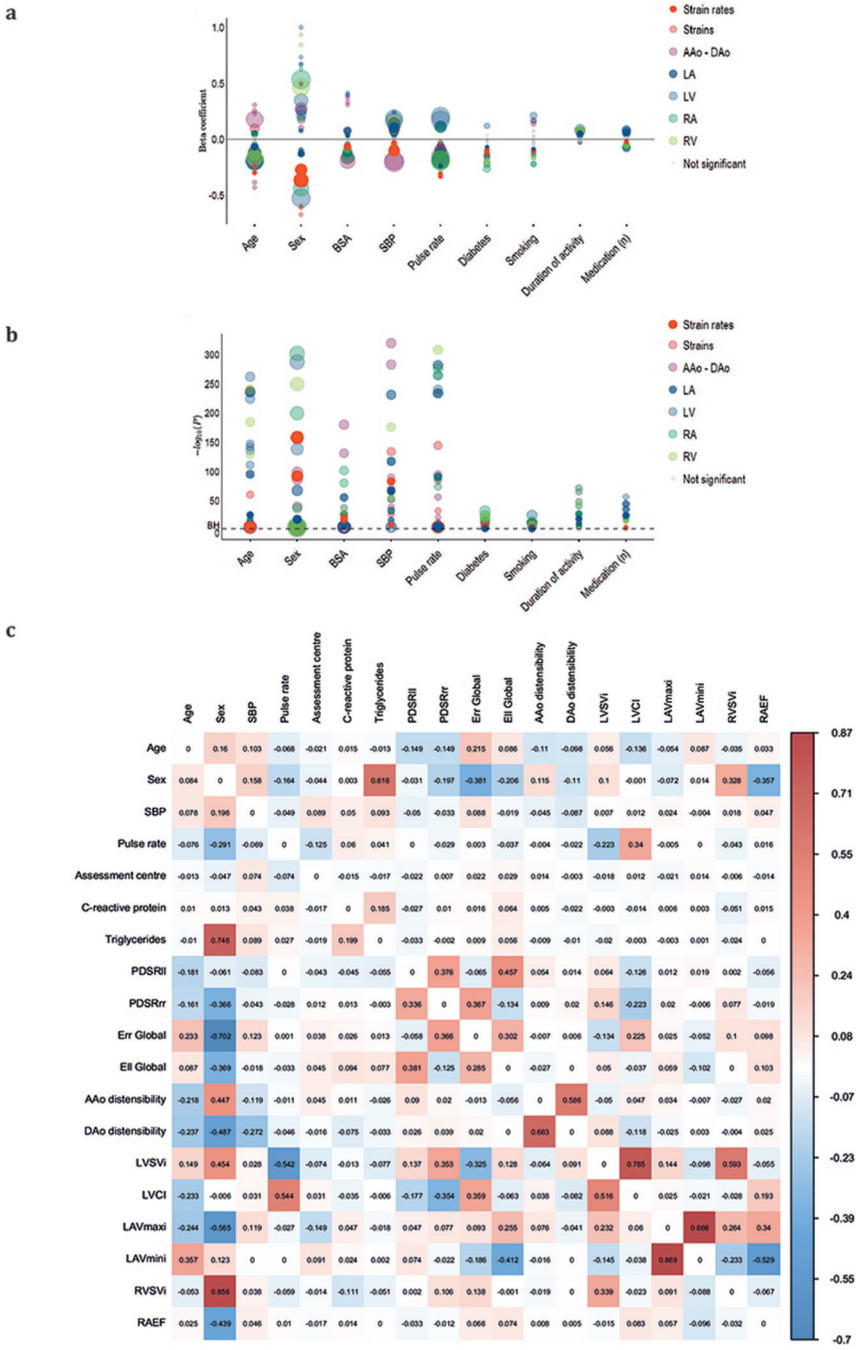
Extended Data Fig. 1.



Extended Data Fig. 2.

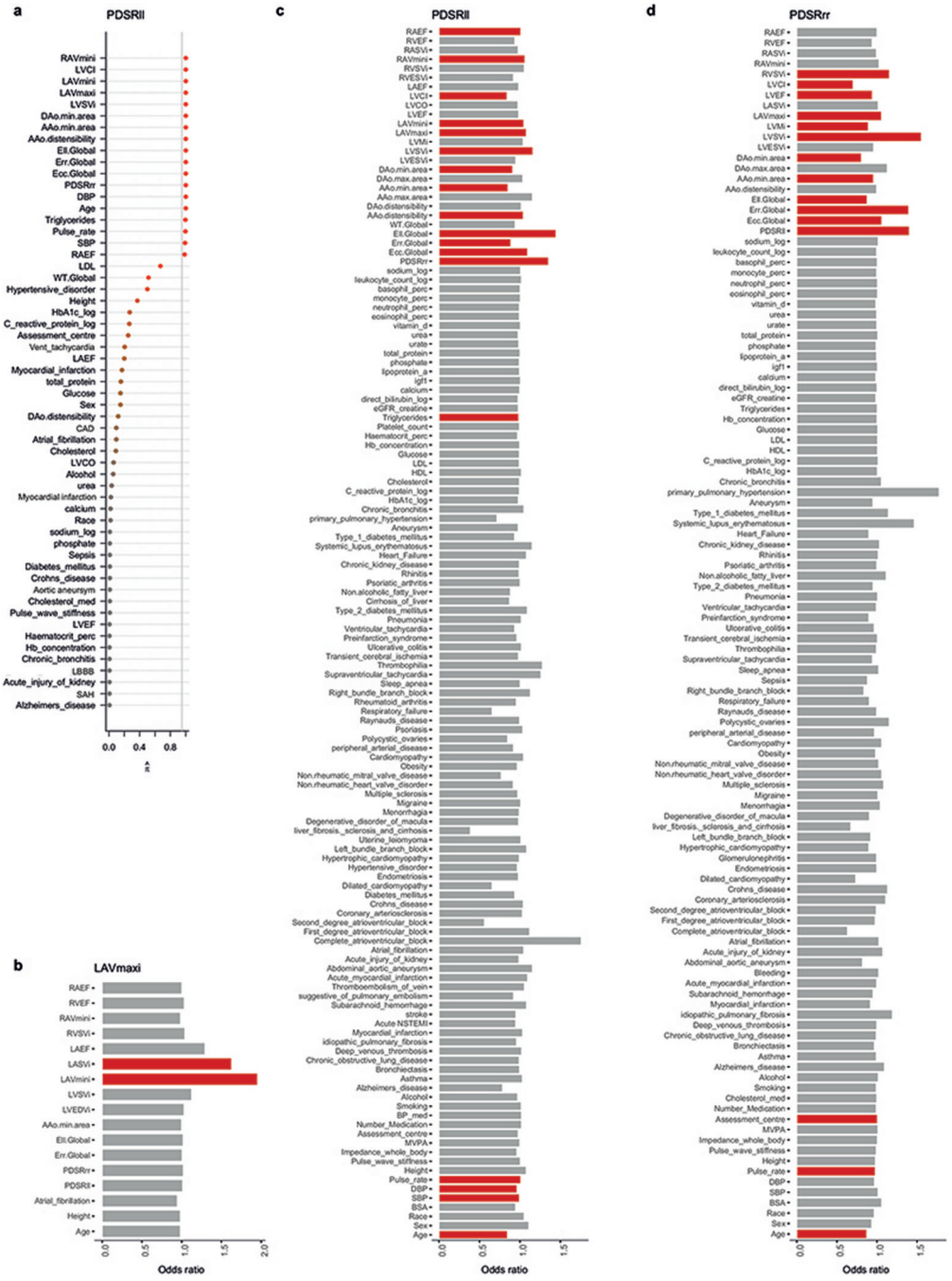
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

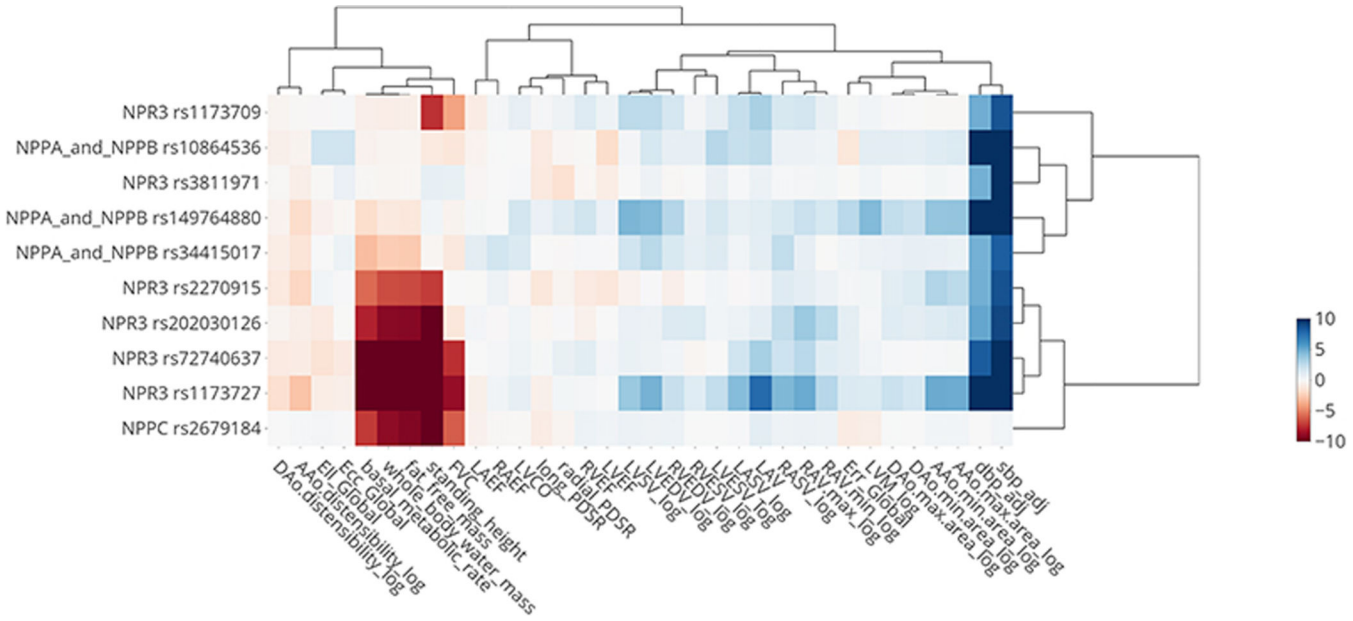


Extended Data Fig. 3.

Europe PMC Funders Author Manuscripts

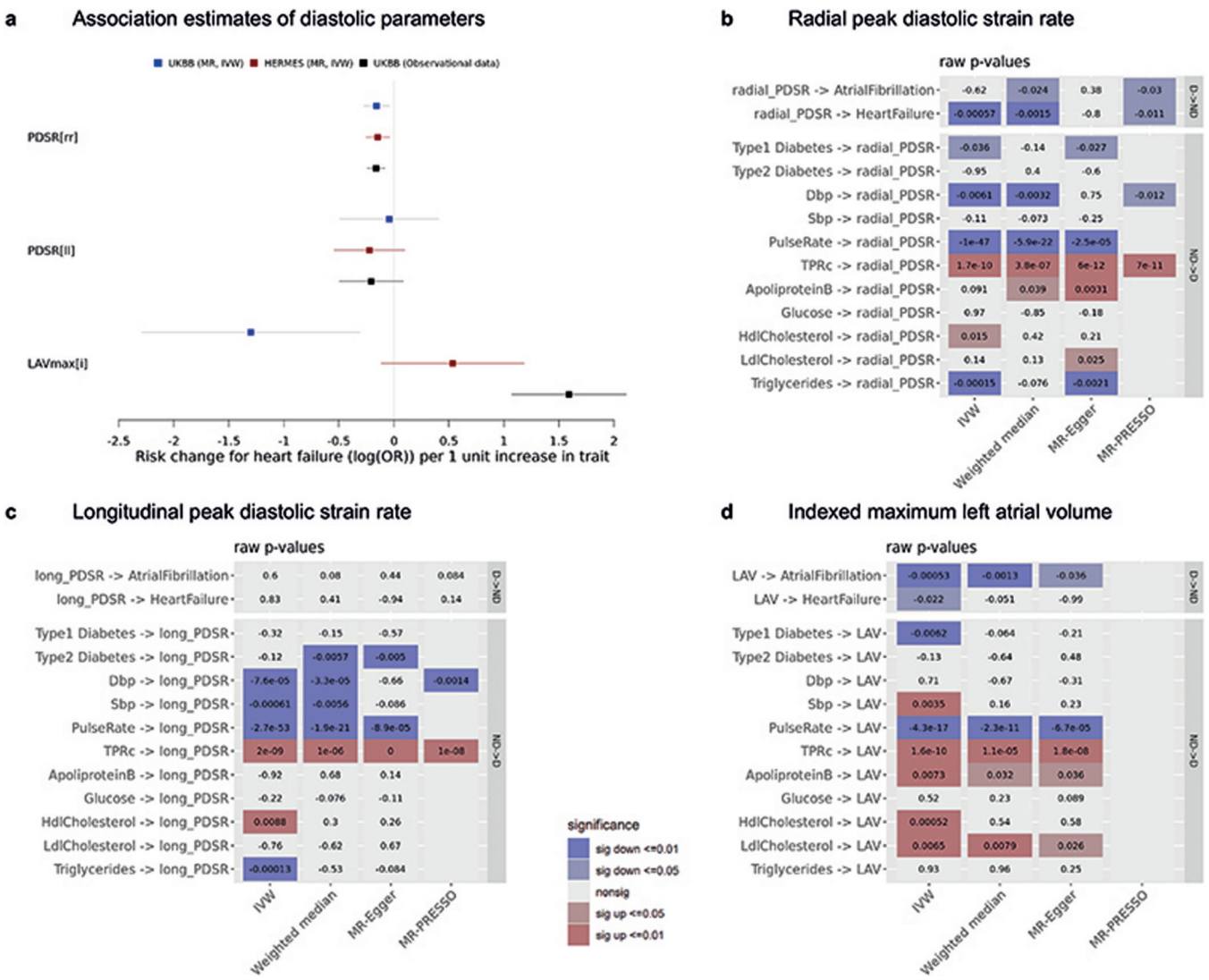


Extended Data Fig. 4.



Extended Data Fig. 5.

Europe PMC Funders Author Manuscripts



Extended Data Fig. 6.

Extended Data Table 1

Baseline characteristics	Mean ± SD or n (%)	Cardiac Characteristics from CMR	Mean ± SD or n (%)
Age (years)	63.6 ± 7.6	**LV wall thickness (mm)**	5.7 ± 0.8
Sex, men, n (%)	18,988 (48%)	**LV end-diastole volume $_{indexed}$ (mL/m^2)**	79 ± 13.8
Race, Nonwhite, n (%)	1,130 (2.8%)	**LV end-systole volume $_{indexed}$ (mL/m^2)**	32.1 ± 8.4
Body mass index (kg/m^2)	26.5 ± 4.4	**LV stroke volume $_{indexed}$ (mL/m^2)**	46.9 ± 8.4
Body surface area (m^2)	1.9 ± 0.2	**LV ejection fraction (%)**	59.6 ± 6.1
Systolic Blood pressure (mmHg)	138.2 ± 18.3	**LV cardiac output (ml)**	5.4 ± 1.2
Diastolic blood pressure (mmHg)	78.6 ± 9.9	**LV cardiac index (ml/m^2)**	2.5 ± 0.5

Baseline characteristics	Mean ± SD or n (%)	Cardiac Characteristics from CMR	Mean ± SD or n (%)
Pulse rate (bpm)	70 ± 12	**LV mass $_{indexed}$ (mL/m^2)**	45.7 ± 8.5
Pulse wave arterial stiffness index (SI)	9.6 ± 2.9	**LA maximum volume $_{indexed}$ (ml/m^2)**	39 ± 11.2
Diabetes mellitus, n (%)	2,432 (6.2%)	**LA minimum volume $_{indexed}$ (ml/m^2)**	15.7 ± 7.5
Heart failure, n (%)	260 (0.66%)	**LA stroke volume $_{indexed}$ (ml/m^2)**	23.3 ± 5.8
Smoking status		**LA emptying fraction (%)**	61.2 ± 9.5
Current, n (%)	1,374 (3.5%)	**RV end-diastole volume $_{indexed}$(ml/m^2)**	83.6 ± 15.2
Previous, n (%)	13,330 (34.1%)	**RV end-systole volume $_{indexed}$ (ml/m^2)**	35.9 ± 9.3
Never, n (%)	24,443 (62.4%)	**RV stroke volume $_{indexed}$ (ml/m^2)**	47.7 ± 8.9
Daily alcohol intake	6,597 (16.7%)	**RV ejection fraction (%)**	57.3 ± 6.1
Duration of physical activity in minutes per day		**RA maximum volume $_{indexed}$ (ml/m^2)**	46.4 ± 13.5
Moderate	53.9 ± 66.2	**RA minimum volume $_{indexed}$ (ml/m^2)**	24.7 ± 9.2
Vigorous	40.3 ± 40.4	**Right atrial stroke volume $_{indexed}$(ml/m^2)**	21.6 ± 6.8
Number of treatment/ medications taken	1.9 ± 2.1	**RA emptying fraction (%)**	47.2 ± 9.5
Blood pressure medication	2,042 (5.2%)	**AAo distensibility $_{indexed}$ ($10^{-3}{\cdot}mmHg^{-1}$)**	0.97 ± 0.63
Cholesterol medication	6,015 (15.2%)	**AAo maximum area (mm^2)**	852.3 ± 188.4
Assessment centre		**AAo minimum area (mm^2)**	775.1 ± 183.9
Cheadle	25,176 (63.6%)	**DAo distensibility $_{indexed}$ ($10^{-3}{\cdot}mmHg^{-1}$)**	1.29 ± 0.8
Reading	4,361 (11%)	**DAo maximum area (mm^2)**	476.7 ± 96.8
Newcastle	10,022 (25.3%)	**DAo minimum area (mm^2)**	418.1 ± 91.6
Laboratory Biochemical Markers		**Strains and Strain rates**	
HbA1c (log (*mmol/mol*))	3.5 ± 0.13	**Peak diastolic longitudinal strain rates ($PDSR_{ll}$, s^{-1})**	1.64 ± 0.6
C-reactive protein (log (*mg/L*))	0.13 ± 1.02		
LDL (*mmol/L*))	3.6 ± 0.8	**Peak diastolic radial strain rates ($PDSR_{rr}$, s^{-1})**	5.71 ± 1.9
Glucose (*mmol/L*)	5.0 ± 0.93	**Global circumferential strain (E_{cc}, %)**	22.3 ± 3.4
Triglycerides (log (*mmol/L*))	0.36 ± 0.51	**Global longitudinal strain (E_{ll}, %)**	18.5 ± 2.8
eGFR cystatin ($mL \cdot min^{-1} \cdot 1.73\ m^{-2}$)	92 ± 12.2	**Global radial strain (E_{rr}, %)**	45.1 ± 8.4

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The study was supported by Bayer AG; Medical Research Council (MC-A658-5QEB0); National Institute for Health Research (NIHR) Imperial College Biomedical Research Centre; British Heart Foundation (NH/17/1/32725, RG/19/6/34387, RE/18/4/34215); Wellcome Trust (107469/Z/15/Z); Academy of Medical Sciences (SGL015/1006); Mason Medical Research Trust grant; and the Engineering and Physical Sciences

Research Council (EP/P001009/1). H.V.M. also receives funding from the Simons Center for Quantitative Biology at Cold Spring Harbor Laboratory.

Data Availability

All raw and derived data in this study is available from UK Biobank (http://www.ukbiobank.ac.uk/). GWAS summary level data are publicly available through the GWAS catalogue (accession number GCST90019012). eQTL data used for variant to gene mapping is available through eQTL Catalogue (https://www.ebi.ac.uk/eqtl/).

Code Availability

The analysis code is freely available on GitHub[43].

References

1. Ladeiras-Lopes R, Araújo M, Sampaio F, Leite-Moreira A, Fontes-Carvalho R. The impact of diastolic dysfunction as a predictor of cardiovascular events: A systematic review and meta-analysis. Rev Port Cardiol. 2019; 38: 789–804. [PubMed: 32001062]
2. Shah AM, et al. Contemporary assessment of left ventricular diastolic function in older adults: The atherosclerosis risk in communities study. Circulation. 2017; 135: 426–439. [PubMed: 27927714]
3. Chiao YA, Rabinovitch PS. The aging heart. Cold Spring Harb Perspect Med. 2015; 5 a025148-a025148 [PubMed: 26328932]
4. Aung N, et al. Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. Circulation. 2019; 140: 1318–1330. [PubMed: 31554410]
5. Wild PS, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. J Clin Invest. 2017; 127: 1798–1812. [PubMed: 28394258]
6. Travers JG, et al. HDAC inhibition reverses preexisting diastolic dysfunction and blocks covert extracellular matrix remodeling. Circulation. 2021.
7. Petersen SE, et al. UK biobank's cardiovascular magnetic resonance protocol. J Cardiovasc Magn Reson. 2016; 18: 8. [PubMed: 26830817]
8. Bai W, et al. A population-based phenome-wide association study of cardiac and aortic structure and function. Nat Med. 2020.
9. Meyer HV, et al. Genetic and functional insights into the fractal structure of the heart. Nature. 2020; 584: 589–594. [PubMed: 32814899]
10. Kermer J, et al. Assessment of diastolic dysfunction: comparison of different cardiovascular magnetic resonance techniques. ESC Heart Fail. 2020; 7: 2637–2649. [PubMed: 32686332]
11. Baroncini LAV, et al. Echocardiographic correlation between right ventricular function and left atrial volume. Arq Bras Cardiol. 2019; 112: 249–257. [PubMed: 30916187]
12. Ushijima T, et al. The actin-organizing formin protein Fhod3 is required for postnatal development and functional maintenance of the adult heart in mice. J Biol Chem. 2018; 293: 148–162. [PubMed: 29158260]
13. Pascual-Figal DA, et al. Sex hormone-binding globulin: a new marker of disease severity and prognosis in men with chronic heart failure. Rev Esp Cardiol. 2009; 62: 1381–1387. [PubMed: 20038404]
14. Barroso MC, et al. Serum insulin-like growth factor-1 and its binding protein-7: potential novel biomarkers for heart failure with preserved ejection fraction. BMC Cardiovasc Disord. 2016; 16: 1–9. [PubMed: 26728597]
15. Kane GC, et al. Progression of left ventricular diastolic dysfunction and risk of heart failure. JAMA. 2011; 306: 856–863. [PubMed: 21862747]

16. Reil J-C, et al. Heart rate reduction by If-inhibition improves vascular stiffness and left ventricular systolic and diastolic function in a mouse model of heart failure with preserved ejection fraction. Eur Heart J. 2013; 34: 2839–2849. [PubMed: 22833515]
17. O'Regan DP. Stiff arteries, stiff ventricles: correlation or causality in heart failure? Circ Cardiovasc Imaging. 2016; 9 e005150 [PubMed: 27353853]
18. Shah S, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nat Commun. 2020; 11: 163. [PubMed: 31919418]
19. Bouthoorn S, et al. The prevalence of left ventricular diastolic dysfunction and heart failure with preserved ejection fraction in men and women with type 2 diabetes: A systematic review and meta-analysis. Diab Vasc Dis Res. 2018; 15: 477–493. [PubMed: 30037278]
20. Ahlberg G, et al. Genome-wide association study identifies 18 novel loci associated with left atrial volume and function. Eur Heart J. 2021; 42: 4523–4534. [PubMed: 34338756]
21. Reddy YN, Carter RE, Obokata M, Redfield MM, Borlaug BA. A simple, evidence-based approach to help guide diagnosis of heart failure with preserved ejection fraction. Circulation. 2018; 138: 861–870. [PubMed: 29792299]
22. Playford D, et al. Diastolic dysfunction and mortality in 436 360 men and women: the National Echo Database Australia (NEDA). Eur Heart J Cardiovasc Imaging. 2020.
23. Moyes AJ, et al. C-type natriuretic peptide co-ordinates cardiac structure and function. Eur Heart J. 2020; 41: 1006–1020. [PubMed: 30903134]
24. van Grootel RWJ, et al. Influence of age and sex on left ventricular diastolic strain analysis. Int J Cardiovasc Imaging. 2019; 35: 491–498. [PubMed: 30377894]
25. Hung C-L, et al. Age- and sex-related influences on left ventricular mechanics in elderly individuals free of prevalent heart failure. Circ Cardiovasc Imaging. 2017; 10 e004510 [PubMed: 28093411]
26. Borbély A, et al. Cardiomyocyte stiffness in diastolic heart failure. Circulation. 2005; 111: 774–781. [PubMed: 15699264]
27. Campbell DJ, et al. Diastolic dysfunction of aging is independent of myocardial structure but associated with plasma advanced glycation end-product levels. PLoS One. 2012; 7 e49813 [PubMed: 23189164]
28. Nauffal V, et al. Genetics of myocardial interstitial fibrosis in the human heart and association with disease. medRxiv. 2021.
29. Grandi AM, et al. Effect of glycemic control on left ventricular diastolic function in type 1 diabetes mellitus. Am J Cardiol. 2006; 97: 71–76. [PubMed: 16377287]
30. Fontes-Carvalho R, Ladeiras-Lopes R, Bettencourt P, Leite-Moreira A, Azevedo A. Diastolic dysfunction in the diabetic continuum: association with insulin resistance, metabolic syndrome and type 2 diabetes. Cardiovasc Diabetol. 2015; 14: 1–9. [PubMed: 25582325]
31. Raisi-Estabragh Z, et al. Left atrial structure and function are associated with cardiovascular outcomes independent of left ventricular measures: a UK Biobank CMR study. Eur Heart J Cardiovasc Imaging. 2021.
32. Aung N, et al. The effect of blood lipids on the left ventricle: a Mendelian randomization study. J Am Coll Cardiol. 2020; 76: 2477–2488. [PubMed: 33213727]
33. Harper AR, et al. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. Nat Genet. 2021. 1–8. [PubMed: 33414547]
34. Frank KF, Bölck B, Erdmann E, Schwinger RH. Sarcoplasmic reticulum Ca2+-ATPase modulates cardiac contraction and relaxation. Cardiovasc Res. 2003; 57: 20–27. [PubMed: 12504810]
35. Chowdhury SA, et al. Modifications of sarcoplasmic reticulum function prevent progression of sarcomere-linked hypertrophic cardiomyopathy despite a persistent increase in myofilament calcium response. Front Physiol. 2020; 11: 107. [PubMed: 32210830]
36. Bonazzola, R; , et al. In: de Bruijne, M; , et al., editors. Image-derived phenotype extraction for genetic discovery via unsupervised deep learning in CMR images; Medical Image Computing and Computer Assisted Intervention - MICCAI 2021; Cham. 2021. 699–708.
37. Surendran P, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat Genet. 2016; 48: 1151–1161. [PubMed: 27618447]

38. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. BMJ. 2020; 368
39. Nagueh SF. Left ventricular diastolic function: Understanding pathophysiology, diagnosis, and prognosis with echocardiography. JACC Cardiovasc Imaging. 2020; 13: 228–244. [PubMed: 30982669]
40. Onishi T, et al. Global longitudinal strain and global circumferential strain by speckle-tracking echocardiography and feature-tracking cardiac magnetic resonance imaging: comparison with left ventricular ejection fraction. J Am Soc Echocardiogr. 2015; 28: 587–596. [PubMed: 25577185]
41. Ito H, et al. Cardiovascular magnetic resonance feature tracking for characterization of patients with heart failure with preserved ejection fraction: correlation of global longitudinal strain with invasive diastolic functional indices. J Cardiovasc Magn Reson. 2020; 22: 1–11. [PubMed: 31898543]
42. Arvanitis M, et al. Linear and nonlinear Mendelian randomization analyses of the association between diastolic blood pressure and cardiovascular events. Circulation. 2021; 143: 895–906. [PubMed: 33249881]
43. Thanaj M, Mielke J, Bender C, Zeng L, O'Regan DP. Genetic and environmental determinants of diastolic heart function. Zenodo. 2022; doi: 10.5281/zenodo.4767044
44. Sudlow C, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015; 12 e1001779 [PubMed: 25826379]
45. Bai W, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson. 2018; 20: 65. [PubMed: 30217194]
46. Cerqueira Manuel D, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. Circulation. 2002; 105: 539–542. [PubMed: 11815441]
47. Bai, W; , et al. Recurrent neural networks for aortic image sequence segmentation with sparse annotations; International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2018. 586–594.
48. Rueckert D, et al. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging. 1999; 18: 712–21. [PubMed: 10534053]
49. Puyol-Antón, E; , et al. Fully automated myocardial strain estimation from cine MRI using convolutional neural networks; 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018. 1139–1143.
50. Bycroft C, et al. The UK biobank resource with deep phenotyping and genomic data. Nature. 2018; 562: 203–209. [PubMed: 30305743]
51. Szustakowski JD, et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat Genet. 2021; 53: 942–948. [PubMed: 34183854]
52. Van Hout CV, et al. Exome sequencing and characterization of 49,960 individuals in the UK biobank. Nature. 2020; 586: 749–756. [PubMed: 33087929]
53. Loh P-R, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015; 47: 284. [PubMed: 25642633]
54. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018; 50: 906–908. [PubMed: 29892013]
55. Ghoussaini M, et al. Open targets genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. 2021; 49: D1311–D1320. [PubMed: 33045747]
56. McLaren W, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016; 17: 122. [PubMed: 27268795]
57. Karczewski KJ, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581: 434–443. [PubMed: 32461654]
58. van Oort S, Beulens JW, van Ballegooijen AJ, Burgess S, Larsson SC. Cardiovascular risk factors and lifestyle behaviours in relation to longevity: a Mendelian randomization study. J Intern Med. 2020.
59. Kerimov N, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet. 2021; 53: 1290–1299. [PubMed: 34493866]

60. Sun BB, et al. Genomic atlas of the human plasma proteome. Nature. 2018; 558: 73–79. [PubMed: 29875488]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

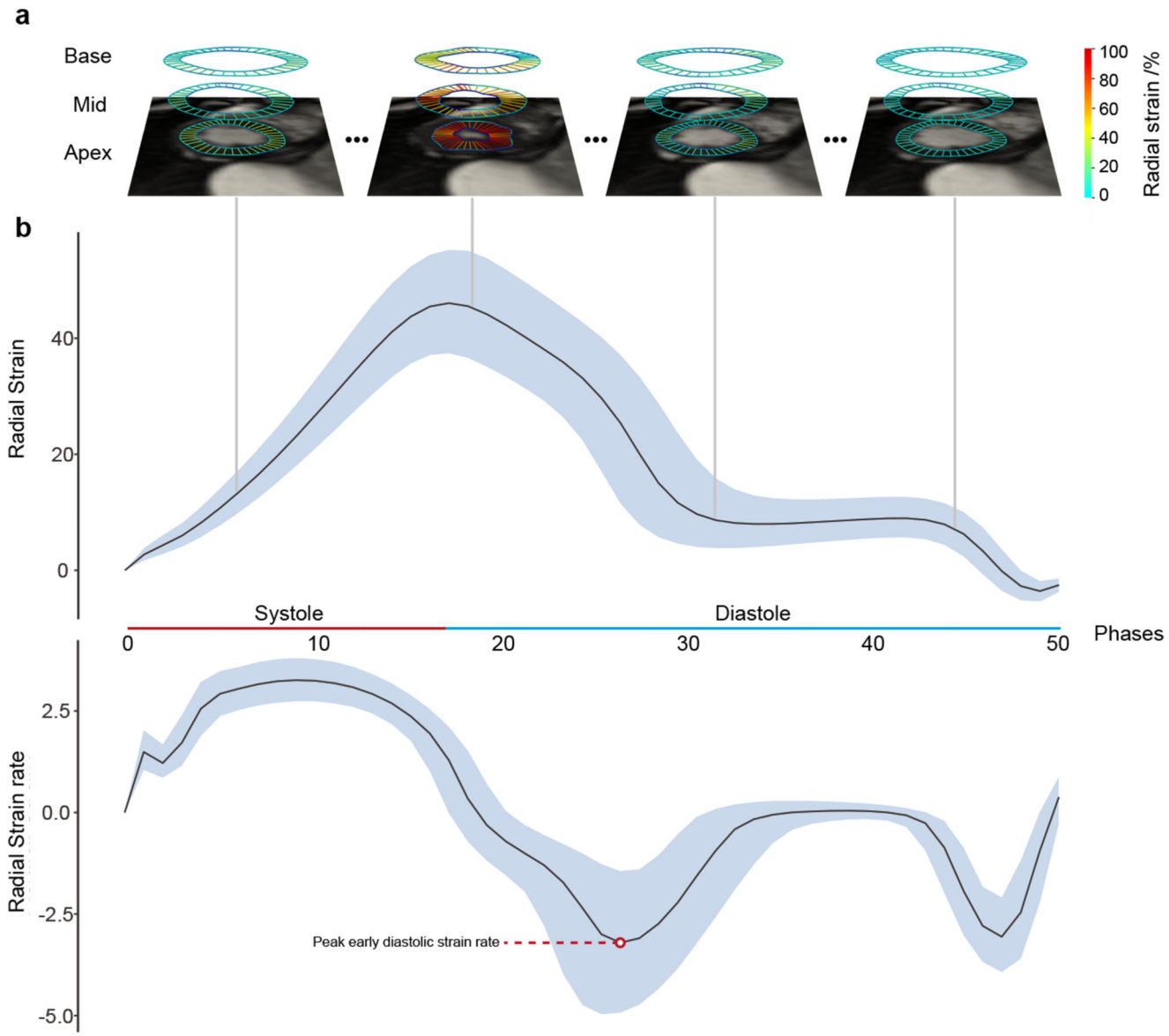


Figure 1. Analysis of cardiac motion.

Motion analysis of cardiac magnetic resonance imaging performed on left ventricular short axis cines. A) An example from one individual where deep learning segmentation and image registration were used to determine the radial components of myocardial deformation. Data from the basal, mid-ventricular, and apical levels are shown at four representative phases from the 50 acquired. B) Radial strain and strain rate (first derivative of strain) for all UK Biobank subjects (median and interquartile ranges, n=39,559 individuals).

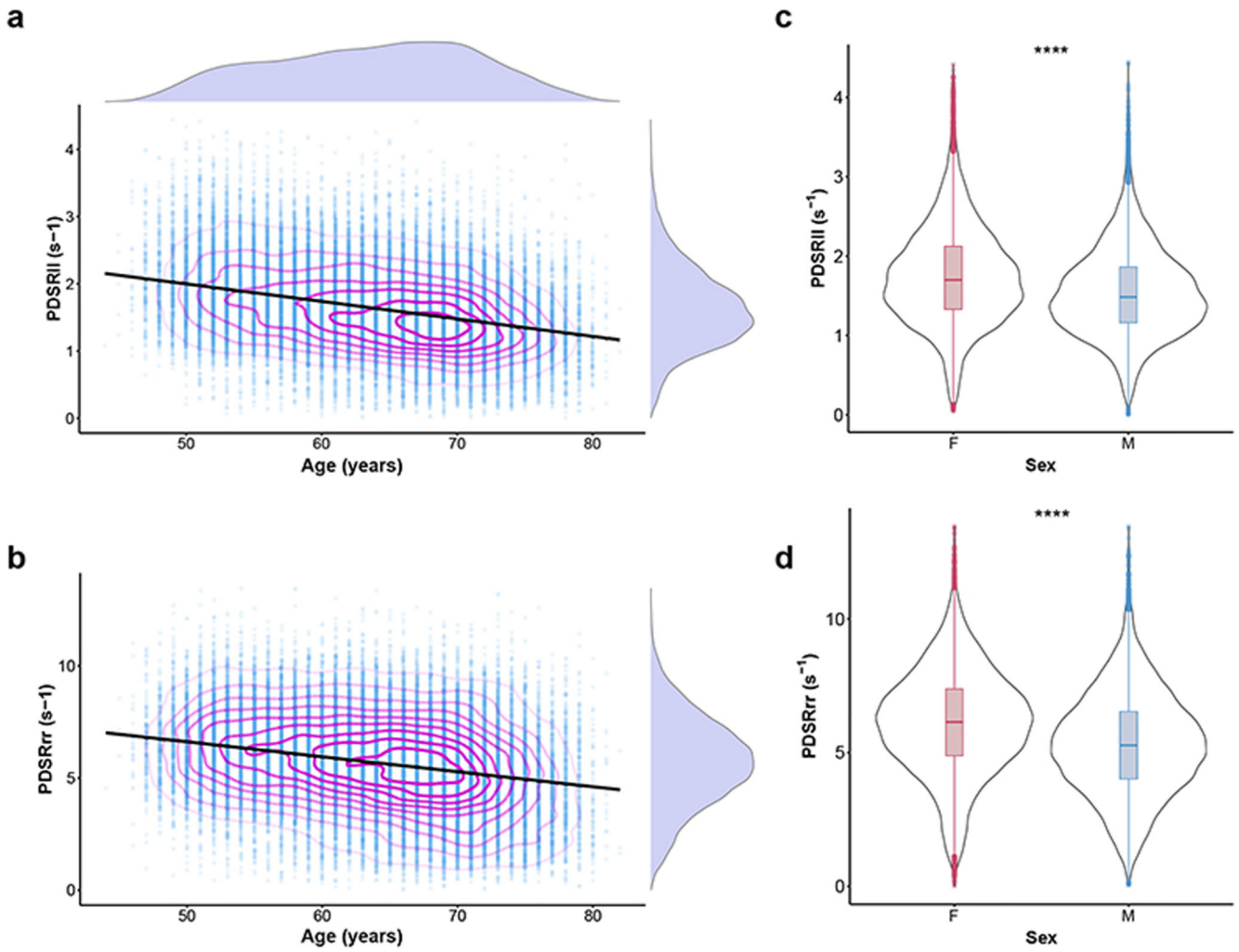


Figure 2. Population strain data.

Scatterplots of a) longitudinal peak diastolic strain rate ($PDSR_{ll}$) (n = 38,923) and b) radial peak diastolic strain rate ($PDSR_{rr}$) with age (n = 38,700); with density contours, linear model fit and marginal density plots. Violin plots of c) longitudinal (n = 38,923) and d) radial (n = 38,700) peak diastolic strain rate with sex; **** $P<10^{-16}$ (Wilcoxon signed-rank test). Boxplots show the median, hinges indicate interquartile ranges (IQR), and whiskers 1.5 × IQR.

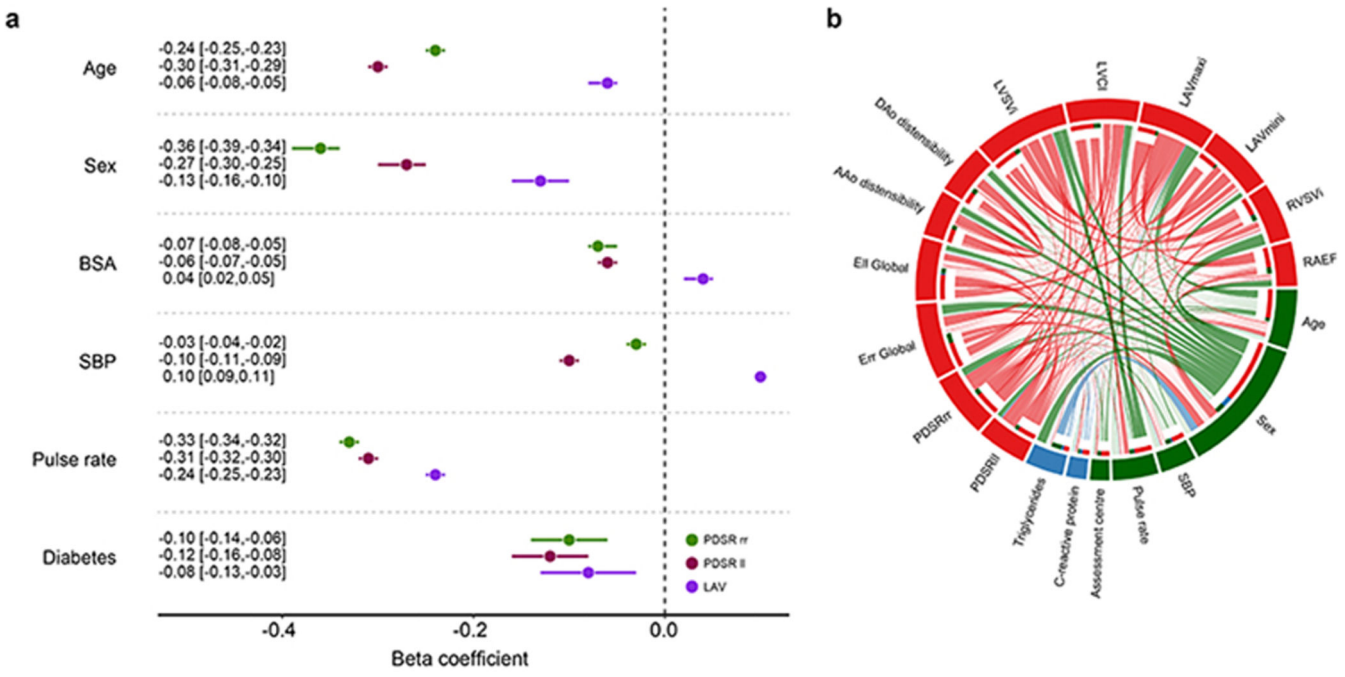


Figure 3. Regression analysis.

a) Multiple linear regression analysis of left ventricular longitudinal peak diastolic strain rate ($PDSR_{ll}$), radial peak diastolic strain rate ($PDSR_{rr}$) and indexed left atrial maximum volume (LAV_{max_i}) with age, sex body surface area (BSA) systolic blood pressure (SBP), pulse rate and diabetes as predictors. All associations were significant after false discovery rate correction. Data are presented as beta coefficient point estimates (95 % confidence intervals). b) Circular plot visualisation of the associations between the imaging (red - $PDSR_{ll}$, $PDSR_{rr}$, global systolic radial strain (E_{rr}), global systolic longitudinal strain (E_{ll}), ascending aortic (AAo) distensibility, descending aortic (DAo) distensibility, indexed left ventricular stroke volume ($LVSV_i$), left ventricular cardiac index (LVCI), LAV_{max_i}, indexed right ventricular stroke volume ($RVSV_i$), and right atrial ejection fraction (RAEF) and the non-imaging phenotypes (green for environmental; blue for biochemical). The strength of the connection between each pair is presented as a ribbon, whose size is proportional to their regression coefficient. All associations with a regression coefficient <0.3 are shown in faint colours (apart from the associations between $PDSR_{ll}$, $PDSR_{rr}$ and LAV_{max_i} and all other phenotypes). The coefficients for the associations of the circular plot are shown in Extended Data Fig 3b. Standardised beta coefficients are shown with units in standard deviations for each variable.

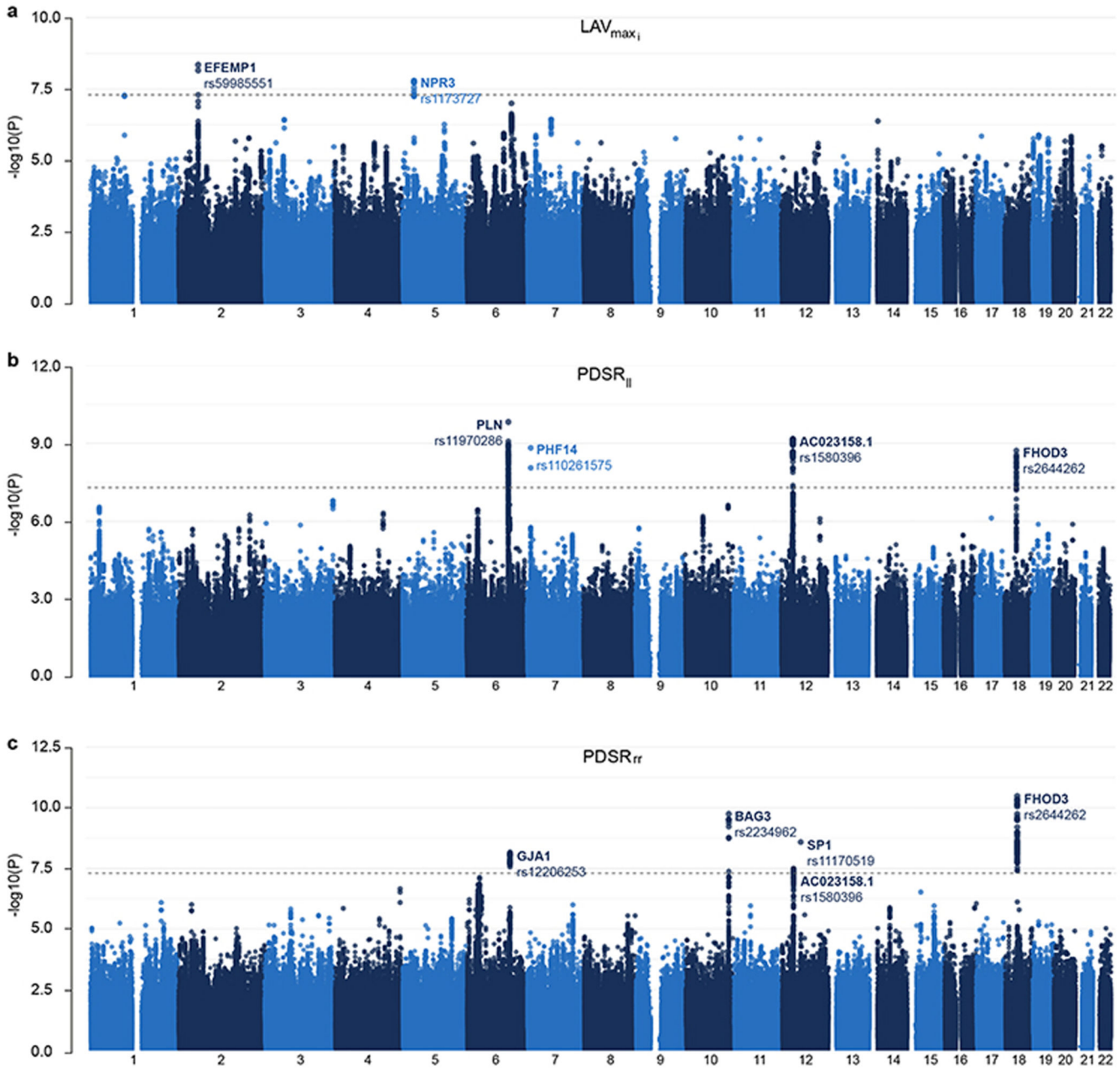


Figure 4. Manhattan plots of the GWAS results for three diastolic function traits

a) indexed left atrial maximum volume (LAV_{max_i}), b) longitudinal peak diastolic strain rate ($PDSR_{ll}$) and c) radial peak diastolic strain rate ($PDSR_{rr}$) (full dataset). This figure shows the -log10(P-value) on the y-axis across all autosomal chromosomal positions (x-axis) from BOLT-LMM. The dotted line indicates genome-wide significance ($P = 5 \cdot 10^{-8}$, $N = 34245$). Significant loci are labeled by their likely causal gene and lead SNP, see Table 1.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

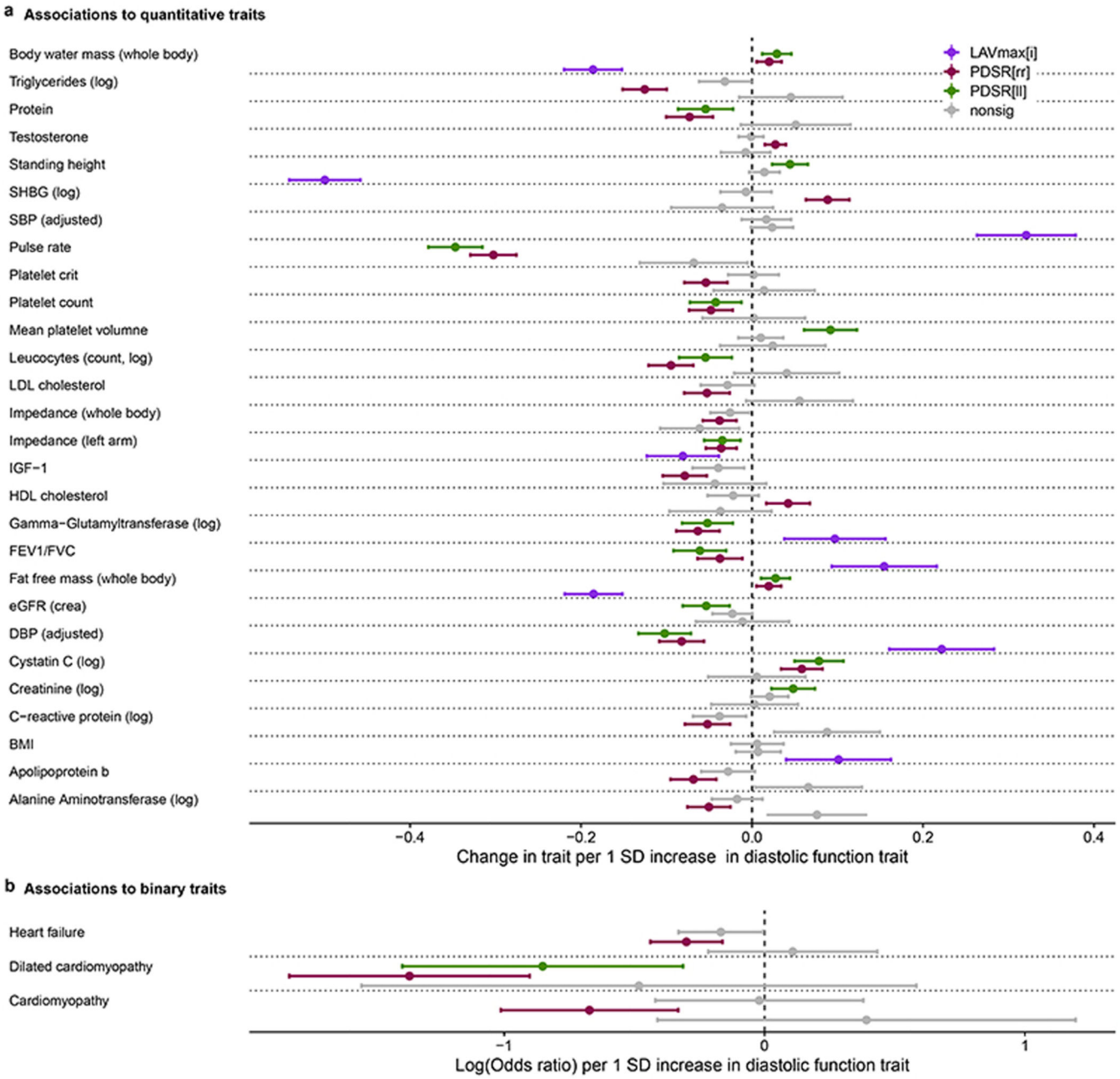


Figure 5. Significant associations of the polygenic instrumental variable scores for diastolic function traits with UK Biobank phenotypes.

a) Quantitative traits that significantly associated with the polygenic instrumental variable scores (PIVSs) of diastolic function (beta coefficient point estimates standardised to change per 1 standard deviation (SD) increase in diastolic function trait with 95 % confidence intervals). b) Binary traits that significantly associated with the PIVSs of diastolic function. Point estimates are Log(Odds ratios) per 1 SD increase in diastolic function trait (95 % confidence intervals). Detailed results, including numerical P-values and 95 % confidence intervals are shown in Supplementary Fig 10. One unit change in the PIVS represents a change of 1 SD in the respective diastolic function trait. All dependent variables (traits)

were standardized representing the change in dependent variable standard deviations for a 1-SD change in the respective measurement. Associations not significant after multiple testing correction (conducted per PIVS) are displayed as grey bars. $PDSR_{ll}$, longitudinal peak diastolic strain rate; $PDSR_{rr}$, radial peak diastolic strain rate; $LAV_{max\,i}$, indexed left atrial maximum volume; SBP, systolic blood pressure; MAP, mean arterial pressure; DBP, diastolic blood pressure; BMI, body mass index; SHBG, sex hormone binding globulin; Log, natural logarithm; SD, standard deviation; nonsig: non-significant. $N = 449{,}263$.

Europe PMC Funders Author Manuscripts

Table 1

Genome-wide Association Results.

Summary information on the lead variants identified from each GWAS analysis and the significant genes from the Loss of Function analysis. For each significant locus across the 3 diastolic phenotypes, variant information, GWAS summary statistics and variant to gene annotation is provided. The evidence column is split by: MS - missense variant; eQTL - colocalisation between the GWAS signal and an eQTL for the gene in a plausible tissue type (see Supplementary Material); M - plausible mechanistic link between the gene and the measured heart phenotypes i.e. the gene function suggests a link to diastolic function; Overall - the confidence of variant to gene mapping given all the available evidence. Loci highlighted in grey are those that reached genome-wide significance in the discovery, validation, and full datasets, loci in white reach suggestive significance in the discovery dataset and genome-wide significance in the full dataset. Further information is provided in the Supplementary Material. (GWAS, genome-wide association study; Chr, Chromosome, Ref, Reference allele; Alt, Alternative allele; MAF, minor allele frequency; $PDSR_{ll}$, longitudinal peak diastolic strain rate; $PDSR_{rr}$, radial peak diastolic strain rate; LAV_{max_i}, indexed left atrial maximum volume; Disc, Discovery; Repl, Replication.)

Lead Variant						GWAS							Annotation		
rsID^{Full}	**Chr**	**Ref**	**Alt**	**MAF**	**Phenotype**	***Estimate*^{Full}**	***SE*^{Full}**	***P*^{Full}**	**Disc**	**Repl**	**Full**	**Locus Genes**	**Closest Gene**	**Likely Causal Gene**	**MS**
rs2234962	10	T	C	0.21		$PDSR_{rr}$	0.1118	0.0175	2.3e-10	Y	Y	Y	*MCMBP, BAG3*	*BAG3*	*BA*
rs2644262	18	T	C	0.28		$PDSR_{rr}$/ $PDSR_{ll}$	0.1087	0.0164	1.7e-11	Y^{rr}	Y	Y	*FHOD3, TPGS2*	*FHOD3*	*FH*
rs11970286	6	C	T	0.45		$PDSR_{ll}$	0.0278	0.0043	1.9e-10	Y	Y	Y	*PLN, CEP85L, SLC35F1*	*PLN*	*PL*
rs1580396	12	C	A	0.46		$PDSR_{rr}$/ $PDSR_{ll}$	0.0807	0.0146	4.1e-8	Y^{rr}	Y	Y	*AC023158.2, AC023158.1, ALG10*	*AC023158.2*	*AC*
rs59985551	2	C	T	0.23		LAV_{max_i}	0.0117	0.0020	5.3e-9	Y	Y	Y	Multiple	*EFEMP1*	*EF*
rs1173727	5	T	C	0.40		LAV_{max_i}	0.0096	0.0017	1.7e-8	N	N	Y	*NPR3, LINC02120*	*LINC02120*	*NP*
rs12206253	6	C	T	0.11		$PDSR_{rr}$	-0.1413	0.0244	8.4e-9	N	N	Y	*HSF2, GJA1, SERINC1*	*GJA1*	*GJ*
rs10261575	7	T	C	0.18		$PDSR_{ll}$	0.0336	0.0056	1.2e-9	N	N	Y	*NDUFA4, PHF14*	*PHF14*	*PH*
rs11170519	12	C	T	0.43		$PDSR_{rr}$	0.0872	0.0146	3.9e-9	N	N	Y	Multiple	*SP1*	*SP*

Predicted Loss of Function Results

	Chr	Carriers	Phenotype	*Estimate*^{Full}	*SE*^{Full}	*P*^{Full}	Cau
	2	187	$PDSR_{rr}$	-0.71	0.14	1.4e-7	*TT*
	6	29	$PDSR_{rr}$	-1.56	0.34	5.6e-6	*LM*